Investigations into the dynamics of paramyxovirus infections by high-throughput sequencing

Elizabeth Bowie Wignall-Fleming

University of Glasgow

A thesis submitted for the degree of Doctor of Philosophy

College of Medical, Veterinary & Life Sciences

University of Glasgow

July 2018

Abstract

The paramyxovirus family can cause a broad spectrum of diseases from mild febrile illnesses to more severe diseases that may require hospitalisation and can in the most serious cases have fatal outcomes. Understanding the virus infection dynamics is fundamental to the development of novel targets for therapeutic and vaccine development.

The advancement of High-throughput sequencing (HTS) has revolutionised biomedical research providing unparalleled opportunities to answer complex questions. In this study we developed a workflow using directional analysis of HTS data to gain a unique opportunity to simultaneously analyse the kinetics of virus transcription and replication for PIV5 strain W3, PIV2, MuV and PIV3. The workflow could be used for the study of all negative strand viruses.

The developed workflow was used to investigate a number of characteristics of paramyxovirus transcription including quantification of the transcription gradient, RNA editing resulting in the generation of non-templated mRNAs and the production of read-through mRNAs. Interestingly, the processivity of the RNA polymerase during transcription was shown to remain consistent throughout the infection amongst all of the viruses analysed.

Additionally, virus replication and the generation of antigenomes were found to occur at early times post infection. This was surprising, as the current model for virus replication requires sufficient levels of NP to be present in the cytoplasm before the virus can enter replicative mode. These results suggest a revision of

this model in which the virus produces local sites of virus transcription and replication in the cytoplasm known as foci and it is the level of NP surrounding the virus genomes at these local sites that dictates the virus ability to enter a replicative mode.

PIV5 strain W3 was shown to supress virus gene expression at late times post infection resulting in the establishment of a persistent infection. The developed workflow was used to analyse the infection dynamics of PIV5. There were no changes in the RNA polymerase processivity of transcription that could account for the suppression of protein synthesis. A comparative analysis of PIV5 strains W3 and CPI+ identified a mutation of a serine to a phenylalanine at position 157 of the P protein in CPI+, a phosphorylation site that when phosphorylated by polo-like kinase 1 (PLK-1) was previously shown to play a role in the inhibition of virus RNA synthesis, that abolished the virus ability to supress protein synthesis and establish a persistent infection. This indicates that phosphorylation of serine at position 157 is responsible for the inhibition of virus gene expression and the establishment of persistence.

Acknowledgements

First and foremost I would like to thank my supervisors Professor Rick Randall and Professor Andrew Davison for the guidance and support throughout my PhD. It would not have been so enjoyable without the members of the Randall lab, namely Dan, Lena, Andri, Liz and Dave who have offered not only knowledge and support when required but also friendship. A special note of thanks to Dan Young who helped me endlessly with lab work.

I would not have entered into a career in bioinformatics without Derek Gatherer, who offered me an internship during which I gained my first glimpse into bioinformatics and which taught me the foundation of knowledge on which this PhD was built. A note of thanks must also go to Carol Leitch, who has offered me support and encouragement during my MRes, PhD and for my future career.

Thank you to my Mum and Dad who have supported me throughout my university career, from undergraduate to masters and now as a PhD student. I would not have been able to achieve any of this without them. I would also like to thank Rona, who has offered support and advice (laughs and coffee!) over a decade of friendship.

Table of Contents

List of Figures

List of Tables

Abbreviations

PIV	parainfluenza viruses
MuV	mumps virus
MeV	measles virus
NDV	Newcastle disease virus
SeV	sendai virus
VSV	vesicular stomatitis virus
RSV	respiratory syncytial virus
HCV	hepatitis C virus
BTV	Blue Tongue Virus
NP	nucleocapsid
P	phosphoprotein
L	large protein
F	fusion protein
HN	haemagglutinin-neuraminidase protein
SH	small hydrophobic
PLK-1	Polo-like kinase 1
IFN	interferon
NNS	negative-sense, non-segmented
RNP	ribonucleoprotein
UTR	untranslated region
nt	nucleotides
Tr	trailer
Le	leader
TrC	antigenome promoter
Gs	gene start cis-acting elements
Ge	gene end cis-acting elements
Ig	intergenic region
ORF	open reading frame
PAMPS	pathogen-associated molecular patterns
PRR	pattern recognition receptors
ds	double-stranded
ss	single-stranded
PKR	protein kinase R
TLR	Toll-like receptors
MAVS	mitochondrial anti-viral signalling protein
RIG-I	retinoic acid inducible gene-I
mda-5	melanoma differentiation-associated gene 5
TBK-1	TANK-binding kinase-1
IKKε	inhibitor of κB kinase ε
IRFs	interferon regulatory factors
IPS-1	IFN-β promoter stimulator 1
JAK	Janus kinases
ISRE	interferon-stimulated response elements
ISGs	IFN-inducible genes
IFITM	IFN-inducible transmembrane family
DI	defective interfering
SSPE	subacute sclerosing panencephalitis
WT	wild-type

ts	temperature sensitive
HTS	High-throughput sequencing
rRNA	ribosomal RNA
BWT	Burrows-Wheeler Transformation
FM-index	full-text in minute space
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
RPKM	Reads Per Kilobase of transcript per Million mapped reads
MOI	multiplicity of infection
HSF	human skin fibroblast cells
pfu	plaque forming units
h p.i.	hours post infection

Chapter one

Introduction

1.0 Introduction

The family *Paramyxoviridae* consists of a group of enveloped, negative-sense, non-segmented RNA viruses that cause a broad spectrum of diseases in a diverse range of hosts, including humans, other mammals, birds, reptiles and fish. The family contains seven genera, examples of viruses in these genera are shown in Table 1.1. These viruses cause disease ranging from asymptomatic and mild febrile illnesses to more severe respiratory diseases such as bronchitis, croup, pneumonia and mumps. These disease examples are caused by human parainfluenza viruses (PIVs) and mumps virus (MuV), and can, as is the case with numerous other paramyxovirus infections, result in hospitalisations especially in vulnerable groups such as young children, the elderly and the immunocompromised. Some viruses in the family confer high mortality rates, including the recently isolated zoonotic Nipah and Hendra viruses. The family also includes some of the most infectious viruses known, such as measles virus (MeV), as well as economically important viruses such as Newcastle disease virus (NDV), which infects poultry, and, until its recent eradication, rinderpest virus in cattle.

Table 1.1. The genera of the family *Paramyxoviridae*, with virus examples.

Genus	Examples of viruses
Respirovirus	Parainfluenza virus 1 and 3, Sendai virus
Rubulavirus	Parainfluenza virus 2 and 5, mumps virus
Morbillivirus	Measles virus, Rinderpest virus
Henipavirus	Hendra virus, Nipah virus
Aquaparamyxovirus	salmon aquaparamyxovirus
Avulavirus	Avian avulavirus 1-19
Ferlavirus	Reptilian ferlavirus

1.1 Viral proteins

The function of the viral proteins is extensively reviewed in Lamb and Parks (2013) and Chattopadhyay et al (2011).

The nucleocapsid protein

The nucleocapsid (NP) protein exists in two states, either as part of the nucleocapsid structure or as free soluble NP (NP^0). The protein is comprised of two domains, the N-terminal N-core and the C-terminal N-tail (Karron and Collins, 2013; Lamb and Parks, 2013). The N-terminal N-core contains a motif that is conserved among most paramyxoviruses. It is essential for interactions with viral RNA and for encapsidation of the viral genome and antigenome in a highly stable, nuclease-resistant complex known as the nucleocapsid (Parks et al., 1992). The N terminus of NP^0 proteins can interact with the phosphoprotein (P) and V proteins to form a soluble complex that prevents NP^0 from interacting non-specifically with cellular RNA (Curran et al., 1995; Horikami et al., 1992; Precious et al., 1995; Spehner et al., 1997; Zhao and Banerjee, 1995). The

binding of NP and V may also play a role in the inhibition of viral RNA synthesis (Lin et al., 2005). The N-tail has not been shown to be necessary for encapsidation, but has been found to play an essential role in the interaction between encapsidated RNA and the RNA polymerase, which is required for RNA synthesis (Buchholz et al., 1994; Heggeness et al., 1980; Horikami et al., 1992; Mountcastle et al., 1974). The N-tail is intrinsically unstructured but forms a folded structure upon interaction with P (Bourhis et al., 2005). The interaction between the N-tail of NP and the C-terminus of the P protein is important for viral replication, as P forms a complex with the large (L) protein to form the RNA polymerase complex.

The accessory proteins

The accessory proteins are non-templated transcripts that are transcribed from the V/P gene. It is the only paramyxovirus gene capable of generating multiple transcripts through a processes known as RNA editing and ribosomal stuttering, which is described in greater detail below. The proteins generated from the non-templated transcripts are often referred to as accessory proteins, and this term can suggest that they are non-essential. However, viruses lacking them often exhibit attenuated growth (Chattopadhyay et al., 2011). The faithful (unedited) copy of the V/P transcript comes from the V gene in rubulaviruses and from the P gene in respiro-, morbilli-, henipa- and avula-viruses. The additional proteins are transcribed by the addition of non-templated G residues in a tract of G residues to create a transcriptional frameshift. In rubulaviruses, the addition of one or two G residues generates the I (sometimes referred to as W) and P transcripts, respectively. In respiro-, morbilli-, henipa- and avula-viruses, the V

and W/D transcripts are generated by the addition of one or two G residues, respectively. The P protein can also transcribe a C protein through an additional open reading frame by ribosomal stuttering. The functions of the accessory proteins are described below.

The phosphoprotein

The P protein is a major component of the RNA polymerase and is essential for viral RNA synthesis. It interacts with the L and NP proteins to perform critical roles in viral transcription and replication (Karron and Collins 2013; Lamb and Parks, 2013). The N terminus of P interacts with NP^0 to form a soluble complex, and prevents NP from interacting with cellular RNA (Curran et al., 1995; De et al., 2000; Horikami et al., 1992; Zhao and Banerjee, 1995). The P protein is also a non-catalytic subunit of the RNA polymerase. It is widely believed that the P subunit acts as a tether to the genomic template for L, which is the catalytic subunit of the RNA polymerase (Hamaguchi et al., 1983). The P protein becomes heavily phosphorylated by cellular kinases at specific serine and threonine residues, principally in the N terminus. The role of phosphorylation of the P protein during RNA synthesis has obtained some contradictory results, reviewed in Fuentes et al., (2010). Studies using SeV using a minigenome systems have reported that up to 90% reduction in the phosphorylation of the P protein exhibits no negative effect on viral RNA synthesis or virus growth (Byrappa et al., 1996; Hu and Gupta, 2000; Hu et al., 1999). However, several of these phosphorylated residues have been shown to have a significant impact on viral RNA synthesis. For example, PKC-mediated phosphorylation has been shown to play a regulatory role for human PIV3, but

not for Sendai virus (SeV) (De et al., 1995; Hu et al., 1999; Hu and Gupta, 2000). Polo-like kinase 1 (PLK1), which is involved in regulating the cell cycle, has been shown to phosphorylate a serine at position 157 of the PIV5 P protein, resulting in inhibition of RNA synthesis (Sun et al., 2009; Timani et al., 2008).

The V protein

This is a multifunctional protein with essential roles as an IFN antagonist and as a regulator of viral RNA synthesis. Its role as an IFN antagonist is discussed further at (section 1.4). The V and P proteins share a common N terminus that allows the protein to possibly interact with the NP and L proteins; this is possibly the mechanism by which it inhibits viral RNA synthesis, as mutations in V can result in increased viral RNA synthesis (Dalenda et al., 1997; Durbin et al., 1999; He et al., 2002; Horikami et al., 1996; Huang et al., 2000; Kato et al., 1997b; Lin et al., 1997; Nishio et al., 2008;, Precious et al., 1998).

The C protein

A number of paramyxoviruses generate the C protein from an alternative open reading frame present in the V/P gene. Human PIV3 produces only one C protein, whereas human PIV1 and SeV produce four (C', C, Y1 and Y2) by two mechanisms, namely leaky ribosomal scanning and ribosomal shunting (Lamb and Parks, 2013; Latorre et al., 1998). The C protein has a role in the regulation of viral RNA synthesis and also serves as an IFN antagonist (Kurotani et al 1998). It does not seem to be essential for viral transcription and replication, but viruses lacking the C gene have shown decreased virus growth (Chattopadhyay

et al., 2011). It has been hypothesized that the C protein interacts with L, but at a different site from that of P, resulting in decreased RNA synthesis (Horikami et al., 1997; Grogan and Moyer, 2001; Smallwood and Moyer, 2004). It has also been suggested that C can interact with the promoter to regulate RNA synthesis (Cadd et al., 1996; Tapparel et al., 1997).

The D, W and I proteins

These proteins are generated via RNA editing and the insertion of 2 G residues (D generated by the respiroviruses) or 1 G residue (I or W for rubulaviruses) at RNA editing site present in the V (rubulavirus) or P genes (respirovirus). The proteins are produced at relatively low abundances during viral infection, and their functions are largely unknown. A study by Horikami et al., (1996) suggested that the W protein, due to its shared N-terminal with P, could interact with free NP^0 proteins and thereby inhibit viral RNA synthesis.

The matrix protein

Matrix (M) proteins are the major structural components of the virus particle and lie beneath the envelope. They interact with the glycoproteins present on the surface of the envelope and the nucleocapsid complex (see the descriptions below) to form the viral particle. The M protein plays an essential role in viral assembly, budding and morphogenesis (Sanderson et al., 1995; Stricker et al., 1994; Takimoto and Portner, 2004; Terrier et al., 2009).

The fusion protein

The fusion (F) protein plays a role in the fusion of the viral envelope to the host cell membrane, allowing the nucleocapsid to be internalized. Initially, the F protein is translated as an inactive precursor, and is cleaved by host cell proteases to form the F1 and F2 polypeptides linked by a disulphide bond. At late times post infection (p.i.) it is widely believed that the newly synthesized F proteins facilitate cytopathic cell-to-cell fusion, forming syncytia (Chattophraya et al., 2011). These syncytia may play an important role in the cell-to-cell transmission of some paramyxoviruses.

The haemagglutinin-neuraminidase glycoprotein

The haemagglutinin-neuraminidase (HN) proteins protrude from the viral envelope. The HN protein is involved in the attachment of the viral particle to host cell membranes by interacting with sialic acid residues (Lamb and Parks, 2013). Additionally, it is thought to interact with the F protein and assist in fusion to host cell membranes. The HN protein plays an essential role in the egress of viral particles from the infected cell through its neuraminidase activity (Mascona, 2005; Porotto et al., 2005).

The large protein

The L protein is a subunit of the viral RNA polymerase, and specifies an enzyme activity that is crucial for the various stages of viral RNA synthesis, including mRNA capping, methylation and polyadenylation (Banerjee, 1987; Banerjee et al., 1991; Karron and Collins, 2013). L interacts with P to form the

RNA polymerase complex. The L proteins of negative-sense, non-segmented (NNS) RNA viruses virus are believed to possess common conserved regions conferring specific functions. Region III is thought to act as a polymerase module because mutations in this region can result in the complete inhibition of viral RNA synthesis (Sleat and Banerjee, 1993; Malur et al., 2002 a,b). Region V carries a specific conserved histidine-arginine nucleotides at position 1227-1228 respectively known as the HR motif that is required for capping of nascent viral mRNAs (Li et al., 2008; Ogino and Banerjee, 2007, 2010; Ogino et al., 2010). However, the functions of some regions remain elusive, even though mutational analyses of SeV L protein regions I, II and IV have shown detrimental effects on viral transcription or replication (Chandrika et al., 1995; Feller et al., 2000; Smallwood et al., 1999, 2002).

1.2 Virus particle

The viral particle of paramyxoviruses is usually spherical and 150-350 nm in diameter. It contains a lipid bilayer envelope that is derived from the host cell membrane during viral particle egress from the infected cell. A schematic of a typical paramyxovirus particle is shown in Fig 1.1. The major component of the viral particle is the M protein, which acts as the principal structural protein in a layer underlying the envelope. The HN and F proteins protrude from the envelope to facilitate viral entry and egress and viral particle fusion with the host cell membrane, respectively. Some paramyxoviruses generate a small hydrophobic (SH) protein, which has been associated with the envelope but is of unknown function. Contained within the viral envelope is the nucleocapsid,

which is also known as the ribonucleoprotein (RNP). This complex is comprised of the NP in a helical structure in which the viral genome is tightly encapsidated. Each NP subunit is bound to its immediate neighbours by N-terminal (N-ter) and C-terminal (C-ter) arms, independent of binding to the RNA (Ruigrok et el., 2011). Each subunit is also associated with 6 nucleotides of the viral RNA genome. As such typically paramyxoviruses favour a genome length consisting of a multiple of 6 nts for effective replication, commonly known as “the rule of six” (Calain and Roux, 1993). The PIV5 nucleocapsid is thought to be comprised of 14 NP subunits per turn, although this may differ amongst paramyxoviruses as SeV is believed to have only 13 NP subunits per turn (Egelman et al., 1989; Lamb and Parks, 2013). Studies have shown that the structure of the nucleocapsid is critical to a functional promoter to initiate virus replication and transcription (reviewed in Fearns and Plemper, 2017).

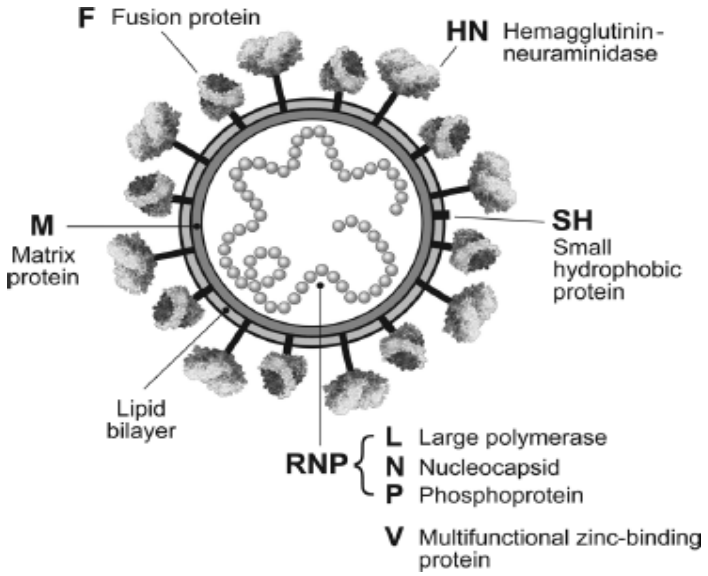


Fig 1.1. A schematic of a parainfluenza virus 5 particle. Nucleocapsid (RNP) is shown inside the virus particle comprised of the negative strand RNA genome encapsidated by NP and associated with L and P proteins. The lipid bilayer and matrix proteins are represented by concentric circles. The HN, F and SH proteins are inserted into the virus membrane. Taken from Lamb and Parks (2013).

The mechanism by which the RNA polymerase recognizes the promoter to initiate RNA synthesis is not fully understood. It is widely accepted that paramyxovirus genomes contain bipartite promoters. The first component is a string of a minimum of 12 conserved nucleotides, known as the internal promoter element (IPE), present at the 3' end of the viral genome. The second component is the presence of tandem repeats in the untranslated region (UTR) of the NP gene between positions 77 and 96, the precise position varies amongst paramyxoviruses. These two promoter elements must be on the correct phase in relation to the encapsidating NP, this is often referred to as hexamer phasing (Cordey and Roux, 2007). A schematic of the two promoter components in relation to the encapsidating NP is provided in Fig 1.2. In rubulaviruses, the second component is a CG at positions five and six of hexamers (NP) 13, 14 and 15 (3'-XXXXCG-5'). In respiroviruses such as SeV and PIV3, it is a C at

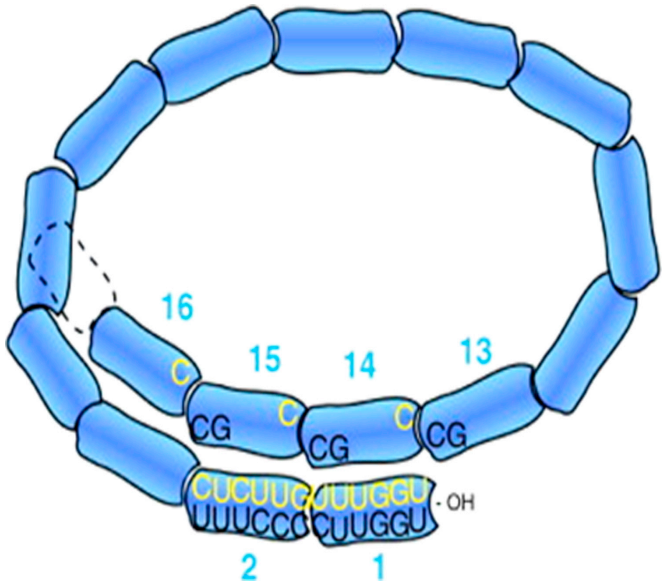


Fig 1.2. A schematic depicting the hexamer phasing of the promoter components for rubula-and respiro-viruses. The NP proteins are shown as blue rectangles, 13 per turn of the nucleocapsid. The internal elements of the promoter, the first is a string of 12 conserved nucleotides at the 3'-end of the virus genome and the second element is repetitive tandem repeats present in the UTR of the NP protein. Rubulavirus second promoter component shown in blue, the respirovirus shown in yellow. Taken from Kolakofsky (2016).

position 1 of hexamers 14, 15 and 16 (3'-CXXXXX-5') (Hoffman and Banerjee, 2000; Murphy and Parks, 1999; Tapparel et al., 1998). The positions of the promoter components on the hexamers place them on the same vertical face as the nucleocapsid helix. Hexamer positions 1, 5 and 6, which seem to be the most important for the second component of the promoter, face away from the RNA-binding groove, whereas positions 2, 3 and 4 face towards the RNA binding grove (Iseni et al., 2002; Gutsche et al., 2015; Alayyoubi et al., 2015). The hexamer phasing of the promoter elements suggests that the RNA polymerase simultaneously interacts with these during the initial stages of promoter recognition for RNA synthesis. The viral genome must then be internalised within the active cavity of the polymerase to initiate RNA synthesis, which would require separation of NP subunits from the viral genome (Maclellan et al., 2007; Tawar et al., 2009). The RNA-dependent RNA polymerase is comprised of one L subunit and four P subunits. The P subunits are non-catalytic and act as a tether for the catalytic L subunit, and P interacts with the C-terminal arm of the NP subunits of the nucleocapsid (Hamaguchi et al., 1983; Horikami et al., 1992). Small amounts of L and P are associated with the nucleocapsid complex present in the virus particle in order to enable viral transcription to occur immediately upon viral entry into the cell. NP subunits that have dissociated from the nucleocapsid in order to allow the viral genome to be internalized into the RNA polymerase cavity are then believed to cross over the surface of the L subunit of the RNA polymerase to reform the helical nucleocapsid structure by associating with the template RNA as it exits the cavity (Liang et al., 2015).

1.3 Paramyxovirus transcription and replication

Paramyxoviruses possess single-stranded, non-segmented, negative-sense RNA genomes. Typically, these are 15000-19000 nucleotides (nt) in length. Different paramyxoviruses possess comparable but not identical cohorts of genes that exhibit analogous functions. A schematic of the genome of prototype virus parainfluenza virus 5 (PIV5) in shown in Fig 1.3. The 3' end of the viral genome contains an extracistronic region of 55-70 nt, which makes up the leader (Le) region and contains the Le promoter elements required for the generation of viral mRNAs and antigenomes. There is also a 5' extracistronic region of 50-161 nt that is known as the trailer (Tr) region that contains the antigenome promoter (TrC) required for generation of antigenomes. Viral mRNAs are transcribed through a stop-start process that is directed by cis-acting elements in the viral genome. Cis-acting elements include the gene start (Gs) and gene end (Ge) sites, which flank the individual genes. The Ge is followed by a U-tract of variable length, which is the site of viral mRNA polyadenylation. Between each gene pair there is an additional cis-acting element known as the intergenic region (Ig) which is composed of a short sequence that is not generally transcribed into mRNA. The Igs vary in sequence and length between the paramyxovirus genera. The respirovirus and morbillivirus contain conserved cis-acting elements, for parainfluenza virus 3 (PIV3) this includes a Gs and Ge sequences of 3'-AGGANNAAAG-5' and 3'-ANTANNA^{5}-5' respectively and a 3'-GAA-5' trinucleotide Ig sequence. Whilst, rubulaviruses including PIV2, MuV and PIV5 possess variable cis-acting elements, both in length and sequence. The Igs can be between 1 and 56 nts in length.

The processes of transcription (generation of viral mRNAs) and replication (generation of antigenomes and newly synthesized viral genome) are common among NNS RNA viruses, and have been reviewed extensively in Banerjee and Barik (1992). As such studies using different NNS RNA viruses, such as the rabies virus (rhabdoviridae), are used to inform the proposed mechanisms by which paramyxoviruses control transcription and replication. A schematic of virus transcription and replication is provided in Fig 1.3. Upon entry into the cell, the virus commences transcription. The small amounts of L and P associated with the nucleocapsid form the RNA polymerase complex, which initiates primary transcription of viral mRNAs. It is believed that viral replication can occur only after sufficient amounts of soluble NP^0 (required for encapsidation of newly synthesized antigenomes and viral genomes) have been produced in the cytoplasm (Curran and Kolakofsky 1999; Howard and Wertz 1989; Patton et al., 1984). The new viral genomes then act as templates for further transcription and replication known as secondary transcription. The precise mechanism by which the virus controls the switch between transcription and replication from a single promoter is yet to be elucidated.

There are two models that have been hypothesized to explain the switch between viral transcription and replication from a single promoter reviewed by Noton and Fearns (2015). The first model suggests that there is a single pool of RNA polymerase that initiates transcription and replication by an identical method at position 1 of the viral genome. The polymerase transcribes a Le transcript, which is then released. After this, the RNA polymerase scans the viral genome further downstream until it finds the Gs signal for the NP gene,

where it initiates and commits to transcription. If a sufficient amount of NP^0 is present in the cytoplasm, the Le transcript would be immediately encapsidated after initiation, which may stabilise the RNA polymerase and commit it to replication (Blumberg et al., 1981; Kolakofsky et al., 2004 Vidal and Kolakofsky, 1989). However, this model is not supported by the observation that increasing the amount of NP^0 did not necessarily increase the level of replication compared to that of transcription (Fearns et al., 1997). Recent studies using a mini-genome system for infections with respiratory syncytial virus (RSV) (family *Orthopneumovirus*) have suggested that the Le promoter contains three distinct regions. The first region, a string of 12 conserved nucleotides, was discussed above as an important element in promoter recognition and is sufficient to initiate transcription and replication, suggesting that it forms the core promoter. Nucleotides at positions 3, 5, 8, 9, 10 and 11 in this element have been shown to be particularly significant, as their mutation led to a reduction of RNA synthesis (Fearns et al., 2002). The second region is comprised of nucleotides 16-34, which are essential for replication but not transcription. The third region corresponds to nucleotides 36-43 at the 5' end of the Le region. It is U-rich and plays a role in transcription efficiency but does not appear to be critical for replication initiation (McGivern et al., 2005). RSV was observed producing short Le transcripts, of approximately 20-25 nts, from position 3 (Tremaglio et al., 2013). Recent studies with VSV have shown that initiation of transcription occurs at position 1 of the genome in the Le and TrC promoters, independent of NP (Morin et al., 2012; Noton et al., 2012). Paramyxoviruses and orthopneumoviruses (RSV) whilst possessing many similarities in the mechanisms of viral RNA synthesis could initiate transcription and replication differently, with paramyxoviruses initiating RNA synthesis from position 1 and

orthomyxoviruses initiating transcription from position 3 of the genome. For paramyxoviruses the RNA polymerase initiates at position 1 of the genome, if insufficient amounts of NP^0 are available encapsidation does not occur and the virus transcribes a short Le transcript, which is then released. The RNA polymerase scans the genome aided by the U-rich region 3 for the downstream NP Gs site to initiate transcription.

Alternatively, the switch between transcription and replication could be controlled by two types of RNA polymerase, transcriptases and replicases. These are composed of the L and P subunits in association with other proteins, and initiate viral transcription and replication, respectively (Banerjee, 2008; Qanungo et al., 2004; Whelan, 2008). Studies of vesicular stomatis virus (VSV) (family rhabdoviridae) have demonstrated the presence of a transcriptase composed of L, P, NP, cellular elongation factor-1-alpha, heat-shock protein 60 and guanylyltransferase (initiating transcription from the N GS site), and a replicase complex composed only L, P and NP (initiating replication from position 1 of the Le region (Keene et al., 1981; Qanungo et al., 2004; Whelan and Wertz, 1999, 2002). However, structural studies have shown that initiation of transcription from an internal site is unlikely because of genome encapsidation by NP (Albertini et al., 2006 a, b; Green et al., 2006).

1.3.1 Virus Transcription

Current evidence suggests that scanning by the RNA polymerase to locate the NP Gs site is important for the initiation of transcription. The RNA polymerase complex then transcribes the gene until it reaches the NP Ge site.

Polyadenylation occurs by stuttering of the polymerase complex over 4-7 U residues following the Ge site. A 5'-capped, 3'-polyadenylated mRNA is then released. The Gs site is thought to contain the signal for the RNA polymerase to carry out capping and cap methylation (Ogino et al., 2005; Stillman and Whitt, 1999; Wang et al., 2007). Studies with RSV show that, if capping is prevented, the RNA polymerase terminates transcription (Liuzzi et al., 2005). Following polyadenylation and release of viral mRNA, the RNA polymerase complex either disengages from the genome or traverses the Ig to reinitiate transcription at the Gs site of the sequential downstream gene. If the RNA polymerase complex disengages from the genome, it can participate in further transcription only by re-attaching to the Le promoter. This transcription mechanism, known as stop-start transcription, produces a transcriptional gradient, with greater quantities of mRNA produced from genes nearer the 3' end of the genome. Thus, the amount of mRNA produced is inversely proportional to the distance of the gene from the 3' end of the genome (Fig 1.3) (Abraham and Banerjee 1976; Ball and White 1976; Collins and Wertz 1983; Huang and Wertz 1982, 1983). The viral mRNAs produced are translated by the host-cell machinery to produce viral proteins.

1.3.2 Read-through mRNAs

During transcription, the RNA polymerase may fail to terminate at the Ge signal. In this circumstance, the RNA polymerase transcribes the Ig and the downstream gene(s), producing a polycistronic mRNA that is also known as a read-through mRNA. The polymerase produces a faithful copy of the genes and the connecting Ig, and there is no evidence of stuttering at the U-tract in

polyadenylation site of the first gene transcript (Gupta and Kingsbury 1985, Wong and Hirano 1986). Polycistronic mRNAs have been found in numerous paramyxo-, rhabdo- and orthopneumoviruses, including RSV, Newcastle disease virus (NDV), measles virus (MeV) and PIV5 (Collins and Wertz, 1983, Hiebert et al., 1985; Paterson et al., 1984; Rozenblatt et al., 1982; Wilde and Morrison, 1984). Much of our knowledge of the generation of read-through mRNAs has been conducted using VSV and RSV, which are rhabdo- and orthopneumoviruses, respectively, and are thought to share common transcription control mechanisms with the paramyxoviruses. Previous studies have suggested that the generation of read-through mRNAs may confer an additional gene expression control mechanism, beyond the generation of the transcription gradient. This mechanism may operate by increasing the access of RNA polymerase to downstream genes or by inhibiting the generation of downstream transcripts of polycistronic mRNAs. Wong and Hirano (1986) showed that generation of P-M transcripts down-regulated the expression of M in MeV. This may indicate that the downstream transcripts are not translated and therefore are a control mechanism for inhibiting the expression of viral proteins. Alternatively, a higher rate of read-through at a specific gene junction may increase RNA polymerase access to downstream genes. In the case of PIV5, the downstream genes included SH, HN and L. Rassa and Parks (1999, 2001) showed that mutations that increased read-through at the M-F gene junction also decreased RNA synthesis. Additionally, as HN has a short half-life on the cell surface, increasing polymerase access to HN gene may be an important factor in maintaining sufficient levels of HN (Leser et al 1996; Rassa and Parks, 1999, 2001). Furthermore, maintaining access to the downstream L gene may be critical to viral replication. Previous studies have shown that an

over- or under-abundance of the L protein can have an inhibitory effect on viral protein synthesis (Finke et al., 2000; He and Lamb 1999; Kato et al., 1999; Parks et al., 2001). It is possible that the control of viral gene expression conferred by the production of polycistronic mRNAs is critical at various stages of the infection. Rassa and Parks (1998) demonstrated a higher generation of read-through transcription early in infection compared to late times post infection. In summary, the generation of read-through mRNAs may have a crucial role in the regulation of viral gene expression, but the exact mechanisms by which the paramyxoviruses maintain this control are unclear.

The generation of polycistronic mRNAs is thought to be directed by cis-acting elements Ge and Ig (Barr et al., 1997 a,b; Kato et al., 1999; Rassa and Park, 1998; Stillman and Whitt, 1997). Previous studies have shown a lack of consensus among viruses in regard to the mechanisms that they use to generate read-through mRNAs. This could be due to the diversity of the cis-acting elements of paramyxoviruses, including the highly conserved elements of the respiroviruses and morbilliviruses, and the high variability of the length and sequence of these elements of the rubulaviruses. Despite this lack of consensus, numerous viruses produce a greater relative abundance of read-through mRNAs at the M-F gene junction. This includes PIV5, PIV1, MeV and PIV3, which synthesize a ~42-80% of M and F transcripts in dicistronic form, indicating that they may have a critical, universal role in efficient virus infection (Bousse et al., 1997; Cattaneo et al., 1987a, b; Rassa and Parks, 1998; Spriggs et al., 1987).

Several studies have demonstrated a crucial role of the Ge sequence in the generation of read-through mRNAs by PIV5, RSV and VSV (Barr et al., 1997b; Hwang et al., 1998; Kuo et al., 1997; Spriggs and Collins, 1986; Rassa and Parks, 1998; Tsurudome et al., 1991). A study by Rassa and Parks (1998) showed that the level of PIV5 M-F read-through mRNAs was decreased to the level of the SH-HN read-through mRNAs by replacing the Ge sequence (10 nt) prior to the U-tract of M with that of the SH Ge sequence. A comparison of the M and SH Ge sequences showed changes in three nucleotides: a G to A at position 5, a U to A at position 8, and a C to U at position 9. Only the change at position 5 was proved to be important in the generation of dicistronic M-F transcripts. Furthermore, a comparison of PIV5 Ge sequences showed that the M Ge is unique and contains a G instead of the collective A at position 5. In the majority of paramyxoviruses the Ge site displays a high level of conservation, suggesting an critical role in paramyxovirus infection (Collins et al.,1996; Kolakofsky et al., 1998).

The polyadenylation site U-tract has also been implicated as a critical factor in determining transcription termination. A minimum length of 7 nt for the U-tract was found to be important in VSV, in which the cis-acting elements are highly conserved, and removing a single U residue decreased the level of transcription termination and thereby generated a higher proportion of read-through mRNAs (Barr et al., 1997b; Hwang et al., 1998). Interestingly, the F Ge of mumps virus (MuV) strain Enders contains a mutation, U to C, which reduces the U-tract from 7 to 4 Us. With this comes a 100% read-through rate and the generation of no detectable SH monocistronic mRNAs (Takeuchi et al., 1991). Combined, these studies suggest that U-tract length is an important factor in polycistronic mRNA

generation for viruses with highly conserved cis-acting elements (VSV), and also those with variable cis-acting elements (MuV). However, a study by Rassa and Parks (1998) showed that, although M in PIV5 (a rubulavirus with variable cis-acting elements) possesses the shortest polyadenylation site, with 4 U residues, increasing this to 6 or 8 Us did not appear to affect the rate of transcript termination (Rassa and Parks, 1998).

The cis-acting element Ig has also been implicated as an essential factor in efficient transcription termination. A study by Bousse et al (2002) using PIV1 and SeV (both respiroviruses with highly conserved cis-acting elements) showed that replacing the SeV M-F Ig with the PIV1 M-F Ig, which produce a low and high proportion of dicistronic mRNAs, respectively, the generation of M-F read-through mRNAs by SeV increased to a level comparable to that observed for PIV1. However, previous studies using SeV and VSV suggested that differences in Igs did not necessarily correlate with changes in transcription termination. Gupta and Kingsbury (1984) showed that the SeV Igs contain a tri-nucleotide sequence of 3'-GAA-5', with the exception of HN-L, which contains 3'-GGG-5'. This mutation did not affect transcription termination at the HN-L Ge site. Studies investigating the importance of Ig in VSV, in which the 3'-GA-5' dinucleotide is present in all Igs with the exception of the NS-M Ig (which contains a G to C mutation), have shown this site to be critical for effective transcription termination (Barr et al., 1997a; Stillman and Whitt, 1997). However, the altered NS-M Ig did not confer an increased proportion of read-through mRNAs (Iverson and Rose, 1981).

A study by Rassa and Parks (1999) suggested that PIV5 cis-acting elements Ig and Ge may work in tandem. A G residue located immediately after the M gene U-tract that contains 4 U residues was found to confer efficient transcription termination for a short U-tract. The rate of transcription termination was comparable to the M gene when the U-tract was increased to contained 6 U residue polyadenylation site and possessed an A residue at position 1 of the Ig. This suggests that the Ig and Ge could function in tandem as a control mechanism for transcription termination. This study also demonstrated, by introducing insertions into the M-F Ig, that the length of Ig may not play a critical role in transcription termination. However, the lengths of insertions were multiples of 6 and thus maintained the hexamer phasing that is essential for effective paramyxovirus RNA synthesis (Rassa and Parks, 1999). Therefore, it is not possible to conclude whether disrupting the position of cis-acting elements in relation to hexamer phasing and the “rule of six” is crucial in transcription termination, as suggested by Kolakofsky et al (1998, 2016). Furthermore, Rassa et al (2000) showed that the spacing between the last residue of the SH gene U-tract and the first residue of the HN gene must be a minimum of 6 nt for effective transcription termination and reinitiation. This suggests that maintaining the spacing between cis-acting elements and their relative positions in relation to hexamer phasing and the “rule of six”, and possibly indicate a crucial role efficient in viral transcription.

The cis-acting elements Ge and Ig and the U-tract polyadenylation site have been implicated in the control of effective transcription termination and the generation of read-through mRNAs. It is hypothesized that the generation of read-through mRNAs plays a critical role in the control of viral gene expression

either by inhibiting the production of monocistronic mRNAs or by increasing access to downstream genes. The lack of consensus over the control mechanisms of transcription termination among the paramyxovirus is possibly not surprising. The differences between the cis-acting elements are diverse, from those of the highly conserved respiroviruses to those of the highly variable rubulaviruses, and it is possible that the individual viruses may exert both common and unique transcription termination controls.

It is not known whether viral proteins are also a factor in transcription termination. A study by Hardy and Wertz (1998) suggested that the M2 ORF1 protein is crucial in increasing the rate of read-through mRNA production by RSV. However, Rassa and Parks (1998) showed that PIV5 proteins N, P and L have no effect on transcription termination, but that other PIV5 viral proteins may be a factor in the generation of polycistronic mRNAs.

1.3.3 RNA editing and the generation of accessory proteins

Paramyxoviruses share the common feature of allowing multiple mRNAs to be transcribed from a single gene through a process known as RNA editing. This occurs in the V (rubularvirus) or P (respiro-, morbilli-, henipa- and avula-viruses) genes, where G residues are inserted at a specific genome position in a proportion of mRNA transcripts, resulting in a transcriptional frameshift. This is thought to occur by slippage of the viral polymerase within a short run of G residues by a similar mechanism as that of polyadenylation (Hausmann et al., 1999a, b; Thomas et al., 1988; Vidal et al., 1990). In rubulaviruses, the V/P gene produces three transcripts: V, which is a faithful copy of the gene; P,

which is generated by insertion of two G residues at position 2338 of the P transcript; and I, which is produced by a single G insertion. As a result, the V, P and I proteins share the same N-terminal sequence but differ in their C-terminal sequences. The proteins produced due to RNA editing differ among viruses in the various paramyxovirus genera. The rubulaviruses, including PIV5, PIV2 and MuV, produce P, V and I transcripts, whereas the respiro-, morbilli-, henipa- and avula-viruses, such as PIV3, generate transcripts for P, V and D through RNA editing. The cis-acting elements that direct RNA editing are conserved and positioned immediately upstream of the RNA editing site. The rubulavirus RNA editing site sequence is 3'-AAAUUCUCCCC-5', whereas in respiroviruses the sequence is 3'-UUUUUUCCC-5'. A study by Thomas et al., (1988) demonstrated that RNA editing, and the insertion of G residues in PIV5, occurs at a 1:1 ratio. Kolakofsky (2016) suggested that conserved hexamer phasing, as is found in the rubula- and morbilliviruses, may control the ratio of 2 and 1 G residue inserts, respectively. However, the lack of conserved hexamer phasing among the respiroviruses, such as SeV and PIV3, results in non-conserved G insertions as SeV inserts 1 G at high frequency, whereas PIV3 inserts between 1 to 6 G residues occur at equal frequency (Hausmann et al., 1999b; Kolakofsky, 2016; Kolakofsky et al., 2005; Iseni et al., 2002).

1.3.4 Viral Replication

Genome replication occurs after initial viral transcription and translation. The RNA polymerase complex attaches to the Le region promoter at the 3' end of the genome, as described above. However, the RNA polymerase complex transcribes the entire length of the genome, ignoring all Gs and Ge sites. This

produces a full-length, faithful, positive-sense copy of the genome known as the antigenome. The antigenome then acts as a template for the production of viral genomes. The 3' end of the antigenome is the complement of the Le region and contains at its end 13 nt that are identical to those at the 3' end of the genome. These are a crucial promoter element for recognition of the RNA polymerase and subsequent RNA synthesis. The newly synthesized genomes and antigenomes are encapsidated by NP^0 to form the nucleocapsid structure. It is thought that simultaneous encapsidation allows the polymerase to ignore Ge signals, thus producing a full-length copy of the template RNA (McGivern et al., 2005; Vidal and Kolakofsky, 1989).

Viral genomes are required at significantly higher levels than antigenomes, as the newly synthesized genomes are assembled into new virus particles. Based on studies by Keller et al (2001, 2003) using mutational analysis and minigenome replicons for PIV5, the TrC promoter on the antigenome is stronger than the Le promoter on the genome and produces a greater number of new viral genomes (Keller et al., 2001; Keller and Parks 2003; Whelan and Wertz 1999). This difference in strength was shown to be the results of 3' cis-acting elements in both promoters (Keller et al., 2001; Keller and Parks, 2003). However, a study by Fearns et al (2000) demonstrated that for RSV the strengths of the Le and TrC promoter are similar.

The replication process of RNA viruses can be erroneous, often producing viral products, such as naked RNA or defective interfering particles (DIs), that can activate the first line in anti-viral defence of the host immune system, the interferon (IFN) response.

