Deciphering the assembly of multi-segment genome complexes in influenza A virus

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat)

eingereicht an der

Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu
Berlin

von

Dipl.-Biophys. Simon Prisner

Präsidentin der Humboldt-Universität zu Berlin:

Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin:

Prof. Dr. Bernhard Grimm

Gutachter:

1. Prof. Dr. Andreas Herrmann
2. Prof. Dr. Matthias Selbach
3. Prof. Dr. Alexander Löwer

Tag der mündlichen Prüfung: 13.07.2017

To boldly go where no man has gone before

— Gene Roddenberry, *Star Trek*

Contents

2 Materials and Methods

Zusammenfassung

Influenza A besitzt ein segmentiertes, acht-strängiges Genom in negativer Orientierung. Die einzelnen Segmente sind in virale Ribonukleoproteinkomplexe (vRNPs) verpackt. Genomische Segmentierung erlaubt es Influenza, zwischen verschiedenen Stämmen Reassortierung zu betreiben, was zur Entstehung von hochgradig virulenten und potentiell pandemischen neuen Stämmen führen kann.

Die Existenz eines Packungsmechanismus wird vermutet, der sicherstellt dass exakt ein Segment jeden Typs in neu knospende Viren verpackt wird. Es gibt Indizien dafür, dass die vRNPs während ihres Wegs vom Nukleus zur Plasmamembran, wo die Knospung stattfindet, Multi-Segment-Komplexe ausbilden, die durch RNA-RNA-Interaktionen, sog. Packungssignale vermittelt werden. Dieser Prozess ist allerdings noch nicht hinreichend verstanden.

In dieser Arbeit wurde eine neue RNA-FISH-Methode namens MuSeq-FISH entwickelt und angewendet, um die spektralen Limitierungen bisheriger Multiplexing-Ansätze zu überwinden und alle vRNA- und mRNA-Spezies vom humanen Stamm A/Panama des Influenza A Virus zu visualisieren. Außerdem wurde ein automatisierter Arbeitsablauf zur Registrierung/Ausrichtung, Punktdetektion, computergestützter Kolokalisationsanalyse und kombinatorischer Analyse der Mikroskopiebilder entwickelt, der auch große Datenmengen verarbeiten kann. Erstmalig wurde damit eine vollständige Kartographierung der Lokalisation und Häufigkeiten alle viralen RNAs in einzelnen Zellen vorgenommen. Aus diesen Daten konnten wir Erkenntnisse zu den Mechanismen und möglichen Hierarchien innerhalb des Packungsprozesses gewinnen. Dazu wurden Reaktionspfade und statistische Analysen von über 60 einzelnen Zellen und mehr als 10^5 einzelner vRNPs herangezogen. Es wurden auch Informationen über die vRNP-Häufigkeiten und deren Unterschiede zwischen Einzelzellen gewonnen, die zeigen dass sich Infektionsumgebungen auch in großer räumlicher Nähe stark unterscheiden und dadurch den Verpackungsmechanismus beeinflussen können. Weiterhin wurde eine Modellierung basierend auf bedingten Wahrscheinlichkeiten genutzt, um Reaktionskonstanten aus statischen FISH-Daten zu erhalten.

Wir haben unsere Analysen zusätzlich auf den aviären Stamm A/Mallard und die reassortanten Stämme A/Pan-M, A/Pan-NS und A/Pan-NSM erweitert, die ein gemischtes Genom aus A/Panama und A/Mallard enthalten. Dabei konnte gezeigt werden, dass sich die Packungsdynamiken und -netzwerke auch zwischen nah verwandten Stämmen erheblich unterscheiden. Heterogene Verpackungsprozesse wurden für diese Stämme observiert, anhand welcher A/Pan-M und A/Pan-NS eher A/Mallard zugeordnet werden konnten.

Ebenfalls wurden erste Schritte unternommen, um die Methode in verschiedener Hinsicht zu erweitern: es zeigte sich, dass MuSeq-FISH und STED-Mikroskopie im Prinzip kombinierbar sind, was auch durch gleichzeitige Detektion von drei vRNA-Segmenten gezeigt werden konnte. MuSeq-FISH wurde auch genutzt, um einzelne Virionen direkt nach deren Eintritt in die Zelle zu färben und auf deren genomischen Inhalt hin zu untersuchen. Dabei fiel auf, dass die Segmente 7 und 8 besonders häufig fehlten, wenn unvollständige Genome detektiert wurden. Außerdem wurde ein Plasmidsystem auf Basis des pHW2000-Vektors für fast alle Segmente von A/Panama umkloniert, welches nun die Expression von mRNA ohne die gleichzeitige Expression von vRNA ermöglicht. In einem ersten Experiment konnte die Funktionalität des Systems gezeigt werden, so dass es potentiell in Transfektionsexperimenten die Untersuchung vom Packungsmechanismus ermöglichen kann, und zwar unter infektionsähnlichen Bedingungen mit beliebig kombinierbaren vRNA-Sets.

Wir erwarten, dass MuSeq-FISH zusammen mit dem automatisierten Arbeitsablauf auch eine nützliche Methode für andere biologische Fragestellungen darstellen wird, besonders wenn es um hochgradig kolokalisierte Untersuchungsobjekte geht. Fundiertes Wissen über den Packungsmechanismus von Influenzaviren kann helfen, die Entstehung von pandemischen Stämmen besser zu verstehen und kann Möglichkeiten aufzeigen, neue antivirale Medikamente zu entwickeln.

Abstract

Influenza A has a segmented genome of eight single-stranded, negative-sense RNAs packed into ribonucleoproteins (vRNPs). This segmentation allows reassortment between different strains with the potential to create highly virulent, pandemic new strains. A packaging mechanism is supposed, ensuring the incorporation of one copy of each segment species into budding virions. En route from the nucleus to budding at the plasma membrane, the vRNPs are thought to form multi-segment complexes via RNA-RNA and RNP-RNP interactions called packaging signals. This process is not yet completely understood.

Here, a new RNA-FISH method (MuSeq-FISH) was introduced to overcome the spectral limits of multiplexing in order to visualize all IAV vRNA and mRNA targets of the human strain A/Panama. An image processing pipeline including image registration, spot detection, automated colocalization analysis and combinatorial analysis was developed, capable of high data throughput. For the first time, a complete map of the localization and abundance of all viral RNAs in individual cells has been generated. This data enabled detailed investigations about the mechanisms and potential hierarchies within the packaging process, which were inferred from pathways and statistical analysis of over 60 individual cells with more than 10^5 vRNP occurrences. We also gained information about the abundance and cell-to-cell heterogeneity of vRNPs among large sets of infected cells, unravelling that infection environments even in neighboring cells differ strongly in segment composition with an impact on packaging. In addition, conditional probability modelling was conducted to infer reaction constants from inherently static FISH data.

We have extended this analysis to the avian strain A/Mallard and the reassortant strains A/Pan-M, A/Pan-NS and A/Pan-NSM, which contain reassorted genomes of A/Panama and A/Mallard. Here we have shown that packaging dynamics and networks differ widely, even among closely related strains. Packaging processes in these strains seemed to be very diverse, however we found A/Pan-M and A/Pan-NS to more closely resemble A/Mallard in terms of packaging.

First steps have been taken to extend the method into different directions: combination of MuSeq-FISH with STED imaging is in principle possible and has been applied for simultaneous detection of three vRNA species. MuSeq-FISH was also applied to single IAV virions directly after cell entry in order to study their genome content, where we found segments 7 and 8 to be lacking most frequently. In addition, a system of pHW2000-based plasmids expressing only mRNA has been created for almost all A/Panama segments. The functionality of this system was shown in a proof of concept, so that its use in transfection experiments can serve as a potential instrument to investigate vRNP packaging in artificial infection-like conditions with reduced vRNAs sets of choice.

MuSeq-FISH together with its image analysis pipeline will be a useful tool also for other biological questions, especially concerning high-grade colocalization. Further understanding of the vRNP packaging in influenza can help us to understand the emergence of pandemic strains and open up paths to new antiviral medication.

1 | Introduction

In this work, the genome packaging of several IAV strains has been investigated and analyzed using RNA-FISH and spinning disk microscopy and subsequent colocalization analysis and modelling approaches. The introduction will therefore cover the biology of the influenza virus (with a special focus on influenza A) and history and recent improvements to in situ hybridization methods.

1.1 Influenza A virus (IAV)

Influenza is an enveloped virus belonging to the *Orthomyxoviridae* family. It is subdivided into three genera, namely influenza A, B and C (IAV, IBV, ICV). IAV and IBV are more similar to each other than both are to ICV. In recent years, claims have been made that there should be a new genus called influenza D (IDV) [1], containing one virus that was found to infect pigs and cattle. In the following, the focus will always be on IAV, and others are only mentioned for the purpose of comparison.

Since the first isolation of a human influenza virus in 1933 [66], it has become more and more apparent that it boasts an enormous capability to evade host immunity and vaccination strategies. Two mutational mechanisms play a role here: (1) antigenic drift – the gradual change of antigens presented to the host, facilitated by the rapid mutational rate of IAV's single stranded RNA genome. (2) antigenic shift – a rapid change of antigens caused by the exchange of genome segments of two or more IAV strains in a process called reassortment. This will be discussed in more detail below.

The social and economic burden caused by annual influenza infections as well as occasional pandemic outbreaks has led to a pronounced research interest. IAV is the genus that by far contributes the most to the number of viral infections [180].

While most human infections are non-lethal, there is a substantial risk when infants, elderly or pregnant people are infected. Nonlethal infections still cause the

victim to be bedridden for days or weeks, which all sums up to a yearly cost of $71-167 billion per year in the US alone and 250 000 - 500 000 deaths worldwide [180]. In pandemic outbreaks, these numbers can easily multiply.

IAV is known for its wide host range, ranging from humans, birds, pigs, horses, ferrets, rats and even to sea lions and bats. However, only waterfowl and possibly bats act as natural reservoirs for the virus [36]. Other possible hosts have emerged via an adaptive process in which the glycoprotein hemagglutinin (HA) is of special importance. Together with the neuraminidase (NA) protein, each IAV strain is characterized by the respective allele of these two antigenic proteins. Currently, there are 16 subtypes of HA and 9 alleles of NA known [49] that are the basis of the well-known nomenclature that is commonly used to classify IAV strains (HxNx). Additional HA alleles 17 and 18 as well as NA alleles 10 and 11 [166] seem to be bat-specific and are therefore termed HA/NA-like alleles.

One main molecular basis of host specificity is the interaction of IAV HA proteins with sialic acids on host cell surfaces. Virions need this interaction for adherence to target cells and it is a prerequisite of a successful infection. Said sialic acids are linked to galactose in an α-2,3- or in an α-2,6-configuration – HA alleles have adapted to preferentially recognize either the α-2,3-configuration (avian IAV) or the α-2,6-configuration (human IAV) [66].

Individual strains of IAV show a highly restricted host range, determined by their respective alleles of HA and NA. Every IAV pandemic in history has been caused by the rise of an antigenically novel strain that overcame the host barrier between other animals and humans: H1N1 for the 1918 Spanish flu, H2N2 for the 1957 Asian flu, H3N2 for the 1968 Hong Kong flu and again H1N1 for the 2009 swine flu. The emergence of such strains is caused by the aforementioned reassortment process, a genomic mixing between at least two different IAV strains (infecting the same cell).

IBV, in contrast, commonly infects humans and seals [132]. Although morphology, genome organization and mutational capabilities are likely to be comparable, no pandemic outbreaks caused by IBV have been reported to date. One reason for this might be the restricted host range.

ICV predominantly infects humans and pigs and the genome is reduced to 7 instead of 8 genomic segments. This reduction stems from the hemagglutinin-esterase-fusion (HEF) protein it possesses, which combines the functionalities of HA and NA found in other influenza genera. ICV infections usually lead to non-lethal upper respiratory illness and sporadically lower respiratory illnesses [165].

IAV virions are typically of spherical or filamentous shape with a diameter of about 100 nm for the spherical type and a length of 300 nm [11] or more for the filamentous type. The filamentous phenotype has been observed more frequently in isolates compared to laboratory strains, suggesting that the common model picture of a spherical IAV virion might actually be an artifact of laboratory research [29]. Both phenotypes are enveloped with a lipid bilayer originating from the plasma membrane of the host cell [152], and are carrying their genomic content in form of viral ribonucleoprotein complexes (vRNPs). The lipid envelope contains the spike proteins HA and NA as well as the ion channel M2 in a ratio of about 40:10:1. The matrix protein M1 forms a layer directly beneath the envelope, which gives support to the vRNPs, themselves being composed of viral RNA (vRNA), the heterotrimeric polymerase complex with its subunits PB2, PB1 and PA as well as numerous copies of the nucleoprotein NP (see Fig. 1). The non-structural proteins NS1 and NS2 are not known to serve a function in complete virions, but have been found there as well [142], maybe just due to diffusion from high concentrations in the cytoplasm.
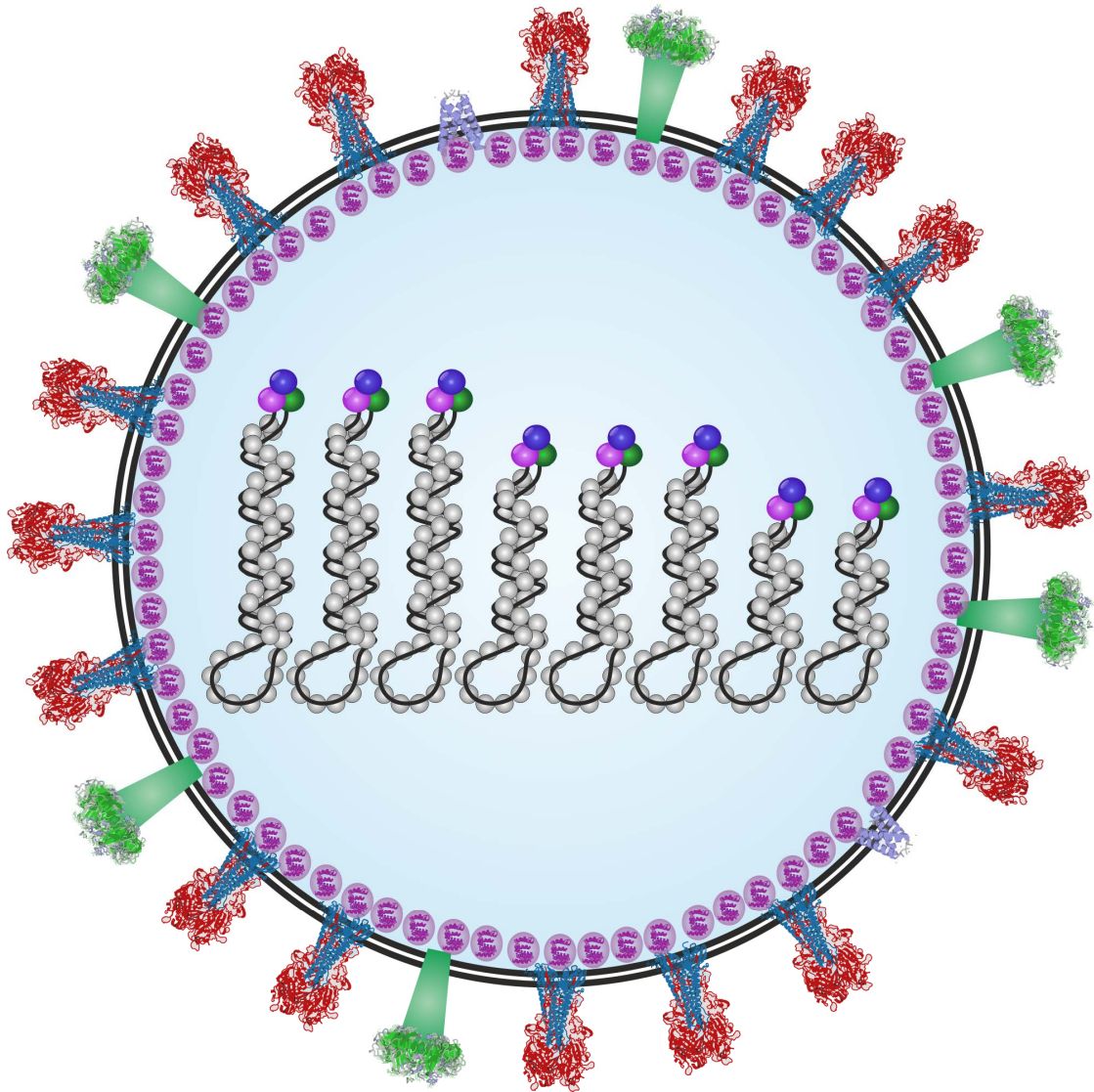
Figure 1 – **Schematic representation of the IAV virion.** The diameter of IAV virions is typically about 100 nm. The viral lipid membrane, which is derived from the host cell plasma membrane, is densely populated with hemagglutinin (HA) trimers and neuraminidase (NA) tetramers and with small amounts of the tetrameric M2 ion channels. The inner lipid monolayer is covered by many copies of M1, which are surrounding the genomic content. The genome is typically composed of one copy of each segment in the form of viral ribonucleoproteins (vRNPs), each consisting of one copy of PB2, PB1 and PA, respectively, as well as multiple copies of NP covering the negative-sense single-stranded RNA. Cartoon representations of HA, NA, M1 and M2 are based on PDB visualization of the respective proteins.

1.2 Genome and proteins of IAV

1.2.1 The IAV genome

IAV comprises a single stranded RNA genome in negative orientation, which is segmented into 8 individual vRNPs. The vRNAs of the segments exhibit different lengths, ranging from 887 to 2341 nucleotides (see Table 1.1) [107]. The genome was first mapped by Ritchey, Palese and Schulmann [143]. These 8 segments are numbered in order of decreasing length. They code for 10 well-studied major IAV proteins among a variety of more recently discovered proteins (reviewed in Vasin et al. [175]), all of which will be described in the respective section below.

RNA species of IAV

At this point it is valuable to introduce the three different kinds of RNA that are synthesized during the IAV life cycle. Viral RNA (vRNA), the cell to cell conveyor of genomic information, is present in negative orientation with respect to messenger RNA (mRNA), which encodes for proteins. In other words, the viral mRNA sequences are each exactly complementary to their respective vRNA coding regions. IAV mRNA, the transcript of vRNA, equals to "regular" cellular mRNA (5'-capped and 3' polyadenylated, although its cap is derived from cellular mRNAs) and is of positive orientation.

For the replication of vRNA, however, an intermediate molecule is needed due to the single-stranded genome. For this reason, in order to produce new negative-sense vRNA from originally negative-sense vRNA, IAV takes the detour of complementary RNA (cRNA) molecules. These are obviously positive-sense, but also become complexed with the polymerase proteins and NP monomers shortly after synthesis [112] and are then called complementary ribonucleoproteins (cRNPs). They are then transcribed "back" to finally yield new vRNPs. The total cRNA level at any given time is relatively low inside an infected cell because they are just transient intermediates [83].

Reconstitution

Notably, the genome of IAV together with a means of transcription and replication is sufficient for the establishment of infection [185]. Virologists have made heavy use of this to generate new viral strains in a process called reverse genetics. Here, one or more plasmids are generated that carry the sequence of one or more proteins, respectively. Binding sites for both polymerase I (PolI promoter) and polymerase II (PolII promoter) of the cellular proteome ensure copying of the sequence in

both opposite directions, yielding mRNA and vRNA. Such reverse genetics systems are also called minireplicon assays and are available for a number of common laboratory strains of IAV. Early versions relied on individual plasmids for each segment, but recent systems reduced this number to yield higher transfection efficiencies. Subsets of the whole system are sufficient in certain cases, as for example vRNP reconstitution, in which only transfection of segments 1-3 and 5 (PB2, PB1, PA and NP) are required for production, but not for export of vRNPs. Reverse genetics can also be used to "rescue" viruses that are deficient in one or more segments, possibly due to attenuating mutations, by transfection of a wild-type copy of the respective segment. In this work, an existing IAV plasmid system for the strain A/Mallard has been modified to express only mRNA.

Advantages of a segmented genome, reassortment only within genus

Genomic segmentation in viruses is common, and influenza is no exception to the trend [69]. The question arises why segmentation would be evolutionary beneficial for a virus. Inevitably, a segmented genome comes with the need to ensure correct packaging of one copy per segment into newly formed virions or, alternatively, with the burden of a low percentage of infectious virions if no such packaging mechanism is in place. In the latter case of so-called random packaging, virions with an incomplete set of genomic segments could only successfully infect a cell in the case of co-infection with one of more other virions that complement the missing segment(s).

However, the huge benefit of segmentation lies in the potential of rapid adaptation and/or antigenic shifts through a mechanism called reassortment. Here, two or more virions of differing strains infect the same host cell, each reproduces within the same cell and is therefore – in principle – able to exchange genomic content in the form of segments. Newly formed virions might be composed of a mixture of any number of segments from virus strain 1 and the remaining segments from virus 2. Reassortment is constantly happening, leading to regular emergence of novel influenza strains. Notably, all IAV pandemics to date stemmed from such reassortant viruses, which helped them overcome the respective host barrier. This process is much more likely to happen in closely related viral strains and is restricted to the respective genus [126]. Thus, IAV and IBV can reassort within their respective genus, but not with each other.

Structure of RNPs

All influenza viruses use structures called viral ribonucleoprotein particles (vRNPs) to pack and protect their vRNA and cRNA, but not mRNA. vRNPs are composed of RNA, the heterotrimeric polymerase complex and numerous copies of NP. The

exact copy number of NP is dependent on the length of the RNA strand, the average coverage density is about 24 nucleotides per copy of NP [89]. A 3-dimensional structure emerges in which the RNA strand is bent and both vRNA 3' and 5' ends partially hybridize and are associated with the heterotrimeric polymerase complex formed by a copy of PB2, PB1 and PA each. Caused by the bending, a so-called "panhandle" structure is visible (see Fig. 2).

Noticeably, and in contrast to some other viruses as for example *Bunyaviridae* [181], the NP proteins are associated with the backbone of the RNA strand, so that the nucleotide sequence is exposed to the outside. Influenza vRNPs are independent entities for both transcription and replication, however they need capped cellular mRNA fragments as primers for transcription and a *trans*-acting RdRp from another vRNP for replication.
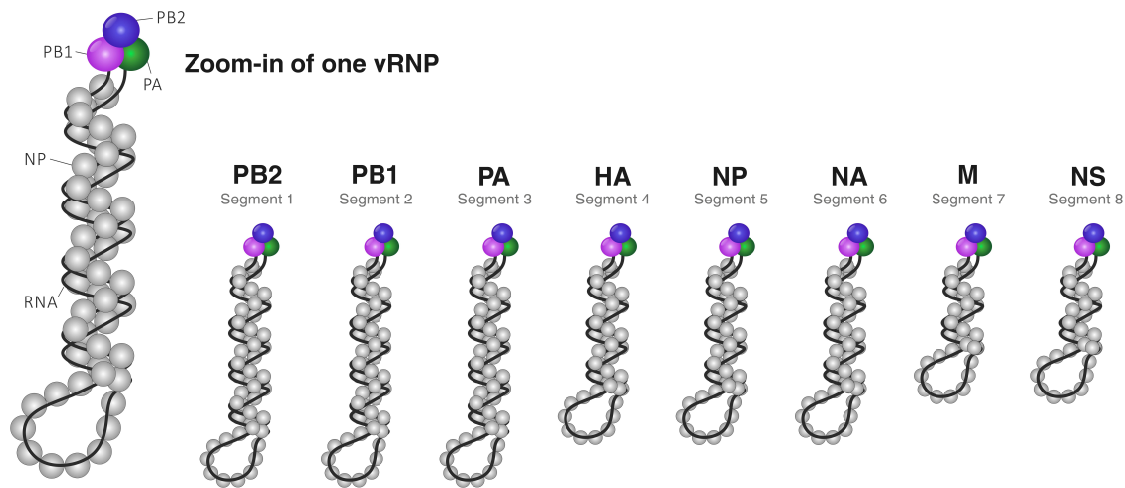
Figure 2 – **Viral ribonucleoproteins (vRNPs) and genomic content of IAV.** The IAV genome is packed into panhandle-structured vRNPs, in which 5' and 3' ends of the RNA strand are bound by the heterotrimeric RNA-dependent RNA polymerase (RdRp). The full length of the RNA is covered by multiple copies of NP. The IAV genome consists of 8 segments, among which PB2, PB1 and PA are treated as segments of "large" size, HA, NP and NA of "intermediate" size and M and NS of "small" size.

Influenza replicates in the nucleus [67], which is rather uncommon for viruses of other genera. The increased complexity of this process seems to be outweighed by several benefits of it: First, the nucleus provides a safe environment for nascent vRNA strands before they are complexed with NP. In the cytoplasm, such a naked RNA strand would be a target to the cytoplasmic RNA receptor retinoic acid inducible gene 1 (RIG-I) [141] and its subsequent antiviral response in the cytoplasm.

Second, the transcription into mRNA uses RNA caps snatched from cellular mRNA which contributes to host shutoff and shifts cellular resources towards viral use. Third, the nuclear environment enables splicing of pre-mRNAs, a process which is used for transcription of segments 7 and 8. Both of them encode two major IAV proteins, the mRNAs of which are untangled by splicing [187].

1.2.2 The proteins

PB2

Polymerase basic subunit 2 (PB2) is the largest protein produced by IAV and part of the viral RNA-dependent RNA polymerase (RdRp), composed of one copy each of PB2, PB1 and PA. The crystal structure of the complete heterotrimeric RdRp was solved in 2014 [138]. It binds cellular mRNAs [10] which are then prone to cap-snatching by PA. So it has been shown to act as a cap-binding protein [9]. It acts in the nucleus and therefore contains several nuclear localization signals (NLS) [117] to ensure nuclear import. Notably, it also seems to be an important virulence factor, as shown by analysis of deleterious mutations on segment 1 which reduced overall viral pathogenicity [65]. It has also been shown that the polymerase can restrict reassortment [94], a process which will be described in more detail below.

PB1

The second subunit of the polymerase is called polymerase basic subunit 1 (PB1) and is the main protein encoded on segment 2. With its catalytic activity, it represents the actual polymerase enzyme within the RdRp [12]. PB1 contains four typical and conserved motifs also found in cellular RNA-dependent RNA polymerases. It has also been described to be a mediator between the other two parts of the RdRp, PB2 and PA, as it contains binding structures for both of them [167]. As with all proteins that need to enter the nucleus, it contains several NLS [122].

PB1-F2 is a more recently discovered protein [26] on segment 2 which is translated from an ORF shifted back by one nucleotide relative to the main PB1 protein. It serves as a virulence factor and is mainly localized in mitochondria but has also been found in the nucleus. It has been detected to promote apoptotic and inflammatory responses of the host cell [190].

PB1-N40, another protein originating from segment 2, essentially shares the amino acid sequence of PB1 but lacks the first 39 amino acids. It is still unclear what the exact functions of this protein are, but mutational studies have shown that

polymerase activity and overall replication kinetics decreased without PB1-N40 [163]. Both novel segment 2 proteins, PB1-F2 and PB1-N40, are restricted to a minority of IAV strains.

PA

Segment 3 mainly encodes the polymerase acidic (PA) subunit of the viral RdRp. It is the smallest subunit of the heterotrimer and was found to be phosphorylated [151]. It is required for polymerase activity and besides that, also offers ATP-binding and helicase activity [37]. A major role lies in the cap-snatching process of cellular mRNAs to generate primers for production of capped viral mRNA. This endonuclease activity had been thought to be mediated by the other subunits until rather recently, when it was confirmed for PA by two groups [39, 189]. Since it is required for assembly of vRNPs taking place exclusively in the nucleus, it is not surprising that it contains several NLS [128].

PA-X is one of the non-canonical protein products of segment 3. It contains the same endonuclease domain as the PA protein, combined with a so-called "X domain" which is encoded by a shifted ORF (+1 nucleotide relative to PA). PA-X was found for all IAV strains for which the occurrence of this protein has been studied and was suspected to play a role in host-cell shutoff, such that it enabled prolonged translation after infection. It is not essential for infection, although a lack of PA-X causes diminished virulence [75]. In addition, there are 2 shorter variants of the PA protein, called PA-N155 and PA-N182, respectively. They share the amino acid sequence of PA, however only starting from residue 155 or 182. Both proteins' functions are unknown, although any polymerase activity could be excluded [120].

HA

Hemagglutinin (HA) is a protein that forms homotrimers on the surface of IAV virions and contains a fusion peptide. It is required for the attachment of virion particles to the host cell membrane [91] via its interaction with sialic acids of glycoproteins and glycolipids. Another function lies in the fusion of endocytosed viruses with the endosomal membrane [127] to allow release of the vRNPs into the cytosol. After production of the precursor HA (termed HA0), it is later extracellularly cleaved into two subunits (termed HA1 and HA2) which are held together via disulfide bonds. The mentioned fusion peptide is only functional after this cleavage process. The trigger pH value that has to be undercut to activate fusion with the endosomal membrane [157] has been shown to differ between different viral strains and is believed to play a role in host cell adaption, as different

host cells exhibit different endosomal pH environments [100].

NP

The nucleoprotein (NP) has a size of 56 kDa and protects vRNA from cellular nuclease digestion. To this aim, it covers the vRNA strand with a ratio of one NP per roughly 24 nucleotides [89]. It is therefore the most abundant protein on vRNPs and often used as a proxy for their detection in e.g. SDS gels. NP carries a net positive charge which is the basis of electrostatic attraction to the negatively charged RNA backbone [84]. For this reason, the nucleotide bases are uncovered and accessible by the RdRp for transcription and replication [6] and potentially also for vRNA-vRNA interactions. NP is monomeric if not complexed with vRNA and exhibits multiple NLS for nuclear import [124]. Subsequent to nuclear entry, it multimerizes upon vRNA interaction, which effectively hides the NLS to prevent re-import. Recently, NP has also been shown to influence IAV genome packaging [44], in a way that amino acids from a IAV-like bat virus abolished genome packaging when introduced into otherwise unaltered IAV.

NA

The neuraminidase (NA) is a 60 kDa exosialidase [156, 174], which means that it is responsible for cleaving sialic acids from the cell during the budding process of newly formed virions. This solves the problem of all budding virions: the virion contains HA proteins, affine for sialic acids on host cells; yet the cell from which it buds exhibits the same sialic acids which would otherwise lead to binding of budded viruses to the parental cell. Nine different subtypes of NA are currently known and characterized for IAV; as well as one subtype for IBV and ICV (as HEF), respectively. NA is present on the plasma membrane of infected cells in homotetrameric complexes. Together with HA, both proteins are also nicknamed "spike proteins" because of their shapes. NA is a drug target of the substances zanamivir and oseltamivir (Tamiflu^{TM}) which inhibit NA activity after budding and thus, viral release after budding [115].

M1/2

M1 is the matrix protein (28 kDa in size) of influenza and forms a layer beneath the lipid membrane of IAV virions [149]. It is imported into the nucleus after its production due to an NLS. Through SUMOylation [184] it becomes affine to NP and thereby associates with vRNPs there [186]. M1 also plays a role in vRNP nuclear export and, moreover, prevents vRNPs to be re-imported into the nucleus [17]. Depending on the viral strain, vRNPs might be covered with M1 most of their life span, from shortly after production until endosomal release [18].

M2 is an ion channel for protons [30] and, as a monomer, represents the smallest protein of IAV with a mass of 11 kDa. However, its functional form is a tetramer, first in the host cell membrane, later in the viral membrane. Upon endocytosis, the pH is acidified and M2 forwards this effect to the inside of the virion, where it leads to vRNP release from M1 [105]. The amino terminus [135] of M2 steers incorporation into budding virions, while the carboxyl terminus [19] promotes viral replication. It is a target of amantadine and rimantadine [61], two of the few small molecules that act as drugs against an influenza infection.

M42 is a variant of the M2 protein where the ectodomain is slightly varied. It is produced through splicing using the same effector, but a different donor site compared to M2. M42 seems to be able to complement M2's function, even though its localization seems to be more directed to the cis-Golgi-network. Only a small subset of tested IAV strains showed expression of M42, so it cannot be essential [183].

NS1/2

The NS1 protein is usually not found in IAV virions, hence the name "non-structural protein 1" [87]. It is responsible for viral defense against host cell and consists of an N-terminal RNA binding domain and a C-terminal effector domain [62]. By interacting with host cell protein kinase R (PKR) [111] and other proteins it suppresses their ability to detect infection and to start the interferon response [40]. As a consequence, cellular counter-measures are attenuated.

NS2, which is also known as the nuclear export protein (NES) referring to its main function, is a small protein of only 14 kDa [178]. Due to this small size, it can be imported into the nucleus without the use of the importin machinery. It drives export of newly synthesized vRNPs into the cytosol through its interaction with the exportin complex Crm1 [125].

NS3 is an isoform of the NS1 protein in which splicing is activated due to an A374G mutation on segment 8. Three beta-strands are spliced out, resulting in a truncated version of NS1. At present, the function of NS3 is unclear – in fact, it might as well be an artifact of host adaptation of human IAV in mice and was only found for a small number of tested strains [153].

Table 1.1 – Overview of IAV A/Panama segments and proteins.

Segment number	Length [nt]	Protein(s)	Size [kDa]
1	2342	PB2	87
2	2342	PB1￿PB1-F2￿PB1-N40	84
3	2233	PA￿PA-X￿PA-N155￿PA-N182	83
4	1762	HA	63
5	1566	NP	56
6	1466	NA	50
7	1027	M1￿M2￿M42	28￿11
8	887	NS1￿NS2￿NS3	27￿14

1.3 Life cycle of IAV from the angle of a vRNP

The following will be a short overview of the influenza A life cycle. Usually, the life cycle is described from the angle of the infected cell, starting with endocytosis of a virion and ending with the budding of newly formed virions. However, as the vRNPs are of special importance in this work and as they form the minimal replication unit (as confirmed by experiments with rescued viruses) they shall be put in the center of this section [71]. So this life cycle will start with the replication of vRNPs in the nucleus and end with their entry into the nucleus of another infected cell.

1.3.1 mRNA transcription

Transcription of IAV mRNAs is accomplished in the nucleus of an infected cell, where vRNPs originating from an IAV virion have been imported. First, the virus has to establish its proteins, so transcription is preferred over replication in the early stages of infection. Transcription into complementary (positive-sense) mRNA copies of the vRNA segments is performed by the viral RdRp which was transported into the nucleus associated with the incoming vRNPs. However, primers are still required for mRNA transcription. In a very slick manner, IAV uses 5' caps from cellular pre-mRNAs, which are cleaved from their original mRNAs in a process called cap-snatching [88] – this process also contributes to host shutoff [177]. The viral RdRp is capable of cleaving about 15 nt from these pre-mRNAs and uses the short oligonucleotide to start transcription of viral mRNA. Transcription ends at the polyadenylation signal, which leads the RdRp to synthesize repeated copies of uridine at the 5' end of the template. It has been noted that the transcription process happens in *cis* [80], meaning that the resident RdRp – the one associated with the parental vRNP – is in charge of transcription and no other vRNP is required.

1.3.2 mRNA export

The resulting mRNAs are then exported from the nucleus and use the cellular translation machinery [112], i.e. cellular ribosomes, for translation of proteins. Thereby protein levels increase over time of infection [99].

1.3.3 Replication

The virus now shifts its activity in the nucleus towards replication instead of transcription. It is still not completely clear how this process is governed and how the underlying kinetics look like. However, it is known that replication kicks in

after transcription has already started and that there might be factors mediating this "switch" [80, 136]. One factor that might play a role is the production of small viral RNAs (svRNAs) [171, 137] as 5'-end by-products of transcription. These svRNAs are thought to accumulate in the cell over the time course of infection and, by allosteric modification of the RdRp, shift RdRp action towards replication. In addition, replication happens in *trans* [47], meaning that the RdRp of a neighboring vRNP is in charge here. This obviously requires a certain concentration of vRNPs in the nucleus and might be another factor for the switch.

During replication, an intermediate (positive-sense) copy is generated which is called cRNA. cRNA shows sequence identity to mRNA but carries the whole sequence, including the complete 5' and 3'-ends. It becomes associated with NP copies and a single RdRp as is the case for vRNA and is then called a cRNP [64]. Such cRNPs are then again copied by an RdRp *in trans* to yield the final (negative-sense) vRNP.

It has been observed that the replication machinery of IAV is lacking a precise proof-reading mechanism. Quite frequently, the RdRp takes a short cut from its RNA strand to the adjacent returning strand, which is possible due to the intertwined secondary structure of vRNPs. In such a case, portions of the sequence including the panhandle loop are omitted from replication, which leads to incomplete, or defective interfering RNAs (DI RNAs) [123, 78]. Such DI RNAs exhibit similar properties as vRNA downstream and are even capable of being packed into newly formed virions. Due to their truncated sequence they will not be able to code for functional proteins and will therefore render such virions non-infectious.

1.3.4 Protein import

Because IAV replicates in the nucleus, there is a need for most of its proteins to be transported there. PB2, PB1 and PA as polymerase subunits as well as NP are crucial for vRNP formation. NS2 as the nuclear export protein and M1 as a mediator between vRNPs and the export machinery are required as well. All these proteins therefore carry at least one nuclear localization signal (NLS) to be imported through the nuclear pore complex (NPC). Cellular importins bind to the NLS and facilitate entry, after which they bind Ran-GTP and release the cargo of viral proteins [54]. It has been shown that PA and PB1 are imported as a heterodimer [38] but separate of PB2.

1.3.5   Nuclear export

It has been observed that nuclear vRNPs tend to associate with chromatin as well as the chromatin-related factors HMGB1 [114] and PARP-1 [106]. The benefit of this association [22] could be spatial proximity to the protein partners needed for vRNP export, namely the cellular export receptor Crm1 [125] and Rcc1. A widely discussed, but also disputed model is the **daisy-chain model** [16] in which vRNPs bind to Crm1 via action of M1 [105] and NEP [131] in a multi-protein-complex.

Export of vRNPs gradually increases through the time course of infection which is caused by a number of factors: increasing HA levels on the membrane promotes vRNP export through activation of the MAPK cascade [101], activation of apoptosis in the infected cell increases the NPC diffusion limit [45] and NS2/NEP levels themselves rise only gradually over time because NS1 is preferred in the respective splicing process [144]. Even though NP contains nuclear localization signals (NLS) which would promote re-import of recently exported vRNPs, this is prevented by a covering layer of M1 [70] which is suspected to shield the respective NLS.

1.3.6   Transport of vRNPs from the nucleus to the plasma membrane

After the vRNPs have been exported from the nucleus, they are frequently detected in close proximity to the microtubule organizing center (MTOC) by immunoflu-orescence [42]. A potential candidate for mediation of this interaction might be Y-box binding protein 1 with affinity towards both vRNPs and microtubules [81]. High vRNP density near the MTOC might also be caused by a high number of recycling endosomes (REs) found in this region, which are vesicles travelling towards the plasma membrane for recycling of proteins and lipids from inwards travelling endosomes. Such REs are specified by carrying copies of the GTPase Rab11 [173]. Rab11 exists in two isoforms, A and B, of which A has been shown to interact with the Rab11 family interacting protein 3 (FIP3) [15, 3] which in turn interacts with PB2 of the IAV RdRp [2, 42]. This mechanism allows vRNPs to "hitch-hike" on REs on their journey towards the plasma membrane. REs typically move along microtubules, but are also able to use actin filaments as a support.

1.3.7   Segment bundling and packaging

This journey on REs was speculated to be key to vRNP packaging. Ultimately, vRNPs of different genome segments have to mingle to form bundles of eight distinct segments comprising the genome of a single virion. Evidence has been emerging that this is a regulated process, and the confined space on REs might

be an ideal ground for vRNP-vRNP interactions to take place. Supporting this, the visualization of IAV genome segments using fluorescence *in situ* hybridization (FISH) has indicated a higher order of complexation with growing distance to the nucleus [90], and that vRNPs occur monomerically in the nucleus [28]. However, a detailed introduction to packaging and bundling will be given below, while for the purpose of the life cycle we will just assume that bundles of vRNPs emerge during RE transport. This mode of transport does not allow the vRNP bundles to directly reach the plasma membrane – the final migratory step after release from Rab11 is possibly taken by simple diffusion [155].

1.3.8   Virus assembly and budding

HA, NA and M2 are the proteins which do not enter the nucleus and instead are directed to the plasma membrane. They are integral membrane proteins both in infected cells as well as in budding virions. Some evidence suggests that HA and NA are preferentially found in so-called lipid rafts [14] while M2 is enriched not within, but on the edges of such rafts [23]. This creates areas on the plasma membrane called *budozones* [82], which are areas of abundant viral proteins and high rates of virion budding. M1 has been shown to interact with all three proteins and also the plasma membrane itself, so it might again serve as an adaptor between vRNPs and viral membrane proteins [68]. Here, packaging signals (in the narrower sense, see below for more details) might come into play and serve as some kind of check point for budding, which is only passed if the arriving genome bundle is complete [133].

Viral budding is physically started by formation of a protuberance of the plasma membrane. In non-physiological conditions, the membrane proteins together and alone can induce curvatures that lead to formation of virus-like particles (VLPs) [23]. So it might be a complex interplay of various factors that promotes budding of virus particles in infected cells, with most of them carrying a cargo of exactly 8 vRNPs. The precise workings of the budding process are still poorly understood.

However, it is known that the final step of budding – membrane abscission – in which the viral membrane is separated from the remaining plasma membrane, is governed by M2 [147]. The timing of this step determines the shape [146] of the resulting virions: the typical spherical shape of most laboratory strains results from abscission directly after the vRNPs are packed, while delayed abscission leads to the filamentous phenotype found in many clinical isolates of IAV.

1.3.9   Virion state

After budding has completed with abscission, a new influenza virion has emerged as an independent infectious unit, given that it contains the complete genome set. The copies of NA on its surface will now use their cleaving activity to cleave sialic acids of the host cell to prevent re-attachment through action of HA. The result is a free-floating virion of spherical or filamentous shape, with a diameter of about 100 nm. The membrane is covered with several HA and NA copies and in comparison only few M2 ion channels. Directly below the membrane, a layer of M1 proteins has been formed and in the core of the virion, the vRNPs are aligned parallel to each other in an ordered bundle. Despite their names, the non-structural proteins NS1 and NS2 (NEP) were also found in low quantities inside the virion. A classical life cycle of IAV infection would end at this point. However, in the angle of the vRNP, only roughly half of its life is over yet.

1.3.10   Binding and cellular entry

Released virions passively diffuse away from their cell of origin and eventually attach to a new, sialic acid exposing cell. Binding is established multivalently by many simultaneous HA trimers binding to sialic acids on glycoproteins on the cell surface. The virus particle is then taken up either through clathrin-dependent endocytosis [24] or (especially in the case of filamentous viruses) macropinocytosis [145]. The virion as a whole ends up in an early endosome, so at this stage two membranes surround the vRNPs. Characteristic of the transition of early to late endosomes (a process called maturation) is the decrease in pH down to about 5.2 [188].

This acidification is propagated through the viral membrane by action of the ion channel M2. Low pH starts a whole program of fusion for IAV, of which the major conformational change of HA is the most important. HA now reveals a fusion peptide which penetrates the endosomal membrane and invokes fusion with the viral membrane [100]. Inside the virion, vRNPs have already detached from the membrane due to reduced affinity to M1 at low pH [17]. As a result, the genomic cargo is released into the cytoplasm of the cell.

1.3.11   Diffusion through cytoplasm

Contrary to the "hitch-hiking" vRNPs on their way to the plasma membrane, incoming vRNPs do not seem to make use of a cellular transportation system during entry. After release from endosomes, the remaining path towards the nucleus seems to be covered by pure diffusion, which is typically completed within the first 30 min of infection [105]. Nuclear localization signals on NP are now

responsible for nuclear import of the vRNPs, mediated by alpha- and beta-importins through the NPC. Using FISH, it has been shown that the segments remain in their bundles before nuclear import, but then dissociate during or directly after they enter the nucleus [28]. This split into monomeric vRNPs representing basic replication units brings the life cycle to an end, or rather to a repeat. Production of new viral mRNA and eventually vRNA begins anew at this point.
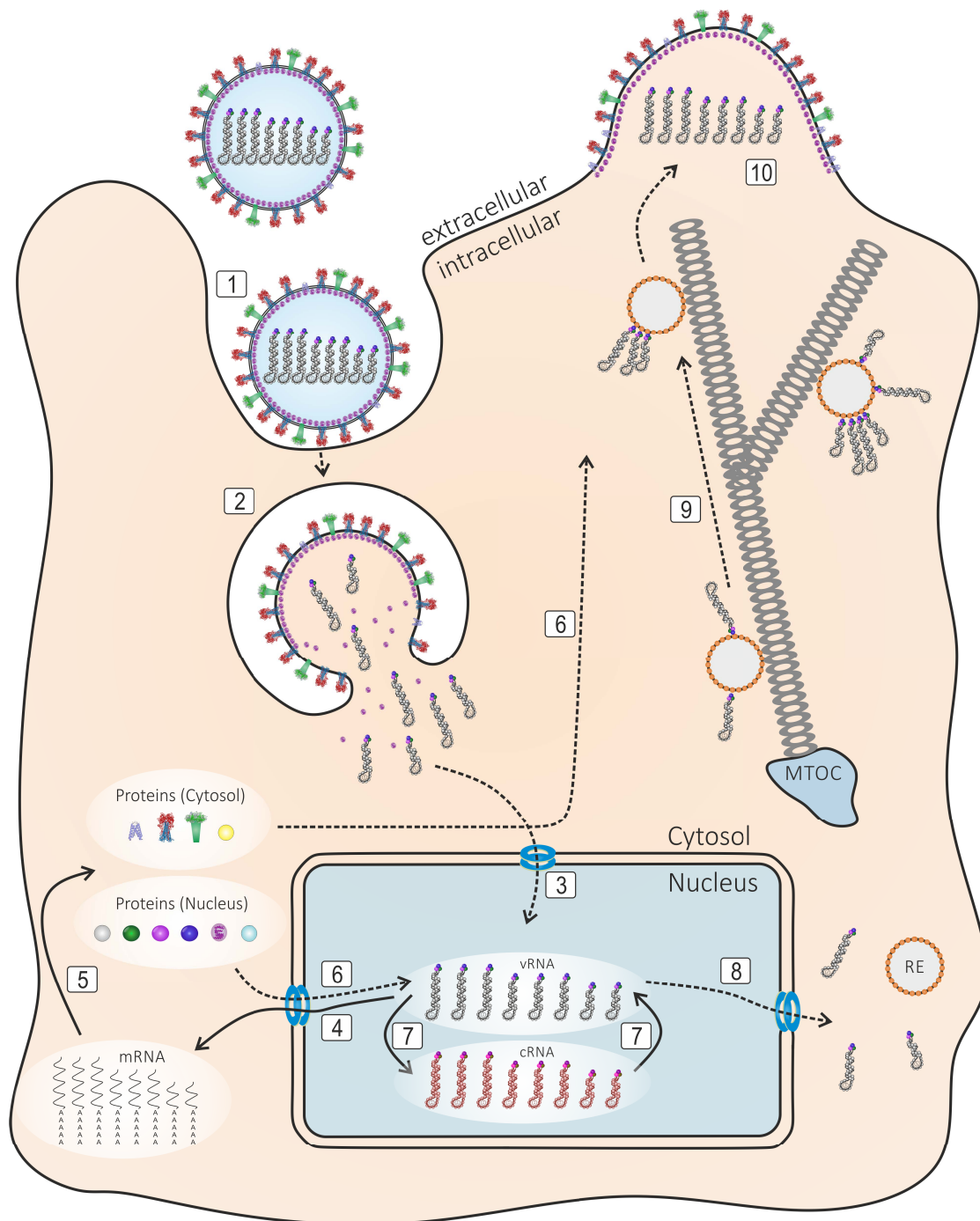
Figure 3 – **Life cycle of IAV.** Cartoon representation of the IAV life cycle. (1) Virus entry into host cell via endocytosis. (2) Fusion of virion with endosomal membrane; release of vRNPs. (3) Nuclear import of vRNPs. (4) Production and export of viral mRNA. (5) Protein translation. (6) Nuclear import of RdRp proteins, NP, M1 and NS2; diffusion and transport to the plasma membrane for other proteins. (7) Replication of vRNPs through cRNP intermediates. (8) Nuclear export of vRNPs. (9) vRNPs transported to the plasma membrane, hitch-hiking on Rab11-recycling endosomes and potentially interacting for genome bundling. (10) Budding and scission of newly formed virions.

1.4 Genome packaging models

1.4.1 Problems of packaging overall, other viruses

The influenza genome is split into 8 (IAV, IBV) or 7 (ICV) independent segments. Each segment replicates as a single unit and there is no covalent binding of segments to each other. This segmentation complicates the viral life cycle and production of new virions, but these challenges seem to be outweighed by the benefits of rapid genetic adaptation through reassortment. Reassortment is the exchange of one or more foreign segments with the genome of a virus, similar to crossing-over events in human reproduction. The reassortment partners might be mutants from the same original strain, but also less related viruses that by coincidence co-infected the same host and cell. Reassortment is known across a variety of viruses [21], and IAV is neither at the top or bottom of the list when it comes to the amount of segments (reviewed in McDonald et al. [108]).

IAV virions provide spatial capacity for about 8 segments to be incorporated. However, this does not necessarily infer that these are 8 unique segments. Several virus families employ a mechanism called *multipartite packaging* (reviewed in Vijaykrishna et al. [176]) where the contents of an individual virion are completely random and multiple infections of the same cell ensure that in total, each segment is present at least once. However, such a random packaging is generally favored in viruses with a low number of segments; mathematics show that with many segments, the number of virions needed for successful infection increases rapidly. In such cases, it might be of evolutionary advantage for a virus to establish an ordered or specific mechanism for packaging in which specific interactions between the genome segments ensure or facilitate packaging of different segments into a single virion.

1.4.2 Packaging signals vs. bundling signals

There has been a long standing debate in the literature about which of the strategies outlined above is the one IAV follows. Evidence to support a specific packaging model has been growing especially in recent years. Numerous regions on the RNA strands – so-called *packaging signals* [53] – have been identified which are thought to mediate vRNP-vRNP interactions [58, 43, 104, 4]. However, it is still unknown how packaging would work in detail and just how specific it is. For example, one could think of a model where a certain backbone of interactions is specific and rigid, but which allows for flexibility at the edges as well.