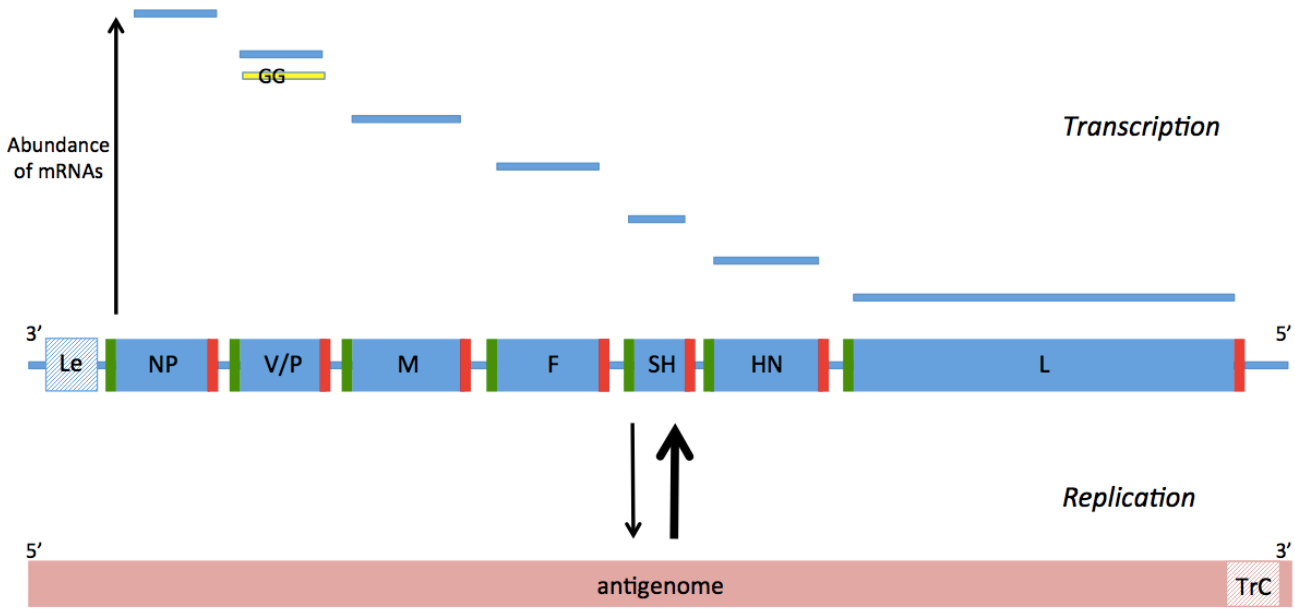


Fig 1.3. A schematic of the PIV5 genome (shown as antisense RNA) indicating the seven genes. The Gs (green) and Ge (red) sites are shown flanking each gene. Genes are represented as blue boxes containing the name of the gene. The replication process and production of the antigenome (shown as sense RNA, red rectangle) are indicated. Transcription is also illustrated, producing 5'-capped, 3'-polyadenylated mRNAs (short blue lines). The transcription gradient displays the abundance of mRNAs produced. The additional P transcript generated via RNA editing and the insertion of two G residues is shown as a yellow bow outlined in blue containing two Gs.

1.4 Evasion of the Host Immune Response

One of the most important determinants of viral pathogenesis is the ability of viruses to circumvent the host immune response. The most effective innate anti-

viral response is the type I IFN system, which can inhibit viral replication and cell-to-cell spread. The type I IFN response can be divided into two phases (induction and signalling), and is extensively reviewed in Stetson and Medzhitov (2006) and Schneider et al., (2014). A schematic overview of the type I IFN induction and signalling pathways is provided in Fig 1.4.

1.4.1 Induction of the Type I IFN Response

The type I IFN response is induced by recognition of pathogen-associated molecular patterns (PAMPS) by host pattern recognition receptors (PRRs). Recognition of PAMPs often occurs through the identification of non-host nucleic acid structures, including viral products generated during (defective) RNA synthesis, such as double-stranded RNA (dsRNA) or uncapped, single-stranded 5' triphosphate RNAs (uncapped ssRNA). Recognition can occur through PRRs such as protein kinase R (PKR) and Toll-like receptors (TLRs), which recognize viral nucleic acid through endocytosis or autophagy. For example, TLR3 and PKR recognize dsRNA, and TLR7 recognizes uncapped ssRNA. Alternatively, viral nucleic acid can be detected in the cytoplasm by RNA helicases such as retinoic acid inducible gene-I (RIG-I) and melanoma differentiation-associated gene 5 (mda-5) (Andrejeva et al., 2004; Yoneyama et al., 2004). The former is believed to associate with short dsRNA (Cui et al., 2004; Hornung et al., 2006; Pichlmair et al., 2006; Takahasi et al., 2008), and the latter interacts with longer dsRNAs (Kato et al., 2008). The activation of RIG-I or mda-5 (or both) leads to the recruitment of mitochondrial anti-viral signalling protein (MAVS) or IFN-β promoter stimulator 1 (IPS-1) (Kawai et al., 2005). The recruitment of these cellular factors leads to the activation of

cytoplasmic kinases such as TANK-binding kinase-1 (TBK-1) and inhibitor of κB kinase ε (IKKε). This then leads to the subsequent phosphorylation and dimerization of interferon regulatory factors (IRFs) and the activation NF-κB, which translocate to the nucleus and activate cellular promoters. This results in the transcription of proinflammatory cytokine genes, including IFN-β (Lamb and Parks, 2013; Randall and Goodbourn, 2008).

1.4.2 Type I IFN Signalling

Upon induction of the type I IFN response, IFN is produced and secreted from the cell. The IFN then binds to its receptor on the cell surface of the infected cell or neighbouring uninfected cells. This results in the phosphorylation of transcription factors STAT1 and STAT2 by Janus kinases (JAK) Tyk2 and JAK1. STAT1 and STAT2 then heterodimerize and associate with IRF-9, forming the transcription factor ISGF3, which translocates to the nucleus and binds to the interferon-stimulated response elements (ISRE) in the promoter region of IFN-inducible genes (Blaszczyk et al., 2016; Horvarth 2000a, b; Raftery and Stevenson, 2017).

1.4.3 Genes induced by IFN

Activation of ISGs results in the expression of numerous anti-viral genes including the production of cytokines to create an anti-viral state (Kotwal et al., 2012). These proteins include the IFN-inducible transmembrane family (IFITM) found principally in late endosomes and lysosomes, most likely affecting virus entry by inhibiting virus fusion with the endosomal membrane (Huang et al.,

2011; Li et al., 2013) and TRIM proteins the functions of which are diverse and range from protein ubiquitination to SUMOylation and ISGylation (Chu and Yang, 2011; Ozato et al., 2008). ISG56, which encodes P56 is the most strongly induced of the ISGs. It interacts with the P48 subunit of the translocation initiation factor (eLF)-3 and specifically targets virus transcription and protein synthesis (Andrejeva et al., 2013; Guo and Sen, 2000).

Other important ISGs responsible for producing an anti-viral state include the multifunctional pro-inflammatory cytokines IL-6 and TNF-α (Cantaert et al., 2010; Ishihara and Hirano, 2002).

1.4.3 Paramyxoviruses and the IFN system

Paramyxoviruses are capable of encoding IFN antagonists that allow them to circumvent host IFN induction and signalling; this has been reviewed extensively by Audsley and Moseley (2013) and Parks and Alexander-Miller (2013). The rubula-, respiro-, morbilli-, henipa- and avulaviruses encode the V protein, which is templated by RNA editing and the insertion of a single G residue. In contrast, the V protein is the faithful copy of the templated gene In the rubulaviruses (He et al., 2002; Komatsu et al., 2004; Poole et al., 2002). An additional protein is generated by respiroviruses and allows the virus to evade the IFN response. The C protein is encoded by alternative read frames in the V/P gene (Komatsu et al., 2004).

Paramyxoviruses produce encapsidated genomes, antigenomes and 5'-capped mRNAs during replication and transcription, and these are not recognised by host cell PRRs. However, numerous studies have shown that paramyxoviruses are strong activators of the IFN system, primarily through the induction of RIG-I

and the production of short dsRNAs (Habjan et al., 2008; Kato et al., 2006; Plumet et al., 2007; Strahle et al., 2007). Paramyxoviruses have also been shown to be inducers of mda-5 (Andrejeva et al., 2004; Berghall et al., 2006; Melchjorsen et al., 2005; Yoneyama et al., 2004, 2005). The precise mechanism by which paramyxoviruses induce type I IFN is yet to be fully elucidated. The activation of PRRs may be due to the generation of some uncharacterised viral products during typical virus transcription and replication. Alternatively, PRRs may be induced by viral products that are generated as a result of defective transcription and replication processes. These products may include defective interfering genomes (DIs), which exist in two types: internal deletion DIs contain the Le and Tr promoter region but lack internal regions of the genome, and copyback DIs are generated when the RNA polymerase switches to the nascent strand during replication, generating a pan-handle dsRNA structure. Copyback DIs have been shown to activate IFN induction in PIV5, SeV and a vaccine strain of MeV (Johnson, 1981; Killip et al., 2013; Poole et al., 2002; Shingai et al., 2007; Strahle et al., 2007).

The V protein of numerous paramyxoviruses, including PIV5, PIV2 and MuV, have been shown to interact with mda-5, thereby inhibiting the type I IFN response (Andrejeva et al., 2004; Childs et al., 2007). The interaction between V protein and mda-5 occurs via the conserved cysteine-rich C terminus of the V protein, which binds to several sites on mda-5 and prevents dsRNA from binding to sites on either the C terminus or the helicase recruitment domain of IRF-3, thereby inhibiting IFNβ production (Childs et al., 2009; Komatsu et al., 2004).

Several studies have suggested that the V protein can block RIG-I induction of type I IFN by acting as a 'decoy substrate', blocking TBK and IKK-ε (rubulaviruses) or IKKα (MeV) and preventing phosphorylation of transcription factors IRF-3 and IRF-7, respectively (Lu et al., 2008; Pfaller and Conzelmann, 2008). A study by Manuse and Parks (2009) suggested that PIV5 was incapable of inhibiting RIG-I induction. However, Childs et al. (2012) demonstrated that RIG-I induction can be inhibited by the V protein interacting with LGP2 (laboratory of genetics and physiology 2), which inhibits induction by RIG-I ligands. Additionally, it has been shown that PIV5 may not be able to block activation of TLR3 and TLR4 (Arimilli et al., 2006).

Studies using mini replicon systems have suggested that the V protein may indirectly inhibit type I IFN induction by down-regulating RNA synthesis (Curran et al.,1991, 1995; Horikami et al., 1997; Lin et al., 2005; Nishio et al., 2008; Parks et al., 2006; Witko et al., 2006). Numerous studies have shown that recombinant viruses with mutations in this protein often display increased viral RNA synthesis (Baron and Barrett, 2000; Delenda et al., 1997; Durbin et al., 1999; Gainey et al., 2008; Kato et al., 1997a, b; Kawano et al., 2001; Schneider et al., 1997; Tober et al., 1998). The control of RNA synthesis may be a mechanism by which the virus decreases the generation of virus products, such as DIs or other, as yet uncharacterised RNA structures, which could induce the IFN response. One proposed mechanism of RNA synthesis control may be the binding of V to NP^0 required for encapsidation of viral genomes and antigenomes (Curran et al., 1995; Precious et al., 1995; Tober et al., 1998). There is some evidence to suggest that PIV5 V protein also binds to Akt1 and inhibits its ability to phosphorylate viral P protein, which has been shown to play

a critical role in the regulation of viral RNA synthesis (Sun et al., 2008; Timani et al., 2008). Furthermore, PIV2 V protein inhibits replication by binding to the L subunit (Nishio et al., 2008).

A number of paramyxoviruses can encode a C protein that acts as an IFN antagonist. Studies using SeV, PIV1 and MeV (which lack the C protein) show an increased induction of type I IFN (Boonyaratanakornkit et al., 2011; Komatsu et al., 2004; Nakatsu et al., 2006; Strahle et al., 2003, 2006). The C protein may function to inhibit RIG-I induction (Strahle et al., 2007). Several studies have suggested that the MeV, SeV and PIV3 C proteins may act as a regulator of the RNA polymerase, thereby limiting the production of dsRNA and indirectly inhibiting IFN induction (Bankamp et al., 2005; Boonyaratanakornkit et al., 2011; Curran et al., 1991, 1992; Cadd et al., 1996; Horikami et al., 1997; Malur et al., 2004; Reutter et al., 2001). Additionally, the C protein may also limit the production of DIs, which have been shown to be potent activators of the IFN induction (Killip et al., 2013; Sanchez-Aparicio et al., 2017). The exact functions of the C protein remain unclear, although PIV1 and PIV3 have been shown to inhibit activation of IRF-3 and the production of IFN-α and IFN-β (Komatsu et al., 2007; Van Cleve et al., 2006).

It has been shown that PIV5 and MuV block IFN signalling by targeting STAT1 for proteosome-mediated degradation, which blocks type I IFN signalling (Didcock et al., 1999a, b; He et al., 2002; Poole et al., 2002; Wansley and Parks 2002; Young et al., 2000). However, PIV2 targets STAT2 for degradation. Targeting STAT for degradation requires the assembly of a cytoplasmic ubiquitin ligase complex. For PIV5 this is composed of the V protein, STAT1,

STAT2, the ultraviolet (UV)-DNA damage repair binding protein DDB1 and a member of the Cullin family of ubiquitin ligase subunits (Lin et al., 1998; Precious et al., 2005, 2007; Ulane et al., 2005).

Alternatively, some paramyxoviruses can inhibit IFN signalling by binding of the V protein to other proteins involved in the phosphorylation or translocation of STAT to the nucleus. The V protein of Hendra and Nipah viruses binds to STAT1, STAT2 and IRF-9, and forms a high molecular weight complex that cannot function, whereas MeV blocks translocation of STATs to the nucleus (Palosari et al., 2003). Interestingly, MeV P protein can also bind STAT1 to prevent phosphorylation (Devaux et al., 2013).

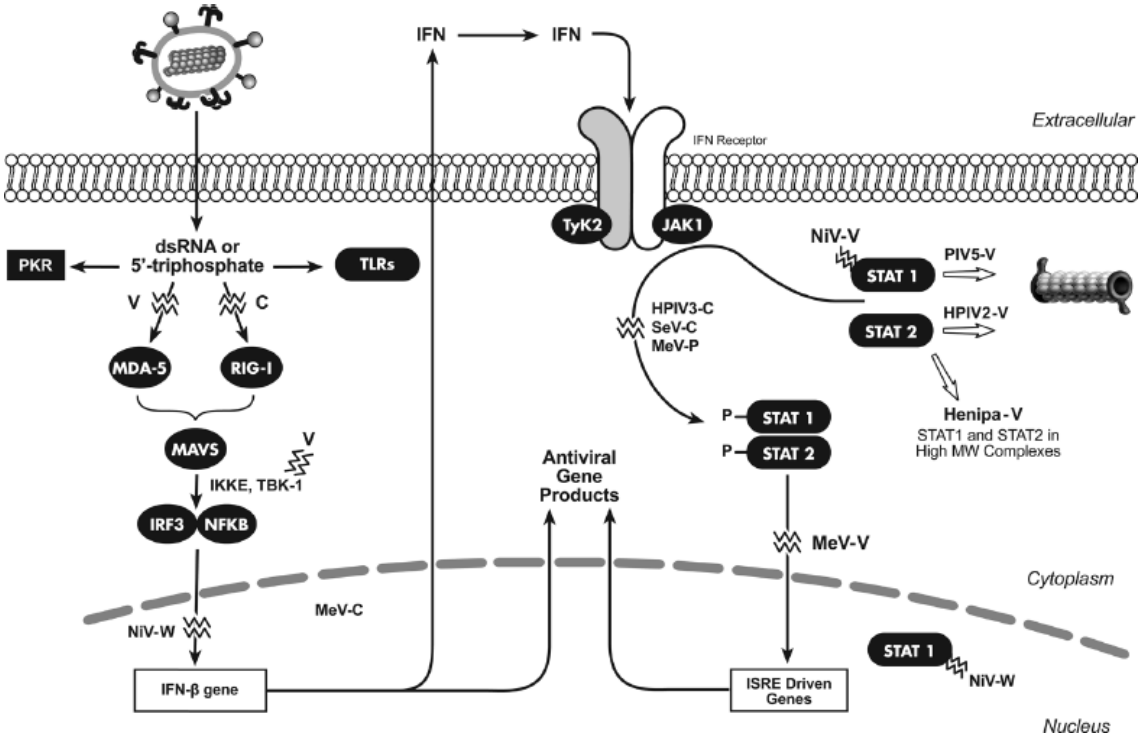


Fig 1.4. A schematic showing the IFN induction and signalling pathways and the paramyxovirus proteins that inhibit the IFN response. Upon entry into the cell the virus produces dsRNA or 5'-triphosphate during virus replication triggering IFN induction through PKR, TLRs, mda-5 and RIG-I. The paramyxovirus V and C proteins inhibit induction of IFN by inhibiting RIG-I and mda-5. Induction of IFN results in a signalling cascade of MAVS and subsequently IRF3 and NFκB resulting in the transcription of IFN-β. The V protein targets TBK-1to inhibit this cascade. Following transcription of IFN-β genes, IFN activates signalling through the JAK-STAT pathway at the cell surface. The paramyxovirus C and V protein target STAT1 and STAT2 for degradation inhibiting stimulation of ISGs.

1.5 Persistent Infections

When a virus infects a host, this typically results in the symptoms of acute disease. During this time, the virus replicates and is shed into the environment. In an acute infection, it would be expected that all virally infected cells would be cleared from the host by the innate and adaptive immune responses approximately within a few weeks of initial infection. An infection that lasts longer is considered to be persistent. This does not necessarily lead to life-long infections like those caused by hepatitis C virus (HCV), but can last a few weeks, months or perhaps years longer than expected; reviewed in Randall and Griffin (2017) and in some cases allow the virus to be transmitted to a new host such as in Ebola (Harrower et al., 2016). Some viruses, including those that generate defective genomes, can spread from cell to cell and thus are capable of persisting in the host without the production of infectious viral particles.

Many paramyxoviruses are capable of establishing a persistent infection in the host. For example, MeV can establish persistent infections in the human brain that can lead to subacute sclerosing panencephalitis (SSPE) many years after the initial infection (Garg, 2002; Kristensson and Norrby, 1986). It has been suggested that persistent MeV infection can lead to life-long immunity for the host, thus conferring a potentially great benefit (Griffin et al., 2012; Lin et al., 2012). The mechanisms by which paramyxoviruses establish persistence remain unknown, but it is known that the virus must accomplish two objectives:

1. The virus must evade the host immune system.
2. The virus must maintain its genome in at least some infected cells while not killing them.

A large abundance of viral genomes may either kill the cell directly by creating a toxic environment including viral RNAs and proteins, or may kill the cell indirectly as a consequence of activating cytotoxic T cells through presentation of viral products by MHC class II molecules. Viral avoidance of killing infected cells may be achieved by low-level replication or cell-to-cell spread (Randall and Griffin, 2017). For viruses that kill cells during infection to establish a persistent infection, the virus must repress virus replication or select viral variants that elicit reduced cytopathogenic activity, such as temperature-sensitive (ts) mutants (Randall and Griffin, 2017).

The generation of DIs has been proposed as a method by which viruses establish persistence and repress virus replication (Johnson, 1981; Roux et al., 1991; Sidhu et al., 1994). DIs contain either TrC or Le and TrC promoter elements, copyback DIs and internal deletion DIs respectively, and therefore can out-compete non-defective genomes and antigenomes for RNA polymerase activity (Roux et al., 1991). Copyback DIs have been shown to induce the host IFN response, thus inhibiting RNA synthesis (Killip et al., 2013). DIs were isolated from brain biopsies of SSPE patients suggesting a causative role in establishing viral persistence (Sidhu et al., 1994). Additionally, studies have shown a correlation between the presence of DIs and the ability of Ebola to establish a persistent infection (Calain et al., 2016; Calain et al., 1999). Alternatively, the virus may be maintained in infected cells with no viral

replication, such as is the case with Blue Tongue Virus (BTV, orbivirus of the reoviridae family) in erythrocytes (MacLachlan et al., 1994; Schwartz-Cornil et al., 2008; Randall and Griffin, 2017; Whetter et al., 1989).

The majority of studies investigating paramyxovirus persistence were conducted using SeV, and the clinical relevance of these infections, if any, has not been demonstrated. Two potentially important factors have been identified that may allow SeV to establish a persistent infection:

1. The presence of DIs.
2. The production of ts mutant viruses.

Studies with SeV have shown that persistent infections can be established by the presence of DIs and/or the selection of ts mutants. Studies by Roux and Holland (1979, 1980) showed that SeV was able to establish persistence through infection with non-defective SeV virus containing DIs. Their presence may confer protection from lytic infection allowing the initial establishment of persistence, with non-defective virus genomes and DIs surviving in a balanced state. If the balance leans towards the non-defective genomes, the cells would enter crisis.

SeV mutants have also often been isolated from persistently infected cells, and the mutations have been found to be ts, enabling the virus to grow well at 33°C but not at 38°C. These mutants were able to establish a persistent infection (Kimura et al., 1975; Nagata et al.,1972). During persistent infection, ts mutants are produced that interfere with the replication of wild-type (WT) non-defective

genomes, resulting in their replacement (Kiyotani et al., 1990; Nagai and Yoshida, 1984; Yoshida et al., 1982). Several mutations have been identified in the ts mutants, in the M, HN and L genes, which are thought to be essential for the establishment of persistence. Additionally, two mutations in the Le were found to confer increased interference capacity (Inoue et al., 2003; Kondo et al., 1993; Nishio et al., 2004; Shimazu et al., 2008). Yoshida et al (1982) observed that cells persistently infected with a stock of SeV that had undergone numerous serial passages in cell culture could enter crisis and cause a high proportion of cell death. The presence of DIs and ts mutants in these persistently infected cell lines was established. Furthermore, the ts mutants were able to establish a persistent infection in the absence of DIs and did not provoke a crisis state.

1.6 High-throughput sequencing

High-throughput sequencing (HTS) is revolutionising biomolecular research, including research on viruses. The advancement of rapid and cost-effective sequencing has allowed the assembly of millions of viral gene sequences and is transforming investigations into the dynamics of viral infections, including areas such as transcriptomics and genomic research (Radford et al., 2012). Advances in HTS technologies have led to the generation of large and complex data sets, with a single sequencing run able to produce millions of short reads. Longer read technologies are available. However, they remain expensive and therefore are not yet widely applied in biomedical research. This study employs short-read sequencing using the Illumina platform, and represents a unique

opportunity to achieve a comprehensive overview of the infection dynamics of paramyxoviruses. There are three steps involved in HTS: the generation of template molecules for sequencing (a process known as library preparation), sequencing and bioinformatic analysis.

1.6.1 Library preparation

The library preparation workflow is illustrated in Figure 1.5. The ability to produce high quality libraries is a crucial step in the generation of HTS data. This process involves the generation of template molecules that are subsequently subjected to sequencing. The numerous library preparation methods reflect the wide range of areas to which HTS methods can be employed, including whole transcriptome sequencing, targeted RNA sequencing and whole genome sequencing. This project employed RNA sequencing using Illumina library preparation kits; namely, total RNA sequencing with nuclear and mitochondrial ribosomal RNA (rRNA) reduction (TruSeq stranded total RNA with Ribo-Zero gold) and targeted mRNA sequencing by polyA selection (TruSeq stranded mRNA library prep kit). These library preparation methods differ in their RNA purification steps (Fig 1.5B) due to their respective target RNA, however subsequent steps are identical.

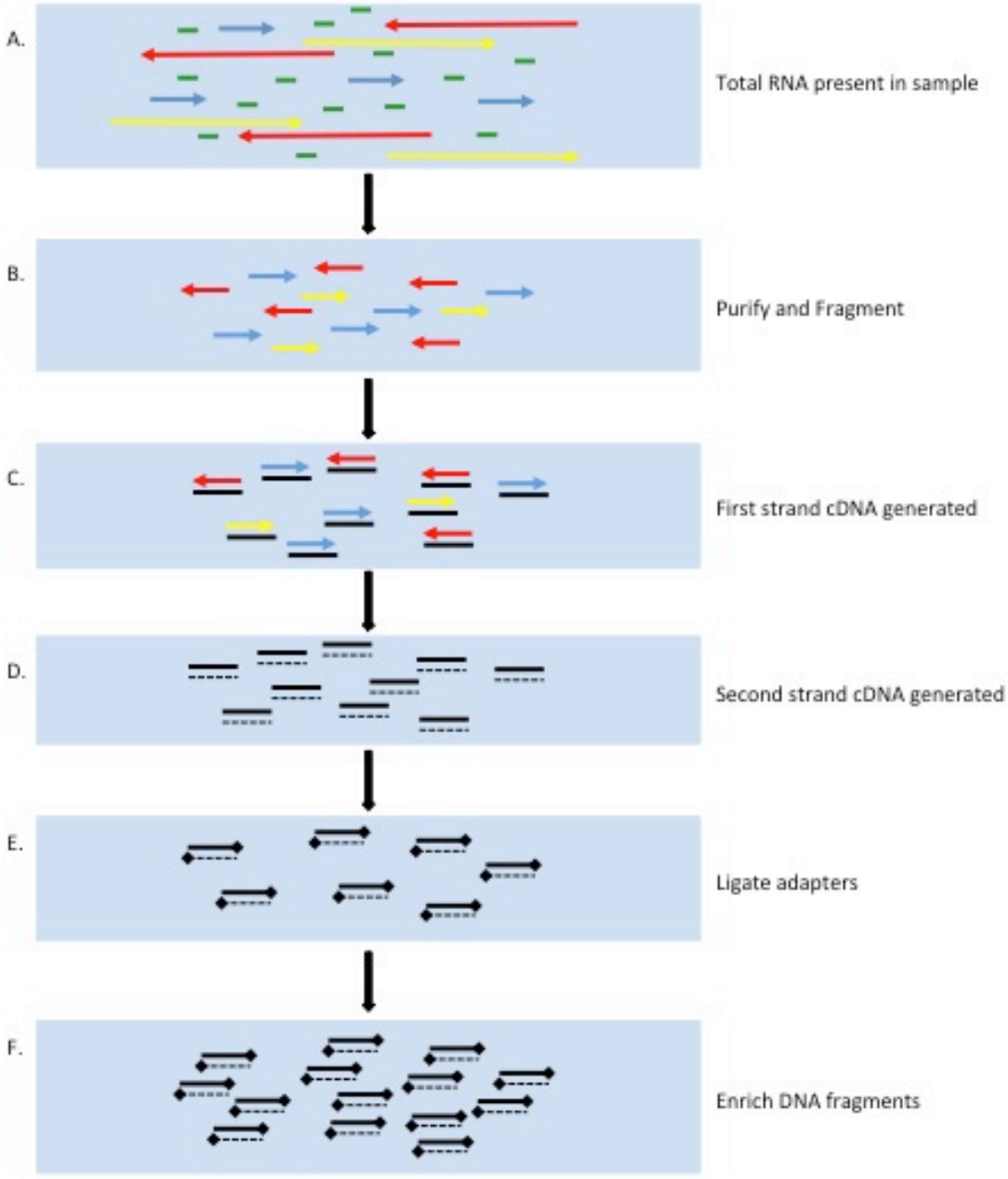


Fig 1.5 A schematic of the workflow of library preparation for HTS. (A) The total RNA present in the sample including host and viral RNAs, different species of RNA are represented as different coloured arrows. (B) RNA is purified either by selecting for target RNA such as in polyA selection or depleting unwanted RNA as in total RNA with ribosomal RNA reduction. The purified RNA is then fragmented. (C) The first strand cDNA is transcribed from the fragmented RNA. (D) The second strand cDNA is transcribed from the first strand cDNA incorporating dUTPs in the place of dTTP thereby retaining strand specificity. (E) Adapters are ligated to the ends of the ds cDNA. (F) The ds cDNA fragments are enriched by PCR amplification.

1.6.1.1 Purification and fragmentation of extracted RNA

Purification of extracted RNA differs dependent on the method of library preparation and the target molecule. This project used both reduction and selection methods to purify the target RNA (Fig 5B). The TruSeq stranded total RNA with Ribo-Zero gold kit depletes nuclear and mitochondrial rRNA. Over 90% of the total RNA in the sample may be composed of these molecules, and removing them allows the remaining RNA (including host mRNA and viral genome and antigenome RNAs and mRNAs) to be enriched. The TruSeq Stranded mRNA Library Prep Kit selects host and viral poly A mRNA using oligo d(T) beads. With both kits, the remaining RNA is fragmented after the RNA purification stage using divalent cations under elevated temperature, and primed for cDNA synthesis.

The fragmented RNA is primed with hexameric oligonucleotides with random sequences and reverse transcribed using reverse transcriptase into first strand cDNA (Fig 5C). The second strand cDNA is generated using the first strand cDNA as a template. Strand specificity is retained during second strand cDNA synthesis by incorporating dUTP in place of dTTP (Fig 5D). This process generates blunt-ended, double-stranded cDNA. The initial template RNA is removed during this procedure.

A single A residue is then incorporated at the 3' ends of the blunt-ended cDNA to prevent the fragments from ligating to one another during subsequent steps. After this, indexed, double-stranded oligonucleotide adapters are ligated to the ends of the cDNA fragments, a T residue at the 3' end of the adapter providing

complementarity to the added A residue incorporated at the 3' ends of the blunt ended cDNA (Fig 5E). This process prepares the cDNA for hybridization to the flow cell used on the Illumina sequencing platform, as discussed in greater detail below.

1.6.1.2 DNA fragment enrichment

The adapter-ligated, double stranded (ds) cDNA is then amplified by PCR using oligonucleotide primers that enrich only the molecules that have adapters at both ends (Fig 5F) by annealing to the ends of the adapters. The number of cycles of PCR amplification is minimized to 15 in order to avoid skewing the representation of sequencing in the library. PCR biases can result in regions with extreme nucleotide compositions being underrepresented in the library, which results in a reduction of read coverage in these regions (Aird et al., 2011; Baskaran et al., 1996; Benita et al., 2003; Dohm et al., 2008; Kozarewa et al., 2009; Oyola et al., 2012).

The library then undergoes quality control and quantification. This is essential for ensuring optimal cluster densities on the flow cell, which are needed in order to generate high quality reads.

1.6.2 DNA Sequencing

There are two approaches to short-read HTS: sequencing by synthesis and sequencing by ligation. This project utilized Illumina sequencing platforms,

which are in the first category. An overview is provided in Figure 1.6 and a video by Illumina can be viewed at https://m.youtube.com/watch?v=fCd6B5HRaZ8. Illumina sequencing is reviewed extensively in Radford et al., (2012).

Illumina technology use a flow cell for carrying out sequencing. The flow cell has multiple channels, which are coated in a layer of oligonucleotide primers that are complementary to the adapters ligated to the ds cDNA fragments in the library. The fragments are first denatured, producing single-stranded (ss) cDNA that is passed through the channels of the flow cell. The ss cDNA anneals to the primers attached to the surface of the channel in the flow cell. This process is repeated until there are millions of cDNA molecules attached to the surface of the flow cell. Thus, the annealed cDNA molecules are immobilized, creating a stable environment for subsequent steps.

Unlabelled nucleotide triphosphates and oligonucleotide primers are added to the flow cell to initiate solid-phase bridge amplification, which can create up to 1000 copies of each individual input cDNA molecule. To determine the sequence of the cDNA fragments, four labelled reversible terminators corresponding to a specific nucleotide (A, T, C and G), primer(s) and DNA polymerase are added to the flow cell. Either single-end or paired-end reads can be generated by using a single or two primers, respectively. A single primer (the forward or F primer) will transcribe the cDNA molecule from the 5’ end, producing single-end reads. Paired-end reads require the use of two primers (the F and the reverse or R primers), which sequentially copy the cDNA molecules from both the 5’ and 3’ ends, respectively. The labelled terminators attach at the relevant cycles and

emit a fluorescent signal unique to each nucleotide after laser excitation. The fluorescence from each cDNA cluster is captured by a CCD camera each time a labelled terminator attaches to a downstream nucleotide. This process is repeated until the sequence of the whole strand of every cDNA molecule has been determined. The number of cycles determines the length of the reads. The data captured from the fluorescent imaging are processed into text files in fastq format, which contain a quality value for each nucleotide. To generate single-end reads only a using a single primer (F) is used producing a single fastq file (R1). Two cycles of sequencing are required for the generation of paired-end reads. The first cycle uses the F primer to produce the R1 fastq file, and then the second cycle uses the R primer to produce the R2 fastq file. The R1 file produces reads that are antisense to the RNA from which they originated, the R2 files produces reads in the same sense as the transcript. The analysis conducted during this study utilized reads that were either 75 or 150 nts in length.

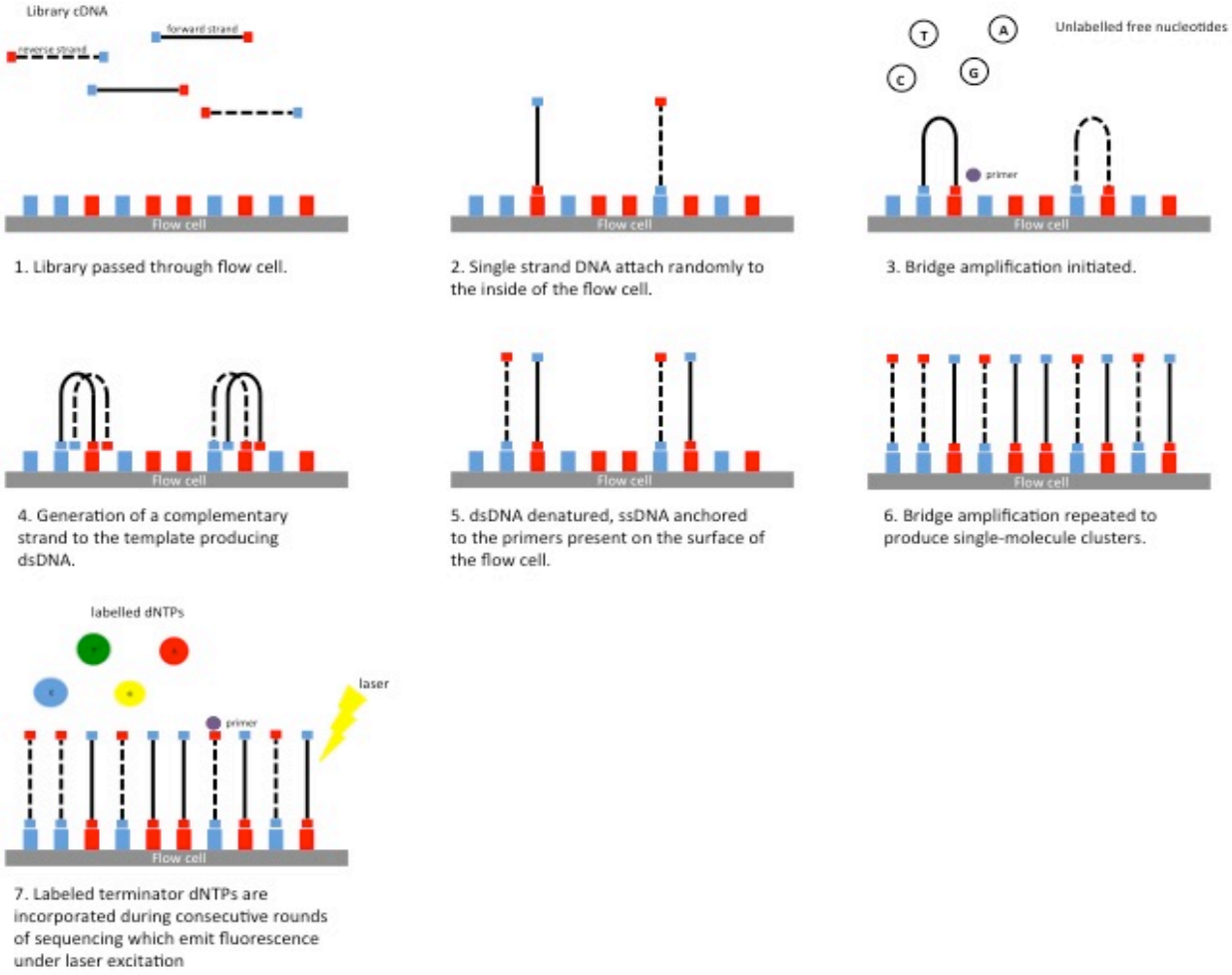


Fig 1.6 Overview of Sequencing by Synthesis used by Illumina sequencing platforms. (1-2) ds cDNA produced during library preparation is denatured and passed through the flow cell attaching to primers complementary to the adapters ligated during library preparation that coat the flow cell surface. (3-6) Unlabelled nucleotide triphosphates and oligonucleotide primers are added to the flow cell to copy the immobilised cDNA fragments. (7) Four labelled reversible terminators corresponding to a specific nucleotide (A, T, C and G) are added to the flow cell, attaching to the nucleotides of the cDNA fragments and emitting fluorescence when excited by a laser that is captured by a CCD camera allowing the sequence of nucleotides of the fragment cDNA to be identified.

1.6.3.1 Read Quality Control

Typically, the first step in HTS analysis is quality control, which is essential for accurate downstream analysis. The millions of short reads that are generated by HTS typically vary in quality since they may contain sequencing errors and artefacts. Each single nucleotide of a read is allocated a quality score, as well as the read as a whole. Nucleotides or reads with low quality scores are removed by trimming and filtering processes. Quality usually decreases over the length of a read, with the 3' end containing a higher proportion of low quality bases. These bases, along with the adapter sequences attached to the ends of cDNA fragments during library preparation, are trimmed from the reads. The reads also undergo filtering to remove those with a low average quality score, reads which are short in length or any containing uncalled (ambiguous) nucleotides.

1.6.3.2 Mapping

Accurate mapping of reads to a reference genome sequence is a crucial step in many bioinformatic pipelines. The use of short-read mapping tools has been reviewed extensively by Hatem et al (2013). The aligning of a read to its correct position in a genome sequence indicates the origin of the read. The majority of mapping programs use hash-based tools or the Burrows-Wheeler Transformation (BWT) to map reads accurately to the reference. Hash-based programmes use have high computational demands but are more sensitive using them to align diverse reads to distantly related references. In contrast, BWT-based programs are rapid and use less computational resource. Since the

paramyxoviruses analysed throughout this project have closely related reference genomes, a BWT-based program was used (Li and Durbin, 2009).

The most commonly used BWT-based alignment software packages are BWA and Bowtie, which were adapted from the Ferragina and Manzini matching algorithm to incorporate an full-text in minute space (FM-index) (Li and Durbin, 2009; Langmead et al., 2009). This transforms the reference genome sequence into an FM-index, which compresses the reference genome sequence and allows the rapid identification of patterns of nucleotides that match those found in the reads. Other alignment programs also utilise the FM-index of the reads, but this has been shown to be less effective than indexing the reference sequence. BWA and Bowtie are the alignment programs that were employed during this project. They differ in their method of mapping. BWA aligns a read by counting the number of matches or mismatches between the read and the corresponding genome position. This is conducted by employing the Bayesian posterior probability to identify the genome position that is most likely to be the correct match. BWA included an additional algorithm that accommodates the mapping of a read with inexact matches to the reference genome position, thus allowing reads containing SNPs, deletions or repeats to be aligned. Bowtie and its successor Bowtie2, which can process reads longer than 50 nt, use a quality threshold (i.e. alignment score) to identify the correct genome position of the read. In general, BWA is the more stringent of the two programs, and therefore was employed throughout this project in cases where the reference genome was closely related to the sequenced sample. However, in cases where several SNPs were present, or in which duplications or deletions were anticipated, Bowtie2 was employed. Accurate mapping is essential in

downstream bioinformatic pipelines, including those involving transcriptomics, which was a principle area of investigation in the analyses described in this report.

1.6.4 Transcriptomics

The ability to quantify a complete set of transcripts promised to provide powerful insights into paramyxovirus gene expression. Gene expression analyses have been conducted previously using microarrays or quantitative PCR. However, HTS provides an alternative method for investigating viral transcription that is known as transcriptomics or RNA-seq analysis (Zhang et al., 2016; Wilhelm and Landry, 2009). This newer technology provides increased specificity and sensitivity and the ability to detect novel transcripts, gene fusions and gene variants, which previous technologies could not do (Zhao et al., 2014). However, to identify transcription products, the reads must be mapped to a reference sequence, and those that span splice junctions will fail to map (Trapnell et al., 2009). Bioinformatic tools are available for detecting reads that map across exon junctions, including alignment software such as TopHat. Such programs are less stringent in mapping reads to a reference genome, compared to BWT alignment tools, and thus allow the mapping of reads that can be matched to two genome positions. However, the reduction in stringency is a potential disadvantage to the accuracy of such analyses. As paramyxoviruses are not believed to use splicing to generate mRNAs, this analysis is not required for transcriptome analysis.

RNA-seq data can also be used to estimate the overall abundance of viral transcripts in order to investigate the kinetics of viral transcription. The data can also be employed to estimate the abundance of the individual viral mRNAs, which effectively allows the quantification of the paramyxovirus transcription gradient and provide an indication of the features of gene expression (Anders and Huber, 2010; Robinson et al., 2010). To analyse the abundance of viral mRNAs, the read counts need to be normalized in order to obtain accurate expression estimates across the genome. There are two features that must be normalized to allow comparisons between genes: gene size and the number of reads mapping to the gene. Firstly, longer genes will generate more reads than shorter genes. The transcripts that are generated are fragmented to a similar size during library preparation, and thus longer transcripts will produce more fragments than shorter transcripts. The fragments are then transcribed into cDNA from which the read data are generated, resulting in a greater number of reads produced by longer genes. Secondly, the cluster density of each sample causes fluctuations in the number of reads generated between samples. These are normalized using values known as FPKM or RPKM (Fragments or Reads Per Kilobase of transcript per Million mapped reads) (Mortazavi et al., 2008). The term mappable read or fragment can be used interchangeably, but the program used to obtained the values throughout the study used fragments, as FPKM. These values represent the relative mRNA abundances. Transcriptomic analysis is accurate when the mRNAs have relatively equal representation, which is usually approached by random priming or RNA ligation during library preparation. The library preparation methods utilized during this study used random priming. However, random priming does result in some coverage

fluctuation over the length of the genome, and this often due to variations in genome composition (Khrameeva and Gelfrand, 2012; Tilak et al., 2018).

1.6.5 Detection of defective interfering genomes

Total RNA sequencing allows the analysis of all aspects of the transcription and replication of negative strand viruses simultaneously, as previously discussed. These methods can also be used to analyse other viral RNA products, such as DIs. Some reads generated from a DI would contain a discontinuity or breakpoint corresponding to the location where the RNA polymerase jumped along the template (internal deletion DI) or switched to the nascent strand (copyback DI) (Fig 1.7). The use of software developed to identify reads that contain a breakpoint causing two regions of the read to map to two different genome positions. In this study, the ViReMa (Virus Recombination Mapper) program was used to identify internal deletion and copyback DIs.

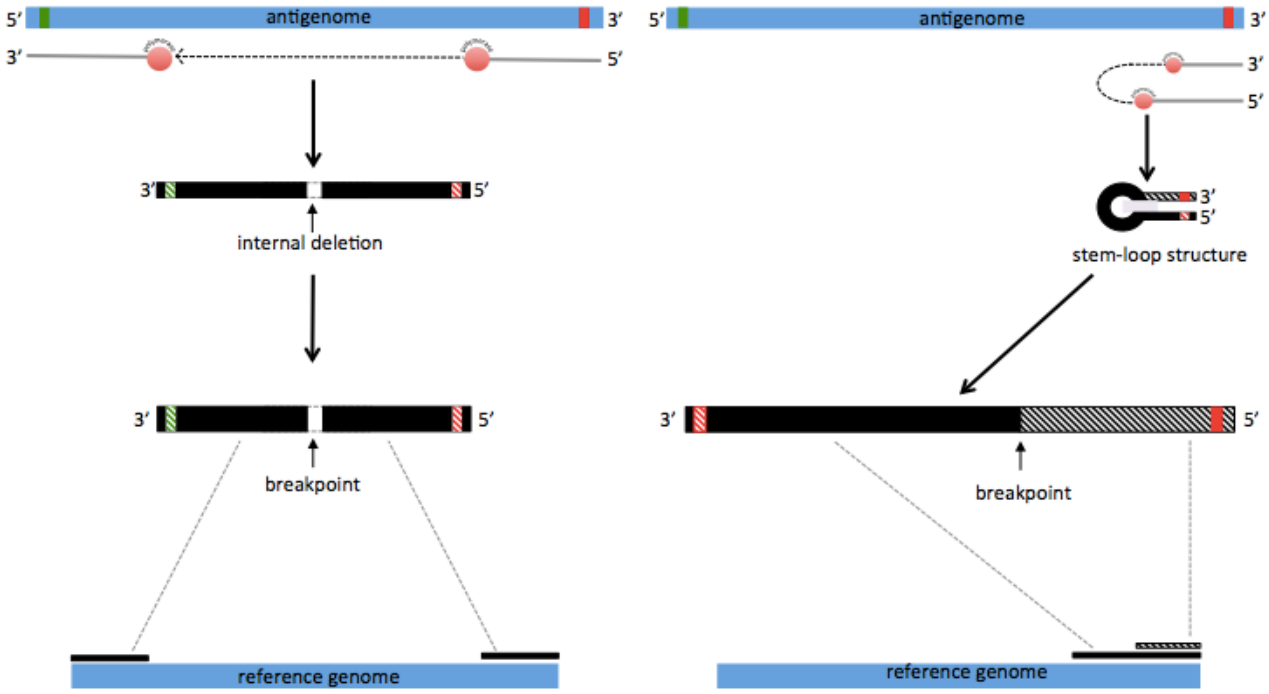


Fig 1.7. A schematic of the generation of internal deletion and copyback DIs, indicating the genome positions to which the components map on the virus genome. The blue boxes represent the antigenome and the reference genome as indicated. RNA polymerase molecules are represented by the red circles, the solid grey lines indicate newly synthesized RNA genomes, and the dotted lines indicate where replication ceased and to which genome position the RNA polymerase jumped. The polymerase jumps along the antigenome in the generation of internal deletion DIs, these therefore contain the 3' and 5' end of the virus genome. During the generation of copyback DIs, the polymerase leaves the antigenome template and attaches to the nascent strand producing a breakpoint, with reads on either side mapping to different regions of the antigenome/genome. The solid black (negative sense) and hatched boxes (positive sense) depict the two regions of the read generated from the DIs and where on the reference genome the reads would map.

1.7 Thesis Objectives

The principal objective of this thesis is the analysis of the dynamics of paramyxovirus infection using HTS. This study developed a pipeline using directional analysis of HTS data to simultaneously analyse virus transcription and replication. This pipeline offers a unique opportunity to investigate the overall infection dynamics of paramyxoviruses and could be applied to the study of other negative sense RNA viruses.

Additionally, the pipeline was employed to analyse a number of characteristics of paramyxovirus transcription including the generation of a transcription gradient, RNA editing and the production of read-through mRNAs showing the versatility and scope of the developed pipeline in transcriptome analysis.

These developed methods of virus transcription and replication analysis were employed to investigate the kinetics of acute and persistent infections as well as possible mechanisms by which the paramyxoviruses can establish a persistent infection.

Chapter two

Materials and Methods

2.1 Infection and treatment of cells

Cells were typically maintained in $25cm^2$ and $75cm^2$ tissue culture flasks (Greiner, U.K.) in DMEM (Dulbecco's modified Eagles's medium; Invitrogen, U.K.) supplemented with 10% (v/v) heat-inactivated FBS (foetal bovine serum, Biowest) and incubated at 37°C/5% CO_2. Either A549 cells (ATCC CRL-1739) or human skin fibroblast (HSF) cells were used during this study. Cells were routinely trypsinised (Trypsin, EDTA; Becton Dickson Ltd., U.K.), and diluted every 4-7 days as appropriate. To stimulate IFN-responsive promoters cells were treated with media supplemented with 10^3 units/ml human IFNα (PBL Biomedical Labs) and incubated overnight. Monolayers of cells were washed with PBS prior to infection to remove traces of serum. Monolayers were infected with virus diluted in DMEM at an appropriate multiplicity (MOI). Virus stocks, for PIV5 strain W3, PIV5 strain CPI+, PIV2, MuV Enders strain and PIV3, were kindly maintained and provided as required by Mr. Dan Young (University of St Andrews, U.K.). The infected monolayers were placed on a rocker for 1-2 hr to allow adsorption of the virus after which the virus inoculum was removed and replaced with DMEM supplemented with 10% (v/v) heat-inactivated FBS and incubated at 37°C/5% CO_2 until harvested.

2.2 Extraction of RNA

Cells were harvested by scraping. The cells were pelleted by centrifugation at 4700 rpm for 5 mins. The supernatant was removed and discarded. The pellet was resuspended in 1 ml of Trizol. An equal volume of ethanol was added. The

RNA was extracted using the Direct-zol RNA miniprep kit (Zymo) as per manufacturers instruction. The sample was directly added to the Zymo-Spin IC column. The sample undergoes several cycles of centrifugation at 13000rpm for 30 secs and washing and eluted in RNase free water.

2.3 Library preparation and high-throughput sequencing

Purification and Selection of RNA

Two methods of library preparation were employed during this study, polyA selection and total RNA with nuclear and mitochondrial rRNA reduction. These methods have different purification/selection of RNA steps described below. Subsequent library preparation steps are identical.

Poly-A selection

The extracted RNA was directionally sequenced using TruSeq Stranded mRNA library Preparation Kit LS protocol (Illumina, U.K.) as per manufacturers instructions. Total RNA was diluted with nuclease-free water to 4µg in a final volume of 50µl. PolyA RNA was selected, adding 50µl oligo-dT magnetic beads to the total RNA and incubated at 65°C for 5 minutes to denature the RNA and 5 minutes at room temperature to allow binding of the polyA RNA to the beads. The sample was then placed on a magnetic stand for 5 minutes selecting the polyA RNA bound beads. The supernatant was discarded. The beads were washed, adding 200µl of Bead Washing buffer. The sample was placed on a magnetic stand for 5 minutes after which the supernatant containing unbound

RNA was discarded (non mRNA). The polyA RNA bound beads were resuspended in 50µl of Elution buffer and incubated at 80°C for 2 minutes to elute mRNA from the beads. The mRNA was rebound to the beads, adding 50µl of Bead Binding Buffer and incubating at room temperature for 5 minutes, to increase specificity and reduce the amount of rRNA that non-specifically binds. The sample was placed on a magnetic stand for 5 minutes at room temperature after which the supernatant was removed and discarded. The beads were resuspended in 200µl of Bead Washing buffer and placed on a magnetic stand for 5 minutes. The supernatant, containing rRNA, was removed and discarded.

Total RNA with rRNA reduction

Total RNA was diluted with nuclease-free water up to 1µg in a final volume of 10µl. To bind the rRNA, 5µl of rRNA binding buffer and RNA removal mix was added to the diluted RNA and incubate at 68°C for 5 mins following which 35µl rRNA removal beads were added and incubate at room temperature for 1 mins then place on magnet for a further 5 mins. The supernatant was transferred to a fresh tube. The RNA was purified by adding 99µl RNAClean XP Beads to the sample and incubating at room temperature for 15 mins. The supernatant was removed and the beads with bound RNA washed twice with 200µl of 70% ethanol. The bead bound RNA was resuspend in 11µl elution solution and incubated at room temperature for 1 mins and transferred to magnet for a further 5 mins. The supernatant containing the rRNA depleted RNA was transferred to a fresh tube.

First Strand cDNA synthesis

The mRNA was eluted, fragmented and randomly primed, adding 19.5µl of Frag, Prime, Finish Mix which contains random hexamers for RT priming to the polyA RNA bound beads and incubated at 94°C for 8 minutes. The beads were placed on a magnetic stand for 5 minutes, after which the supernatant was transferred to a fresh tube. The fragmented RNA was then reverse transcribed using reverse transcriptase and random primers, the First Strand Synthesis Act D Mix and SuperScript II (Invitrogen, U.K.) was mixed at a ratio of 9:1µl, 8µl of which was added to the fragmented mRNA. Actinomycin D is included in the First Strand Synthesis Act D Mix to prevent spurious DNA-dependent synthesis but allows RNA-dependent synthesis improving strand specificity. The sample was incubated at 25°C for 10 minutes, 42°C for 15 minutes and 70°C for 15 minutes.

Synthesis of Second Strand cDNA

The RNA template was removed and a second strand cDNA replacement strand generated, adding 20µl of Second Strand marking Master Mix to the sample and incubate at 16°C for 1 hour. Blunt ended dsDNA was generated incorporating dUTP in place of dTTP to quench the second strand during amplification as the polymerase cannot incorporate past this nucleotide.

AMPure XP beads were used to separate the ds cDNA from the Second Strand Marking Master Mix. AMPure XP beads was added to the sample and incubated at room temperature for 15 minutes and then placed on a magnetic stand for 5 minutes. The supernatant was removed and discarded. The beads

were washed twice with 200µl of 80% ethanol and the supernatant was discarded. The beads were air dried at room temperature for 10 minutes. The beads were eluted in 17.5µl Resuspension Buffer, incubated at room temperature for 2 minutes and placed on the magnetic strand for 5 minutes. The supernatant containing the RNA was transferred to a fresh tube.

Ligating Adapters

An ‘A’ nucleotide was added to the 3’ end of the blunt cDNA fragments, adding 12.5µl of A-Tailing Mix to the sample, to prevent them from ligating to one another during adapter ligation reaction and providing a complementary overhang to the ‘T’ nucleotide present on the 3’ end of the adapter for ligation. The sample was incubated at 37°C for 30 minutes and 70°C for 5 minutes. The Ligation mix, RNA Adapter Index and resuspension buffer were added at a volume of 2.5µl to the sample. The sample was then incubated at 30°C for 10 mins after which 5µl of Stop Ligation Buffer was added. A further AMPure XP bead clean up was conducted as previously described, using 42µl of Ampure XP beads. The sample was resuspended in 52.5µl of resuspension buffer.

An additional clean up step was conducted by adding 52.5µl of SPRI to the resuspended sample and incubated at room temperature for 5 mins and then placed on a magnetic stand for 5 mins. The supernatant was removed and discarded. The beads were washed twice using 200µl of 80% ethanol, incubate at room temperature for 30 seconds. The supernatant was discarded. The beads were allowed to air dry at room temperature for 10 minutes. The beads were resuspended in in 22.5µl Resuspension Buffer, incubated at room

temperature for 2 minutes and placed on the magnetic strand for 5 minutes. The supernatant containing the cDNA was transferred to a fresh tube.

Enrich DNA Fragments

The DNA fragments containing the adapter molecules on both end are selectively enriched by PCR. The PCR Primer Cocktail and PCR Master Mix are added at 5µl and 25µl respectively to the sample. The sample is incubated at 98°C for 30 seconds (secs). The samples then undergo 15 cycles of 98°C for 10 secs, 60°C for 30 secs and 72°C for 30 secs. The samples are then held at 72°C for 5 mins. The number of PCR cycles is kept to a minimum to avoid skewing the library representation. This is followed by an Ampure XP clean up, as previously described, using 50µl of Ampure XP beads and resuspending in 15µl of resuspension buffer.

Validate Library

Quality control and quantification of the DNA library must then be performed. To achieve high quality data optimal cluster densities must be reached which requires accurate quantification conducted using the dsDNA HS Assay Kit with the Qubit Fluorometer as per manufacturers instructions. The Qubit buffer containing, 198µl of Qubit dsDNA HS Buffer and 1µl of Qubit dsDNA HS Reagent per sample, is prepared. To individual assay tubes 199µl of the Qubit buffer solution and 1µl of each sample is mixed thoroughly by vortexing for 2-3 secs (for the standards add 190µl of Qubit buffer to 10µl of standard 1 and standard 2). The samples were then incubate at room temperature for 2 mins.

The standards were firstly read by the fluorometer (Thermo Fischer Scientific UK) to provide accurate markers for the sample quantification. The individual dsDNA library samples were then quantified.

The quality of the samples was analysed using the Agilent Technologies 2100 Bioanalyzer using a DNA-specific chips Agilent DNA 1000 or 5000. Loading 1µl of buffer to 1µl of sample. The size and purity of the library can be observed. A pure library should display a single peak representing a fragment size of approximately 260bp.

Normalize and Pool Libraries

The individual dsDNA library samples are normalized to 10 nM and pooled in equal volumes to prepare them for cluster generation (see section Introduction 1.6.1 for description of cluster generation during Illumina sequencing). The samples were then run on the MiSeq or Next-Seq platforms (Illumina).

2.4 Bioinformatic Analysis

The sequencing data was demultiplexed. The reads were trimmed to remove adapter sequences and filtered to remove low quality reads using trim_galore software. The read quality was ensured to have a quality (Q score) of 30, 1 nucleotide sequencing error in 1000, using fastqc analysis.

2.4.1 Bioinformatic analysis of virus transcription and replication

A bioinformatic pipeline was developed for the analysis of virus transcription and replication. The reads contained in the R1 and R2 files generated from the HTS described above were independently mapped to a reference genome using BWA alignment software. The reference genomes for PIV5 strain W3, PIV5 strain CPI+, PIV2 and MuV are available from GenBank, accession numbers are JQ743318.1, JQ743321.1, NC_003443.1 and GU980052.1. A consensus sequence was constructed for PIV3 and used as a reference genome as the available PIV3 reference genomes did not provide sufficient read coverage across all of the genome of the strain used due to the presence of several mutations. The aligned reads were split based on directionality using the flag field of the sam file that indicates read orientation. The viral mRNA/antigenome reads from the R1 and R2 files were concatenated, producing a file that contained reads generated exclusively from viral mRNAs/antigenomes. A similar process was carried out for viral genome reads. The viral genome and mRNA/antigenome reads were then mapped separately to the respective reference sequence using BWA software. The number of viral mRNA/antigenome and viral reads was ascertained from these alignments. The alignments were visualized using Tablet (Milne et al., 2013). The viral mRNA and antigenome reads are of the same orientation and therefore cannot be separated by directional analysis. The contribution of antigenomes to the viral mRNA cannot be more than the average coverage of reads of the L mRNA. The contribution of antigenome reads to the transcription gradient was not significant and therefore would not interfere with downstream bioinformatic analysis. As the antigenome reads did not significantly contribute to the positive

sense reads, these were referred to as only viral mRNA reads in subsequent analysis. The remaining nuclear and mitochondrial ribosomal RNA (rRNA) reads were removed from the total cellular reads by aligning the reads contained in the R1 and R2 files to reference genomes for 18S, 28S, 5S, 5.8S and mitochondrial rRNA (Accession numbers NR_003286.2, NR_003287, X51545, J01866, NC_012920). The unaligned reads were quantified and referred to as total cellular reads in downstream analysis. To calculate the abundance of viral mRNAs and genomes the number of viral mRNAs and genome reads was compared to the total cellular reads.

2.4.2 Analysis of virus transcription

The abundances of viral mRNAs was calculated using FPKM values obtained by RSEM software (Li and Dewey 2011). The coordinates of the genes required for the program was provided by manually creating gtf files. This method cannot distinguish between the transcripts generated via RNA editing and the insertion of G residues at the RNA editing site. To ascertain the ratio of G inserts the reads overlapping the RNA editing site were identified using a 10 nt search string upstream and downstream of the RNA editing site. The number of reads overlapping the RNA editing site containing 1-12 G insertions were quantified and compared to the total number of reads overlapping the RNA editing site.

To quantify the number of reads that cross the Ig sub alignments were extracted from the previously described viral mRNA alignment generated using BWA software. The average coverage of the genes was ascertained by aligning reads to the viral genes using BWA software that was subsequently visualized

in Tablet, from which the average coverage of the genes could be determined. The generation of read-through mRNAs was calculated by quantifying the number of reads that cross the Ig and comparing this to the average coverage of the gene immediately upstream.

The antigenome reads were identified by quantifying the reads that crossing the Le region into the NP gene. A sub-alignment was extracted from genome positions 44-54 nts from the genome and mRNA/antigenome read BWA alignments. The sum of the reads present at these positions was quantified and the ratio of genome and antigenomes reads was determined.

2.4.3 Establishing the presence of defective interfering genomes

The presence of DIs was established using ViReMa (Virus Recombination Mapper) software (Routh and Johnson 2013). The software was run on the R1 and R2 files separately. ViReMa detects recombination events by identifying reads with a breakpoint, therefore, the read would map to two different areas of the genome. The identification of reads containing different breakpoints allowed the identification and quantification of different DI populations.

2.5 Immunofluorescence

Cells were grown in monolayers on 10mm coverslips. The cells were fixed in 5% formaldehyde, 2% sucrose in PBS for 10 mins, before permeabilisation with 0.5% NP-40, 10% sucrose in PBS. The cells underwent repeated washes with

1% FBS in PBS. The cells were incubated with the appropriate diluted primary antibody for 1 h. The cells underwent repeated washes again, following which the cells ere incubated for 1 h with a secondary antibody of Texas Red or FITC conjugated secondary antibody (Oxford Biotechnology Ltd., U.K.) with the DNA-binding flurochrome 4', 6-diamidino-2-phenylindole (DAPI; 0.5µg/ml; Sigma-Aldrich Ltd., U.K.). All reactions were conducted at room temperature in a humidified chamber. The coverslips were washed in PBS and mounted on slides using Citifluor AF-1 mounting solution (Citifluor Ltd., U.K.). Immunofluorescence was visualised using a Nikon Microphot-FXA microscope.

2.6 ^{35}S-methionine labelling of proteins

Cells were grown in 25cm^2 flasks and infected with the appropriate virus at an MOI of 10. The cells were metabolically labelled at specified times for 1 h on a rocking platform with ^{35}S-methionine (25µCi/25cm^2 flask; Amersham International Ltd., U.K.). The cells were washed in PBS, following which disruption buffer was added to the cells which were then sonicated and heated to 100° for 3 mins. The polypeptides present in total cell extracts were separated by SDS-PAGE through a 4-12% polyacrylamide gel, and the labeled polypeptides were visualized using a phosphoimager

2.7 Immunoblots of proteins

Cells were grown in 25cm^2 flasks and infected with the appropriate virus at an MOI of 10. The cells were washed in PBS, following which disruption buffer was added to the cells which were then sonicated and heated to 100° for 3 mins.

The polypeptides present in total cell extracts were separated by SDS-PAGE through a 4-12% polyacrylamide gel. The proteins were transferred to PVDF (polyvinylidene difluoride) membrane, which had been activated with methanol for 1 min. The membrane was blocked for 1hr at room temperature in blocking buffer. The membrane underwent several washes using washing buffer. The membrane was incubated with the appropriate diluted primary antibody for 1 h. The membrane underwent repeated washes again, following which the membrane was incubated for 1 h with an appropriate secondary antibody (Li-Cor). The antibody-bound proteins were then visualised using Odyssey Clx imager (Li-Cor).

2.8 Nucleocapsid Purification

Cells were grown in 300cm^2 flasks and infected with the appropriate virus at an MOI of 10. Infected cells were harvested and washed with PBS. The cells were then scrapped into 10 mls of PBS and pelleted at 700 rpm for 10 min. The pellet was resuspended in ice-cold lysis buffer (150 mM NaCl, 50 mM Tris-HCl, 0.6% NP-40), incubated on ice for 5 mins and centrifuged at 4,200 rpm for 5 min at 4°C. The supernatant was transferred to a fresh tube and incubated on ice. EDTA was added at a concentration of 6 mM. Linear 25 to 35% cesium chloride gradients were prepared. The lysed cells were overlaid on the gradients which were centrifuged overnight at 175,000 rpm at 12°C. Nucleocapsid bands were harvested and the RNA extracted using TRIzol (Invitrogen) according to the manufacturer's instructions.

Chapter three

Technical

Previous research has demonstrated that paramyxoviruses control viral gene expression by stop-start transcription and the generation of a transcription gradient, with genes located at the 3' end of the viral genome producing the most abundant mRNAs and genes positioned at the 5' end producing the least abundant mRNAs. PIV5 strain W3 transcription was investigated using HTS to analyse the generation of viral mRNAs and the transcription gradient. Using this analysis, a bioinformatic pipeline was established that could be employed to analyse the transcriptome of additional paramyxoviruses as well as other negative strand RNA viruses.

3.0 Pipeline for analysing virus transcription

Human skin fibroblast (HSF) cells were infected with PIV5 strain W3 at an MOI of 50 pfu/cell. RNA was extracted at 18 hours post-infection (h p.i.). The extracted RNA was subjected to library preparation using TruSeq stranded mRNA library preparation kit (Illumina) which isolates mRNAs by polyA selection using oligo(dT) beads. This library preparation method generates directional sequencing data, which retain information on the strand orientation of the selected RNA. Paramyxoviruses have single-stranded, negative sense RNA genomes, from which positive sense viral mRNAs and antigenomes are transcribed or replicated. Directional sequencing allows the reads generated by HTS to be separated based on their directionality, thus allowing the origin of the read from viral genomes (-ssRNA) or viral mRNAs (+ssRNA) and antigenomes (+ssRNA) to be established. The reads generated from viral mRNAs and

antigenomes cannot be distinguished from each other, as these are of the same positive sense orientation. Directional analysis of the sequencing data is only required if the orientation of the reads must be determined for downstream bioinformatic analysis. As polyA selection should have only selected mRNAs, the orientation of the reads would not have to be determined in this experiment, and consequently the data were not initially subjected to directional analysis.

Sequencing was conducted on a MiSeq (illumina) platform using two primers, an antisense (forward) primer and a sense (reverse) primer, from which two files (R1 and R2, respectively) were produced (see Introduction section 1.6). These files were concatenated and contained 6,523,498 reads. The reads present in the R1 and R2 files were generated from all of the mRNAs present in the cell at the time of RNA extraction, including both host and viral mRNAs; these are referred to as total cellular reads. The total cellular reads were mapped to the PIV5 strain W3 reference sequence (Accession number JQ743318.1) by alignment using the BWA software, and the alignment was visualised using the Tablet software. A screen shot of the alignment visualized in Tablet is provided in Fig 3.1.

The reads that map to the PIV5 strain W3 reference genome are of viral origin and account for 4.7% of the total cellular reads. A greater number of aligned reads mapped to genes transcribed from the 3'-end of the genome, NP and V/P, acquire a greater number of aligned reads, suggesting a greater abundance of the mRNAs from these genes. However, genes downstream of M display an approximately equal coverage of reads, implying that a similar abundance of mRNAs was generated. This flat transcription gradient for genes

downstream of M does not correlate with the previously reported decreasing gradient from the 3' to the 5' end of the genome.

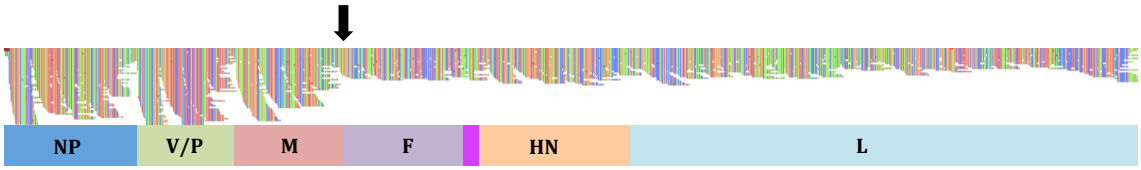


Fig 3.1. BWA alignment of the PIV5 strain W3 transcriptome in HSF cells at 18 h p.i. visualized in Tablet, using the reads from the concatenated R1 and R2 files. The RNA was subjected to polyA selection prior to library preparation followed by sequencing using the MiSeq platform (Illumina). Coloured boxes indicate approximate gene positions and contain the names of the genes. The individual coloured vertical bars represent the coverage (number of reads) aligned to each nucleotide of the reference sequence. Each nucleotide is represented by a different colour, G is red, A is green, T is blue and C is yellow. The overall coverage of a gene indicates the abundance of viral mRNA produced from that gene. The black arrow indicates the boundary between the M and F genes.

One possible explanation for this finding is the presence of contaminating viral reads that were generated from the viral genome. It is possible that during RNA extraction and polyA selection, viral mRNAs hybridised to complementary viral genomes. Thus, the subsequent analysis of viral mRNAs would incorporate reads from hybridised viral genomes. Reads that are of viral genome origin would also align to the reference genome, thereby increasing the read coverage at that genome position, interfering with the analysis of the abundances of viral mRNAs, and distorting the transcription profile of PIV5 strain W3.

To investigate whether viral genome sequences were present in the polyA selected RNA samples, directional analysis was conducted to distinguish reads originating from viral genomes from those originating from viral mRNAs/antigenomes. The use of the antisense and sense primers to generate the R1 and R2 files allows the orientation of the RNA molecule from which each read originated to be determined. Splitting the reads based on directionality separates the reads generated from viral genomes from those generated from viral mRNAs/antigenomes. The R1 and R2 files were individually mapped to the PIV5 strain W3 reference sequence by alignment using the BWA software. The aligned reads were then split based on directionality (as summarised in Fig 3.2A), using the flag field of the SAM alignment file that indicates read orientation. The viral mRNA/antigenome reads from the R1 and R2 files were concatenated, producing a file that contained reads generated exclusively from viral mRNAs/antigenomes. A similar process was carried out for viral genome reads. A schematic of the bioinformatic process, which separates the reads based on directionality, is provided in Fig 3.2B. The viral genome and mRNA/antigenome reads were then mapped separately to the PIV5 strain W3 reference sequence by alignment using the BWA software, and visualised in Tablet. A screen shot of the alignments are provided in Fig 3.3.

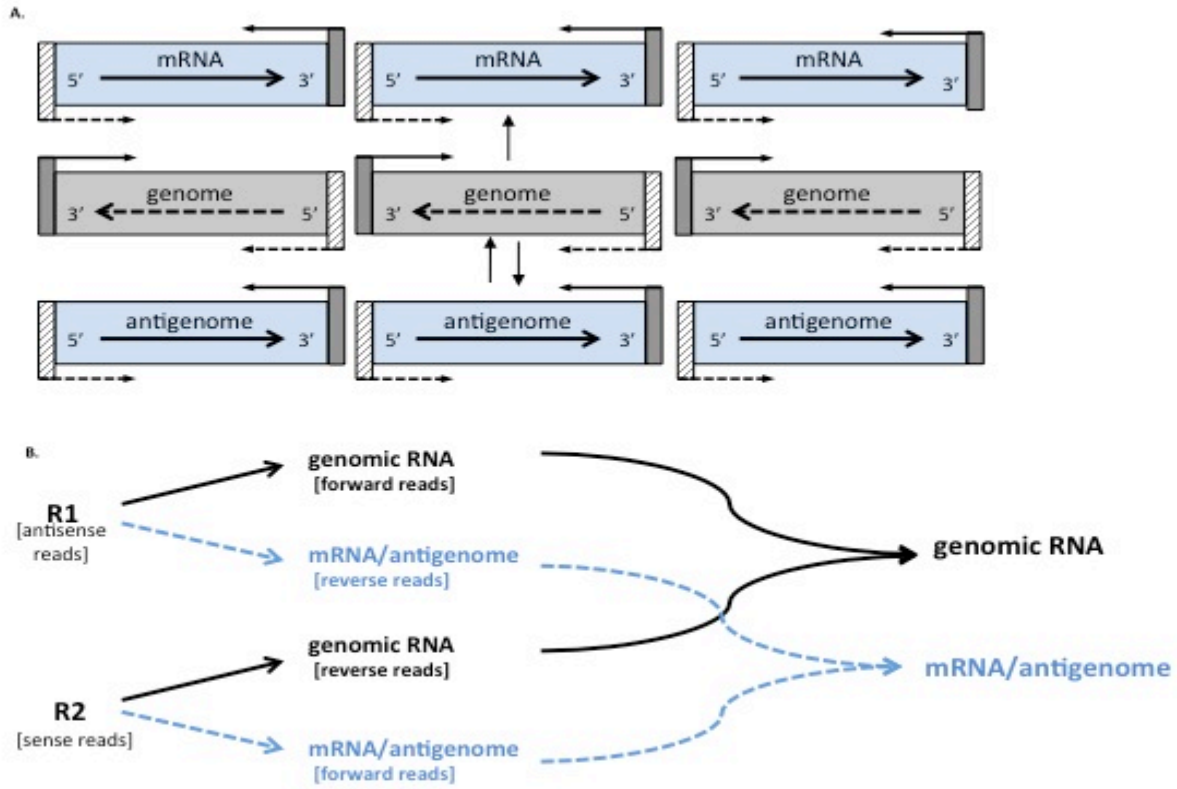


Fig 3.2. Schematics depicting the sorting of reads into those originating from viral genomes and those originating from viral mRNA/antigenomes. (A) The directionality of reads relative to the negative sense viral genome (grey rectangles) and positive sense viral mRNA and antigenome (blue rectangles) generated during library prep and subsequent sequencing. The arrows showing the 5' to 3' direction in the centre of the rectangles indicate read orientation. The directionality of the adapters used during sequencing to generate the R1 (forward primer) and R2 (reverse primer) files are shown at the ends of the rectangles, with the 3' primer represented by a solid grey rectangle and the 5' primer by a hatched rectangle. (B) Directional splitting of the forward and reverse reads from the R1 and R2 files. The viral mRNA/antigenome reads were concatenated into a single file, and similarly the viral genome reads.

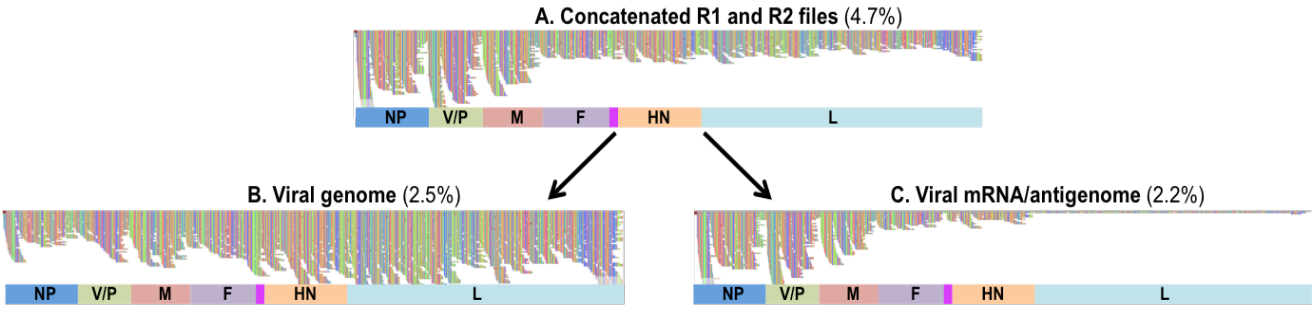


Fig 3.3. BWA alignments of the PIV5 strain W3 transcriptome in HFS cells at 18 h p.i. visualised in Tablet, using reads sorted directionally. The individual coloured vertical bars represent the coverage (number of reads) aligned to each nucleotide of the reference sequence. Each nucleotide is represented by a different colour, G is red, A is green, T is blue and C is yellow. The percentage of reads present in the alignments compared to total cellular reads is provided. Coloured boxes indicate approximate gene positions and contain the names of the genes. (A) Aligned reads from the concatenated R1 and R2 files (non-directional analysis). (B) Viral genome reads generated from directional splitting of the R1 and R2 files (directional analysis). (C) Viral mRNA/antigenome reads generated from directional splitting of the R1 and R2 files (directional analysis).

The alignment produced using directional analysis of the viral genome reads (Fig 3.3B) revealed that approximately 2.5% of total cellular reads aligned to the PIV5 strain W3 reference sequence. This indicates the presence of a significant amount of viral genomes present in the sample. The viral mRNA/antigenome reads were 2.2% of total cellular reads. The alignment of viral mRNA/antigenome reads (Fig 3.3C) reveals a decreasing transcription gradient, with the greater coverage of reads aligning to the genes at the 3' end of the genome (NP and V/P) and a lesser coverage of reads mapping to the L gene at

the 5’ end. The data confirm that a significant number of viral genomes had co-purified with the viral mRNAs during library preparation, despite the use of poly A selection. This was most likely due to hybridisation of viral genomic RNA to viral mRNAs during RNA extraction and polyA selection, as previously discussed.

Although directional analysis demonstrated the presence of viral genomes and permitted it to be quantified, it did not establish the potential presence of viral antigenome, as this has the same polarity as viral mRNAs. If a proportion of the mRNAs selected by oligo(dT) beads had hybridised to viral genomes, it is possible that subsequently the viral genomes had hybridised to antigenomes. If a significant proportion of reads were of antigenome origin, this could have compromised the accuracy of the PIV5 strain W3 transcription analysis. The antigenome is a full length complementary copy of the genome, and therefore should be present at a relatively constant coverage level in an alignment of the mRNA/antigenome reads against the PIV5 strain W3 reference sequence. The coverage of the antigenome can be estimated from the coverage of reads in the Le region, which should not contain any mRNA reads as it is located outwith the gene boundaries. Some models for paramyxovirus transcription suggest that the Le region produces a short transcript. However, due to fragment size selection during library preparation and the length of reads generated from sequencing (in this analysis, read length was approximately 300 nt), these short transcripts would not have been represented in the data. Additionally, the reads originating in the Le region cross the NP gene boundary, ensuring that they are of antigenome origin. The Tr region could not be used to estimate the antigenome reads, as coverage is reduced at the 5’ end of the genome due to

the requirement for random priming and a minimum fragment size during library preparation and sequencing. Therefore, the average coverage of 10 nt at the 5' end of Le (positions 45-54) was calculated, in order to reduce this effect. The average coverage at positions 45-54 was calculated as 15.6 reads. A rearranged Lander/Waterman equation (Lander and Waterman 1988) [coverage = (length of read x number of reads)/length of genome] was used to estimate the number of reads attributed to viral antigenome throughout the full length of the alignment. The average length of a read was extracted from the alignment and calculated as 300 nt, and the length of the PIV5 strain W3 single-stranded genome is 15246 nt. Therefore, the number of reads attributed to the antigenome was 792.8, which comprises 0.01% of the total cellular reads. Furthermore, the coverage of antigenome reads cannot have exceeded that of the L gene (Fig 3.3C) extended over the whole genome, calculated as 3794 read contributing 0.06% of total cellular reads. In conclusion, it is unlikely that these low numbers of antigenome reads would have compromised the transcriptome analysis.

These results demonstrate that directional analysis is essential for accurate paramyxovirus transcriptome research using HTS data, and led to the establishment of a general bioinformatic pipeline for analysing negative strand virus transcription and gene expression analysis. After removing viral genome reads from the PIV5 strain W3 transcription profile and showing that antigenomes make a negligible contribution to viral mRNA/antigenome reads, a decreasing gradient of mRNAs from the 3' end of the genome was revealed (Fig 3.3C).

3.1.1 Optimisation of library preparation

Although viral genome sequences were shown to be present in the above analysis, the method used to select mRNAs was not appropriate for quantifying viral genomes. As an alternative, it would be possible to use the TruSeq stranded total RNA with Ribo-Zero human/mouse/rat library preparation kit (Illumina), which depletes cytoplasmic-coded ribosomal RNA (rRNA), rather than polyA selection. This method could be used together with directional sequencing to quantify both viral mRNAs and viral genomes. Furthermore, this procedure could be used to detect the presence of defective interfering viral genomes, which are discussed in chapter 6. The ability to analyse viral transcription and replication from the same data offers a unique opportunity to achieve a complete overview of the dynamics of viral infection.

To determine whether rRNA depletion of total RNA preparations could be used for the accurate analysis of viral mRNAs, which can be at low levels in the cell, a study was conducted to compare two library preparation kits: the TruSeq stranded mRNA Library preparation kit (involving polyA selection, as previously described) and the TruSeq stranded total RNA library preparation kit (involving rRNA depletion). Library preparation was conducted on the same samples to exclude sampling bias. RNA was extracted from A549 cells infected with PIV5 strain W3 at an MOI of 10 pfu/cell at 6, 12 and 18 h p.i. Following sequencing, the data were directionally analysed as previously described. Furthermore, since the selection or depletion of target RNA is not completely efficient during library preparation, any remaining rRNA reads were removed from the total cellular reads by alignment using the BWA software, for improved accuracy when calculating the abundances of viral RNAs. Additionally, the percentage of

mitochondrial rRNA reads present in each sample was also calculated, and these were removed from the total cellular reads. Table 3.1 shows the percentages of cytoplasmic and mitochondrial rRNA reads compared to total cellular reads, and the percentages of viral mRNA reads before and after removal of cytoplasmic and mitochondrial rRNA reads. These results indicate that the cytoplasmic and mitochondrial rRNA reads can have a significant effect of the quantification of viral mRNA abundance. The samples subjected to polyA selection showed significant proportions of cytoplasmic and mitochondrial rRNA reads (up to 3.04 and 8.5%, respectively). The rRNA depleted samples also showed significant proportions of cytoplasmic and mitochondrial rRNA reads (up to 1.88 and 13.2%, respectively). The impact on the accurate quantification of the abundance of viral mRNAs due to the presence of cytoplasmic and mitochondrial rRNA reads can be seen especially for the samples isolated at 12 and 18 h p.i. After removal of cytoplasmic and mitochondrial rRNA reads, the apparent abundance of viral mRNAs increased as much as 1%. The abundance of mitochondrial RNA reads is particularly apparent in the rRNA reduction approach, and indicates that a method that reduces both cytoplasmic and mitochondrial rRNA would be more effective.

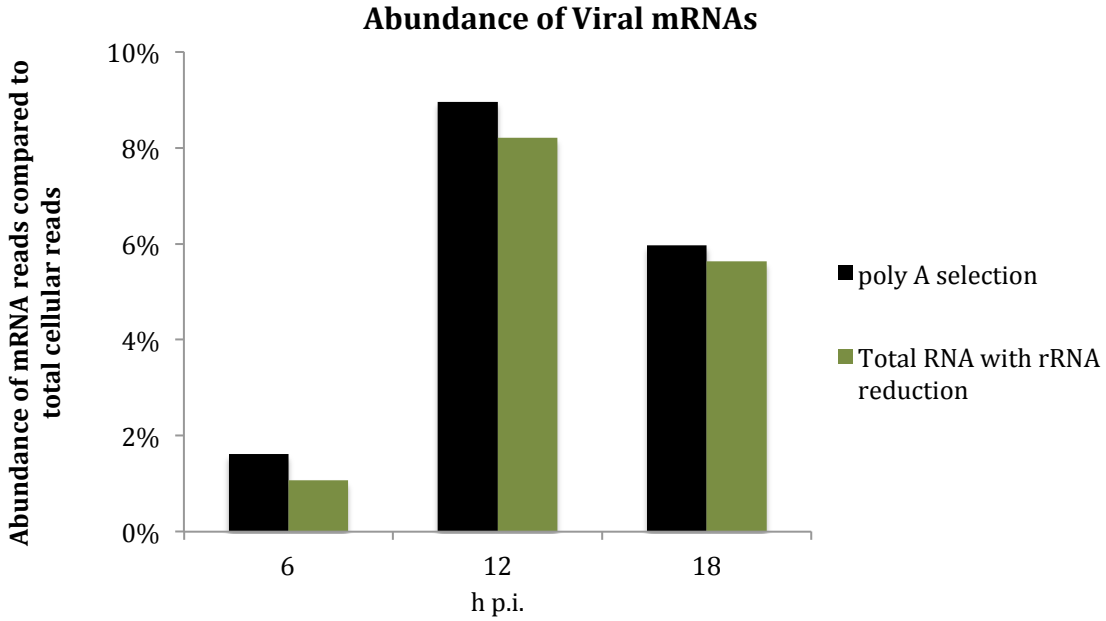
When cytoplasmic and mitochondrial rRNA reads were eliminated from the analysis, the abundance of viral mRNAs compared to total cellular RNA reads (with the cytoplasmic and mitochondrial rRNA reads removed) was calculated for each library preparation method (Fig 3.4A). To ensure that the coverage of reads when aligned to the PIV5 strain W3 reference genome was sufficient, which is critical for downstream analysis, the viral mRNA reads were aligned to the reference genome (Fig 3.4B). There was very little difference between the

library preparation methods in the relative abundance of viral mRNAs (Fig 3.4A), although rRNA depletion generated slightly less viral mRNA at each time point, compared to polyA selection. However, this decrease is not sufficient enough to compromise analysis of the abundance of viral mRNAs, as can be seen in Fig 3.4B. Alignments of viral mRNA reads to the reference PIV5 strain W3 genome showed little difference between the methods. For example, the mRNA profile at 6 h p.i. for polyA selected RNA is essentially indistinguishable from that for rRNA depleted RNA (Fig 4B).

Table 3.1. Percentages of viral mRNA reads compared to total cellular reads before and after cytoplasmic and mitochondrial rRNA reads were removed from total cellular reads from the data obtained using polyA selection or rRNA depletion library preparation methods.

	Viral reads before removal of cytoplasmic and mitochondrial rRNA reads		Percentage of cytoplasmic and mitochondrial rRNA reads compared to total cellular reads					Viral reads after removal of cytoplasmic and mitochondrial rRNA reads	
	Viral mRNA reads		Cytoplasmic rRNA		Mitochondrial rRNA			Viral mRNA reads	
h p.i.	PolyA selection	rRNA depletion	PolyA selection	rRNA depletion	PolyA selection	rRNA depletion		PolyA selection	rRNA depletion
6	1.46	1.02	1.62	0.36	8.5	3.8	→	1.62	1.07
12	8.27	7.22	1.55	0.23	6.1	11.8		8.96	8.21
18	5.35	4.78	3.04	1.88	7.3	13.2		5.97	5.63

A



B

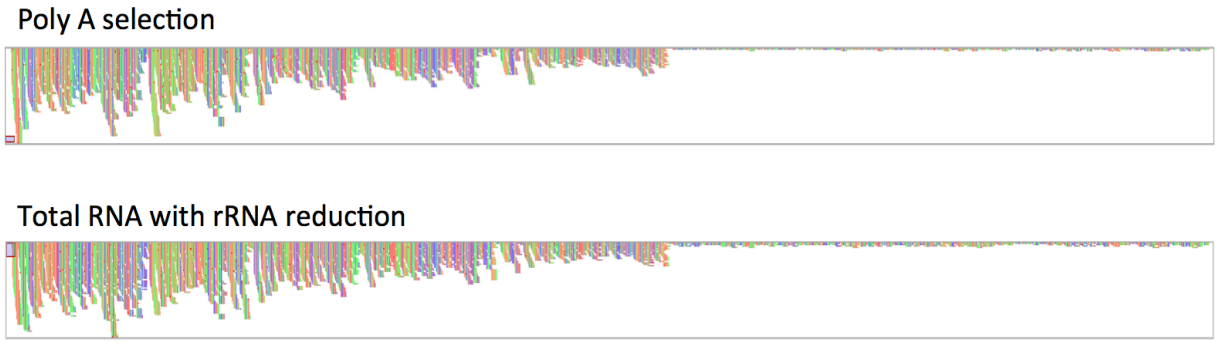


Fig 3.4. A comparison of the abundance of viral mRNAs using polyA selection or total RNA with rRNA reduction library preparation methods. (A) Abundance of PIV5 strain W3 mRNA compared to total cellular reads as obtained using polyA selection or rRNA reduction. RNA was extracted from infected A549 cells at 6, 12 and 18 h p.i. and subjected to library preparation and subsequent sequencing and directional analysis. (B) Screen shots of BWA alignments for data obtained at 6 h p.i. visualised in Tablet.