It has also been pointed out that there might be a distinction between bundling

signals and actual packaging signals. Bundling signals would be parts of the RNA sequence of a segment which are important for a specific interaction with one or more other segments. The concerted action of all bundling signals leads to formation of a vRNP bundle. In the narrower sense, packaging signals now only refer to such parts of the RNA sequence that mediate an interaction of this vRNP of an associated bundle with the budozone at the plasma membrane to establish incorporation of this bundle into a budding virion.

1.4.3   Identification of packaging signals

Bundling signals have been identified for a variety of IAV strains using methods like bandshift experiments or virus titrations after introduction of mutations or deletions (reviewed in Giese et al. [60], Isel et al. [74]). Generally, these signals are concentrated towards the 5' and 3' ends of the sequences and occur often, but not exclusively in untranslated regions [33]. The nature of the bundling interactions is not completely understood, but mediator protein(s) have been virtually ruled out by proteomic investigation of virions [164]. Instead, they are suspected to work via direct RNA-RNA-interactions [58], possibly on the level of secondary structures. Transfection of viral plasmids with a reporter segment [179, 52], carrying a fluorescent protein instead of the genomic sequence of the respective segment, but with remaining packaging signals allowed pinpointing their regions along the segment. Another approach was the use of competitive reverse genetics by transfecting 9 plasmids in total with one segment transfected twice using plasmids coding for different viral strains [31]. Using this, it was possible to assess the proportion of produced virions that carried the foreign version of this segment, so that high proportions were associated with compatible packaging signals. Obviously these approaches were also combined with mutational analysis on the transfected plasmids [58].

All in all, the available evidence is plentiful, but originates from the use of different methods and viral strains, complicating comparisons. So far, no consensus pathway has been obtained, leading to the speculation that either (a) packaging pathways are rather redundant within each viral strain or (b) packaging pathways differ strongly between viral strains, or (C) even a mixture of both.

Table 1.2, lists the interactions between segments as described in the literature. Most research has been focused on the well-established A/PR8 and A/WSN laboratory strains. Very extensive data on binary interactions was generated by Fournier and Gavazzi on the strains Finch and England, however using band-shift assays on naked vRNA so that the information contained in the quarternary structure of vRNPs was not covered.

Table 1.2 – Overview of IAV packaging signals described in literature.

Observation	Comment	Virus strain	Reference
S1↔S7, S2↔S6, S3↔S4, S3↔S6, S4↔S7, S4↔S8, S5↔S8, S6↔S7	vRNA↔vRNA interactions in vitro using naked vRNA	A/Moscow/10/99 (H3N2)	Fournier et al. [50]
S1↔S2, S1↔S6, S2↔S3, S2↔S4, S2↔S6, S2↔S8, S3↔S5, S3↔S8, S4↔S5, S4↔S7, S4↔S8, S5↔S7, S5↔S8	vRNA↔vRNA interactions in vitro using naked vRNA	A/Finch/England/ 2051/91(H5N2)	Gavazzi et al. [58]
S1, S3, S5 and S7 essential for packaging	S2, S4, S6 and S8 are recruited	A/PR/8/34 with ICV HEF segment	Gao et al. [57]
S1-S3 affect all segments except S4 and S6	Signals in conserved CRs, especially S1 crucial for virus production	A/WSN/33	Liang et al. [97, 96], Muramoto et al. [119], Marsh et al. [103], Dos Santos Afonso et al. [41]
S2 and S8 interact, disruption leads to empty viruses	Signal in CR, partially conserved in avian H5N2 but not in human H1N1 and H3N2	A/Finch/England	Gavazzi et al. [58]
S3 affects S5 incorporation and vice versa	Signal in conserved S3 CR Signal in S5 codons 464-466	A/PR/8/34	Marsh et al. [103], Hutchinson et al. [73]
S4 affects all segments except S1	S3, S5, S6 and especially S2 affected, all segments except S1 affected	A/WSN/33, A/PR/8/34	Marsh et al. [102]
S7	Signals in CR influenced other segments and virus production	A/PR/8/34	Hutchinson et al. [72]
S6 does not affect packaging of other segments	Virus production enhanced if all segments were present	A/PR/8/34 with modified ICV HEF	Gao et al. [56]

1.4.4 How many vRNPs are packed into a virion?

Electron microscopy and even tomography have further increased our knowledge about the genome composition of complete IAV virions [161, 48]. Usually, these particles have been imaged after budding, when still attached to their former host cell. Here, numerous studies have confirmed that the vRNPs are aligned parallel to each other and in a highly ordered "7+1" arrangement [129, 121], where a central vRNP is surrounded by 7 others. Restricted by the method, it was not possible to identify the exact identity of the segments, but using their lengths as proxies some information was gained nevertheless [129]. The central segment was consistently found to be one of the long ones and some other conformations were supposedly forbidden. However, the order of the segments was never exactly the same in two virions and seems to be at least partly flexible.

Partially contradicting these findings of completely packed virions, virological work with infections using a low MOI found only a small share of virions infectious on their own, as measured by complete protein expression [13]. However, one has to keep in mind that for a successful infection it is necessary, but not sufficient to have a complete genome set. Further steps that need to be taken are successful endosomal escape, import of all 8 vRNPs into the nucleus and genomic stability of the complete genome, all of which are quite independent of the packaging process. DI RNAs might also contribute to the observed effect.

Building on the data mentioned above, a *hierarchical packaging model* was proposed [119, 51, 50]. Here, some segments might be more important in the packaging process than others – clearly a central segment (if there is a clear central segment) plays a crucial role. Then there might be intermediate segments directly interacting with the central segment and dead end segments interacting with these intermediates. The dead ends would then be segments that could reassort quite easily, as no other segments rely on their bundling signals. Intermediate segments might reassort in blocks together with their dependent dead end segments.

1.5 Single molecule fluorescence *in situ* hybridization (smFISH)

1.5.1 Historical overview

Ever since the establishment of histology and antibody stainings, the identification and – to a certain degree – information about abundance have become possible for cellular structures and proteins. For nucleic acids, it was traditionally more

complicated to yield similar results. This has been changing since the onset of the new century and will be briefly covered in this section.

The use of antibodies has already been established in the 1940s [93], and special antibodies have also been used to fluorescently tag nucleic acids [148]. However, reliable and easy detection of DNA [172] and RNA had to wait until the introduction of (F)ISH, standing for (fluorescence) in situ hybridization. In situ refers to the binding of probes at the exact place of the target in the cell, so that information about its localization is achieved. Hybridization denotes that ISH probes are themselves made of (sometimes modified) nucleic acids, which then hybridize with the respective target.

Early ISH probes did not carry a fluorophore, but were radiolabeled instead [55]. When fluorescence came into play [7], FISH advanced in terms of resolution, safety and speed. However, probes were still large and microscopy equipment needed high photon yields, so that detection of DNA was the primary use at the time (DNA-FISH).

The advance of nucleic acid synthesis on a routine, low-cost basis allowed for sets of small (around 20 nt) probes [139], each carrying one or more fluorophores (commercialized as Stellaris^{TM} FISH), as depicted in Fig. 4. Now it was also feasible to stain cellular RNA (RNA-FISH) with sufficient photon yield for modern charge-coupled device (CCD) cameras [46]. Because single RNA molecules were now detectable, methods of this sensitivity are often referred to as single molecule FISH (smFISH). Transcript counting became possible, as well as studies about alternate splicing when using suitable probe sets. The sequences of probe sets (typically 30-48 individual oligonucleotides) used can be easily generated by web tools such as the Stellaris^{TM} Probe Designer nowadays. Such algorithms conduct an NCBI BLAST search against known sequences in the organism of interest to minimize potential non-specific binding. For most targets, sequence ambiguity is high enough at a length of about 20 nt to yield a number of potential specific probe candidate sequences.

FISH, both used for the detection of DNA as well as RNA, has now become a routinely used method in clinical diagnosis [93]. It allows the investigation of the detailed genomic structure of a tissue sample, which is most important for the case of cancer. Macroscopically identical tumors can often be sub-classified into different variants depending on the type of genomic aberration that occurred in the onset of cancer [35]. The identification of the correct subtype can improve both diagnosis and therapy. The drawback of rather long, sometimes overnight, incubation times for FISH experiments have been addressed by the introduction of
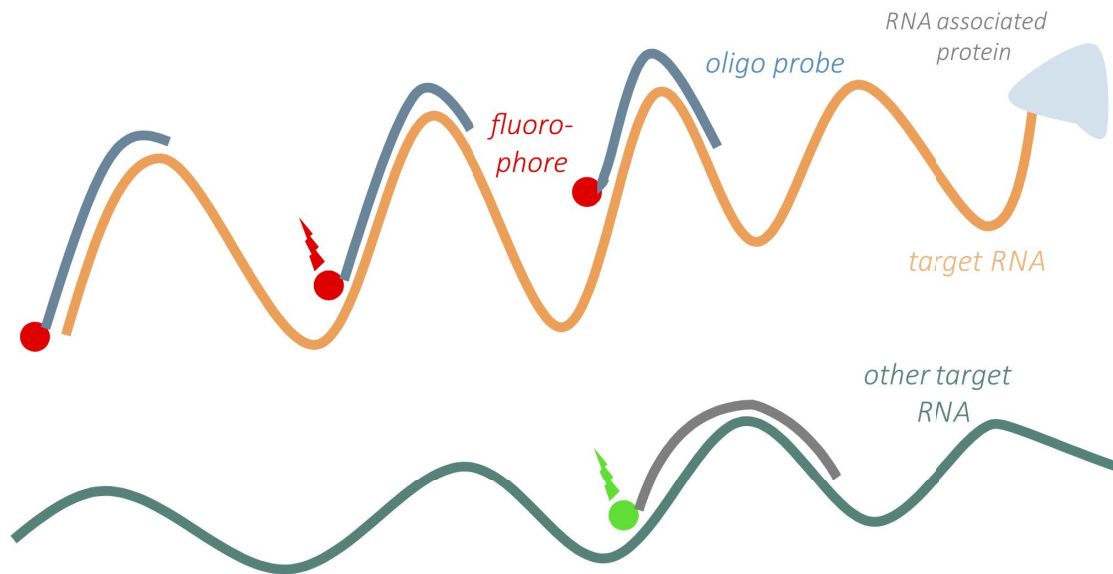
Figure 4 – **Principle of RNA-FISH.** A set of 30-48 short (20 nt long) oligonu-
cleotides (blue, gray) coupled to typically one fluorophore each (red, green) is
hybridized to the target RNA of interest (orange, green) and can be used to specif-
ically stain all cellular RNAs of this type. Multiplexing is possible, e.g. by the
use of another fluorescent dye coupled to another set of probes, as depicted for an
exemplary second target.

Turbo-FISH, which allows reduction of staining times down to a few minutes by the
use of different fixation conditions and high concentrations of detection probes [154].

In recent years, work in the field has been focused on overcoming certain lim-
itations of the technique, like limited multiplexing capabilities, the use of fixated
sample material and limited resolution.

1.5.2   Multiplexing

In microscopy, multiplexing refers to the number of simultaneous stainings which
can be performed independently of each other on the same sample. As with antibody
stainings, FISH offers the intrinsic ability to multiplex using different fluorophores
coupled to different probe sets. Each probe set can then be excited using a unique
laser source and its own filter sets on the microscope. However, due to the width
of excitation and emission spectra of fluorophores, availability and cost of lasers
and suitable fluorophores, there is a limit of typically 4-5 spectrally resolvable
fluorescence channels to that. Several additions to the standard FISH protocol
have been published that try to push this boundary in various ways: Barcoding

refers to several ideas in which researchers have used probe sets labeled in several colors, but unique combinations [92, 76]. When using a small number of basic colors, the number of overlaid colors can be significantly expanded (for example, 3 basic colors A, B, C yield a combinatory space of 7 overlaid colors A, B, C, AB, AC, BC, ABC). The multiplexing capability of barcoding can be stretched to almost infinite numbers when switching to temporal barcoding, termed seqFISH [98]. Here, samples are repeatedly stained after DNAse cleavage of the probes used in previous FISH runs. In subsequent runs, completely different, yet full-length probe sets could be used to have linear extension multiplexing capability. However, it is also possible to use partial probe sets of individual targets in subsequent runs and then mathematically reconstruct their identity based on the combinations of colors observed over time. This approach scales exponentially with the number of repeated stainings and has been – improved by image correlation (corrFISH) – envisioned to be used for complete transcriptomics experiments [32]. Similar (yet different in the details) is an approach termed multiplexed error-robust FISH (MERFISH) which relies on redundant basic probes ubiquitously binding to a large set of mRNAs which are then specified by probes coding for "words" [25, 113]. In subsequent imaging runs, different words are probed and binary answers are given – existent or not for a given mRNA molecule. Bleaching is performed between different runs to achieve signal removal. Transcriptomics-like analyses have also been conducted using a branched DNA (bDNA) approach, in which a relatively small probe is then secondarily extended by a DNA branch covered with fluorophores [5]. This method allows bright visualization of relatively small parts of RNA, but in this case not directly increased multiplexing abilities – the manifold analysis of different RNAs was performed in different samples. FISSEQ was originally introduced as a method to amplify and detect mRNA on solid substrate and was recently extended for gene expression profiling in cells. In situ amplicons of 27 nt length were imaged using FISH and allowed for the sequencing of over 4000 genes [77].

Multiplexing with classical antibody stainings is somewhat more complicated. Due to the extremely high affinity between antibodies and their epitopes, there is no current solution to elute bound antibodies from a sample without irreversible damage to it. The MxIF approach used chemical inactivation of the fluorescent dyes to repeatedly delete the fluorescent signal for another staining run [59]. Compatibility with DNA-FISH has been claimed, but not proven so far. In fact, use of harsh chemicals might likely interfere with the workings of FISH.

Regardless of the multiplexing method, a multitude of probe sets and fluorophores will be present in most extensive FISH experiments. Because FISH probes are covalently linked to their fluorophores, it is not easy to "recycle" probes for use

with another fluorophore. Strictly speaking, this is an obstacle that can be eas-
ily overcome with sufficient time and effort, but quite possibly many interesting
FISH experiments have not been conducted because the expected outcome was
not beneficial enough to justify costly probe design, synthesis and coupling. A
recent approach has addressed this problem in the style of antibody stainings,
which usually consist of two steps: a primary antibody recognizes the epitope on
a protein of interest, and a versatile secondary antibody is fluorescently marked
and is only fixed to the species in which the primary antibody was raised. Most
laboratories working in immunofluorescence stock a range of secondary antibodies
with dyes of interest and recognizing the typical species used to raise antibodies
in. The transfer of this concept to the world of FISH has been called smiFISH
[168]. During probe design, a specific non-hybridizing tail is added to each probe.
The tail sequence (an example would be a FLAP tag) should not be affine to any
known sequence in the host of interest. Secondary probes are designed as well,
which carry complementary sequences to the tail of the normal (primary) probes.
Such a system separates detection from staining and allows for a convenient change
in detection colors. Additionally, it makes the synthesis of the primary probes
comparatively cheap because no modifications are involved.

1.5.3 STED-FISH

Interestingly, stimulated depletion emission microscopy (STED), which allows
resolutions in optical microscopy down to tens of nanometers [85], has rarely been
combined with RNA-FISH [191]. In a recent study, the replication of herpes
simplex virus 1 (HSV-1) has been studied in detail using two FISH probes imaged
at super-resolution [95]. This might be due to rather low spot counts in most
RNA-FISH assays, so that individual spots' point spread functions are highly
unlikely to overlap. In comparison, copy numbers of viral RNA can reach very
high numbers, so that super-resolved applications of RNA-FISH might well yield
additional benefit.

1.5.4 Complementary methods

It should not be overlooked that complementary methods for fluorescent detection
of RNA exist.

One such method, the MS2 system, leverages the MS2 coat protein from the
bacteriophage MS2 which shows an affinity for a certain RNA hairpin [8]. Introduc-
ing repeats of this hairpin into an mRNA of interest and simultaneous transfection
of a fusion protein of MS2 with a fluorescent protein like GFP allows for fluorescent
mRNA detection. A great advantage of the MS2 system is its ability to function in

live cell imaging. However, there are also a number of obstacles involved: cloning might be strenuous and time-consuming, the dynamics of the target RNA might be changed by the introduction of the hairpins and steric hindrance of the MS2 proteins might pose a problem in certain situations, for example with densely coated RNA found in IAV vRNPs. Free-floating fluorescently marked MS2 protein also causes relatively high background fluorescence.

Other approaches to yield live-cell RNA stainings are molecular beacons [170] and forced intercalation (FIT) probes [158], both of which are essentially modified oligonucleotide probes that have been designed to significantly enhance their fluorescence upon binding to their target DNA or RNA sequence to allow wash-free labelling. Usually the oligonucleotide backbone is modified by substitution of one or more PNA or LNA bases which partially shield the probe from cellular RNAse activity. Said fluorescence enhancement is achieved by a quencher which is moved away by conformational change during binding to the target (in the case of molecular beacons) or a fluorophore releasing energy by rotation in the unbound state and by fluorescence in the bound state (in the case of FIT probes). Both approaches require the probes to be delivered to cells if applied for live-cell imaging. This can be achieved by microinjection or permeabilization via treatment with streptolysin-O which induces reversible pores in the cellular membrane [150?].

Quantum dots, characterized by their superior photostability and versatility, have also been attached to probes [20]. As a rather recent invention, they currently suffer from comparatively large size which complicates cell entry and target accessibility. Once these issues are tackled, they might pose an interesting alternative to the above mentioned methods.

1.6 Image analysis of (highly multiplexed) FISH data

1.6.1 General image analysis for smFISH

RNA-FISH images are often compared to a sky full of stars, hence the commercial name Stellaris^{TM}. This means that, at least when viewed on a non-super-resolution setup, detected RNAs occur as spots (often referred to as *puncta*) shaped by the point spread function (PSF) of the actual RNA shape. Image analysis of smFISH is therefore essentially spot detection.

In many applications of FISH (especially DNA-FISH), there is no need for advanced image analysis: the outcome of a staining is simply whether signal is detected

or not, possibly with the additional information of the cellular localization. If a scarcely expressed mRNA is stained using smFISH, it is often possible to assess transcription levels via counting by eyesight. However, more complex problems require sophisticated image analysis software and workflows.

1.6.2   Deconvolution and Laplace filtering

If not only the number of spots, but also their precise localization is of interest, imaging should be performed on a confocal microscope which yields superior resolution. The sample can be scanned along the z-axis by taking several images (slices) in different heights. Deconvolution can help to pinpoint the actual localization by analyzing the acquired spot through several slices of an image stack, obscured by its PSF [109]. Out-of-focus photons are used in the process to reconstruct the image without a PSF as good as possible. Alternatively, or in addition to deconvolving the image stack, filtering e.g. via as a Laplace filter can be applied. Such filtering can enhance the contrast of the image and improve the extraction of image features, i.e. the FISH spots. Obviously, care has to be taken to not misuse filtering settings to generate false positive signals.

1.6.3   Thresholding and point detection, FISH-quant

There are numerous actual spot detection packages available online, like comDet [181], StarSearch [139] and FISH-quant [116]. Basically, the user sets thresholding parameters to separate background fluorescence from actual signals by assigning a threshold value. FISH-quant uses primarily the brightness of an object for thresholding, but also computes a "quality score" which includes parameters like the local contrast. All structures that cross the barrier of the threshold are considered for further analysis. A variety of measures is computed, most importantly the localization in 3D (if the software supports image stacks). FISH-quant even supports sub-resolution localization which is possible through the assumption of a spherical PSF – if enough pixels are covered by the PSF of a detected spot, its center can be taken as a more accurate measure than the brightest pixel. For convenience, some software packages offer user-based or automatic detection of cells and nuclei to really yield a single-cell resolution.

1.6.4   Multiplexing

Interestingly, none of the freely available image analysis softwares support multiplexing out of the box. Multiplexing by means of spectral barcoding in a single FISH run can be assessed by comparing all analyzed channels side by side, but will still need downstream algorithms if a large number of spots has to be processed.

The situation is even more complicated for methods which include multiple staining runs. An ideal analysis software with multiplexing capability would need to import sets of images of the same sample position, but from different runs (stained with different probe sets), and merge the data in a meaningful way. Depending on the way such repeated FISH runs have been acquired, the need for image registration might arise [110]. The sample position might be recovered approximately, but not exactly at each run, and registration is the process to align such images with a little offset to each other. Implementations of registration algorithms are available in MATLAB and as ImageJ plugins. For FISH images, it is typically advisable to use feature-based registration, where certain features are identified in each single image and transition matrices are computed to match those features onto all other images as good as possible. Transition matrices are discriminated by their degrees of freedom, from translation to rigid (translation + rotation), and similarity (rigid + scaling) to affine (similarity + stretching and shearing). For FISH experiments, rigid transformations should usually be sufficient for registration.

1.7 Aim of the project

It is clear that influenza biology has progressed quite extensively in the study of its proteins and life cycle, as well as *in vitro* studies of its RNA biology and -interplay. However, insights into localization and cell specific expression of viral RNA have been limited because of lacking tools for the study of large numbers of RNA at the same time.

Therefore, the first aim of this work was to overcome the spectral limits of multiplexed RNA imaging. A method was sought for that would allow for comparatively easy and reliable staining of large numbers of RNA species, as they typically occur in an IAV infection: 8 vRNAs and 10 mRNAs.

A second aim was then to use the highly multiplexed data to unravel the mystery of IAV genome packaging. As discussed in detail above, it is a long-standing question whether IAV packages its genome in an orderly fashion, possibly even hierarchical, or in a completely or partially random way. Evidence for selective packaging has been emerging over the last couple of years, however many of the studies involved relied on *in vitro* methods like gel bandshift assays or were limited in the number of segments to study in parallel. Such technical limitations prevented an overall picture of the packaging mechanism and needed to be overcome in this project.

2 | Materials and Methods

2.1   Materials

2.1.1   Chemicals and enzymes

- **UltraPure DNase/RNase-Free Distilled Water,** Thermo Fisher Scientific, Hennigsdorf, Germany

- **DNase I, RNase-free,** Thermo Fisher Scientific, Hennigsdorf, Germany

- **10% formalin,** Sigma-Aldrich, St. Louis, MS, USA

- **Triton X-100,** Sigma-Aldrich, St. Louis, MS, USA

- **Acetylated BSA (antibody staining),** Sigma-Aldrich, St. Louis, MS, USA

- **Dextran sulfate sodium salt from *Leuconostoc spp.*,** Sigma-Aldrich, St. Louis, MS, USA

- **Ethanol (EtOH),** Carl Roth, Karlsruhe, Germany

- **20x Saline-sodium citrate buffer (SSC),** Carl Roth, Karlsruhe, Germany

- **Deionized formamide,** Merck Millipore, Darmstadt, Germany

- **Ribonucleoside vanadyl complex (VRC),** New England BioLabs, Frankfurt, Germany

- **α-NP-FITC mouse anti IAV NP primary monoclonal antibody conjugated to FITC,** MAB8257F, Merck Millipore, Darmstadt, Germany

- **Superscript IV (SSIV) reverse transcriptase,** Thermo Fisher Scientific, Hennigsdorf, Germany

- **5x SSIV buffer,** Thermo Fisher Scientific, Hennigsdorf, Germany

- **0.1 M Dithiolthreitol (DTT),** Thermo Fisher Scientific, Hennigsdorf, Germany

- **dNTPs mix, 10 mM each,,** Thermo Fisher Scientific, Hennigsdorf, Germany

- **Ribonuclease H (RNase H),** Thermo Fisher Scientific, Hennigsdorf, Germany

- **RNasin,** Promega, Mannheim, Germany

- **Salmon sperm DNA (SSD),** Sigma-Aldrich, St. Louis, MS, USA

- **KAPA SYBR FAST Mastermix (SYBR green),** Peqlab, Erlangen, Germany

- **Bovine serum albumin (BSA),** A7979, Sigma-Aldrich, St. Louis, MS, USA

- **SYBR®Safe DNA Gel Stain,** Thermo Fisher Scientific, Hennigsdorf, Germany

- **O'GeneRuler DNA Ladder Mix, ready-to-use,** SM1173, Thermo Fisher Scientific, Hennigsdorf, Germany

- **Orange DNA Loading Dye (6X),** Thermo Fisher Scientific, Hennigsdorf, Germany

- **Phusion High-Fidelity DNA Polymerase,** Thermo Fisher Scientific, Hennigsdorf, Germany

- **DpnI,** Thermo Fisher Scientific, Hennigsdorf, Germany

- **Bacto™ agar, Bacto™ tryptone, Bacto™ yeast extract,** BD Biosciences, Heidelberg, Germany

- **TurboFect Transfection Reagent,** Thermo Fisher Scientific, Hennigsdorf, Germany

2.1.2 Solutions

- **PBS+:** Dulbecco's Phosphate-Buffered Saline with Ca^{2+} and $^{Mg2+}$, Pan Biotech, Aidenbach, Germany

- **PBS-:** Dulbecco's Phosphate-Buffered Saline without Ca^{2+} and $^{Mg2+}$, Pan Biotech, Aidenbach, Germany

- **Trypsin solution:** Trypsin 0.05 % and EDTA 0.02 % in PBS-, Pan Biotech, Aidenbach, Germany

- **DMEM:** Dulbecco's Modified Eagle Medium, Pan Biotech, Aidenbach, Germany

- **DMEM+:** DMEM with additional 1 % penicillin/streptomycin solution and 2 mM L-glutamine, Pan Biotech, Aidenbach, Germany

- **DMEM++:** DMEM+ with additional 10 % fetal bovine serum (FBS), Pan Biotech, Aidenbach, Germany

- **LB agar:** 35 g/l LB agar + 100 µg/ml Ampicillin, Carl Roth, Karlsruhe, Germany

- **LB medium:** 35 g/l LB medium + 100 µg/ml Ampicillin, Carl Roth, Karlsruhe, Germany

- **Infection medium:** DMEM+ with additional 0.2 % BSA

- **Cryo medium:** 80 % FBS, 10 % DMEM, 10 % DMSO

- **Blocking buffer:** 2x SSC, 0.2 % acetylated BSA, 2 mM VRC

- **Hybridization buffer:** 2x SSC, 10 % formamide, 10 % dextrane sulfate, 2 mM VRC

- **FISH wash buffer:** 10 % formamide in 2x SSC, 2 mM VRC

- **Probe removal buffer:** 80 % formamide, 20 % H_2O, 2 mM VRC

- **TBE buffer:** 11 g Tris, 5.5 g boric acid, 4 ml 0.5 M EDTA in H_2O

2.1.3 Consumables

- **Filter tips 10 µl, 20 µl, 1000 µl,** Starlab, Hamburg, Germany

- **15 ml, 50 ml falcon tubes,** SARSTEDT, Nümbrecht, Germany

- **1.5 ml reaction tubes, nuclease-free, low binding,** SARSTEDT, Nümbrecht, Germany

- **Cryo tubes (1.6 ml),** SARSTEDT, Nümbrecht, Germany

- **T25 cm², T75 cm², T175 cm² cell culture flasks,** SARSTEDT, Nümbrecht, Germany

- **6-well, 24-well plates,** SARSTEDT, Nümbrecht, Germany

- **2 ml, 5 ml, 10 ml serological pipettes,** SARSTEDT, Nümbrecht, Germany

- **15 ml, 50 ml falcons, nuclease-free,** Greiner Bio-One, Kremsmünster, Austria

- **LightCycler 480 white Multiwell Plate 96,** Roche, Berlin, Germany

- **ibidi µ-Slide VI 0.4 (#80606),** ibidi, Munich, Germany

- **ibidi µ-Slide Angiogenesis (#81506),** ibidi, Munich, Germany

- **ibidi µ-Slide 8 well (#80826),** ibidi, Munich, Germany

2.1.4   Fluorescent dyes

- **ATTO550 NHS-ester (#AD 550-35),** ATTO-TEC, Siegen, Germany

- **Abberior STAR 635P NHS-ester (#07679),** Sigma-Aldrich, St. Louis, MS, USA

- **DAPI (#D9542),** Sigma-Aldrich, St. Louis, MS, USA

2.1.5   Primers and sequences of FISH probes

Primers used for generation of mRNA-only plasmid system of A/Panama are described in the results part of this work. An extensive list of primers used for reverse transcriptions, for qRT-PCR and of all sequences of the FISH probe sets used can be found in the thesis of Ivan Haralampiev [63].

2.1.6   Kits

- **Plasmid Plus Maxi Kit,** Qiagen, Hilden Germany

- **RNeasy Mini Kit,** Qiagen, Hilden Germany

- **QIAquick Gel Extraction Kit,** Qiagen, Hilden Germany

2.1.7 Devices

- **VisiScope scanning disk confocal laser microscope 405, 488, 561 and 640 nm diode lasers Halogen lamp with 525/50 nm excitation and 600/50 nm emission filters,** Visitron Systems, Puchheim, Germany

- **CSU-W1 Confocal Spinning Disc unit,** Yokogawa, Tokio, Japan

- **IX-83 inverted fluorescence microscope,** Olympus, Hamburg, Germany

- **60x/1.2 UPlanSApo water objective,** Olympus, Hamburg, Germany

- **60x/1.45 UPlanSApo oil objective,** Olympus, Hamburg, Germany

- **100x/1.3 UPlanFLN oil objective,** Olympus, Hamburg, Germany

- **100x/1.3 UPlanFLN oil objective,** Olympus, Hamburg, Germany

- **Bioprecision2 motorized X-Y-stage,** Ludl EP, Exton, PA, USA

- **NanoPrecision fast Piezo Z-stage insert,** Ludl EP, Exton, PA, USA

- **Evolve 512 back-illuminated EMCCD camera (512x512 pixels),** Photometrics, Tucson, USA

- **iXon 888 EMCCD camera (1024x1024 pixels),** Andor, Belfast, Northern Ireland

- **LightCycler 480 II,** Roche Diagnostics, Mannheim, Germany

- **NanoDrop 2000,** Peqlab, Erlangen, Germany

- **Incubator New Brunswick Galaxy 170 S,** Eppendorf, Hamburg Germany

- **Thermomixer Comfort,** Eppendorf, Hamburg Germany

- **Centrifuge Avanti J-20XP,** Beckman Coulter, Krefeld, Germany

- **Rotor JLA10.500,** Beckman Coulter, Krefeld, Germany

- **HERAsafe biological safety cabinet,** HERAEUS Instruments, Hanau, Germany

- **DNA/RNA UV-Cleaner UVC/T-M-AR,** Kisker Biotech, Steinfurt, Germany

- **Mini-Sub Cell GT Systems (agarose electrophoresis),** Bio-Rad, München, Germany

- **UV transilluminator,** Vilber Lourmat, Eberhardzell, Germany

2.1.8   Cell lines and bacterial strains

- **A549 human epithelial lung cell,** ATCC CCL-185

- **MDCKII Mardin-Darby canine kidney cell,** Gerrit van Meer, Utrecht University

- **HEK293 human embryonal kidney cell,** Molecular biophysics lab, HU Berlin

- ***E. coli* DHS,** Molecular biophysics lab, HU Berlin

All viral strains used in this work were provided by the group of Thorsten Wolff at the Robert-Koch-Institute (RKI) Berlin. Preparations were done by Anne Sadewasser and Gudrun Heinz in the group of Thorsten Wolff.

2.2   Methods

2.2.1   Cell culture

Cells were thawn from stocks stored in liquid nitrogen to start a new culture. Such cryo tubes from the stock were thawn and then added to 13 ml pre-heated DMEM+ medium to reduce the DMSO concentration from the cryo medium. Cells were then pelleted via centrifugation (5 min, 300 g) and again added to DMEM+ and transferred to a T25 cell culture flask. After a few rounds of passage (see below) they were transferred to normal T75/T175 cell culture flasks if needed.

Prolonged culture of cells was done in T75 or T175 cell culture flasks in the incubator (37°C and 5 % CO_2). Cells were passaged every two or three days to prevent growth until 100 % confluency. Passaging/seeding consisted of two wash steps using PBS-, followed by 10-20 min incubation with pre-heated Trypsin solution in the incubator. The now suspended cells were taken up in 8 ml DMEM+ to inactivate the Trypsin solution and were returned to the cell culture flask in the desired amount (passaging) and/or added to microscopy dishes for experiments (seeding) in desired concentrations. Finally, DMEM+ was added to the cell culture flask to reach normal volume again. Cell culture was not stretched over more than 15 propagations to avoid genomic changes. Every 5 propagations, the cell culture

flask was exchanged for a new one.

If needed, cells were frozen after trypsination and a centrifugation step (5 min, 300 g). After resuspension in cryo medium, 1 ml aliquots in cryo tubes were stored in liquid nitrogen.

2.2.2   Infection, Fixation, Permeabilization

Infections were performed on cells which were seeded on the day before. Infection suspension was prepared, using infection medium and adding the amount of virus stock needed. To calculate the amount of virus needed for a given number of cells at a desired multiplicity of infection (MOI), the following equation was used.

$$MOI = \frac{\# \ of \ virus \ particles}{\# \ of \ cells} \tag{2.1}$$

Cells were washed with PBS+ and put on ice. Infection suspension was added to the cells and 15 min adheration time on ice followed. At 4Â°C, IAV cannot enter cells as no fusion takes place, however HA proteins can attach at cellular sialic acids. After the adheration period, the cells were transferred to the incubator (37°C and 5% CO2) to allow synchronous fusion and infection. All infection times refer to this step. 45 min after infection start, the cells were washed with PBS+ again to remove unbound virus and plain infection medium without virus was added.

After the desired infection time, cells were washed again with PBS+ and then fixated using 4% PFA solution or 10% formaline for 10 min at RT. Cells were again washed twice using PBS+.

Permeabilization was either performed overnight at -20°C using 70% EtOH (RNAse-free) in water or 10-20 min at RT using a 0.5% Triton-X-100 solution. In the case of Triton-X-100, extensive washing with PBS+ followed to ensure that no detergent would remain in the microscopy dish.

2.2.3   Immunostaining

Samples were regularly stained using antibodies and/or DAPI. For antibody incubation, the sample was washed using PBS+, after which 0.2% BSA solution (RNAse-free) was added for 30 min of blocking at RT to avoid off-target binding of the antibody. Thereafter, the solution was exchanged for primary antibody diluted at a desired rate (usually 1:500) in 0.2% BSA solution (RNAse-free). Incubation for 45-120 min at RT followed. Some antibodies were already coupled to a fluorophore, so that no secondary antibody staining was necessary. In all other cases,

two washes with PBS+ followed (15 min each, RT) after which the antibody incubation was repeated as described above, just with a matching secondary antibody.

In any case, the sample was protected from light exposure from the moment that fluorophores were added.

If nuclear staining was required, the sample was exposed to 500 nM DAPI solution for 10-15 min at RT and washed twice with PBS+ afterwards.

2.2.4   smFISH and MuSeq-FISH

For FISH staining, the sample was first put into 2x SSC buffer with 2 mM vanadyl ribonucleoside complex (VRC) added to it. The hybridization mixture was prepared as follows: an aliquot of pre-hybridization buffer containing dextran sulfate and SSC was thawed. Freshly thawn formamide and water (RNAse-free) were added, so that the final hybridization buffer contained 10% formamide, 2x SSC, 2 mM VRC and 10% dextran sulfate. FISH probes were added to the hybridization buffer to reach a final concentration of 100 nM for each probe set. Before applying the hybridization mixture to the sample, it was briefly heated to 60°C to melt secondary structures. The actual hybridization lasted 2-4 h in the incubator. Microscopy dishes of small volume were wrapped in wet tissues to prevent evaporation of the hybridization mixture.

After hybridization, the sample was washed twice with FISH wash buffer and 15 min in the incubator (37°C, 5% CO2) each time. Microscopy was conducted in 2x SSC buffer with 2 mM VRC added, for details see section below. After images of one staining were acquired, the bound FISH probes from this staining were removed using pre-heated probe removal buffer 15-20 min in the incubator (37°C, 5% CO2). Subsequently, the high formamide content was washed twice with 2x SSC and microscopy for control pictures without bound probe was performed if needed. Finally, the sample was ready for another round of staining.

For MuSeq-FISH, this cycle of events was repeated up to 12 times.

2.2.5   Image acquisition

All images except for STED super resolution images in this work were acquired on a spinning disk confocal microscope. STED microscopy was done by Maria Loidolt, AG Hell, Max-Planck-Institut für Biophysikalische Chemie, Göttingen using a custom-built setup that uses a supercontinuum laser source for concurrent detection in four spectral channels via APDs, the outputs of which were then later

translated to the imaging channels using spectral unmixing [182].

Diode lasers at the wavelengths of 405 nm (for DAPI and DyLight 405 illumination), 488 nm (for Atto 488, FITC and Alexa Fluor 488 illumination), 561 nm (for Atto 550 illumination) and 640 nm (Star 635P and Alexa Fluor 647 illumination) were used for microscopy. All microscopy images were acquired as z-stacks with a spacing of 400 nm between slices and a total number of 26-29 slices.

MuSeq-FISH requires images taken at the exact same positions within the sample for each of the sequential stainings. Therefore "landmarks" were chosen at the edge of a microscopy dish where unique impurities in the dish itself or cells in this area would be easily recognizable for the next run. All other microscopy positions would then be at a defined distance in x-direction away from the "landmark". For some experiments, the relative positions were calculated and loaded into VisiView software to directly find the saved positions. During microscopy, care was taken to ensure best possible matching to earlier acquisition positions in order to ease later image registration.

2.2.6   Image analysis and spot detection

All images of the same positions, but stemming from different MuSeq-FISH runs were grouped together and sub-pixel image registration in x and y directions was performed using a custom written ImageJ script by Matthias Schade (AG Molecular Biophysics, HU Berlin), which itself made use of the Template Matching and Slice Alignment plugin by Tseng et al. [169]. Registration in z-direction was done in MetaMorph by optical comparison of the respective z-slices.

FISH-quant detection software by Florian Mueller (Institut Pasteur, Paris) was then used for spot detection in all images [116]. Cell boundaries and nuclei positions were marked using the in-built cell manager of FISH-quant. Using the registration data from above, the cell positions were adjusted for each MuSeq-FISH run. The detection settings, especially concerning the thresholding, were set so that control images showed none or minimal spot counts and were then kept constant for all images to ensure reproducibility. Spot detection was then performed on all acquired images in batch mode.

2.2.7   Colocalization analysis

The actual colocalization analysis was implemented as a custom written R script, the workings of which will be explained in the following:

Initially, a set of individual spot detections from FISH-quant belonging to the same MuSeq-FISH image position was loaded into R. Helper scripts in R with re-formatting routines made it possible to use data from multiple detections in the same dataset, a feature that is not normally available in FISH-quant. Next, registration data from above was loaded into the script and then applied to the respective images, therefore making the spot detections comparable between different MuSeq-FISH runs. The maximum displacement of images was calculated and the same area was excluded from further analysis because one or more images did not contain this area.

An optional step was then to transform images to generate negative control colocalization. If used, this option would rotate images by 90°, 180°, 270°, flip them in x- or y-direction or a combination of those. This resulted in images which were all distorted to each other and should show colocalization at a basal level only, corresponding to the spot density in the images.

Spot listings for each image were then scrambled to avoid any bias of "early" detections or images, i.e. in the upper left corner where the algorithm would otherwise start searching for centers of colocalization. Also, in line with the analysis procedure of Lakdawala et al. [90] all nuclear spots were excluded from the analysis because we did not expect vRNPs to directly interact after their replication.

If targets were stained redundantly in two different MuSeq-FISH runs using two different flurophores, the spots from both acquisitions were compared to each other. Only spots detected in both independent stainings were kept for further analysis and statistics were generated to determine the quality of both stainings. This procedure could only be applied to A/Panama because the stainings using the PCA Star 635P fluorophore were of insufficient quality for all other viral strains.

After all these preparatory procedures, the actual colocalization could begin. The scrambled list of spot detections for the respective MuSeq-FISH image set was processed spot by spot. For each spot, all other spots in close proximity (a cylinder with a radius of 300 nm in x and y-directions and a height of 1000 nm in z-direction) were binned with the original spot into a "multi-segment-complex" (MSC) with consecutive numbers. Spots contained in this MSC were then flagged as "used" and were not considered for further colocalizations – this ensured that each spot would only be part of one MSC.

An optional feature to increase accuracy was to perform a second colocalization analysis after completion of the first one. Here, all MSCs below a given rank

(number of spots in this MSC) were unflagged and therefore made available for colocalization again. For all remaining MSCs (those of high enough rank), centers of mass were calculated and those were used as starting points for the second colocalization analysis. This removed the restriction that MSCs always had to be centered in one of its spots. However, this feature was computationally very intense and was therefore not used in the analysis described as accuracy improvements were negligible.

The resulting MSCs were checked for their contents. If an MSC contained two copies of the same segment i.e. two detections from the same image), this biologically not expected result was treated as a detection artifact. Such MSCs were removed from the data set and treated as an indicator of detection and colocalization quality. With the settings used, about 5% of MSCs were discarded in this step.

Finally, the data set was saved for later reference, all relevant statistics were saved in Excel file format and the *ggplot2* package of R was used to generate various plots to interpret the data. Among those are a histogram of MSC sizes, binary colocalization heat maps for all segment sizes, centroid measures such as MSC diameters and angles, cell measures such as bar and jitter plots to display segment abundancies for each cell, and localization plots of the detected spots in various color themes and 2D/3D.

2.2.8   Network analysis

The network analysis can be seen as an extension to the aforementioned colocalization analysis. All detected MSCs were binned based on their segment content. To simplify the packaging model as far as possible, we assumed that MSCs have a maximum size of 8, that segments do not occur more than once within an MSC, and that the order of segments is irrelevant to characterize an MSC (i.e. segment combinations are permutation-insensitve). With these assumptions, the total number of states within the packaging network or packaging process is 255 (8 for rank 1, 28 for rank 2, 56 for rank 3, 70 for rank 4, 56 for rank 5, 28 for rank 6, 8 for rank 7, 1 for rank 8). For connections between those states, we assumed that no back-reactions occur (i.e. a complex is stable once it has assembled) and that two educt MSCs react to a product MSC. In addition, the product MSC should follow the assumptions stated above. Depending on the order of the reactions allowed, this resulted in 1016 (only monomeric additions allowed, i.e. one educt MSC is always of rank 1) to 3025 (up to tetrametric additions allowed) connections within the network.

These theoretical thoughts were implemented in another custom written R script

using the *igraph* package [34]. Here, we used the aggregated data of all analyzed images per viral strain. A network graph object was created where the nodes correspond to states of IAV packaging whereas the edges are the connections allowed between those nodes. Nodes received a "weight" attribute which was equal to the number of detections of this specific segment combination in the microscopy data, normalized among combinations of the same rank. Edge weights were not easy to assign at this stage because the microscopy data was essentially a snapshot picture, with no kinetics involved. To ease graphical assessment of the network graph, a proxy measure for edge weights was found as the normalized product (within each rank) of both end node weights. Using those weights, a heatmap color-coded network graph representation could be plotted using in-built *igraph* routines.

To analyze whether some edges are preferred in the network (i.e. whether reaction constants vary), an iterative simulation model was implemented in R. Here, IAV packaging was simulated by an initially empty network (node counts = 0 for all MSCs). Monomeric MSC counts were then set to the total spot count of the respective segment. Now the simulation proceeded with one reaction happening per iteration, with the laws of mass action governing the choice which reaction to take. Therefore, the probability of a reaction to happen would be

$$p = \frac{\# \ educt \ 1 \ * \ \# \ educt \ 2}{(\# \ of \ all \ spots)^2} \tag{2.2}$$

One of the educts was always a monomeric MSC and the counts of these monomeric MSCs were fixed. All other counts were dynamic, so that each reaction increased one count and decreased another. Simulations were run for 10^7 iterations so that mass propagated through the network. Each 5000 iterations, measures like the cross-correlation to the microscopy data network were calculated. After completion of the simulation the iteration with the highest cross-correlation was then used for further analysis; usually this value peaked during the simulation and then decreased again.

2.2.9   Conditional probability modelling

An approach was taken to estimate k-values in the IAV packaging network from intrinsically static data based on microscopy at only one time point. The implementation was accomplished by Max Schreiber (University of Leipzig, Germany) in Matlab by Mathworks. The general concept behind this modelling approach was to estimate the "real" underlying k-values by factoring out the influence based on segment availability, i.e. mass action kinetics. Reaction order was restricted to monomeric additions, back-reactions were excluded and each pair of MSC ranks

(e.g. rank 2 to 3, rank 3 to 4...) was treated as an individual problem to reduce complexity.

2.2.10   RNA extraction

In order to generate samples for qRT-PCR (described below), total RNA extractions were performed on virus-infected A549 cells at various time points after infection (2-24 h p.i.). Seeding and infection of cells was performed either in 6-well or 24-well format with an MOI of 5. After infection, cells were washed with PBS+ and RNA extraction was performed according to Qiagen RNeasy Mini Kit instructions. For each time point and viral strain, three samples were processed to yield biological triplicates. Elution was done using RNAse-free water and the resulting total RNA concentration for each sample was measured using a Nanodrop 2000 spectrophotometer. Samples were then either directly used for reverse transcription or stored at -80°C.

2.2.11   Reverse transcription

Reverse transcription is a method to specifically generate complementary DNA from an RNA template. SuperScript IV^{TM} Reverse Transcriptase (Life Technologies) was used according to supplier instructions. 500 ng RNA was used in each reaction, together with 1 µl of 50 µM primer. Two reactions were set out to reverse transcribe viral mRNA and vRNA, respectively. For viral mRNA, an anchored oligo-dT primer was used for the annealing step and incubation lasted 15 min at a temperature of 50°C. For vRNA, a primer targeting the conserved 5' end of the vRNA sequences was used (so called UNI12 primer, sequence: AGCRAAAGCAGG) and incubation lasted 20 min at a temperature of 55°C. RNAse H was used to remove residual RNA after the transcription had finished. After reverse transcription, the produced DNA samples were stored either at -20°C or -80°C.

2.2.12   Quantitative Real-Time PCR (qRT-PCR)

Polymerase chain reaction (PCR) is the theoretical and practical basis of quantitative real-time PCR (qRT-PCR) [118]. Basically, optimized polymerase enzymes are used in repeated cycles to generate very high copy numbers of a specific piece of DNA governed by the choice of primers. qRT-PCR extends the possibilities of the method by adding a fluorescent dye, SYBR FAST (KAPA Biosystems) in this case. SYBR FAST intercalates with double-stranded DNA which is the result of a successful PCR and increases its brightness in this conformation. This means that a readout of the fluorescence over the time course of the PCR allows for an estimation of the total copy number in the reaction tube and also to infer the

Table 2.1 – qRT-PCR Scheme.

Substance	Volume [µl]
10 µM PCR primer fwd	0.5
10 µM PCR primer rev	0.5
SYBR Fast MasterMix (2x)	5
H_2O (RNase-free)	1
DNA Template	3

starting concentration. A base line curve is needed for calibration, which consists of the same DNA sequence in defined concentration – in this work, pHW2000 plasmids of the minireplicon system carrying the viral genome segments were used to generate the standard curves.

The actual qRT-PCR experiments were performed on a LightCycler 480II (Roche) using the SYBR FAST qPCR Master Mix (KAPA Biosystems) in a reaction mixture as follows:
Plasmids for standard curves were obtained by the group of Thorsten Wolff, RKI Berlin. Using ddH2O, dilutions from 10^{10} copies/µl down to 10^0 copies/µl were created and used in the qRT-PCR experiments. Care was taken that melting curves showed only a single peak and that the reaction efficiency stayed in the expected range for qRT-PCR measurements.

The primers were chosen so that the PCR products would result in a length of about 200 nt.

Results from qRT-PCR experiments were then analyzed and visualized using an R script kindly provided by Fabian Jolmes (AG Herrmann, HU Berlin, Germany).

2.2.13 Polymerase Chain Reaction (PCR), gel electrophoresis and gel extraction

Standard PCR was used to clone pHW2000 plasmids carrying one IAV A/Panama gene segment each to ensure only mRNA would be processed. Original pHW2000 plasmids capable of vRNA and mRNA production were obtained by the group of Thorsten Wolff, RKI Berlin. An approach originally designed for plasmid system creation for new viral strains described by Stech et al. in 2008 [159] was adapted for our approach to remove PolI promoter sites from the already existing A/Panama pHW2000 plasmids.

Basically, cloning was a two-step process. In a first step, the coding sequences of each gene segment were amplified using the pHW2000 plasmids as templates, thereby generating so-called megaprimers. Primers in these reactions were flanking the gene segment and provided overlap sequences with other parts on the plasmid sequence. So each primer consisted of a conserved pHW part, followed by a partially conserved Uni12/Uni13 part depending on 5' or 3' end and a specific part for each segment, as described by Stech et al.

The following PCR program and mixture were used for generation of the megaprimers:

Table 2.2 – PCR scheme

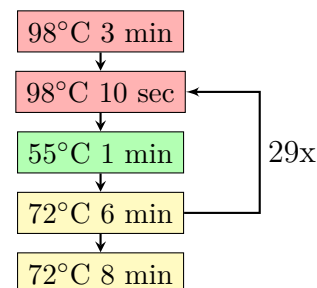| Substance	Volume [µl]
H_2O	35.5
5x Phusion HF buffer	10
10mM dNTP	1
20 µM Primer 1 (U12+specific)	1.25
20 µM Primer 2 (U13+specific)	1.25
100 mg/µl A/Panama pHW2000	0.5
Phusion HF DNA-Polymerase	0.5

Figure 5 – PCR
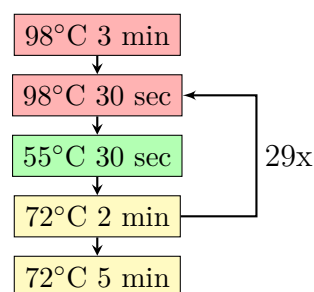
The PCR products were then mixed with Orange DNA loading dye (6x) and applied to a 1% agarose gel for separation by length. SYBR Safe was used as an unspecific DNA marker and O'GeneRuler DNA Ladder Mix as a DNA ladder. An electric field was applied using a Mini-Sub Cell GT Systems device set to 80 V. Gel electrophoresis was run for 45 min, after which the bands were inspected under UV light and those of correct size were cut out using a razor knife. Extraction was performed using QIAquick Gel Extraction Kit according to manufacturer's instructions. Concentrations were measured on a NanoDrop 2000 spectrophotometer.

The second step of cloning consisted of the insertion of the produced megaprimers into the original pHW2000 plasmids in a way in which the PolI promoter would be removed from the resulting plasmids, but the neighboring polyadenylation site would be retained. The primers used in the first PCR were designed in a way that the megaprimers were already able to hybridize with the original plasmid sequence, so that an adapted QuikChange protocol could be applied. Program and mixture were as follows:

Table 2.3 – 2nd PCR scheme   **Figure 6** – 2nd PCR

Substance	Volume [µl]
H_2O	33
5x Phusion HF buffer	10
10 mM dNTP	1
Template pHW2000 (empty)	varied (100 ng)
Megaprimer	varied (100 ng)
Phusion HF DNA-Polymerase	0.5

98°C 3 min
98°C 10 sec
55°C 1 min
72°C 6 min
29x
72°C 8 min

Afterwards, a DpnI digestion followed, where 2 µl DpnI were added to each sample and incubated for 2 h at 37°C. Digested DNA samples were then stored in the fridge at 4°C for later use.

2.2.14 Transformation, plasmid prep. and sequencing

Resulting PCR products were transformed into α-competent *E. coli* of the strain DHS. For this, the bacteria were taken from their -80°C stock, thawed on ice and then gently mixed with 10 µl of the respective PCR product. Incubation on ice of 5-10 min followed. Then, a 42°C heat shock was applied for 45 s, after which cells were transferred back to ice for 2-5 min to rest. 500 µl LB media without any antibiotics was gently added to each reaction tube, and incubation for 45 min at 37°C followed. Petri dishes prepared with LB agar and ampicillin resistance were pre-warmed at 37°C during this time. Bacteria were then gently pipetted together with LB solution onto one petri dish each and the solution was thoroughly dispersed using glass beads. Finally, incubation overnight at 37°C and 5% CO2 followed. On the next day, grown colonies could be picked.

100 ml Erlenmeyer flasks were filled with 5 ml LB with ampicillin and were inoculated with a single colony each by dropping a pipette tip which was dipped onto the colony before. Incubation at 37°C and 5% CO2 for 16 h at a shaking incubator followed. Plasmid preparation (Miniprep) was performed using the QIAprep Spin Miniprep Kit and according to manufacturer's instructions.

Sequencing was done using the LightRun service of GATC Biotech (Konstanz, Germany). 5 µl of each miniprep sample was mixed with 5 µl of one of the respective primers (5 µM concentration) for the given segment, which were used for generation of the megaprimer PCR amplifications. The resulting solution was sent to GATC Biotech and results were compared to the reference sequence using ApE –

A plasmid editor (by M. Wayne Davis).

2.2.15   Transfection

Transfections were always done using the human embryonal kidney cell line HEK293 of low passage number. Cells were seeded in µ-slide 8 well ibiTreat microscopy dishes (ibidi, Munich, Germany) and were grown for 24 h before transfection was started. Transfection was performed according to manufacturer's instructions. Therefore, the transfection mixture consisted of transfection medium which is identical to DMEM+, where plasmids to be transfected were added in amounts so that the total DNA content would be 1 µg and the same volume of transfection reagent was added as well. The mixture was then briefly vortexed and incubated for 20 min at RT. It was then added to the cells and each well was filled up using transfection medium so that a final volume of 200 µl was reached. Transfection then lasted 30 h in the incubator (37°C and 5% CO2), after which the cells were fixated, permeabilized and subjected to immunostaining, as described above. A Rab11-RFP plasmid provided by Chris Höfer (AG Veit, FU Berlin) was used as a transfection control in transfected samples and was also the only transfected plasmid in mock-transfections.

3 | Results

The data shown here are the outcome of a highly collaborative project. The establishment of MuSeq-FISH as a method and all FISH experiments have been conducted together with Ivan Haralampiev. Figures and results concerning this part are also part of his thesis and are covered in more detail there [63]. Image analysis and data processing were conducted by myself, but in close teamwork and with many helpful insights by Ivan Haralampiev. The conditional probability modelling was designed in collaboration with Ivan Haralampiev and Max Schreiber and implemented entirely by Max Schreiber. All RT-PCR data was acquired by Kalle Jongen (A/Panama strain and mock control) and Vanessa Körner (all other viral strains) under the joint supervision of Ivan Haralampiev and myself [79, 86]. The proof-of-concept experiment for the transfection of mRNA-only A/Panama plasmids was performed by Malte Hilsch under the supervision of Ivan Haralampiev and myself.

3.1 Generation of a set of mRNA-only IAV A/Panama plasmids

3.1.1 Cloning

In this first part of the thesis, an existing system of A/Panama plasmids should be re-cloned to change their expression pattern from both vRNA and mRNA to only mRNA of the respective segment. The plasmid system consisted of 8 pHW2000-based plasmids, each of which was equivalent to an A/Panama segment in actual viral infection. Apart from the coding sequence and an ampicillin resistance, these plasmids contained two promoters, a RNA-polymerase I (PolI) and a RNA-polymerase II (PolII) promoter. In addition, a polyadenylation site was present to allow produced mRNA to receive a poly-A tail.

As with typical IAV plasmid systems, each segment is capable of production of vRNA and mRNA from the same coding sequence, making use of the different

59

reading directions of PolI and PolII polymerases. RNA-polymerase II, which is the cellular enzyme producing mRNA copies of DNA templates, is attracted by the PolII promoter and then reads along 3' to 5' direction on the DNA template strand, generating 5' to 3' mRNA. In contrast, RNA-polymerase I normally produces rRNA and runs along 5' to 3' direction on the DNA template strand, in this case producing vRNA from the plasmid system. Usually it is preferable for a plasmid system to provide generation of both kinds of RNA. However, in this case it was required to generate an mRNA-only plasmid system for A/Panama to achieve the possibility to transfect a subset of vRNA species which are complemented by mRNA-only plasmids to allow successful vRNP assembly and export. Such transfections could serve as a useful tool to further investigate IAV genome packaging by the use of MuSeq-FISH which will be described in detail in the next part.

Generation of such plasmids was planned according to a protocol described by Stech et al. [159]. Design of primers was planned so that the PolI promoter would be lacking from the resulting plasmid, but the polyadenylation site would be retained (see Fig. 7). Primers used for cloning were 25 nt long and contained 12 nt complementary to start or end of the coding sequence and remaining 13 nt overlap with the pHW2000 sequence as highlighted in Fig. 7. Effectively, the PolI promoter highlighted in red was removed from the resulting plasmids. Due to high sequence conservation of the different IAV segments, only three primers were sufficient:

- **Fwd PB2/PB1/PA:** GAAGTTGGGGGGG AGCGAAAGCAGG

- **Fwd HA/NP/M/NS:** GAAGTTGGGGGGG AGCGAAAAGCAGG

- **Rev all:** CAGCGAGCTCTAGC AGTAGAAACAAGG

Two cloning steps were required, the first of which amplified the coding sequences flanked by overlapping sequences on the pHW2000 plasmid sequence. PCR was performed as described in the methods part and the resulting PCR products were separated on an agarose gel as depicted in Fig. 8.
Bands were then cut from the gel and subjected to a gel extraction, after which the concentrations were measured (Table 3.1).
In a second step, these products were then used as so-called megaprimers in another QuikChange type PCR reaction, using empty pHW2000 plasmid as template. After DpnI digestion, the resulting plasmids were transformed into *E. coli* and plasmid preparation was performed. Sequencing confirmed that for all segments except the one coding for NA, the cloning process was successful. Therefore the resulting plasmids were ready to be used in transfection experiments.

```
red: polI promoter region
yellow: BsmBI-sites
blue: lacZ
light blue: BGHpA polyadenylation signal sequence
brown-green: overlap megaprimers
```


Figure 7 – **Annotated sequence of a part of the pHW2000 plasmid.** Excerpt of the pHW2000 plasmid sequence and cloning strategy. Primers were designed to hybridize with the ochre-colored parts marked in the sequence, thereby keeping the polyadenylation site (light blue) but removing the PolI promoter region (red). Coding sequence (dark blue) was different depending on the segment coded on the particular plasmid.

Table 3.1 – DNA concentrations of bands cut out from agarose gel.

Segment	Concentration [mg / µl]
PB2	128
PB1	125
PA	107
HA	69
NP	95
NA	92
M	131
NS	123
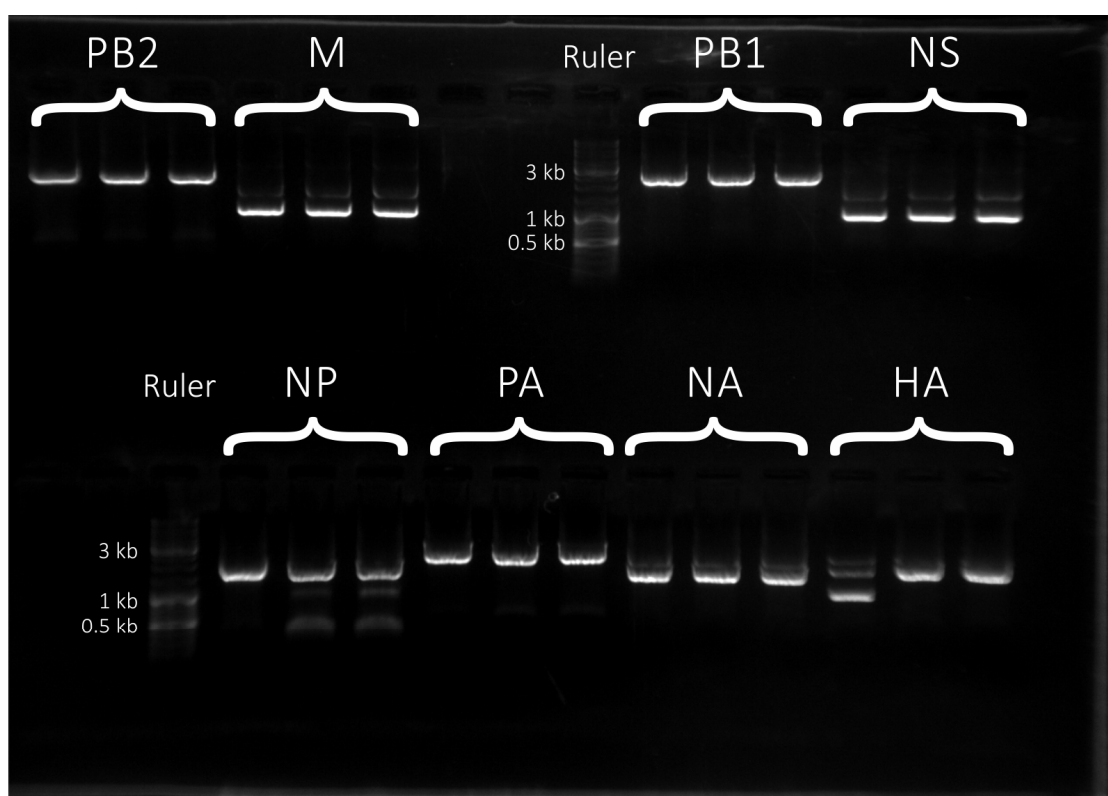
Figure 8 – **Agarose gel of the PCR products during plasmid re-cloning.** Each segment was amplified in triplicates and PCR products were separated on an agarose gel. Megaprimers of all segments could be successfully amplified, although depending on segment type secondary bands of side-products could be detected as well. The left lane for HA segment showed mainly production of a side-product and was not used. For all other lanes, the most prominent band was corresponding to the expected PCR product and was employed.

3.1.2 Microscopy

In a proof-of-concept experiment, the newly created mRNA-only A/Panama plasmid coding for NP was transfected together with the remaining standard plasmids of the A/Panama plasmid system and Rab11-RFP as a control into HEK293 cells. Transfection lasted for 30 h, after which the cells were fixated, permeabilized and antibody staining against NP was performed. In addition, a mock-transfection using only transfection reagent and Rab11-RFP, was conducted in parallel. Results from fluorescence microscopy are shown in Fig. 9.

Transfection of NP-mRNA-only plasmid and standard plasmids + Rab11-RFP
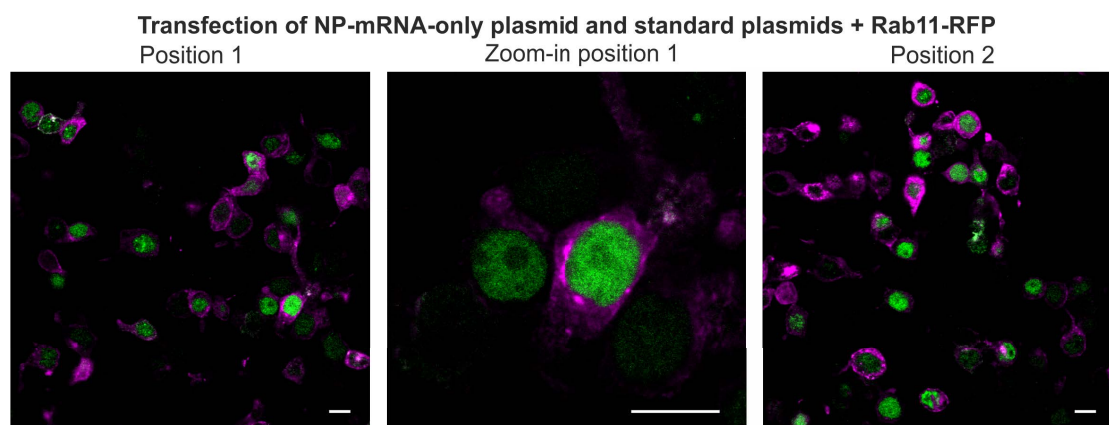Position 1   Zoom-in position 1   Position 2

Figure 9 – **Confocal microscopy of transfected mRNA-only NP plasmid among others in HEK293 cells.** Cells were transfected with A/Panama mRNA-only NP plasmid, all other normal A/Panama pHW2000 plasmids (mRNA and vRNA) and Rab11-RFP. After 30 h incubation, fixation, permeabilization and antibody staining, cells were analyzed under the microscope. α-NP-FITC is shown in green, Rab11-RFP signal is shown in magenta. Not all cells expressed the two signals, and while Rab11 showed a cytosolic localization, NP was restricted to the nucleus. Scale bars 10 µm.

Immunostaining was successful, as confirmed by the positive signal of the control staining of Rab11, which was present both in transfected and mock-transfected samples. We were also able to detect NP in the transfected sample, however only in some cells, indicating that transfection efficiency was below 100%. The presence of NP confirmed that mRNA was correctly transcribed into mRNA from the new mRNA-only plasmid and further translated into protein copies. However, localization of NP was restricted to the nucleus, which is observed when vRNP assembly or export are defective, likely caused by incomplete transfection of the whole plasmid system.

3.2 IAV viral RNA detection using MuSeq-FISH (Panama strain)

The major part of this dissertation was the detection (this section) and analysis (next section) of IAV genome localization and colocalization in infected A549 cells through the introduction of MuSeq-FISH as a novel method to allow comparatively easy multiplexing of FISH data. A brief overview of the method will be given at the beginning. The focus of this work will be on image analysis and modelling of the acquired data sets.

3.2.1 Establishing multiple sequential FISH (MuSeq-FISH)

As described above, ordinary FISH as it is currently routinely used is limited to a multiplexing capability equal to the number of spectrally resolvable fluorescent channels available at the used microscopy setup. Clearly this prohibits the detection of all IAV viral RNAs in the same cells, as there are 8 species of vRNA and additionally 10 species of major mRNA. To overcome this problem, we established multiple sequential FISH (MuSeq-FISH) which allows higher multiplexing by way of sequential stainings. After each FISH staining (which will be referred to as a "run" from now on), images were acquired following standard procedures. However, treating the samples with a high concentration of formamide (80%) in aqueous solution for a 15 min at 37°C, we were able to almost completely remove the bound FISH probes. We then confirmed the removal by microscopy and stained for different targets, but using the same colors as before, which set the beginning of a new FISH run. A typical run within a multi-run MuSeq-FISH experiment lasted 4 h, which means that a complete MuSeq-FISH experiment will be lengthy. As RNA is measured here, the sample quality is endangered by the action of ubiquitous RNAses and this will ultimately limit the number of successful runs which can be performed. In our experiments, we regularly conducted 12 runs over a time of 2-4 days, and sample as well as image quality was still sufficient by the end of it (control experiments are described in the thesis of Ivan Haralampiev [63]). Using only two spectral channels, we were able to perform up to 24 stainings (staining against 8 vRNA segments in 2 colors and additional staining against 8 mRNAs).

3.2.2 vRNA and mRNA - extensive data at 10 h.p.i

Using our MuSeq-FISH approach, we infected A549 cells with A/Panama at an MOI of 5. 10 h p.i., the cells were fixated and immunostaining against NP was performed (FITC-labeled primary mouse antibody). The nuclei were also fluorescently marked by DAPI. Then, FISH staining was conducted as outlined above,

which yielded fluorescent images of all viral mRNAs except for M2 and NS2 and all vRNAs. M2 and NS2 had to be omitted because of the very short sequence of the spliced transcripts, while the M1 and NS1 probes in principle hybridize with their respective mRNAs as well as with unspliced pre-mRNAs. The mRNAs were stained either using Atto 550 dye or PCA Star 635P, while vRNAs were stained twice each to increase robustness of later image analysis.

The resulting images and overlays are depicted in Fig. 10. For the first time, a mapping of abundance and localization of IAV viral RNA in cells has been accomplished. Unusual for FISH, but expected for IAV infections, the cells were densely filled with viral RNA molecules so that high numbers of spots could be seen. All of the cells were infected with A/Panama, which was expected at an MOI of 5.

vRNA - 550

PB2 vRNA	PB1 vRNA	PA vRNA	HA vRNA	Overlay

NP vRNA	NA vRNA	M vRNA	NS vRNA

vRNA - 635P

PB2 vRNA	PB1 vRNA	PA vRNA	HA vRNA	vRNA 550 + 635P

NP vRNA	NA vRNA	M vRNA	NS vRNA

Immunofluorescence and nuclear staining

α-NP	α-NP + vRNA 550	DAPI	DAPI + vRNA 550

viral mRNA

PB2 mRNA	PB1 mRNA	PA mRNA	HA mRNA	Overlay

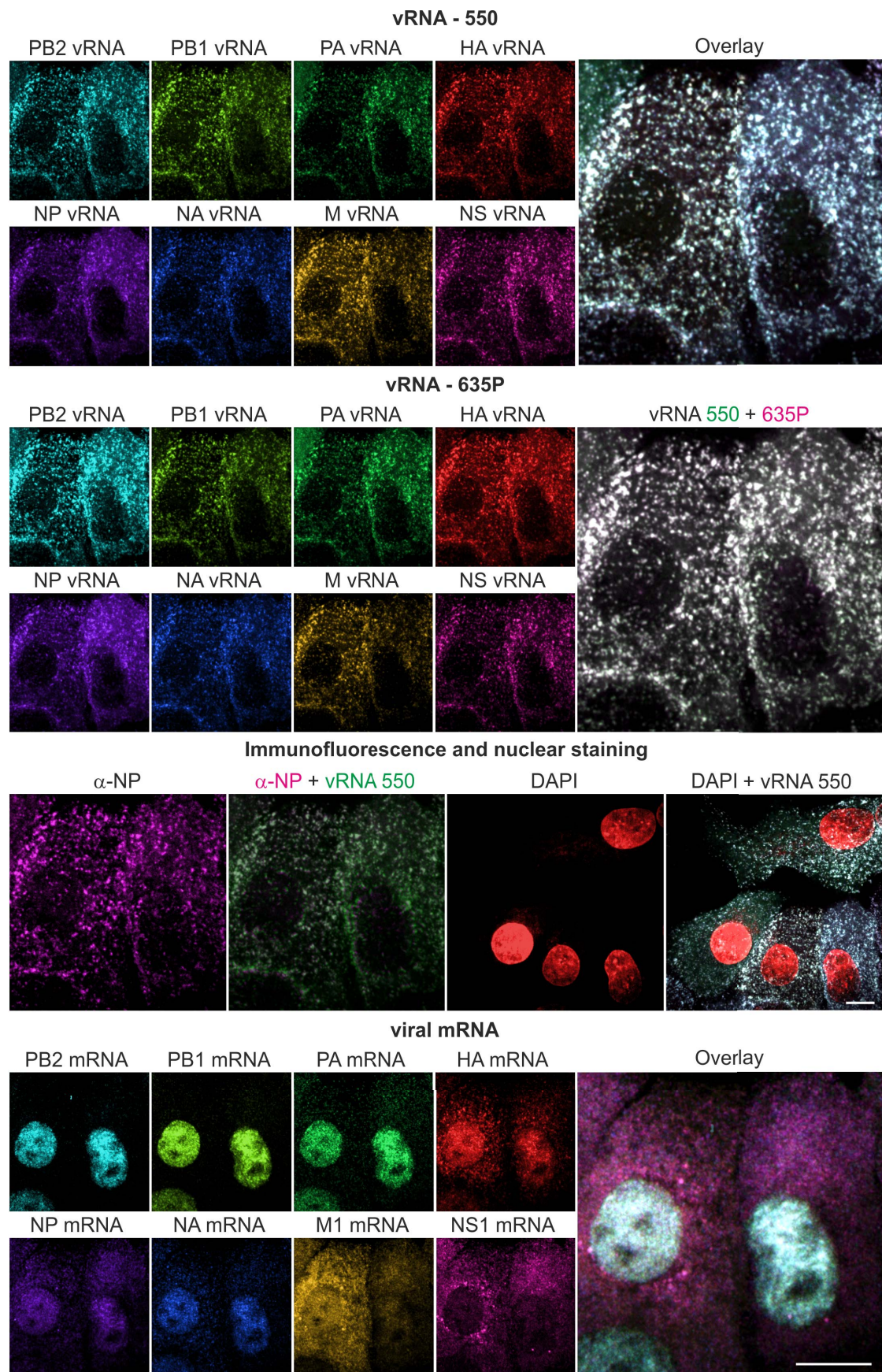| NP mRNA	NA mRNA	M1 mRNA	NS1 mRNA

Figure 10 – **Confocal microscopy of A/Panama infected A549 cells 10 h p.i. and stained for viral vRNA and mRNA.** A549 cells were infected with A/Panama at an MOI of 5 and fixated, permeabilized, DAPI treated and subjected to antibody staining of NP. Subsequently, MuSeq-FISH was conducted, staining all vRNAs and mRNAs. vRNA staining was conducted both using ATTO 550 coupled probe sets as well as PCA Star 635P coupled probe sets (upper two rows). Large numbers of vRNA spots were observed, highly colocalizing among all segments but with a certain cell-to-cell variability visible in the overlay image. Stainings using the second fluorophore validated the results and were in close agreement with the other channel. vRNA localization was mainly cytosolic. mRNA stainings (lower row) were diverse, with PB2, PB1, PA, and to a lesser degree NA showing a predominantly nuclear localization while the other segments were found in both compartments or mainly in the cytosol. Scale bars 10 µm.

All mRNA species except for M2 and NS2 were probed and all of them could be visualized. Signal strength was much weaker compared to vRNA stainings and – while points were still discernible – generally showed a diffuse pattern. We found a primarily – almost exclusive – nuclear localization of the three mRNA segments that encode for the RdRp: PB2, PB1 and PA. The other mRNA species were found both in nucleus and cytosol, while NS1 mRNA seemed to be highly enriched in the cytoplasm.

The vRNA signal was consistently brighter compared to mRNA and shows high degrees of colocalization even when assessed by eyesight. Localization patterns were highly similar across different vRNA segments and cytosolic localization seemed to be preferred.

Both mRNA and especially vRNA stainings exhibited a certain preference for one or more segments in any given cell, which resulted in slightly tinted instead of white colored overlay images (where white indicates high colocalization). Depending on the cell of interest, different such colorings were found.

3.3 Image analysis and deciphering the genome assembly process

Building on the data acquired in the experiments outlined before, digitalization of the data and computer-aided combinatorial analysis were the next steps. This section describes the development and characterization of a colocalization workflow built around a custom-written R script that is capable of colocalization analysis with highly multiplexed data.

3.3.1 Development of an image analysis workflow to assess colocalization

Images were registered using an ImageJ plugin and MetaMorph. Subsequently, spot detection was performed using FISH-quant. Cell outlines were painted and nuclei were marked so that detected spots could be assigned to a specific cell and to nuclear or cytoplasmic origin. FISH-quant is by itself not capable of multiplexing or colocalization analysis (see methods section), so that all downstream analysis was coded in R.

A script was developed that basically reads in multiplexed data and assigns spot from various detections to the segments that were stained. Close proximity of spots to each other was treated as colocalization and interaction, and such spots would then be assigned to a common MSC identifier. Outputs of the analysis were generated as plots using the *ggplot2* package or as Excel sheets using the *xlsx* package.

Using the computational power of the colocalization script, a total of 480 image stacks from an extensive MuSeq-FISH experiment of A/Panama were processed. Using only 69 cells which were evenly illuminated and were of normal shape and appearance, about 385 000 spots were loaded into the colocalization analysis. A total of about 98 000 MSCs, containing one to eight segments resulted from this, the details of which will be shown below.

3.3.2 Spot detection measures

For a first assessment of the colocalization quality, a quantitative measure for this was sought. Any artifacts that would arise form sub-optimal image registration should result in slightly displaced spots of one or more imaging channels (corresponding to IAV segments) to the rest of the data. Thus, the center of mass for each MSC was calculated, see equation below, where N is the number of segments

in the MSC and the index k runs over all segments in the MSC. Y-coordinates were caluclated accordingly, while z-coordinates of spots were omitted for the purpose of this analysis.

$$(Center\ of\ mass)_x = \frac{1}{N} * \sum_k x_k \tag{3.1}$$

Using the center of mass for each MSC, the distance of each spot as well as its 2D angle towards it could be calculated and then plotted in a histogram, as shown in Fig. 11 (analysis of an exemplary microscopy position). Most segments showed a distribution with a marked peak relatively close to the center of mass (25-60 nm distance), with a shoulder towards higher distances up to 250 nm. Segments with a certain offset stemming from sub-optimal registration showed less clear peaks and a distribution shifted towards higher distances. The same holds true for angles towards the center of mass – assuming random organization of spots within an MSC all angles should be populated equally. A strong deviance from this again served as an indication of an offset to the other microscopy channels. All in all, even though one or more segments frequently showed signs of sub-optimal registration, distances were still within the cylinder volume that allowed colocalization to be detected.
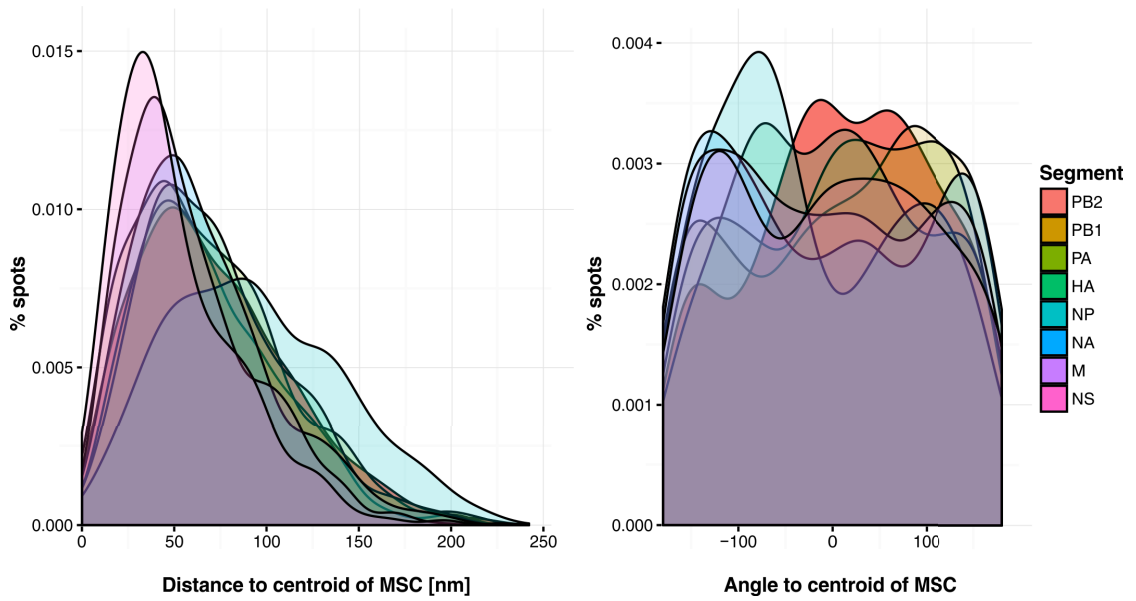
Figure 11 – **Spot detection measures of analyzed images.** All found MSCs for spots detected in A/Panama were analyzed to find their center of mass (centroid) in x and y directions. Then, distributions of distances (left) and angles (right) of spots to this centroid were calculated for each segment. High distances and uneven distributions of angles indicated sub-optimal image registration and a bias towards a certain shift of all spots of this segment in comparison to the other segments. Here, segment 5 (NP) showed the least optimal registration as its distribution was shifted towards higher distances. However, all spots still lay well within the colocalization radius of 300 nm so that no colocalizing spots were lost in the analysis.

3.3.3   Patterns of recognized MSCs

All spots in the resulting data set are still present with their 3D coordinates and cell and nucleus/cytoplasm identifiers. An exemplary data set of a single image position, with nuclear spots excluded, is shown in Fig. 12. As already noted in Fig. 10, the spot density is very high and spots seem to be dispersed all over the cytoplasm. In this representation, it looks like there would be a decrease in spot density towards the edges of cells, however this is mainly due to smaller height (and therefore less volume to contain spots) at the outskirts of cells. However, in several cases increased densities were found at a hotspot structure close to the nuclear membrane which might likely correspond to the microtubule organizing center (MTOC).
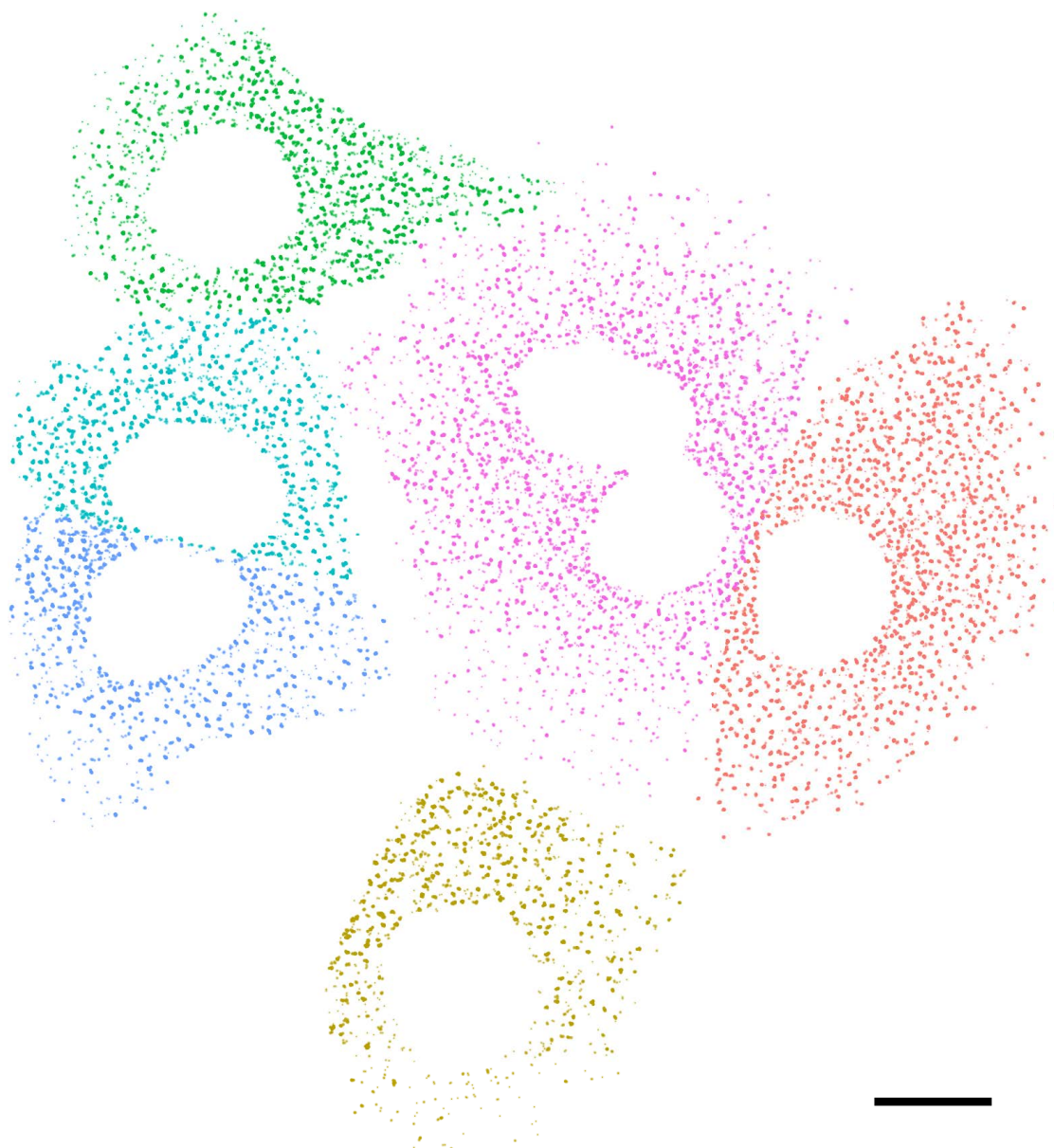
Figure 12 – **Single cell spot clouds recognized by image analysis.** A typical imaging position as seen in its digital representation as spot clouds. A total of 6 cells were imaged for this position and cell outlines were drawn to identify each cell (marked by color). Nuclei were also marked and then excluded for the analysis. Dividing cells (pink) were treated as one cell due to the shared cytosol. Opaqueness of the signal corresponds with colocalization rank (the number of segments in an MSC) of the detected MSCs, so that transparent spots represent MSCs of low colocalization rank. When taking into account that cells are thicker near their nuclei, no obvious patterns of highly colocalized MSCs were observed. Scale bar 10 µm.

3.3.4 Cell-to-cell variability of vRNA abundances

Next, we investigated segment detections at single cell level. We initially expected segment abundancies to be roughly equimolar as this would be the most energy-efficient way of producing a segmented genome that in the end needs to be present in equal amounts. However, we found that in many cases cells varied significantly from such a distribution (Fig. 13). Shares of individual segments were in a range of about 3-20% of total segment counts of cells, with a few cells exhibiting markedly higher shares of segments 2 and 7, corresponding to PB1 and M genes of IAV. However, the mean values (corresponding to a bulk view of the data) of all segments are in a range of 11-14% of total segment count per cell, which matches with an expected value of 12.5% at an equimolar distribution.
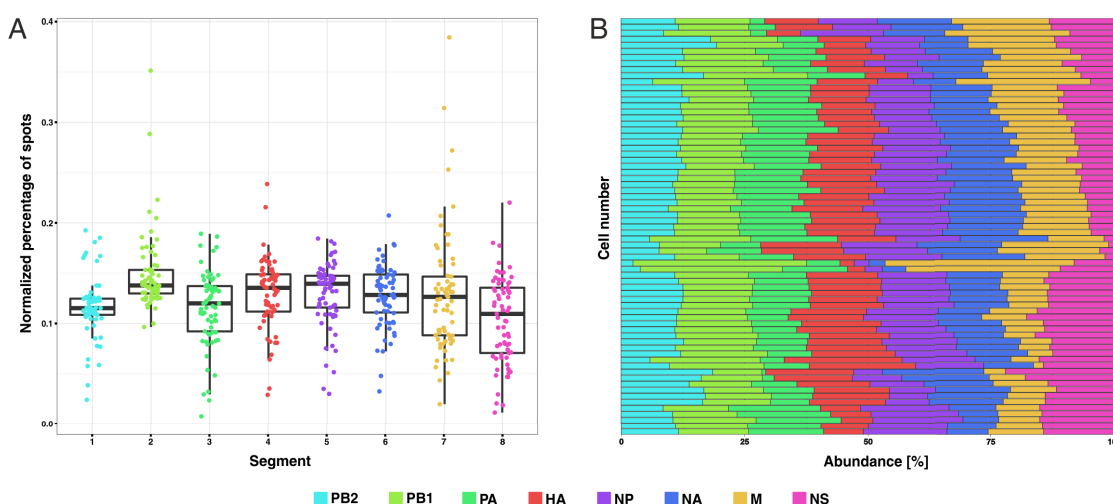

Figure 13 – Cell-to-cell variability of vRNA abundances in A/Panama-infected A549 cells 10 h.p.i. All detected vRNA spots of selected cells were the basis of this representation. (A) Jitter plot showing the variation of genomic content distribution among different cells; each spot corresponds to the share of one segment of the total genome content of the given cell, boxes show the median (thick line) and upper and lower quartile of each data set. The expected average value for equimolar distribution would be 12.5%. Data are highly variable, but median values more or less represent equal expression in bulk. (B) Cells are listed on y-axis and their genomic content is represented by the share of the bar along the x-axis. Most cells showed expression of segments in a more or less even distribution, but some exceptions were observed.

3.3.5 Distribution of MSC ranks

We were also interested in the distribution of MSC ranks, i.e. the amount of segments within an MSC. This question was of major interest to assess whether genome packaging can be followed using such colocalization data – if the vast majority of MSCs was fully packed, no dynamics would be observed.

Colocalization analysis revealed that colocalization ranks of detected MSCs show a U-shaped distribution for A/Panama at 10 h p.i. For the 69 selected cells, about 385 000 points were assigned to MSCs and then binned for their respective MSC ranks (Fig. 14). There was a certain share of completely unpackaged monomeric vRNA segments, to be seen towards the left of the figure. MSCs of intermediate rank, corresponding to genome bundles that started to condensate but did not (yet) finish this process, are also present but seem to play a minor role. Finally, towards the right of the distribution, numbers rise again for completely or almost completely bundled MSCs.
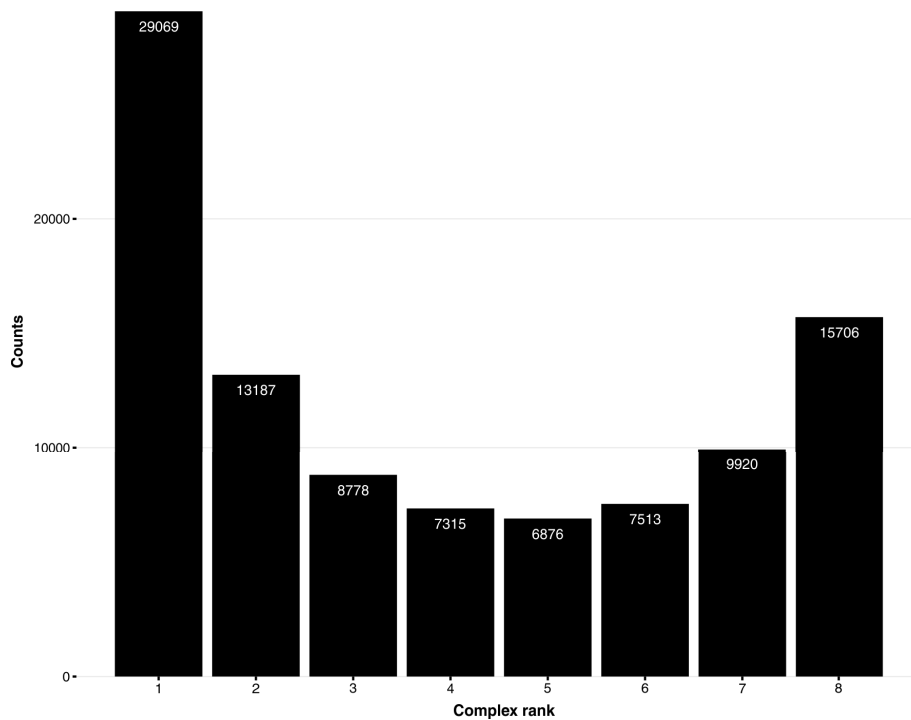

Figure 14 – **Histogram of MSCs detected for A/Panama, grouped by their colocalization rank.** Bars represent the number of MSCs detected among all selected cells. Monomers constitute the highest fraction, directly followed by completely packaged MSCs of rank 8. All in all, a U-shaped distribution emerged which cannot be explained by random colocalization.

3.3.6 Segment availability in MSCs of specific ranks

To find out more about the importance of individual vRNPs within the packaging process, different approaches were followed. One rather basic approach is to analyze the MSCs of a certain rank, as shown in Fig. 15, and look deeper into their genomic content. In a very easy bundling hierarchy, it could be possible that some segments are forbidden in a sterical or other way, or that other segments might be mandatory for the condensation process. Such segments should be highly over- or underrepresented in MSCs of low or high ranks, respectively. Our analysis showed the probability to find a certain segment in MSCs of ranks 1-8, together with a dashed line representing the probabilities in the case of equal chances of incorporation at any stage. Spots from all selected cells were included. It can be seen that most segments follow this line very closely, indicating that this comparatively easy analysis did not shed further light on the question which segments might act as condensation points. The only exceptions from this rule are segments 2 (PB1) and 8 (NS), where PB1 progressed with a preference for incorporation at early stages and NS showed a contrary preference for late incorporation.
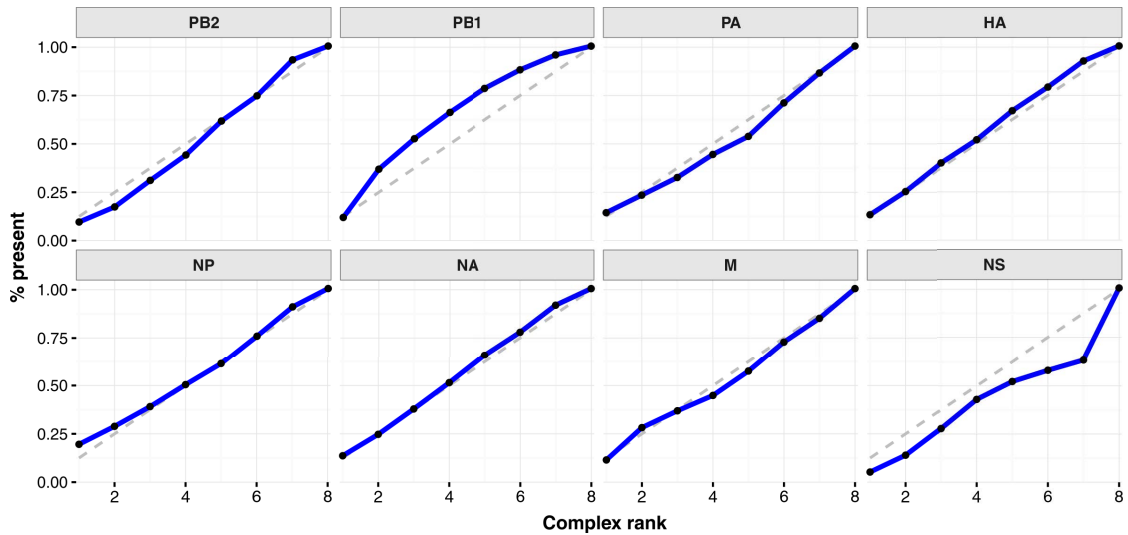
Figure 15 – **MSC inclusion probability of A/Panama segments in dependence of the complex rank.** For all detected MSCs in the A/Panama dataset, probabilities to find a certain segment (facets) in dependence of the complex rank (x-axis) were plotted. The dashed grey line represents the expected value when segment into MSC incorporation probability does not depend on the rank of the existing MSC. Most segments followed this expected line very closely, with the exception of PB1, which was preferred for early MSC incorporation and NS which was less likely to be incorporated into small MSCs.

3.3.7   Network representations of combinatorial data

Next, we followed a network approach to go into more detail of MSC compositions. MuSeq-FISH, contrary to other methods, allows access to full information about the content of such MSCs and reconstruction of a complete genome packaging network. Such a packaging network consists of vertices and edges, where all potential combinations of segments (i.e., the possible MSCs) populate the vertices and condensation reactions between them represent the edges of this graph. To simplify the network and to incorporate known and suspected facts from the literature, a number of assumptions were taken.

- MSCs are only allowed to contain one copy of each segment and the order of segments is irrelevant, i.e. combinations are permutation-insensitive

- Only condensation reactions are allowed, no dissociations

- The order of condensation reactions was restricted to monomeric additions in most cases, except for the iterative simulation where all potential reaction orders with MSCs up to rank 4 were explored

Based on these assumptions, the reaction network consists of 255 vertices and follows a "diamond-shape". Low and high ranks comprise comparatively few potential combinations, whereas rank 4 shows maximal differentiation with 70 potential combinations. Connecting these vertices, 1016 to 3025 edges exist depending on the order of reactions allowed: higher order reactions add to lower order reactions so that the complexity of the network increases.

The experimental data was incorporated into the framework of the network. For vertices, the data is directly accessible from the microscopy data – the values correspond to the number of MSCs that were observed for the given combination. However, edges represent kinetics, which are not immediately extractable from data acquired in fixated cells, i.e. in a frozen state. This issue was addressed by conditional probability modelling (see below), but for the purpose of network generation a simple proxy was implemented to calculate edge data:

$$Weight(edge) = norm.\ weight(vertex\ 1)\ *\ norm.\ weight(vertex\ 2) \qquad (3.2)$$

Following this approach, network graphs were generated to easily visualize the vast amount of data (Fig. 16). Here, the order of reactions was restricted to one (corresponding to reactions in which one of the two educts is strictly monomeric) to retain visual clearness of the output graph. Both vertices and edges are heat-map color-coded and normalized within each rank, so that red coloring corresponds

to vertices and edges which are most elevated within a given rank. All circles corresponding to vertices from a certain MSC rank were grouped together, so that eight such groups are formed (with the last "group" being composed of the single octameric fully bundled genome complex).

Such networks give comparatively easy insights into very complex and multidimensional data. As expected, the highest number of important vertices and edges is found towards the "wide" side of the network at intermediate MSC ranks. Options for reactions are highest in this regime. Towards the sides of the network, at low and high MSC ranks, the picture becomes clearer and specific vertices and edges are of elevated levels.

A notion of importance might be that several important vertices from different MSC ranks are not randomly dispersed, but instead connected to each other in biologically meaningful ways, such that monomeric additions of segments link highly elevated combinations. In fact, such pathways can almost fully be tracked throughout the system – they appear singular towards the beginning and end of the condensation process and become more redundant and splitted in the intermediate MSC rank regime.

Such observations become more obvious when switching to the single cell level. The first graph is a convolution of data gathered from the complete data set of 69 cells and therefore contains high counts, increasing significance. However, the data was acquired with single cell resolution, and the same networks can be examined for such single cells. While vertices are less populated and the amount of noise and uncertainty increases in these situations, we found that single cell networks looked consistently more "tidy", with one prominent "pathway" as a very obvious feature emerging. However, it is important to note that vertices within such prominent "pathways" are only partially conserved, so that some combinations seemed to play an important role in all settings while other combinations were of importance only in a subset of cells or singular examples. Two such exemplary cells are depicted in the lower part of Fig. 16.
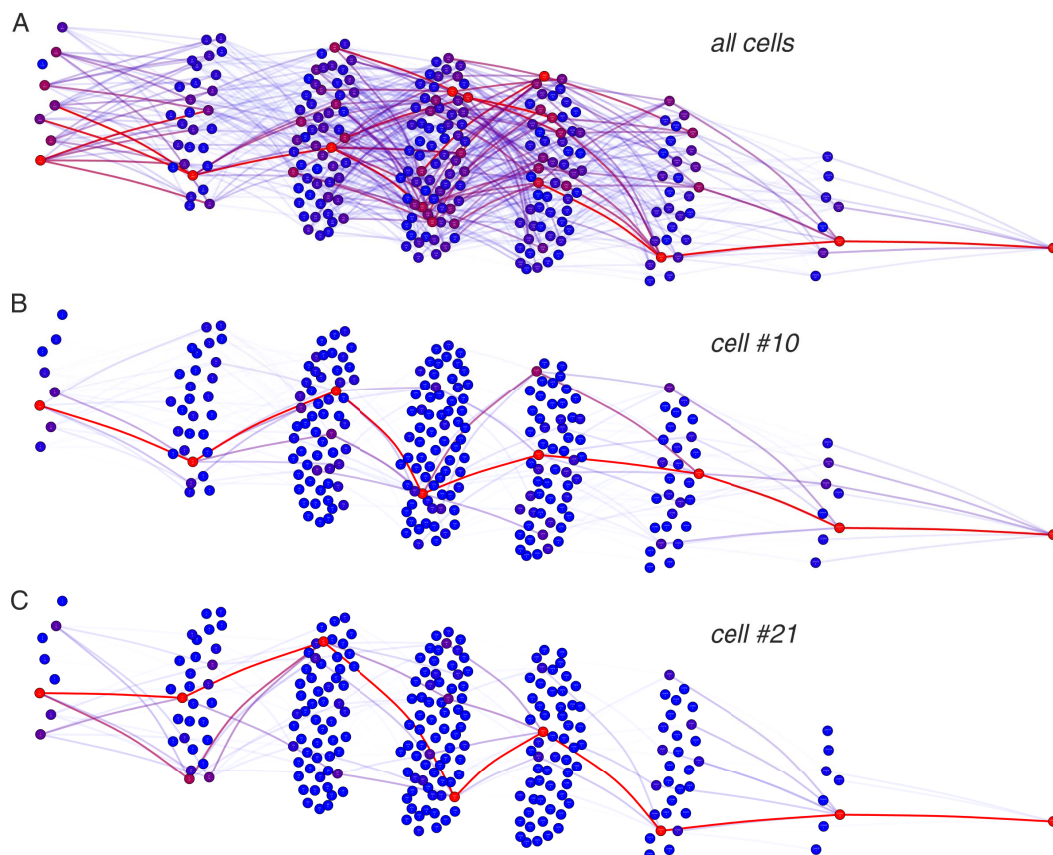
Figure 16 – **The complete packaging network of A/Panama.** Abundances of all combinations found in the detected MSCs were put into a network representation of IAV genome packaging. The network consists of 255 states, i.e. the possible combinations, which are represented as circles. They are grouped by MSC rank, starting with the 8 monomeric segments on the left and ending with a single circle of the fully packaged MSC of rank 8 on the right. 1016 monomeric addition reactions exist between them and are represented by lines connecting the circles. Both combinations and lines are heat-map color coded so that bright red indicates the combination of high abundance within a given MSC rank (high weight). Reaction weights were calculated by multiplying the weights of the adjacent combinations. (A) Packaging network for all A/Panama data. A crowded impression is given, although there are only a few bright red connections on each MSC rank. These connections also usually form a more or less continuous pathway from ranks 1 to 8, with possible redundant side-tracks along the way. (B) and (C) Packaging networks for two individual cells. Both are less crowded than the overall bulk data, indicating that for a given environment in a single cell, a certain pathway of packaging is highly preferred over alternative routes.