The advantage of using rRNA reduction rather than polyA selection is that it facilitates the quantification of the abundance of virus genome reads (Fig 3.5). The abundance of viral genome reads gradually increased between 6 and 18 h p.i. from 0.09 to 1.42% of total cellular reads. Additionally, the proportion of antigenomes was estimated by counting the coverage of reads at positions 45-54 and using the Lander-Waterman equation as previously described. The proportion of antigenomes present at 6, 12 and 18 h p.i. was estimated as 0.07, 0.21 and 0.16%, respectively, of total cellular reads. Antigenome read coverage accounted for a very low proportion of total cellular reads, and, in any case, could not exceed mRNA/antigenome coverage of the L gene, and therefore did not significantly affect the observed transcription gradient (Fig 3.4B). Consequently, the numbers of mRNA/antigenome reads calculated in later experiments are referred to solely as mRNA reads.

In conclusion, using rRNA reduction instead of polyA selection offers a powerful opportunity to analyse viral transcription and replication in parallel during infection.

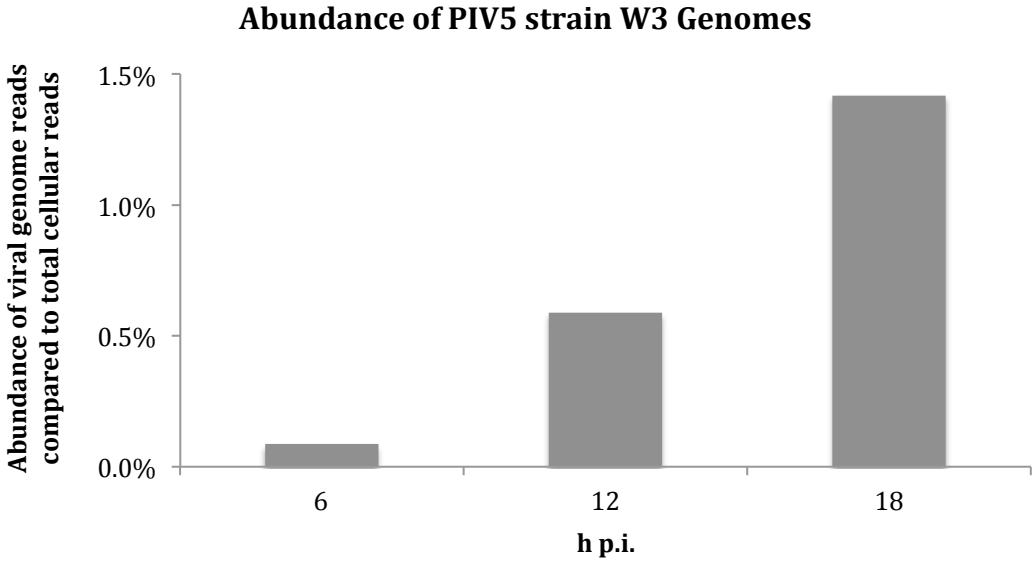


Fig 3.5. Abundance of viral genome reads compared to total cellular reads after remaining cytoplasmic and mitochondrial rRNA reads had been removed. The data were directionally analysed using the bioinformatic pipeline established previously.

3.1.2 Optimised pipeline for analysing transcription and replication

The analysis above used the TruSeq Stranded Total RNA Ribo-zero human/mouse/rat library preparation kit, which depletes human/mouse/rat cytoplasmic rRNA. However, as shown in Table 3.1, a significant proportion of total cellular reads were of mitochondrial rRNA origin, which could impact the accuracy of the analysis. Therefore, all subsequent experiments were conducted using the TruSeq Stranded Total RNA Ribo-Zero Gold library preparation kit, which also depletes mitochondrial rRNA. In addition, following library preparation, samples were sequenced using the Illumina NextSeq

platform, which generates more reads (up to 260 million) per run. Following sequencing, the most essential step for accurate analysis of viral transcription and replication was directional analysis using the estabished bioinformatic pipeline, which allows reads derived from mRNA to be separated from those generated from genomic RNA.

3.1.3 Establishing the presence of defective interfering genomes

The repression of viral RNA synthesis and the subsequent establishment of persistent infections by paramyxoviruses, as well as the induction of IFN, has been associated with the generation of copyback DIs (see Introduction section 1.5). To determine whether the established workflow described above using total RNA with rRNA reduction could be employed to reliably detect the presence of DIs in RNA extracted from infected cells without the need for nucleocapsid purification, a comparative study was conducted. A549 cells were infected with a DI rich virus stock of PIV5 strain VΔC, at an MOI of 10 pfu/cell. At 24h p.i. either total cellular RNA was extracted by Trizol, or viral nucleocapsids were first purified on a CsCl gradient prior to RNA extraction. The extracted RNA was subjected to cytoplasmic and mitochondrial rRNA reduction library preparation followed by sequencing. The data was then subjected to analysis by ViReMa (Virus Recombination Mapper) software that identifies recombination events also known as breakpoints where the RNA polymerase during virus replication has jumped along the template to produce an internal deletion DI or alternatively switched to the nascent strand to generate a copyback DI. Reads generated from these DIs would contain two regions that map to different positions of the genome. ViReMa identifies reads

that contain breakpoints, thereby identifying internal deletions and copyback DIs (see Introduction Figure 1.7 for information regarding the generation of copyback and internal deletion DIs). The software identifies the position of the recombination event and quantifies the number of reads containing the recombination. This allows the identification and quantification of distinct DI populations. There were no significant internal DIs identified from the nucleocapsid purified or not nucleocapsid purified samples. Several populations of copyback DIs were identified (Table 3.2).

Table 3.2. The identification of distinct copyback DIs populations, the genome and antigenome position of the breakpoint, the number of reads which are generated from the DIs and the proportions of the populations

Nucleocapsid purified				No Nucleocapsid purification			
antigenome	genome	number of reads	Proportion of the DI population	antigenome	genome	number of reads	Proportion of the DI population
14041/2	15025/6	181	33%	14041/2	15025/6	157	32%
14824/5	15159/60	207	38%	14824/5	15159/60	180	37%
14869	15157	30	6%	14869	15157	35	7%
15021	14046	124	23%	15021	14046	114	23%

There was no significant difference between the results obtained from total RNA extracted from infected cells and RNA isolated from purified nucleocapsids. ViReMa identified 4 identical DIs populations present in both samples (Table 3.2). Furthermore, the proportion of DIs to genomes in the two samples was nearly identical. Two of the four identified DI, with junctions at positions 14827-15157 and 14873-15153, matched within +/- 4 nts those determined by Killip et al (2013), who used a similar, but not identical, preparation of PIV5 strain VΔC to identify DI populations. These results demonstrate that ViReMa software can

be employed to determine the presence of DIs in samples that have not undergone nucleocapsid purification and can therefore be utilized on HTS data generated using the pipeline for the analysis of virus transcription and replication described above.

3.2 Analysis of Parainfluenza Virus 5 replication and transcription

The bioinformatic pipeline established above was used to investigate PIV5 strain W3 transcription and replication during the course of infection. A549 cells were infected with PIV5 strain W3 at an MOI of 10 pfu/cell, and total cellular RNA was extracted at 0, 6, 12, 18 and 24 h p.i. The RNA was subjected to cytoplasmic and mitochondrial rRNA reduction, and libraries were made and sequenced. The data were then directionally analysed as previously described. As the efficiency of cytoplasmic and mitochondrial rRNA depletion during library preparation can vary between samples, the remaining reads generated from cytoplasmic and mitochondrial rRNA were removed from the total cellular reads. To estimate the relative abundance of viral genomes and viral mRNAs in infected cells, the numbers of relevant reads, obtained from aligning the reads to the PIV5 strain W3 reference genome using the BWA software, were compared to the numbers of total cellular reads (Fig 3.6). This was conducted in triplicate, from which the standard deviation was calculated and depicted as error bars. A Table of the number of total cellular, viral genome and viral mRNA reads at each time point is provided in Appendix I.

There was an initial surge of transcriptional activity during the initial stages of infection, as the rate of transcript production between 6 and 12 h p.i. increased

nine-fold (Fig. 3.6). This activity peaked between 12 and 18 h p.i. at 4.59 and 4.34% of total cellular reads, respectively, after which the abundance of viral mRNAs decreased to 2.31%. As this analysis reports the overall abundance of transcripts and not the active rate of mRNA production, it is likely that the rate of viral transcription decreases after 12 h p.i.

The abundance of viral genome reads gradually increased between 6 and 24 h p.i. from 0.03 to 1.82%, suggesting a gradual increase in the rate of viral genome production over the initial 24 h of infection. At the earlier time point of 6 h, the abundance of viral genome reads had increased from the input virus at 0 h p.i. suggesting a low rate of new viral genome production. The generation of new viral genomes would require the prior production of antigenomes and viral protein NP for encapsidation. A more detailed investigation of viral replication during the initial stages of infection is presented in section 3.2.2.

These results demonstrate that, during the first 18 h of infection, transcription is the principal viral process, with an initial surge in activity between 6 and 12 h p.i. This correlates with the widely accepted model of paramyxovirus infection, in which the virus commences transcription immediately upon entering the cell. By 24 h p.i., the abundance of viral genome production is comparable to that of mRNAs, suggesting a change in virus RNA polymerase activity in which transcription is no longer the primary process and the RNA polymerase has been in a replicative mode. To gain further insight into how the virus controls the switch between transcription and replication, the abundance of individual viral mRNAs was analysed, in order to identify changes in viral transcription

(indicating protein expression) that might play a role. All bioinformatic analyses were conducted using the data set described above.

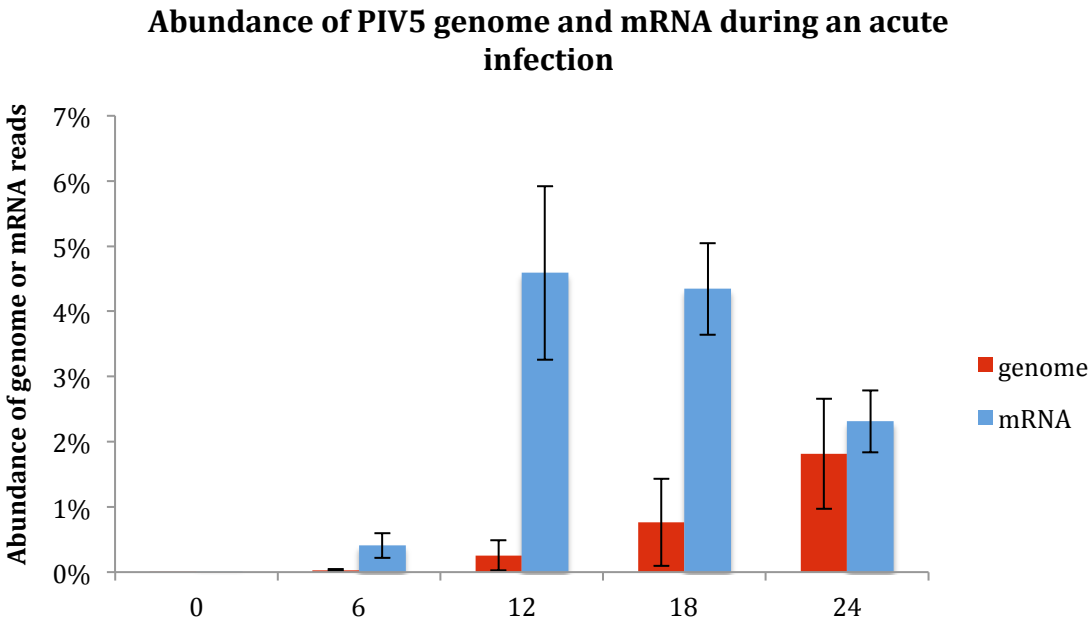


Fig 3.6. Abundance of PIV5 strain W3 viral genome and mRNA reads at 0, 6, 12, 18 and 24 h p.i. compared to the number of total cellular reads after removal of cytoplasmic and mitochondrial rRNA reads. The data were directionally analysed using the established bioinformatic pipeline. The bars show standard deviation values based on three independent experiments.

3.2.1 Analysis of viral transcription

The abundance of viral mRNAs was investigated in order to identify changes that might account for the switch in viral transcription and replication during PIV5 strain W3 infection. This was done by analysing the abundance of viral mRNAs using FPKM values (fragments per kilobase of transcript per million mapped reads). FPKM values normalise the abundance of transcripts

generated from individual genes to account for differences in gene length, thus allowing the transcription of different genes to be compared (see Introduction section 1.6). The abundances of viral mRNAs were calculated using RSEM software (Fig 3.7).

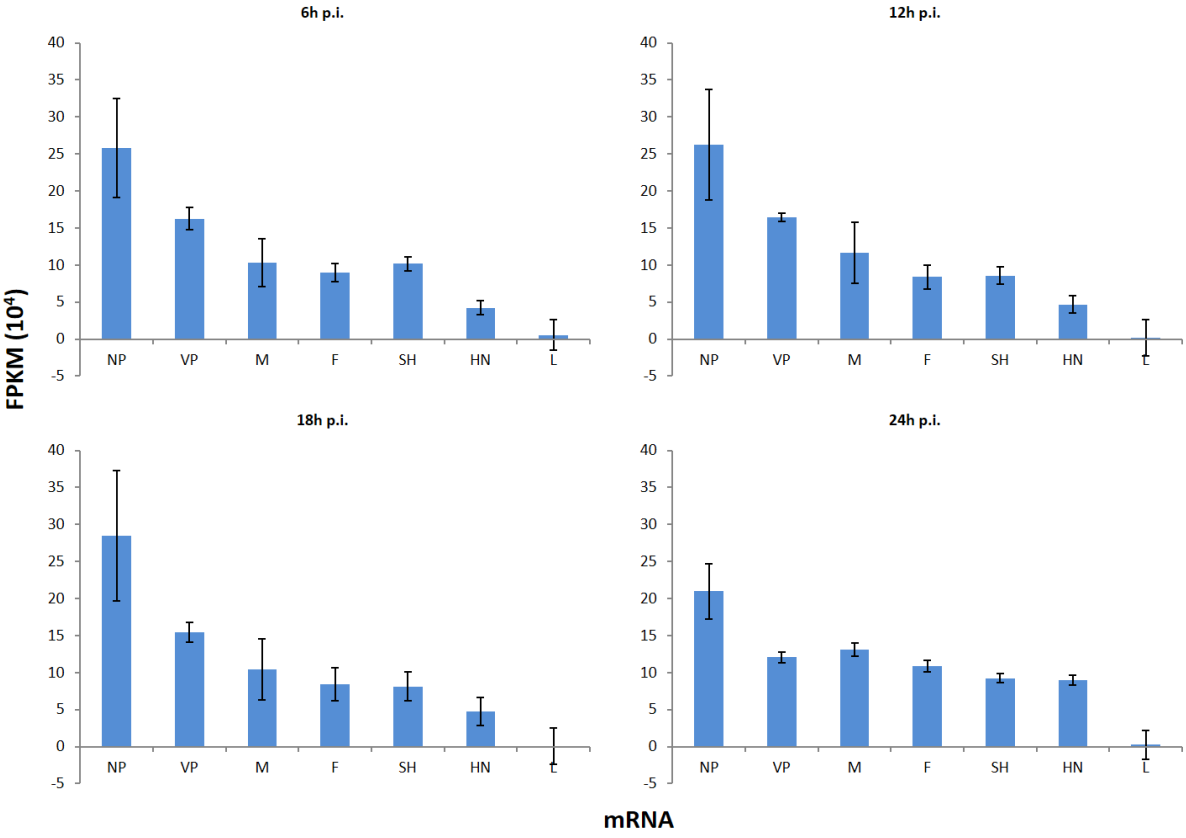


Fig 3.7. Transcription of PIV5 strain W3 genes at 6, 12, 18 and 24 h pi.. The data were directionally analysed using the established bioinformatic pipeline. The FPKM values on the *y*-axes were obtained using the RSEM software. The bars show standard deviation values based on three independent experiments.

While the earlier time points of 6, 12 and 18 h p.i. exhibited little difference in the abundance of viral mRNA transcripts, all displayed a transcriptional gradient

with higher abundance of mRNA from genes at the 3' end compared to the 5' end of the genome. There was a plateauing of the abundance of viral V/P, M, F, SH and HN mRNAs at 24h p.i. However, this could be due to the stability of transcripts rather than changes in virus gene expression. This suggests that there are no differences in virus gene expression that could account for the switch between transcription and replication.

RNA editing and generation of the V, I and P mRNAs

The V and P mRNA transcripts, which are generated differentially via RNA editing in a tract of 4 G residues at position 2340 in the PIV5 strain W3 genome, cannot be distinguished using the method describe above. The abundance of these transcripts was estimated by identifying reads that overlap the RNA editing site. The reads were identified in the R1 and R2 data files by using a 10 nt search string upstream and downstream of the editing site, upstream string 3'-TTAAGAGGG-5' and downstream string 3'-GCAGGGATC-5', and reads containing up to 7 additional G residues were counted. Their abundance was expressed as a mean percentage of the total number of reads overlapping the RNA editing site in three replicate experiments (Table 3.3). To estimate the abundance of the V and P transcripts, reads containing 0 and 0+3 G inserts (generating V mRNAs) and 2 and 2+3 G inserts (generating P mRNAs) were identified. Additionally, a third transcript encoding a third potential protein (I), the function of which has not been defined, was estimated by counting reads containing 1 and 1+3 G inserts. These numbers were compared to the total number of reads overlapping the RNA editing site (Fig 3.8).

The proportions of G insertions made by the RNA polymerase remain stable throughout infection (Table 3.3). The number of reads containing either 0 (V mRNA) or 2 G inserts (P mRNA) were the most abundant, contributing >93% of total reads, with the ratio of V to P mRNAs being approximately 2:1 (Fig 3.8). Only 1-2% of reads had an insertion of 1 G residue (I mRNA), and insertions of more than 3 G inserts were also rare (<3%). These data demonstrate the precision and stability of the RNA editing mechanism employed by PIV5 strain W3 in generating the V, P and I transcripts.

Collectively, these results demonstrate a consistent mechanism of transcription throughout infection. There were no obvious changes in the abundance of viral mRNAs, including those generated by RNA editing. Furthermore, the stability of viral mRNA generation suggests a consistency to the RNA polymerase processivity.

Table 3.3 Mean percentages of G inserts compared with the total number of reads overlapping the V/P RNA editing site.

	Percentage of G inserts							
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
6	61.3	1.3	33.2	2.1	1.7	0.5	0	0
12	61.2	1.6	33.1	1.9	1.6	0.5	0.03	0
18	60.2	1.2	34.6	2.3	1.2	0.5	0.01	0
24	63.1	0.9	32.6	1.8	1.2	0.4	0.02	0

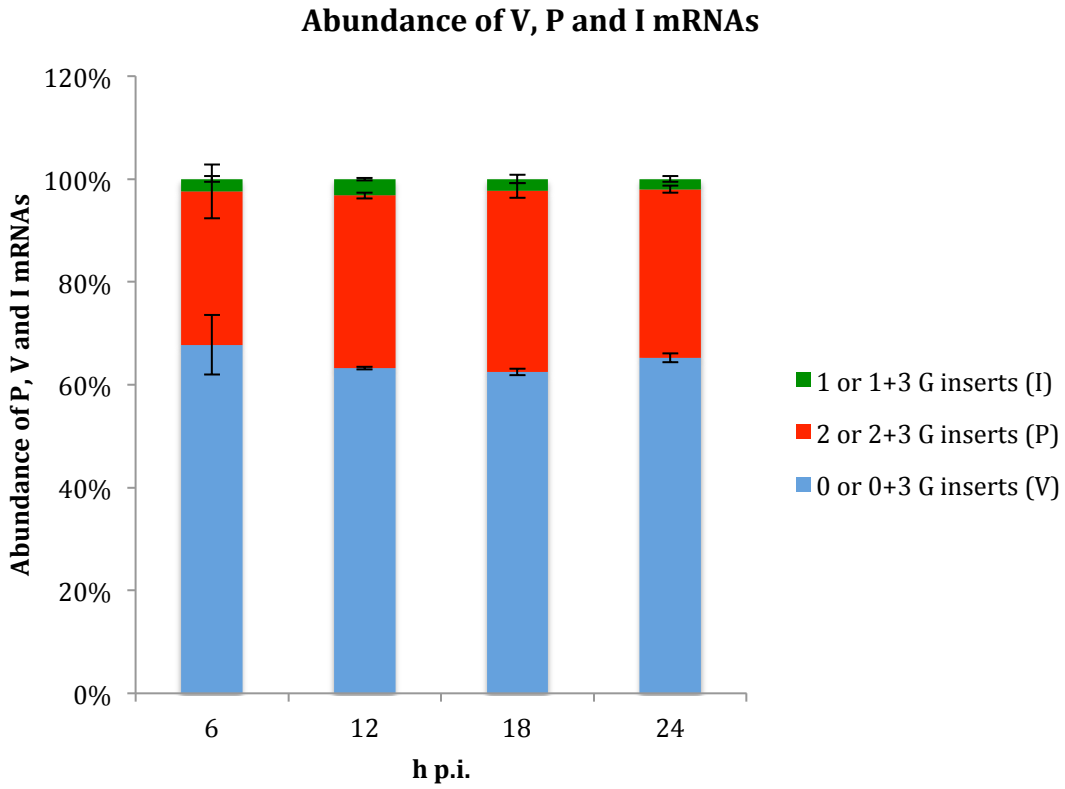


Fig 3.8. Abundance of V, P and I mRNAs generated by PIV5 strain W3 during infection. The number of reads that were generated from the V, P and I mRNAs were calculated: 0 and 0+3 G inserts (V), 2 and 2+3 G inserts (P) and 1 and 1+3 G inserts (I) from the R1 and R2 files using a 10 nt search string upstream and downstream of the RNA editing site. The bars show standard deviation values based on three experiments.

Generation of read-through mRNAs

In order to investigate the efficiency of transcription termination and re-initiation at the gene junctions, which is an additional method of control for viral gene expression, the generation of read-through mRNAs was analysed. Read-through mRNAs are generated when the RNA polymerase fails to terminate transcription at the Ge sequence (cis-acting transcription termination signal) and

continues to transcribe the Ig (cis-acting element located between Ge and Gs) and the subsequent downstream gene(s), producing a bi- or polycistronic mRNA. To investigate the generation of read-through mRNAs, the reads overlapping the Igs, from the Ge of the upstream gene to the Gs of the downstream gene, were calculated by extracting a sub-alignment from the mRNA read alignment previously generated using the established bioinformatic pipeline. This was compared to the average coverage of the gene immediately upstream of the Ig (Fig 3.9). There was a greater proportion of read-through mRNAs produced at 6 h p.i. However, some of these reads could be of antigenome origin. Such reads would affect the intergenic region at a constant level across the genome and therefore not bias the overall pattern of read-through mRNA generation, but they could account for the overall increase in the abundance of read-though mRNAs at 6 h p.i. The differences in the overall abundance of read-through mRNAs could also be influenced by their turn-over rate and stability. The overall pattern of read-through mRNA generation remained stable throughout the infection. The M-F Ig always produced a higher proportion of read-through mRNAs. The NP-V/P and V/P-M Igs produced greater proportions of read-through mRNAs compared to the downstream SH-HN and HN-L Igs.

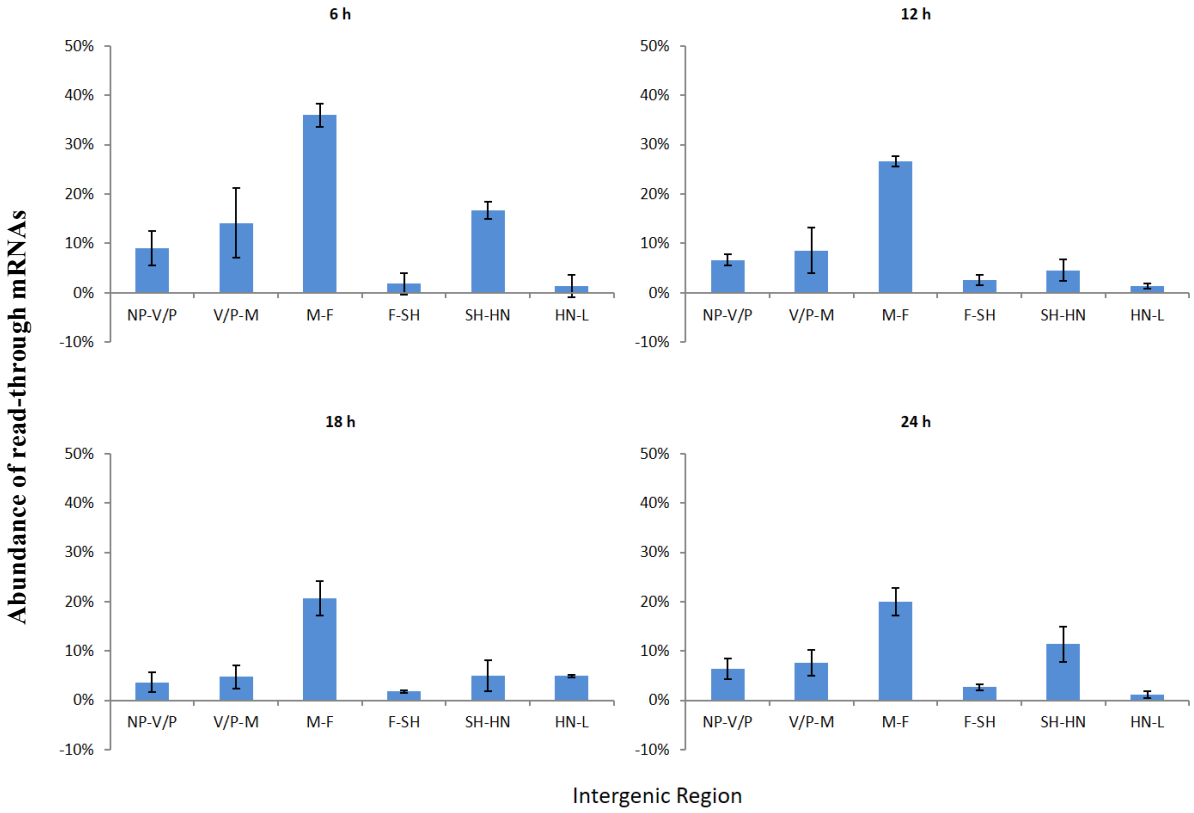


Fig 3.9. Analysis of the generation of read-through mRNAs of PIV5 strain W3. The average coverage of reads present in the intergenic regions was compared to the average coverage of reads of the gene immediately upstream. The bars show standard deviation values based on three independent experiments.

3.2.2 Analysis of viral replication

As previously demonstrated, the abundance of viral genomes gradually increases during infection (Fig 3.6). New virus genomes are synthesized via the generation of the viral antigenome, which is a faithful positive sense copy of the negative sense viral genome produced by the RNA polymerase. The RNA polymerase then binds to the TrC promoter at the 3' end of the viral antigenome

to synthesize new viral genomes. The TrC promoter is widely accepted as being a stronger promoter to promote the production of new virus genomes.

In general, viral mRNA and antigenome reads cannot be distinguished by the established bioinformatics pipeline. The production of viral antigenome was estimated by calculating the reads generated from the Le region, which lies outwith the PIV5 strain W3 gene boundaries and should not generate reads from viral mRNA (Fig 3.10A). As mentioned earlier, there is evidence that some paramyxoviruses generate a short Le transcript which has the potential of interfering with the estimation of viral antigenome production. To summarise, the method of library preparation and sequencing would not generate reads from short RNAs (the Le transcripts have been estimated to be 1-55 nts in length), and, in addition, all of the reads included in the quantification of viral antigenomes cross the NP gene junction and therefore must be of viral antigenome origin. A search string of 10 nt in the 3' Le region (positions 44-54) was used to calculate the number of reads generated from this location, which must be of viral antigenome origin. Using the Lander/Waterman equation, the number of viral antigenome reads was estimated at each time point and compared to the number of total cellular reads (Fig 3.10B).

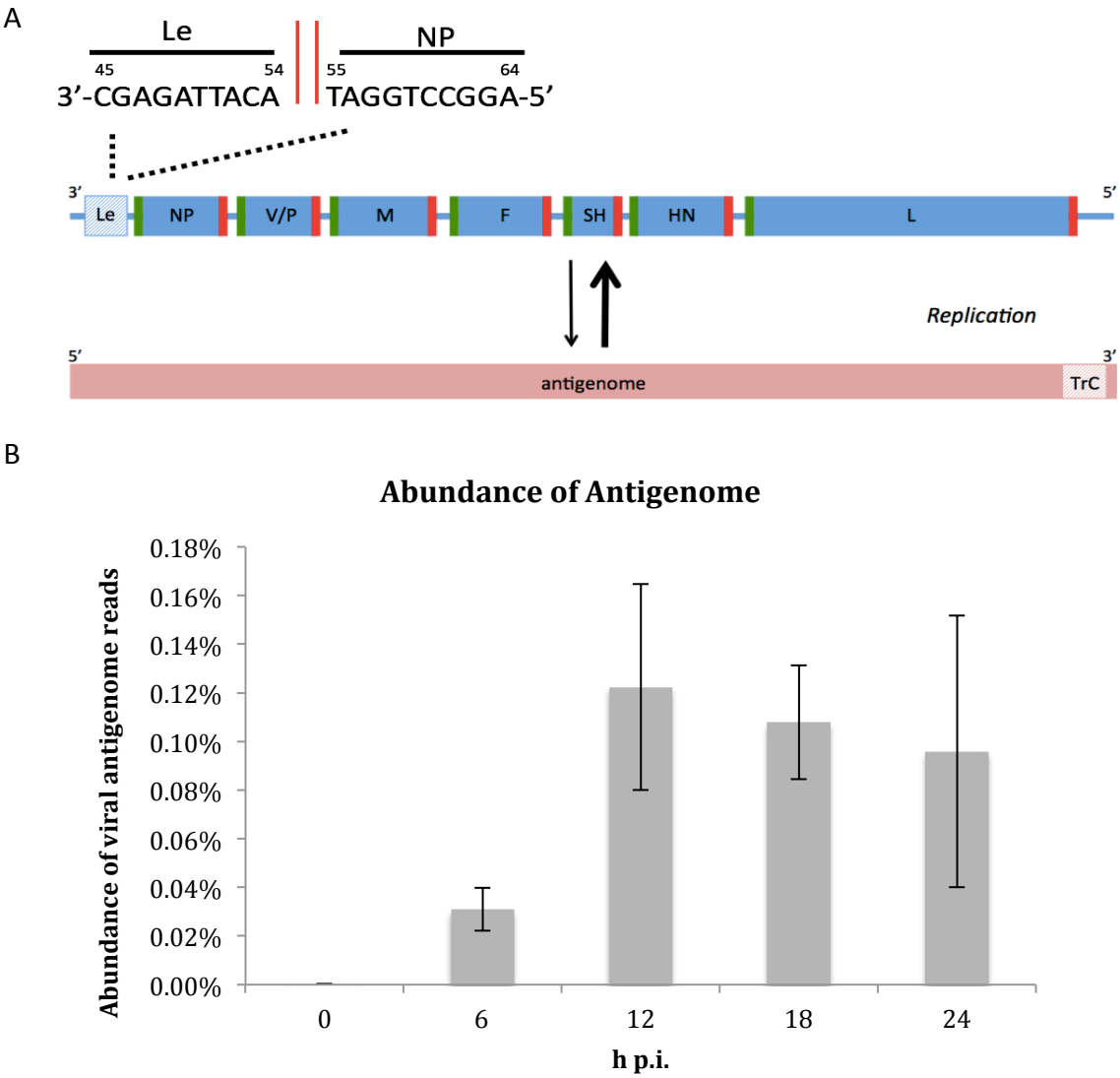


Fig 3.10. Ratio of PIV5 strain W3 virus genomes to antigenomes (A) A schematic of the PIV5 strain W3 genome, depicting the positions of the genes (blue rectangles containing names of genes, not to scale) and Gs and Ge sequences (green and red rectangles, respectively. The Le region from which the abundance of antigenomes was estimated, position 44-54 is indicated, as is the NP gene boundary shown as a double red line and 10 nt downstream of the boundary, the sequence and position of nucleotides is indicated. (B) The abundance of viral antigenome reads calculated from Le positions 44-54, compared to total cellular reads from which cytoplasmic and mitochondrial rRNA reads had been removed. The reads were directionally analysed as previously described and aligned to the PIV5 strain W3 reference sequence by BWA alignment software. The bars show standard deviation values based on three experiments.

The abundance of viral antigenomes increased between 0 and 12 h p.i. to 0.12% of total cellular reads. After this there was a slight and gradual decrease to 24 h p.i. This suggests the abundance of viral antigenomes remains at very low levels throughout the 24 h infection. However, the use of random primers during library preparation (which do not prime truly randomly), and the dependence on reads in a short region of the genome that is not represented in mRNAs, can only lead to an approximation of the proportion of viral antigenomes. Interestingly, the increase in the abundance of viral antigenomes between 0 and 6 h p.i. mirrors that of the abundance of viral mRNAs (Fig 3.6), which also peaks at 12 h p.i. This analysis suggests that a small proportion of antigenomes, and therefore potentially newly synthesized viral genomes, is being generated at very early times p.i.

The generation of viral antigenomes and newly synthesized viral genomes was investigated by ascertaining the ratio of genome to antigenome reads calculated at Le positions 44-54 (Fig 3.11). At the point of infection, 0 h p.i., viral genomes far outnumbered viral antigenomes at a ratio of 9 to 1, but by 6 h p.i. this had been reversed. Thereafter, the abundance of viral genome increased and that of the antigenome decreased. These results suggest that in the first 6 h of infection the virus is engaged in viral antigenome generation. It is commonly thought that the TrC promoter must be stronger than that of the Le promoter in order to facilitate the production of more viral genomes than antigenomes. These results suggest three mechanisms. Firstly, that the RNA polymerase undergoes modification to allow a switch between viral antigenome

production earlier in infection and viral genome generation later in infection. Secondly, there are other viral or host cell proteins that act to change the strength of the promoter. Lastly, there could be several RNA polymerase molecules copying a single viral genome at the same time, producing several viral antigenome molecules from a single template, which could account for the antigenome generation at early times p.i. and the switch in the proportion of genome to antigenomes. In addition, the generation of viral antigenomes and new viral genomes requires the production of the viral NP protein for RNA encapsidation. Previous research has indicated that viral replication cannot occur until a sufficient level of NP has been reached in the cytoplasm to allow encapsidation, suggesting that NP may have a critical role to play in the switch between transcription and replication. The abundance of viral mRNAs at 6 h is low (0.41% of total cellular reads; Fig 3.6), indicating that the level of NP in the cytoplasm available for encapsidating antigenomes and new viral genomes may also be low.

To investigate the synthesis and relative abundance of NP, A549 cells were infected at an MOI of 10 pfu/cell with PIV5 strain W3 and metabolically labelled with ^{35}S-methionine at 3, 6, 9, 12 and 18 h p.i. (Fig 3.12A). In agreement with NGS data on the relative abundance of viral mRNAs (Fig 3.6), maximum rates of NP synthesis were observed at 12-18 h p.i.. However, low levels of NP synthesis were also be observed at 6 h p.i. These results were supported by immunoblot analysis showing that relatively low levels of NP accumulated by 6 h p.i. (Fig 3.12B). These results were unexpected, in that viral antigenome synthesis (which requires encapsidation) had started by 6 h p.i., and yet

maximum rates of mRNA production and NP synthesis were not reached until 12 h p.i., and viral genome synthesis increased until 24 h p.i..

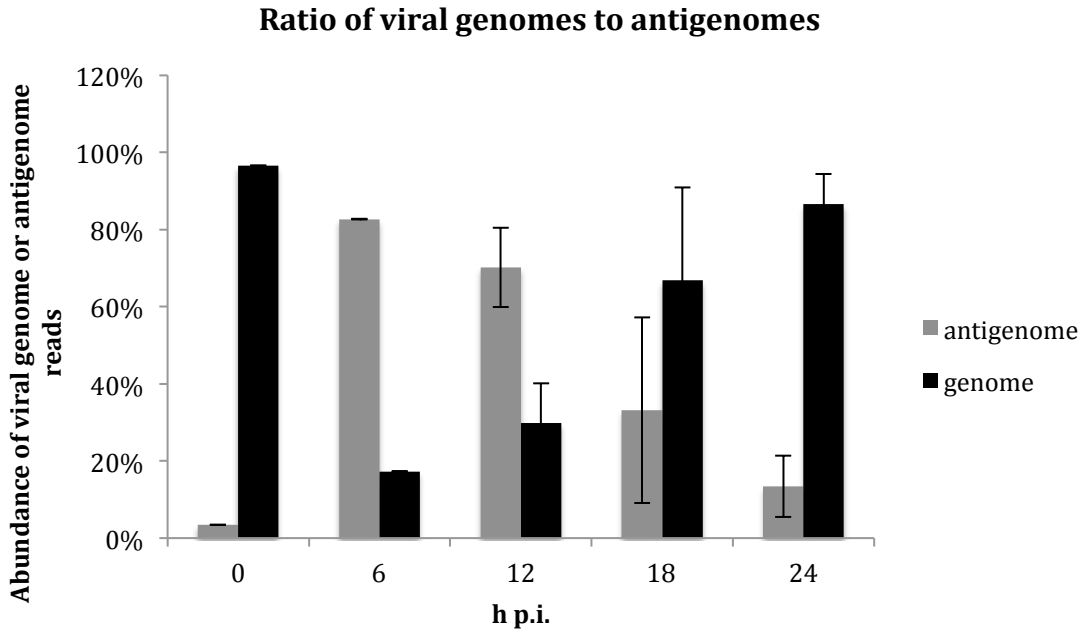


Fig 3.11. Proportions of viral genome and viral antigenome reads in positions 45-54 of the Le region, expressed as a percentage of the total. The data was directionally analysed using the bioinformatic pipeline previously described. The viral genome and mRNA/antigenome reads were then aligned to the PIV5 strain W3 reference sequence by BWA software. The bars show standard deviation values based on three independent experiments.

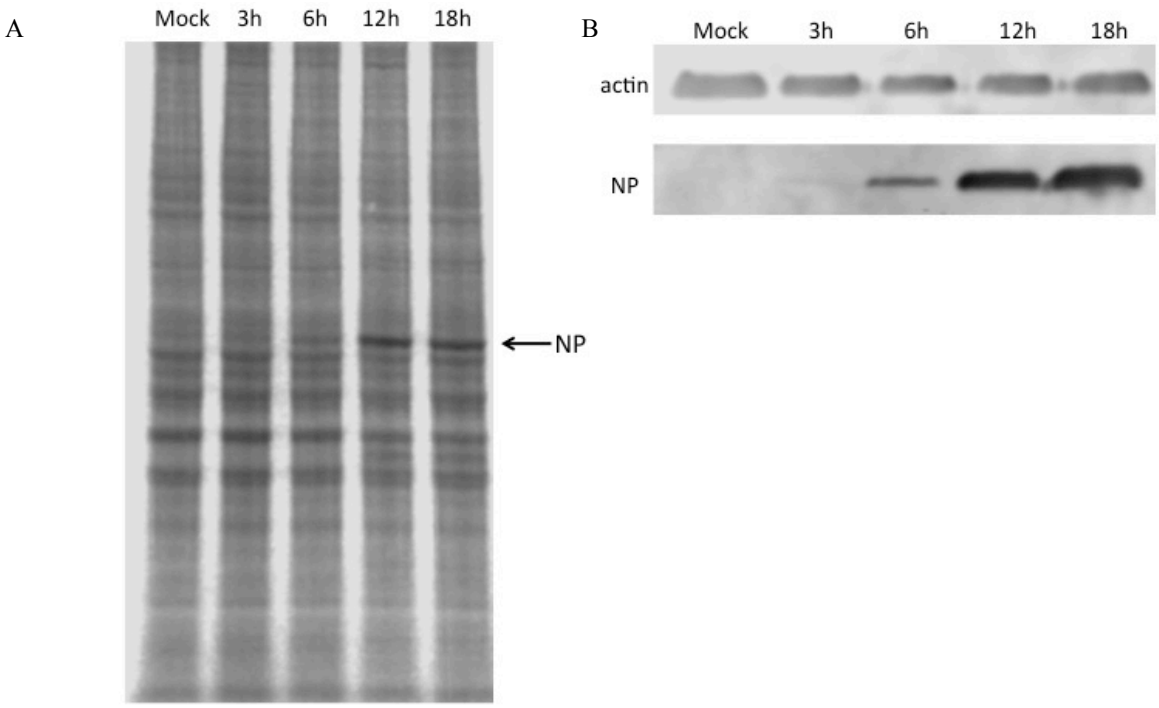
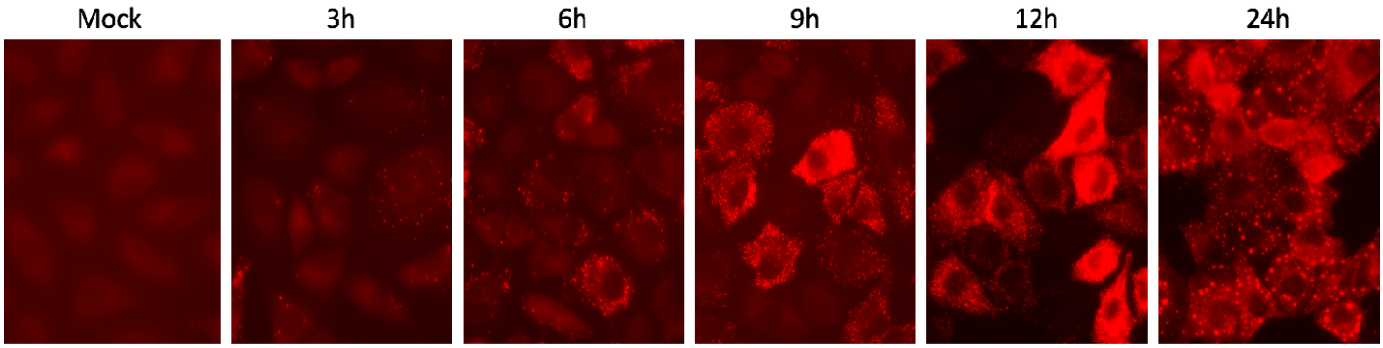


Fig 3.12. Production of viral NP at early time points during infection. (A) A549 cells were infected with PIV5 strain W3 at an MOI of 10 pfu/cell. The infected cells were metabolically labelled for 1 h using ^{35}S-methionine. The polypeptides present in the total cell extracts were separated by SDS-PAGE on a 4-12% polyacrylamide gel, and the labelled polypeptides were detected using a phosphoimager. (B) A549 cells were infected with PIV5 strain W3 at an MOI of 10 pfu/cell. The polypeptides present in the total cell extracts were separated by SDS-PAGE on a 4-12% polyacrylamide gel. Anti-NP and anti-actin antibodies were used to detect the polypeptides and imaged using LiCor software.