3.3.8   Iterative simulation of a random-packaging hypothesis

Interpretation of network graphics is somewhat tampered by the fact that certain segments show higher abundance overall than others. Preferred combinations among MSCs of higher ranks indeed often contained segments of high monomeric abundance. To investigate whether the observed network structure solely stems from different starting conditions (i.e. abundances of the monomeric segments), a simulation was set up to emulate generation of a network starting from monomers. Such a simulation will essentially try to mimic IAV genome bundling following given assumptions.

The simulation was shaped in a way so that it would resemble completely non-specific, or random, genome bundling. In such a situation, all edges would be "weighted" in the same way and all share the same k-value of exactly 1. So the factor driving a network angled towards specific combinations would be solely different starting conditions as we found when looking at the monomeric segments in the data. Implementation of the simulation was done in R, partly making use of the igraph package. The network was initialized by setting the monomeric MSC vertices to values equal to the total abundance found for a certain segment throughout all MSCs, because this represents the actual mass contributed by this segment. Now, in an iterative manner, reactions took place that resulted in mass flow from combinations of lower rank to such of higher rank, solely governed by the rules of mass action. In this case, the assumption of monomeric additions was given up, or at least extended: separate simulations were conducted for reaction orders up to one, two, three and four (four being the highest reaction order conceivable for this given system). In fact, this restriction had almost no effect on the outcome, so that the data shown here is for the least restricted simulation with all reactions allowed. The simulation was run for 107 iterations and after completion, the state with the highest cross-correlation of network values to the microscopy data as a reference network was chosen for further analysis. This meant that the simulated data was as close to the observed situation as possible, given the assumptions stated above and also assuming that reaction preferences are solely determined by mass action.

The resulting network distributions are shown in overlay with the microscopy (reference) network distributions in Fig. 17. Each MSC rank is listed as a separate facet, and all potential combinations of the respective rank are shown, ordered by combination abundance in the reference dataset. To a varying extent, for all MSC ranks the reference dataset shows much higher variability compared to the simulated dataset. Simulated combinations within one MSC rank differ by about a

factor of 2, while the reference dataset shows variation up to a factor of 20. This indicates that an iterative model assuming random packaging, solely governed by kinetics of mass action, is not able to reproduce the observed data.
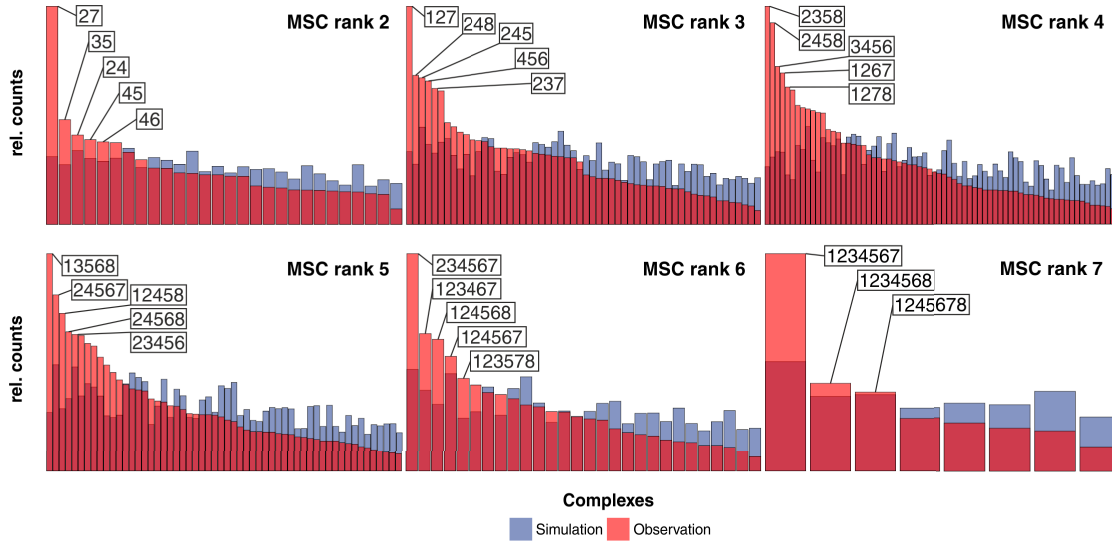

Figure 17 – **Abundances of all observed and simulated segment combinations for A/Panama shown by MSC rank.** Each possible segment combination is depicted by overlaid bars (red = observed data, blue = simulated data) and grouped by MSC rank. Combinations are sorted left to right by abundance as seen in observed (microscopy) data. Simulated data stems from the iterative model assuming reactions governed by mass action (i.e. all k-values = 1). Most abundant combinations within each MSC rank are named. Observed data was found to be of higher variance compared to the simulated dataset, implicating non-random or specific packaging, applicable to all MSC ranks. Segments 2 and 4 seemed to play a dominating role as confirmed by their appearance in many upregulated combinations.

The deviation between observed data and the simulated distribution was present at all MSC ranks, but most pronounced at ranks 2, 3, 6 and 7, which might hint at more restricted packaging at these condensation steps and more redundant packaging at MSC ranks 4 and 5.

3.4 Conditional probability modelling to estimate k-values in the assembly process

All previous visualizations were essentially composed of static data and only indirectly inferred kinetics of the condensation reactions. To further address this issue, conditional probability modelling was designed. The model was implemented in Matlab and split up the complete network into sub-parts of two adjacent MSC rank groups, respectively. Step by step, the complete network was covered with the exception of MSC rank 8 which could not be processed due to technical limitations in the model. Using the abundances of all possible combinations on each MSC rank group, the model fitted variable reaction constants (k-values) for each reaction to achieve the best fit matching our experimental FISH data. Single cells were treated as different training conditions, assuming that each cell uses part, but not necessarily the complete solution space for packaging in terms of actual mass flow. However, k-values should be constant among all cells as we assume that these are specific, at least for a given combination of viral strain and host cell line.

A best fit was achieved and resulted in a set of k-values for all monomeric reactions. Once this dataset was available, it was also possible to follow the network along the most preferred reactions (i.e. the ones with the highest k-value within this MSC rank group). Doing this, an overlay representation of the best 10 paths through the network was plotted. Both the k-values (color-coded in heat map coloring) and these paths are depicted in Fig. 18. The distribution of k-values was of approximate Gaussian shape (data not shown), so that a wide range of condensation reactions should be hypothetically possible, with a subset of those indicating higher preferences. The representation of the best paths was surprising in a way that these paths did not necessarily overlap with the ones that were indicated by the standard network representation. Some of the bright red states were included in these paths, especially in the early stages of genome bundling, but others were omitted. Of course, it has to be kept in mind that this view intentionally tries to simulate an environment where segment availability does not play a role for packaging, i.e. a cell with equal supplies for all segments. This was almost never the case for the cells we investigated, as shown above (Fig. 13).
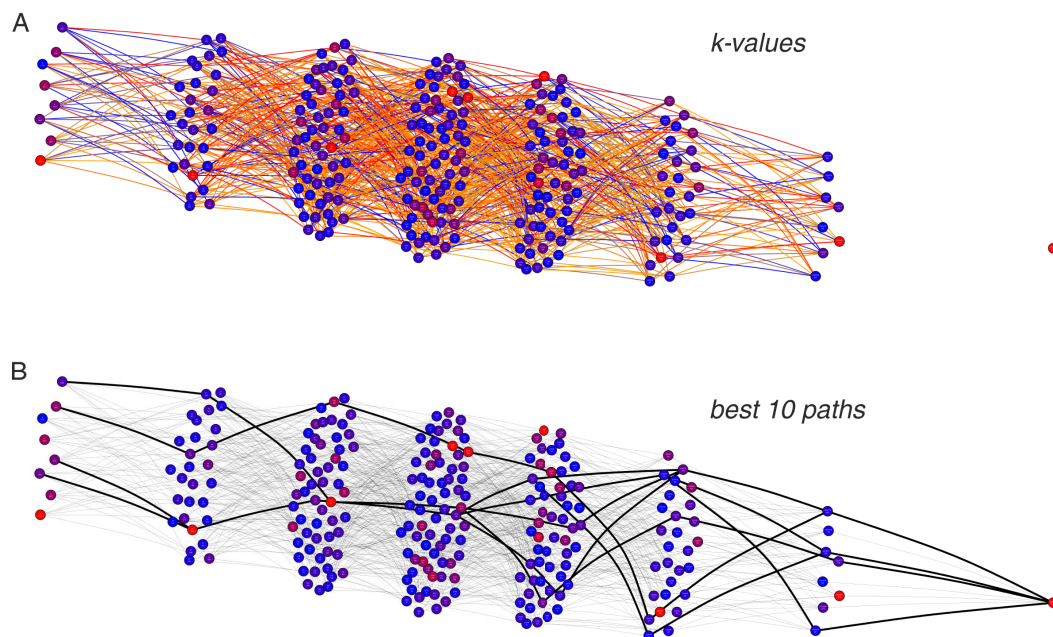
Figure 18 – **Network representation of fitted k-values and best packaging paths for A/Panama based on conditional probability modelling.** Network graphs are identical to Fig. 16, however coloring of monomeric addition reactions was based on results from conditional probability modelling here. Basically, modelling removed the bias of segment availability to identify underlying k-values despite the static nature of the data. (A) Reactions are heat-map color coded to represent the relative size of reactions' k-values within one transition from one MSC rank to the next (red is strongest, orange intermediate and blue lowest). A distribution of a large range of k-values over the whole spectrum was observed, indicating that no single set of clearly preferred reaction pathways exists. (B) 10 best paths in terms of highest k-values were plotted in bold black lines in the network. A limited, yet redundant set of possible reactions emerged. These pathways are not necessarily overlapping with highly abundant combinations (bright red circles).

3.5   Supportive qRT-PCR analysis over the time course of infection

The data obtained from the FISH experiments and analysis described above were complemented by a dataset based on quantitative real-time PCR (qRT-PCR) over the time course (4-12 h) of infection of IAV strain A/Panama in A549 cells at an MOI of 5 (see Fig. 19). All experiments were conducted by bachelor student Kalle Jongen (AG Molecular Biophysics, HU Berlin, Germany). Data analysis was performed by Kalle Jongen with advice and scripting by Fabian Jolmes (AG

Molecular Biophysics, HU Berlin, Germany) [79].

To generate the data, A549 cells were infected with A/Panama virus and after the respective times of infection subjected to RNA extraction. This RNA was then subjected to specific reverse transcription into DNA, dilutions of which served as samples for the actual qRT-PCR experiments. Primers were designed by Ivan Haralampiev and targeted specific parts of vRNA and mRNA sequence of all IAV segments. Standard curves were obtained using pHW2000-based plasmids carrying the sequence of one A/Panama segment each, which were provided by the group of Thorsten Wolff, RKI Berlin. Biological triplicates were used in all experiments and each triplicate was measured three times in qRT-PCR, thereby also generating technical triplicates.

The vRNA of all segments was found to be monotonously increasing in abundance over the time course of infection. The general trend was very similar across all of them, even though segments 3 (coding for PA), 5 (coding for NP) and 2 (coding for PB1) showed elevated levels compared to the other segments.

Levels for all mRNA species were constantly higher compared to the respective vRNA species, except for segment 3 where vRNA was more abundant. The mRNAs of the RdRp subunits showed lower abundance in comparison to the other segments, which matched the observation during FISH measurements that those mRNAs were predominantly found within the nucleus. Most segments show increasing mRNA content until 10 h p.i. and a decline towards the last measured time point at 12 h p.i. Results of the RT-PCR experiments have been described in detail in the bachelor thesis of Kalle Jongen.
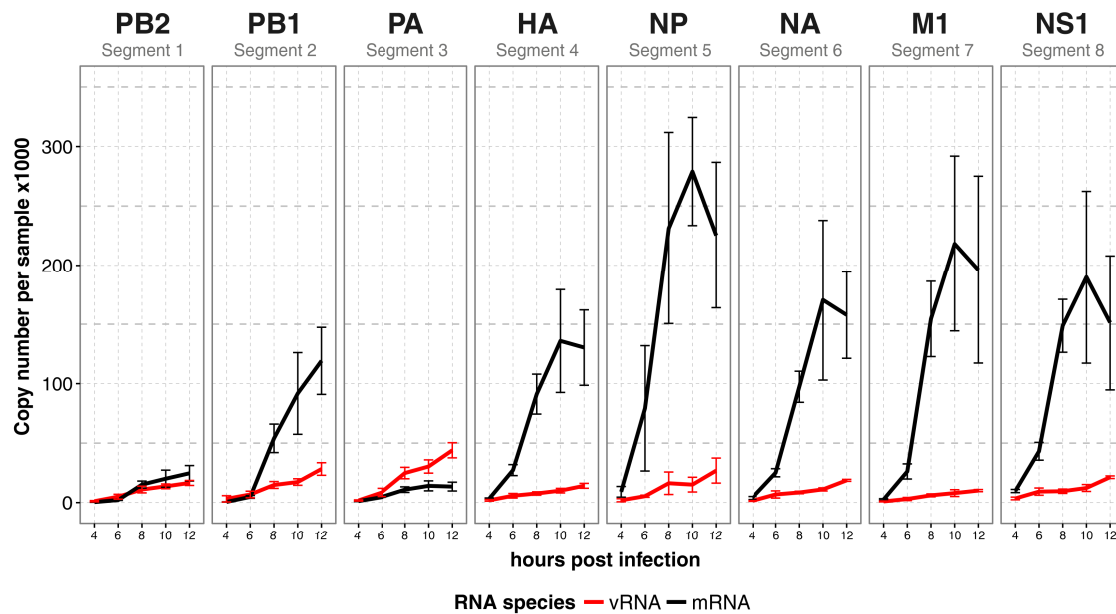
Figure 19 – **qRT-PCR measurements of A/Panama vRNA and mRNA abundances over the time course of infection.** A549 cells were infected with A/Panama at an MOI of 5 for different time spans. Total RNA was extracted, followed by reverse transcription into DNA using either vRNA- or mRNA-specific primers. RNA content was then measured using qRT-PCR, where pHW2000 plasmids of all segments served as standard curves. Experiments were conducted with three biological and three technical replicates each. Abundances of mRNA (black) and vRNA (red) over a time course of 4 to 12 h p.i. are shown. mRNA expression was higher than vRNA expression except for segment 3 (PA). All curves show growth over time, except for a decrease from 10 to 12 h p.i. of HA, NP, NA, M1 and NS1 mRNAs. Error bars represent standard error of the mean.

3.6   Comparison of the acquired data with reassortant strains of A/Panama and A/Mallard

Besides the A/Panama studies, we extended our analysis to other viral strains to investigate the question of genome packaging conservation during evolution. Namely, we used the avian IAV strain A/Mallard (H3N2) and three reassortant viruses composed of segments both from A/Panama and A/Mallard. A reassortant virus containing segments 1-6 and 8 originating from A/Panama and segment 7 (coding for the M1 and M2 proteins) originating from A/Mallard was named Pan-M; another virus with essentially the A/Panama genome but segment 8 (coding for the NS1 and NS2 proteins) from A/Mallard was named Pan-NS. Additionally, a double reassortant virus with segments 1-6 from A/Panama and 7 and 8 from A/Mallard, named Pan-NSM, was measured. Fig. 20 gives an overview of names and genomic composition of the five viral strains that were used in this work.
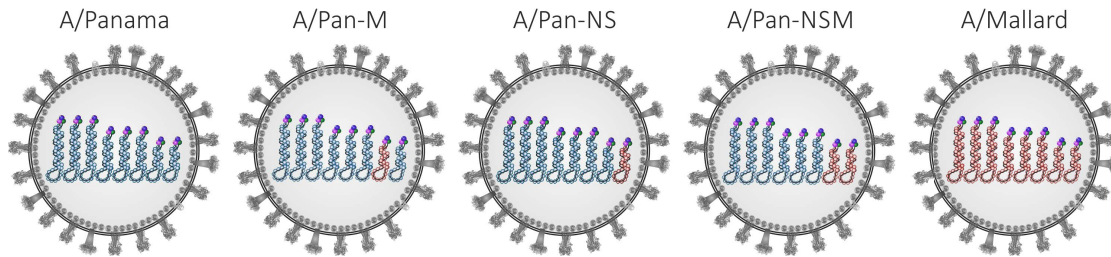

Figure 20 – **Overview of the viral strains used in this study.** In total, 5 viral strains were used in the work presented here. A/Panama (left) is a human H3N2 virus, A/Mallard (right) is an avian H3N2 virus. A/Pan-M, A/Pan-NS and A/Pan-NSM are reassortant viruses containing A/Mallard genome segments in the environment of A/Panama. The replaced segments are 7 for A/Pan-M, 8 for A/Pan-NS and 7+8 for A/Pan-NSM. Genomic segments originating from A/Panama are marked in blue in this representation, segments originating from A/Mallard are marked in red.

Microscopy and subsequent analysis for all viral strains was performed exactly as described above for A/Panama. However, different probe sets had to be used for vRNA and mRNA stainings of A/Mallard due to lack of conservation in the genomic sequences of the two viral strains. In the case of Pan-M, Pan-NS and Pan-NSM a mixture of the probe sets designed for A/Panama and A/Mallard was used, so that the probes used were corresponding to the parental strain for the segment of interest.

FISH stainings for A/Mallard and all reassortant strains were performed using redundant double stainings of vRNA with both fluorophores ATTO 550 and PCA Star 635P and single stainings of mRNA using a mixture of the fluorophores. However, only ATTO 550 stainings of vRNA were used in further colocalization

analysis due to inferior performance of some A/Mallard specific probe sets coupled to PCA Star 635P.

qRT-PCR was conducted to yield information about abundances of vRNA and mRNA during the time course of infection and was performed as described for A/Panama. The measurements were restricted to segments 5 (NP), 7 (M) and 8 (NS) due to availability of plasmids serving as standard curves. All RT-PCR experiments described in this part were conducted by Vanessa Körner (AG Molecular Biophysics, HU Berlin, Germany) [86].

3.6.1   A/Mallard

For the avian virus A/Mallard (H3N2), three independent experiments were conducted from which 54 evenly illuminated cells of normal appearance were selected for spot analysis. About 200 000 spots were detected in total, which could be colocalized into 85 000 MSCs.

Mapping all viral RNA species in A/Mallard

As for A/Panama, we were able to successfully stain all viral vRNA and mRNA using our MuSeq-FISH approach (Fig. 21. The probe set for A/Mallard was therefore validated. Similar to our previous results, we again saw spot-like structures for both RNA species distributed throughout the cells, but with a preference for cytosolic localization in case of vRNA. In the case of mRNA, the polymerase segments PB2, PB1 and PA, and to a lesser extent NA showed high abundance in the nucleus, although this effect was attenuated compared to A/Panama and cytosolic spots appeared for all of them as well. We also noted a certain tendency for large and bright areas larger than the usual spot size which occurred preferentially in some stainings and in certain parts of cells, most notably in close proximity of the nucleus which likely represented the MTOC.
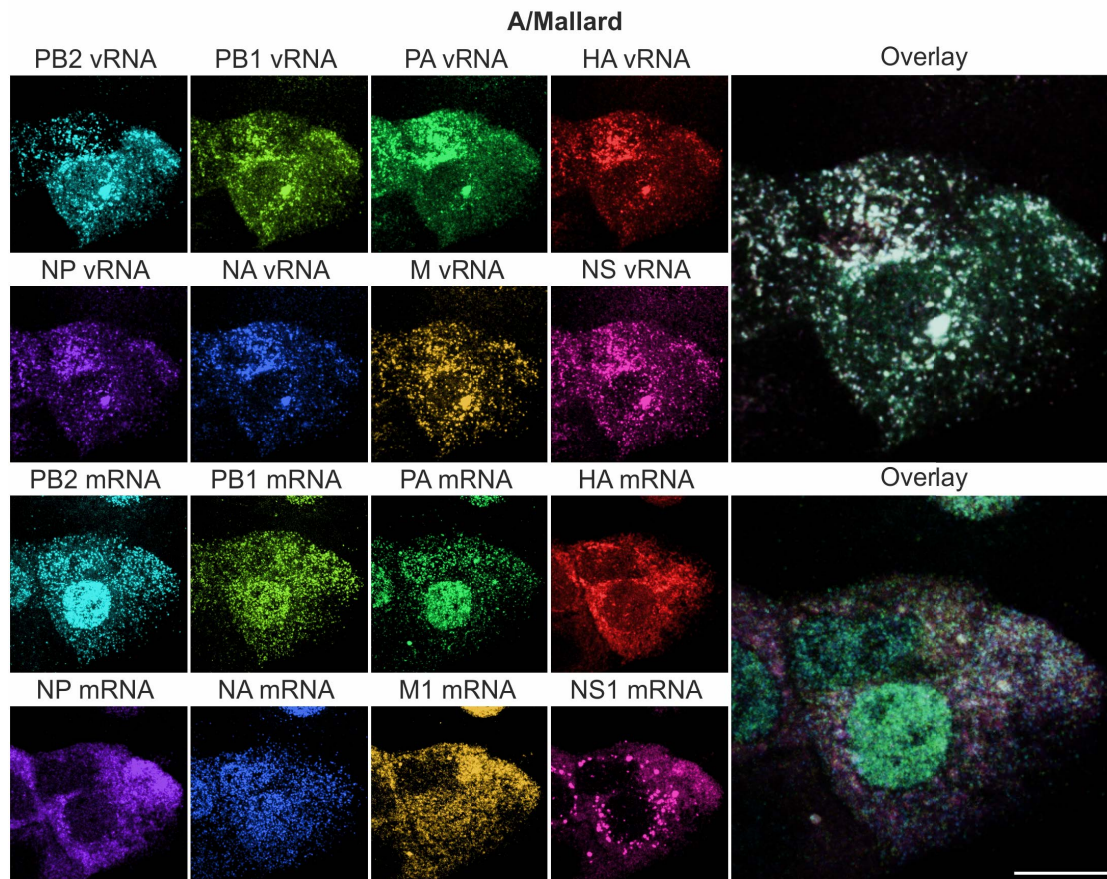
Figure 21 – **Confocal microscopy of A/Mallard infected A549 cells 10 h p.i. and stained for viral vRNA and mRNA.** A549 cells were infected with A/Mallard at an MOI of 5 and fixated, permeabilized, DAPI treated and subjected to antibody staining of NP. Subsequently, MuSeq-FISH was conducted, staining all vRNAs and mRNAs. vRNA staining was conducted using ATTO 550 coupled probe sets (upper row). Large numbers of vRNA spots were observed, highly colocalizing among all segments but with a certain cell-to-cell variability visible in the overlay image. mRNA stainings (lower row) were diverse, with PB2, PB1, PA, and to a lesser degree NA showing a predominantly nuclear localization while the other segments were found in both compartments or mainly in the cytosol. Scale bar 10 µm.

Cell-to-cell variability of vRNA abundances

In the case of A/Mallard, the cell-to-cell variability of vRNA abundance turned out to be higher compared to A/Panama (see Fig. 22, compare Fig. 13). Segments 2 and 7 were attenuated in most cells, while segments 4, 5 and 8 were highly expressed. However, variance of the data was also higher compared with A/Panama, to a point where some cells expressed no or nearly no copies of a certain segment. A small number of cells was dominated by expression of one segment, comprising up to 40% of all spots in such cases.
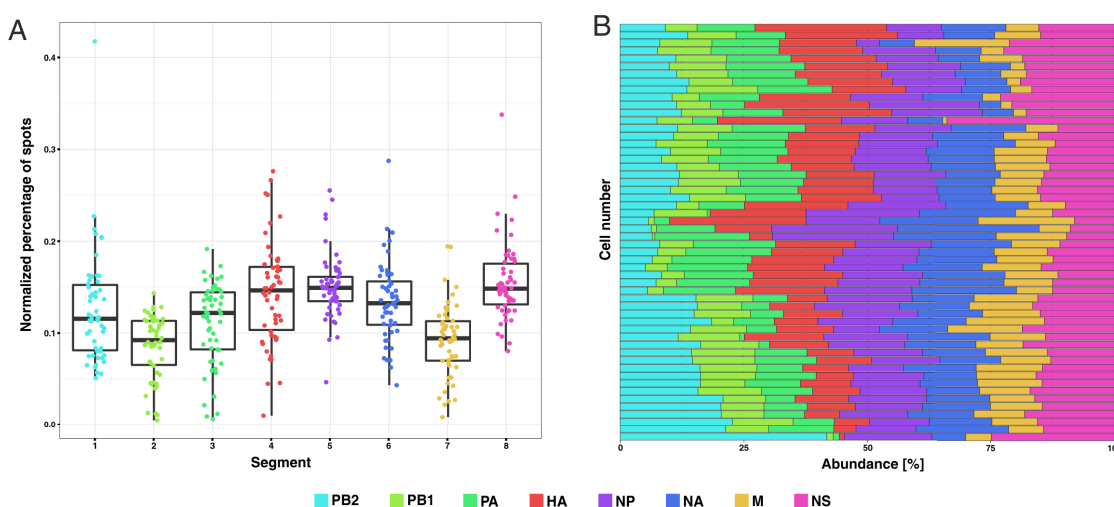

Figure 22 – **Cell-to-cell variability of vRNA abundances in A/Mallard-infected A549 cells 10 h.p.i.** All detected vRNA spots of selected cells were the basis of this representation. (A) Jitter plot showing the variation of genomic content distribution among different cells; each spot corresponds to the share of one segment of the total genome content of the given cell, boxes show the median (thick line) and upper and lower quartile of each data set. The expected average value for equimolar distribution would be 12.5%. Data are highly variable, but median values more or less represent equal expression in bulk. However, outlier cells are present and in general, segments 4, 5 and 8 are upregulated while segments 2 and 7 are less expressed. (B) Cells are listed on y-axis and their genomic content is represented by the share of the bar along the x-axis. Some cells showed expression of segments in a more or less even distribution, but others showed irregular expression, dominated by only one or two segments.

Distribution of MSC ranks

We also analyzed overall colocalization for the microscopy data of A/Mallard (see Fig. 23). Similar to the results of A/Panama, the monomeric spots for which no

colocalization was detected constituted the largest fraction. However, in contrast to the U-shaped distribution observed for A/Panama, A/Mallard exhibited a distribution resembling an exponential decay. Colocalization was less present and no preference for completely packaged genome bundles was observed.
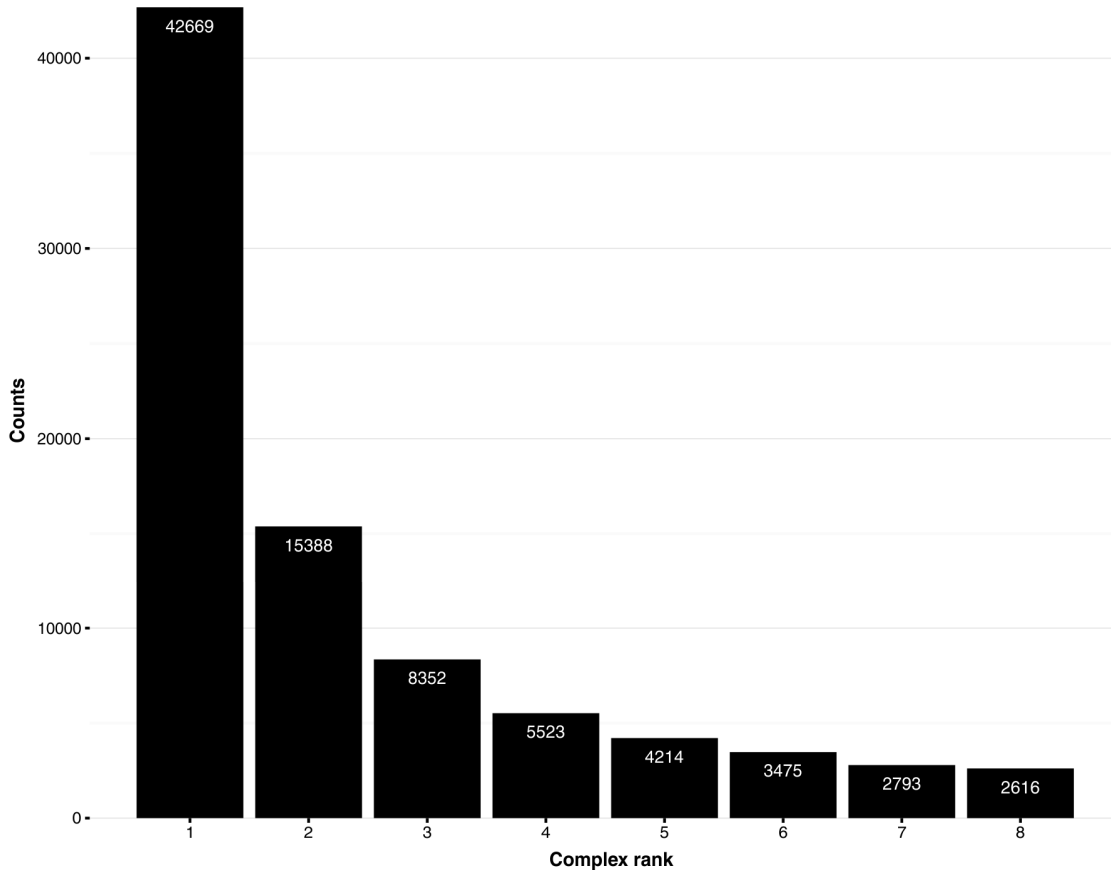
Figure 23 – **Histogram of MSCs detected for A/Mallard, grouped by their colocalization rank.** Bars represent the number of MSCs detected among all cells with a certain rank. Monomers constitute the highest fraction, and numbers decline towards higher ranks in a fashion resembling exponential decay. Values do not necessarily prove specific packaging, but imply much stronger colocalization than expected from random distribution.

Network representations of combinatorial data

Network analysis was performed as described in detail for A/Panama (Fig. 24). Again, the network graph for all cells looked rather crowded, with many connections of intermediate strength visible and some strong paths discernible. Strong connections tended to be in complete graphs, however many side-paths and a high

degree of overall redundancy was detected. Graphs for two exemplary single cells looked less crowded, but still a multitude of paths seemed to play a role, more than what we saw for A/Panama.

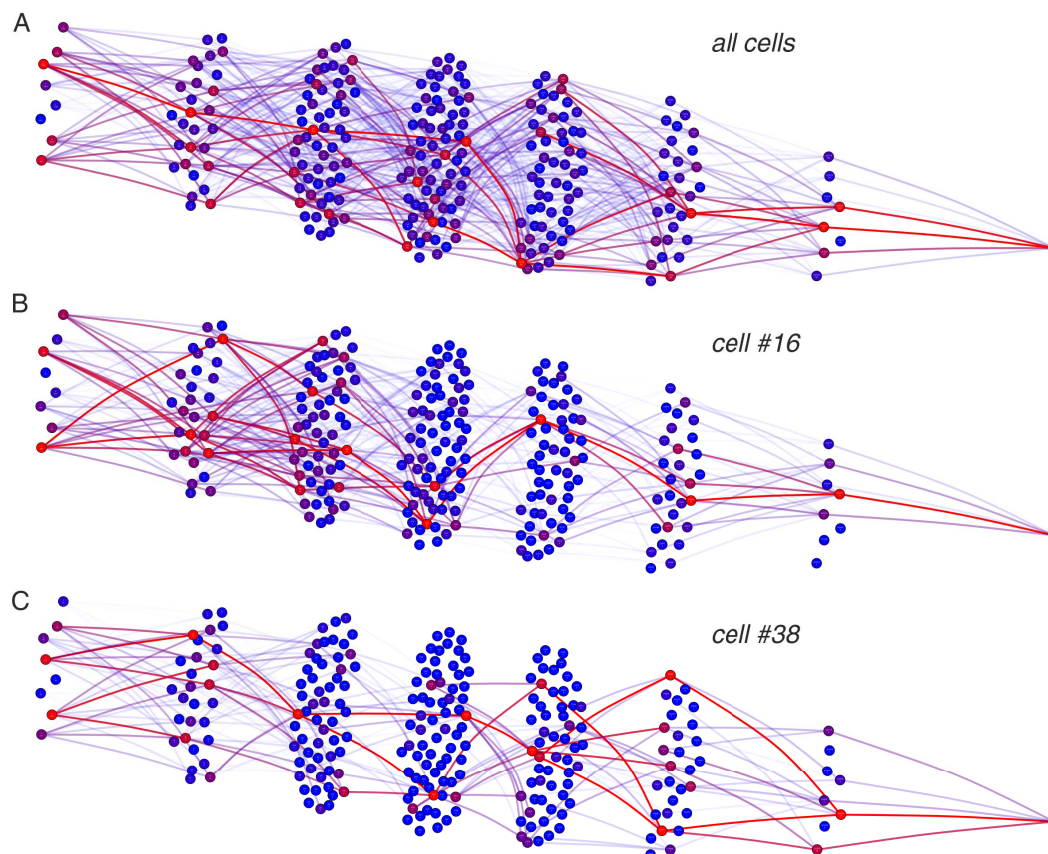**Figure 24** – **The complete packaging network of A/Mallard.** Abundances of all combinations found in the detected MSCs were put into a network representation of IAV genome packaging. The network consists of 255 states, i.e. the possible combinations, which are represented as circles. They are grouped by MSC rank, starting with the 8 monomeric segments on the left and ending with a single circle of the fully packaged MSC of rank 8 on the right. 1016 monomeric addition reactions exist between them and are represented by lines connecting the circles. Both combinations and lines are heat-map color coded so that bright red indicates the combination of highest abundance within a given MSC rank (high weight). Reaction weights were calculated by multiplying the weights of the adjacent combinations. (A) Packaging network for all A/Mallard data. A highly crowded impression is given, with many bright connections between all MSC ranks. However, some of these connections also form a more or less continuous pathway from ranks 1 to 8, with possible redundant side-tracks along the way. (B) and (C) Packaging networks for two individual cells. Both look less crowded, indicating that for a given environment in a single cell, certain pathways of packaging are preferred over alternative routes.

Iterative simulation of a random-packaging hypothesis

Again, we performed our iterative simulation to test against pure random packaging, depicted in Fig. 25. While the distribution of observed combinations was more homogenous compared to A/Panama, and simulated distributions showed increased variance caused by stronger segment availability bias in single cells, we still observed that experimental and simulated data showed different distributions among most MSC ranks, most notably 4 and 5. For other MSC ranks, e.g. 2 and 7, no clear distinction to the simulation was visible so that we could not prove that specific packaging plays a role at these steps in the packaging process. All in all, in the case of A/Mallard, segment availability seemed to be a major determinant of the order of condensation reactions, which might outshine underlying packaging preferences.
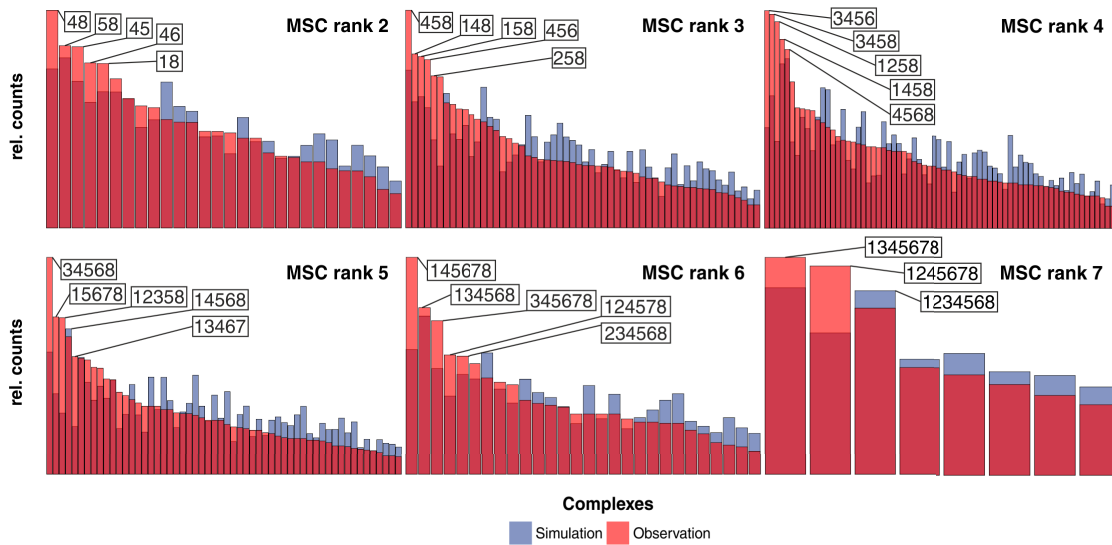
Figure 25 – **Abundances of all observed and simulated segment combinations for A/Mallard shown by MSC rank.** Each possible segment combination is depicted by overlaid bars (red = observed data, blue = simulated data) and grouped by MSC rank. Combinations are sorted left to right by abundance as seen in observed (microscopy) data. Simulated data stems from the iterative model assuming reactions governed by mass action (i.e. all k-values = 1). Most abundant combinations within each MSC rank are named. Observed data was found to be of higher variance compared to the simulated dataset, implicating non-random or specific packaging, applicable at least for MSC ranks 3-5. However, differences are much smaller compared to A/Panama. Segments 4 and 8 seemed to play a dominating role as confirmed by their appearance in many upregulated combinations.

Conditional probability modelling

Conditional probability modelling to fit k-values to the monomeric addition reaction was performed as described for A/Panama. Again, the fit itself was not very concise, rather suggesting that a broad range of reactions are hypothetically possible (Fig. 26). However, selection of the 10 best paths in terms of k-values was markedly different from the situation in A/Panama and moreover showed a single preferred pathway up to MSC rank 4. In later condensation steps, two parallel, yet interconnected pathways emerged.
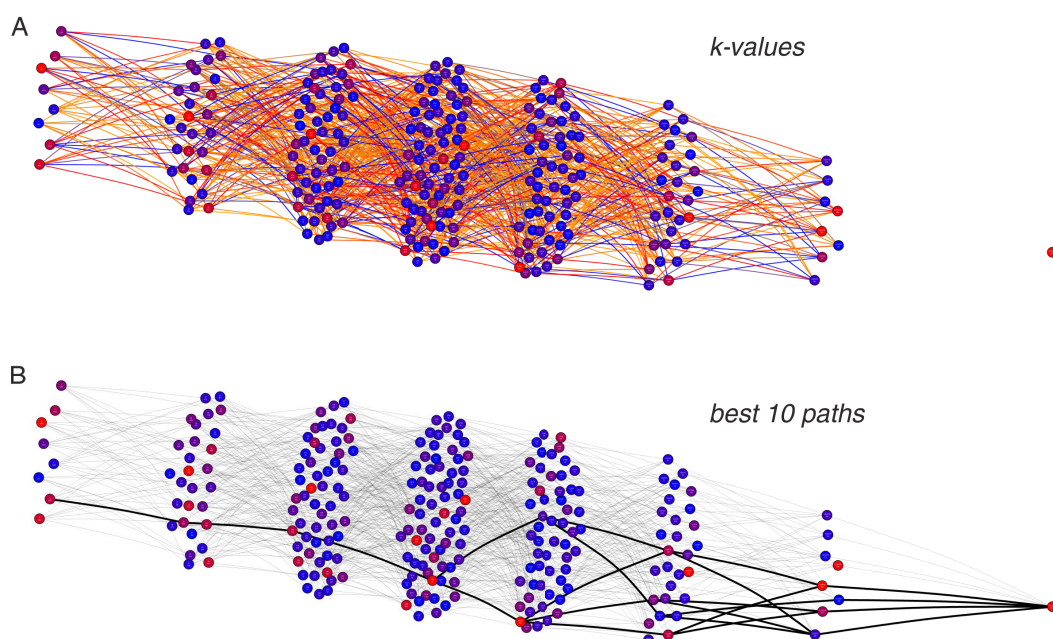
Figure 26 – **Network representation of fitted k-values and best packaging paths for A/Mallard based on conditional probability modelling.** Network graphs are identical to Fig. 24, however coloring of monomeric addition reactions was based on results from conditional probability modelling here. Basically, modelling removed the bias of segment availability to identify underlying k-values despite the static nature of the data. (A) Reactions are heat-map color coded to represent the relative size of reactions' k-values within one transition from one MSC rank to the next (red is strongest, orange intermediate and blue lowest). A distribution of a large range of k-values over the whole spectrum was observed, indicating that no single set of clearly preferred reaction pathways exists. (B) 10 best paths in terms of highest k-values were plotted in bold black lines in the network. A limited, yet redundant set of possible reactions emerged. These pathways are not necessarily overlapping with highly abundant combinations (bright red circles).

qRT-PCR data

qRT-PCR data of vRNA as well as mRNA abundances was obtained as described above (see Fig. 27). Copy numbers were strongly reduced compared to A/Panama. mRNA was consistently more abundant than the respective vRNA counterpart, however the difference of the two species was less pronounced than in A/Panama. Curves show a roughly linear increase in abundance apart from a stronger increase in M vRNA and a plateau phase for NS1 and NS2 mRNAs.
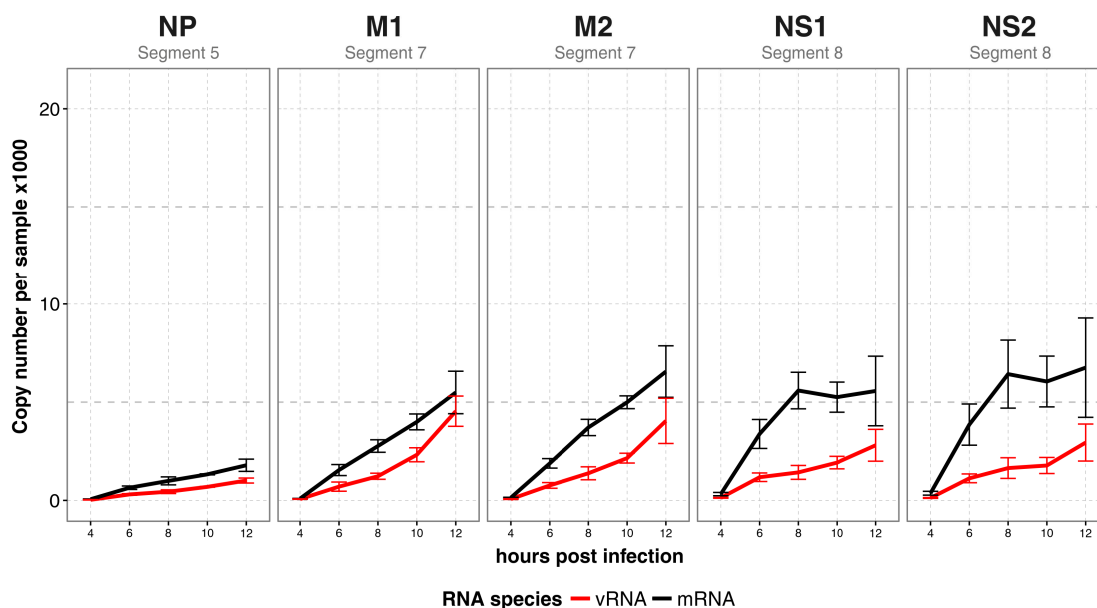
Figure 27 – **qRT-PCR measurements of A/Mallard vRNA and mRNA expression over the time course of infection.** A549 cells were infected with A/Mallard at an MOI of 5 for different time spans. Total RNA was extracted, followed by reverse transcription into DNA using either vRNA- or mRNA-specific primers. RNA content was then measured using qRT-PCR, where pHW2000 plasmids of all segments served as standard curves. Availability of such plasmids limited analysis to the segments shown here. Experiments were conducted with three biological and three technical replicates each. Abundances of mRNA (black) and vRNA (red) over a time course of 4 to 12 h p.i. are shown. mRNA expression was higher than vRNA in all cases. All curves show growth over time, except for a saturation level from 8 to 10 h p.i. for NS1 and NS2 mRNAs. Error bars represent standard error of the mean.

3.6.2 A/Pan-M

Mapping all viral RNA species in A/Pan-M

First, A549 cells were infected with the reassortant virus A/Pan-M at an MOI of 5 and infection was allowed to proceed for 10 h. After fixation and permeabilization, MuSeq-FISH was conducted, basically using the probe sets designed for A/Panama with the exception of vRNA and mRNA of segment 7, where the A/Mallard probe set was used.

MuSeq-FISH yielded successful stainings of all viral RNAs (Fig. 28). We observed strong vRNA signals and weaker mRNA signals overall. vRNA was almost not present in the nuclei and PB2 was strongly expressed in all cells, which is also reflected by the coloring of the overlay image. On the mRNA side, there was a strong tendency for the RdRp coding segments (1-3) to be localized in the nuclei of cells, while this could be seen much less for NA, a segment that showed similar behavior in some other viral strains. The other mRNAs were predominantly found in the cytosol.
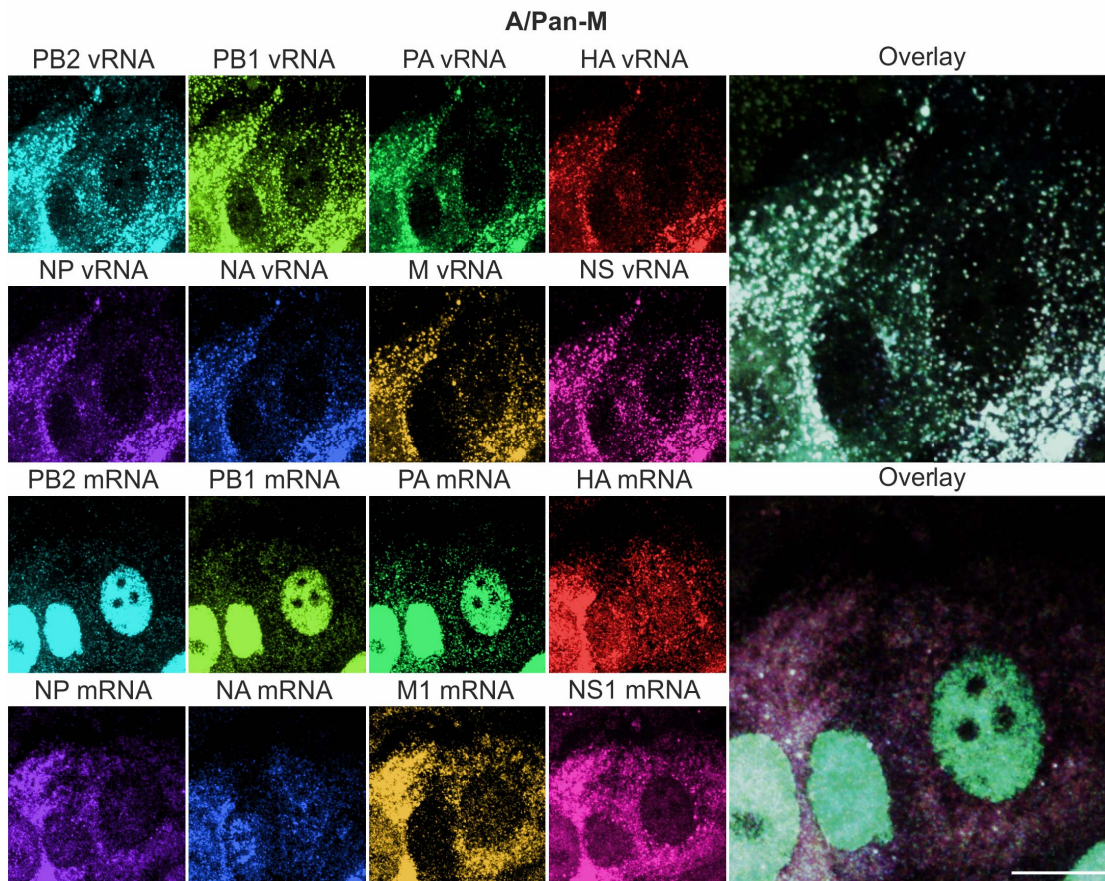
Figure 28 – **Confocal microscopy of A/Pan-M infected A549 cells 10 h p.i. and stained for viral vRNA and mRNA.** A549 cells were infected with A/Pan-M at an MOI of 5 and fixated, permeabilized, DAPI treated and subjected to antibody staining of NP. Subsequently, MuSeq-FISH was conducted, staining all vRNAs and mRNAs. vRNA staining was conducted using ATTO 550 coupled probe sets (upper row). Large numbers of vRNA spots were observed, highly colocalizing among all segments with no apparent cell-to-cell variability visible in the overlay image. mRNA stainings (lower row) were diverse, with PB2, PB1, PA, and to a lesser degree NA showing a predominantly nuclear localization while the other segments were found in both compartments or mainly in the cytosol. Scale bar 10 µm.

Cell-to-cell variability

For A/Pan-M, the variability of vRNA segments within single cells was found to be dominated by strong expression of PB2 (see Fig. 29), which in one exceptional case comprised almost the complete genomic content. Even when ignoring this extreme cell, a high number of cells contained high amounts (15-65%) of PB2. In contrast, segments 2, 3, 4 and 7 were found to be consistently underrepresented. The remaining segments were produced in roughly the expected amounts (one eighth of total genomic content). Notably, facet B in the figure also indicates that PB2 and NP are mostly high expressed when the other is of lower expression and vice-versa.
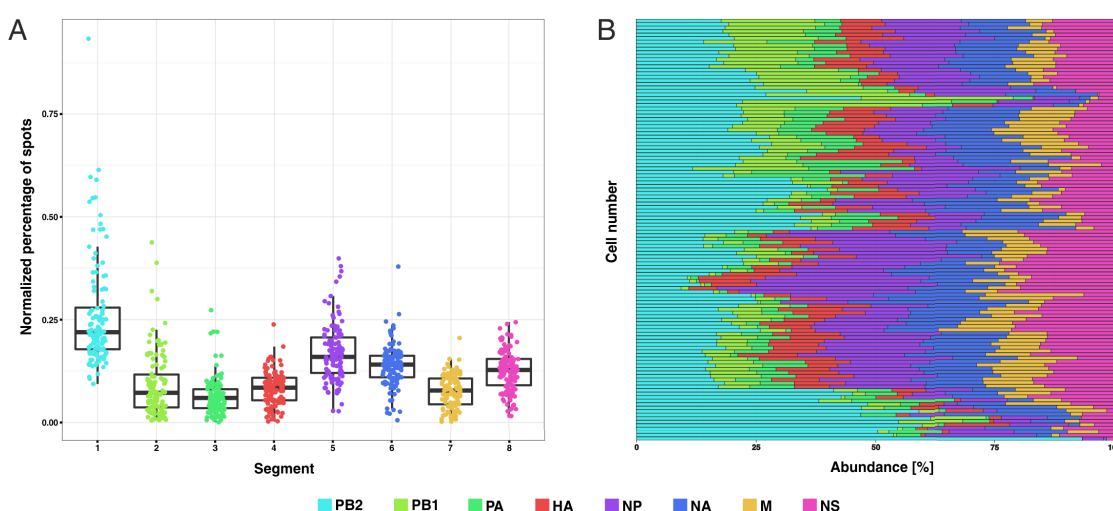
Figure 29 – **Cell-to-cell variability of vRNA abundances in A/PanM-infected A549 cells 10 h.p.i.** All detected vRNA spots of selected cells were the basis of this representation. (A) Jitter plot showing the variation of genomic content distribution among different cells; each spot corresponds to the share of one segment of the total genome content of the given cell, boxes show the median (thick line) and upper and lower quartile of each data set. The expected average value for equimolar distribution would be 12.5%. Data are highly variable, with segment 1 highly expressed and segments 5 and 6 also showing higher than average expression. Segments 2, 3 4 and 7 were less abundant. Outlier cells were present as well. (B) Cells are listed on y-axis and their genomic content is represented by the share of the bar along the x-axis. Some cells showed expression of segments in a more or less even distribution, but most showed irregular expression, dominated by only one or two segments.