The analysis described above showed that there was an increase in the levels of viral genome and antigenomes early in infection, and that by 6 h p.i. the amount of viral antigenomes exceeded that of genomes. It was also clear that viral replication had begun by 6 h p.i., at a time when there were only very low levels of NP in infected cells. As a result, it was hypothesised that it was the local concentrations of NP surrounding viral genomes, rather than overall levels in the cytoplasm, that controlled the switch from transcription to replication. Local foci of virus replication, also known as inclusion or Negri bodies, have been shown to form in some Mononegavirales virus infections such as Rabies (Nikolic et al 2017). To investigate whether the formation of these foci plays a role in PIV5 strain W3 viral RNA synthesis, A549 cells were infected with PIV5 strain W3 at an MOI of 1 pfu/cell, and infected cells were fixed for immunostaining at 3, 6, 9, 12 and 24 h p.i. An anti-NP antibody was used to detect the presence of the viral NP protein in the cytoplasm by immunofluorescence (Fig 3.13). At early times in infection (3 and 6 h p.i.), NP was detected at a restricted number of cytoplasmic foci within infected cells (Fig 3.13B). By 9 h p.i., viral NP and the foci had significantly increased in number and were dispersed throughout the cell. By 24h p.i., the foci had aggregated to form larger inclusion bodies. These results suggest that the virus forms local foci as centres of virus transcription and replication, where the level of NP may have a critical role to play in the control of RNA synthesis. During the initial stages of infection, viral transcription and replication are localised at these sites. At later time points, by 9 h p.i., the foci had grown in number and dispersed throughout the cytoplasm. This expansion correlates with the significant increase in viral mRNA levels between 6 and 12 h p.i. (Fig 3.6). Finally, by 24 h

p.i. the foci had aggregated to form larger foci, the formation of which may play a role in the establishment of persistence, creating stable environments for the maintenance of viral genomes that require little or no viral transcription (see Chapter 6 for further information on persistent infections).

A



B

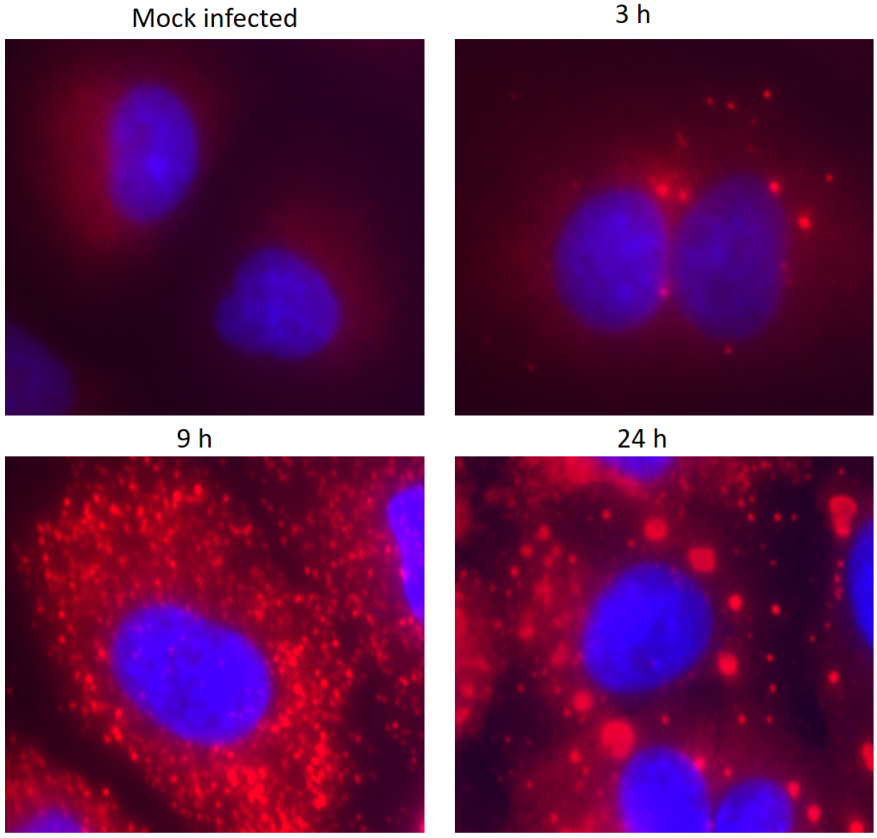


Fig 3.13. Immunofluorescence staining of viral NP in the cytoplasm of A549 cells infected with PIV5 strain W3. A) A549 cells were grown on coverslips and infected with PIV5 W3 at an MOI of 1 pfu/cell. The cells were fixed and immunostained using anti-NP antibodies and DAPI (not included in panel A) at 3, 6, 9, 12 and 24 h p.i. and visualized using an Evos microscope. (B) Images taken from those described above in panel A of time points 3, 9 and 24 h p.i. by zooming in and including DAPI staining nucleic acid

Chapter four

Analysis of the transcription and replication of PIV2, MuV and PIV3

There are numerous clinically important paramyxoviruses, many of which can infect humans, including PIV2 and MuV (rubulaviruses) and PIV3 (respirovirus), which can cause anything from mild respiratory infections to more serious illnesses, including bronchitis and mumps. Understanding the infection dynamics that underlie the diseases caused by these pathogens is important for the identification of new therapeutic targets.

As discussed in the Introduction (section 1.3), virus transcription and replication is widely accepted to be directed by cis-acting elements, namely the Gs, Ig and Ge. The cis-acting elements in the rubulaviruses and respiroviruses differ significantly. Rubulaviruses (PIV5, PIV2 and MuV) contain cis-acting elements that vary in sequence and length, whereas the respiroviruses contain conserved elements that, in the case of PIV3, include Gs and Ge sequences of 3'-AGGANNAAAG-5' and 3'-ANTANNA^{5}-5' respectively and a 3'-GAA-5' trinucleotide Ig sequence.

The methods developed to study the transcription and replication of the prototypic virus PIV5 strain W3 (see chapter 3) were employed to investigate human isolates of the two rubulaviruses PIV2 strain GREER (Accession number AF533012.1) and MuV strain Enders (Accession number GU980052) and the respirovirus PIV3 (reference genome obtained by de novo assembly). A549 cells were infected with PIV2, MuV or PIV3 at an MOI of 10 pfu/cell. Total cellular RNA was extracted at 6, 12, 18 and 24 h p.i. and subjected to rRNA reduction and library preparation followed by Illumina sequencing. As the efficiency of cytoplasmic and mitochondrial rRNA reduction can vary among samples, the remaining cytoplasmic and mitochondrial rRNA reads were

removed from the total cellular reads by Bowtie2 alignment. The data were then directionally analysed using the established bioinformatic pipeline, as previously described (chapter 3). This analysis was conducted as three independent biological experiments, from which the standard deviations were calculated and are represented as error bars where appropriate. The exception is PIV3 6 h p.i. for which due to the time constraints of this project, a single sample was obtained but replicates are currently being generated to ensure a more representative result. The results were compared to those obtained from the analysis of PIV5 strain W3 transcription and replication (chapter 3.2). All subsequent analyses were conducted on the data sets described above.

4.0 Analysis of PIV2 and MuV (Rubulavirus) and PIV3 (Respirovirus) transcription and replication

To estimate the kinetics of PIV2, MuV and PIV3 viral genome and mRNA production, the numbers of viral genome and mRNA reads, obtained by aligning the total cellular reads to the appropriate reference sequence using the BWA alignment software, were compared to the number of total cellular reads (Fig 4.1). A table showing the numbers of total cellular, viral genome and mRNA reads for each virus is provided in Appendix I.

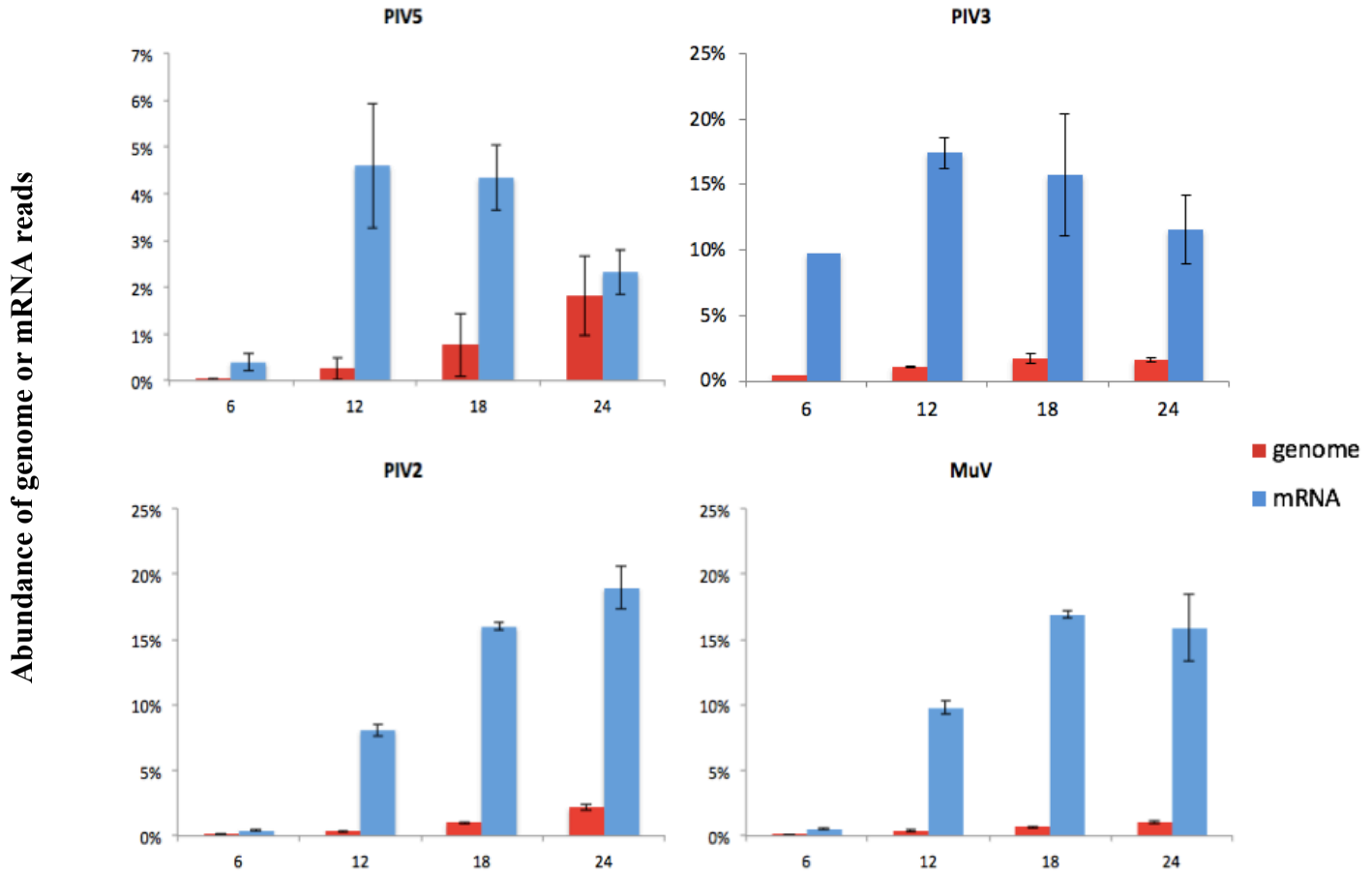


Fig 4.1. The abundance of PIV5 strain W3, PIV2, MuV and PIV3 viral genomes and mRNA reads at 6, 12, 18 and 24h p.i. The data was directionally analysed. The abundance of viral genomes and mRNAs were compared to total cellular reads from which the remaining rRNA reads had been removed. The analysis was conducted in triplicate error bars represent the standard deviation except PIV3 6 h p.i. for which only a single sample was obtained.

The paramyxoviruses display diverse transcriptional profiles. As previously discussed, the abundance of PIV5 strain W3 (rubulavirus) mRNAs increased greatly in the first 12 h of infection, peaking at 12 h p.i. when it comprised 4.59% of total cellular reads. After this, the abundance of viral mRNAs decreased until 24 h p.i. when it had halved to 2.3% of total cellular reads. Interestingly, PIV2 and MuV (also rubulaviruses) displayed a significant increase in mRNA generation between 6 and 12 h p.i., from 0.4 and 0.5% of total cellular reads to 8 and 9.8%, respectively. In comparison, PIV3 exhibited a greater abundance of mRNAs by 6 h, when 9% of total cellular reads were of viral mRNA origin, suggesting a significant increase in transcriptional activity than that observed with the rubulaviruses.

The abundance of mRNAs for PIV2 and MuV peaked at 24 and 18 h, respectively, generating 18.96% of total cellular reads for PIV2 and 16.9% for MuV. This reflects an almost four-fold difference in peak mRNA abundance between PIV5 strain W3 and PIV2 and MuV, and is perhaps a cause of the cytopathogenicity of the infection. PIV2 and MuV are known to kill almost 100% of infected cells by 48 h p.i. (observational data). Conversely, PIV5 strain W3 establishes a persistent infection in cells, with the majority surviving up to 96 h p.i. (see chapter 5 for more information). There was a slight decrease in the abundance of MuV mRNAs between the peak at 18 h p.i. and 6 h later at 24 h p.i. A decrease in the abundance of viral mRNAs was also observed for PIV3 (respirovirus) with the peak of 17.4% at 12 h p.i. declining to 11.5% at 24 h p.i. Interestingly, cells infected with PIV3 also die by 48 h p.i. (observational data), again suggesting a possible correlation between high levels of viral mRNA generation and cytopathogenicity.

The abundance of viral genomes displayed the same pattern in PIV5 strain W3, PIV2, MuV and PIV3, with a gradual increase in the abundance of viral genomes between 6 and 24 h p.i. The abundance of viral genomes at the early time point of 6 h p.i. may indicate the generation of newly synthesised genomes or, alternatively, the presence of input genomes. Unfortunately, a 0 h p.i. time point was not included for PIV2, MuV and PIV3 as it was for PIV5 strain W3 where there was a clear increase in the abundance of virus genomes from the input virus at 0 h p.i. to 6 h pi. (chapter 3 Fig 3.6). As was previously discussed for PIV5 strain W3 (chapter 3) if new viral genomes were being generated at early time points, this would require the production of the viral antigenome and sufficient levels of NP in the cytoplasm to allow encapsidation. A more detailed analysis of viral replication is available in section 4.3.

4.1 Analysis of PIV2, MuV and PIV3 Transcription

As described for PIV5 strain W3, the abundance of viral mRNAs was analysed to ascertain any changes in the transcriptional gradient indicating a change in viral gene expression that occurs over time. This analysis was conducted using FPKM values, which normalize the abundance of mRNAs generated from different samples and from genes of different lengths (see Introduction section 1.6.4). FPKM values were obtained using the RSEM software (Fig 4.2).

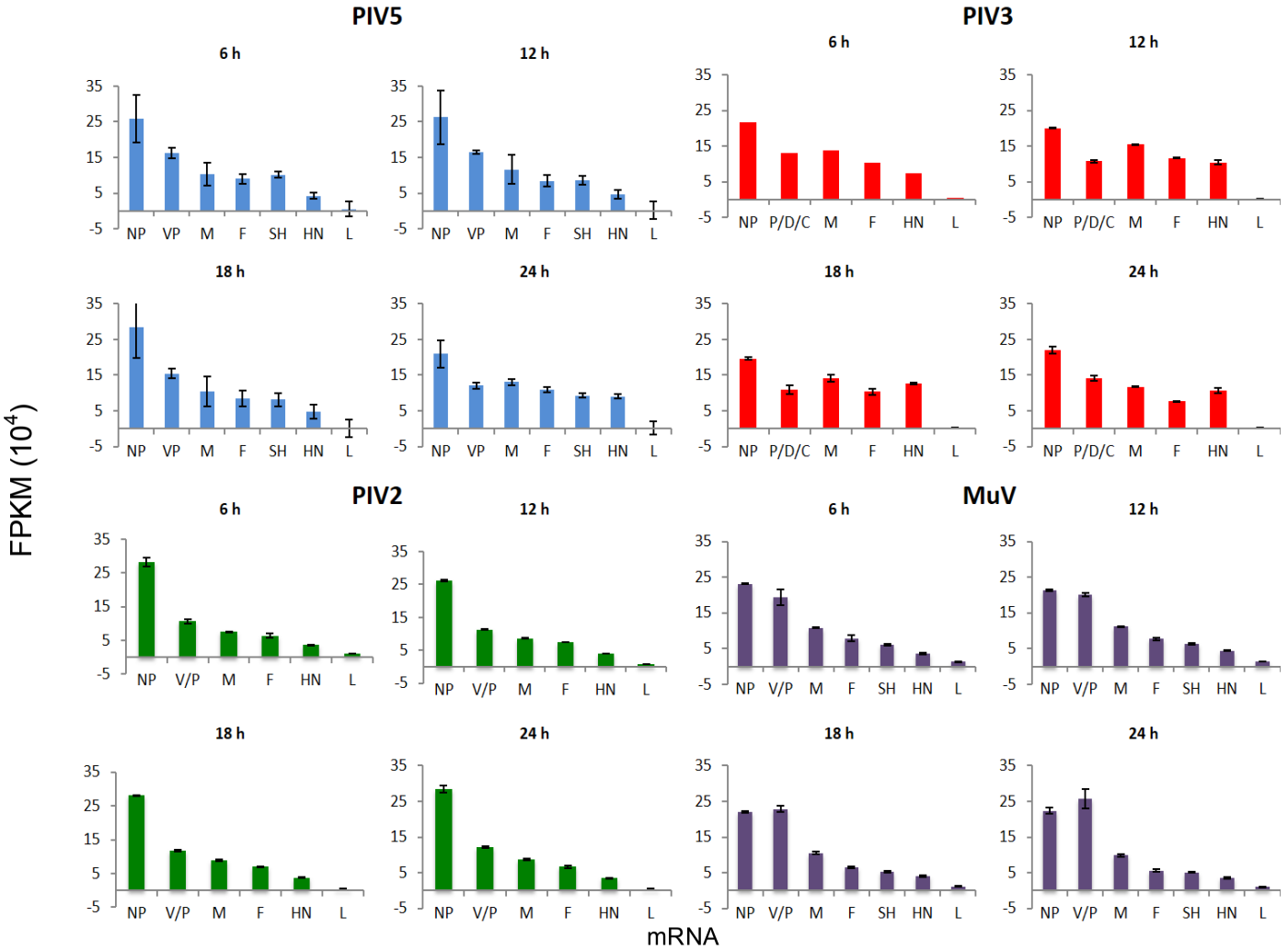


Fig 4.2. Transcription of PIV5 strain W3, PIV2, MuV and PIV3 genes at 6, 12, 18 and 24 h pi.. The data were directionally analysed using the established bioinformatic pipeline. The FPKM values on the *y*-axes were obtained using the RSEM software. The bars show standard deviation values based on three independent experiments except PIV3 6 h p.i. for which only a single sample was obtained.

The relative abundance of individual viral mRNAs remained relatively stable throughout infection. The transcription profile depicts the transcription gradient displayed by the individual viruses, which generally has a negative slope from the 3' to the 5' end of the genome. However, there are differences between the viruses in transcription profile. PIV2 demonstrated a sharp decrease in expression between NP and V/P, and the genes downstream of V/P showed a less sharp decrease and therefore a shallower transcription gradient. In contrast, MuV displayed a more equal level of expression between NP and V/P at 6 and 12 h p.i., with the abundance of V/P more obviously greater than that of NP at the later time points of 18 and 24 h p.i. This was most likely due to differences in mRNA stability rather than to changes in RNA polymerase processivity. The genes downstream displayed a clearly negative gradient. PIV3 displays a flatter transcription gradient between the P/D/C, M, F and HN. Interestingly, all four viruses show a significant decrease in the abundance of the HN and L mRNAs, more obviously observed for PIV5 strain W3 and PIV3 that exhibit decreases of more than eight-fold and 14-fold in the abundance of HN to L mRNAs, respectively, compared to less than a two-fold decrease in the abundances of mRNAs from upstream genes. The significance of any changes observed over time was analysed using a two-way anova test, all P values were shown to be not significant.

Analysis of RNA editing and the generation of accessory protein transcripts

The analysis described above did not allow the mRNAs generated via RNA editing to be distinguished from each other. This involves the mRNAs encoding accessory proteins transcribed by insertion of additional G residues into a tract of G residues specified by the viral genome at the RNA editing site. This site was analysed to ascertain the proportions of G inserts encoding accessory proteins, which are named P and I for PIV5, PIV2 and MuV (rubulaviruses) and D and V for PIV3 (respirovirus), as these ratios cannot be estimated using the RSEM software. An additional accessory protein C is produced by respiroviruses, but this is generated via an additional open reading frame and not via RNA editing, and therefore could not be included in the analysis. The rubulaviruses produce a faithful copy of the gene (0 or 0+3 G inserts), generating the V mRNA, and insertion of 1 or 1+3 and 2 or 2+3 residues generates frameshifts that produce the I and P mRNAs, respectively. PIV3 (respirovirus) transcribes a faithful copy of the gene (0 or 0+3 G inserts) to produce the P mRNA. The insertion of 1 or 1+3 G residues generates a frameshift that produce the V mRNA. Similarly the addition of 2 or 2+3 G inserts produces the D mRNA. The reads overlapping the RNA editing site were identified from the relevant R1 and R2 read files by using a 10 nt search string upstream and downstream of the RNA editing site, and the number of reads containing up to 15 G inserts was calculated. As with PIV5 strain W3 7 G inserts was identified as the upper limit for these data sets. The mean number of G inserts produced from each replicate is shown in Table 4.1.

The pattern of G insertions is remarkably different among PIV2, MuV and PIV3. The pattern for PIV2 (rubulavirus) was similar to that of PIV5 strain W3, and displayed a ratio of 2:1 for 0 or 0+3 and 2 or 2+3 G inserts, which generate the V and P mRNAs, respectively, and together accounted for approximately 98% reads overlapping the RNA editing site. The I mRNA is generated by 1 or 1+3 G inserts and accounted for less than 2% of reads. In contrast, MuV (rubulavirus), displayed an approximately equivalent number of mRNAs with either 0 and 2 G inserts, the former being only 5% greater than the latter. Additionally, the proportion of 1 or 1+3 G inserts generating the I mRNA was 5%, compared to <2% in PIV2. MuV (rubulavirus) transcripts also produced a greater number of 3 and 4 G inserts, amounting to approximately 8-9% of reads overlapping the RNA editing site and generating the V and I transcripts, respectively. PIV3 (respirovirus) produced 0, 1 and 2 or G inserts at a ratio of approximately 5:3:1, generating the P, V and D mRNAs, respectively. These account for 96-99% of reads overlapping the RNA editing site.

Table 4.1. Total reads containing G inserts present in reads overlapping the RNA editing site in PIV2, MuV and PIV3

		Number of G Inserts							
		0	1	2	3	4	5	6	7
PIV2	6	73.6%	0.2%	25.3%	0.8%	0.0%	0.0%	0.0%	0.0%
	12	75.8%	1.2%	22.0%	0.9%	0.1%	0.0%	0.0%	0.0%
	18	75.9%	1.1%	21.9%	1.0%	0.1%	0.0%	0.0%	0.0%
	24	76.8%	1.1%	21.0%	0.9%	0.1%	0.0%	0.0%	0.0%
MuV	6	40.7%	5.9%	43.6%	4.8%	3.8%	1.0%	0.2%	0.1%
	12	46.1%	5.1%	38.8%	5.5%	3.9%	0.5%	0.0%	0.0%
	18	47.0%	4.4%	38.5%	5.8%	3.6%	0.7%	0.1%	0.0%
	24	47.6%	5.1%	39.0%	4.8%	3.0%	0.4%	0.1%	0.0%
PIV3	6	59.5%	27.1%	9.2%	3.4%	0.8%	0%	0%	0%
	12	50.7%	33.8%	11.3%	3.1%	1.1%	0%	0%	0%
	18	52.7%	31.0%	11.9%	3.2%	1.1%	0%	0%	0%
	24	54.7%	29.5%	11.8%	3.2%	0.8%	0%	0%	0%

The viruses displayed different proportions of G inserts. However, these proportions remained stable throughout infection for each virus. To quantify the relative abundance of accessory protein transcripts, reads containing 0 or 0+3, 1 or 1+3 and 2 or 2+3 G inserts were counted and compared to the total number of reads overlapping the RNA editing site (Fig 4.3).

The generation of accessory transcripts remains stable throughout the infection. The viruses display different rates of generation for transcripts produced via RNA editing. PIV5, PIV2 and MuV (rubulaviruses) generated mRNAs at a ratio for P:V of 2:1, 3:1 and 1.5:1, respectively. The I mRNA (1 or 1+3 G inserts) was generated at levels of less than 1% for PIV5 and PIV2. However, approximately 10% of MuV transcripts overlapping the RNA editing site originated from the I mRNA, suggesting that this mRNA may have a function. PIV3 generated transcripts at a ratio of approximately 5:3:1 for the P (0 or 0+3 Gs), V (1 or 1+3 Gs) and D (2 or 2+3 Gs) mRNAs, respectively. The significance of any changes observed over time was analysed using a two-way anova test that showed no significant differences over time.

These results demonstrate that the pattern of G insertions and the generation of accessory protein transcripts remains stable throughout infection for each virus but differs between viruses. This suggests that each viral RNA polymerase operates a precision in inserting G residues at the RNA editing site.

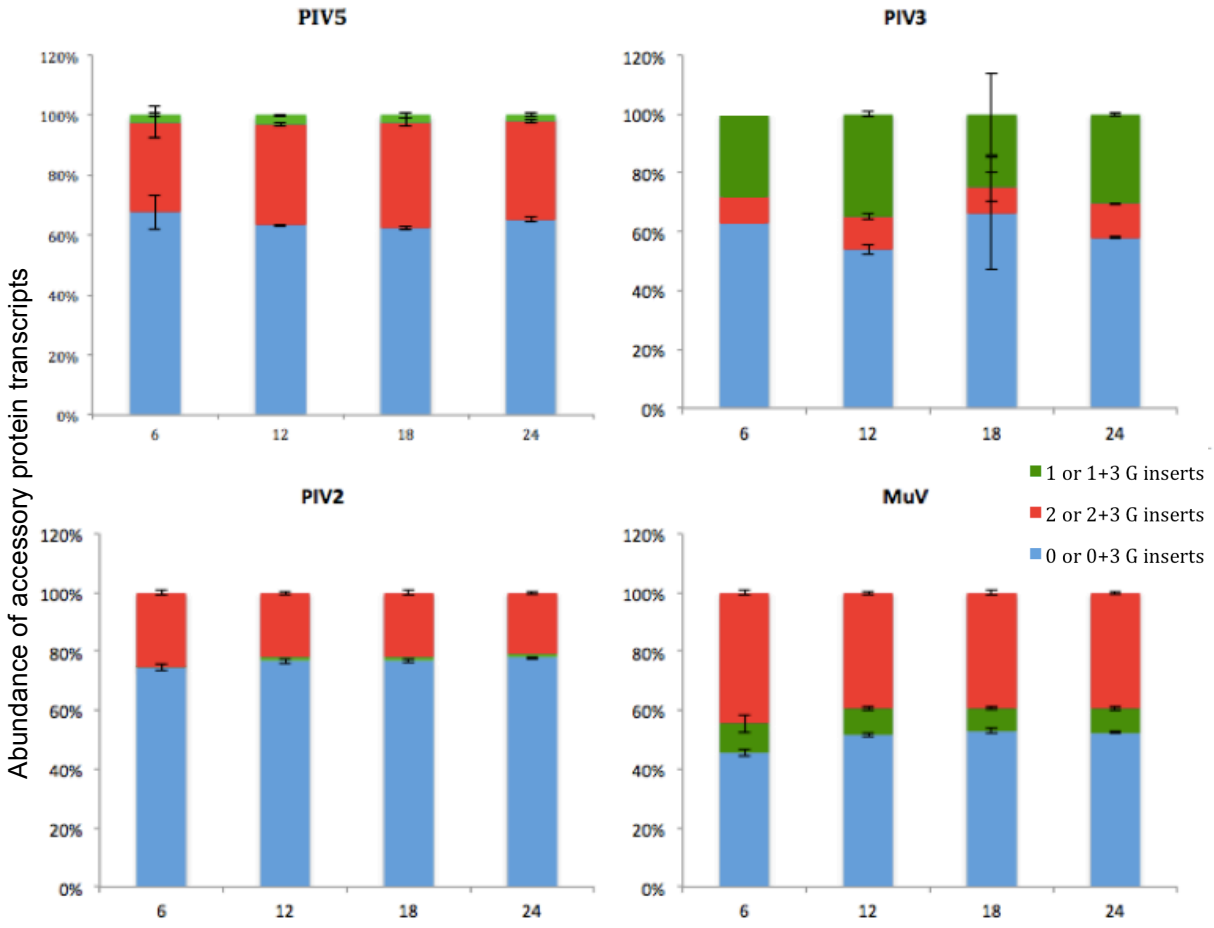


Fig 4.3 Percentage abundances of accessory proteins generated by PIV2 and MuV (rubulaviruses; V, P and I mRNAs) and PIV3 (respirovirus; P, V and D mRNAs) at 6, 12, 18 and 24 h p.i. The number of reads generated from the RNA editing site was calculated using a 10 nt search string immediately upstream and downstream of the site. The number of G inserts present in the reads overlapping the RNA editing site which generated the V, P and I mRNA transcripts was calculated, 0 and 0+3 G inserts (V or P for rubulaviruses and respiroviruses respectively), 2 and 2+3 G inserts (P or D for rubulaviruses and respiroviruses respectively) and 1 and 1+3 G inserts (I or V for rubulaviruses or respiroviruses respectively).

Generation of read-through mRNAs

The generation of read-through mRNAs has been proposed as a secondary mechanism by which paramyxoviruses can control virus gene expression. Read-through mRNAs are generated when the RNA polymerase fails to terminate transcription at the Ge sequence and continues transcribing the Ig and subsequent gene(s), producing a polycistronic mRNA. The generation of read-through mRNAs was analysed by counting the reads overlapping the Ig regions by extracting sub-alignments from the previously discussed mRNA read BWA alignment and comparing these to the average coverage of the gene immediately upstream (Fig 4.4). Overall read-through mRNA abundance can be influenced by the presence of viral antigenomes, as reads from mRNAs and antigenomes cannot be separated. However, viral antigenomes were calculated as being present at a very low abundance (see chapter 3.1), which would not significantly affect read-through analysis. Furthermore, viral antigenome reads would display a relatively consistent coverage throughout the alignment and therefore should affect the quantification of read-through mRNAs at an equivalent level at each time point. The overall pattern of read-through mRNA generation remained relatively stable throughout the various infections (Fig 4.4). PIV2 and PIV3 displayed a pattern of transcription termination and reinitiation similar to that of PIV5 strain W3, with the M-F Ig consistently generating a greater proportion of read-through mRNAs. However, the abundance of read-though mRNAs generation is different, approximately 20-36% in PIV5 W3 indicating that over two-thirds of transcripts are terminated at the Ge signal. Whilst approximately >97% of M-F transcripts are read-throughs in PIV3 and PIV2 infections, displaying a transcription termination rate <3%. The generation

of M-F read-through mRNAs increased slightly during PIV3 infection, from 73% at 6 h p.i. to 100% at 18 and 24 h p.i. This observed increase in read-through production at the M-F Ig is supported by numerous studies (see Introduction 1.3.1). The generation of read-through mRNAs at the NP-P and P-M Igs of PIV5, PIV2 and MuV (rubulaviruses) and PIV3 (respirovirus) showed transcription termination rates of >88%. The transcription termination rate at the PIV2 F-HN Ig was lower, with over a third of transcripts being read-through. There was also a slightly higher rate of read-through mRNA generation at the SH-HN Ig in PIV5 strain W3 infection.

MuV demonstrates a different pattern of read-through mRNA generation, with the highest proportion of read-through detected at the F-SH Ig, where 100% of transcripts are dicistronic. Indeed, previous studies have demonstrated that there are no monocistronic SH mRNAs produced by MuV Enders strain (Takeuchi et al., 1991). The rate of transcription termination at the MuV M-F Ig is similar to that of PIV5 strain W3, where approximately two-thirds of transcripts are terminated. MuV also showed a greater abundance of read-through mRNAs at the SH-HN Ig, where a third of transcripts are read-through.

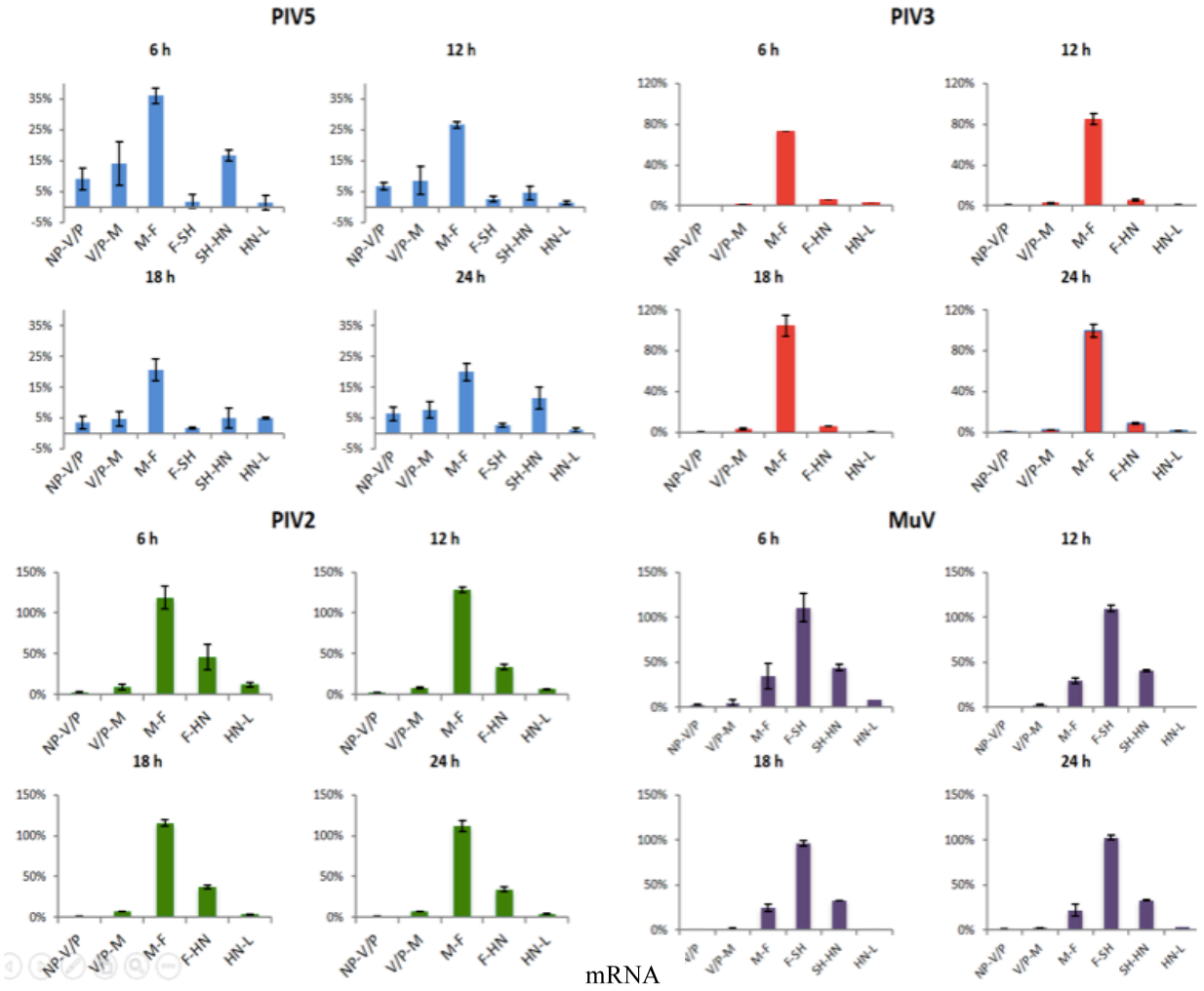


Fig 4.4. Analysis of the generation of read-through mRNAs for PIV5 strain W3, PIV2, MuV and PIV3 at the intergenic regions compared to the average coverage of the gene immediately upstream. The bars show standard deviation values based on three independent experiments except PIV3 6 h p.i. for which only a single sample was obtained.

4.1 Analysis of the Kinetics of Viral Replication

As demonstrated previously, the abundance of PIV5 strain W3, PIV2, MuV and PIV3 viral genomes increased during infection (Fig 4.1). It is notable that a three-fold increase in the abundance of viral genome occurred between 6 and 12 h p.i. for PIV2, MuV and PIV3, whereas PIV5 strain W3 displayed a ten-fold increase between 6 and 12 h p.i. This suggests that viral replication occurs at early times during infection, concurrent with significant increases in mRNA abundance, which occurred between 6 and 12 h p.i. for PIV5 strain W3, PIV2 and MuV. Therefore, there is not a sharp demarcation in timing between viral transcription and replication. Although the amount of viral genomes at 6 h p.i. suggests that viral replication may have occurred between 0 and 6 h p.i., this was not confirmed directly as a 0 h p.i. for PIV2, MuV and PIV3 was not obtained. Replication was shown to occur between 0 and 6 h p.i. for PIV5 strain W3 (chapter 3).

As discussed previously, the viral mRNA and antigenome reads cannot be distinguished from each other by the established bioinformatics pipeline as they are both positive-sense. The production of antigenome was estimated for PIV5 strain W3 (see chapter 3.0) by calculating read coverage in the Le region, which is not transcribed into mRNA. However, coverage of genome and antigenome reads in this region of the PIV2, MuV and PIV3 genome was low at early time points and therefore introduced too much variation to allow confident estimates of the abundance of antigenome or the ratio of genomes to antigenomes. Despite the lack of coverage at the 3' end of the genome, reads were detected

that cross the Le region, the gene boundary and the NP gene, indicating the presence of antigenomes at early times p.i.

If viral replication does occur at very early times p.i., it would require sufficient amounts of NP to encapsidate the newly synthesized genomes and antigenomes, prior to maximum rates of viral mRNA synthesis. To investigate the generation of NP at early times, ^{35}S-methionine labelling was conducted. A549 cells were infected with PIV5 strain W3, PIV2, MuV or PIV3 at an MOI of 10 pfu/cell were metabolically labelled for 1 h with ^{35}S-methionine at 3, 6, 12, 18 and 24 h p.i. (Fig 4.5). The production of NP was observed at 12 h p.i. for PIV2, MuV and PIV3. NP was detected at low levels earlier during infection with PIV5 strain W3, with a band visible at 6 h p.i.. It should be noted that the sensitivity of the assay may not allow a detection of very low levels of proteins including NP therefore the absence of a band at early times p.i. does not correlate with absence of the protein.

The widely accepted model for paramyxovirus replication requires NP to be present at sufficient levels to allow encapsidation of the antigenome and new viral genomes, and the level of NP in the cytoplasm is thought to be a determining factor in the switch between viral transcription and replication. However, peak transcription and translation of NP did not occur until 12 – 24h p.i. during PIV strain W3, PIV2, MuV and PIV3 infection (Fig 4.1). Despite this, there was a clear increase in the abundance of viral genomes between 6 and 12 h p.i. (Fig 1). Therefore, either the overall cellular concentration of NP required for viral replication is be relatively low (much less than that observed at later times p.i.), or NP may localise in areas of viral replication. To investigate

the cytoplasmic distribution of NP during infection, A549 cells were infected with PIV3 and MuV at an MOI of 5 pfu/cell. The cells then were fixed and immunostained using anti-NP antibodies at 3, 6, 12 and 24 h p.i. (Fig 4.6). Unfortunately, similar studies could not be undertaken with PIV2, as there was no antibodies against NP available.

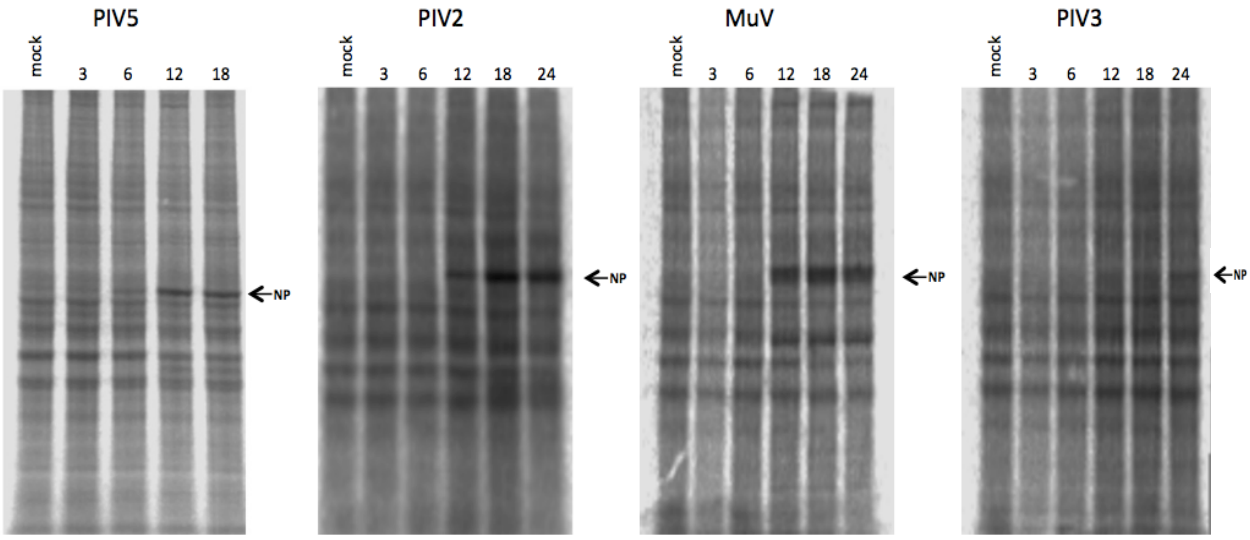


Fig 4.5. Investigating the production of viral NP at early time points during infection. (A) A549 cell monolayers were infected at an MOI of 10 with PIV5 strain W3, PIV2, MuV or PIV3. The infected cells were labelled with ^{35}S-methionine for 1 h at 3, 6, 12, 18 and 24 h p.i. The polypeptides present in the total cell extracts were separated by SDS-PAGE through a 4-12% gel, and visualized using a phosphoimager.

These preliminary results suggest that the distribution of NP during MuV and PIV3 infections is significantly different. NP displayed a distribution during MuV infection similar to that observed during PIV5 strain W3 infection (see chapter 3 Fig 3.13), with localisation into small cytoplasmic foci that may correspond to areas of viral transcription and replication. By 6 and 12 h p.i., these foci had

dispersed relatively equally throughout the cytoplasm, and larger inclusion bodies were clearly visible by 24 h p.i. During PIV3 infection, the foci were restricted to one area of the cell up to 6 h p.i., and by 12 h had dispersed, displaying a restricted grouping of foci into clear regions of the cytoplasm. By 24 h p.i., the foci were more equally dispersed throughout the cell, but, unlike the situation with PIV5 strain W3 and MuV, there was no evidence of the formation of inclusion bodies.