Distribution of MSC ranks

Colocalization analysis of A/Pan-M data was performed as described above and is presented in Fig. 30. Monomers constituted the largest fraction, while the size of the other fractions steadily declined with increasing MSC rank. The distribution was roughly similar to exponential decay. Compared to the parental strains A/Panama and A/Mallard, the data resembled those acquired for A/Mallard. However, packaging efficiency seemed to be higher in the case of A/Pan-M as the ratio of completely packaged MSCs of rank 8 to monomeric spots doubled compared to A/Mallard.
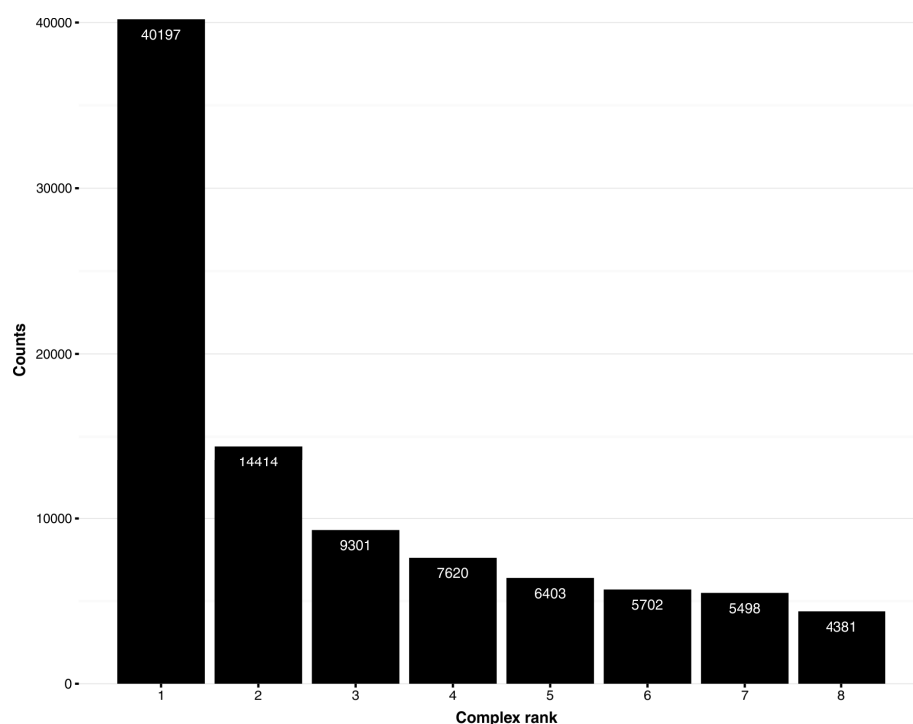
Figure 30 – **Histogram of MSCs detected for A/Pan-M, grouped by their colocalization rank.** Bars represent the number of MSCs detected among all cells with a certain rank. Monomers constitute the highest fraction, and numbers decline towards higher ranks in a fashion resembling exponential decay. Values do not necessarily prove specific packaging, but imply much stronger colocalization than expected from random distribution. In comparison to A/Mallard with a similar distribution, the ratio of completely packaged MSCs of rank 8 to the fraction of monomers is higher.

Network representations of combinatorial data

When plotting A/Pan-M data in a network graph, a comparatively easy layout emerged as depicted in Fig. 31. One major path with limited side paths was visible. In each rank, only a limited number of segment combinations seemed to play a major role, details of which will be presented in the subsequent section. Interestingly, one of the exemplary single cell networks showed a more crowded network which might indicate that this cell did possibly not follow the general packaging scheme for A/Pan-M. Another exemplary cell confirmed that not all cells were dominated by high numbers of segment 1; in the case of this cell segments 1 and 5 were expressed in high numbers.
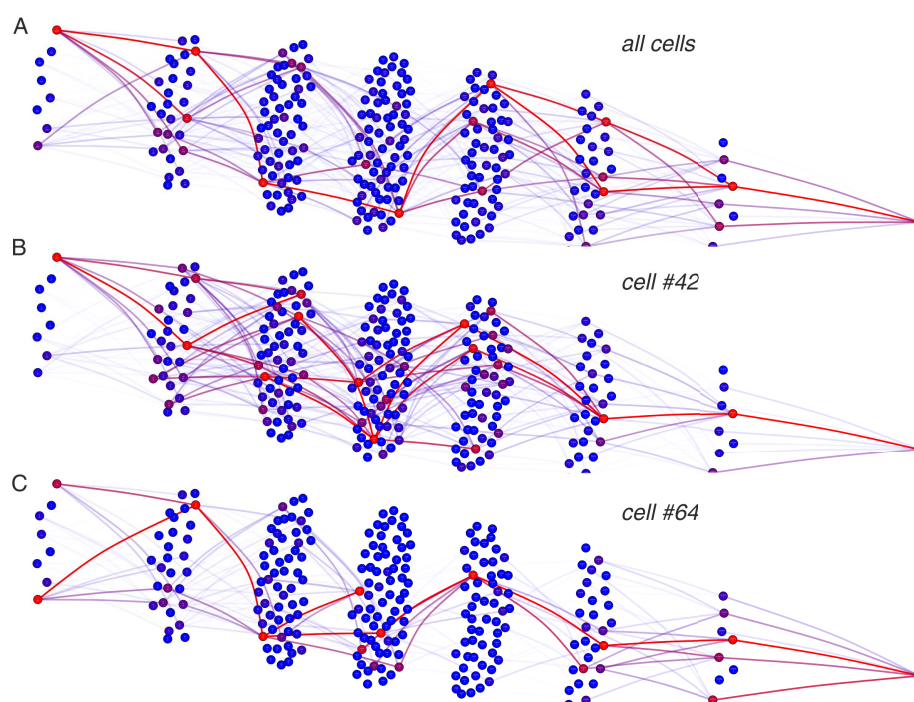
Figure 31 – **The complete packaging network of A/Pan-M.** Abundances of all segment combinations found in the detected MSCs were put into a network representation of IAV genome packaging. The network consists of 255 states, i.e. the possible combinations, which are represented as circles. They are grouped by MSC rank, starting with the 8 monomeric segments on the left and ending with a single circle of the fully packaged MSC of rank 8 on the right. 1016 monomeric addition reactions exist between them and are represented by lines connecting the circles. Both combinations and lines are heat-map color coded so that bright red indicates the combination of highest abundance within a given MSC rank (high weight). Reaction weights were calculated by multiplying the weights of the adjacent combinations. (A) Packaging network for all A/Pan-M data. A comparatively tidy impression is given, and there are only a few bright red connections on each MSC rank. These connections also form a more or less continuous pathway from ranks 1 to 8, with possible redundant side-tracks along the way. (B) and (C) Packaging networks for two individual cells. One cell actually shows a more complicated packaging network. The other looks more tidy, indicating that for a given environment in a single cell, a certain pathway of packaging is highly preferred over alternative routes.

Iterative simulation of a random-packaging hypothesis

Similar findings held true in our iterative simulation approach in Fig. 32. Observed distributions are highly variable and a handful of highly expressed combinations were found among all MSC ranks except rank 7. Segment 1 (PB2) was without exception present in such preferred combinations. Even though the iterative simulation took into account the high availability of segment 1, it was not able to reproduce the observed differences assuming random packaging. This suggested that the importance of this segment could not be explained by high production rates alone. The observed effect was most pronounced in lower and intermediate MSC ranks and was not highly significant among almost complete packed genome bundles of rank 7.
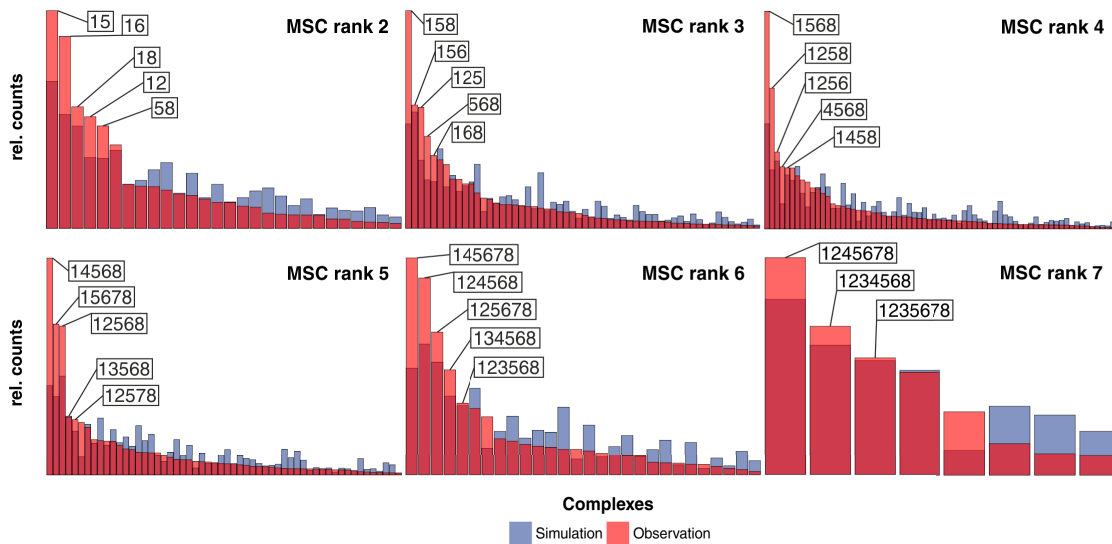
Figure 32 – **Abundances of all observed and simulated segment combinations for A/Pan-M shown by MSC rank.** Each possible segment combination is depicted by overlaid bars (red = observed data, blue = simulated data) and grouped by MSC rank. Combinations are sorted left to right by abundance as seen in observed (microscopy) data. Simulated data stems from the iterative model assuming reactions governed by mass action (i.e. all k-values = 1). Most abundant segment combinations within each MSC rank are named. Observed data was found to be of higher variance compared to the simulated dataset, implicating non-random or specific packaging, applicable at least for MSC ranks 2-6. However, part of the differences might be explained by segment availability. Segments 1 and 8 seemed to play a dominating role as confirmed by their appearance in many upregulated combinations.

Conditional probability modelling

Again, modelling was performed using conditional probabilities in the same way as for the other viral strains. Fig. 33 depicts that k-values were again quite broad, and even more balanced as seen for other strains, indicated by the prevalence of orange among the reactions. Plotting of the 10 preferred pathways through the network yielded several ways through the network with reactions connecting one such "stream" to the other. It was also evident that most segment combinations included in those pathways were highly abundant segment combinations also found in earlier analyses.
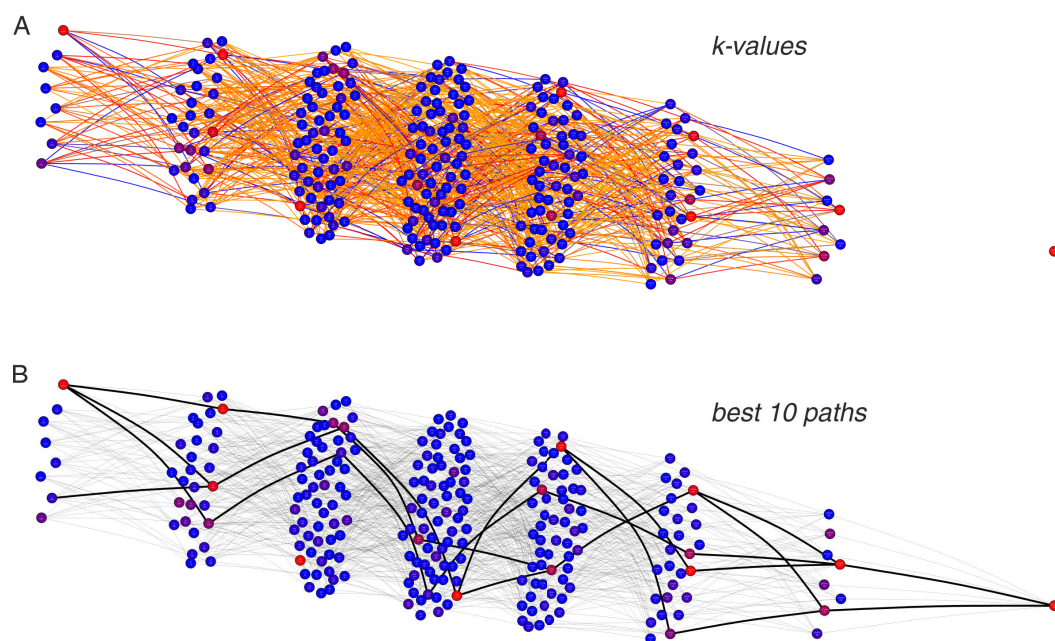
Figure 33 – **Network representation of fitted k-values and best packaging paths for A/Pan-M based on conditional probability modelling.** Network graphs are identical to Fig. 31, however coloring of monomeric addition reactions was based on results from conditional probability modelling here. Basically, modelling removed the bias of segment availability to identify underlying k-values despite the static nature of the data. (A) Reactions are heat-map color coded to represent the relative size of reactions' k-values within one transition from one MSC rank to the next (red is strongest, orange intermediate and blue lowest). A distribution of a large range of k-values over the whole spectrum was observed, indicating that no single set of clearly preferred reaction pathways exists. (B) 10 best paths in terms of highest k-values were plotted in bold black lines in the network. A limited, yet redundant set of possible reactions emerged. These pathways are not necessarily overlapping with highly abundant combinations (bright red circles).

qRT-PCR data

Time course data in bulk for A/Pan-M provided unexpected insights especially into the later stages of infection (see Fig. 34). Among all segments, and both RNA species, a steep drop in RNA levels at 10 h p.i. could be observed, with an increase towards 12 h p.i. visible thereafter. vRNA levels always remained lower than the respective mRNA complement, except for the case of M1 in late infection phases (10-12 h p.i.). Another notable case is the late infection phase of NS1 mRNA, which almost dropped to the vRNA level.
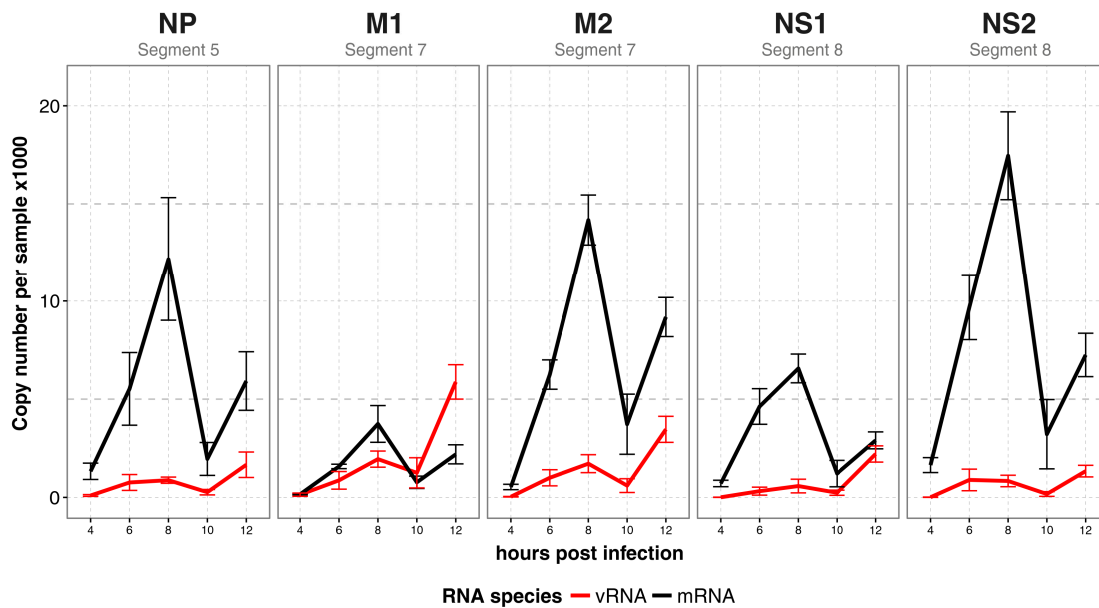
Figure 34 – **qRT-PCR measurements of A/Pan-M vRNA and mRNA expression over the time course of infection.** A549 cells were infected with A/Pan-M at an MOI of 5 for different time spans. Total RNA was extracted, followed by reverse transcription into DNA using either vRNA- or mRNA-specific primers. RNA content was then measured using qRT-PCR, where pHW2000 plasmids of all segments served as standard curves. Availability of such plasmids limited analysis to the segments shown here. Experiments were conducted with three biological and three technical replicates each. Abundances of mRNA (black) and vRNA (red) over a time course of 4 to 12 h p.i. are shown. mRNA expression was higher than vRNA in all cases except for M1. All curves show a strong decline in mRNA levels and – to a lesser degree vRNA levels – from 8 to 10 h p.i. From 10 to 12 h p.i., an increase is visible again. Error bars represent standard error of the mean.

3.6.3 A/Pan-NS

Mapping all viral RNA species in A/Pan-NS

Similar to the procedure for the earlier described viral strains, we conducted MuSeq-FISH on A/Pan-NS infected cells. Here, essentially the A/Panama probe set was used but segment 8 was stained using the A/Mallard probes both for mRNA and vRNA. Much fewer vRNA spots were detected compared to all strains presented above, however, a very high degree of colocalization was still visible from inspection of the overlay image (Fig. 35). On first glance, cell-to-cell variability was not visible in A/Pan-NS based on the homogenous appearance in overlay images of all vRNA channels. mRNA was detected in comparatively low numbers and sometimes appearing in large, bright spots which might correspond to areas of high ribosomal activity. Again, the RdRp coding segments, M1 and NA exhibited a preference for nuclear localization, although we also detected mRNA of these segments in the cytosol. Overall, cell shape of the infected cells was aberrant and often showed a rounded phenotype.
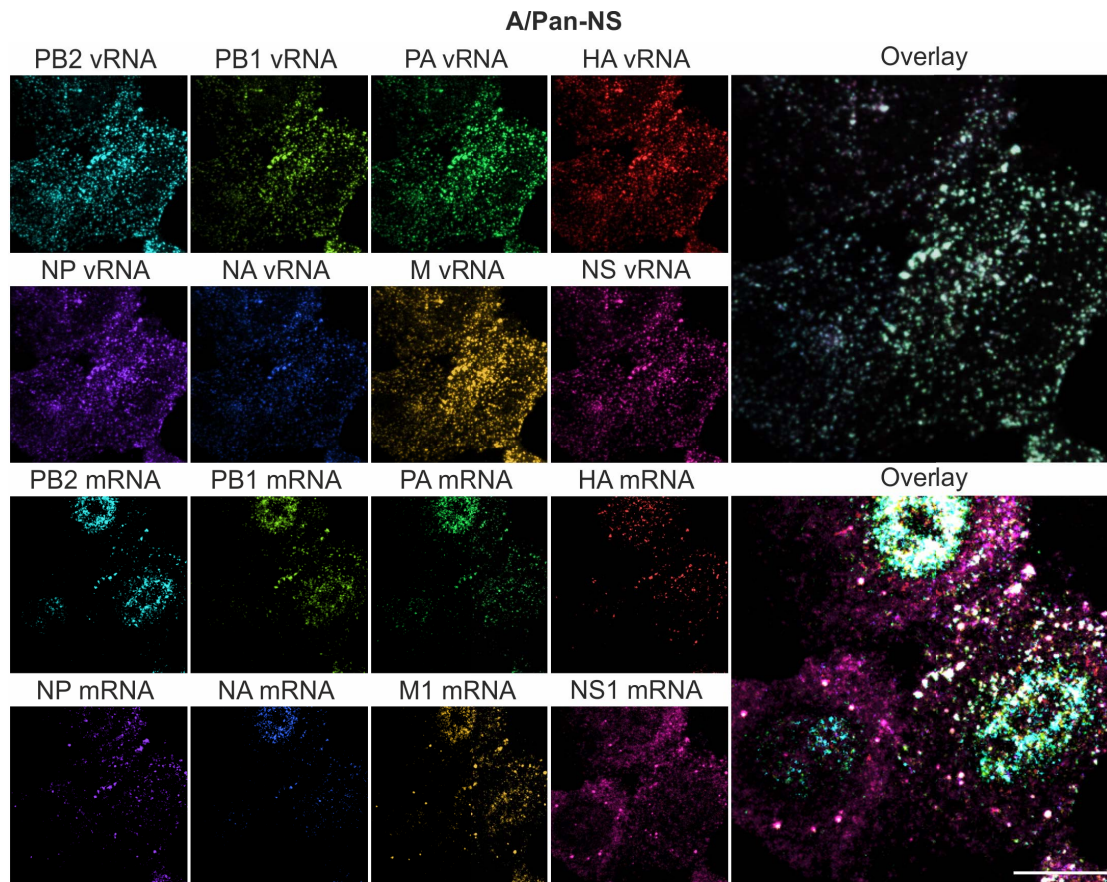
Figure 35 – **Confocal microscopy of A/Pan-NS infected A549 cells 10 h p.i. and stained for viral vRNA and mRNA.** A549 cells were infected with A/Pan-NS at an MOI of 5 and fixated, permeabilized, DAPI-treated and subjected to antibody staining of NP. Subsequently, MuSeq-FISH was conducted, staining all vRNAs and mRNAs. vRNA staining was conducted using ATTO 550 coupled probe sets (upper row). Intermediate numbers of vRNA spots were observed, highly colocalizing among all segments but with a certain cell-to-cell variability visible in the overlay image. Some larger speckles were visible mainly, but not exclusively around the nuclear membrane. mRNA stainings (lower row) were diverse, with PB2, PB1, PA, and to a lesser degree NA showing a predominantly nuclear localization while the other segments were found in both compartments or mainly in the cytosol. mRNA signal levels were overall rather low compared to other viral strains. Scale bar 10 µm.

Cell-to-cell variability

A/Pan-NS showed an unusual distribution of cell-to-cell variabilities compared to the other viral strains (Fig. 36). One the one hand, facet B in the figure illustrates that there was no predominant segment and overall, cells were more or less well-equipped with a number of different vRNA segments. However, facet A emphasizes that indeed systematic differences between expression of different segments were detected, so that segments 1, 4, 6 and 8 were slightly more abundant while segments 2, 3 and 7 were least present and segment 3 was completely absent in several cells. A number of outlier cells showed high contents of either segments 4, 5 and 7.
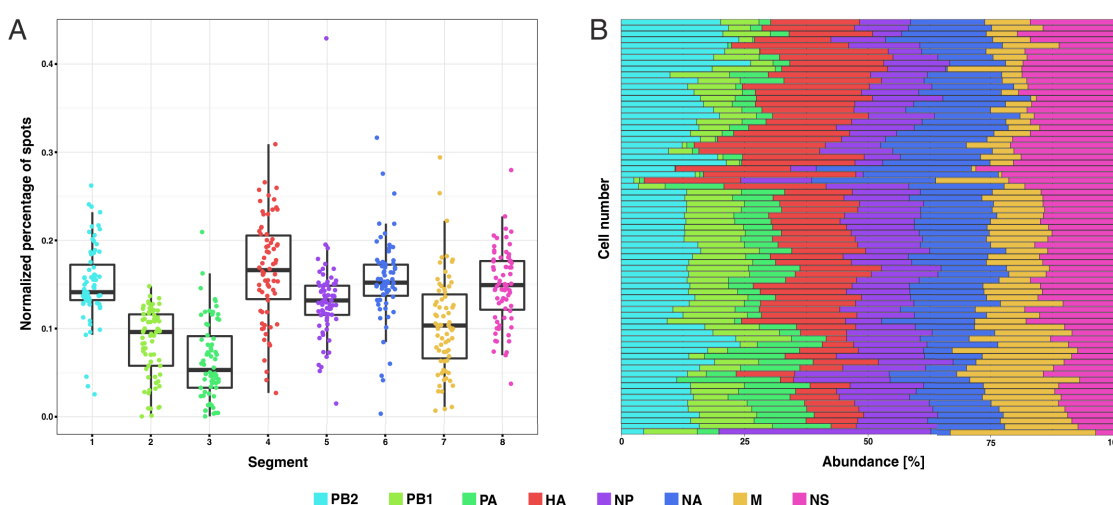
Figure 36 – **Cell-to-cell variability of vRNA abundances in A/PanNS-infected A549 cells 10 h.p.i.** All detected vRNA spots of selected cells were the basis of this representation. (A) Jitter plot showing the variation of genomic content distribution among different cells; each spot corresponds to the share of one segment of the total genome content of the given cell, boxes show the median (thick line) and upper and lower quartile of each data set. The expected average value for equimolar distribution would be 12.5%. Data are highly variable so that outlier cells are found and segments 2, 3 and 7 are less expressed than others. (B) Cells are listed on y-axis and their genomic content is represented by the share of the bar along the x-axis. Some cells showed expression of segments in a more or less even distribution, but others showed irregular abundances, dominated by only one or two segments.

Distribution of MSC ranks

In the case of A/Pan-NS, the histogram of MSC ranks (see Fig. 37) again contained a large fraction of non-colocalizing monomeric spots, followed by a roughly linear

decrease starting from MSCs with rank two. There was even a slight increase from MSC ranks 6 to 7, followed by a comparatively strong drop towards rank 8. Overall, the distribution was comparable to the one observed for A/Mallard.
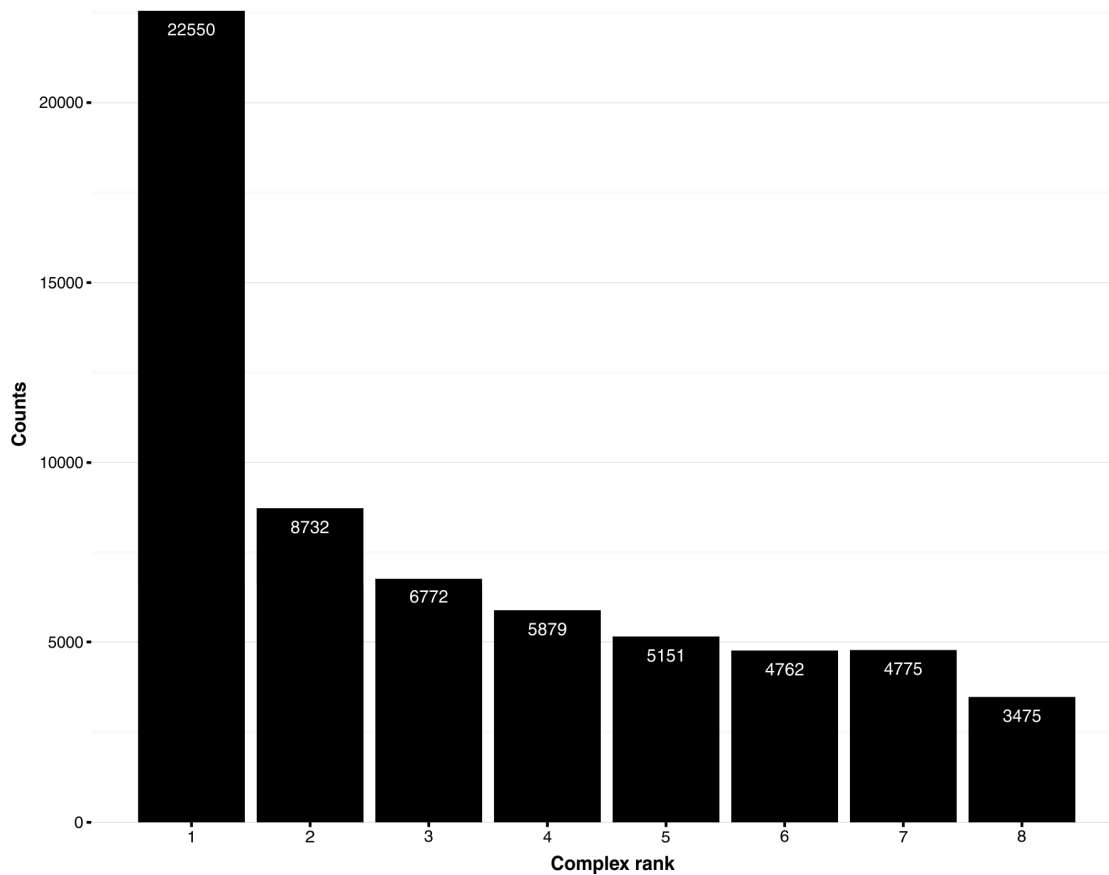
Figure 37 – **Histogram of MSCs detected for A/Pan-NS, grouped by their colocalization rank.** Bars represent the number of MSCs detected among all cells with a certain rank. Monomers constitute the highest fraction, and numbers decline towards higher ranks in a fashion resembling exponential decay. Values do not necessarily prove specific packaging, but imply much stronger colocalization than expected from random distribution. In comparison to A/Mallard with a similar distribution, the ratio of completely packaged MSCs of rank 8 to the fraction of monomers is higher.

Network representations of combinatorial data

Plotting the generated data in network representation (Fig. 38), a few clear paths emerged and other connections were largely attenuated. More often than for the other viral strains, A/Pan-NS exhibited clear "dead ends" or "detours",

where strong connections were observed that did not progress with equally strong connections later in the packaging process and vice-versa. One of the exemplary single cell data networks actually showed a more crowded packaging network while the other basically offers a single highly enriched route with one major side-track at intermediate MSC ranks.
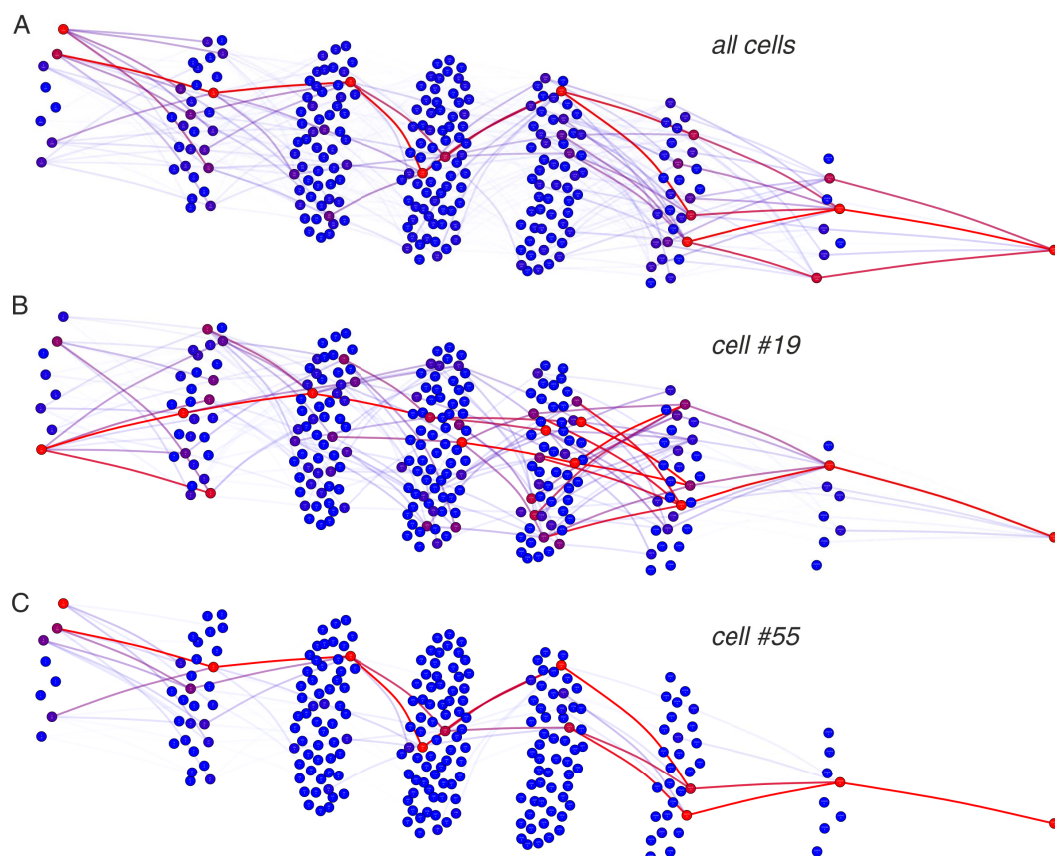
Figure 38 – **Network representations of the complete packaging network of A/Pan-NS.** Abundances of all segment combinations found in the detected MSCs were put into a network representation of IAV genome packaging. The network consists of 255 states, i.e. the possible combinations, which are represented as circles. They are grouped by MSC rank, starting with the 8 monomeric segments on the left and ending with a single circle of the fully packaged MSC of rank 8 on the right. 1016 monomeric addition reactions exist between them and are represented by lines connecting the circles. Both combinations and lines are heat-map color coded so that bright red indicates the combination of highest abundance within a given MSC rank (high weight). Reaction weights were calculated by multiplying the weights of the adjacent combinations. (A) Packaging network for all A/Pan-NS data. A comparatively tidy impression is given, and there are only a few bright red connections on each MSC rank. These connections also form a more or less continuous pathway from ranks 1 to 8, with possible redundant side-tracks along the way. (B) and (C) Packaging networks for two individual cells. One cell shows a different packaging network which is not represented by the summary network. The other follows the summary network and looks more tidy, indicating that for a given environment in a single cell, a certain pathway of packaging is highly preferred over alternative routes.

Iterative simulation of a random-packaging hypothesis

Simulation results regarding distributions of A/Pan-NS data in case of random packaging showed higher variability compared to A/Panama, yet lower than the observed data of A/Pan-NS (Fig. 39). An exception to the rule is posed by MSC rank 7, where the variability of the simulation is even higher, indicating potential balancing mechanisms by restrictions posed by packaging interactions. Overall, especially for MSC ranks 2-5, it is clearly visible that random packaging is not sufficient to explain the observed distributions of possible combinations, thereby refuting purely random packaging for A/Pan-NS. Segment 4 played an important role in almost all combinations of elevated levels, as it was also the most abundant segment for most cells.
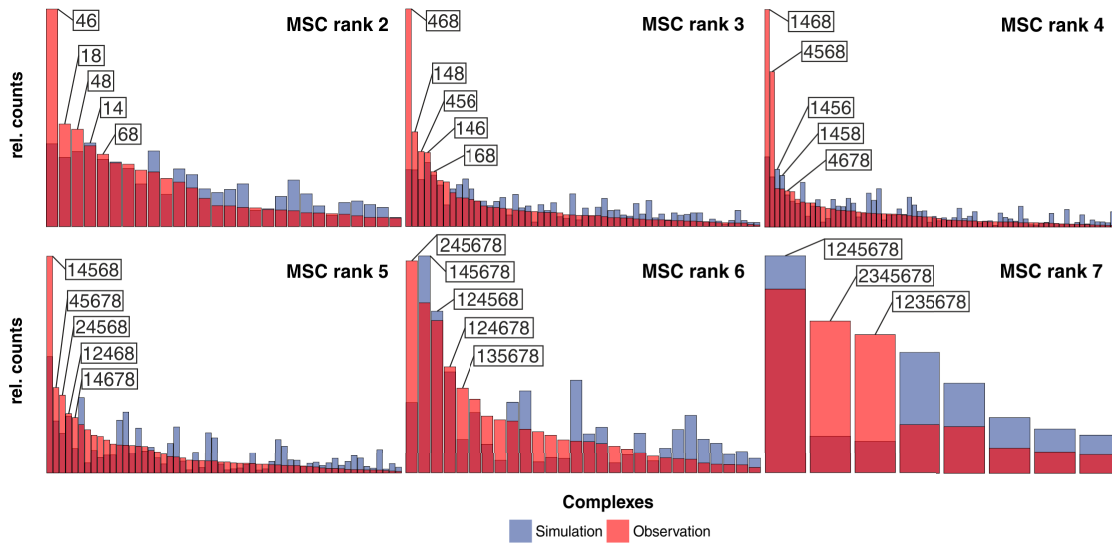
Figure 39 – **Abundances of all observed and simulated segment combinations for A/Pan-NS shown by MSC rank.** Each possible segment combination (ranks 2-7) is depicted by overlaid bars (red = observed data, blue = simulated data) and grouped by MSC rank. Combinations are sorted left to right by abundance as seen in observed (microscopy) data. Simulated data stems from the iterative model assuming reactions governed by mass action (i.e. all k-values = 1). Most abundant segment combinations within each MSC rank are named. Observed data was found to be of higher variance compared to the simulated dataset, implicating non-random or specific packaging, applicable at least for MSC ranks 2-6. For Pan-NS, some segment combinations were extremely preferred, such as for example 46, 468 and 1468. Segments 4, 6 and 8 seemed to play a dominating role as confirmed by their appearance in many upregulated combinations.

Conditional probability modelling

Conditional probability modelling for A/Pan-NS (Fig. 40) showed a broad range of fitted k-values among all MSC ranks, again pointing towards a redundant packaging model in which a number of different pathways towards complete packaging is conceivable. The best pathways showed high conservation on MSC ranks 2, 3 and 5, where a single combination was chosen for all of them. Higher flexibility was observed to exist later in packaging, with MSC ranks 6 and 7 being the ones of highest variability in combinations.
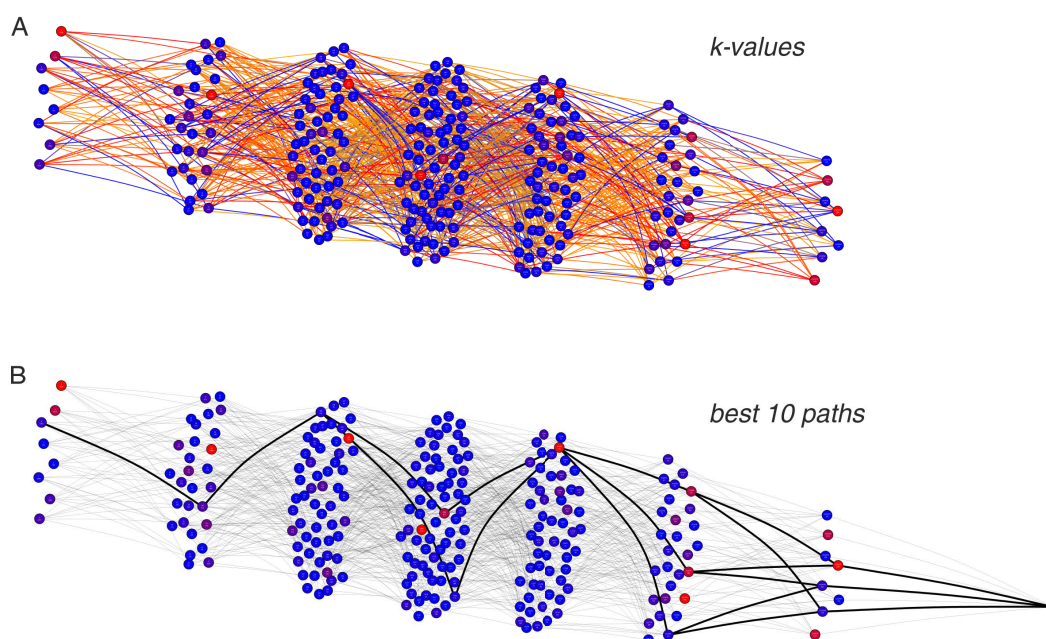
Figure 40 – **Network representation of fitted k-values and best packaging paths for A/Pan-NS based on conditional probability modelling.** Network graphs are identical to Fig. 38, however coloring of monomeric addition reactions was based on results from conditional probability modelling here. Basically, modelling removed the bias of segment availability to identify underlying k-values despite the static nature of the data. (A) Reactions are heat-map color coded to represent the relative size of reactions' k-values within one transition from one MSC rank to the next (red is strongest, orange intermediate and blue lowest). A distribution of a large range of k-values over the whole spectrum was observed, indicating that no single set of clearly preferred reaction pathways exists.(B) 10 best paths in terms of highest k-values were plotted in bold black lines in the network. A limited, yet redundant set of possible reactions emerged. These pathways are not necessarily overlapping with highly abundant combinations (bright red circles).

qRT-PCR data

In qRT-PCR time series experiments on A/Pan-NS infected samples, mRNA levels were consistently higher than vRNA levels for all segment species, except for NS1 which showed nearly identical abundance for the two RNA species (see Fig. 41). M2 was the most highly expressed mRNA, reaching nearly 15 000 copies 12 h.p.i. Notably, vRNA levels of all investigated segments were almost undetectable until 8 h p.i. and rose only moderately afterwards to a maximum value of about 4 000 copies at 12 h p.i.
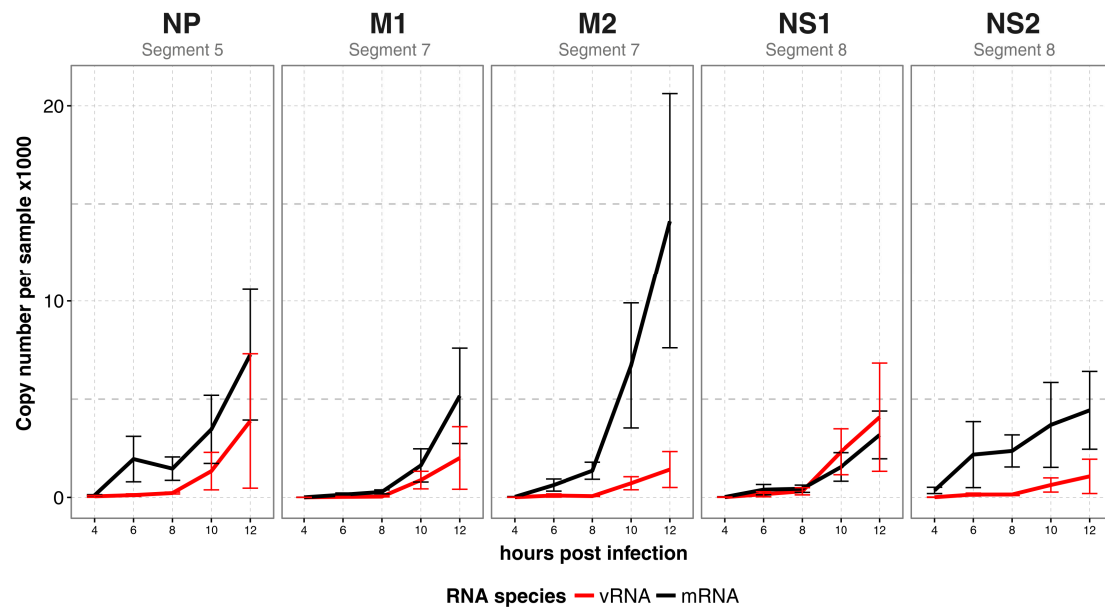
Figure 41 – qRT-PCR measurements of A/Pan-NS vRNA and mRNA expression over the time course of infection. A549 cells were infected with A/Pan-NS at an MOI of 5 for different time spans. Total RNA was extracted, followed by reverse transcription into DNA using either vRNA- or mRNA-specific primers. RNA content was then measured using qRT-PCR, where pHW2000 plasmids of all segments served as standard curves. Availability of such plasmids limited analysis to the segments shown here. Experiments were conducted with three biological and three technical replicates each. Abundances of mRNA (black) and vRNA (red) over a time course of 4 to 12 h p.i. are shown. mRNA expression was higher than vRNA in all cases except for NS1. All curves show constant growth over time, except for a decline in mRNAs of NP and NS1 between 8 and 10 h p.i. vRNA levels of all segments are almost indistinguishable from 0 for the first three time points. Error bars represent standard error of the mean.

3.6.4 A/Pan-NSM

Mapping all viral RNA species in A/Pan-NSM

For A/Pan-NSM, infection and staining was conducted identically to the other viral strains. A bright FISH signal was visible for all vRNA segments (Fig. **??**). Numerous points were detected spreading throughout the cells and A/Pan-NSM showed the strongest phenotype in terms of large, bright vRNA speckles with a higher diameter than usual spots. Such speckles were often observed close to the nucleus, but were also found throughout the cells. All vRNA channels were similar in terms of spot numbers and of comparable brightness, leading to no observable cell-to-cell variability. For mRNA, we also detected high signal intensity and a multitude of spots in all channels. In contrast to all other viral strains, all mRNAs showed a strong tendency towards nuclear localization, where they exhibited very bright, almost evenly colored fluorescence signal. This effect was slightly decreased, but still clearly present for NP and NS. For A/Pan-NSM, no aberrant cell shapes were observed during microscopy.
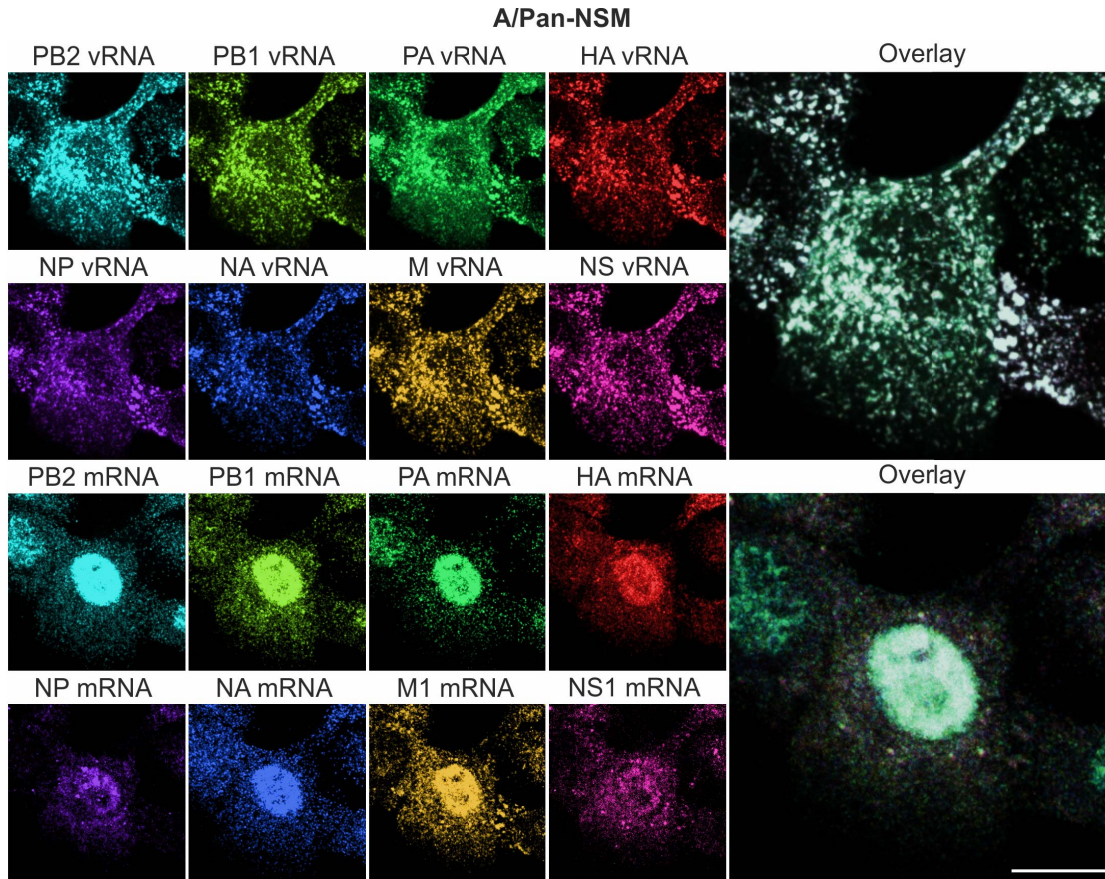
Figure 42 – **Confocal microscopy of A/Pan-NSM infected A549 cells 10 h p.i. and stained for viral vRNA and mRNA.** A549 cells were infected with A/Pan-NSM at an MOI of 5 and fixated, permeabilized, DAPI treated and subjected to antibody staining of NP. Subsequently, MuSeq-FISH was conducted, staining all vRNAs and mRNAs. vRNA staining was conducted using ATTO 550 coupled probe sets (upper row). Very large numbers of vRNA spots were observed, highly colocalizing among all segments but with a certain cell-to-cell variability visible in the overlay image. mRNA stainings (lower row) were diverse, with PB2, PB1, PA, and to a lesser degree NA, HA and M1 showing a predominantly nuclear localization NP and NS were found in both compartments or mainly in the cytosol. Signal intensity of mRNA stainings was higher compared to other viral strains. Scale bar 10 µm.

Cell-to-cell variability

Regarding vRNA segment abundance at single cell level (Fig. 43), we found average values close to an equimolar ratio. Segments 1, 3, 5 and 6 were slightly higher expressed, while segments 2, 4 and 7 were lower expressed. Less evenly distributed cells were still present, especially a group of cells where PB2 was the dominant segment. Segment 7 was the least observed segment, with a few cells showing strongly underrepresented content of it. However, the general picture illustrated that this viral strain had the most consistent and equal distribution of vRNA types. This is also clearly depicted in facet B, where it can be seen that the vast majority of cells could be treated as one single group within normal biological variance.
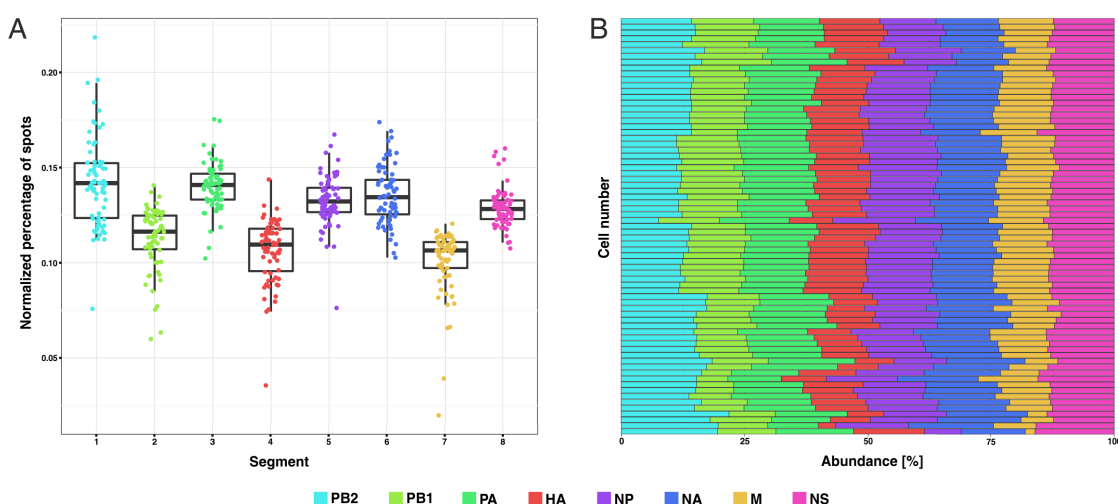
Figure 43 – **Cell-to-cell variability of vRNA abundances in A/Pan-NSM-infected A549 cells 10 h.p.i.** All detected vRNA spots of selected cells were the basis of this representation. (A) Jitter plot showing the variation of genomic content distribution among different cells; each spot corresponds to the share of one segment of the total genome content of the given cell, boxes show the median (thick line) and upper and lower quartile of each data set. The expected average value for equimolar distribution would be 12.5%. Data are less variable than for the other viral strains, with median values more or less representing equal expression in bulk. However, outlier cells are present and in general, segments 1, 3, 5 and 6 are upregulated while segments 2, 4 and 7 are less expressed. (B) Cells are listed on y-axis and their genomic content is represented by the share of the bar along the x-axis. Some cells showed expression of segments in a more or less even distribution, but others showed irregular expression, dominated by only one or two segments.

Distribution of MSC ranks

Results of the colocalization analysis – binning of spots into MSCs – showed a U-shaped distribution similar to A/Panama, which was unique among the reassortant virus strains and A/Mallard (see Fig. 44). Monomers are by far the highest fraction, followed by dimeric MSCs. However, fully packaged octameric MSCs were third, indicative of a functional packaging process. Spots in octameric MSCs constituted roughly a quarter of monomeric MSCs, a ratio which was lower than for A/Panama where it was roughly one half.
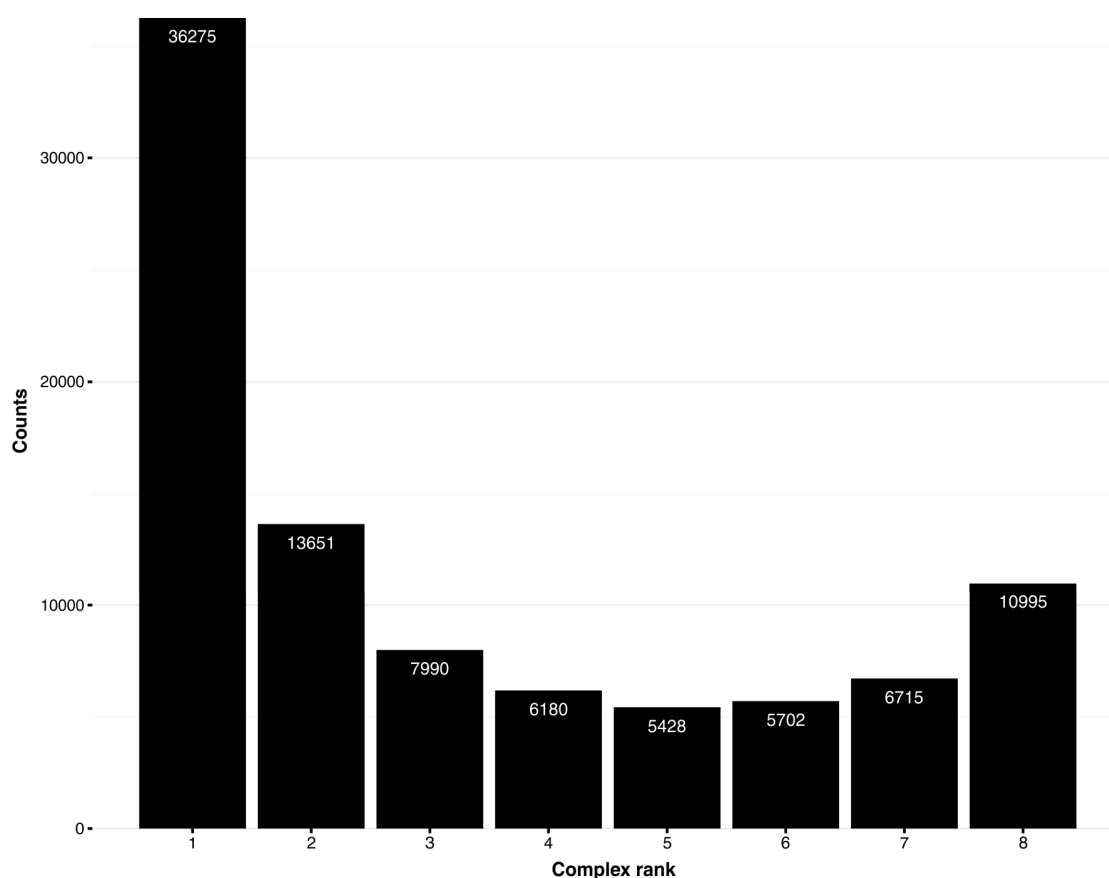
Figure 44 – **Histogram of MSCs detected for A/Pan-NSM, grouped by their colocalization rank.** Bars represent the number of MSCs detected among all cells with a certain rank. Monomers constitute the highest fraction, followed by dimers and then by completely packaged MSCs of rank 8. All in all, a U-shaped distribution emerged which cannot be explained by random colocalization.

Network representations of combinatorial data

In the network representation for all cells (Fig. 45), A/Pan-NSM showed a high variability or redundance of pathways, indicated by many combinations colored in bright red. This held especially true for MSCs of lower ranks towards the left side of the network graph. However, pentameric MSC complexes seemed to be mainly comprised of combination 13568, also described in the section below. In contrast, the network graphs for two exemplary single cells showed very tidy and clear pathways with limited detours from a prominently emerging main path.
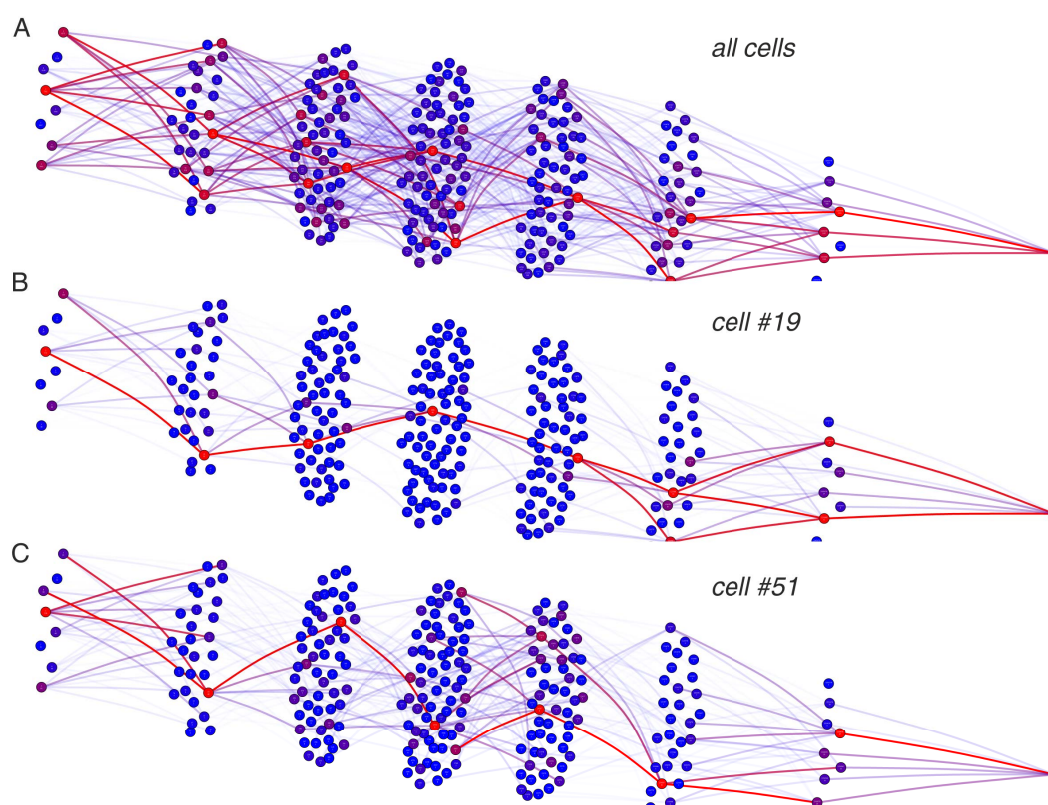
Figure 45 – **Network representations of the complete packaging network of A/Pan-NSM.** Abundances of all combinations found in the detected MSCs were put into a network representation of IAV genome packaging. The network consists of 255 states, i.e. the possible combinations, which are represented as circles. They are grouped by MSC rank, starting with the 8 monomeric segments on the left and ending with a single circle of the fully packaged MSC of rank 8 on the right. 1016 monomeric addition reactions exist between them and are represented by lines connecting the circles. Both combinations and lines are heat-map color coded so that bright red indicates the combination of highest abundance within a given MSC rank (high weight). Reaction weights were calculated by multiplying the weights of the adjacent combinations. (A) Packaging network for all A/Pan-NSM data. A crowded impression is given, with several bright red connections on each MSC rank. These connections mostly form a more or less continuous pathway from ranks 1 to 8, with possible redundant side-tracks along the way. (B) and (C) Packaging networks for two individual cells. Both look much tidier, indicating that for a given environment in a single cell, a certain pathway of packaging is highly preferred over alternative routes.

Iterative simulation of a random-packaging hypothesis

We inspected the detected segment combinations on all MSC ranks more closely and compared them to a distribution computed by iterative simulation based on a random packaging assumption, as we had done before for the other viral strains (Fig. 46). Especially segment 1, and later segment 8, frequently occurred in upregulated combinations along all MSC ranks, indicating a special importance for the packaging process. Comparison with the simulated dataset showed that the hypothesis of random packaging could be refuted for A/Pan-NSM, as variance in abundance in the observed dataset was consistently higher compared to the simulated one. However, the data also showed that a multitude of combinations was upregulated, with the sole exception of MSC size 5 where one combination (13568) clearly stood out.
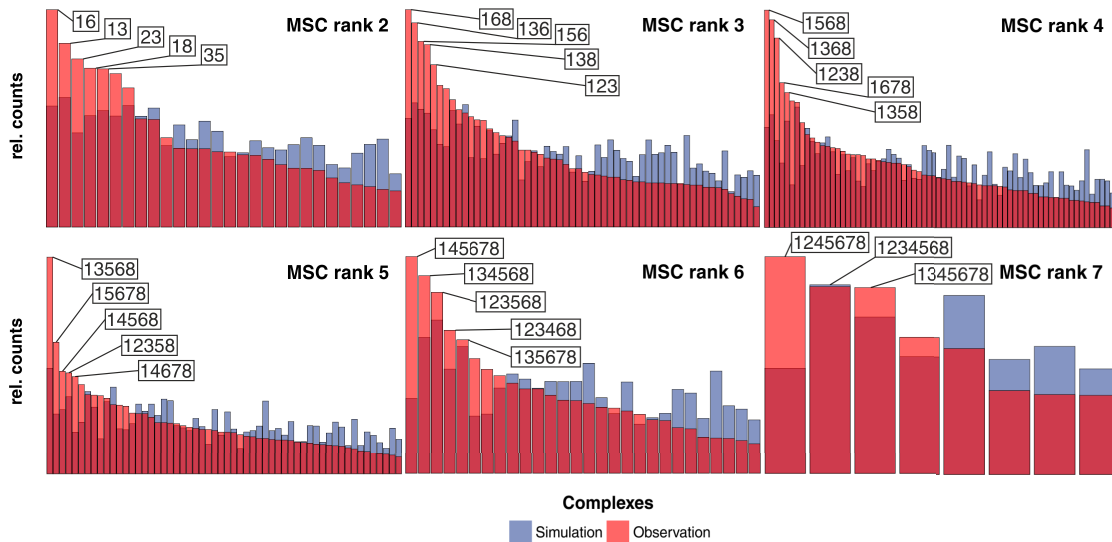
Figure 46 – **Abundances of all observed and simulated segment combinations for A/Pan-NSM shown by MSC rank.** Each possible segment combination is depicted by overlaid bars (red = observed data, blue = simulated data) and grouped by MSC rank. Combinations are sorted left to right by abundance as seen in observed (microscopy) data. Simulated data stems from the iterative model assuming reactions governed by mass action (i.e. all k-values = 1). Most abundant combinations within each MSC rank are named. Observed data was found to be of higher variance compared to the simulated dataset, implicating non-random or specific packaging, applicable at least for MSC ranks 2-6. However, differences are smaller compared to A/Panama. Segments 1 and 6 seemed to play a dominating role as confirmed by their appearance in many upregulated combinations.

Conditional probability modelling

In the same way as for the other viral strains, conditional probability modelling was conducted. The representation of all k-values colored in a heat-map style (Fig. 47) depicted a large number of intermediate k-values, so that no clear separation into preferred and discriminated reactions was seen. This was largely in line with the results for the other viral strains. Looking at the courses of the best pathways, they converged for MSC ranks 2, 3 and 6 but showed redundancy for other parts of the condensation process. Some, but by no means all of the highly expressed combinations from the earlier analyses were recovered in this representation.
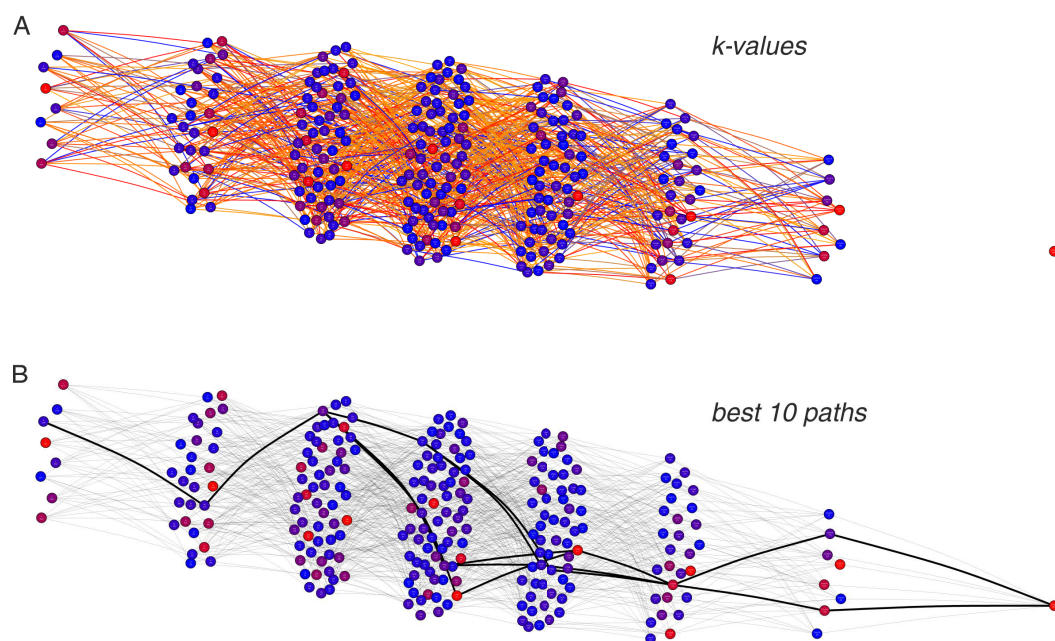
Figure 47 – **Network representation of fitted k-values and best packaging paths for A/Pan-NSM based on conditional probability modelling.** Network graphs are identical to Fig. 45, however coloring of monomeric addition reactions was based on results from conditional probability modelling here. Basically, modelling removed the bias of segment availability to identify underlying k-values despite the static nature of the data. (A) Reactions are heat-map color coded to represent the relative size of reactions' k-values within one transition from one MSC rank to the next (red is strongest, orange intermediate and blue lowest). A distribution of a large range of k-values over the whole spectrum was observed, indicating that no single set of clearly preferred reaction pathways exists. (B) 10 best paths in terms of highest k-values were plotted in bold black lines in the network. A limited, yet redundant set of possible reactions emerged. These pathways are not necessarily overlapping with highly abundant combinations (bright red circles).

qRT-PCR data

All measured abundances for A/Pan-NSM RNA species were considerably higher compared to the other reassortant strains and A/Mallard (Fig. 48, please note different scaling on y-axis). Maximum values of about 120 000 copies were reached for NP mRNA 6 h p.i. Apart from this difference, the time course also showed unexpected behavior, such as oscillations of RNA amounts in NP and M2 or even decreasing RNA levels for NS1 and NS2. For the M segment, levels of mRNA and vRNA were almost identical; the other segments showed a higher share of the respective mRNA as we have seen for most other viral strains.
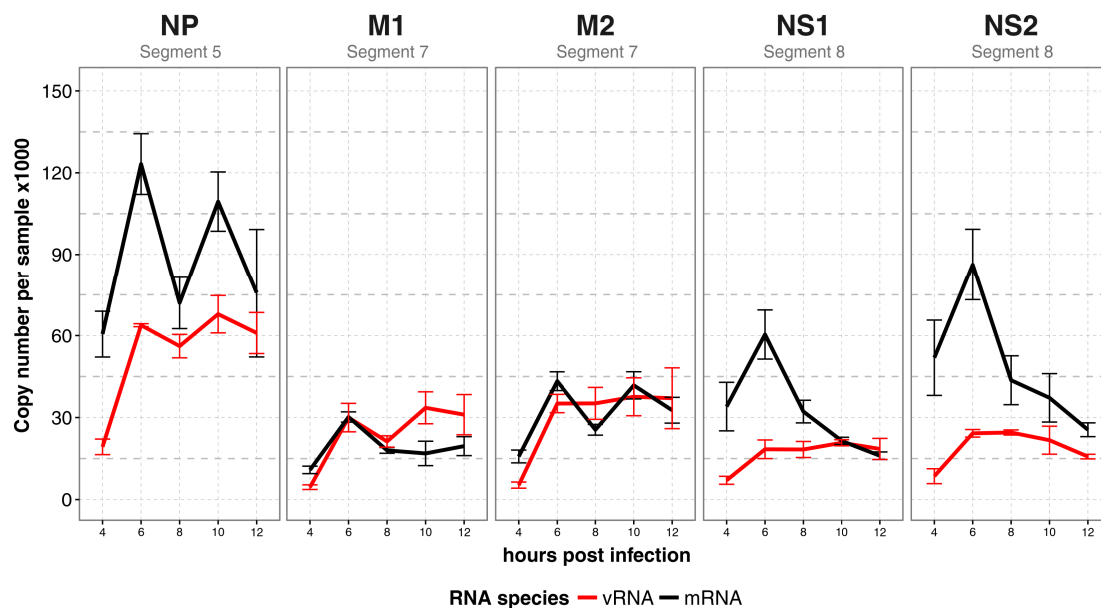
Figure 48 – **qRT-PCR measurements of A/Pan-NSM vRNA and mRNA abundances over time course of infection.** A549 cells were infected with A/Pan-NSM at an MOI of 5 for different time spans. Total RNA was extracted, followed by reverse transcription into DNA using either vRNA- or mRNA-specific primers. RNA content was then measured using qRT-PCR, where pHW2000 plasmids of all segments served as standard curves. Availability of such plasmids limited analysis to the segments shown here. Experiments were conducted with three biological and three technical replicates each. Abundances of mRNA (black) and vRNA (red) over a time course of 4 to 12 h p.i. are shown. mRNA expression was higher than vRNA for NP and NS, while equal levels were observed for the M segment. Most curves show a decline in abundances at 8 h p.i., which is then followed by increase and second decline for NP, stabilization for M and further decline for NS. Error bars represent standard error of the mean.

3.7 Application of MuSeq-FISH to stain single IAV virions

3.7.1 Staining of intact IAV virions in A549 cells

Another application of MuSeq-FISH in the context of IAV was the study of complete virions. To this aim, A549 cells were incubated on ice with A/Panama, A/Pan-M and A/Pan-NS at an MOI of 5. After attachment, the samples were transferred to the incubator to allow entry via endocytosis. Fixation shortly followed (20 min p.i.), then permeabilization and MuSeq-FISH staining were performed as described above. Due to the very short "infection" time, virion particles might have fused with their enclosing endosomes, but vRNP content was expected to still be bundled [27], so that the original genome content of stock virions could be assessed by colocalization of spots representing the different vRNA segments. Obviously, no mRNA staining was conducted in this case as IAV virions do not enclose their mRNA molecules into particles.

Microscopy results are depicted in Fig. 49. Sparse, individual dots were acquired for all imaging channels, but signal intensity was very low compared to the FISH images before. Therefore, especially the signal of M vRNA looked more diffuse than the others and apparent colocalization represented an underestimation of the real situation. Even though a number of approximately five virions per cell was initially estimated (MOI 5), the observed spots per cell turned out to be much higher, in the range of about 40 per cell for A/Panama and even higher in the case of A/Pan-M. However, it has to be kept in mind that the MOI was calculated using titers from plaque tests, where the number of infectious particles, not of virions, is measured.
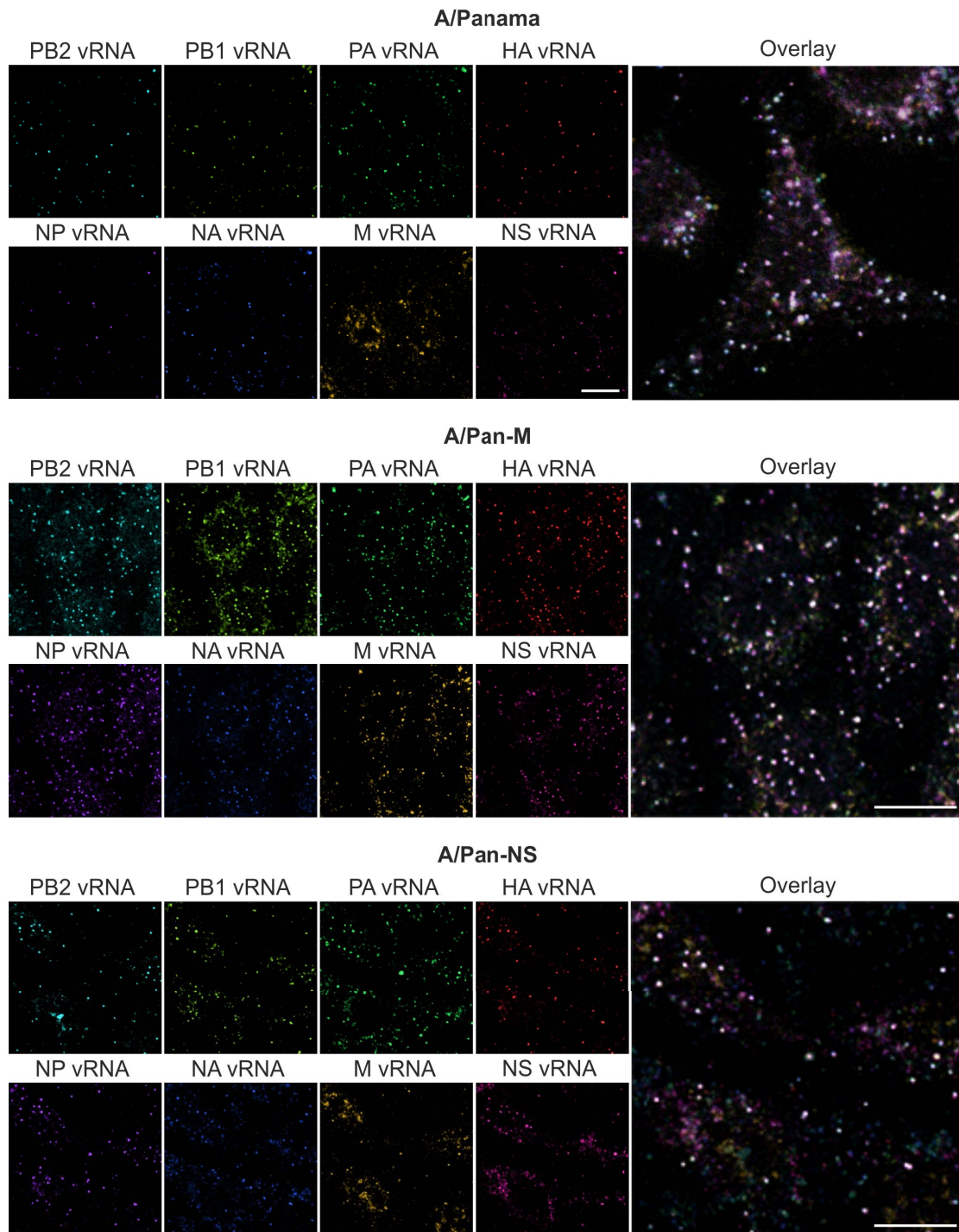
Figure 49 – **Confocal microscopy of A/Panama, A/Pan-M and A/Pan-NS infected A549 cells 20 min p.i. and stained for viral vRNA.** A549 cells were infected with A/Panama, A/Pan-M and A/Pan-NS at an MOI of 5, fixated and permeabilized. Subsequently, MuSeq-FISH was conducted, staining all vRNAs. vRNA FISH was conducted using ATTO 550 coupled probe sets (upper row). Lower numbers of vRNA spots compared to 10 h.p.i. data were observed, highly colocalizing among all segments. Numbers of virion particles seemed to be much higher than indicated by MOI, this effect was most present for A/Pan-M. Scale bars 10 μm.

3.7.2   Distribution of MSC ranks

To assess the genomic content of the virions we had imaged, segment colocalization was performed. Due to low signal intensities, we could not use the automated image analysis workflow and instead reverted to visual inspection, hence this analysis was restricted to lower numbers of spots. vRNA content of virions that had entered three individual cells was analyzed and yielded the distribution depicted in Fig. 50. A strong contrast to the infection of A/Panama 10 h p.i. is visible, with the majority of spots found in MSCs of rank 7 and 8 and the highest fraction contributed by MSC rank 7. We only found very low numbers of MSCs of monomeric, dimeric and trimeric size. These data suggest that genome packaging in A/Panama is not perfect, so that more than half of the virion particles might actually carry an incomplete set of genomic segments possibly caused by imperfect genome packaging.

Data for early infections of A/Pan-M and A/Pan-NS were treated accordingly. Due to low numbers of analyzed spots, this data is only of indicatory nature. However, it looks as if all strains do not perfectly package 8 segments into every single virion. Data of A/Pan-M hinted at best packaging and A/Pan-NS at the worst.
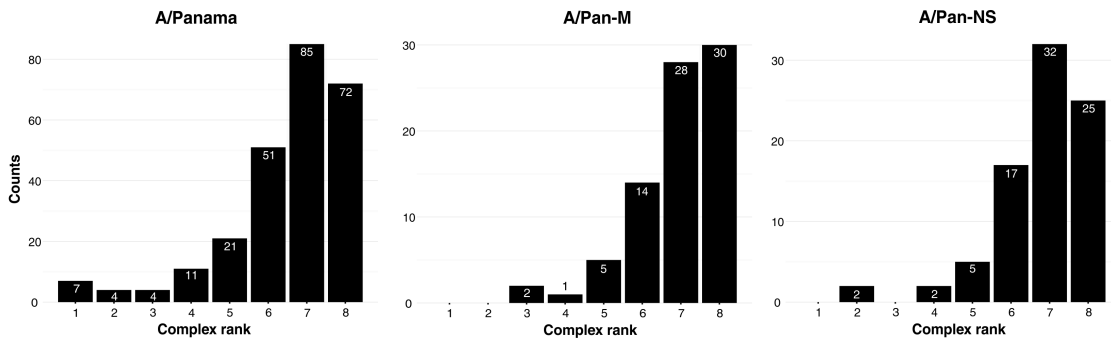
Figure 50 – Histogram of MSCs detected for A/Panama, A/Pan-M and A/Pan-NS, grouped by their colocalization rank. Bars represent the number of MSCs detected with a certain size. MSCs of ranks 7 or 8 constituted the highest fraction, and overall MSCs showed a much higher rank compared to infection a later time points. However, only a part of the detected MSCs corresponded to completely packaged virions.

3.7.3 Observed combinations within MSCs

A similar analysis of the segment combination content was performed as for the infections 10 h p.i. However, iterative simulation as well as network graphs were irrelevant at this early time point of infection, as the underlying data were not intermediates of a dynamically growing genome bundle, but instead static contents of individual virion particles. Due to the small number of analyzed spots, Fig. **??** is restricted to MSC sizes 5-7. Combinations for smaller MSC sizes were only populated by single detections (data not shown). Even for sizes 5-7, it is obvious that the small amounts of data points could only provide indicative data. However, even with such small numbers analyzed, a strong preference for A/Panama and A/Pan-NS MSCs of size 7 to lack segment 8 was observable, followed by the MSC lacking segment 7. For A/Pan-M it was vice-versa, so that segment 7 was lacking most often. This was further supported by the fact that in all strains the single most abundant combination for MSCs of size 6 (i.e. virions lacking two genomic segments) was the one where both these segments, 7 and 8, were missing. Taken together, we could find that virions of all investigated viral strains seem to contain incomplete genomes rather frequently, and that incorporation of especially segments 7 and 8 seems to be less robust than for the other segments.
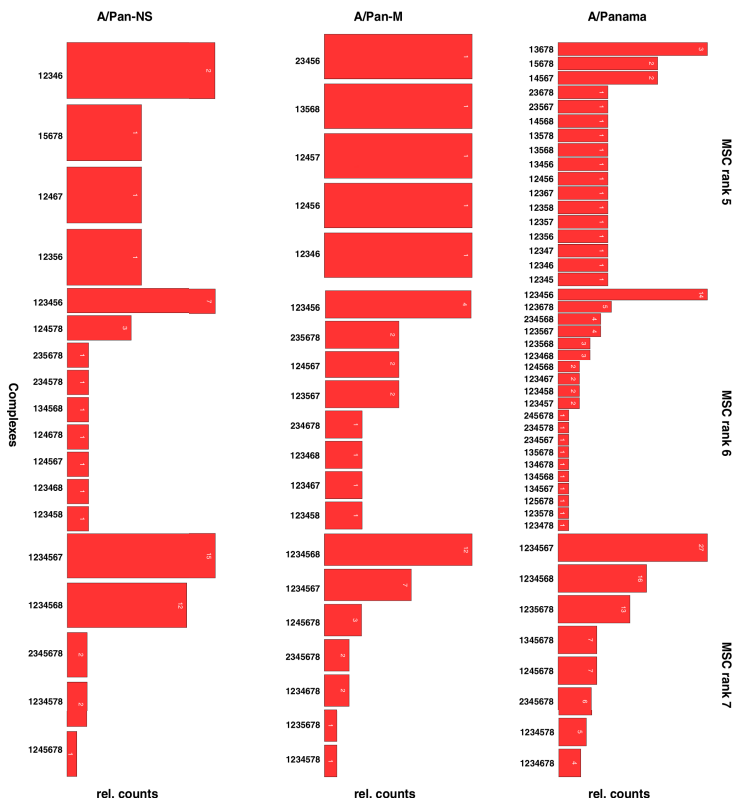
Figure 51 – **Abundances of all observed segment combinations for A/Panama, A/Pan-M and A/Pan-NS shown by MSC rank.** Each detected segment combination (ranks 5-7) is depicted by bars and grouped by MSC rank. Combinations are sorted left to right by abundance. Data is only of indicatory nature due to low numbers of analyzed spots, however a trend emerged suggesting that packaging of segments 7 and 8 was disfavored. In addition, the reassortant strains showed that the respective foreign segment originating from A/Mallard was lacking most in MSCs of rank 7. N = 255 for A/Panama, N = 80 for A/Pan-M, N = 83 for A/Pan-NS.

4 | Discussion

In this work, we have introduced a novel multiplexing protocol for multiple sequential RNA-FISH (MuSeq-FISH), related yet distinct in comparison to alternative techniques that were introduced in recent years [76, 98, 77, 32, 113].

MuSeq-FISH allowed us to visualize and map – to the best of our knowledge – for the first time the complete viral genome of several IAV strains including the seasonal human pathogen A/Panama (H3N2). In contrast to previous studies, we were not limited to a subset of the genomic segments which enabled in-depth analysis of infections in individual cells. We observed a significant cell-to-cell variability in terms of vRNA abundances even among neighboring cells, a dimension that is completely omitted when using bulk methods like RT-PCR.

In addition, we have established an image analysis pipeline for semi-automated image registration, colocalization analysis up to iterative modelling for large amounts of imaging data. All steps in said analysis workflow have been described in this work and can be potentially adapted to other biological questions in which a multitude of – potentially colocalizing – targets play a role. We have used our analysis workflow to investigate genome packaging of various IAV strains and could rule out the hypothesis of random packaging as described by Chou et al. [28]. We also provide a number of potential preferred packaging pathways for the different viral strains which showed different "profiles", or "fingerprints" for each of the measured strains.

We have extended our core analysis to additionally yield proof-of-principle data of STED microscopy of IAV infected cells in collaboration with others (work by Maria Loidolt, MPI Göttingen, Germany), to generate a conditional probability model of IAV genome packaging reaction rates (work by Max Schreiber, University of Leipzig, Germany) and complementary RT-PCR time series data for all viral strains (work by Kalle Jongen and Vanessa Körner, AG Molecular Biophysics, HU Berlin, Germany).

Another aim of this thesis has been the generation of mRNA-only plasmids for A/Panama based on a provided pHW2000-based plasmid system for A/Panama which may serve as a useful tool in further MuSeq-FISH experiments on IAV genome packaging.

First, the mRNA-only plasmids of A/Panama will be discussed, after which the focus of the first parts in this section will be on A/Panama where most extensive data are available. An in-depth comparison of the different investigated viral strains will be the topic of a following part.

4.1 Generation of a set of mRNA-only IAV A/Panama plasmids

In this part of the project, a set of mRNA-only A/Panama plasmids on the pHW2000 vector should be created to serve as a future tool for transfection experiments with MuSeq-FISH staining. Starting from plasmids expressing both vRNA and mRNA of the respective segments, we followed an approach by Stech et al. [159] as described above and were successful in re-cloning all plasmids except for NA. Plasmid sequences were confirmed by sequencing.

Such a set of mRNA-only expressing plasmids, in effect an mRNA-only plasmid system, can be used in combination with the standard system consisting of 8 plasmids coding for both mRNA and vRNA. It is always possible to only transfect parts of the system, corresponding to a limited set of proteins available in the cell, which is used in many routine applications in virology. However, in the context of genome packaging, it is important to yield functional and interacting vRNPs with a defined set of segments. Creation and export of vRNPs can only be achieved if at least PB2, PB1, PA and NP (required for vRNP assembly) together with M and NS (required for export) are transfected, which limits the number of segments to be omitted from a transfection to 2 (HA and NA) and the number of potential vRNA combinations to 4 (all vRNAs, all vRNAs except HA, all vRNAs except NA, all except HA+NA). By the use of a combination of mRNA-only plasmids with standard pHW2000 plasmids, it can be ensured that all needed proteins are available intracellularly while the vRNA content is restricted to any number of standard segments transfected. Therefore, the number of possible experiments is potentially increased to 255, which is exactly the number of segment combinations observed in our colocalization analysis.

In a proof-of-concept experiment, we transfected the NP mRNA-only plas-

mid, while all other segments originated from the A/Panama standard plasmid system. NP expression was indeed confirmed via antibody staining, indirectly proving that the removal of the PolI promoter did not negatively influence the plasmid's ability to express mRNA. However, NP location was limited to the nucleus, indicating that vRNP assembly and/or export were perturbed because exported vRNPs would be detected as NP-positive spots in the cytosol. The likely cause for this is incomplete transfection, so that one or more plasmids coding for proteins needed for assembly and export were lacking from the imaged cells. This is also supported by the fact that transfection efficiency for the plasmids that were directly imaged, i.e. mRNA-only NP and Rab11-RFP was far below 100%. As five other plasmids need to be transfected for successful export of vRNPs, chances are rather low to observe a full transfection.

Therefore, the results have shown that tools are at hand to investigate specific parts of the A/Panama genome packaging network using MuSeq-FISH, but transfection needs to be optimized and extensively validated first. These will be valuable next steps for the future.

4.2   Mapping the viral genome of several IAV strains

Using our MuSeq-FISH approach, we were able to obtain highly multiplexed fluorescence images of IAV infected cells. This gave us access to a holistic view on viral RNA household in the model system of infection in the human lung cell line A549. It also allowed us to study the interplay of vRNA and mRNA of the same segments as well as vRNAs and mRNAs among different segments.

4.2.1   Viral mRNA

For mRNA of A/Panama, we found differences in abundance as well as localization using MuSeq-FISH. The mRNA copies of the RdRp subunits were almost exclusively detected within the nuclei of cells, which held true to a lesser degree for HA and NA mRNA. In contrast, NP and M1 mRNAs were evenly distributed throughout nuclei and cytosol, while NS1 mRNA showed a preference for cytosolic localization. Unfortunately, it was not possible to stain M2 and NS2 mRNAs due to the short length of their sequences, inadequate for design of a sufficient number of FISH probes for detection. The biological implications of these localization patterns remain unclear, however it might be speculated that it is no coincidence to find three mRNA species – PB2, PB1 and PA – closely related in function all

enriched in the nuclei.

As the mRNA sequence is essentially identical to the sequence of complementary RNA (cRNA) transiently produced during replication of vRNA, our probe sets inevitably stained both. However, cRNA has never been reported to leave the nucleus [83], so that any cytosolic signal is expected to stem from mRNA with very high confidence. Concerning the nucleus, one can argue that mRNA levels should be highly enriched in comparison to cRNA according to literature, so that at least a vast majority of the signal would truly be originating from mRNA [83].

Using the resolution of confocal microscopy, we were able to resolve single spots of mRNA, supposedly representing single RNA molecules. However, the density of viral mRNA turned out to be extremely high, so that the otherwise spot-like FISH signal sometimes partially resembled an illuminated area more than a typical FISH staining of cellular mRNAs, in which only a dozens or hundreds of mRNA molecules are found per cell [139].

4.2.2   Staining of single IAV virions

In other experiments, stainings of the vRNA content of recently endocytosed (20 min p.i.) IAV virions of the strains A/Panama, A/Pan-M and A/Pan-NS have been performed and imaged. Detection of all viral genomic segments using MuSeq-FISH was successful, however fluorescence signal brightness was reduced compared to experiments discussed above. Spot numbers differed between strains even though the MOI was kept constant, leading to the assumption that the ratio of infectious particles capable of reproduction might be lower in the reassortant strains compared to A/Panama. Automated analysis using our image analysis pipeline was not possible in this case due to low signal to noise ratio, so that manual inspection of comparatively low numbers of spots (80-250) was performed using FIJI, which will be discussed below.

4.2.3   Reliability of FISH

Fluorescence *in situ* hybridization is an intrinsically error-prone method as has been mentioned in many studies before [113, 32]. Such errors are both types, false-positives and false-negatives, however it is arguably more likely that hybridization – and therefore detection – fails in comparison to a situation in which several fluorophores aggregate without target. Cellular mRNA is covered with interacting proteins and might also be affiliated with one or more ribosomes during translation when stained – which are all factors of steric hindrance which might impair correct hybridization of FISH probes. In the case of influenza, and especially its vRNA,

the situation becomes more complex as the RNA forms strong secondary structures (the panhandle structure) within vRNPs. We were interested in genome packaging of IAV, which is assumed to take place by means of direct RNP-RNP interactions. So it is easy to envision that parts of the RNA sequence are masked from FISH probes once such interactions are formed. In effect, we cannot pinpoint the error rate present in our experiments, but it is evident that false negatives are an issue in such assays. The problem becomes even more prominent as we were interested in colocalization of high orders, so that the chance to detect a completely packaged MSC of rank 8 is in fact the detection rate for a single species to the power of 8. It is therefore of utmost importance to keep in mind that all data presented here are most likely an underestimation of actual condensation states. Nevertheless, we believe that highly multiplexed FISH as implemented by MuSeq-FISH still is the best approach currently available to identify and colocalize high numbers of different RNA targets. In addition, there is no indication in the literature to assume that certain segments should suffer from a higher error rate in FISH experiments than others. If true, homogenous error rates would mean that no bias is introduced if spot misdetections of all segment types happen with the same probability. Therefore, all data and analyses on combination compositions of MSCs should reflect the actual workings of IAV packaging, even though absolute numbers might not.

4.2.4   Cell-to-cell variation

For the first time, we have detected and quantified cell-to-cell variations in IAV infection in terms of segment availability. MuSeq-FISH, in this context, allowed to unambiguously extract proportions of the individual segments to the total viral RNA content in given cells. Among cells showing the expected behavior of almost equimolar distributions of vRNA segments, we also found cells that showed strong overexpression of one or more segments with respect to the others. Some segments were more likely to show this overexpression, but this was a trend, not a rule. In the case of A/Panama, average values of segment expression were almost equal, which is consistent with the known picture from any bulk analysis [134]. So we speculate that this is probably not a systematic preference for some segments, but rather a stochastic process that is aided by the small numbers involved in infection. Even at MOI = 5, a given cell is likely to be infected by only a handful of virion particles. If a certain segment happens to be lacking in one or more of these, or fails to enter the nucleus due to degradation, this might give a head start for other segments' replication. Ultimately, a distorted composition of newly synthesized segments will be harmful for the virus as cellular resources are "wasted" for the production of superfluous vRNPs. However, it seems as if the system tolerates at least a certain degree of tilt in vRNA content.

4.3 Implications of the data concerning the genome packaging

The real gem within the extensive FISH data lay in its possibility to investigate the long-standing question of IAV genome packaging without compromises. As described above (see introduction), extensive data has been acquired in recent years in favor of regulated packaging, at least to a certain degree [97, 96, 51, 58]. However, the techniques used operated either at bulk level (like co-reassortment analyses) and/or *in vitro* (like bandshift assays) [51]. It is quite likely that a complex tertiary and quaternary structure like the IAV vRNP contributes to interactions just as the primary sequence does. Hence, it was of foremost importance to extend existing assays to actual vRNP-vRNP interactions. Electron microscopy and tomography have provided that, yet without conclusive identification of segments' identity [130, 161]. FISH has been used for this purpose before, but was restricted to a maximum number of four segments concurrently imaged [28, 90]. Hence, only a subset of segments could be analyzed at the same time, with the additional burden that segment combinations could not be reliably detected – any apparent combination could well contain additional segments which were not stained at the same time. Using our multiplexed approach, we could overcome this barrier and even conduct stainings for each segment in two different colors in the case of A/Panama. We then only used spots occurring in both stainings for further analysis, thereby increasing robustness of our approach.

However, RNA-FISH is an inevitably error-prone method, especially when it comes to false negatives, i.e. RNA molecules not sufficiently stained even though they are present in the sample. The rate of false negatives has been estimated to be around 20% [162], which is a ballpark number high enough to significantly affect results. This problem is somewhat put into perspective when we assume similar rates of false positives of all segments, so that no systematic bias would be introduced into our analysis. What remains is a systematic underestimation of actual MSC ranks and packaging efficiency.

4.3.1 Distribution of MSC ranks

Plotting all detected MSCs into a histogram revealed that within A549 cells 10 h p.i. with A/Panama at an MOI of 5, all kinds of intermediate, not yet fully bundled genome complexes were found. Therefore, we were able to assume a steady-state for genome bundling, allowing us to investigate an essentially dynamic process by means of microscopy on fixated cells at a single time point. In fact, the temporal dimension which was lacking in the experiments was replaced by the

spatial dimensions, where we found MSCs in all intermediate steps on their way to growing bundles.

We consistently found a high number of monomeric spots, where we were not able to find any colocalizing segments. This number is highly dependent on thresholding levels, i.e. stringency of acceptance criteria for a spot during spot detection. It could easily be reduced by increasing the threshold value (data not shown), however we intentionally chose a relatively low threshold in order to keep as many data points as possible early in analysis. Further down the analysis pipeline, data integrity and validity were ensured by both a second independent staining for each segment and by the rejection of MSCs containing double copies of any segment.

The characteristic U-shaped distribution of the population of MSC ranks (compare Fig. 14) strikingly confirmed the existence of a packaging mechanism. Occurrence of completely packaged octameric MSCs was preferred over MSCs of intermediate rank, which cannot be explained by purely random colocalization. The relatively low abundance of MSCs of intermediate ranks, especially ranks 4-6, might also indicate that bundling is a relatively fast process in the lifetime of a vRNP, in a way that vRNPs bundle quite rapidly once they have reached the cytosol and are ready for interaction with other vRNPs.

Next, we set out to learn more about the contribution of individual segments during the packaging process. Numerous studies have reported that certain segments are more important than others, possibly serving as mediators between adjacent segments within the interaction network. It might also be speculated that only a subset of segments can serve as condensation cores of bundling. We calculated the probability of each segment to be found within an MSC of a given rank to identify such preferred segments. If specific condensation cores existed, it would be expected that MSCs of small rank, only starting to bundle, would preferentially contain such segments. However, we found that for all segments (aside from a slight preference of segment 2 and a negative preference for segment 8 for small MSCs) probabilities grow essentially identically and according to the expected value for random incorporation. This is a clear indication that IAV does not seem to rely on specific condensation cores during the bundling process, or in other words, that segments do not need to be "primed" by a previous interaction to allow subsequent interactions. Hierarchical packaging is not disfavored by this finding – we merely learn that sub-sections of the packaging network might be able to assemble on their own.

4.3.2 Network analysis

The analysis discussed above was segment-specific, but not yet combination-specific. Therefore, the next step to a more granular analysis was to look at the distributions of specific combinations in the packaging network. We therefore generated a network consisting of 255 states (all potential segment combinations) and about 1000 - 4000 edges connecting them (assembly reactions). The states were then populated with the data extracted from our FISH experiments, so that the weight of each state in the network was set by the observed abundance of this specific combination.

Here, we assumed that combinations were permutation-insensitive, so that a colocalization of e.g. segments 1, 2 and 3 was identical to 2, 1 and 3. In other words, combinations were completely defined by their genomic content and invariant to any spatial organization. This assumption was necessary as the resolution of confocal microscopy, and in fact any kind of optical microscopy, is not sufficient to detect spatial organization. However, efficient packaging can hardly be envisioned when incompatible arrangements of vRNPs play a role, so that we are confident that this assumption is biologically sensible.

For the number of possible reactions, we simplified the network assuming that reactions are irreversible and – for some analyses – that only monomeric additions are possible. These assumptions could not be backed because there is no literature data available on the kinetics of vRNP assembly. Nevertheless, we believe that data analysis was not impaired by the simplifications, while on the other hand they allowed for modelling attempts that would have been impossible without restrictions due to the enormous complexity of the network and thus, computational costs. In fact, reversible reactions would only play a role if an MSC would react differently after dissociation of one segment, which would imply some kind of "memory" within a genome bundle. This is rather unlikely, given that vRNP-vRNP interactions are probably mediated by means of RNA-RNA base pair interactions and no covalent binding. Similar reasoning holds true for only allowing monomeric additions: they only restrict modelling success if certain condensation steps *require* the addition of already pre-bundled sub-complexes. For example, if the reaction (12) + (34) → (1234) is possible, but also reaction paths like (12) + (3) → (123) and (123) + (4) → (1234), then the dimeric addition is completely equivalent to two monomeric additions and can be treated as a combination of such *basic* reactions.

Those network graphs were created both for the complete dataset and additionally for data of each cell, thereby exploiting the single cell level of the data. States/combinations within the same MSC rank were grouped together

and heat-map coded to visualize the states with the highest relative abundance within this MSC rank. Connecting edges were colored similarly, using a weight indicator determined by the weights of both endpoints for the edge. This approach yielded graphs that allowed for easy visual inspection of the highly complex dataset.

In the case of A/Panama, we observed a number of highly abundant combinations within each MSC rank among a high number of other combinations with moderate or very low expression levels. This means that combinations of segments are not randomly dispersed but rather confined to a subset of the complete packaging network, which is again a strong indication for non-random packaging following specific rules. While classical binary interactions (which are the result of e.g. bandshift experiments) are reflected in the group of MSC rank 2, our data expanded interaction information to all subsequent condensation steps as well.

When interpreting such network data, it has to be kept in mind that these are actually kinetics inferred from a static time point. Therefore, it cannot be ruled out that highly expressed states might in fact not be intermediates of productive genome bundling, but instead dead end combinations which accumulate because they fail to produce virions and leave the cell. We rejected this hypothesis based on the fact that highly expressed combinations of one MSC rank usually had counterparts in higher MSC ranks containing the same segments, which is a strong indication that ongoing reactions are indeed possible.

In many cases, highly expressed combinations of neighboring ranks were actually composed of the same segments, with an additional segment in the MSC of higher rank. In fact, many complete or almost complete "packaging paths" could be followed by eyesight when tracking bright red states in the network graphs. This also supports our initial assumption to restrict visualization to monomeric additions, which seem to be sufficient to at least explain a majority of the packaging observed. If condensation reactions of higher order would be required at any step, this would be reflected in network graphs by gaps in packaging paths. We only rarely observed bright red states with no successor in combinations of the next rank, which indicates that dead ends in packaging seem to occur only to a minor degree or not at all.

Interestingly, network graphs of single cells regularly showed less crowded networks in comparison to bulk data, meaning that only one or a few combinations were clearly upregulated at each MSC rank. While the network graphs emerging from bulk data indicate that *several* – parallel but heavily interconnected – pathways to full packaging exist, many single cell network graphs indicated the existence of a strong *single* packaging pathway. Why should different cells follow

different paths for genome condensation? One potential answer lies in the observed cell-to-cell variability discussed above. We have learned that infections proceed differently even in neighboring cells, with some segments being much higher expressed than others. From here, we can speculate that a number of potential packaging paths exist, while segment availability governs which of those is actually taken in every cell. In fact, we could show that highly expressed combinations in lower MSC ranks tend to contain exactly those segments which are highly abundant, i.e. packaging tends to start with segments of high availability. In sum, we believe that packaging happens in a confined solution space and should actually be predictable in cells with a given segment composition, but that the actual packaging network for any given viral strain is much wider and offers high degrees of redundancy.