Collectively, these results indicate that, similar to PIV5 strain W3, viral replication is occurring at early times p.i.. The formation of foci during MuV and PIV3 infections suggest that it is the level of NP surrounding the viral genomes at these local sites of virus transcription and replication that dictates the RNA polymerase ability to enter a transcriptive or replicative mode. Interestingly, the viruses belonging to different genera exhibit distinctive patterns of distribution around the infected cell possibly indicating genera specific mechanisms of dispersal.

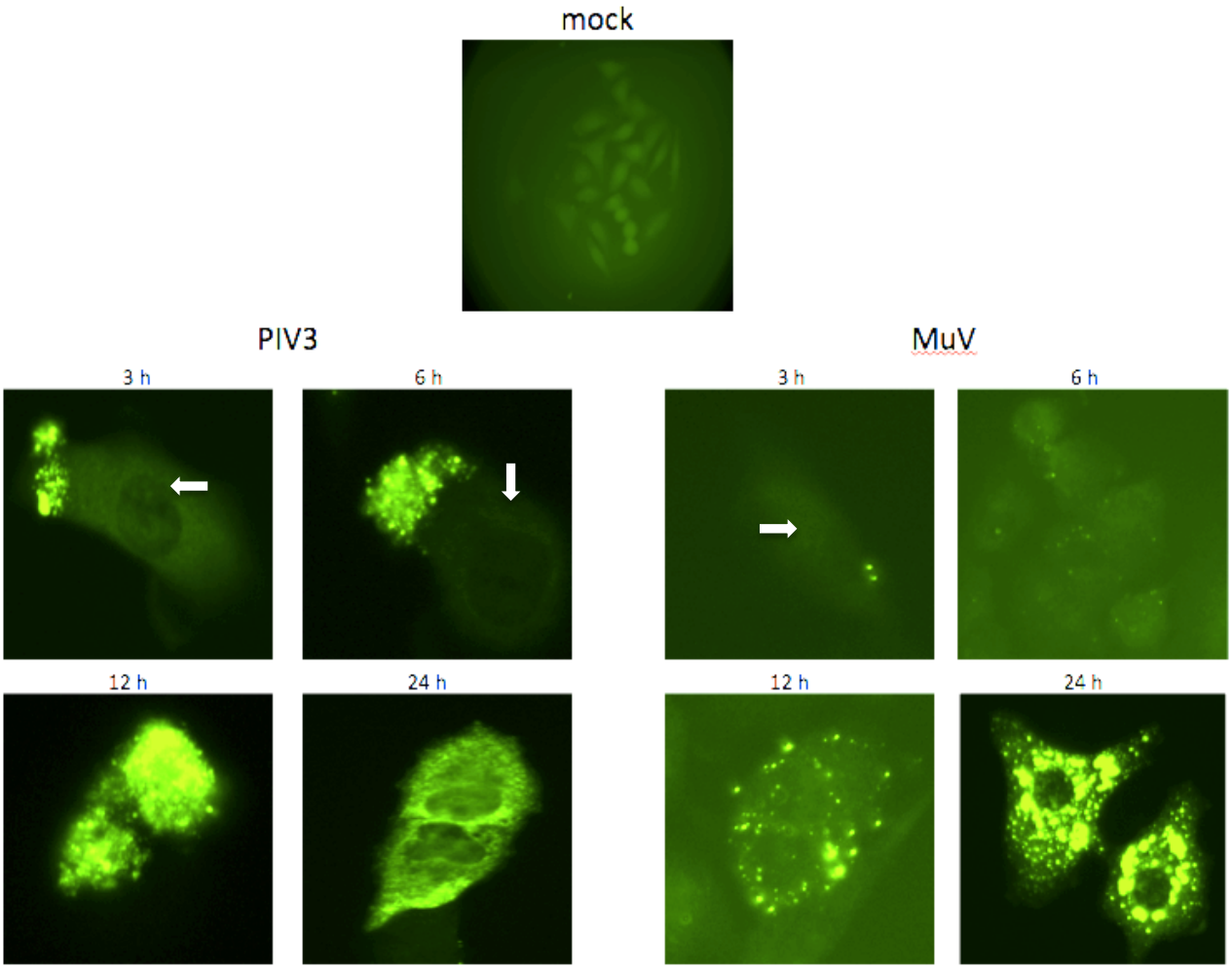


Fig 4.6. Immunofluorescence analysis of the localization of NP during infection. A549 cells monolayers were infected with PIV3 and MuV at an MOI of 2. The cells were fixed and immunostained using an anti-NP antibody at the times specified. The white arrows indicate the nucleus.

Chapter five

Analysis of persistence of PIV5 infected A549 cells

Many paramyxoviruses are capable of establishing persistent infections *in vitro* and *in vivo*. MeV infections can cause SSP in patients several years after the acute infection has been eliminated by the immune response. Defective MeV genomes have been isolated from brain lesions of people who have gone on to develop SSPE (Garg 2002, Kristensson and Norrby 1986, Randall and Griffin 2017). SeV have been well documented as being able to establish persistent infections in vitro (see Introduction section 1.6 for full details). These infections can cause chronic or reactivated diseases that can have serious clinical outcomes. However, the mechanisms by which paramyxoviruses are capable of establishing persistence are not well understood. A PIV5 model was used to investigate possible mechanisms.

5.0 Analysis of persistent infection of A549 cells by PIV5

Previous observations in Professor Randall's laboratory had shown that, following infection of A549 cells with the W3 strain of PIV5, the majority of cells survived the infection and became persistently infected. This was confirmed by phase contrast microscopy of A549 cells infected with PIV5 at an MOI of 10 pfu/cell at 24 and 96 h p.i. (Fig 5.1A). Immunostaining of the cells using anti-NP and anti-HN antibodies revealed that the majority of cells were positive for NP and HN at 24 h p.i. (Fig 5.1B). By 96 h p.i., the majority of cells were positive for NP but <50% were positive for HN. Since the half-life of HN is only 2-4h compared to the half life of NP which is >24h (Leser et al., 1996), this suggested that viral protein synthesis may be repressed in the majority of infected cells at late times p.i.

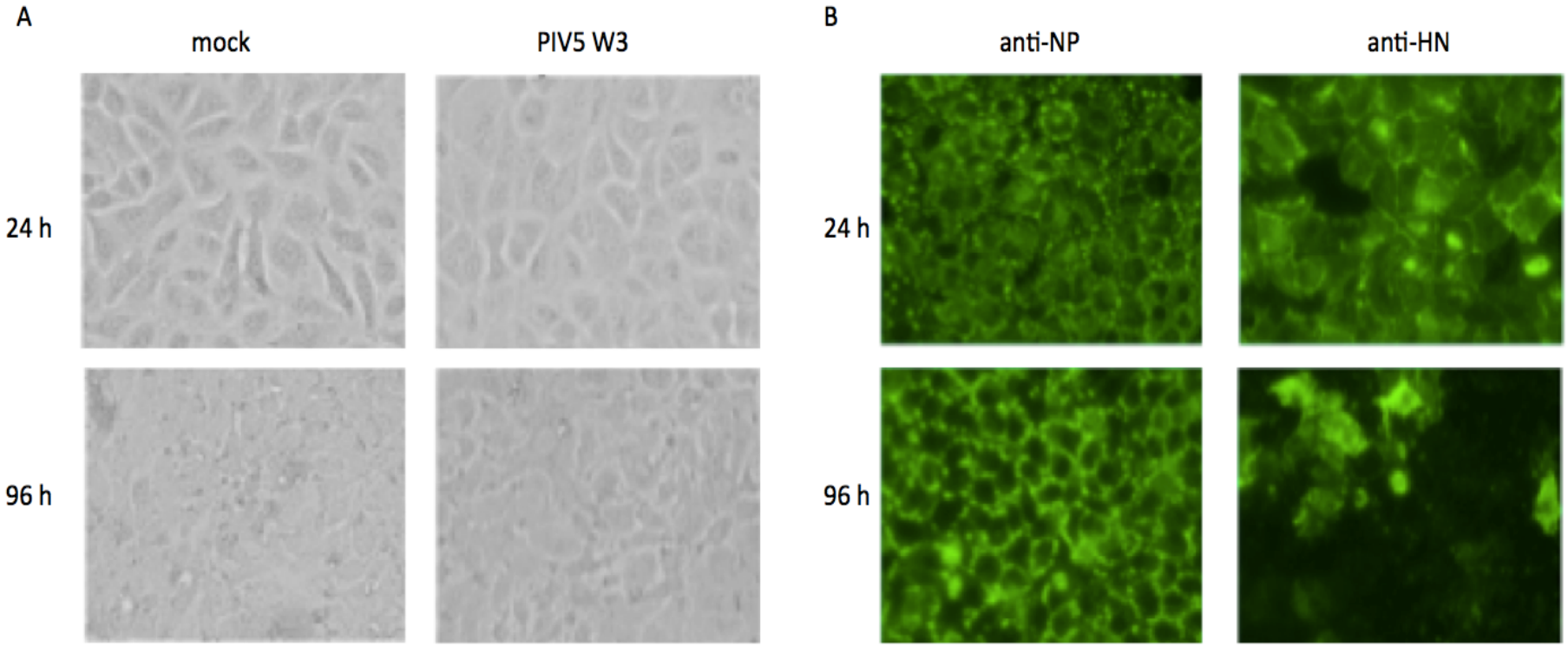


Fig 5.1. Phase contrast microscopy and immunofluorescence of A549 cells infected with PIV5 strain W3 at 24 and 96 h p.i.. (A) At 24 and 96 h p.i., cells were imaged by phase contrast microscopy. (B) A549 cells were grown on coverslips and infected with PIV5 W3 at an MOI of 10 pfu/cell. The cells were fixed and immunostained using anti-NP and anti-HN antibodies at 24 and 96 h p.i. and visualized using an Evos microscope. Conducted by Dan Young

To assess the kinetics of viral protein synthesis during infection, A549 cells were infected with PIV5 strain W3 at an MOI of 10 pfu/cell and metabolically labelled with ^{35}S-methionine at various times p.i. (Fig 5.2). Viral NP and M were clearly detected at 24 h p.i., indicating active translation of the respective mRNAs. At 48 h p.i., the production of NP and M had significantly decreased, and NP and M were no longer detected at 72 and 96 h p.i., demonstrating that viral protein synthesis is repressed at late times p.i.

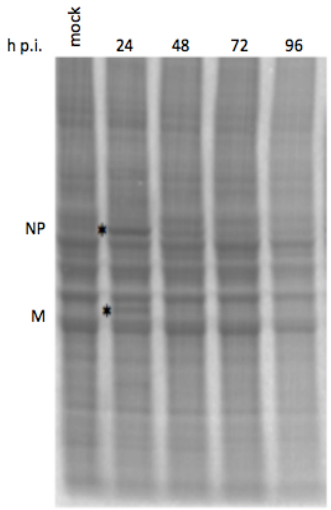


Fig 5.2. Radioactive labelling of PIV5 strain W3 infected A549 cells at 24, 48, 72 and 96 h p.i.. Monolayers of A549 cells were grown in T25 flasks and were mock-infected or infected with PIV5 strain W3 at an MOI of 10 pfu/cell. At 24, 48, 72 and 96 h p.i., the cells were metabolically labelled for 1 h using ^{35}S-methionine. The polypeptides present in total cell extracts were separated by SDS-PAGE through a 4-12% polyacrylamide gel, and the labelled polypeptides were visualized using a phosphoimager.

To determine whether the repression of viral protein synthesis observed at late times p.i. is reflected in the abundance of viral mRNAs, or whether translation of viral mRNAs is specifically repressed, A549 cells were infected with PIV5 strain W3 at an MOI of 10 pfu/cell and total RNA was extracted at 6, 12, 18, 24, 48 and 96 h p.i. and subjected to cytooplasmic and mitochondrial rRNA reduction. Sequencing libraries were then prepared and analysed on the Illumina platform. Directional data were obtained using the bioinformatic pipeline described previously (see chapter 3.0). The abundances of viral mRNA and genome reads were then compared to total cellular reads (Fig 5.3), having removed remaining nuclear and mitochondrial rRNA reads. The time points 6, 12, 18 and 24 h p.i. had also been analysed previously as part of an investigation into the acute paramyxovirus infections (see chapter 3).

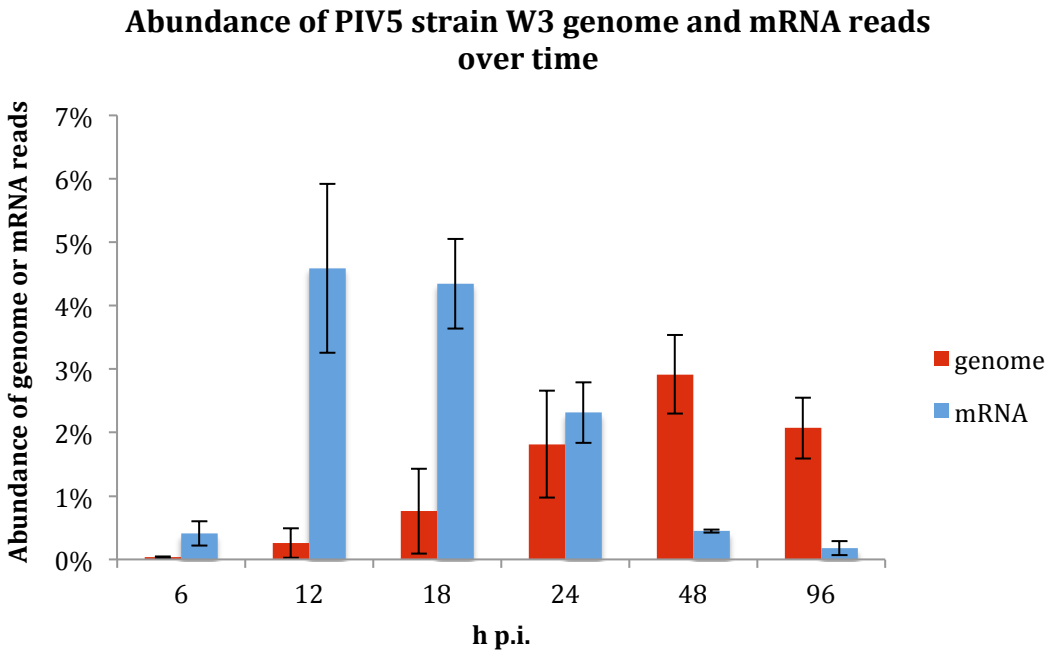


Fig 5.3. The abundance of PIV5 strain W3 viral mRNAs and viral genomes over time. The data was directionally analysed using the established bioinformatic pipeline. The number of viral mRNA and viral genome reads was compared to total cellular reads from which the remaining cytoplasmic and mitochondrial rRNA were removed at each time point. The error bars are based on three independent experiments.

As previously described (chapter 3.0), there was an initial significant increase in viral mRNA production between 6 and 12 h p.i. the abundance of viral mRNAs increasing 4-fold and peaking at 4.59% of total cellular reads. Subsequently, there was a significant decrease in the abundance of viral mRNAs to 0.18% of total cellular reads at 96 h p.i. In contrast, the abundance of viral genomes gradually increased to 48 h p.i., at which point they constituted 2.9% of total cellular reads. Thereafter, the level of viral genomes decreased slightly to 2% at 96 h p.i.. These results demonstrates that shut-off of viral protein synthesis at

late times p.i. occurred because viral transcription was repressed, and not because of loss of viral genomes or because of specific inhibition of the translation of viral mRNAs.

Viral transcription was analysed further in order to identify changes that might account for the repression of viral transcription. This utilised the data set described above in which A549 cells had been infected with PIV5 W3 at an MOI of 10 pfu/cell. Total RNA was extracted at 6, 12, 18, 24, 48 and 96 h p.i. and subjected to cytoplasmic and mitochondrial rRNA reduction and directional sequencing as described above. As the time points up to 24 h p.i. were analysed in chapter 3.0, the following analysis considers only the later time points between 24 and 96 h p.i., during which viral protein synthesis was repressed and the virus established a persistent infection.

5.0.2 Transcription and replication of PIV5 strain W3 during establishment of persistent infection

To identify whether changes in the transcription profile of PIV5 strain W3 could be responsible for the subsequent repression of viral transcription, the relative abundance of individual viral mRNAs was determined. The abundance of viral

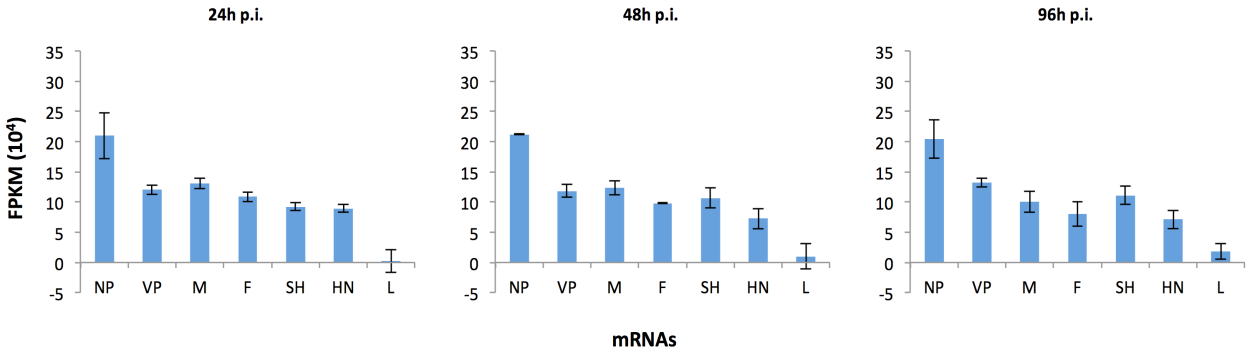


Fig 5.4. The abundances of PIV5 strain W3 mRNAs at 24, 48 and 96 h p.i. The abundances of viral mRNAs was quantified using FPKM values obtained using RSEM software. Standard deviation values are based on three independent experiments.

mRNAs was calculated using FPKM values (fragments per kilobase of transcript per million mapped reads). FPKM values normalise the abundance of transcripts generated from individual genes in order to account for differences in gene length and number of reads generated due to clustering during sequencing, thus allowing the transcription of different genes to be compared. The abundances of viral mRNAs were calculated by using RSEM software (Fig 5.4).

As described previously (see chapter 3.1 Fig 3.6) PIV5 strain W3 exhibits a decreasing transcription gradient at early times p.i., with the gene at the 3' end of the genome (NP) producing the most abundant mRNA and the gene at the 5' end of the genome (L) producing the least abundant mRNA. The levels of the V/P, M, F and HN mRNAs exhibit a slight plateaued gradient at 24 and 48 h p.i. However, this analysis was conducted on the overall abundance of mRNAs and could reflect the stability of the transcripts and not necessarily changes in viral

transcription, especially as a plateauing of these mRNAs was not observed at 96 h p.i., at which point a decreasing gradient, similar to that observed at 6, 12 and 18 h p.i., was observed. A slight increase observed in the abundance of the SH mRNA, compared with the upstream F mRNA at 48 and 96 h p.i., could also reflect mRNA stability. The relative stability of paramyxovirus viral mRNAs is not well documented and is an area of research which requires future research. Overall, the transcription profile of PIV5 strain W3 did not change significantly between early and late times p.i. suggesting that changes on the relative abundance of viral mRNAs is not responsible for the establishment of persistence. An two-way anova was conducted to test the significance of any observed changed, no significance was identified.

RNA editing and production of V, P and I transcripts

The analysis described above could not distinguish among the V, P and I mRNAs, which are generated via RNA editing and the insertion of 2 or 1 non-templated G nucleotides, respectively, in a tract of 4 G residues in the V/P gene (position 2340 in the genome). As previously described for early times p.i. (chapter 3.2), the abundance of these transcripts was estimated by identifying reads that overlap the RNA editing site, by counting reads containing up to 7 additional G residues. The abundance of each type of read was expressed as a mean percentage of the total number of reads overlapping the RNA editing site in three replicate experiments (Table 5.1). The numbers of reads in the individual replicates is provided in Appendix II. To estimate the abundance of the V, P and I transcripts, reads were identified containing 0 and 0+3 G inserts (generating V mRNAs), 1 and 1+3 G inserts (generating I mRNAs) and 2 and

2+3 G inserts (generating P mRNAs). These numbers were compared to the total number of reads overlapping the RNA editing site (Fig 5.5).

The proportions of G insertions made by the RNA polymerase remained stable throughout infection (Table 5.1). Reads containing either 0 (V mRNA) or 2 G inserts (P mRNA) were the most abundant, contributing >97% of total reads. Less than 2% of reads contained a single G residue (I mRNA). Reads containing more than 3 G inserts contributed to <3.2% of reads. These results are the same as those observed at the earlier time points. These data demonstrate stability of the RNA editing mechanism employed by PIV5 strain W3 throughout the infection, suggesting that there are no changes in the abundance of accessory protein transcripts that could account for the transcriptional “shut off”. To ensure there were no significant differences between 24 and 96 h p.i. a two-way anova with post-hoc tests was conducted generating a p value of no significance suggesting there were no significant differences to account for the establishment of persistence.

Table 5.1 Abundance of Reads Containing G Inserts that Overlap the RNA Editing Site

	Number of G inserts							
h p.i.	0	1	2	3	4	5	6	7
24	63.1%	0.9%	32.6%	1.8%	1.2%	0.4%	0.0%	0.0%
48	67.4%	0.9%	29.0%	1.9%	0.9%	0.1%	0.0%	0.0%
96	76.0%	1.1%	21.1%	0.6%	0.6%	0.5%	0.0%	0.0%

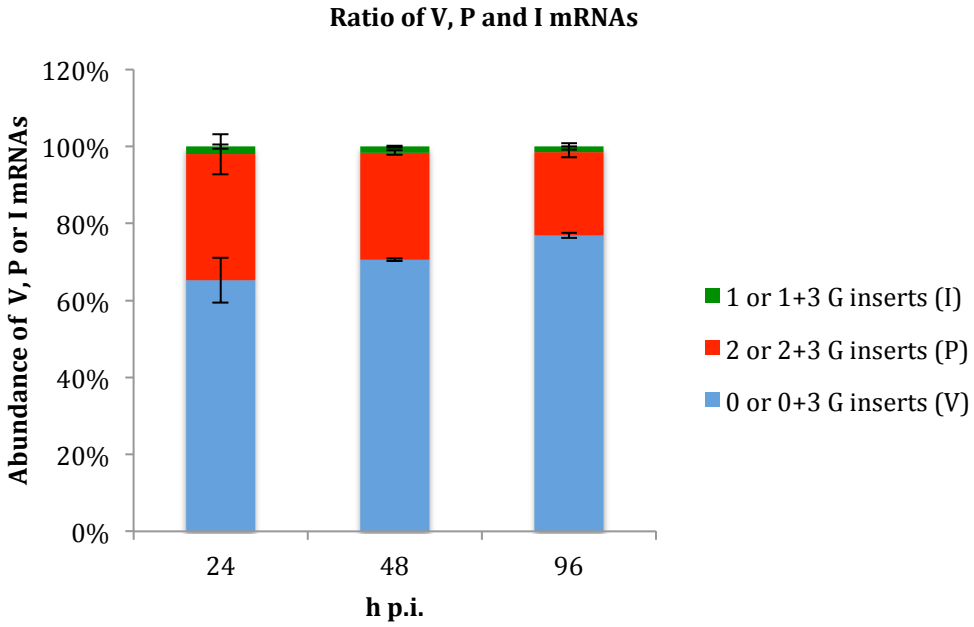


Fig 5.5. The relative abundances of V, P and I mRNAs at 24, 48 and 96 h p.i. The number of reads that were generated from the V, P and I mRNAs were calculated: 0 and 0+3 G inserts (V), 2 and 2+3 G inserts (P) and 1 and 1+3 G inserts (I). The bars show standard deviation values based on three independent experiments.

Generation of read-through mRNAs

The generation of read-through mRNAs occurs when the RNA polymerase fails to terminate transcription at the Ge site, instead continuing to transcribe the Ig and downstream gene(s) to generate a polycistronic mRNA. It has been hypothesised that generation of read-through mRNAs is a mechanism by which the virus controls gene expression. To investigate how the generation of read-through mRNAs could contribute to repression of viral transcription, the number

of reads overlapping the Ig were compared to the average coverage of the gene immediately upstream (Fig 5.6). The coverage at 96 h p.i. was to low for accurate analysis, but the overall pattern of polycistronic mRNA production is similar at 24 and 48 h p.i. There is a greater proportion of read-through mRNAs at 24 and 48 h p.i. in the M-F Ig compared to the NP-V/P, V/P-M, F-SH and HN-L Igs. A greater proportion of read-through RNAs at the M-F Ig was observed at earlier time points. It is a well characterised observation of many paramyxoviruses. The mechanism is not well understood, but it is thought that it may serve to inhibit the expression of the F protein or allow the RNA polymerase greater access to the downstream genes SH, HN and L. The read through at NP/V/P, V/P-M and SH-HN is relatively equal. Unfortunately, at 48 h p.i. the F-SH Ig displays greater variation most likely due to random priming.

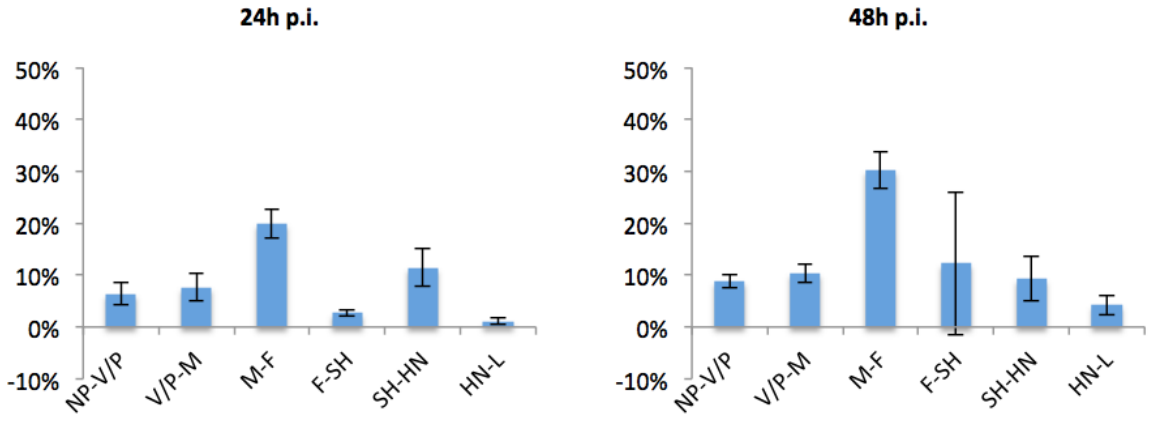


Fig 5.6. Analysis of the generation of read-through mRNAs at the intergenic regions of PIV5 strain W3 compared to the average coverage of the gene immediately upstream. The bars show standard deviation values based on three independent experiments.

Collectively, analysis of PIV5 strain W3 transcription at the later time points (24, 48 and 96 h p.i.) indicates that the processivity of the RNA polymerase does not significantly change during the course of the infection. This further suggests that modulation of RNA polymerase processivity is not responsible for repression of viral transcription and establishment of persistent infection.

The generation of defective interfering genomes (DIs) produced as an erroneous by-product of viral replication has been associated with the ability of paramyxoviruses to establish persistent infections (see Introduction section 1.5). DIs are shorter than the full-length viral genome and contain either the TrC promoter or both the TrC and Le promoters, giving them a replicative advantage

over the viral genome. Competition for RNA polymerase between viral genomes and DIs could result in repression of virus transcription. Furthermore, copyback DIs are dsRNA products and therefore could induce the host IFN response, which could also result in repression of viral transcription. The generation of DIs was investigated by bioinformatic analysis of HTS data; more details are provided in chapter 3 section 3.1.3. The analysis was conducted using ViReMa (Viral Recombination Mapper) software that identifies recombination events by detecting reads that map to two separate areas of the genome. This can be applied to the identification and production of DIs as reads that are generated from DIs would also map to two separate regions of the genome. The analysis was conducted at each time point, but no reads which map to two separate areas of the genome were detected, demonstrating that significant levels of DIs were not generated during the course of the PIV5 strain W3 infection described above. Therefore, DIs do not play a role in repression of viral transcription under these experimental conditions.

To investigate whether repression of viral transcription could be due to the host IFN response, A549 cells were pretreated with IFN or mock-treated prior to infection with PIV5 strain W3 at an MOI of 10 pfu/cell. RNA was extracted at 6, 12, 18, 24 and 96 h p.i. and subjected to polyA selection and library preparation followed by Illumina sequencing. The data were directionally analysed using the established bioinformatic pipeline, and the abundance of viral mRNAs was calculated (Fig 5.7). Establishment of the IFN response in A549 cells infected with PIV5 strain W3 delayed the peak of mRNA abundance by 24-48 h p.i. The abundance of viral mRNAs in pretreated cells peaks at similar levels to that of cells not pretreated with IFN, 7% and 6% of total cellular reads at 48 and 12 h

p.i. respectively. By 96 h p.i. the abundance of viral mRNAs was shown to be <0.23% of total cellular reads in pretreated and non-pretreated cells demonstrating a similar transcriptional “shut-off”. This experiment would require repetition to ascertain a more representative result. These preliminary results demonstrate that the repression of virus transcription at late times p.i. is not IFN dependent. This conclusion is supported by following the expression of virus proteins in A549 cells, cultured in the presence of Ruxoltinib, which blocks IFN signalling, and in A549/Npro cells, which cannot make IFN. In both cases, the switch off of viral protein synthesis follows the same kinetic pattern as that observed above in A549 cells (R.E. Randall & D.F Young, personal communication).

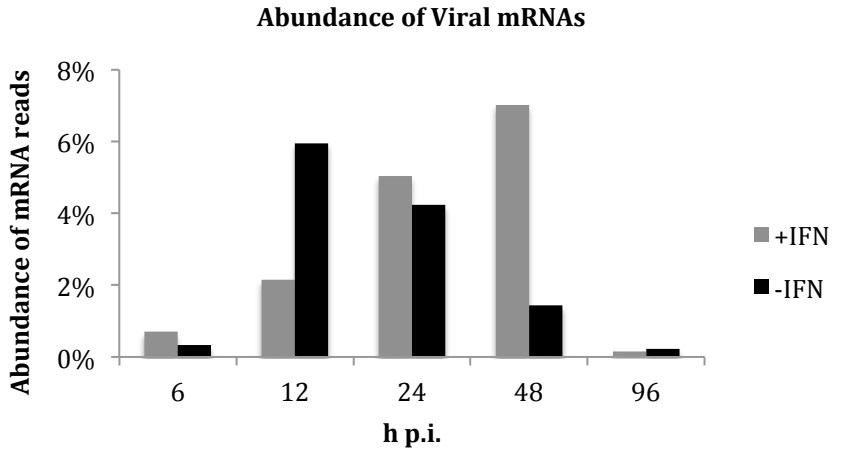


Fig 5.7. Kinetics of viral mRNA production in A549 cells infected with PIV5 strain W3 at an MOI of 10 pfu/cell in the presence and absence of 12 h pretreatment with IFN. The data were directionally analysed using the established bioinformatic pipeline. The number of viral mRNA reads was compared to total cellular reads from which the cytoplasmic and mitochondrial rRNA reads had been removed.

5.0.2 Comparison of PIV5 strain W3 and PIV5 strain CPI+

In contrast to the W3 strain of PIV5, previous observations by Professor Randall's group had shown that in cells infected with the PIV5 CPI+ strain the vast majority of cells die during the later stages of infection. Also, in contrast to PIV5 strain W3, which established a persistent infection easily, it is difficult to establish cell lines that are persistently infected with PIV5 strain CPI+. To investigate whether the high death rate of CPI+-infected cells is due to a lack of viral transcription repression at later times p.i., A549 cells were infected with PIV5 strain CPI+ at an MOI of 10 pfu/cell. RNA was extracted at 6, 12, 18, 24,

48 and 96 h p.i. and subjected to cytoplasmic and mitochondrial rRNA reduction library preparation and sequencing. The data were directionally analysed using the established bioinformatic pipeline (see chapter 3.0), and the abundance of viral genomes and mRNAs was compared to total cellular reads from which remaining rRNA reads had been removed (Fig 5.8A). Overall, PIV5 strain CPI+ demonstrated a transcription profile significantly different from that of PIV5 strain W3. W3 and CPI+ exhibited similar levels of viral mRNA generation between 6 and 12 h p.i., reaching 4.59% and 4.86% of total cellular reads, respectively. However, after 12 h p.i. the abundance of viral mRNAs continued to increase during PIV5 strain CPI+ infection, the overall abundance of mRNAs peaking at 17.3% at 18 h p.i., which is a more than four-fold increase in comparison to the peak abundance of PIV5 strain W3 viral mRNAs. The most significant increase in viral mRNAs during CPI+ infection occurred between 12 and 18 h, rising from 4.86 to 17.3% of total cellular reads. The abundance of viral mRNAs then remained approximately constant at 24 and 48 h, before halving to 8.86% of total cellular reads at 96 h p.i. The reduction in the abundance of viral mRNAs by 96 h p.i. reflects the death of the majority of infected cells.

The abundance of viral genomes for PIV5 strain CPI+ increased over time. There was a sharper increase in the abundance of viral genomes between 48 and 96 h p.i. from 0.36% to 4.78% of total cellular reads. This is different from W3, for which the abundance of viral genomes gradually increased up to 48 h p.i. to 2.9% of total cellular reads. This analysis would need to be repeated for PIV5 strain CPI+ to ascertain a representative result.

To corroborate these observations, A549 cells were infected with PIV5 strain W3 or CPI+ at an MOI of 10 pfu/cell or mock-infected. The cells were metabolically labelled using ^{35}S-methionine at 24, 48 and 72 h p.i. (Fig 5.8B). NP was observed as being generated at 24, 48 and 72 h p.i. during a PIV5 strain CPI+ infection, whereas PIV5 strain W3 again demonstrated a clear transcriptional shut-off with NP only observed at 24 h p.i. The radioactive labelling experiment thus corroborated the transcriptional results.

PIV5 strain CPI+ transcription was analysed as previously conducted for PIV5 strain W3, including assessing the abundance of viral mRNAs, templated and non-templated, and the generation of read-through mRNAs during the CPI+ infection was analysed as described for W3 (Appendix III). The abundances of viral mRNAs, both templated and non-templated, remained stable throughout the infection. The generation of read-through mRNAs also remains constant over time. Taken together these results suggest that, similar to W3, the RNA polymerase processivity does not change over the course of the infection.

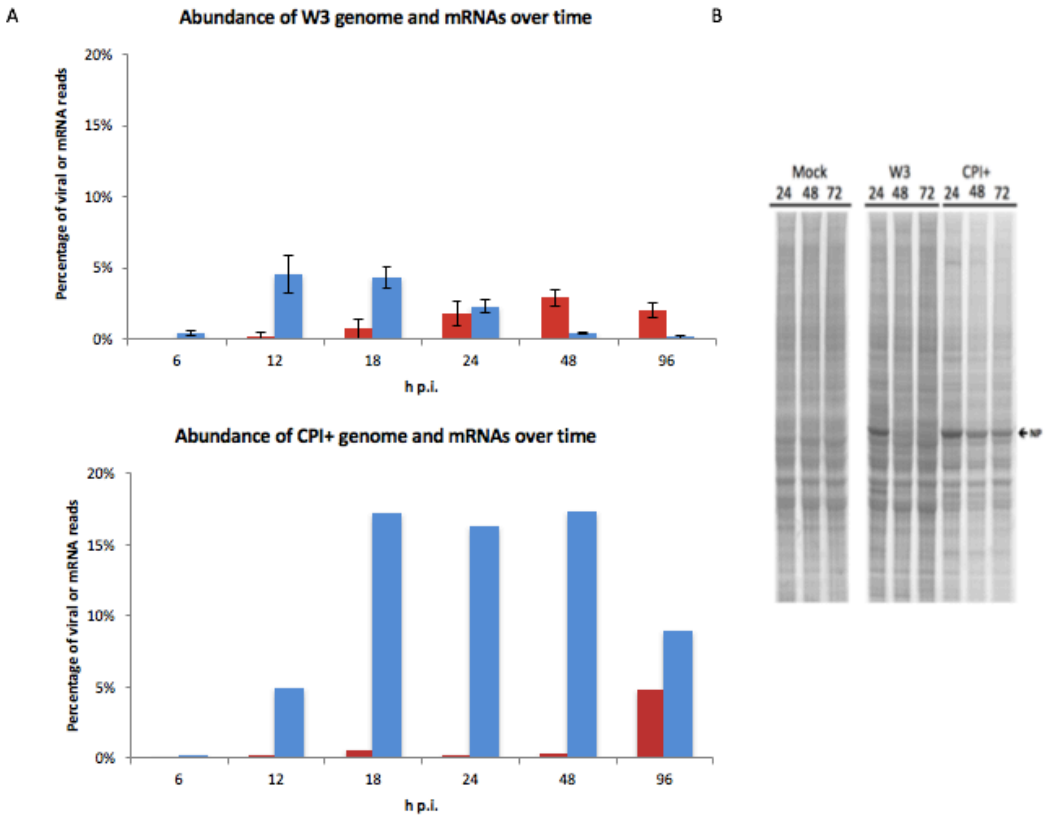


Fig 5.8. Kinetics of viral mRNA and genome production during PIV5 strain W3 or CPI+ infection. The data were directionally analysed using the established bioinformatic pipeline. The number of viral mRNA and viral genome reads was compared to total cellular reads from which the remaining cytoplasmic and mitochondrial rRNA reads were removed. The error bars displayed in the W3 panel are based on three independent experiments. No replicate data were available for CPI+. (B) Monolayers of A549 cells were infected or mock-infected with PIV5 W3 at an MOI of 10 pfu/cell. At 24, 48 and 72 h p.i. the cells were metabolically labelled for 1 h using ^{35}S-methionine. The polypeptides present in the total cell extracts were separated by SDS-PAGE on 4-12% polyacrylamide gels, and the labelled polypeptides were visualized using a phosphoimager. NP protein bands are indicated.

Although previous observations made by Professor Randall's group has shown that >90% of the cells die following infections with PIV5 strain CPI+, it was possible to establish cell lines persistently infected with this strain. This required the surviving cells to be cultured for many weeks without subculturing, replacing the culture medium regularly. Eventually the surviving cells started to divide and could be passaged, although for the first few passages they often showed CPE, for example fusion foci within the monolayers. The observations that, in striking contrast, the majority of cells infected with PIV5 strain W3 survived the infection, and that viral transcription was switched off in W3-infected cells but not in CPI+-infected cells, suggest that shut-off of viral transcription facilitated the early establishment of persistence by the latter. Therefore, it was of interest to determine whether viral transcription was also repressed in PIV5 strain CPI+ persistently infected cells, and, if so, whether this was due to the selection of mutant viruses, the presence of DIs or some other mechanism. To investigate this, total RNA was extracted from cells persistently infected with PIV5 strain W3 or CPI+ and subjected to rRNA reduction library preparation and sequenced. The data were then directionally analysed, as in previous experiments. The abundances of viral mRNAs and genomes were measured in relation to total cellular reads, from which remaining rRNA reads had been removed (Table 5.2). The PIV5 strain W3 data show a greater abundance of viral genomes, at 2.5% compared to viral mRNAs, at 0.03%. In contrast, CPI+ demonstrate a greater abundance of viral mRNAs, at 6.3%, compared to viral genomes, at 4%. These results indicate that PIV5 strain CPI+ generates a much higher proportion of viral mRNAs than PIV5 strain W3 in persistently infected cells. The abundance of viral mRNAs generated by PIV5 strain CPI+ is approximately 2.5 fold less compared to the peak viral mRNA abundance

observed in the lytic during the time course described above at 18-48 h p.i. (Fig 5.8A). The abundance of viral mRNAs has decreased at 96 h p.i. during the lytic infection most likely reflecting infected cell death. The decrease in the abundance of viral mRNAs in the persistently infected cell lines compared to the lytic infection indicates repression of viral transcription.

Table 5.2. Percentage of viral genome reads and viral mRNA reads compared to total cellular reads in PIV5 strain W3- and CPI+- persistently infected cell lines.

	genome	mRNA
W3	2.5	0.03
CPI+	4.0	6.3

The relative abundances of individual viral mRNAs were analysed by using FPKM values as previously described, using the RSEM software (Fig 5.9). The transcriptional profiles of the two PIV5 strains display some interesting differences. The PIV5 strain W3 transcription gradient is largely negative, decreasing from the NP gene generating the most abundant mRNAs to the L gene producing the least abundant mRNAs. However, the F gene produced a low abundance of mRNAs compared to downstream genes SH and HN. In addition, there is a sharp decrease in viral mRNA abundance between the HN and L genes. In contrast, PIV5 strain CPI+ produced more V/P mRNAs than NP mRNAs. The M, F, HN and L genes display a negative gradient, with a substantial decrease between the F and HN genes, and there was a greater abundance of V/P, M and F mRNAs in the PIV5 strain CPI+ infection, whereas

PIV5 strain W3 generated a greater abundance of HN mRNAs. The level of L mRNAs was relatively equivalent between the PIV5 strain W3 and CPI+ infections. The numerous differences between the transcription profiles of strains W3 and CPI+ during persistent infection could reflect differences in mRNA stability, or they may suggest a significant difference in the infection dynamics of the two strains and indicate that the mechanism of establishing persistent infections may be specific to the strain. This analysis would have to be repeated to obtain a representative result.

To investigate whether the selection of mutant viruses may have been involved in the establishment of persistent infections, the BWA alignments of the sequence reads generated from the genomes of PIV5 strain W3 and CPI+ in persistently infected cell lines were compared to the alignments of genome reads from the W3 and CPI+ lytic infections described above at 6 and 96 h p.i. If the selection of virus mutations were involved in the observed switch off of virus protein synthesis observed in cells infected with the W3 strain of PIV5 and the establishment of persistence, there should have been evidence of the mutations present in infected cells at 96 h p.i. However, no mutations were identified either by 96h p.i. or in persistently infected cells lines. Whilst no mutations were detected in cells infected with CPI+ at 96h p.i., several polymorphic mutations were identified in the CPI+ persistently infected cell lines but all of these, but one, were synonymous mutations. The exception was located at position 13093 of the genome (amino acid residue 1560 of the L protein), an A to T change in 17% of the reads that would change a phenylalanine residue in the L protein to a leucine residue.

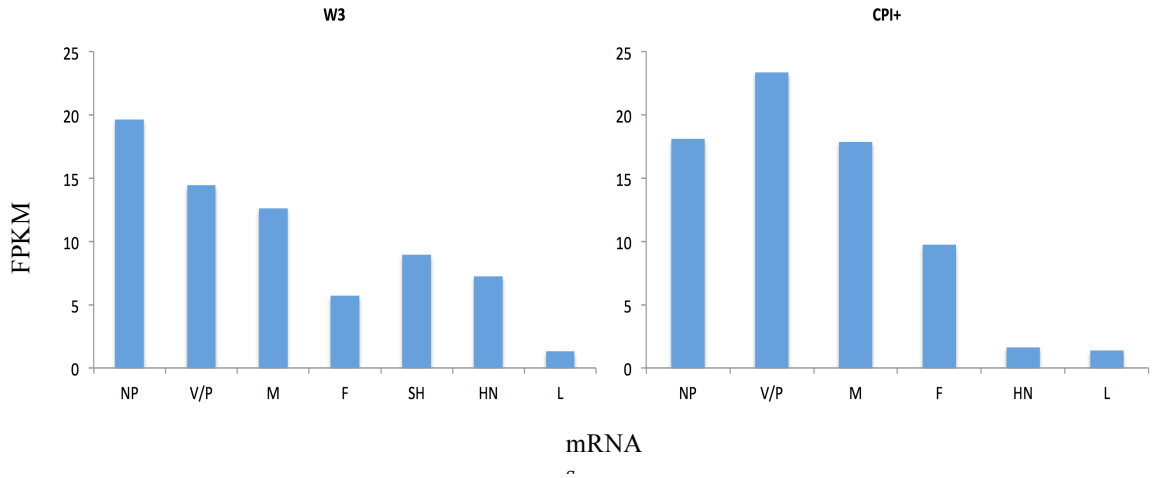


Fig 5.9. Abundances of W3 or CPI+ mRNAs in persistently infected cell lines. The data were directionally analysed using the established bioinformatic pipeline. The abundances of viral mRNAs were quantified by using FPKM values obtained using the RSEM software.