4.3.3   Rejecting random packaging, iterative modelling

Network representations of the combinatorial data already indicated that packaging should be non-random. However, it could be speculated that such a pattern might emerge just due to the specific starting conditions of a given cell. In such a case, genome bundling would follow basic kinetics of mass action, while the actual interaction preferences (or k-values) of condensation reactions would be all equal (i.e. 1). To rule out this possibility, an iterative modelling approach was implemented, trying to populate the packaging network starting from monomers in the observed concentrations for each cell. All reactions were governed solely by mass action kinetics and the simulation was allowed to proceed until a maximum value for cross-correlation to the observed distribution from microscopy data was reached. As shown in Fig. 17, both distributions differ widely: while the iterative modelling approach showed a preference for combinations with highly available segments, those were only upregulated by a factor of about 2, while we observed differences by a factor of 10 or more in the experimental data. Therefore, we conclude that those differences cannot be explained by segment availability due to differential replication. This refutes simple mass-action kinetics and in turn means that different k-values for reaction paths seem to exist. Given the manifold different paths of which highly elevated levels of intermediates were found, we believe that a packaging model of multiple, redundant and interconnected pathways towards complete genome bundling explains the data best.

4.3.4   Conditional probability modelling

We speculated that some information in the network might be hidden by uneven abundances of vRNPs. Cells dominated by one or more segments will most likely also include them early in the packaging process due to mass action kinetics.

Hence, we sought a method to eliminate the contribution of segment availability and resorted on conditional probability modelling, which was implemented by Max Schreiber (University of Leipzig, Germany) in Matlab. Several assumptions, similar to those described above, had to be made in order to reduce complexity of the model. For example, only monomeric additions were included and each transition step from one MSC rank to the next was treated independently. After fitting, a complete set of k-values for each reaction was obtained, together with "best paths" resulting from reactions with the highest k-values on each MSC rank (see Fig. 18). The distribution of k-values was rather broad and approximately Gaussian shaped, indicating that no clear preference for certain reactions based on RNP-RNP interactions was detected. Rather, in agreement with our previous results, a multitude of reactions seemed to be possible in theory, of which certain pathways are chosen based on segment availability in a precise infection environment. The best pathways of A/Panama only overlapped to a small fraction with data based on abundances of combinations, however agreement was much better for the other viral strains investigated (see below).

All in all, the chosen modelling approach introduced a novel tool into genome packaging research and therefore stands in parallel to existing and validated methods. It should be treated as additional and indicative evidence, but with caution. Due to the step-wise fitting from low to high MSC ranks, we expect the data of higher ranks to be less precise. There is generally a need for validation of such approaches on systems where k-values are accessible, as for example metabolic networks. Indeed, metabolic network modelling is an evolved field with many software packages available, such as CellNetAnalyzer [160] or Data2Dynamics [140]. However, classical metabolic modelling assumes states to be of very temporary nature, which is a good assumption for metabolites, but possibly not for vRNPs. Therefore, abundances of metabolites are usually not part of the model and are replaced by endpoint data, such as monomers and complete octameric complexes in the case of IAV packaging. These data were not sufficient to model the system using classical software tools available.

4.3.5   Creation of putative, preferred pathways

Putative packaging pathways for A/Panama were compiled, even though the two sources of data (microscopy data vs. conditional probability modelling) did not show large agreement with each other and were therefore treated separately (Fig. 52). These data do not necessarily include all highest k-values/abundances, but only cover those which lead to a consistent pathway through the complete model, from monomers to fully packaged genome bundles. Modelling suggested that three packaging routes are favored, independent to a certain degree but with

connecting reactions possible. Therefore, it can be speculated that such pathways do not represent completely different modes of packaging, but rather describe essentially the same process with potentially different sub-networks as starting points. Amounts of segment combinations however suggested a more concise pathway, starting with segment combination 27 which then splitted into three strongly intertwined paths.
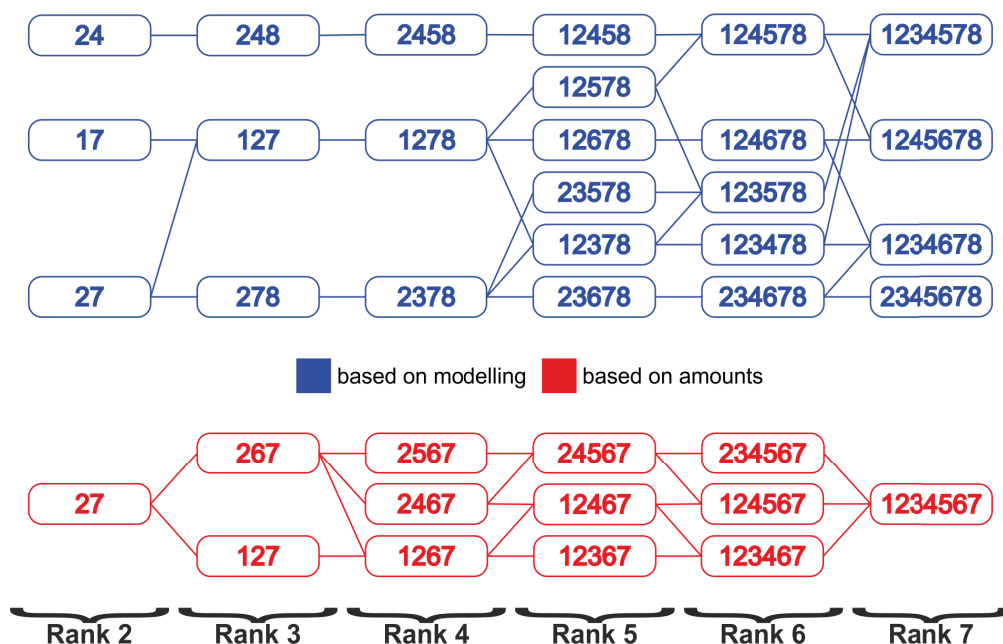
Figure 52 – **Proposed pathway diagram for preferred pathways of A/Panama genome packaging.** Pathway diagram shows preferred pathways based on two data sources, (1) abundances of combinations in microscopy data, shown in red and (2) "fastest" paths as determined by highest k-values from conditional probability modelling, shown in blue. In both cases, only MSCs/reactions were included that were part of the best 10 end-to-end pathways leading from monomers to fully packaged octameric complexes. Ranks 1 and 8 were omitted for improved readability. For A/Panama, a complex network emerged from modelling data which might reflect uncertainty in the model. Both approaches show only limited overlap.

4.3.6   Comparison with the literature

Even though the study of IAV packaging mechanisms has seen a lot of interest in the last years and decades (see above), it is challenging to put the data on A/Panama packaging into perspective. RNA-FISH has only been used with

a limited set of probes, thereby obscuring some interactions [28, 90]. Other approaches, like extensive mapping of binary interactions for the strains A/Finch and A/Moscow, have been limited to *in vitro* interactions of naked RNA and should therefore be treated with caution. However, due to the extensive data sets provided for these two strains, a comparison has been drawn in Fig. 53. Here, interaction strengths in our data have been inferred from their fitted k-values and ordered by strength. As depicted, the strong interaction of segments 2 and 7 was not observed in either A/Finch [58] or A/Moscow [50], however many other interactions in the upper third were actually shared with A/Finch. A/Moscow showed less agreement, which is surprising given that it is a strain of the H3N2 subtype like A/Panama.
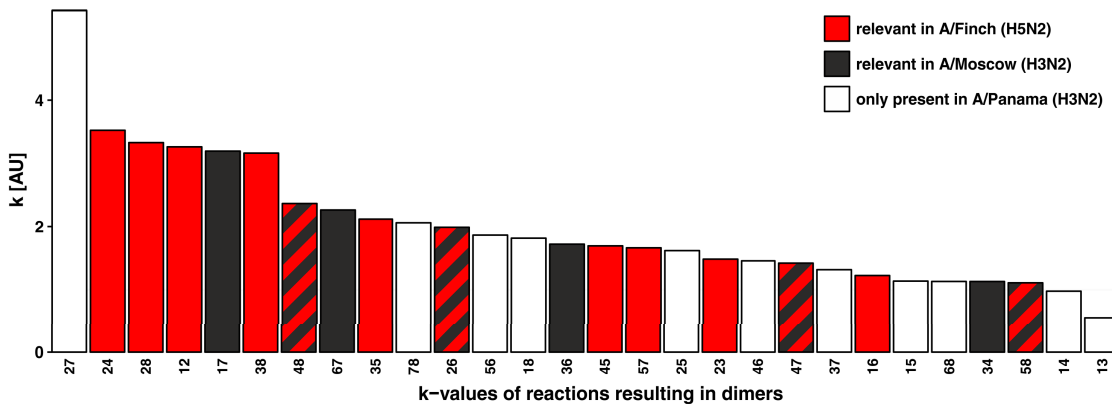
Figure 53 – **Comparison of observed dimeric interactions in A/Panama with those observed in A/Finch and A/Moscow.** Bar plot shows the relative k-values for the creation of each possible dimeric MSC as fitted by conditional probability modelling for A/Panama. Dimeric interactions for A/Finch and A/Moscow have been obtained by Fournier et al. and Gavazzi et al., respectively. Surprisingly, A/Panama interactions show more overlap with those reported for A/Finch, even though both A/Panama and A/Moscow are H3N2 viruses and therefore seem to be more closely related.

We can only speculate about possible reasons, one of them being that packaging interactions might have evolved differently from the rest of virus genetics. In fact, the interaction network is first of all an "internal" feature of the virus and not necessarily related to its interactions with the host cell, as the hemagglutinin and neuraminidase alleles are. Furthermore, bundling signals use parts of the viral RNA that partially also code for mRNA and later for proteins. So there is clearly another factor driving the evolution of motifs on bundling signals, as first of all integrity of RNA and proteins need to be ensured for infectious viruses. In general,

it can be said that proper understanding of genome packaging requires interaction measures of higher orders than dimers, which our data has now entered. Future studies will be able to compare interaction strengths on this level, which might open the way for more insights.

4.3.7 How redundant is the system, evolutionary advantages of redundancy

Summing up results from our various analyses, we can conclude that IAV genome packaging – at least for the strains that were investigated – is neither random nor completely determined. The validity of such results has been discussed from a technical perspective already, but can also be seen from the angle of evolutionary advantage of a viral strain over time. Having a segmented genome comes with the additional burden to ensure packaging of as many infectious virion particles as possible, but on the other hand allows faster evolution through the reassortment process. Many viruses, not only among *Orthomyxoviridae*, leverage segmentation of their genome, as described in the introduction of this work. The question of how exactly genome packaging is achieved is tightly intertwined with the one of segmentation. Basically, an evolutionary successful virus balances high replication rates, genomic stability against deleterious mutations and a high degree of adaptability towards new environmental challenges. One can assume that implementation of a segmented genome and a packaging mechanism consume some of the exploited resources of the host cell and therefore attenuate the viral replication potential, while at the same time they allow for fast adaptation. Likewise, a strongly restricted packaging mechanism that confines packaging to a single pathway might increase stability and reduce the amount of incorrectly packaged genome bundles. However, a flexible pathway might be more resilient towards odd segment availability within single infected cells and, even more importantly, facilitate reassortment with other viral strains. As shown in many studies [97, 96, 50, 58], successful reassortment requires compatible packaging signals, so that we can assume that redundancy in the packaging process makes it easier to reach such compatibility by fewer required mutations. If the environment of the virus changes, e.g. the population of the host organism declines, reassortment can lead to survival of the virus as it allows a switch of host organisms.

4.4   Differences between different reassortant strains of Panama IAV

4.4.1   Expectations

In addition to the measurements for the A/Panama strain, we have also conducted experiments on the avian H3N2 strain A/Mallard and three reassortant viruses containing a mixed genome set originating from A/Panama and A/Mallard (see Fig. 20). Measurements on different viral strains provided insights on how strongly packaging is conserved among related strains and how reassortment is able to re-wire such a system. To keep other variables constant, all experiments were performed in human A549 lung cells, even though they do not represent an optimal system for the avian strain A/Mallard. It is known that A549 cells are susceptible, yet not permissive for the A/Mallard virus, so that infection occurs but the virus fails to replicate. This led to the expectation that A/Mallard and its related reassortant strains might perform weaker infections than A/Panama.

4.4.2   Differences in localization and abundance of vRNA/mRNA

For all other viral strains, we were able to successfully repeat our MuSeq-FISH staining protocol, resulting in complete mapping of vRNA and mRNA abundances and localization pattern at 10 h p.i. in A549 cells.

While we indeed found A/Mallard infections to yield fewer vRNA spots, this was not the case for the reassortant viruses. A/Pan-M and especially A/Pan-NSM showed very high densities of vRNA spots all over infected cells, sometimes even higher than for A/Panama. In terms of mRNA, abundance was highest in A/Pan-M and very low in A/Pan-NS. In A/Pan-NS, a few very bright spots were detected in mRNA measurements, which might correspond to aggregates. However, this finding was not further investigated. Preferential RdRp (PB2, PB1, PA) mRNA localization in nuclei, as found in A/Panama, was confirmed by the experiments with A/Pan-M and A/Mallard. For A/Pan-NSM, HA and M1 mRNAs were also mostly located in nuclei in addition to the aforementioned RdRp mRNAs. For A/Pan-NS, signal levels were lower compared to the other strains impeding gain of conclusive information about localization, however a trend towards nuclear localization was also seen here.

These data showed that for all investigated virus strains, infection of A549 cells succeeded, even though permissiveness in A549 cells was negative at least for

A/Mallard (Thorsten Wolff, RKI Berlin, personal communication). It might be speculated that high signal levels and spot counts, especially in A/Pan-M and A/Pan-NSM viruses, might stem from more virion particles per cell in our experiments compared to A/Panama. All experiments were conducted with an MOI of 5, using viral titers that were measured in plaque assays on MDCK cells. Possibly, the number of plaques yielded for the reassortant viruses might not reflect the actual number of viral particles since only infectious and thus, virions comprising a full genome, induce plaque induce plaque formation. This would result in highly elevated virion numbers for the affected strains, which is backed by our observation of very high spot density in single virus experiments with A/Pan-NSM (data not shown). It remains an open question to the field whether one should try to balance plaque numbers and virion particles observed in FISH experiments of different viral strains.

4.4.3   Differences in cell-to-cell variation

All other viral strains, except for A/Pan-NSM, showed higher variability in segment abundances compared to A/Panama. This indicates that the replication process might have been disturbed by the introduction of foreign genomic material in the case of the reassortants. We found that the broad pictures of segment abundances in A/Mallard and A/Pan-NS looked very similar and shared high variance in segment 4 abundance between different cells and comparatively weak abundances of segments 2 and 7. Yet the opposite is the case for A/Panama where these two segments are of special importance. A/Pan-M was dominated by very high levels of segment 1, followed by segments 5, 6 and 8 which were also preferred. The bar plot in this case revealed that a distinct group of cells is visible to the bottom of the distribution where segment 1 contributed up to half of total genome content (Fig. 29). For A/Pan-NSM, variation among different cells was almost non-existent, while variation among segment expression levels was clearly visible, although within a small range (median values per segment between 11-14%). Expression of segments 1, 3, 5, 6 and 8 was preferred.

Heterogeneity in IAV infections to the degree we have reported here will have an impact on the interpretation of many experiments in the field. It implies that measurements of only a small number of cells might not be valid, as any effects detected in such a small population could be restricted to this exact population. The question remains whether such a degree of variation is a by-product of weak regulation in IAV replication patterns or might actually fulfill a role in virus fitness by probing different reaction environments. It may be caused by variable starting numbers of vRNPs of different segments, easily envisioned when a small number of virions infect a cell and stochastically, less copies of some segments

reach the nucleus. However, too little is known about the actual situation of *in vivo* infections of adjacent cells, where very high MOI values might be present during most infection phases which would even out cell-to-cell variability due to the law of high numbers. In this context, the extremely low cell-to-cell variation in A/Pan-NSM may be easily explained by the fact that virion numbers might have been elevated up to 10-fold compared to A/Panama.

4.4.4   Differences in colocalization

When looking at colocalization strengths as indicated by the histograms of MSC ranks, A/Mallard, A/Pan-M and A/Pan-NS all showed weaker bundling capability in comparison to A/Panama. For the reassortants, this might have been caused by their artificial genomic composition and sub-optimal interplay of A/Panama- and A/Mallard-derived segments in bundling. However, this is contradicted by the results of A/Pan-NSM, where a strong preference for complete bundling was observed, similar to A/Panama. Taken together, histograms for all viral strains showed distributions which cannot be explained by pure stochastic colocalization, as confirmed by colocalization analysis of non-bundling viral mRNA (data not shown).

4.4.5   Differences in network graphs and modelling

While the network graphs of A/Mallard and A/Pan-NSM were very crowded, showing a multitude of preferred reactions, graphs of A/Pan-M and A/Pan-NS were very tidy, even on bulk level using data of all cells. The reason for that remains unknown; however a plausible explanation might lie in their reassorted genome. If one assumes that segment 7 (in A/Pan-M) and segment 8 (in A/Pan-NS) derived from A/Mallard do not perfectly fit into the genomic environment of A/Panama, this might reduce the solution space for packaging considerably. In effect, it might resemble the situation in single cells of e.g. A/Panama where similarly tidy diagrams were obtained – where the solution space would be reduced not by means of genetic incompatibilities but instead by means of segment availability. The situation in A/Pan-NSM, exhibiting a wide solution space again, might be reasoned by the fact that both segments 7 and 8 are derived from A/Mallard here, so that these might form a functional sub-unit of bundling with less friction towards the A/Panama genomic part.

Moreover, we consistently observed that highly abundant combinations tended to form continuous pathways through the networks, indicating that most observed combinations are likely to be intermediates on their way to higher orders of bundling instead of dead ends.

After simulation of random packaging for all viral strains, the available data strongly favored specific packaging for most MSC ranks. Exceptions to this rule were seen mostly for A/Mallard and for MSC rank 7 of most virus strains. Sterical reasons might play a role in the very late stages of genome bundling: in a situation where almost all segments are already present in an MSC, the order of segments might already be well-established, so that mass action kinetics play an increasing role for the last few segment additions. Interestingly, the viral strains differ widely in their most important dimeric combinations, which then evens out towards MSC ranks 4 and 5, where a lot of conservation in the process is visible. This is discussed in more detail in the following section.

4.4.6 Putative, preferred pathways

As for A/Panama, we combined insights from conditional probability modelling and amounts of segment combinations, restricted them to reactions constituting full pathways and compiled putative preferred pathways for each viral strain (see Fig. **??**). The resulting pathways are comparatively concise in contrast to A/Panama. In addition, agreement between both data sources was significantly higher in these cases. For A/Mallard, packaging seems to start with segments 4, 5 and 6 or 8 depending on data source. Agreement is reached on MSC rank 5. Segments 1, 5 and 6 or 8 (depending on data source) seem to play a pivotal role for A/Pan-M, and partial agreement between model and observation can be seen throughout the network. In the case of A/Pan-NS, segment 6 and 8 are important, while observations suggest adding segment 4. Agreement is visible from MSC rank 4 onwards. Finally, for A/Pan-NSM, segments 6 and 1 or 8 (depending on data source) are found early in packaging; partial agreement of data sources is seen from MSC rank 4.

Even though pathways are different for each viral strain and moreover, in contrast to A/Panama packaging, some common features are present. For example, the exact combination 14568 is used by A/Mallard, A/Pan-M and A/Pan-NS, while A/Pan-NSM uses the related combination 13568. It is also evident that even though parallel pathways might exist, as seen in A/Pan-M or A/Pan-NSM, these are still strongly connected to each other, so that many different routes of packaging will likely be taken, determined by the next available segment for addition.
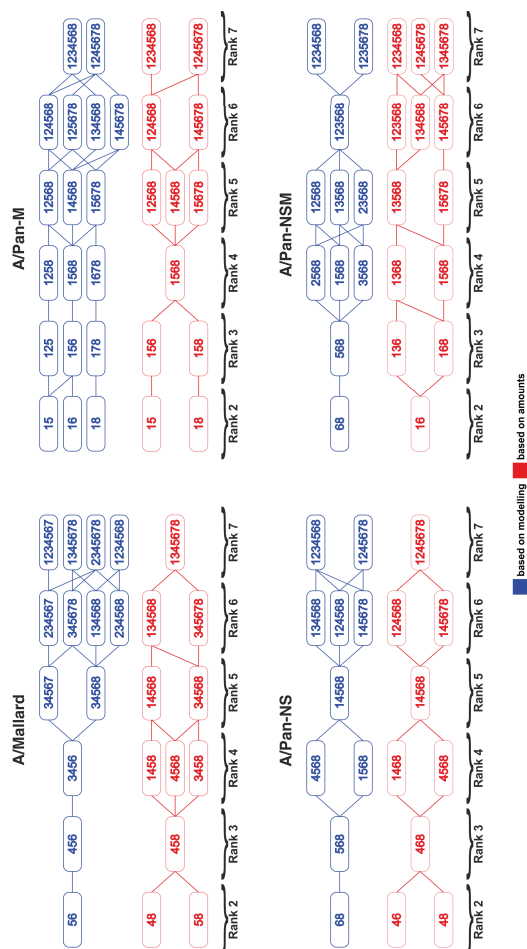
Figure 54 – **A/Mallard, A/Pan-M, A/Pan-NS and A/Pan-NSM genome packaging based either on MSC abundances or conditional probability modelling.** Pathway diagram shows preferred pathways based on two data sources, (1) abundances of combinations in microscopy data, shown in red and (2) "fastest" paths as determined by highest k-values from conditional probability modelling, shown in blue. In both cases, only MSCs/reactions were included that were part of the best 10 end-to-end pathways leading from monomers to fully packaged octameric complexes. Ranks 1 and 8 were omitted for improved readability. For these viral strains, networks from both data sources were tidier compared to A/Panama and also showed good agreement between both data sources.

4.4.7 Summary

All in all, results from the different viral strains remained highly diverse. FISH stainings and data analysis were successful and were in good agreement with A/Panama results on a visual level. Genome packaging of all other viral strains except for A/Pan-NSM seemed to be impaired as indicated by histograms of MSC sizes. For A/Pan-NSM, highly elevated virion numbers in infection might explain its better packaging performance in comparison to A/Mallard with a wild-type genome set. Another reason might be that reassortment of both segments 7 and 8 yielded an improved interplay of the two and helped the whole genome packaging process. Network analysis as well as conditional probability modelling revealed that significant differences between all viral strains exist, but on the other hand some packaging constraints, like highly enriched combinations, seemed to be conserved among the reassortants. Taking in the results of all available data, one can conclude that (1) A/Mallard performs weaker in genome packaging than A/Panama, (2) A/Pan-M and A/Pan-NS also show impaired genome packaging similar to A/Mallard, (3) A/Pan-NSM seems to exhibit functional genome packaging similar to A/Panama, (4) A/Pan-M and A/Pan-NS resemble A/Mallard in terms of e.g. appearance of stainings and distribution of highly elevated combinations.

5 | Conclusion and Outlook

5.1   Conclusion

Taken together, the results presented in this thesis have shown that (1) MuSeq-FISH is a versatile and comparatively easy method for implementation of multiplexing capability to a standard FISH assay, (2) MuSeq-FISH is a promising method to study genome packaging in influenza viruses and (3) IAV genome packaging, at least for the five viral strains covered here, is neither random nor completely determined.

MuSeq-FISH has been extensively used in this project, both on vRNA and mRNA targets of IAV. In addition, other systems, such as hantaviruses and yeast cells, have been tested in proof-of-concept experiments with promising results (data not shown). Together with the image analysis pipeline especially tailored for processing highly multiplexed data and freely available upon request, we believe that MuSeq-FISH will become a beneficial tool in the field.

The question of the nature and workings of IAV genome packaging was not satisfactorily answered here, in terms of an exact and specific pathway as a result. We have learned that such an answer most likely does not represent the actual workings – actually the solution seems to lie in a highly flexible, yet specific and definitely not random packaging network. As we have seen, this network is also dependent on the viral strain and even on the infection environment of a given cell. For the situations investigated here, we could rule out random packaging and contour the solution space. Future work in this field will most likely have to give up the simple view of "reaction pathways" and evolve towards a holistic and integral view of packaging space.

5.2   Outlook

5.2.1   Super-resolution microscopy

One important issue in assessing colocalization in IAV is whether it actually implies physical interaction or similar localization by chance. In this work, we treated segment spots with a maximum distance of 300 nm in the x-y plane to each other as "interacting", a similar value compared to other FISH studies in the field [90]. Confocal resolution will not be significantly more precise than these values, while it is clear that colocalization in such areas might be an over-interpretation of the data. We are confident that this was not the case here, at least for the majority of colocalizations detected because (1) we corrected for MSCs with double occurrences of any segment which only comprised $< 5\%$, (2) assessment of non-colocalizing targets such as viral mRNA showed much weaker MSC ranks even for very high densities (data not shown) and (3) resulting strong pathways in network diagrams would be highly unlikely if no functional interaction was underlying.

Nevertheless, to find out more about this issue and also to evaluate a potential combination of MuSeq-FISH with super resolution stimulated emission depletion (STED) microscopy, first proof-of-concept experiments have been undertaken together with Maria Loidolt and Franziska Winter (AG Hell, MPI for Biophysical Chemistry, Göttingen, Germany) to test STED microscopy on our stained IAV samples (see Fig. 55). While microscopy was successful in principle, optimal application of MuSeq-FISH would need precise control over sample position along the z-axis in subsequent runs, which was not achieved yet. Therefore, only one run of a typical MuSeq-FISH experiment was conducted here.

Comparison of confocal and STED resolutions showed a significant increase in detail levels, and in some cases revealed additional spots hidden in the point spread function (PSF) of confocal microscopy, an indicator of data that would have been analyzed differently. However, a major fraction of spots remained single *puncta*, so that one must ask whether the gain in resolution outweighs the much higher imaging speed in confocal or (in this case) spinning disk set-ups.
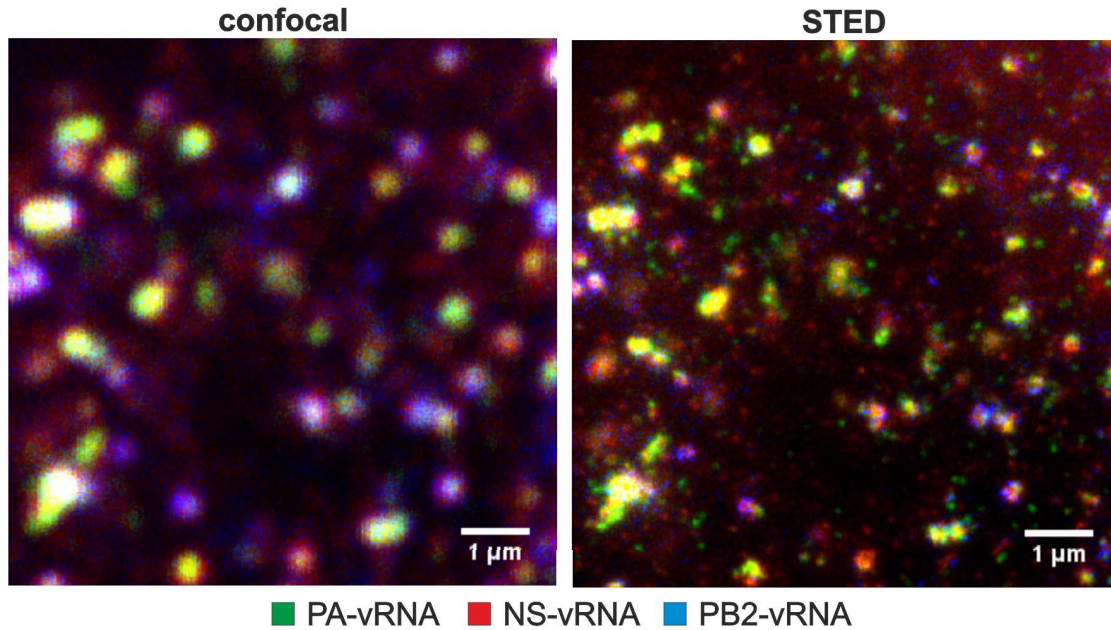
Figure 55 – **Comparison of confocal and STED imaging of three distinct vRNA segments.** A549 cells were infected with A/Panama at an MOI of 5 and fixated 10 h.p.i. and subsequently permeabilized. Then, FISH staining was conducted for PA, NS and PB2-vRNAs. Images were taken by Maria Loidolt using a custom-built STED setup, both without (left) and with (right) the use of the STED laser for resolution enhancement. Obviously, STED resolution looked much crisper, and in some cases revealed multiple spots hidden within the PSF of confocal imaging. The STED image shows more counts of PA-vRNA (green), which is caused by different dynamic ranges and not actual spot counts. However, most data would be detected similarly using the spot detection and image analysis pipeline described for the experiments above.

5.2.2 Live cell imaging in multiple channels

We have seen that the lack of actual kinetics data hindered the detailed analysis of preferred pathways in this work. Therefore, future studies will greatly benefit from the introduction of a highly multiplexed live cell imaging system. Sequential multiplexing, as is the basis of MuSeq-FISH will not be an option for such an approach as the key to understanding packaging kinetics would lie in direct observations of the complete genomic set of a virus, i.e. labelling all IAV vRNAs in the same run. RNA-FISH in its standard version is not capable of live cell imaging, because it requires a wash step to remove unbound probes which otherwise lead to a strong background signal. Techniques like the substitution of the always-on standard fluorophores bound to the probes with switchable dyes (like FIT-probes [158])

which increase in brightness upon target binding have led to the establishment of wash-free protocols. However, these currently suffer from problematic entry into live cells. In addition, the variety of colors to choose from is limited and would most likely need to increase to allow high levels of multiplexing. Finally, multiplexing without sequential stainings requires (in the case of IAV) to at least distinguish eight different spectral channels. There are approaches in spectral unmixing, where spectral fingerprints of different fluorophores are employed to distinguish even dyes with partially overlapping spectra. One can also envision using different lifetimes of dyes to increase precision of such fingerprints. While such approaches look bright in theory, they still await robust technical implementation for the acquisition of reliable data.

5.2.3 Comparison with more viral strains

Another key insight from our analysis was the strong variability of packaging pathways, even among relatively similar viral strains of IAV. Comparison with the literature revealed that pronounced differences were also present to a variety of laboratory strains that had been investigated before [57, 119, 103, 73], and as well among those strains themselves. Subsequent studies should therefore focus on using MuSeq-FISH for a complete staining of viral RNA for heavily studied viruses like A/WSN, A/PR8, A/Finch and A/Moscow to allow better comparison. Design and synthesis of strain-specific probe sets will be lengthy and expensive, but might be eased by the use of systems like the recently introduced smiFISH [168], where dyes can be partially reused. Once several probe sets are at hand, reassortment within cells could be visualized when co-infecting samples with two viral strains known to frequently reassort. Such studies will be very beneficial for our understanding of why or why not specific segments reassort more easily, or some segments prefer to reassort preferably together with other segments.

5.2.4 Comparison with sub-networks of A/Panama

It will also be worthwhile to optimize the usage of the produced mRNA-only A/Panama plasmid system. A completely functional and easy-to-use plasmid system that allows to transfect defined combinations of vRNA in an otherwise functional infection environment will be very beneficial for the further investigation of genome packaging in A/Panama. The system could in principle also be implemented for plasmid systems for other viral strains.

6 | Bibliography

[1] New virus gets official name, influenza D, 2016. URL `https://www.sciencedaily.com/releases/2016/09/160901140433.htm`.

[2] Maria Joao Amorim, Emily a Bruce, Eliot K C Read, Agnes Foeglein, Robert Mahen, Amanda D Stuart, and Paul Digard. A Rab11- and microtubule-dependent mechanism for cytoplasmic transport of influenza A virus viral RNA. *Journal of virology*, 85(9): 4143–56, may 2011. ISSN 1098-5514. doi: 10.1128/JVI.02606-10. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3126276{&}tool=pmcentrez{&}rendertype=abstract`.

[3] Nicholas W Baetz and James R Goldenring. Rab11-family interacting proteins define spatially and temporally distinct regions within the dynamic Rab11a-dependent recycling system. *Molecular biology of the cell*, 24(5):643–58, mar 2013. ISSN 1939-4586. doi: 10.1091/mbc.E12-09-0659. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3583667{&}tool=pmcentrez{&}rendertype=abstract`.

[4] Steven F Baker, Aitor Nogales, Courtney Finch, Kevin M Tuffy, William Domm, Daniel R Perez, David J Topham, and Luis Martínez-Sobrido. Influenza A and B intertypic reassortment through compatible viral packaging signals. *Journal of virology*, (July), jul 2014. ISSN 1098-5514. doi: 10.1128/JVI.01440-14. URL `http://www.ncbi.nlm.nih.gov/pubmed/25008914`.

[5] Nico Battich, Thomas Stoeger, and Lucas Pelkmans. Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nature methods*, 10(11):1127–33, nov 2013. ISSN 1548-7105. doi: 10.1038/nmeth.2657. URL `http://www.ncbi.nlm.nih.gov/pubmed/24097269`.

[6] F Baudin, C Bach, S Cusack, and R W Ruigrok. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *The EMBO Journal*, 13(13):3158, 1994.

147

[7] J G J Bauman, J Wiegant, P Borst, and P Van Duijn. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. *Experimental cell research*, 128 (2):485–490, 1980.

[8] Edouard Bertrand, Pascal Chartrand, Matthias Schaefer, Shailesh M. Shenoy, Robert H. Singer, and Roy M. Long. Localization of ASH1 mRNA Particles in Living Yeast. *Molecular Cell*, 2(4):437–445, 1998. ISSN 10972765. doi: 10.1016/S1097-2765(00)80143-4. URL http://www.ncbi.nlm.nih.gov/pubmed/9809065{%}5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S1097276500801434.

[9] Dieter Blaas, Erik Patzelt, and Ernst Kuechler. Cap-recognizing protein of influenza virus. *Virology*, 116(1):339–348, 1982.

[10] Michele Bouloy, Stephen J Plotch, and Robert M Krug. Both the 7-methyl and the 2'-O-methyl groups in the cap of mRNA strongly influence its ability to act as primer for influenza virus RNA transcription. *Proceedings of the National Academy of Sciences*, 77(7):3952–3956, 1980.

[11] Nicole M. Bouvier and Peter Palese. The biology of influenza viruses. *Vaccine*, 26(SUPPL. 4):49–53, 2008. ISSN 0264410X. doi: 10.1016/j.vaccine.2008.07.039.

[12] Janet Braam, Ismo Ulmanen, and Robert M Krug. Molecular model of a eucaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription. *Cell*, 34(2):609–618, 1983.

[13] Christopher B Brooke, William L Ince, Jens Wrammert, Rafi Ahmed, Patrick C Wilson, Jack R Bennink, and Jonathan W Yewdell. Most influenza a virions fail to express at least one essential viral protein. *Journal of virology*, 87(6):3155–62, mar 2013. ISSN 1098-5514. doi: 10.1128/JVI.02284-12. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3592173{&}tool=pmcentrez{&}rendertype=abstract.

[14] Deborah A Brown and John K Rose. Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. *Cell*, 68(3):533–544, 1992.

[15] Emily a Bruce, Paul Digard, and Amanda D Stuart. The Rab11 pathway is required for influenza A virus budding and filament formation. *Journal of virology*, 84(12):5848–59, jun 2010. ISSN 1098-5514. doi: 10.1128/JVI.00307-10. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2876627{&}tool=pmcentrez{&}rendertype=abstract.

[16] Linda Brunotte, Joe Flies, Hardin Bolte, Peter Reuther, Frank Vreede, and Martin Schwemmle. The nuclear export protein of H5N1 influenza A viruses recruits Matrix 1 (M1) protein to the viral ribonucleoprotein to mediate nuclear export. *Journal of Biological Chemistry*, 289(29):20067–20077, 2014.

[17] Matthew Bui, Gary Whittaker, and Ari Helenius. Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins. *Journal of virology*, 70(12):8391–8401, 1996.

[18] Matthew Bui, Elizabeth G Wills, Ari Helenius, and Gary R Whittaker. Role of the influenza virus M1 protein in nuclear export of viral ribonucleoproteins. *Journal of virology*, 74(4):1781–1786, 2000.

[19] Maria R Castrucci and Yoshihiro Kawaoka. Reverse genetics system for generation of an influenza A virus mutant containing a deletion of the carboxyl-terminal residue of M2 protein. *Journal of virology*, 69(5):2725–2728, 1995.

[20] PokMan Chan, Tony Yuen, Frederique Ruf, Javier Gonzalez-Maeso, and Stuart C Sealfon. Method for multiplex cellular detection of mRNAs using quantum dot fluorescent in situ hybridization. *Nucleic acids research*, 33(18): e161—-e161, 2005.

[21] Lin Chao. Evolution of sex in RNA viruses. *Trends in ecology & evolution*, 7 (5):147–151, 1992.

[22] Geoffrey P. Chase, Marie Anne Rameix-Welti, Aurelija Zvirbliene, Gintautas Zvirblis, Veronika Götz, Thorsten Wolff, Nadia Naffakh, and Martin Schwemmle. Influenza virus ribonucleoprotein complexes gain preferential access to cellular export machinery through chromatin targeting. *PLoS Pathogens*, 7(9), 2011. ISSN 15537366. doi: 10.1371/journal.ppat.1002187.

[23] Benjamin J Chen, George P Leser, Eiji Morita, and Robert a Lamb. Influenza virus hemagglutinin and neuraminidase, but not the matrix protein, are required for assembly and budding of plasmid-derived virus-like particles. *Journal of virology*, 81(13):7111–23, jul 2007. ISSN 0022-538X. doi: 10.1128/ JVI.00361-07. URL http://www.pubmedcentral.nih.gov/articlerender. fcgi?artid=1933269{&}tool=pmcentrez{&}rendertype=abstract.

[24] Benjamin J Chen, George P Leser, David Jackson, and Robert A Lamb. The influenza virus M2 protein cytoplasmic tail interacts with the M1 protein and influences virus assembly at the site of virus budding. *Journal of virology*, 82 (20):10059–10070, 2008.

[25] Kok Hao Chen, Alistair N Boettiger, Jeffrey R Moffitt, Siyuan Wang, and Xiaowei Zhuang. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science (New York, N.Y.)*, 348(6233): aaa6090, 2015. ISSN 1095-9203. doi: 10.1126/science.aaa6090. URL http://www.ncbi.nlm.nih.gov/pubmed/25858977{%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4662681.

[26] Weisan Chen, Paul A Calvo, Daniela Malide, James Gibbs, Ulrich Schubert, Igor Bacik, Sameh Basta, Robert O'Neill, Jeanne Schickli, Peter Palese, and Others. A novel influenza A virus mitochondrial protein that induces cell death. *Nature medicine*, 7(12):1306–1312, 2001.

[27] Yi-ying Chou, Reza Vafabakhsh, Sultan Doganay, Qinshan Gao, Taekjip Ha, and Peter Palese. One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 109(23): 9101–6, jun 2012. ISSN 1091-6490. doi: 10.1073/pnas.1206069109. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3384215{&}tool=pmcentrez{&}rendertype=abstract.

[28] Yi-ying Chou, Nicholas S Heaton, Qinshan Gao, Peter Palese, Robert H Singer, Robert Singer, and Timothée Lionnet. Colocalization of different influenza viral RNA segments in the cytoplasm before viral budding as shown by single-molecule sensitivity FISH analysis. *PLoS pathogens*, 9(5): e1003358, jan 2013. ISSN 1553-7374. doi: 10.1371/journal.ppat.1003358. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3649991{&}tool=pmcentrez{&}rendertype=abstract.

[29] C M Chu, I M Dawson, and W J Elford. Filamentous forms associated with newly isolated influenza virus. *Lancet*, 1(6554):602, 1949. ISSN 01406736. doi: 10.1016/S0140-6736(49)91699-2.

[30] F Ciampor, C A Thompson, S Grambas, and A J Hay. Regulation of pH by the M2 protein of influenza A viruses. *Virus research*, 22(3):247–258, 1992.

[31] Joanna C A Cobbin, Chi Ong, Erin Verity, Brad P Gilbertson, Steven P Rockman, and Lorena E Brown. Influenza virus PB1 and neuraminidase gene segments can cosegregate during vaccine reassortment driven by interactions in the PB1 coding region. *Journal of virology*, 88(16):8971–8980, 2014.

[32] Ahmet F Coskun and Long Cai. Dense transcript profiling in single cells by image correlation decoding. *Nature Methods*, (June):1–6, 2016. ISSN 1548-7091. doi: 10.1038/nmeth.3895. URL http://www.nature.com/doifinder/10.1038/nmeth.3895.

[33] Bernadette Crescenzo-Chaigne, Cyril V. S. Barbezange, Stéphane Léandri, Camille Roquin, Camille Berthault, and Sylvie van der Werf. Incorporation of the influenza A virus NA segment into virions does not require cognate non-coding sequences. *Scientific Reports*, 7(August 2016):43462, 2017. ISSN 2045-2322. doi: 10.1038/srep43462. URL `http://www.nature.com/articles/srep43462`.

[34] G Csárdi and T Nepusz. The igraph software package for complex network research. *InterJournal Complex Systems*, 1695:1695, 2006.

[35] Chenghua Cui, Wei Shu, and Peining Li. Fluorescence in situ hybridization: cell-based genetic diagnostic and research applications. *Frontiers in Cell and Developmental Biology*, 4, 2016.

[36] Replication Cycle and Sequence Comparisons. Evolution and ecology of influenza A viruses. *Australian Veterinary Journal*, 69(9): 235–235, 1992. ISSN 0005-0423. doi: 10.1111/j.1751-0813.1992. tb09940.x. URL `http://www.blackwell-synergy.com/doi/abs/10.1111/j.1751-0813.1992.tb09940.x`.

[37] Susana de la Luna, Concepción Mart\'\inez, and Juan Ort\'\in. Molecular cloning and sequencing of influenza virus A/Victoria/3/75 polymerase genes: sequence evolution and prediction of possible functional domains. *Virus research*, 13(2):143–155, 1989.

[38] Tao Deng, Jane Sharps, Ervin Fodor, and George G Brownlee. In vitro assembly of PB2 with a PB1-PA dimer supports a new model of assembly of influenza A virus polymerase subunits into a functional trimeric complex. *Journal of virology*, 79(13):8669–8674, 2005.

[39] Alexandre Dias, Denis Bouvier, Thibaut Crépin, Andrew A McCarthy, Darren J Hart, Florence Baudin, Stephen Cusack, and Rob W H Ruigrok. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*, 458(7240):914–918, 2009.

[40] Nicola R Donelan, Christopher F Basler, and Adolfo Garc\'\ia-Sastre. A recombinant influenza A virus expressing an RNA-binding-defective NS1 protein induces high levels of beta interferon and is attenuated in mice. *Journal of virology*, 77(24):13257–13266, 2003.

[41] Emmanuel Dos Santos Afonso, Nicolas Escriou, India Leclercq, Sylvie Van Der Werf, and Nadia Naffakh. The generation of recombinant influenza A viruses expressing a PB2 fusion protein requires the conservation of a

packaging signal overlapping the coding and noncoding regions at the 5 prime end of the PB2 segment. *Virology*, 341(1):34–46, 2005. ISSN 00426822. doi: 10.1016/j.virol.2005.06.040.

[42] Amie J Eisfeld, Eiryo Kawakami, Tokiko Watanabe, Gabriele Neumann, and Yoshihiro Kawaoka. RAB11A is essential for transport of the influenza virus genome to the plasma membrane. *Journal of virology*, 85 (13):6117–26, jul 2011. ISSN 1098-5514. doi: 10.1128/JVI.00378-11. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3126513{&}tool=pmcentrez{&}rendertype=abstract.

[43] Boris Essere, Matthieu Yver, Cyrille Gavazzi, Olivier Terrier, Catherine Isel, Emilie Fournier, Fabienne Giroux, Julien Textoris, Thomas Julien, Clio Socratous, Manuel Rosa-Calatrava, Bruno Lina, Roland Marquet, and Vincent Moules. Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 110(40): E3840–8, oct 2013. ISSN 1091-6490. doi: 10.1073/pnas.1308649110. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3791739{&}tool=pmcentrez{&}rendertype=abstract.

[44] Mindaugas Juozapaiti Étori Aguiar Moreira, Anna Weber, Hardin Bolte, Larissa Kolesnikova, Sebastian Giese, Seema Lakdawala, Martin Beer, Gert Zimmer, Adolfo García-Sastre, Martin Schwemmle. A conserved influenza A virus nucleoprotein code controls specific viral genome packaging. pages 1–24, 2016. ISSN 2041-1723. doi: 10.1038/ncomms12861.

[45] Lavina Faleiro and Yuri Lazebnik. Caspases disrupt the nuclear-cytoplasmic barrier. *J Cell Biol*, 151(5):951–960, 2000.

[46] A. M. Femino. Visualization of Single RNA Transcripts in Situ. *Science*, 280(5363):585–590, 1998. ISSN 00368075. doi: 10.1126/science.280.5363.585. URL http://www.sciencemag.org/cgi/doi/10.1126/science.280.5363.585.

[47] E Fodor. The RNA polymerase of influenza a virus: mechanisms of viral transcription and replication. *Acta virologica*, (2012):113–122, 2012. doi: 10.4149/av. URL http://europepmc.org/abstract/MED/23600869.

[48] Judith M Fonville, Nicolle Marshall, Hui Tao, John Steel, and Anice C Lowen. Influenza virus reassortment is enhanced by semi-infectious particles but can be suppressed by defective interfering particles. *PLoS Pathog*, 11(10): e1005204, 2015.

[49] Ron a M Fouchier, Vincent Munster, Anders Wallensten, Theo M Bestebroer, Sander Herfst, Derek Smith, F Guus, Björn Olsen, Albert D M E Osterhaus, and Guus F Rimmelzwaan. Characterization of a Novel Influenza A Virus Hemagglutinin Subtype (H16) Obtained from Black-Headed Gulls Characterization of a Novel Influenza A Virus Hemagglutinin Subtype (H16) Obtained from Black-Headed Gulls. *J. Virol.*, 79(5):2814–2822, 2005. doi: 10.1128/JVI.79.5.2814.

[50] Emilie Fournier, Vincent Moules, Boris Essere, Jean-Christophe Paillart, Jean-Daniel Sirbat, Annie Cavalier, Jean-Paul Rolland, Daniel Thomas, Bruno Lina, Catherine Isel, and Roland Marquet. Interaction network linking the human H3N2 influenza A virus genomic RNA segments. *Vaccine*, 30(51): 7359–67, dec 2012. ISSN 1873-2518. doi: 10.1016/j.vaccine.2012.09.079. URL `http://www.ncbi.nlm.nih.gov/pubmed/23063835`.

[51] Emilie Fournier, Vincent Moules, Boris Essere, Jean-Christophe Paillart, Jean-Daniel Sirbat, Catherine Isel, Annie Cavalier, Jean-Paul Rolland, Daniel Thomas, Bruno Lina, and Roland Marquet. A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic acids research*, 40(5):2197–209, mar 2012. ISSN 1362-4962. doi: 10.1093/nar/gkr985. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid= 3300030{&}tool=pmcentrez{&}rendertype=abstract`.

[52] Ken Fujii, Yutaka Fujii, Takeshi Noda, Yukiko Muramoto, Tokiko Watanabe, Ayato Takada, Hideo Goto, Taisuke Horimoto, and Yoshihiro Kawaoka. Importance of both the coding and the segment-specific noncoding regions of the influenza A virus NS segment for its efficient incorporation into virions. *Journal of virology*, 79(6):3766–3774, 2005.

[53] Yutaka Fujii, Hideo Goto, Tokiko Watanabe, Tetsuya Yoshida, and Yoshihiro Kawaoka. Selective incorporation of influenza virus RNA segments into virions. *Proceedings of the National Academy of Sciences*, 100(4):2002–2007, 2003.

[54] Gülsah Gabriel, Astrid Herwig, and Hans-Dieter Klenk. Interaction of polymerase subunit PB2 and NP with importin alpha is a determinant of host range of influenza A virus. *PLoS Pathog*, 4(2):e11, 2008.

[55] Joseph G Gall and Mary Lou Pardue. Formation and Detection of Rna-Dna Hybrid Molecules in Cytological Preparations*. *Proc Natl Acad Sci U S A.*, 63(1):378–383, 1969. ISSN 0027-8424. doi: 10.1073/pnas.63.2.378.

[56] Q Gao, E W Brydon, and P Palese. A seven-segmented influenza A virus expressing the influenza C virus glycoprotein HEF. *J Virol*, 82(13):6419–6426, 2008. ISSN 1098-5514. doi: 10.1128/jvi.00514-08. URL http://jvi.asm.org/content/82/13/6419.full.pdf.

[57] Qinshan Gao, Yi-Ying Chou, Sultan Doganay, Reza Vafabakhsh, Taekjip Ha, and Peter Palese. The influenza A virus PB2, PA, NP, and M segments play a pivotal role during genome packaging. *Journal of virology*, 86(13):7043–51, jul 2012. ISSN 1098-5514. doi: 10.1128/JVI.00662-12. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3416350{&}tool=pmcentrez{&}rendertype=abstract.

[58] Cyrille Gavazzi, Matthieu Yver, Catherine Isel, Redmond P Smyth, Manuel Rosa-Calatrava, Bruno Lina, Vincent Moulès, and Roland Marquet. A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proceedings of the National Academy of Sciences of the United States of America*, 110(41): 16604–9, oct 2013. ISSN 1091-6490. doi: 10.1073/pnas.1314419110. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3799358{&}tool=pmcentrez{&}rendertype=abstract.

[59] Michael J Gerdes, Christopher J Sevinsky, Anup Sood, Sudeshna Adak, Musodiq O Bello, Alexander Bordwell, Ali Can, Alex Corwin, Sean Dinn, Robert J Filkins, Denise Hollman, Vidya Kamath, Sireesha Kaanumalle, Kevin Kenny, Melinda Larsen, Michael Lazare, Qing Li, Christina Lowes, Colin C McCulloch, Elizabeth McDonough, Michael C Montalto, Zhengyu Pang, Jens Rittscher, Alberto Santamaria-Pang, Brion D Sarachan, Maximilian L Seel, Antti Seppo, Kashan Shaikh, Yunxia Sui, Jingyu Zhang, and Fiona Ginty. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proceedings of the National Academy of Sciences of the United States of America*, 110 (29):11982–7, jul 2013. ISSN 1091-6490. doi: 10.1073/pnas.1300136110. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3718135{&}tool=pmcentrez{&}rendertype=abstract.