As previously discussed the presence of DIs has been associated with the repression of viral RNA synthesis and the establishment of persistent infections. Therefore cell lines persistently infected with PIV5 strains W3 or CPI+ were investigated for the presence of DIs. No DIs were identified in the W3 persistently infected cells. However, whilst DIs were not detected during the lytic CPI+ infection, DIs were detected in the CPI+ persistently infected cells lines. The analysis quantifies the number of reads that map to two different regions and identifies their genome positions (see chapter 3.0.3 for further details). This allows the abundance of DIs to be ascertained. The DIs make up approximately 0.02% of total viral reads in cells persistently infected with CPI+. This finding suggests that the generation of DIs in the small proportion of CPI+ infected cells

that survive the infection may play a role in the establishment of persistence by CPI+.

5.1 Analysis of the mechanism of repression of viral protein synthesis

Previous work has shown that phosphorylation of the P protein, which is a subunit of the RNA polymerase complex, plays a significant role in the regulation of polymerase activity. A study by Sun *et al* (2013) demonstrated the importance of the phosphorylation of a serine residue at position 157 in the P protein as a mechanism for down-regulating viral gene expression). The amino acid sequences of the P proteins of PIV5 strains W3 and CPI+ were aligned in order to identify non-synonymous differences that could account for differences in the viral transcription and replication profiles. This analysis showed that at position 157 PIV5 strain W3 has a serine residue whereas strain CPI+ has a phenylalanine residue. Interestingly, analysis of a number of other PIV5 isolates showed that the incidence of either a serine or a phenylalanine at position 157 occurred at approximately a rate of 1:1.

To investigate whether phosphorylation of the P protein at serine 157 may be responsible for the shut-off of viral transcription observed in W3-infected cells and the establishment of persistence, a recombinant virus (rPIV5-W3(F157)) using the W3 backbone was generated by Professor Goodbourn (St George's Hospital, London), in which the serine at position 157 had been mutated to a phenylalanine. Analysis of the consensus sequence of the recombinant virus showed that this was the only mutation present in the genome.

To assess cell survival after infection, A549 cells were infected at an MOI of 10 pfu/cell with either the W3 parental virus with a serine residue at position 157 (PIV5-W3(S157)) or rPIV5-W3(F157). Phase contrast microscopy revealed that, in contrast to the parental virus (PIV5-W3(S157)), >90% of cells had died by 72 h p.i. following infection with rPIV5-W3(F157) (Fig 5.10B). Additionally, the infected cells were metabolically labelled using ^{35}S-methionine at 24 and 72 h p.i. As previously observed (Fig 5.1-3), PIV5-W3(S157) displayed repression of viral transcription at later times p.i., whereas rPIV5-W3(F157) demonstrated no observable shut-off of viral protein synthesis at later times p.i.

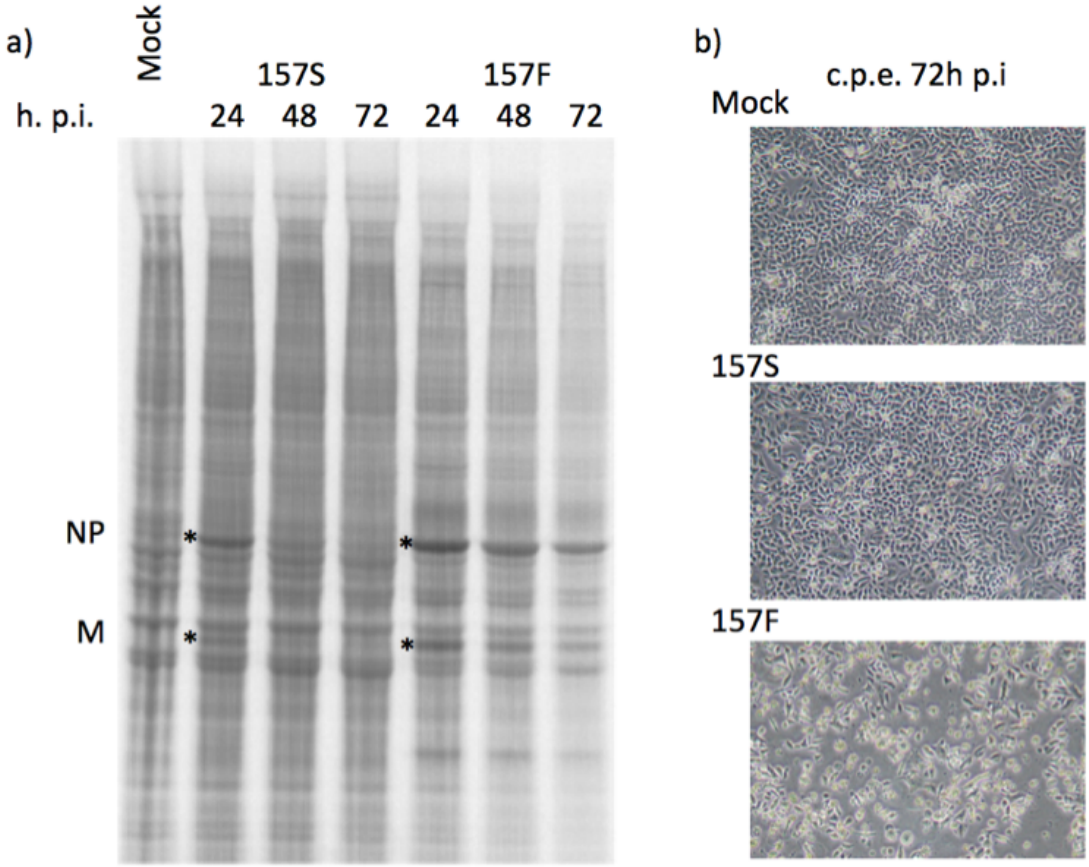


Fig 5.10. Analysis of protein synthesis of PIV5-W3(S157) and rPIV5-W3(F157) (A) A549 cells were infected with PIV5 W3 (PIV5-W3(S157)) or rPIV5-W3(F157) at an MOI of 10 pfu/cell. At 24, 48 and 72 h p.i. cells were metabolically labelled for 1 h using ^{35}S-methionine. The polypeptides present in the total cell extracts were separated by SDS-PAGE on a 4-12% polyacrylamide gel, and the labelled polypeptides were detected using a phosphoimager. (B) A549 cells were mock-infected or infected with PIV5 W3 or rPIV5-W3(F157) at 10 pfu/cell and imaged by phase contrast microscopy.

These results identify a single point mutation resulting in replacement of the serine residue at position 157 in the P protein by a phenylalanine residue to be responsible for the ablation of the transcriptional shut-off observed in PIV5 strain W3.

To investigate the effect of the mutation of a serine to a phenylalanine on virus transcription, A549 cells were infected with either PIV5-W3(S157) or rPIV5-W3(F157) at an MOI of 10 pfu/cell. RNA was extracted at 24 h p.i. and subjected to cytoplasmic and mitochondrial rRNA reduction library preparation followed by sequencing. The sequencing data was then directionally analysed using the established bioinformatic pipeline. The abundance of viral mRNAs and genomes was compared to total cellular reads after the remaining nuclear and mitochondrial rRNA reads were removed (Fig 5.11).

The abundance of viral mRNAs of PIV5-W3(S157) and rPIV5-W3(F157) displayed very different profiles at 24 h p.i. PIV5-W3(S157) accounts for 2.3% of total cellular reads. In contrast, rPIV5-W3(F157) viral mRNAs accounted for 12.3% of total cellular reads, a six-fold increase compared to PIV5-W3(S157). These results strongly suggest that phosphorylation of a serine at position 157 of the P protein is responsible for the inhibition of viral transcription. Interestingly, the abundance of viral genomes is relatively equal, at 1.8% and 2.4% of total cellular reads for PIV5-W3(S157) and rPIV5-W3(F157) respectively.

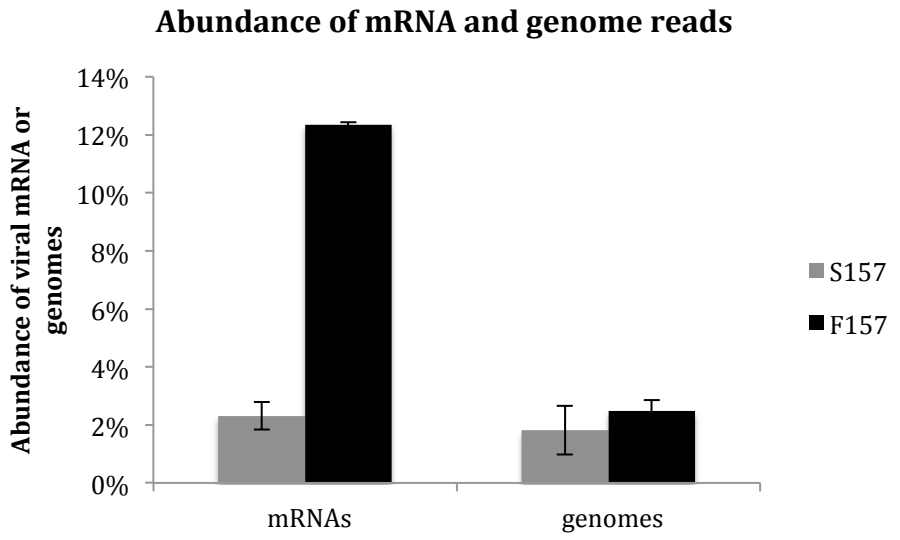


Fig 5.11. Abundance of viral mRNA and genome reads for PIV5-W3(S157) and rPIV5-W3(F157) at 24 h p.i.. The data was directionally analysed using the established bioinformatic pipeline. The abundance of viral mRNA and genome reads was compared to the total cellular reads from which remaining cytoplasmic and mitochondrial rRNA reads were removed. Standard deviations were obtained from three independent experiments.

5.2 Analysis of PIV5 persistence in AGS cells

Previous observations have demonstrated the presence of a persistent PIV5 infection in the commercially available AGS cell line (Young et al., 2007). This cell line was derived from a human gastric adenocarcinoma and is commonly used in biomedical research. Since the persistently infected cells do not exhibit cytopathology, this could indicate that the virus may be defective. The genetic characteristics of the virus were determined. The virus (PIV5-AGS) was isolated

from the AGS cells, and A549 cells were infected with PIV5-AGS at an MOI of 10 pfu/cell. Total RNA was extracted at 24 h p.i. from these cells and from the persistently infected AGS cell line. The RNAs were subjected to polyA selection and libraries were sequenced on the MiSeq platform (Illumina). The data were directionally analysed using the established bioinformatic pipeline, and the genome sequences were determined (as significant abundance of virus genomes were co-purified along with the polyA tail mRNAs). The data were originally aligned against the PIV5 strain W3 reference sequence (accession no. JQ743318) using Bowtie2 software and visualized using Tablet, and differences were identified in order to derive the final sequences. The consensus sequences revealed only two differences at positions 6910 and 7445, which were polymorphic in both genomes and were comprised of G and A residues. These results show that the virus in the persistently infected AGS cells was able to replicate autonomously, although it gave a small plaque phenotype on Vero cells (D.F. Young, personal communication).

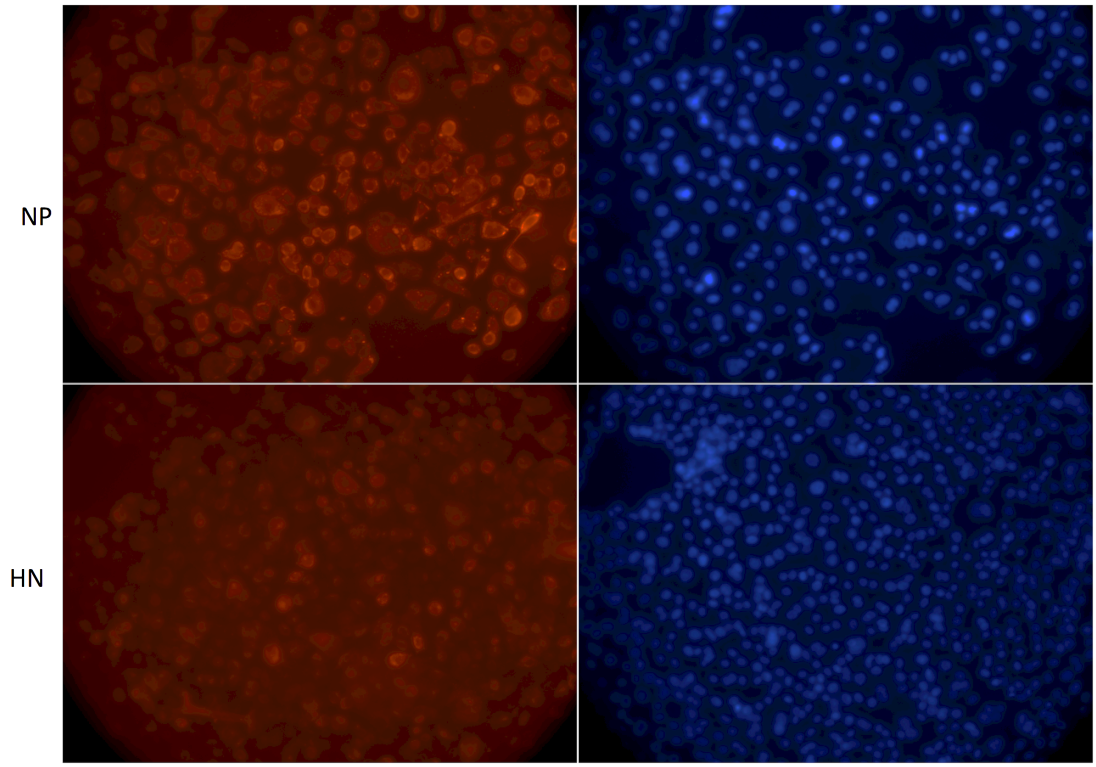


Fig 5.12. Immunoflurescence of AGS persistently infected cells. Persistently infected cell were grown in monolayers on coverslips for 24 h, and then fixed and immunostained using anti-NP and anti-HN antibodies (red). DAPI was used as a counter stain for nuclei (blue). The cells were visualized using an Evos microscope.

The genome sequence of the AGS virus revealed the presence of a serine at position 157 in the P protein. Immunofluorescence was conducted to ascertain whether AGS-infected cells display a similar NP and HN profile, which indicate whether the infected cells are actively transcribing the virus, similar to that of PIV5 strain W3, potentially indicating a similar mechanism by which persistence is established. A549 cells persistently infected with AGS were fixed and

immunostained at 24 and 96 h p.i. using anti-NP and anti-HN antibodies (Fig 5.12). Similar to the situation with PIV5 strain W3 (Fig 5.1), NP was detected in the majority of infected cells at 96 h p.i. Also, as observed in cells infected with PIV5 strain W3, HN was detected in only a proportion of cells, indicating that at late times p.i. only these cells were transcribing and replicating the viral genome. Although further work needs to be carried out, these results suggest that the presence of a serine at position 157 of the P protein in the AGS virus may result in the repression of viral transcription at late times p.i., thereby facilitating the establishment of persistence in the commercially available cells. Unfortunately, the persistently infected cell line could not undergo analysis of virus transcription and replication using the bioinformatic pipeline previously described due to technical reasons.

A phylogenetic analysis was also conducted to determine the relatedness of strain AGS to the 17 other PIV5 strains for which genome sequences are available, using the MEGA7 software (Fig 13) (Kumar et al., 2016). The phylogenetic tree suggests that strain AGS is most closely related to human PIV5 isolates, indicating that the virus may have been present in the original biopsy or that it may have been contaminated in the laboratory during isolation of the AGS cell line.

Taken together these results show that the presence of a serine a position 157 of parainfluenza viruses may be an indicator of the virus ability to easily establish a persistent infection through the repression of viral protein synthesis at late time p.i.. A characteristic that can have long reaching consequences such as the ability to persistently infect a commercially available cell line.

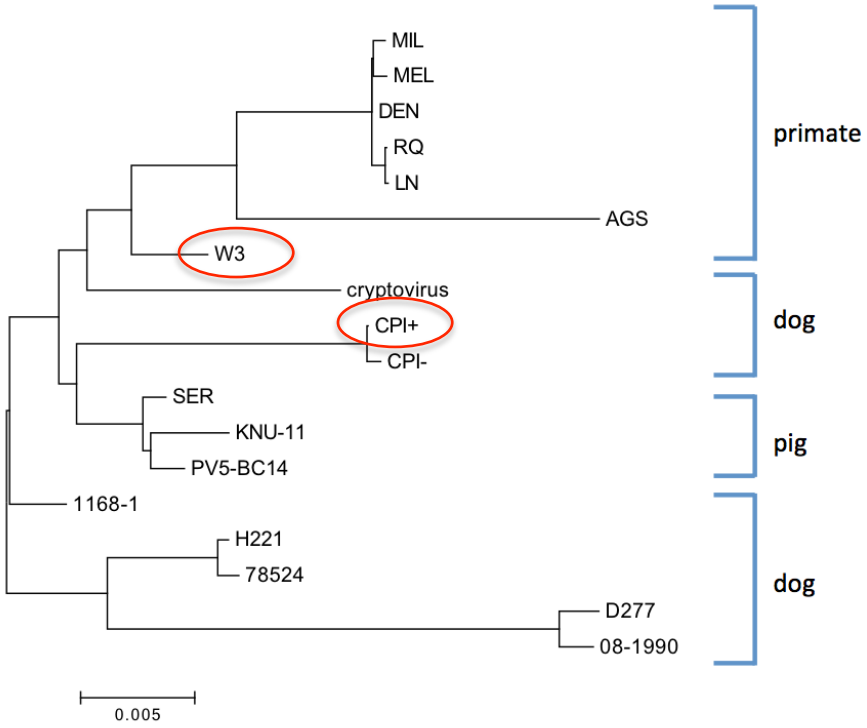


Fig 5.13. Phylogenetic analysis of the complete genome sequences of PIV5 strains. Constructed using a general time reversible model using MEGA 7 software. The animals from which the strains were isolated are provided. The red circles indicate strains used in this study.

Chapter six

Discussion

6.0 Establishment of a pipeline for the analysis of negative-strand RNA viral transcription and replication

The development of HTS has revolutionized biomedical research, including virology. There are numerous library preparation and sequencing methods that can be employed to answer specific biological questions in great detail. For research into negative-strand viruses HTS offers unparalleled opportunities to gain an understanding of the dynamics of viral infection.

Previous research has demonstrated that paramyxoviruses control viral gene expression by stop-start transcription and the consequent generation of a transcription gradient, with genes at the 3' and 5' ends of the genome generating the most and least abundant mRNAs, respectively. Traditionally, gene expression studies and investigations of transcription by paramyxoviruses have been conducted using ultra violet-treated viruses and northern blot analysis (Abraham and Banerjee, 1976; Ball and White, 1976; Collins and Wertz, 1983; Huang and Wertz, 1982, 1983). More modern methods include the use of microarray analysis or quantitative PCR, for which the target nucleic acid sequence must be known. Owing to the ability to assemble data de novo into genomes of previously unknown sequence, HTS does not require any such prior knowledge. Additionally, unlike earlier methods, HTS can identify novel transcripts, gene fusions and gene variants with great precision. Typically, in gene expression and transcriptome studies using HTS, the total mRNA of infected cells is isolated by polyA selection and then sequenced, thereby generating an overview of all of the viral mRNAs present in the cell at the time of RNA extraction.

The initial aim of this study was to investigate whether the pattern of paramyxovirus transcription changed during infection. Initial investigations employed polyA selection of mRNAs from infected cells using the TruSeq stranded mRNA library preparation kit (Illumina), which isolates RNAs containing polyA tracts by hybridisation to oligo(dT) beads. This library preparation method generates directional data, which can, but does not automatically, facilitate bioinformatical separation of sequence reads based on directionality. For paramyxoviruses (and this could be applied to all negative-strand RNA viruses), bioinformatic steps allow the separation of viral genome (negative sense) reads from viral mRNAs and antigenome reads (positive sense). A preliminary investigation of the PIV5 strain W3 transcriptome, in which directional analysis was not conducted, generated a relatively flat transcription gradient in genes downstream of the M gene, which was unexpected as it did not agreed with published work showing a decreasing transcription gradient from the NP to the L gene (Karron and Collins, 2013; Lamb and Parks, 2013). However, after conducting directional analysis of the same data and separating viral genome reads from viral mRNA and antigenome reads, the true transcription profile was revealed as consistent with that published previously.

Viral mRNA and antigenome reads cannot be separated using directional analysis as they are both positive sense. However, the contribution of antigenome reads to the transcriptional gradient appeared to be minimal and must have been less than the read coverage of the L gene, which was very low. I concluded from this that the shape of the transcription gradient and further

downstream analysis of the read data would not be significantly affected by the presence of a small proportion of antigenome reads. In summary, PIV5 strain W3 displayed a transcription gradient from the 3' to the 5' end of the genome, with genes at the 3' end (starting with NP) producing the most abundant mRNAs and genes at the 5' end (ending with L) producing the least abundant mRNAs. Over 50% of the viral reads isolated by polyA selection were of viral genome origin. These results highlight a major disadvantage of isolating viral mRNAs by polyA selection, as a significant proportion of viral reads in this fraction originated from genomes, which probably co-purified with viral mRNAs through molecular hybridization during RNA extraction and polyA selection.

Although viral genomes co-purified with polyA RNAs, this method could not be used to accurately quantify viral genomes. Therefore, a study was conducted to ascertain whether sequencing total infected cellular RNA, following depletion of cytoplasmic and mitochondrial rRNA, would facilitate the quantification of not only viral mRNAs but also viral genomes and, potentially, DIs. A comparison of library preparation methods, using polyA selection or rRNA reduction, was undertaken. Although there was a very slight decrease (<1%) in the abundance of viral mRNA reads generated by the latter method, this was not significant to downstream analysis. Furthermore, the coverage observed in read alignments using both methods was sufficient for downstream analysis even at 6 h p.i., when the abundance of viral mRNAs was at its lowest. These results revealed that total RNA with rRNA reduction library preparation could be employed to quantify the relative abundance of viral mRNAs and therefore the viral transcriptional gradient. Additionally, the abundance of viral genomes could be quantified, and the presence of DIs could be established.

These experiments demonstrated the advantages of using rRNA reduction, stranded sequencing and directional analysis for investigating the relative abundance of viral mRNAs. Additionally, the use of rRNA reduction allows the simultaneous quantification of viral mRNAs and genomes. The establishment of this workflow offered a powerful opportunity to investigate both the transcription and replication of paramyxoviruses, and could be applied to other negative-strand viruses. Previously, the analysis of transcription and replication was conducted using traditional approaches such as mini-replicon systems and mutational analysis. These can be employed to answer specific questions such as the importance of particular regions of the genome, and still have an essential role to play in negative-strand virus research. However, they cannot provide a complete overview of the infection dynamics of the virus. The workflow established in this study utilizes HTS to analyse viral transcription and replication simultaneously, thus providing unique insights into negative-strand virus infections.

6.1 Analysis of viral transcription for the rubulaviruses PIV5, PIV2 and MuV and the respirovirus PIV3

The pipeline described above was employed to analyse simultaneously the transcription and replication of the prototype paramyxovirus, PIV5 strain W3 (a rubulavirus). Initially, this analysis was conducted over a 24 h time course in order to determine viral infection dynamics during an acute infection. The

analysis was then extended to human isolates of two additional rubulaviruses, PIV2 and MuV, and a respirovirus, PIV3.

6.1.1 Kinetics and abundance of viral mRNAs and genomes

Although PIV5 strain W3, PIV2, MuV and PIV3 all showed a clear transcriptional gradient of mRNA abundance, as described in detail below, there were differences in the kinetics of transcription between these viruses. PIV5 strain W3 displayed a significant increase in the abundance of viral mRNAs between 6 and 12 h p.i., increasing from 0.4 to 4.5%, respectively. The peak of viral mRNA abundance occurred at 12 h p.i. There was a slight decrease in the abundance of viral mRNAs at 18 and 24 h p.i., by which time the viral mRNA reads accounted for 2.3% of total cellular reads. Interestingly, PIV2 and MuV exhibited a similar transcription pattern at 6, 12 and 18 h p.i.. Between 6 and 12 h p.i., PIV2 and MuV displayed an increase in the abundance of viral mRNAs from 0.4 and 0.5 to 8 and 9.8%, respectively. The abundance of viral mRNAs then approximately doubled between 12 and 18 h p.i. to 16 and 16.9% for PIV2 and MuV, respectively. This increase represents the peak abundances of viral mRNAs for MuV, after which time there was a slight decrease to 15.9% of total cellular reads by 24 h p.i. However, the abundance of PIV2 mRNAs continued to increase to 19% at 24 h p.i. PIV3 transcription occurred at the earliest times p.i., with 9.7% of total cellular reads being of viral mRNA origin by 6 h p.i. Thereafter, viral mRNA abundance continued to rise until 12 h p.i., at which time viral mRNA reads made up 17.5% of total cellular RNA reads. At later times in infection, the abundance of PIV3 mRNAs decreased slightly, so that by 24 h p.i. they contributed 11.5% of total cellular reads.

The abundance of viral genomes exhibited a similar pattern for all four viruses, gradually increasing over time and not rising above 2.9% of total cellular reads. Interestingly, the abundance of viral genomes did not seem to be dependent on the abundance of viral mRNAs, as PIV5 strain W3 displayed a peak abundance of transcripts that was four-fold less compared to the other viruses, whereas these viruses had comparable genome abundances. Additionally, growth curves conducted using PIV5 strain W3 and PIV2 (appendix III) demonstrated that they had released newly synthesized viruses in equivalent amounts at 24 h p.i. (7×10^6 and 1×10^7 pfu/ml respectively), indicating that an increase in the abundance of viral transcripts did not correlate to an increase in the generation of new viruses.

6.1.2 Intracellular spread of infection during viral replication

The abundance of viral genomes increased approximately three-fold between 6 and 12 h p.i. for all four viruses. The detection of viral replication at early times in infection was unexpected, as the current model requires there to be sufficient amounts of NP^0 available in the cytoplasm for encapsidation of new synthesized viral genomes and antigenomes to allow the virus RNA polymerase to switch from a transcription to a replicative mode. Unfortunately, a 0 h time point was not taken for PIV2, MuV or PIV3, and therefore it was not possible to ascertain whether viral replication was occurring prior 6 h p.i.. Any future analysis should include this time point. PIV5 strain W3 exhibited an increase in the abundance of viral genomes between 0 and 6 h p.i. demonstrating viral replication prior to 6 h. Radioactive labeling of infected cells showed that NP was being generated

by each virus at early times, although at 6 and 12 h p.i. it was present only at low abundances. However, immunofluorescence studies showed that even by 3–6 h p.i. NP was present, and, rather than being evenly distributed throughout the cytoplasm, was present in a small number of cytoplasmic foci that were restricted in their location within infected cells. As infection proceeded, the number of foci increased to a point at which the cytoplasm was full of them. It has been suggested for PIV5 that the cytoplasmic foci contain the NP, P and L proteins, as well as genomic RNA, and may represent areas of active viral transcription and replication (Fearns et al., 1994; Carlos et al., 2009). There is also evidence that other members of the order *Mononegavirales*, including rabies virus (discussed further below), may replicate in similar foci. If, as seems likely, the cytoplasmic foci are sites of viral transcription and replication, the data in combination suggests a model for paramyxovirus replication in which, following infection, viral transcription occurs initially at (random) sites within the cytoplasm where incoming genomes reside. The viral mRNA is then translated on ribosomes localized around the viral genome, resulting in a local build-up of replicative proteins until the concentration of NP^0 and other viral proteins in the vicinity of the foci becomes sufficient to facilitate the switch from transcription to genome replication. Thereafter, the newly replicated genomes are dispersed to new sites within the cell, and so the process continues until there are replicative sites distributed throughout the cytoplasm. The proposed new model of virus transcription and replication is summarized in figure 6.1. In this model, although viral replication occurs at localized sites very early in infection, maximal levels of viral transcription may not be observed until later times, when there are more replication foci. Also, there would presumably be asynchrony at different foci as to whether a switch between viral transcription and replication has occurred.

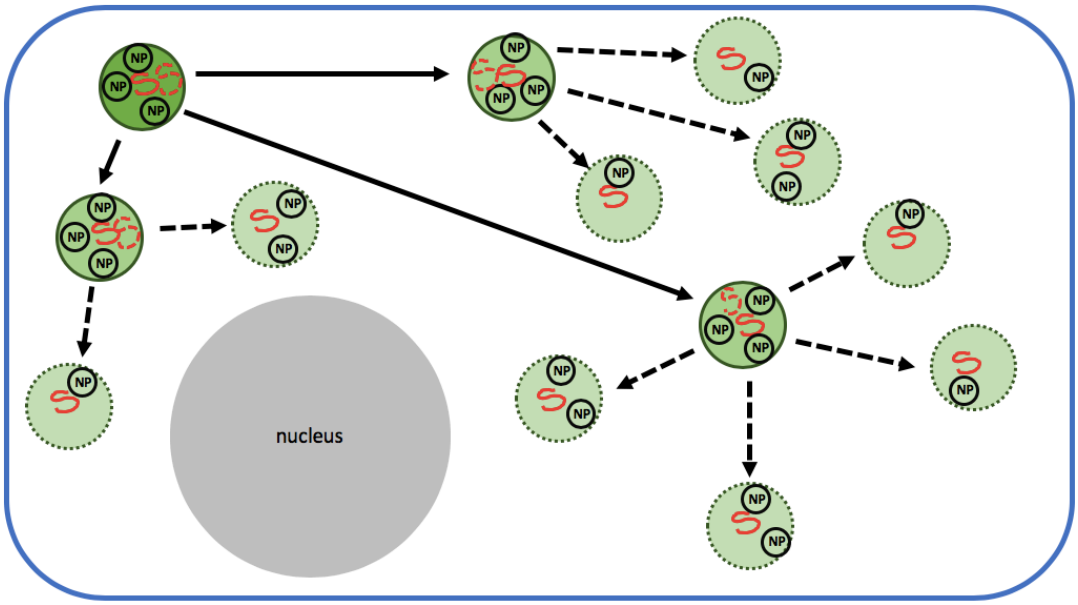


Fig 6.1 A schematic of the new model of paramyxovirus transcription and replication. The blue box represents the infected cell membrane and the grey circle indicates the nucleus. The green circles represent foci of virus transcription and replication. The darker green circle with the dark green solid outline is the foci generated from the infecting virus nucleocapsid complex. The solid red lines represent the virus genomes which are transcribed by the virus polymerase complex and translated by host cell ribosomes to produce virus proteins including NP (represented by black circles). A build up of NP at the local sites and not in the cytoplasm allows the virus to switch to replicative mode and produce antigenomes (represented by red dashed lines) and new virus genomes that require encapsidation with NP. The new encapsidated virus genomes then spread throughout the cell to create new foci (secondary foci) as indicated by the solid arrows. Where the process is repeated. The dashed arrows indicated the nucleocapsid spread from the secondary foci to multiple new foci indicated by the light green circles with the dashed outer line. The foci are eventually dispersed throughout the cytoplasm.

Although a restricted number of cytoplasmic foci were observed at early times in infection with all three viruses analysed (PIV2 could not be included in this analysis), preliminary results suggested that there are differences in how viral infection spreads intracellularly within infected cells. For PIV5 and MuV, there were initially only a few small cytoplasmic foci, presumably initiating at sites of incoming viral genomes. As their number increased, the foci appeared to be randomly distributed through the cell. For PIV5 strain W3, the cell became full of small cytoplasmic foci between 6 and 12 h p.i.. Thereafter, as viral transcription plateaued and then declined, there was a decrease in the total number of foci because they appeared to fuse and form larger cytoplasmic bodies. Similarly in MuV infected cells at late times in infection, large cytoplasmic foci accumulated. The formation of these large inclusion bodies may provide a stable environment for the maintenance of viral genomes and protect the genomes from the host immune response. This would allow infected cells to survive the infection and be reactivated at later times. The establishment of persistence is discussed further below.

The pattern of viral spread within infected cells appears to be different for PIV3, which replicates faster than PIV5 strain W3 and MuV. Indeed, the amount of viral mRNAs within PIV3-infected cells at 6 h p.i. was 18-fold higher that that in cells infected with PIV5 strain W3 or MuV, and this correlated with a higher abundance of NP. Furthermore, although numerous cytoplasmic foci were detected in PIV3-infected cells by 3-6 h p.i., rather than being evenly distributed through the cytoplasm these foci were localized to a specific region of the cell, from which the infection appeared to spread out rapidly. Also, there was no evidence for the formation of larger inclusion bodies in PIV3.

The differences in the intracellular spread of cytoplasmic foci formed by PIV3, compared to PIV5 strain W3 and MuV, suggests a different mechanism by which the virus genomes are dispersed throughout the cell and form new foci for viral replication and transcription. This may be dependent on the genera, PIV5 strain W3 and MuV being rubulaviruses whereas PIV3 is a respirovirus. The way in which rabies virus (a rhabdovirus) spreads intracellularly during an infection has recently been investigated by Nikolic et al., (2017). Following replication in inclusion bodies (known as negri bodies for this virus), newly encapsidated genomes are ejected from inclusion bodies and dispersed throughout the cell by microtubules (Lahaye et al., 2009). The ejected nucleocapsid complexes subsequently form new negri bodies. When the formation of microtubules was disrupted, the encapsidated viral genomes were not distributed and remained close to the original inclusion body. Eventually, newly formed negri bodies fused with the original body, creating a single larger inclusion body.

During the course of these studies, there was no evidence for the formation of large inclusion bodies in cells infected with PIV3. However, Zhang et al., (2017) observed that large PIV3 inclusion bodies were formed from the fusion of smaller inclusion bodies and required acetylated α-tubulin. Additionally, the formation of the larger inclusion bodies led to an increase in viral replication. The difference between Zhang results and the results obtained from this study could be due either to strain differences in the virus used, or to cell-type specific differences. Interestingly, a study by Zhang et al., (2013) demonstrated that mutation of a leucine to an alanine residue at position 478 of NP abolished the

ability of PIV3 to form inclusion bodies through distrusting the interaction of the disordered C-terminal tails of NP present in NP-RNA and the C-terminal end of P. However, the N^0-P interaction was not affected by this mutation. This study also shows that formation of inclusion bodies was critical to RNA synthesis. Overall, these studies suggest that the NP-RNA-P interaction, and possibly the interaction of this complex with host tubulins, plays a critical role in the formation of inclusion bodies. It is also of note that the NP proteins of PIV5 strain W3 share between 50-60% identity with their orthologue in PIV2 and MuV (rubulaviruses) but share little common sequence with their orthologue in PIV3 (respirovirus) (Lamb and Parks, 2013). In contrast, the C-terminal domain of P, which is responsible for binding to the N-RNA complex, has a relatively well conserved secondary structure in members of the family Paramyxoviridae (Lamb and Parks 2013). If there is a critical role for the NP-RNA-P interaction in the formation and dispersal of viral genomes, it is possible that, given the conserved secondary structure of the P protein that mutations in the NP protein, such as that observed for PIV3 by Zhang et al., (2013), this could account for the differences in the mechanism of viral genome dispersion in the cytoplasm between rubulaviruses (PIV5 strain W3 and MuV) and respiroviruses (PIV3).

6.1.3 Viral transcription profiles

Analysis of the transcriptional profiles conducted by calculating the abundances of virus mRNAs at each time point effectively allows the quantification of the transcription gradient. Any changes observed in the transcript abundances may indicate a change in viral RNA polymerase activity and therefore in the control of viral transcription and ultimately gene expression. All four viruses clearly

displayed a transcription gradient, with genes at the 3' end of the genome (starting with NP) producing the most abundant mRNAs, and those located towards the 5' end (ending with L) generating the least abundant. The transcription gradient of each virus remained relatively unchanged throughout the infection, and any slight changes may reflect differences in the stability and turnover of viral mRNAs. Therefore, from this analysis there was no evidence of temporal control of relative mRNA abundance.

Although for any given virus the transcriptional gradient remained unchanged during infection, there were clear differences in the shape of the gradient between the four viruses. PIV2 displayed a significant decrease in the relative abundance of V/P mRNAs compared to the upstream NP mRNA. In genes downstream of V/P, a consistent, continuous decrease in the transcription gradient was observed. For MuV, there were similar levels of the NP and P/V mRNAs early in infection, but at later time the V/P gene displayed a greater abundance of transcripts compared to the upstream NP gene. This may have been due to differences in the stability of viral mRNAs. There was a significant decrease in the abundance of mRNAs between the V/P and M genes, with genes downstream of V/P displaying a decreasing transcriptional gradient. PIV3 exhibited a flatter gradient in the P/D/C, M, F and HN genes. A flatter transcription gradient was also observed at 24 h p.i. during PIV5 strain W3 infection. However, this most likely reflects differences in the stability of mRNAs and not changes in the processivity of the RNA polymerase as by 96 h p.i. (observed during the persistence chapter analysis Fig 3) PIV5 strain W3 again exhibits a decreasing gradient.

A common feature of all of the viruses is a significant decrease in relative abundance of L mRNAs compared to HN mRNAs (from the upstream gene), the latter being at least two-fold less abundant. The decrease in the abundance of L mRNA was particularly obvious for PIV3, for which there was an eight-fold decrease in the abundance between the HN and L genes. This was unexpected, since PIV3 exhibits a level of conservation in the Ig. Therefore, if the level of L mRNA is dictated by the rate of transcription termination and re-initiation at the HN-L Ig region, this cis-acting element cannot be primarily responsible for determining the ratio of HN to L mRNAs. It is possible that sequences within the Ge of HN and the Gs of L are involved, but these are also conserved in PIV3. Also, the rate of read-through at the HN-L Ig is not significantly lower compared to the other genes, indicating that the rate of termination and re-initiation is similar and does not account for the lower abundance of L mRNAs. Alternatively, the lower abundances of L mRNA may indicate that, during transcription, the RNA polymerase disengages from the genome not only at gene junctions but also (randomly) within genes. If this is the case, then any truncated mRNAs would not be polyadenylated and consequently would be rapidly degraded (Dreyfus and Regnier, 2002; Wickens et al., 1997). Given the very long length of the L gene, only a very small proportion of L mRNAs may be full-length, polyadenylated and therefore stable, resulting in the observed large difference between the relative abundance of HN and L mRNAs.

6.1.4 Read-through mRNAs and transcription termination

The precision of RNA polymerase processivity can be observed in the generation of read-through mRNAs, which are generated when the RNA polymerase fails to terminate transcription at the Ge signal, instead continuing to transcribe the Ig and downstream gene(s) and releasing a polycistronic mRNA. The overall pattern of read-through mRNA generation remains stable throughout infection for all four viruses. There are significant similarities among the viruses in read-through mRNA generation. PIV5 strain W3, PIV2 and PIV3 displayed a greater abundance of M-F read-through mRNAs compared with the other gene junctions. However, the abundance of the dicistronic M-F mRNA was different. For PIV5 strain W3, approximately one third of transcripts were read-through, whereas PIV2 and PIV3 displayed a much higher proportion (approximately 98%). These results are particularly interesting because the cis-acting elements that direct effective transcription termination in paramyxoviruses differ significantly between rubulaviruses and respiroviruses. These results suggest that the viruses have evolved specific mechanisms to retain the proportion of M-F read-through mRNAs, thus suggesting that this has an essential role in viral growth.

The results for PIV5 strain W3 displayed similar relative abundances of read-through mRNAs to those observed by Rassa and Parks (1998). Read-through transcripts at NP-V/P, V/P-M and SH-HN were produced in a relatively equal proportions, thus corroborating the observation made by Rassa and Parks (1998), although the proportions obtained in that study were higher (~15%) than in this study (~9%). The HTS results displayed an approximately three-fold

greater abundance of read-through mRNA generation at the M-F genes compared to the V/P-M genes, and this remained relatively consistent throughout infection. However, unlike the Rassa and Parks study, which found read-through production at the M-F genes to be higher at early times in infection compared to late times (approximately 50% and 20%, respectively), the results observed here showed that the relative abundance of read-through generation at the M-F genes remained relatively constant. In contrast to the other viruses, MuV Enders strain displayed a 100% read-through rate at the F-SH genes, confirming previous results showing that this strain does not produce any monocistronic SH mRNAs (Takeuchi et al., 1991, 1996). Takeuchi et al., (1991) detected no bicistronic SH-HN mRNAs and relatively abundant (~40%) read-through at the SH-HN genes similar to the results obtained here. MuV Enders strains also displayed a high level of M-F read-through mRNAs (~27%), similar to that observed for PIV5 strain W3.

Numerous studies have corroborated a higher proportion of read-through mRNAs at the M-F genes in a number of different paramyxoviruses including NDV, MeV and PIV5 (Hiebert et al., 1985; Paterson et al., 1984; Rozenblatt et al., 1982; Wilde and Morrison, 1984). Two hypotheses have been put forward to explain why paramyxoviruses have evolved a mechanism to achieve this. Firstly, that this inhibits expression of the F protein, thus facilitating a decrease in the cytopathic effects of infection. An increase in F protein abundance has been shown to increase cell-to-cell fusion (Horvath and Lamb, 1992a, b). Therefore, inhibiting expression of the F protein may be beneficial for viral growth by decreasing this cytopathogenic effects. A study by Lingemann et al., (2015) suggested that in PIV3, where the Ge, Ig and Gs signals are highly

conserved, the high rate of M-F read-through was due to an 8 nt insert in the M gene Ge signal that is also present in the strain used in this study. A mutant virus in which this insert was deleted displayed a decrease in the read-through at the M-F junction and a subsequent increase in expression of the F protein, which resulted in an increase in plaque size probably due to a higher rate of cell-to-cell fusion. Additionally, F is a neutralizing and protective antigen during PIV3 infection, and increased expression of F due to a decrease in M-F read-through resulted in an increase in F-specific serum IgG antibody titres, suggesting that the down-regulation of F may be a control mechanism for helping the virus to evade the host immune response (Lingemann et al., 2015).

The second hypothesis is that the greater abundance of M-F read-through transcripts may act to increase the access of RNA polymerase to the downstream genes HN and L (and SH in some paramyxoviruses). A study by Parks et al., (2001) using minigenomes in which the Ig of the M-F gene junction had been altered showed that an increase in the generation of M-F read-through mRNAs resulted in an increase in expression of the L protein, resulting in an decrease in viral growth and lower viral titres compared to WT PIV5. A study by He and Lamb., (1999) showed that switching the HN and L Igs led to greater expression of the L protein but conferred slower viral growth and smaller plaques. Additionally, Shabman et al., (2013) demonstrated that, for Ebola virus (filoviridae), under- or over-expression of L resulted in inhibition of viral transcription. However, studies using recombinant SeV and rabies virus (rhabdovirus) showed that an increase in the level of L increased virus growth (Finke et al., 2000; Kato et al., 1999). These studies suggest that the amount of L, and presumably the ratio of L to P and NP, is critical for optimal viral growth.