[60] Sebastian Giese, Hardin Bolte, and Martin Schwemmle. The Feat of Packaging Eight Unique Genome Segments. *Viruses*, 8(6):165, 2016. ISSN 1999-4915. doi: 10.3390/v8060165. URL http://www.mdpi.com/1999-4915/8/6/165.

[61] Ruo Xu Gu, Limin Angela Liu, and Dong Qing Wei. Structural and energetic analysis of drug inhibition of the influenza A M2 proton channel. *Trends in Pharmacological Sciences*, 34(10):571–580, 2013. ISSN 01656147. doi: 10.

1016/j.tips.2013.08.003. URL http://dx.doi.org/10.1016/j.tips.2013.08.003.

[62] Benjamin G Hale, Richard E Randall, Juan Ortín, and David Jackson. The multifunctional NS1 protein of influenza A viruses. *The Journal of general virology*, 89(Pt 10):2359–76, oct 2008. ISSN 0022-1317. doi: 10.1099/vir.0.2008/004606-0. URL http://www.ncbi.nlm.nih.gov/pubmed/18796704.

[63] Ivan Haralampiev. *Multiplexed Fishing of Influenza A Virus Genomic RNAs Opens Access to Influenza Segment Bundling and a Novel Pan-Selective IAV Infection Marker*. PhD thesis, Humboldt University Berlin, 2017.

[64] Eriko Hatada, Masakazu Hasegawa, Jun Mukaigawa, Kazufumi Shimizu, and Ryuji Fukuda. Control of influenza virus gene expression: quantitative analysis of each viral RNA species in infected cells. *The Journal of Biochemistry*, 105(4):537–546, 1989.

[65] Masato Hatta, Peng Gao, Peter Halfmann, and Yoshihiro Kawaoka. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science*, 293(5536):1840–1842, 2001.

[66] a J Hay, V Gregory, a R Douglas, and Y P Lin. The evolution of human influenza viruses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 356(i): 1861–70, 2001. ISSN 0962-8436. doi: 10.1098/rstb.2001.0999. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1088562{&}tool=pmcentrez{&}rendertype=abstract.

[67] Carolyn Herz, Edward Stavnezer, Robert M Krug, and Theodore Gurney. Influenza virus, an RNA virus, synthesizes its messenger RNA in the nucleus of infected cells. *Cell*, 26(3):391–400, 1981.

[68] Malte Hilsch, Björn Goldenbogen, Christian Sieben, Chris T Höfer, Jürgen P Rabe, Edda Klipp, Andreas Herrmann, and Salvatore Chiantia. Influenza a matrix protein m1 multimerizes upon binding to lipid membranes. *Biophysical Journal*, 107(4):912–923, 2014. ISSN 15420086. doi: 10.1016/j.bpj.2014.06.042.

[69] Edward C Holmes, Elodie Ghedin, Naomi Miller, Jill Taylor, Yiming Bao, Kirsten St George, Bryan T Grenfell, Steven L Salzberg, Claire M Fraser, David J Lipman, and Others. Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol*, 3(9):e300, 2005.

[70] E. C. Hutchinson and E. Fodor. Nuclear import of the influenza A virus transcriptional machinery. *Vaccine*, 30(51):7353–7358, 2012. ISSN 0264410X. doi: 10.1016/j.vaccine.2012.04.085. URL http://dx.doi.org/10.1016/j.vaccine.2012.04.085.

[71] Edward C Hutchinson and Ervin Fodor. Transport of the influenza virus genome from nucleus to nucleus. *Viruses*, 5(10): 2424–46, oct 2013. ISSN 1999-4915. doi: 10.3390/v5102424. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3814596{&}tool=pmcentrez{&}rendertype=abstract.

[72] Edward C Hutchinson, Martin D Curran, Eliot K Read, Julia R Gog, and Paul Digard. Mutational analysis of cis-acting RNA signals in segment 7 of influenza A virus. *Journal of virology*, 82(23): 11869–79, dec 2008. ISSN 1098-5514. doi: 10.1128/JVI.01634-08. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2583641{&}tool=pmcentrez{&}rendertype=abstract.

[73] Edward C Hutchinson, Helen M Wise, Katerine Kudryavtseva, Martin D Curran, and Paul Digard. Characterisation of influenza A viruses with mutations in segment 5 packaging signals. *Vaccine*, 27(45):6270–5, oct 2009. ISSN 1873-2518. doi: 10.1016/j.vaccine.2009.05.053. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2771075{&}tool=pmcentrez{&}rendertype=abstract.

[74] Catherine Isel, Sandie Munier, and Nadia Naffakh. Experimental Approaches to Study Genome Packaging of Influenza A Viruses. *Viruses*, 8(8):218, 2016. ISSN 1999-4915. doi: 10.3390/v8080218. URL http://www.mdpi.com/1999-4915/8/8/218.

[75] B W Jagger, H M Wise, J C Kash, K-A Walters, N M Wills, Y-L Xiao, R L Dunfee, L M Schwartzman, A Ozinsky, G L Bell, and Others. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science*, 337(6091):199–204, 2012.

[76] Lars Martin Jakt, Satoko Moriwaki, and Shinichi Nishikawa. A continuum of transcriptional identities visualized by combinatorial fluorescent in situ hybridization. *Development (Cambridge, England)*, 140(1):216–25, jan 2013. ISSN 1477-9129. doi: 10.1242/dev.086975. URL http://www.ncbi.nlm.nih.gov/pubmed/23175635.

[77] Authors Je, Hyuk Lee, Evan R Daugharthy, Jonathan Scheiman, Reza Kalhor, Derek T Peters, Brian M Turczyk, Adam H Marblestone, and Samuel A

Inverso. Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science*, (December), 2013.

[78] Philip A Jennings, John T Finch, Greg Winter, and James S Robertson. Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? *Cell*, 34(2):619–627, 1983.

[79] Kalle Jongen. *Charakterisierung des Influenza A RNA-Haushalts im Infektionsverlauf von A549 Zellen mittels qRT-PCR*. PhD thesis, Humboldt University Berlin, 2016.

[80] Núria Jorba, Rocío Coloma, and Juan Ortín. Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. *PLoS pathogens*, 5(5):e1000462, may 2009. ISSN 1553-7374. doi: 10.1371/journal.ppat.1000462. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2682650{&}tool=pmcentrez{&}rendertype=abstract`.

[81] Atsushi Kawaguchi, Ken Matsumoto, and Kyosuke Nagata. YB-1 functions as a porter to lead influenza virus ribonucleoprotein complexes to microtubules. *Journal of virology*, 86(20):11086–95, oct 2012. ISSN 1098-5514. doi: 10.1128/JVI.00453-12. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3457152{&}tool=pmcentrez{&}rendertype=abstract`.

[82] Atsushi Kawaguchi, Mikako Hirohama, Yoshimi Harada, Suguru Osari, and Kyosuke Nagata. Influenza Virus Induces Cholesterol-Enriched Endocytic Recycling Compartments for Budozone Formation via Cell Cycle-Independent Centrosome Maturation. *PLOS Pathogens*, 11(11):e1005284, 2015. ISSN 1553-7374. doi: 10.1371/journal.ppat.1005284. URL `http://dx.plos.org/10.1371/journal.ppat.1005284`.

[83] Eiryo Kawakami, Tokiko Watanabe, Ken Fujii, Hideo Goto, Shinji Watanabe, Takeshi Noda, and Yoshihiro Kawaoka. Strand-specific real-time RT-PCR for distinguishing influenza vRNA, cRNA, and mRNA. *Journal of Virological Methods*, 173(1):1–6, 2011. ISSN 01660934. doi: 10.1016/j.jviromet.2010.12.014. URL `http://dx.doi.org/10.1016/j.jviromet.2010.12.014`.

[84] O. Kistner, K. Muller, and C. Scholtissek. Differential phosphorylation of the nucleoprotein of influenza A viruses. *Journal of General Virology*, 70(9): 2421–2431, 1989. ISSN 00221317. doi: 10.1099/0022-1317-70-9-2421.

[85] T a Klar, E Engel, and S W Hell. Breaking Abbe's diffraction resolution limit in fluorescence microscopy with stimulated emission depletion beams of

various shapes. *Phys. Rev.*, 64(6 Pt 2):066613, 2001. ISSN 1063-651X. doi: 10.1103/PhysRevE.64.066613.

[86] Vanessa Körner. *Vergleichende qRT-PCR und MuSeq-FISH Untersuchungen von Influenza A wildtyp- und reassortanten-RNA hinsichtlich ihrer Abundanz und Lokalisation in A549-Zellen während des Infektionsverlaufes*. PhD thesis, 2016.

[87] Robert M Krug and Adolfo García-Sastre. The NS1 protein: A master regulator of host and viral functions. In *Textbook of Influenza*, pages 114–132. John Wiley & Sons, Ltd, 2013. ISBN 9781118636817. doi: 10.1002/9781118636817. ch7. URL http://dx.doi.org/10.1002/9781118636817.ch7.

[88] Robert M. Krug, Barbara A. Broni, and Michele Bouloy. Are the 5' ends of influenza viral mRNAs synthesized in vivo donated by host mRNAs? *Cell*, 18(2):329–334, 1979. ISSN 00928674. doi: 10.1016/0092-8674(79)90052-7.

[89] Alice Labaronne, Christopher Swale, Alexandre Monod, Guy Schoehn, Thibaut Crepin, and Rob W H Ruigrok. Binding of RNA by the nucleoproteins of influenza viruses A and B. *Viruses*, 8(9):1–14, 2016. ISSN 19994915. doi: 10.3390/v8090247.

[90] Seema S Lakdawala, Yicong Wu, Peter Wawrzusin, Juraj Kabat, Andrew J Broadbent, Elaine W Lamirande, Ervin Fodor, Nihal Altan-Bonnet, Hari Shroff, and Kanta Subbarao. Influenza a virus assembly intermediates fuse in the cytoplasm. *PLoS pathogens*, 10(3): e1003971, mar 2014. ISSN 1553-7374. doi: 10.1371/journal.ppat.1003971. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3946384{&}tool=pmcentrez{&}rendertype=abstract.

[91] George P Leser and Robert A Lamb. Influenza virus assembly and budding in raft-derived microdomains: a quantitative analysis of the surface distribution of HA, NA and M2 proteins. *Virology*, 342(2):215–227, 2005.

[92] J. M. Levsky. Single-Cell Gene Expression Profiling. *Science*, 297(5582): 836–840, 2002. ISSN 00368075. doi: 10.1126/science.1072241. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1072241.

[93] J. M. Levsky. Fluorescence in situ hybridization: past, present and future. *Journal of Cell Science*, 116(14):2833–2838, 2003. ISSN 0021-9533. doi: 10.1242/jcs.00633. URL http://jcs.biologists.org/cgi/doi/10.1242/jcs.00633.

[94] Chengjun Li, Masato Hatta, Shinji Watanabe, Gabriele Neumann, and Yoshihiro Kawaoka. Compatibility among polymerase subunit proteins is a restricting factor in reassortment between equine H7N7 and human H3N2 influenza viruses. *Journal of virology*, 82(23):11880–11888, 2008.

[95] Zhuoran Li, Ce Fang, Yuanyuan Su, Hongmei Liu, Fengchao Lang, Xin Li, Guijun Chen, Danfeng Lu, and Jumin Zhou. Visualizing the replicating HSV-1 virus using STED super-resolution microscopy. *Virology journal*, 13 (1):65, 2016. ISSN 1743-422X. doi: 10.1186/s12985-016-0521-7. URL `http://virologyj.biomedcentral.com/articles/10.1186/s12985-016-0521-7`.

[96] Yuhong Liang, Taoying Huang, Hinh Ly, Tristram G Parslow, and Yuying Liang. Mutational analyses of packaging signals in influenza virus PA, PB1, and PB2 genomic RNA segments. *Journal of virology*, 82 (1):229–36, jan 2008. ISSN 1098-5514. doi: 10.1128/JVI.01541-07. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2224372{&}tool=pmcentrez{&}rendertype=abstract`.

[97] Yuying Liang, Ying Hong, and Tristram G Parslow. cis-Acting packaging signals in the influenza virus PB1, PB2, and PA genomic RNA segments. *Journal of virology*, 79(16):10348–10355, 2005.

[98] Eric Lubeck, Ahmet F Coskun, Timur Zhiyentayev, Mubhij Ahmad, and Long Cai. Single-cell in situ RNA profiling by sequential hybridization. *Nature methods*, 11(4):360–1, apr 2014. ISSN 1548-7105. doi: 10.1038/nmeth.2892. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4085791{&}tool=pmcentrez{&}rendertype=abstract`.

[99] Ian M Mackay, Katherine E Arden, and Andreas Nitsche. Real-time PCR in virology. *Nucleic acids research*, 30(6):1292–1305, 2002.

[100] Caroline M. Mair, Tim Meyer, Katjana Schneider, Qiang Huang, Michael Veit, and Andreas Herrmann. A Histidine Residue of the Influenza Virus Hemagglutinin Controls the pH Dependence of the Conformational Change Mediating Membrane Fusion. *Journal of virology*, 88(22):13189–200, 2014. ISSN 1098-5514. doi: 10.1128/JVI.01704-14. URL `http://www.ncbi.nlm.nih.gov/pubmed/25187542{%}5Cnhttp://jvi.asm.org/cgi/doi/10.1128/JVI.01704-14`.

[101] Henju Marjuki, Mohammad I Alam, Christina Ehrhardt, Ralf Wagner, Oliver Planz, Hans-D Klenk, Stephan Ludwig, and Stephan Pleschka. Membrane accumulation of influenza A virus hemagglutinin triggers nuclear export of the

viral genome via protein kinase alpha-mediated activation of ERK signaling. *Journal of Biological Chemistry*, 281(24):16707–16715, 2006.

[102] Glenn a Marsh, Raheleh Hatami, and Peter Palese. Specific residues of the influenza A virus hemagglutinin viral RNA are important for efficient packaging into budding virions. *Journal of virology*, 81(18): 9727–36, sep 2007. ISSN 0022-538X. doi: 10.1128/JVI.01144-07. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid= 2045411{&}tool=pmcentrez{&}rendertype=abstract.

[103] Glenn a Marsh, Raúl Rabadán, Arnold J Levine, and Peter Palese. Highly conserved regions of influenza a virus polymerase gene segments are critical for efficient viral RNA packaging. *Journal of virology*, 82 (5):2295–304, mar 2008. ISSN 1098-5514. doi: 10.1128/JVI.02267-07. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid= 2258914{&}tool=pmcentrez{&}rendertype=abstract.

[104] Nicolle Marshall, Lalita Priyamvada, Zachary Ende, John Steel, and Anice C Lowen. Influenza virus reassortment occurs with high frequency in the absence of segment mismatch. *PLoS Pathog*, 9(6):e1003421, 2013.

[105] Kelsey Martin and Ari Heleniust. Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. *Cell*, 67(1):117–130, 1991.

[106] Daniel Mayer, Kaaweh Molawi, Luis Martínez-Sobrido, Alexander Ghanem, Stefan Thomas, Sacha Baginsky, Jonas Grossmann, Adolfo García-Sastre, and Martin Schwemmle. Identification of cellular interaction partners of the influenza virus ribonucleoprotein complex and polymerase complex using proteomic-based approaches. *Journal of Proteome Research*, 6(2):672–682, 2007. ISSN 15353893. doi: 10.1021/pr060432u.

[107] J W McCauley and B W Mahy. Structure and function of the influenza virus genome. *The Biochemical journal*, 211(2):281–94, 1983. ISSN 0264-6021. doi: 10.1002/9781118636817.ch8. URL http://www.ncbi.nlm.nih.gov/pubmed/6191756{%}5Cnhttp://www. pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1154358.

[108] Sarah M. McDonald, Martha I. Nelson, Paul E. Turner, and John T. Patton. Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nature Reviews Microbiology*, (May), 2016. ISSN 1740-1526. doi: 10.1038/nrmicro.2016.46. URL http://www.nature.com/doifinder/10. 1038/nrmicro.2016.46.

[109] James G McNally, Tatiana Karpova, John Cooper, and José Angel Conchello. Three-dimensional imaging by deconvolution microscopy. *Methods*, 19(3): 373–385, 1999.

[110] Ralf Mikut, Thomas Dickmeis, Wolfgang Driever, Pierre Geurts, Fred A Hamprecht, Bernhard X Kausler, Mar\'\ia J Ledesma-Carbayo, Raphaël Marée, Karol Mikula, Periklis Pantazis, and Others. Automated processing of zebrafish imaging data: a survey. *Zebrafish*, 10(3):401–421, 2013.

[111] Ji-Young Min, Shoudong Li, Ganes C Sen, and Robert M Krug. A site on the influenza A virus NS1 protein mediates both inhibition of PKR activation and temporal regulation of viral RNA synthesis. *Virology*, 363(1): 236–43, jun 2007. ISSN 0042-6822. doi: 10.1016/j.virol.2007.01.038. URL `http://www.ncbi.nlm.nih.gov/pubmed/17320139`.

[112] Arne Moeller, Robert N Kirchdoerfer, Clinton S Potter, Bridget Carragher, and Ian a Wilson. Organization of the influenza virus replication machinery. *Science (New York, N.Y.)*, 338(6114):1631–4, dec 2012. ISSN 1095-9203. doi: 10.1126/science.1227270. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3578580{&}tool=pmcentrez{&}rendertype=abstract`.

[113] Jeffrey R. Moffitt, Junjie Hao, Guiping Wang, Kok Hao Chen, Hazen P. Babcock, and Xiaowei Zhuang. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proceedings of the National Academy of Sciences*, 113(39):201612826, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1612826113. URL `http://www.pnas.org/lookup/doi/10.1073/pnas.1612826113`.

[114] Dorothée Moisy, Sergiy V Avilov, Yves Jacob, Brid M Laoide, Xingyi Ge, Florence Baudin, Nadia Naffakh, and Jean-Luc Jestin. HMGB1 protein binds to influenza virus nucleoprotein and promotes viral replication. *Journal of virology*, 86(17):9122–9133, 2012.

[115] Anne Moscona. Global transmission of oseltamivir-resistant influenza. *New England Journal of Medicine*, 360(10):953–956, 2009.

[116] Florian Mueller, Adrien Senecal, Katjana Tantale, Hervé Marie-Nelly, Nathalie Ly, Olivier Collin, Eugenia Basyuk, Edouard Bertrand, Xavier Darzacq, and Christophe Zimmer. FISH-quant: automatic counting of transcripts in 3D FISH images. *Nature Methods*, 10(4):277–278, 2013. ISSN 1548-7105. doi: 10.1038/nmeth.2406. URL `http://dx.doi.org/10.1038/nmeth.2406{%}5Cnpapers3://publication/doi/10.1038/nmeth.2406`.

[117] J Mukaigawa and DEBI P Nayak. Two signals mediate nuclear localization of influenza virus (A/WSN/33) polymerase basic protein 2. *Journal of virology*, 65(1):245–253, 1991.

[118] Kary Mullis, Fred Faloona, Stephen Scharf, R K Saiki, G T Horn, and H Erlich. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In *Cold Spring Harbor symposia on quantitative biology*, volume 51, pages 263–273. Cold Spring Harbor Laboratory Press, 1986.

[119] Yukiko Muramoto, Ayato Takada, and Ken Fujii. Hierarchy among viral RNA (vRNA) segments in their role in vRNA incorporation into influenza A virions. *Journal of . . .*, 2006. doi: 10.1128/JVI.80.5.2318. URL `http://jvi.asm.org/content/80/5/2318.short`.

[120] Yukiko Muramoto, Takeshi Noda, Eiryo Kawakami, Ramesh Akkina, and Yoshihiro Kawaoka. Identification of novel influenza A virus proteins translated from PA mRNA. *Journal of virology*, 87(5):2455–62, 2013. ISSN 1098-5514. doi: 10.1128/JVI.02656-12. URL `http://www.ncbi.nlm.nih.gov/pubmed/23236060`.

[121] Sumiho Nakatsu. Complete and Incomplete Genome Packaging of Influenza A and B Viruses. *mBio*, 7(5):1–7, 2016. ISSN 2150-7511. doi: 10.1128/mBio.01248-16.Editor.

[122] Surend T Nath and DEBI P Nayak. Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1). *Molecular and cellular biology*, 10(8):4139–4145, 1990.

[123] D P Nayak, K Tobita, J M Janda, A R Davis, and B K De. Homologous interference mediated by defective interfering influenza virus derived from a temperature-sensitive mutant of influenza virus. *Journal of virology*, 28(1):375–386, 1978.

[124] Gabriele Neumann, Maria R Castrucci, and Yoshihiro Kawaoka. Nuclear import and export of influenza virus nucleoprotein. *Journal of virology*, 71(12):9690–9700, 1997.

[125] Gabriele Neumann, Mark T Hughes, and Yoshihiro Kawaoka. Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *The EMBO journal*, 19(24):6751–6758, 2000.

[126] Alexey D Neverov, Ksenia V Lezhnina, Alexey S Kondrashov, and Georgii A Bazykin. Intrasubtype reassortments cause adaptive amino acid replacements in H3N2 influenza genes. *PLoS Genet*, 10(1):e1004037, 2014.

[127] Karl G Nicholson, Robert G Webster, Alan J Hay, and Others. *Textbook of influenza.* Blackwell Science Ltd, 1998.

[128] Amelia Nieto, Susana de la Luna, Juan Bárcena, Agustin Portela, and Juan Ortin. Complex structure of the nuclear translocation signal of influenza virus polymerase PA subunit. *Journal of general virology*, 75(1):29–36, 1994.

[129] Takeshi Noda and Yoshihiro Kawaoka. Packaging of influenza virus genome: robustness of selection. *Proceedings of the National Academy of Sciences of the United States of America*, 109(23): 8797–8, jun 2012. ISSN 1091-6490. doi: 10.1073/pnas.1206736109. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3384147{&}tool=pmcentrez{&}rendertype=abstract`.

[130] Takeshi Noda, Yukihiko Sugita, Kazuhiro Aoyama, Ai Hirase, Eiryo Kawakami, Atsuo Miyazawa, Hiroshi Sagara, and Yoshihiro Kawaoka. Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nature communications*, 3:639, jan 2012. ISSN 2041-1723. doi: 10.1038/ncomms1647. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3272569{&}tool=pmcentrez{&}rendertype=abstract`.

[131] Robert E O'Neill, Julie Talon, and Peter Palese. The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *The EMBO journal*, 17(1):288–296, 1998.

[132] A D M E Osterhaus, G F Rimmelzwaan, B E E Martina, T M Bestebroer, and R A M Fouchier. Influenza B virus in seals. *Nature*, 288(1995):1051–1053, 2000.

[133] Makoto Ozawa, Sarmila Basnet, Lisa M Burley, Gabriele Neumann, Masato Hatta, and Yoshihiro Kawaoka. Impact of amino acid mutations in PB2, PB1-F2, and NS1 on the replication and pathogenicity of pandemic (H1N1) 2009 influenza viruses. *Journal of virology*, 85(9):4596–4601, 2011.

[134] Peter Palese and David W Kingsbury. *Genetics of influenza viruses.* Springer Science & Business Media, 2012.

[135] Eun K Park, Maria R Castrucci, Allen Portner, and Yoshihiro Kawaoka. The M2 ectodomain is important for its incorporation into influenza A virions. *Journal of virology*, 72(3):2449–2455, 1998.

[136] Jasmine T Perez, Andrew Varble, Ravi Sachidanandam, Ivan Zlatev, Muthiah Manoharan, Adolfo Garc\'\ia-Sastre, and Others. Influenza A virus-generated small RNAs regulate the switch from transcription to replication. *Proceedings of the National Academy of Sciences*, 107(25):11525–11530, 2010.

[137] Jasmine T Perez, Ivan Zlatev, Shilpa Aggarwal, Sailakshmi Subramanian, Ravi Sachidanandam, Baek Kim, Muthiah Manoharan, and Others. A small-RNA enhancer of viral polymerase activity. *Journal of virology*, 86(24): 13475–13485, 2012.

[138] Alexander Pflug, Delphine Guilligay, Stefan Reich, and Stephen Cusack. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature*, 2, nov 2014. ISSN 0028-0836. doi: 10.1038/nature14008. URL http://www.nature.com/doifinder/10.1038/nature14008.

[139] Arjun Raj and Alexander van Oudenaarden. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, 135(2):216–226, 2008. ISSN 00928674. doi: 10.1016/j.cell.2008.09.050.

[140] A. Raue, B. Steiert, M. Schelker, C. Kreutz, T. Maiwald, H. Hass, J. Vanlier, C. Tönsing, L. Adlung, R. Engesser, W. Mader, T. Heinemann, J. Hasenauer, M. Schilling, T. Höfer, E. Klipp, F. Theis, U. Klingmüller, B. Schöberl, and J. Timmer. Data2Dynamics: A modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics*, 31(21):3558–3560, 2014. ISSN 14602059. doi: 10.1093/bioinformatics/btv405.

[141] Jan Rehwinkel, Choon Ping Tan, Delphine Goubau, Oliver Schulz, Andreas Pichlmair, Katja Bier, Nicole Robb, Frank Vreede, Wendy Barclay, Ervin Fodor, and Others. RIG-I detects viral genomic RNA during negative-strand RNA virus infection. *Cell*, 140(3):397–408, 2010.

[142] J C Richardson and R K Akkina. NS2 protein of influenza virus is found in purified virus and phosphorylated in infected cells. *Archives of Virology*, 116(1):69–80, 1991. ISSN 1432-8798. doi: 10.1007/BF01319232. URL http://dx.doi.org/10.1007/BF01319232.

[143] M B Ritchey, P Palese, and J L Schulman. Mapping of the influenza virus genome. III. Identification of genes coding for nucleoprotein, membrane protein, and nonstructural protein. *Journal of virology*, 20(1):307–13, 1976. ISSN 0022-538X. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=354990{&}tool=pmcentrez{&}rendertype=abstract.

[144] Nicole C Robb, David Jackson, Frank T Vreede, and Ervin Fodor. Splicing of influenza A virus NS1 mRNA is independent of the viral NS1 protein. *Journal of General Virology*, 91(9):2331–2340, 2010.

[145] Jeremy S Rossman and Robert a Lamb. Influenza virus assembly and budding. *Virology*, 411(2):229–36, mar 2011. ISSN 1096-0341. doi: 10.1016/j.virol.2010.12.003. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3086653{&}tool=pmcentrez{&}rendertype=abstract`.

[146] Jeremy S Rossman, Xianghong Jing, George P Leser, Victoria Balannik, Lawrence H Pinto, and Robert A Lamb. Influenza virus m2 ion channel protein is necessary for filamentous virion formation. *Journal of virology*, 84 (10):5078–5088, 2010.

[147] Jeremy S Rossman, Xianghong Jing, George P Leser, and Robert a Lamb. Influenza virus M2 protein mediates ESCRT-independent membrane scission. *Cell*, 142(6):902–13, sep 2010. ISSN 1097-4172. doi: 10.1016/j.cell.2010.08.029. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3059587{&}tool=pmcentrez{&}rendertype=abstract`.

[148] George T Rudkin and B D Stollar. High resolution detection of DNA–RNA hybrids in situ by indirect immunofluorescence. 1977.

[149] Rob W H Ruigrok, Lesley J Calder, and Stephen A Wharton. Electron microscopy of the influenza virus submembranal structure. *Virology*, 173(1): 311–316, 1989.

[150] Philip J Santangelo, Aaron W Lifland, Paul Curt, Yukio Sasaki, Gary J Bassell, Michael E Lindquist, and James E Crowe. Single molecule–sensitive probes for imaging RNA in live cells. *Nature methods*, 6(5):347–349, 2009.

[151] J J Sanz-Ezquerro, J Fern, T Sierra, T Arag, J Ortega, J Ort, G L Smith, A Nieto, and Others. The PA influenza virus polymerase subunit is a phosphorylated protein. *Journal of general virology*, 79(3):471–478, 1998.

[152] Peter Scheiffele, Anton Rietveld, Kai Simons, and Thomas Wilk. Influenza Viruses Select Ordered Lipid Membrane Influenza Viruses Select Ordered Lipid Domains during Budding from the Plasma Membrane. 274(4):2038–2044, 1999. ISSN 0021-9258. doi: 10.1074/jbc.274.4.2038.

[153] Mohammed Selman, Samar K Dankar, Nicole E Forbes, Jian-Jun Jia, and Earl G Brown. Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. *Emerging Microbes & Infections*, 1(11):e42, 2012. ISSN 2222-1751. doi:

10.1038/emi.2012.38. URL `http://www.nature.com/doifinder/10.1038/emi.2012.38`.

[154] Sydney M Shaffer, Min-Tzu Wu, Marshall J Levesque, and Arjun Raj. Turbo FISH: a method for rapid single molecule RNA FISH. *PloS one*, 8(9): e75120, jan 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0075120. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3774626{&}tool=pmcentrez{&}rendertype=abstract`.

[155] Megan L. Shaw, Kathryn L. Stone, Christopher M. Colangelo, Erol E. Gulcicek, and Peter Palese. Cellular proteins in influenza virus particles. *PLoS Pathogens*, 4(6):1–13, 2008. ISSN 15537366. doi: 10.1371/journal.ppat.1000085.

[156] Y A Shtyrya, L V Mochalova, and N V Bovin. Influenza virus neuraminidase: structure and function. *Acta naturae*, 1(2):26–32, 2009. ISSN 2075-8251. URL `http://www.ncbi.nlm.nih.gov/pubmed/22649600{%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3347517`.

[157] J J Skehel, P M Bayley, E B Brown, S R Martin, M D Waterfield, J M White, I A Wilson, and D C Wiley. Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion. *Proceedings of the National Academy of Sciences*, 79(4):968–972, 1982.

[158] Elke Socher, Dilip V. Jarikote, Andrea Knoll, Lars Röglin, Jens Burmeister, and Oliver Seitz. FIT probes: Peptide nucleic acid probes with a fluorescent base surrogate enable real-time DNA quantification and single nucleotide polymorphism discovery. *Analytical Biochemistry*, 375(2):318–330, 2008. ISSN 00032697. doi: 10.1016/j.ab.2008.01.009.

[159] Jürgen Stech, Olga Stech, Astrid Herwig, Hermann Altmeppen, Jana Hundt, Sandra Gohrbandt, Anne Kreibich, Siegfried Weber, Hans Dieter Klenk, and Thomas C. Mettenleiter. Rapid and reliable universal cloning of influenza A virus genes by target-primed plasmid amplification. *Nucleic Acids Research*, 36(21), 2008. ISSN 03051048. doi: 10.1093/nar/gkn646.

[160] Klamt Steffen, Saez-Rodriguez J., Gilles Ernst D., Steffen Klamt, Julio Saez-Rodriguez, and Ernst D Gilles. Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biology*, 1(1): 1–13, 2007. ISSN 1752-0509. doi: 10.1186/1752-0509-1-2. URL `http://dx.doi.org/10.1186/1752-0509-1-2`.

[161] Yukihiko Sugita, Hiroshi Sagara, Takeshi Noda, and Yoshihiro Kawaoka. Configuration of viral ribonucleoprotein complexes within the influenza A

virion. *Journal of virology*, 87(23):12879–84, dec 2013. ISSN 1098-5514. doi: 10.1128/JVI.02096-13. URL http://www.ncbi.nlm.nih.gov/pubmed/24067952.

[162] Rose Tam, Lindsay S Shopland, Carol V Johnson, J McNeil, and J B Lawrence. Applications of RNA FISH for visualizing gene expression and nuclear architecture. *FISH: A practical approach*, 260:93–117, 2002.

[163] Sandra Tauber, Yvonne Ligertwood, Marlynne Quigg-Nicol, Bernadette M. Dutia, and Richard M. Elliott. Behaviour of influenza A viruses differentially expressing segment 2 gene products in vitro and in vivo. *Journal of General Virology*, 93(4):840–849, 2012. ISSN 00221317. doi: 10.1099/vir.0.039966-0.

[164] Kazuki Tawaratsumida, Van Phan, Eike R Hrincius, Anthony a High, Richard Webby, Vanessa Redecke, and Hans Häcker. Quantitative Proteomic Analysis of the Influenza A Virus Nonstructural Proteins NS1 and NS2 during Natural Cell Infection Identifies PACT as an NS1 Target Protein and Antiviral Host Factor. *Journal of virology*, 88(16):9038–9048, 2014. ISSN 1098-5514. doi: 10.1128/JVI.00830-14.

[165] RM Taylor. Studies on survival of influenza virus between epidemics and antigenic variants of the virus. *American journal of public health and the nation's health*, 39(2):171–8, 1949. ISSN 0002-9572. doi: 10.2105/AJPH.39.2.171. URL http://www.ncbi.nlm.nih.gov/pubmed/18111593{%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1527782.

[166] Suxiang Tong, Xueyong Zhu, Yan Li, Mang Shi, Jing Zhang, Melissa Bourgeois, Hua Yang, Xianfeng Chen, Sergio Recuenco, Jorge Gomez, Li Mei Chen, Adam Johnson, Ying Tao, Cyrille Dreyfus, Wenli Yu, Ryan McBride, Paul J. Carney, Amy T. Gilbert, Jessie Chang, Zhu Guo, Charles T. Davis, James C. Paulson, James Stevens, Charles E. Rupprecht, Edward C. Holmes, Ian A. Wilson, and Ruben O. Donis. New World Bats Harbor Diverse Influenza A Viruses. *PLoS Pathogens*, 9(10), 2013. ISSN 15537366. doi: 10.1371/journal.ppat.1003657.

[167] Tetsuya Toyoda, Djanybek M Adyshev, Makoto Kobayashi, Akira Iwata, and Akira Ishihama. Molecular assembly of the influenza virus RNA polymerase: determination of the subunit-subunit contact sites. *Journal of General Virology*, 77(9):2149–2157, 1996.

[168] Nikolay Tsanov, Aubin Samacoits, Racha Chouaib, Abdel-Meneem Traboulsi, Thierry Gostan, Christian Weber, Christophe Zimmer, Kazem Zibara, Thomas Walter, Marion Peter, Edouard Bertrand, and Florian Mueller.

smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Research*, page gkw784, 2016. doi: 10.1093/NAR/GKW784.

[169] Qingzong Tseng, Irene Wang, Eve Duchemin-Pelletier, Ammar Azioune, Nicolas Carpi, Jie Gao, Odile Filhol, Matthieu Piel, Manuel Théry, and Martial Balland. A new micropatterning method of soft substrates reveals that different tumorigenic signals can promote or reduce cell contraction levels. *Lab on a Chip*, 11(13):2231, 2011. ISSN 1473-0197. doi: 10.1039/c0lc00641f. URL http://xlink.rsc.org/?DOI=c0lc00641f.

[170] Sanjay Tyagi, Fred Russell Kramer, and Others. Molecular beacons: probes that fluoresce upon hybridization. *Nature biotechnology*, 14(3):303–308, 1996.

[171] Jennifer L Umbach, Hui-Ling Yen, Leo L M Poon, and Bryan R Cullen. Influenza A virus expresses high levels of an unusual class of small viral leader RNAs in infected cells. *MBio*, 1(4):e00204—-10, 2010.

[172] M. Van Der Ploeg. Cytochemical nucleic acid research during the twentieth century. *European Journal of Histochemistry*, 44(1):7–42, 2000. ISSN 1121760X.

[173] Sven C D van Ijzendoorn. Recycling endosomes. *Journal of cell science*, 119 (Pt 9):1679–81, may 2006. ISSN 0021-9533. doi: 10.1242/jcs.02948. URL http://www.ncbi.nlm.nih.gov/pubmed/16636069.

[174] J. N. Varghese and P. M. Colman. Three-dimensional structure of the neuraminidase of influenza virus A/Tokyo/3/67 at 2 armstrong resolution. *Journal of Molecular Biology*, 221:473–486, 1991. ISSN 00222836. doi: 10.1016/0022-2836(91)80068-6.

[175] A. V. Vasin, O. A. Temkina, V. V. Egorov, S. A. Klotchenko, M. A. Plotnikova, and O. I. Kiselev. Molecular mechanisms enhancing the proteome of influenza A viruses: An overview of recently discovered proteins. *Virus Research*, 185: 53–63, 2014. ISSN 18727492. doi: 10.1016/j.virusres.2014.03.015. URL http://dx.doi.org/10.1016/j.virusres.2014.03.015.

[176] Dhanasekaran Vijaykrishna, Reshmi Mukerji, and Gavin J. D. Smith. RNA Virus Reassortment: An Evolutionary Mechanism for Host Jumps and Immune Evasion. *PLOS Pathogens*, 11(7):e1004902, 2015. ISSN 1553-7374. doi: 10.1371/journal.ppat.1004902. URL http://dx.plos.org/10.1371/journal.ppat.1004902.

[177] Frank T Vreede and Ervin Fodor. The role of the influenza virus RNA polymerase in host shut-off. *Virulence*, 1(5):436–439, 2010.

[178] Alister C Ward, Laura A Castelli, Anna C Lucantoni, Jacinta F White, Ahmed A Azad, and Ian G Macreadie. Expression and analysis of the NS2 protein of influenza A virus. *Archives of virology*, 140(11):2067–2073, 1995.

[179] Tokiko Watanabe, Shinji Watanabe, Takeshi Noda, Yutaka Fujii, and Yoshihiro Kawaoka. Exploitation of nucleic acid packaging signals to generate a novel influenza virus-based vector stably expressing two foreign genes. *Journal of virology*, 77(19):10575–10583, 2003.

[180] WHO. Influenza (Seasonal) Fact Sheet, 2016. URL `http://www.who.int/mediacentre/factsheets/fs211/en/`.

[181] Paul J. Wichgers Schreur and Jeroen Kortekaas. Single-Molecule FISH Reveals Non-selective Packaging of Rift Valley Fever Virus Genome Segments. *PLOS Pathogens*, 12(8):e1005800, 2016. ISSN 1553-7374. doi: 10.1371/journal.ppat.1005800. URL `http://dx.plos.org/10.1371/journal.ppat.1005800`.

[182] Franziska Ruth Winter. *Multicolour STED nanoscopy with hyperspectral detection*. PhD thesis, 2016. URL `http://archiv.ub.uni-heidelberg.de/volltextserver/20199/{%}5Cnhttp://archiv.ub.uni-heidelberg.de/volltextserver/20199/1/Dissertation{_}FranziskaWinter{_}Nov2015{_}v2.pdf{%}5Cnhttp://archiv.ub.uni-heidelberg.de/voll`.

[183] Helen M. Wise, Edward C. Hutchinson, Brett W. Jagger, Amanda D. Stuart, Zi H. Kang, Nicole Robb, Louis M. Schwartzman, John C. Kash, Ervin Fodor, Andrew E. Firth, Julia R. Gog, Jeffery K. Taubenberger, and Paul Digard. Identification of a Novel Splice Variant Form of the Influenza A Virus M2 Ion Channel with an Antigenically Distinct Ectodomain. *PLoS Pathogens*, 8 (11), 2012. ISSN 15537366. doi: 10.1371/journal.ppat.1002998.

[184] Chung-Yi Wu, King-Song Jeng, and Michael M-C Lai. The SUMOylation of matrix protein M1 modulates the assembly and morphogenesis of influenza A virus. *Journal of virology*, 85(13):6618–6628, 2011.

[185] Kunitoshi Yamanaka, Akira Ishihama, and Kyosuke Nagata. Reconstitution of influenza virus RNA-nucleoprotein complexes structurally resembling native viral ribonucleoprotein cores. *Journal of Biological Chemistry*, 265 (19):11151–11155, 1990. ISSN 00219258.

[186] Zhiping Ye, Teresa Liu, Daniel P Offringa, Jonathan McInnis, and Roland A Levandowski. Association of influenza virus matrix protein with ribonucleoproteins. *Journal of virology*, 73(9):7467–7473, 1999.

[187] Ashley York and Narin Hengrung. Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. *Proceedings of the . . .*, pages 1–8, 2013. doi: 10.1073/pnas.1315068110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1315068110. URL `http://www.pnas.org/content/110/45/E4238.short`.

[188] Akihiko Yoshimura and S Ohnishi. Uncoating of influenza virus in endosomes. *Journal of virology*, 51(2):497–504, 1984.

[189] Puwei Yuan, Mark Bartlam, Zhiyong Lou, Shoudeng Chen, Jie Zhou, Xiaojing He, Zongyang Lv, Ruowen Ge, Xuemei Li, Tao Deng, and Others. Crystal structure of an avian influenza polymerase PAN reveals an endonuclease active site. *Nature*, 458(7240):909–913, 2009.

[190] Dmitriy Zamarin, Adolfo García-Sastre, Xiaoyao Xiao, Rong Wang, and Peter Palese. Influenza virus PB1-F2 protein induces cell death through mitochondrial ANT3 and VDAC1. *PLoS Pathogens*, 1(1):0040–0054, 2005. ISSN 15537366. doi: 10.1371/journal.ppat.0010004.

[191] Zhengjian Zhang, Andrey Revyakin, Jonathan B Grimm, Luke D Lavis, and Robert Tjian. Single-molecule tracking of the transcription cycle by sub-second RNA detection. *eLife*, 3:e01775, jan 2014. ISSN 2050-084X. doi: 10.7554/eLife.01775. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3901038{&}tool=pmcentrez{&}rendertype=abstract`.

7 | Appendix

7.1 Conditional probability modelling

Here we describe a method to investigate whether there are certain MSCs and segments, which bind together noticeably frequent or infrequent.

Firstly, we want to introduce some notations. Let n denote the total number of possible segments and $N := \{1, ..., N\}$. We identify the subsets of N with the MSCs. For $l = \{1, ..., n-1\}$ we assume the $\binom{n}{l}$ MSCs of rank l to be enumerated, such that for fixed $l \in \{1, ..., n-1\}$ and $i \in \left\{1, ..., \binom{n}{l}\right\}$ there is a unique MSC $c^{l,i} \subset N$ of rank l. (For convenience choose $c^{1,i} = \{i\}$ for the MSCs of rank 1.) Now assume we have collected data for M different cells. We denote by $n_m^{l,i}$ the observed number of occurrences of MSC $c^{l,i}$ in the mth cell. So for the mth cell the number

$$p_m^{l,i} := \frac{n_m^{li}}{\sum_{j=1}^{\binom{n}{l}} n_m^{l,i}} \tag{7.1}$$

is the probability that a randomly chosen MSC of rank l equals the MSC $c^{l,i}$.

Secondly, for known probabilites $p_m^{1,i}$ of the segments, i.e. MSCs of rank 1, we will introduce a naïve straight-forward model to estimate all the probabilities $p_m^{l,i}$ with $l \geq 2$. Later we will extend this model and use the naïve model as baseline to evaluate whether or not there are noticeably frequent or infrequent segment-MSC-fusions as mentioned at the beginning of this chapter.Assuming there is no such noticeable fusion and regarding the rules stated in methods section, we can tell that in a cell m for an MSC $c^{l,i} \subset N$ the probability to fuse next with segment $c^{1,j} \in N \backslash c^{l,i}$ would only be dependent on the observed segment probabilities $p_m^{1,i}$ that belong to segments which do not already belong to $c^{l,i}$. That

means in cell m the stated probability equals the conditional probability

$$p_m(c^{1,j} \mid c^{1,j} \notin c^{l,i}) = \frac{p_m^{1,j}}{1 - \sum_{k \in c^{l,i}} p_m^{1,k}}$$

which only depends on the segment probabilities $p_m^{1,i}$. Thus, by summing over all possibilities to create an MSC $c^{l+1,r}$, we get in cell m:

$$p_m^{l+1,r} = \sum_{c^{l+1,r} = c^{l,i} \cup c^{1,j}} p_m^{l,i} \frac{p_m^{1,j}}{1 - \sum_{k \in c^{l,i}} p_m^{1,k}} \tag{7.2}$$

In the following, we drop the above assumption and try to find a model which is able to reasonably quantify the deviation from the naïve approach. The idea is based on the following Ansatz: We model

$$p_m(c^{1,j} \mid c^{1,j} \notin c^{l,j}) \sim k^{l,i,j} \frac{p_m^{1,j}}{1 - \sum_{k \ inc^{l,i}} p_m^{1,k}} \tag{7.3}$$

with $k^{l,i,j} > 0$. Thus $k^{l,i,j} > 1$ indicates increased and $k^{l,i,j} < 1$ indicates decreased probability for fusion of MSC $c^{l,i}$ and segment $c^{1,j}$. Therefore we get

$$p_m^{l+1,r} \sim \sum_{c^{l+1,r} = c^{l,i} \cup c^{1,j}} k^{l,i,j} p_m^{l,i} \frac{p_m^{1,j}}{1 - \sum_{k \in c^{l,i}} p_m^{1,k}} \tag{7.4}$$

and also taking into account that what we want to model is a probability density which must sum to one, after normalization we end up with

$$p_m^{l+1,r} = \sum_{c^{l+1,r} = c^{l,i} \cup c^{1,j}} \hat{k}^{l,i,j} p_m^{l,i} \frac{p_m^{1,j}}{1 - \sum_{k \in c^{l,i}} p_m^{1,k}} \tag{7.5}$$

with

$$\hat{k}^{l,i,j} = \frac{k^{l,i,j}}{\sum_{s=1}^{\binom{n}{l+1}} \sum_{c^{l+1,s} = c^{i,i} \cup c^{i,j}} k^{l,i,j} p_m^{l,i} \frac{p_m^{1,j}}{1 - \sum_{k \in c^{l,i}} p_m^{1,k}}} \tag{7.6}$$

to keep notation short.

Having modeled our data that way, we can try to find values $k^{l,i,j}$ which lead to a better explanation of the observed data. This was implemented as nonlinear least squares problem with box constraints: for $l = 1, ..., n - 2$ and given $p_m^{1,j}, (p_m^{l,j}, (p_m^{l+1,j}$, find optimal $k^{l,i,j} \geq 0$ such that

$$\sum_{m=1}^{M} \sum_{r=1}^{\binom{n}{l+1}} p_m^{l+1,r} - \sum_{c^{l+1,r} = c^{l,i} \cup c^{1,j}}^{\binom{n}{l+1}} \left(\hat{k}_m^{l,i,j} p_m^{l,i} \frac{p_m^{1,j}}{1 - \sum_{k \in c^{l,i}} p_m^{1,k}} \right)^2 \tag{7.7}$$

is minimized.

The next step in our analysis is to investigate whether or not there are certain successions of segments $1, ..., N$ to form MSC N that have increased probability. Our idea is, to assign each possible succession, also called path in the following, a score. The score is simply the product of the $k^{l,i,j}$ which belong to the MSC-segment-fusions of that path. One pitfall to consider is, that for fixed l and i, multiplying $k^{l,i,j}$ (for all $j \notin c^{l,i}$) by a common factor, we end up with the same values for $\hat{k}_m^{l,i,j}$ in our optimization problem. Therefore for each l and i we normalized $k^{l,i,j}$ to sum $n - l$, i.e. $\sum_{j \notin c^{l,i}} k^{l,i,j} = n - l$. By this choice $k^{l,i,j} > 1$ indicates increased frequency and $k^{l,i,j} < 1$ decreased frequency compared to the naïve model. Note that a different choice of the normalization strategy (for instance such that the geometric mean is one) would be possible.

Publications

1. **Prisner, S.**[*], Haralampiev, I.[*], Jolmes, F., Schade, M., Schreiber, M., Loidolt, M., Jongen, K., Chamiolo, J., Winter, F., Seitz, O., Herrmann, A., *Deciphering the assembly of multi-segment genome complexes in influenza A virus by the introduction of MuSeq-FISH.* (2017). *In preparation.*

2. Haralampiev, I.[*], Schade, M.[*], Chamiolo, J., Jolmes, F., **Prisner, S.**, Witkowski, P.T., Behrent, M., Hövelmann, F., Wolff, T., Seitz, O., Herrmann, A., *Influenza A viral genome switching from transcription to replication studied at single host cell level.* (2017). *In preparation.*

3. Jin, S., Kedia, N., Illes-Toth, E., Haralampiev, I., **Prisner, S.**, Herrmann, A., Wanker, E.E., Bieschke, J., *Amyloid-β(1-42) Aggregation Initiates Its Cellular Uptake and Cytotoxicity.* J Biol Chem. (2016). 291(37):19590-606.

---

[*]equal contribution

Acknowledgements

I would like to thank, first of all, Ivan Haralampiev, for the very close and fruitful collaboration on many projects covered in this thesis, for joint supervision of numerous students and for a great time in the lab.

I would also like to thank Andreas Herrmann for being a inspiring and motivating supervisor, who allowed me to work as independently as possible, yet always provided support when needed.

Additionally I would like to thank Kalle Jongen, Nia Abbas, Michael Cheng, Malte Hilsch, Vanessa Körner and Marie Behrendt who joined the project as interns, bachelor or master students and were great people to work with. Partners from collaborations, namely Jasmine Chamiolo, Maria Loidolt, Franziska Winter and Max Schreiber also deserve special thanks!

Thanks also go to Matthias Schade, Fabian Jolmes, Chris T. Höfer and all colleagues from the group of Molecular Biophysics for a very nice working environment in which science is our common interest, but not the only one.

It was wonderful to be part of the German-Israeli research school Sign-Gene, so I also want to thank the coordinators Sabine Löwer and Sandra Krull, all fellow PhD students from the program, and especially my co-supervisor Amit Meller and his wonderful lab that hosted me twice with very warm welcomes. I also have to thank the SignGene program for allowing me to broaden my horizons through frequent visits to Israel, even more frequent scientific conferences and financing courses for my personal development.

And of course a scientific environment is not everything – I want to warmly thank Laura Jakubowitz for bearing with me in this time, even moving to Berlin, when sometimes work called for experiments day and night.

Selbständigkeitserklärung

Hiermit versichere ich, Simon Prisner, dass ich die vorliegende Dissertation selbständig und nur mit den angegebenen Hilfsmitteln verfasst habe. Ich erkläre ausdrücklich, dass ich *sämtliche* in der Arbeit verwendeten fremden Quellen und Grafiken als solche kenntlich gemacht habe.

In Kooperation erhobene Daten und Ergebnisse wurden im Text markiert. Dazu gehören insbesondere die Etablierung und Nutzung von MuSeq-FISH in enger Zusammenarbeit mit Ivan Haralampiev, die Durchführung der qRT-PCR-Experimente durch Kalle Jongen und Vanessa Körner sowie die mathematische Modellierung der Netzwerkdaten von Influenza durch Max Schreiber.

Die Arbeit wurde bei keiner anderen Institution eingereicht. Ich besitze keinen Doktorgrad und habe mich an keiner anderen Stelle um einen Doktorgrad beworben. Die dem Promotionsverfahren zugrunde liegende Promotionsordnung ist mir bekannt.