Furthermore, increasing access to downstream genes downstream from M/F may counteract the relatively short half-life of the HN protein, as shown by Leser et al., (1996).

6.1.5 RNA editing and transcription of the P/V/C/D(I) gene

The P/V/C gene of paramyxoviruses encodes multiple proteins, primarily by an RNA editing process that results in the insertion of non-templated G residues at the RNA editing site and causes translational frameshifts. It is thought that this occurs through RNA polymerase stuttering in a tract of G residues by a similar mechanism as that occurring during polyadenylation (Hausmann et al., 1999). This results in the production of the V, I and P mRNAs (rubulaviruses) generated from the insertion of 0 or 0+3 G residues, 1 or 1+3 G residues and 2 or 2+3 G residues, respectively. The respiroviruses (PIV3) produce the P, V and D transcripts from the insertion of 0 or 0+3 G residues, 1 or 1+3 G residues and 2 or 2+3 G residues, respectively. The respiroviruses also produce the C protein, but this is generated through an additional ORF close to the 3' end of the P gene and therefore cannot be quantified by HTS.

In general, for a given virus the proportions of transcripts generated via RNA editing remained stable throughout viral infection. A recent study by Qui et al (2016) demonstrated an increase in RNA editing at early times in NDV, suggesting that this stability might not apply to all paramyxoviruses. There were clear differences between the viruses in the ratio of mRNAs produced with different numbers of G residue insertions. Although work by Thomas et al., (1998) showed that for PIV5 the insertion of 0 and 2 G residues (producing V

and P mRNA) occurs at a ratio of approximately 1:1, the analysis of HTS data clearly demonstrated that PIV5 strain W3, as well as PIV2, displayed a ratio of 2:1 of V:P mRNA generated through 0 and 2 G residue inserts, respectively. These transcripts accounted for 98% of all the transcripts generated from the P/V gene, and the generation of I mRNAs accounted for <2%. In contrast, MuV demonstrated a more equivalent generation of V and P transcripts, and approximately 5% of mRNAs were I transcripts. PIV3 RNA editing processes generated transcripts containing 0, 1 and 2 G residues at a ratio of approximately 5:3:1 and generating the P, V and D mRNAs. This result for PIV3 is in contrast to that observed by Kolakofsky et al., (2005), who found PIV3 inserts of 1 to 6 G residues at equal frequency.

It is thought that the generation of transcripts in the observed proportions usually reflects the functional needs for the translated proteins. If this is the case, these results suggest that, for PIV2 and PIV5 strain W3, there is a requirement for more V protein than P. This may be because the V proteins are IFN antagonists and therefore it is important to produce them in a burst as early as possible in infection. In contrast, although the P protein is an essential component of the viral RNA polymerase complex, high level of expression of P may be required only once virus replication and secondary transcription have been initiated. Only a small number of transcripts would results in the I protein, which has a stop codon soon after the RNA editing site resulting in a protein with only a few unique amino acids after the N-terminal domain common to P/V/I. No function has been ascribed to the I protein, and, as I mRNA is only present at very low levels, it suggests that it may be non-essential for virus growth.

The generation of accessory protein transcripts by PIV3 is particularly interesting. Although an ancestral ORF is present in the V mRNA, there are two stop codons downstream of the editing site that would result in the production of a truncated V protein that is unlikely to function as an IFN antagonist. Consequently, it is generally accepted that V has no function in PIV3 infection. However, the results obtained from this analysis show that one third of P/D/C transcripts would encode the truncated V protein, a significantly higher proportion than would encode D, or by comparison, the truncated I proteins of PIV2 and PIV5. Such a high proportion of V mRNA generation suggests it may retain an important function in PIV3 replication. Although the primary function of the V protein of paramyxoviruses has been linked to the role as IFN antagonists, it may also play important roles in the control of viral transcription and replication. For example, several studies have shown that, for a number of paramyxoviruses, including SeV, PIV2 and PIV5, the V protein, through its N-terminal P/V common domain, can interact with NP and L to inhibit RNA synthesis (Horikami et al., 1996; Nishio et al., 2008). Also, a study by Precious et al., (1995) demonstrated that PIV5 V may have a role in keeping NP^0 soluble prior to encapsidation of the viral genome or antigenome, and thus the V protein of PIV3 may have a similar function.

The analysis of RNA editing, in particular the insertions of higher proportions of 2 G residues compared to 1 G residue by PIV2 and PIV5 strain W3, emphasizes the precision of the RNA editing mechanism by the viral RNA polymerases. Hausmann et al (1999) demonstrated that the cis-acting element that directs insertion of the G residues is a 6 nt sequence upstream of the AnGn

tract of the RNA editing site. The tri-nucleotide immediately upstream of this tract is of particular importance, as mutations in it alter the frequency at which G residues are inserted at the RNA editing site. Iseni et al., (2002) also suggested that hexamer phasing may play an important role in the ability of cis-acting elements to direct RNA polymerase activity during transcription, including RNA editing although this has not been characterized. Kolakofsky (2016) suggested that as the NP monomers are displaced from the nucleocapsid complex by the RNA polymerase they are known to move over the surface of the L subunit and reform the nucleocapsid complex as the template viral RNA exits the active cavity of the RNA polymerase and it is the NP monomers interaction with L (the catalytic subunit of the polymerase) as they move across the surface of L that dictates the RNA polymerase activity namely stuttering and pausing at specific sequnces. The cis-acting elements responsible for RNA editing must be in the correct hexamer phase position for efficient RNA polymerase function. The RNA editing and hexamer phasing of the rubulaviruses is highly conserved (3′-A_3UUCUC_4-5', with the hexamer phasing 3′-NAAAUU CUCCCC-5'). These conserved elements may reflect a fitness selection of the virus to allow high rates of transcription for mRNAs that contain 2 G inserts and whose translated proteins have an essential role in viral growth. Interestingly, the 3 nt upstream of the conserved editing site, which are within the 6 nt cis-acting sequence identified by Hausmann et al., (1999), are different between PIV5, PIV2 and MuV (CUA, UAU and CUU. respectively). These tri-nucleotide sequences upstream of the highly conserved RNA editing site may determine the differences in the insertion of G residues observed among the rubulaviruses. Conversely, there is no conservation of hexamer phasing in the respiroviurses, and that this is reflected in the non-conserved nature of the insertion of G

residues at the RNA editing site that differs between viruses in this genus. This can be observed in SeV where the insertion of 1 G residue is made at a higher frequency whilst human PIV3 the frequency of 1 to 6 G insertions was suggested to occur at a ratio of 1:1 (Galinski et al., 1992; Kolakofsky et al., 2005; Pelet et al., 1991). However, the results obtained here demonstrate that this is not the case for human PIV3, and that the proportion of G insertions for this virus reflects a more precise mechanism of RNA editing by the RNA polymerase.

Taken together, the above analysis suggests that there is no temporal control of paramyxovirus transcription. RNA polymerase processivity directed by cis-acting elements in PIV5 strain W3, PIV2, MuV and PIV3 remained stable throughout the infection, with the relative abundances of viral mRNAs (templated and non-templated), remaining relatively stable. The differences in the transcription profiles of the viruses may indicate that cis-acting elements directing the activity of the RNA polymerase have evolved to generate abundances of viral mRNAs, and therefore of their translated proteins, that reflect the growth requirements of the virus. Paramyxovirus transcription and replication are directed by cis-acting elements such as the gene start (Gs), intergenic region (Ig) and gene end (Ge). The cis-acting elements in the rubulaviruses and respiroviruses differ significantly. The sequences of the Gs and Ge of the genes found in the viruses used in this study are shown in Appendix IV. These show that the rubulavirus Gs sequences uniformly start with an a residue. The individual viruses contain between 1 and 4 conserved residues compared to the respirovirus PIV3 Gs that contains 8 conserved residues. The Igs of the rubulaviruses differ greatly in their sequences. The

PIV5 strain W3 and PIV2 show that V/P-M and M-F and HN-L contain longer Igs compared to the other genes. In comparison MuV contains shorter Igs. PIV3 contains a trinucleotide 3'-GAA-5' Ig. Interestingly, the Ge of all of the viruses show very little conservation which could account for the differences in the rate of read-through between the different genes. However, this does not offer an insight into the observed higher rate of read-through between the M-F genes observed in all four viruses. The identification of sequences that are essential to maintaining the virus transcription profiles is beyond the scope of this project and requires further investigation. Studies by Rassa and Parks (1998-2001) suggested that cis-acting elements such as the Ge and the length of the U-tract may work in tandem to direct RNA polymerase activity. Kolakofsky et al., (2005, 2016) and Iseni (2002) suggested that hexamer phasing is crucial for the RNA polymerase to recognize the cis-acting elements that direct viral transcription and RNA editing. Hexamer phasing has been shown to be crucial in initiation of RNA synthesis at the Le promoter for paramyxoviruses (see Introduction section 1.2). The differences between the cis-acting elements of viruses in different paramyxovirus genera may indicate the employment of different mechanisms of control, even though this may result in similar outcomes as was observed for the increased read-through at the M-F genes. These cis-acting elements do not seem to elicit a temporal control of RNA polymerase processivity, indicating a role for other factors, such as host cell factors that could be involved in changes to viral transcription kinetics during infection.

The role of cellular proteins that could play a role of paramyxovirus transcription and replication is not well characterized. Studies have shown that heat shock protein 70 and 90 (Hsp70 and Hsp90) played a role in MeV and MuV

polymerase maturation through interactions with the L protein, inhibition of which resulted in a reduction in the virus ability to replicate (Bloyet et al., 2016; Connor et al 2007; Katoh et al., 2016). Other studies have shown that cellular kinases such as PLK-1 Akt-1 can inhibit virus RNA synthesis through phosphorylation of the P protein, the role of these shall be discussed in greater detail below in section 6.3.

Together, the analysis described above may provide the opportunity to generate a more accurate virus gene expression profile by combining the FPKM values with the percentage RNA editing and read-through values. However, due to time constraints this could not be attempted here but shall be included in a larger study which is currently being undertaken.

6.2 Ratio of genome to antigenome during viral replication

The analysis of the ratios of genome to antigenome conducted on PIV5 strain W3 demonstrated some interesting preliminary observations of the dynamics of viral replication at early times in infection. The analysis presented here suggests a complete reversal of the genome to antigenome ratio between the point of infection and early times in infection. Input virus displayed a genome:antigenome ratio of 9:1, but by 6 h p.i. this had almost completely reversed to a ratio of 1:4. This finding is supported by the study of Irie et al., (2014), which demonstrated a similar change in SeV infections. This was correlated to the presence of the viral C protein, with the Le promoter being stronger that the TrC promoter in viruses lacking C expression. However, in the

presence of C this was reversed so that the TrC promoter was stronger. Thus, at early times in infection, prior to the synthesis of large amounts of the C protein, the Le replication promoter may be stronger, resulting in more antigenomes than genomes being produced. After transcription and translation of the C gene, the strength of the TrC promoter may increase, resulting in the generation of more genomes than antigenomes. However, PIV5 and other rubulaviruses do not encode a C protein, and therefore a different viral protein with a similar function may be responsible for the observed change. Alternatively, the switch in the ratio of genomes to antigenomes may be due to modifications to the RNA polymerase at early versus late times in infection. However, modifications of the RNA polymerase could possibly result in changes in viral transcription dynamics over time, as both processes use the same promoter and RNA polymerase complex. However, the results obtained during this study showed that the processivity of the RNA polymerase remains remarkably stable throughout the course of a 96 h infection. Therefore, modifications in the RNA polymerase are an unlikely source of the changes in genome to antigenome ratio. Alternatively, the switch may be the result of multiple RNA polymerases simultaneously functioning on a single viral genome template. This may result in an initial accelerated proliferation of antigenomes compared to genomes.

6.3 Transcription and replication are repressed during establishment of PIV5 persistence

Initial phase contrast microscopy and immunofluorescence experiments showed that the majority of cells infected with PIV5 strain W3 survived the

infection. HTS revealed that, after 24 h p.i., viral genomes were maintained throughout the infection at a relatively constant level, accounting for approximately 1-3% of total cellular reads. HTS analysis and radioactive labeling of infected cells demonstrated that viral transcription and protein synthesis were progressively shut off after 12-24 h p.i. Thus, although at 12 h p.i. viral mRNA reads accounted for 4.5% of total cellular reads, by 96 h p.i. they represented <0.2%. The conclusion that viral transcription had been shut off in the majority of cells at late times in infection was supported by immunostaining the cells for the presence of the HN protein. Unlike the NP protein, which is stable for days, HN has a relatively short half-life of a few hours (Leser et al., 1996) and thus the observation that cells were positive for NP but negative for HN at 96 h p.i. indicated that viral transcription was repressed in HN-negative cells. However, the pattern of the viral transcriptional gradient remained relatively similar over this time period, suggesting that changes in RNA polymerase processivity were not responsible for this shut-off. Preliminary results obtained from conducting an identical time course experiment in cells in which an IFN response had been established by pretreating the cells with IFN prior to infection with PIV5 strain W3 suggested that the repression of protein synthesis at later times was not IFN-dependent, as infected cells either pretreated with IFN or not pretreated exhibited a similar "shut-off" by 96 h p.i. This experiment would need to be repeated in order to obtain a reproducible result. Additionally, the presence of DIs has been shown to play a role in the repression of viral protein synthesis (Huang and Baltimore, 1970; Roux et al., 1991). However, bioinformatic analysis showed no evidence for the generation of DIs during PIV5 strain W3 infection over a 96 h period.

PIV5 strain CPI+ shares >99% nucleotide sequence identity with PIV5 strain W3 (strain CPI+ lacks the SH gene). However, phase contrast microscopy and immunofluorescence experiments revealed that the majority of cells infected with PIV5 strain CPI+ die by 96 h p.i. and, unlike PIV5 strain W3, persistent infections are not easily established. Preliminary results obtained from HTS analysis of PIV5 strain CPI+ showed that the kinetics of transcription are significantly different from those of PIV5 strain W3. The abundance of viral mRNAs peaked at 18 h p.i., at 17.3% of total cellular reads. This represents a four-fold increase in peak viral mRNA abundances compared to PIV5 strain W3. By 96 h p.i., the abundance of viral mRNAs in the few surviving CPI+-infected cells was still higher than that of viral mRNAs expressed by PIV5 strain W3 at its peak transcription rate, accounting for 8.86% of total cellular reads. An analysis of viral transcription similar to that described above was also conducted on PIV5 strain CPI+-infected cells, in order to investigate the transcriptional gradient, the efficiency of RNA editing and the abundance of read-through mRNAs. However, no obvious difference between the overall patterns of transcription for strains W3 and CPI+ was observed that could account for the higher level of viral mRNA generation in CPI+-infected cells or why this was not repressed at later times in infection, as was observed for PIV5 strain W3.

Differences in the overall levels of viral transcripts may reflect differences in the cytopathogenic nature of the infection. Previous observations in Professor Randall's laboratory have demonstrated that the majority of cells infected at a high MOI with PIV2, MuV and PIV3 die by 48 h p.i. The results obtained in this study show a similar outcome for cells infected with PIV5 strain CPI+, whereas

the majority of PIV5 strain W3-infected cells survived the infection and easily established a persistent infection. The more than four-fold increase between the peak abundance of mRNA reads observed in PIV2, MuV, PIV3 and PIV5 strain CPI+ compared to PIV5 strain W3 may reflect a more cytopathogenic infection, in which RNA synthesis is not repressed, resulting the induction of cytokines and ultimately cell death.

Although most cells infected with PIV5 strain CPI+ died at late times in infection, a small proportion of cells survived and eventually began to grow while being persistently infected with the virus. To determine whether any viral mutations were selected during the establishment of CPI+ persistently infected cells, perhaps resulting in viruses exhibiting an inhibition of viral transcription at later times in infection, the reads generated from the viral genome of PIV5 strain CPI+ in persistently infected cells were analysed. A single mutation was identified in the L gene at the nucleotide 13093 (amino acid residue 1560 of the L protein) which would result in a phenylalanine to leucine change in the protein. However, this mutation was not present in any of the reads aligning at that genome position at 96 h p.i. during the PIV5 strain CPI+ time course analysis, and therefore is unlikely to represent a population of viruses that had been selected for its ability to establish persistence. HTS analysis showed that the abundance of viral genomes in PIV5 strain CPI+ persistently infected cells accounted for 4% of total cellular reads, whereas in W3 persistently infected cells it accounted for 2.5%. However, the abundance of viral mRNAs in these cells contributed 6.3% of the total reads compared to 0.03% in W3 persistently infected cells. There were slight differences in the transcriptional gradient between strain W3- and strain CPI+-infected cells that may result from

differences in the mechanisms by which the two viruses establish persistent infections. Slight differences in the abundance of viral mRNAs may be the result of differences in the stability and translation of mRNAs. As this was a single experiment, it would need to be replicated to ensure reproducible results. Strikingly, although no DIs were detected in PIV5 strain W3 persistently infected cells, they accounted for approximately 4% of the viral reads in PIV5 strain CPI+-infected cells. Given that DIs have previously been linked to repression of RNA synthesis and the establishment of persistent infections (Huang and Baltimore, 1970; Roux et al., 1991), it may be the presence of PIV5 strain CPI+ DIs reduces viral replication in CPI+-infected cells to a level compatible with cell survival and the establishment of a persistent infection. A similar mechanism for the establishment of persistence was recently identified by Xu et al., (2017), which showed that in SeV infection although the majority of cells die a small sub population of cells infected with DIs survive by inducing pro-survival factors TNF22/TRAF1 resulting in the establishment of a persistent infection. Although there was no clear evidence for the presence of DIs at 96 h p.i. following a high MOI infection with PIV5 strain CPI+, >99% of infected cells died and DIs may have been generated in the small proportion of surviving cells at a low level that precluded their detection during bioinformatic analysis. Collectively, these results indicate that, despite the high percentage of sequence identity between the two PIV5 strains, different mechanisms may have been involved in establishing persistence.

6.3.1 Analysis of differences in the transcription kinetics of PIV5 strains W3 and CPI+

A major question that arose from this work was the molecular basis for differences in the kinetics of viral transcription between strains CPI+ and W3 and for their different abilities to establish persistent infections. Since numerous studies have shown a role for the P protein as a regulator of viral RNA synthesis (Fuentes et al., 2010) any differences in the P proteins between these strains were of particular interest. Of note was a difference at amino acid residue 157 of the P protein. PIV5 strain W3 contained a serine residue at this position, whereas PIV5 strain CPI+ contained a phenylalanine. This was particularly interesting, as work by Sun et al., (2009) showed that a serine at position 157 could be phosphorylated by cellular kinase PLK-1, resulting in a reduction in viral gene expression. PLK-1 is a serine/threonine protein kinase that plays an essential role in the modulation of the cell cycle (Eckerdt and Strebhardt, 2006). It binds to its target by a SSP motif, which is present at position 157. To investigate whether this difference explains the biological properties of strains W3 and CPI+, a recombinant virus was produced by the Professor Steve Goodbourn laboratory using the PIV5 strain W3 backbone but containing a phenylalanine residue at position 157 of the P protein. This virus, rPIV5-W3(F157), displayed no shut-off of transcription of protein synthesis at later times in infection, and, furthermore, the majority of infected cells had died by 72 h p.i. This result strongly suggests that phosphorylation of a serine at position 157 of the P protein is primarily responsible for the observed transcriptional shut-off at later times during PIV5 strain W3 infection, and that it plays a significant role in the establishment of a persistent infection. Additionally, another cellular serine/threonine kinase, AKT1, has been shown to play a role

in the regulation of viral gene expression, but a direct interaction between this kinase and the P protein was not demonstrated (Sun et al., 2008). Interestingly, a study by Pickar et al., (2014) showed a similar result for MuV, in which phosphorylation of a threonine residue at position 101 of the P protein was shown to inhibit viral RNA synthesis. Any future study could include the use of phosphomimetics, which utilizes the structural similarity of amino acids which cannot be phosphorylated to other phosphorylated amino acids namely serine, tyrosine and threonine. By replacing the latter with the former, the non phosphorylated amino acids mimicks the phosphorylated amino acid with minimal structural changes to the protein thereby allowing characterization of the role of phosphorylation in the activity of the protein. This method could be employed to better characterise the phosphorylation of the P protein at specific sites that could play a role in the establishment of persistence in PIV5 and/or other paramyxoviruses.

6.4 Role of inclusion bodies in the establishment of persistence

Preliminary immunofluorescence results demonstrated the formation of inclusion bodies at later times in infection with PIV5 strain W3. The production of inclusion bodies in PIV5 strain W3 may be correlated to the repression of viral RNA synthesis observed at later times in infection. A study by Chatziandreou et al., (2002) demonstrated that, in cells persistently infected with PIV5 strain CPI-, which is sensitive to IFN, RNA synthesis was repressed and the viral nucleocapsids formed inclusion bodies, thus possibly indicating a role in the inclusion body formation for viruses in which RNA synthesis is

repressed for genome maintenance. The P protein has been shown to play a role in the formation of inclusion bodies in PIV5 and PIV3, as previously discussed (Fearns et al., 1996; Carlos et al., 2009). Therefore, it is possible that phosphorylation of the P protein and the role of this protein as a subunit of the RNA polymerase that interacts with NP-RNA results in an inhibition of viral transcription, thereby resulting in the formation of inclusion bodies that provide stable environments for viral genome maintenance with little or no active viral transcription and the establishment of a persistent infection. Inclusion bodies were observed forming at later times during MuV infection. Although, the MuV Enders strain displayed a higher rate of viral mRNAs generation compared to PIV5 strain W3, and a higher rate of cell death, the formation of inclusion bodies may indicate the selection of viruses in which virus RNA synthesis is repressed in order to establish a persistent infection in a small population of surviving cells.

Appendix

Appendix I

PIV5 W3

	replicate 1			replicate 2			replicate 3		
	Number of reads								
h p.i.	total	genome	mRNA/ antigenome	total	genome	mRNA/ antigenome	total	genome	mRNA/ antigenome
6	24545471	11173	138513	974158	301	3617	11241664	1014	19136
12	18849455	36268	430352	25764654	9150	560326	16162663	21008	407089
18	12081460	161454	484769	7915267	49911	359860	15159506	15155	433554
24	23949505	393355	593056	10010849	236496	185184	36672534	286058	620565
48	14482897	302389	59625	18630069	588167	71790	15779643	437731	67635
96	13047657	153064	7478	16998498	256430	19781	11154444	226663	25796

PIV5 CPI+

	Number of reads		
h p.i.	total	genome	mRNA/ antigenome
6	1556041	447	2446
12	1783243	4813	86616
18	1875001	10302	321583
24	2908255	7301	474134
48	3151489	11413	544228
96	2839633	135840	251488

PIV2

	replicate 1			replicate 2			replicate 3		
	Number of reads								
h p.i.	total	genome	mRNA/ antigenome	total	genome	mRNA/ antigenome	total	genome	mRNA/ antigenome
6	8018801	7769	28397	7163318	6267	21888	7728774	9765	29422
12	8478801	27336	616243	9292656	24211	663075	9292656	38367	667594
18	9714622	92686	667594	11035478	112434	1609576	8746096	77701	1267845
24	8746096	77701	1267845	10311290	227561	1700395	11893320	208620	1941779

MuV

	replicate 1			replicate 2			replicate 3		
	Number of reads								
h p.i.	total	genome	mRNA/ antigenome	total	genome	mRNA/ antigenome	total	genome	mRNA/ antigenome
6	9460206	7811	40321	9556086	8707	51659	11296106	8868	52656
12	13146022	56772	1252717	10698006	36931	967728	35266827	115396	3219964
18	8179449	58678	1346625	18946597	108082	3079500	9683981	65887	1518073
24	11476071	128810	1496159	10896027	96173	1505079	14134696	146000	2481818

PIV3

	replicate 1			replicate 2			replicate 3		
	Number of reads								
h p.i.	total	genome	mRNA/ antigenome	total	genome	mRNA/ antigenome	total	genome	mRNA/ antigenome
6	17771226	70198	1626468						
12	16383759	163150	2628102	10508456	163150	1719353	9991857	108441	1812590
18	30097446	605364	4574802	21375979	339993	4136137	28506916	372510	3112494
24	7596687	118325	934804	11619756	151691	951656	16413354	256176	1697518

Appendix II

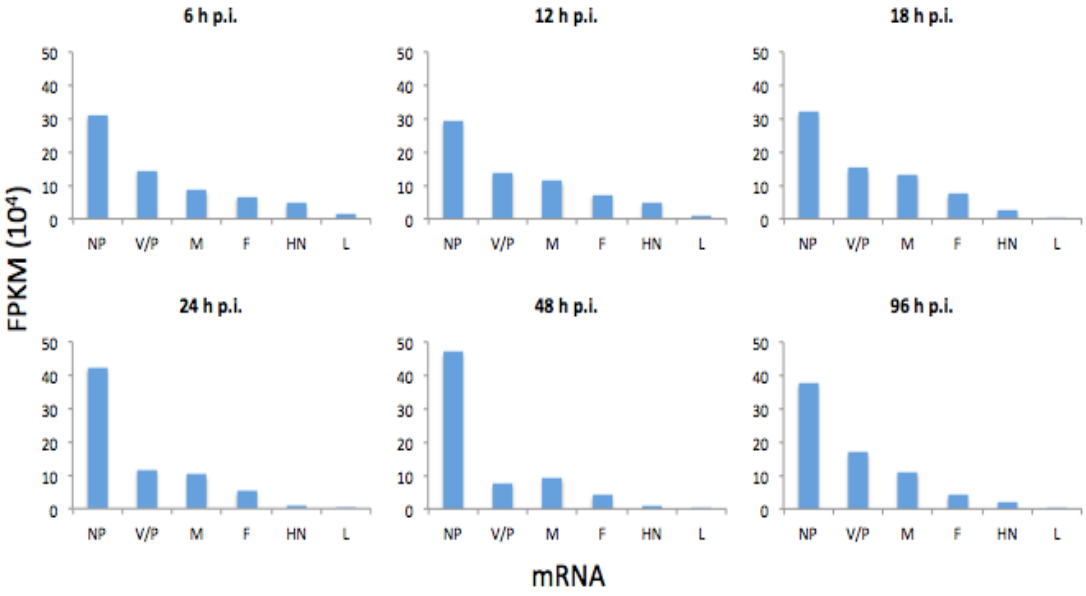


Fig I .Transcription of PIV5 strain CPI+ genes at 6, 12, 18 , 24, 48 and 96 h pi.. The data were directionally analysed using the established bioinformatic pipeline. The FPKM values on the *y*-axes were obtained using the RSEM software

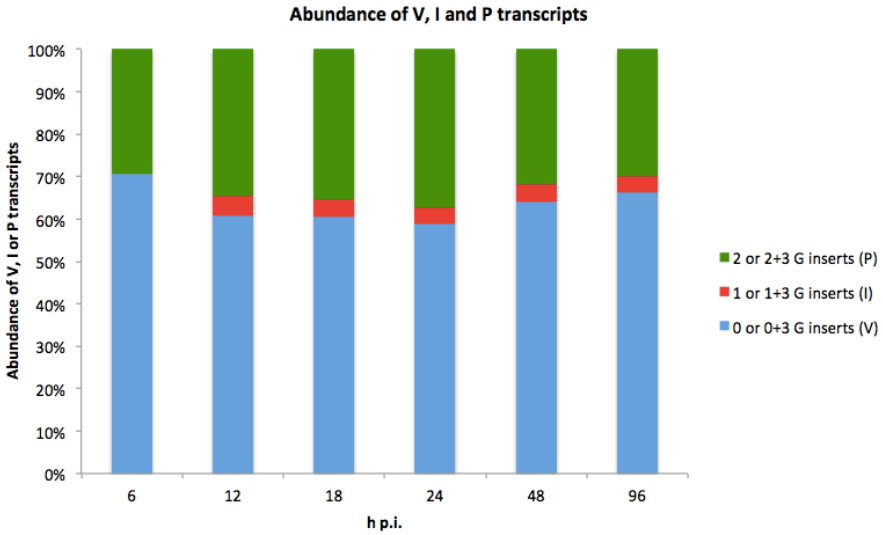
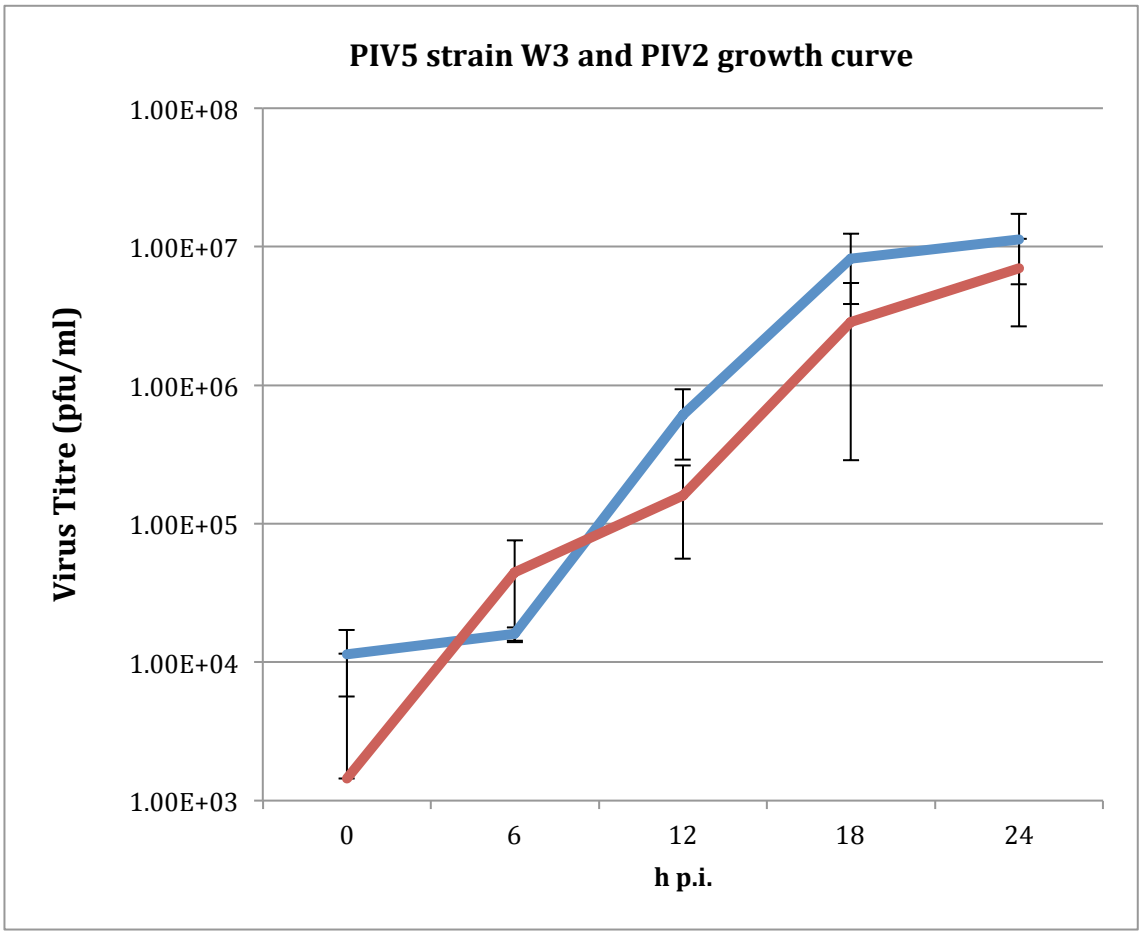


Fig II. Abundance of V, P and I mRNAs generated by PIV5 strain W3 during infection. The number of reads that were generated from the V, P and I mRNAs were calculated: 0 and 0+3 G inserts (V), 2 and 2+3 G inserts (P) and 1 and 1+3 G inserts (I) from the R1 and R2 files using a 10 nt search string upstream and downstream of the RNA editing site.

Appendix III



Appendix IV – Table showing the Gs and Ge sequence of each gene, including the length of the U tract depicted by A followed by the number of As found in the U-tract and the length of the Ig

PIV5 W3	gene	gene start	gene end	length of Ig (Ge to downstream Gs)
	NP	aggtccggaa	ggctttaaagA^{7}	1
	V/P	aggcccggac	agggttttagA^{6}	16
	M	aagcccgaac	tgattcaaagA^{4}	23
	F	aagcacgaac	aatttttaagA^{7}	4
	SH	aggaccgaac	tattttaaagA^{6}	1
	HN	aggcccgaac	tggttttaagA^{6}	13
	L	aggccagaat	ttagtttaagA^{6}	-
	Consensus	a*g***g*a*	****t***ag	

PIV2	gene	gene start	gene end	length of Ig (Ge to downstream Gs)
	NP	agattccggt	ataatttaagA^{6}	3
	V/P	aggcccggac	taaatttaatA^{7}	46
	M	aggacacaat	tcaatctaacA^{7}	28
	F	aggccaaaat	acaatttaagA^{6}	8
	HN	aggtgtcgta	gatatttaagA^{6}	42
	L	aggccagaat	tttatttaagA^{5}	-
	Consensus	ag********	***at*taa*	

MuV	gene	gene start	gene end	length of Ig (Ge to downstream Gs)
	NP	aagcccggaa	agtctttaagA^{6}	2
	V/P	aggcccggaa	cctattaaatA^{6}	1
	M	aagcacgaac	gaaattatagA^{6}	1
	F	aagcctagaa	tcctgcaattA^{6}	7
	SH	aagaatgaat	agacataaagA^{6}	4
	HN	agccagaaca	gattattaagA^{6}	1
	L	aggccagaat	tccttttaagA^{6}	-
	Consensus	a*********	**********	

PIV3	gene	gene start	gene end	length of Ig (Ge to downstream Gs)
	NP	aggattaaag	aataaataagA^{5}	3
	P/D/C	aggattaaag	tacaaataagA^{5}	3
	M	aggattaaag	aaggataatcA^{5}	3
	F	aggacaaaag	ttaaaattatA^{6}	3
	HN	aggagtaaag	agggaaatatA^{6}	3
	L	aggagcaaag	aaaaagtaagA^{5}	-
	Consensus	agga**aaag	****a*****	

References

ABRAHAM, G. & BANERJEE, A. K. 1976. Sequential transcription of the genes of vesicular stomatitis virus. *Proc Natl Acad Sci U S A,* 73**,** 1504-8.

AIRD, D., ROSS, M. G., CHEN, W. S., DANIELSSON, M., FENNELL, T., RUSS, C., JAFFE, D. B., NUSBAUM, C. & GNIRKE, A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol,* 12**,** R18.

ALAYYOUBI, M., LESER, G. P., KORS, C. A. & LAMB, R. A. 2015. Structure of the paramyxovirus parainfluenza virus 5 nucleoprotein-RNA complex. *Proc Natl Acad Sci U S A,*112**,** E1792-9.

ALBERTINI, A. A., CLAPIER, C. R., WERNIMONT, A. K., SCHOEHN, G., WEISSENHORN, W. & RUIGROK, R. W. 2007. Isolation and crystallization of a unique size category of recombinant Rabies virus Nucleoprotein-RNA rings. *J Struct Biol,* 158**,** 129-33.

ALBERTINI, A. A., WERNIMONT, A. K., MUZIOL, T., RAVELLI, R. B., CLAPIER, C. R., SCHOEHN, G., WEISSENHORN, W. & RUIGROK, R. W. 2006. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science,* 313**,** 360-3.

ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biol,* 11**,** R106.

ANDREJEVA, J., CHILDS, K. S., YOUNG, D. F., CARLOS, T. S., STOCK, N., GOODBOURN, S. & RANDALL, R. E. 2004. The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN-beta promoter. *Proc Natl Acad Sci U S A,* 101**,** 17264-9.

ANDREJEVA, J., NORSTED, H., HABJAN, M., THIEL, V., GOODBOURN, S. & RANDALL, R. E. 2013. ISG56/IFIT1 is primarily responsible for interferon-induced changes to patterns of parainfluenza virus type 5 transcription and protein synthesis. *J Gen Virol,* 94**,** 59-68.

ARIMILLI, S., ALEXANDER-MILLER, M. A. & PARKS, G. D. 2006. A simian virus 5 (SV5) P/V mutant is less cytopathic than wild-type SV5 in human dendritic cells and is a more effective activator of dendritic cell maturation and function. *J Virol,* 80**,** 3416-27.

AUDSLEY, M. D. & MOSELEY, G. W. 2013. Paramyxovirus evasion of innate immunity: Diverse strategies for common targets. *World J Virol,* 2**,** 57-70.

BALL, L. A. & WHITE, C. N. 1976. Order of transcription of genes of vesicular stomatitis virus. *Proc Natl Acad Sci U S A,* 73**,** 442-6.

BANERJEE, A. K. 1987. The transcription complex of vesicular stomatitis virus. *Cell,* 48**,** 363-4.

BANERJEE, A. K. 2008. Response to "Non-segmented negative-strand RNA virus RNA synthesis in vivo". *Virology,* 371**,** 231-3.

BANERJEE, A. K. & BARIK, S. 1992. Gene expression of vesicular stomatitis virus genome RNA. *Virology,* 188**,** 417-28.

BANERJEE, A. K., BARIK, S. & DE, B. P. 1991. Gene expression of nonsegmented negative strand RNA viruses. *Pharmacol Ther,* 51**,** 47-70.

BANKAMP, B., WILSON, J., BELLINI, W. J. & ROTA, P. A. 2005. Identification of naturally occurring amino acid variations that affect the ability of the measles virus C protein to regulate genome replication and transcription. *Virology,* 336**,** 120-9.

BARON, M. D. & BARRETT, T. 2000. Rinderpest viruses lacking the C and V proteins show specific defects in growth and transcription of viral RNAs. *J Virol,* 74**,** 2603-11.

BARR, J. N., WHELAN, S. P. & WERTZ, G. W. 1997. cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. *J Virol,* 71**,** 8718-25.

BASKARAN, N., KANDPAL, R. P., BHARGAVA, A. K., GLYNN, M. W., BALE, A. & WEISSMAN, S. M. 1996. Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. *Genome Res,* 6**,** 633-8.

BENITA, Y., OOSTING, R. S., LOK, M. C., WISE, M. J. & HUMPHERY-SMITH, I. 2003. Regionalized GC content of template DNA as a predictor of PCR success. *Nucleic Acids Res,* 31**,** e99.

BERGHÄLL, H., SIRÉN, J., SARKAR, D., JULKUNEN, I., FISHER, P. B., VAINIONPÄÄ, R. & MATIKAINEN, S. 2006. The interferon-inducible RNA helicase, mda-5, is involved in measles virus-induced expression of antiviral cytokines. *Microbes Infect,* 8**,** 2138-44.

BLASZCZYK, K., NOWICKA, H., KOSTYRKO, K., ANTONCZYK, A., WESOLY, J. & BLUYSSEN, H. A. 2016. The unique role of STAT2 in constitutive and IFN-induced transcription and antiviral responses. *Cytokine Growth Factor Rev,* 29**,** 71-81.

BLUMBERG, B. M., LEPPERT, M. & KOLAKOFSKY, D. 1981. Interaction of VSV leader RNA and nucleocapsid protein may control VSV genome replication. *Cell,* 23**,** 837-45.

BOONYARATANAKORNKIT, J., BARTLETT, E., SCHOMACKER, H., SURMAN, S., AKIRA, S., BAE, Y. S., COLLINS, P., MURPHY, B. &

SCHMIDT, A. 2011. The C proteins of human parainfluenza virus type 1 limit double-stranded RNA accumulation that would otherwise trigger activation of MDA5 and protein kinase R. *J Virol,* 85**,** 1495-506.

BOURHIS, J. M., RECEVEUR-BRÉCHOT, V., OGLESBEE, M., ZHANG, X., BUCCELLATO, M., DARBON, H., CANARD, B., FINET, S. & LONGHI, S. 2005. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. *Protein Sci,* 14**,** 1975-92.

BOUSSE, T., MATROSOVICH, T., PORTNER, A., KATO, A., NAGAI, Y. & TAKIMOTO, T. 2002. The long noncoding region of the human parainfluenza virus type 1 f gene contributes to the read-through transcription at the m-f gene junction. *J Virol,* 76**,** 8244-51.

BOUSSE, T., TAKIMOTO, T., MURTI, K. G. & PORTNER, A. 1997. Elevated expression of the human parainfluenza virus type 1 F gene downregulates HN expression. *Virology,*232**,** 44-52.

BUCHHOLZ, C. J., RETZLER, C., HOMANN, H. E. & NEUBERT, W. J. 1994. The carboxy-terminal domain of Sendai virus nucleocapsid protein is involved in complex formation between phosphoprotein and nucleocapsid-like particles. *Virology,* 204**,** 770-6.

BYRAPPA, S., PAN, Y. B. & GUPTA, K. C. 1996. Sendai virus P protein is constitutively phosphorylated at serine249: high phosphorylation potential of the P protein. *Virology,*216**,** 228-34.

CADD, T., GARCIN, D., TAPPAREL, C., ITOH, M., HOMMA, M., ROUX, L., CURRAN, J. & KOLAKOFSKY, D. 1996. The Sendai paramyxovirus accessory C proteins inhibit viral genome amplification in a promoter-specific fashion. *J Virol,* 70**,** 5067-74.

CALAIN, P., MONROE, M. C. & NICHOL, S. T. 1999. Ebola virus defective interfering particles and persistent infection. *Virology,* 262**,** 114-28.

CALAIN, P. & ROUX, L. 1993. The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *J Virol,* 67**,** 4822-30.

CALAIN, P., ROUX, L. & KOLAKOFSKY, D. 2016. Defective interfering genomes and Ebola virus persistence. *Lancet,* 388**,** 659-60.

CANTAERT, T., BAETEN, D., TAK, P. P. & VAN BAARSEN, L. G. 2010. Type I IFN and TNFα cross-regulation in immune-mediated inflammatory disease: basic concepts and clinical relevance. *Arthritis Res Ther,* 12**,** 219.

CARLOS, T. S., YOUNG, D. F., SCHNEIDER, M., SIMAS, J. P. & RANDALL, R. E. 2009. Parainfluenza virus 5 genomes are located in viral

cytoplasmic bodies whilst the virus dismantles the interferon-induced antiviral state of cells. *J Gen Virol,* 90**,** 2147-56.

CATTANEO, R., REBMANN, G., BACZKO, K., TER MEULEN, V. & BILLETER, M. A. 1987a. Altered ratios of measles virus transcripts in diseased human brains. *Virology,* 160**,**523-6.

CATTANEO, R., REBMANN, G., SCHMID, A., BACZKO, K., TER MEULEN, V. & BILLETER, M. A. 1987b. Altered transcription of a defective measles virus genome derived from a diseased human brain. *EMBO J,* 6**,** 681-8.

CHANDRIKA, R., HORIKAMI, S. M., SMALLWOOD, S. & MOYER, S. A. 1995. Mutations in conserved domain I of the Sendai virus L polymerase protein uncouple transcription and replication. *Virology,* 213**,** 352-63.

CHATTOPADHYAY, S., ESPER, F. & BANERJEE, A. K. 2011. *The Biology of Paramyxoviruses,* USA, Caister Academic Press.

CHATZIANDREOU, N., YOUNG, D., ANDREJEVA, J., GOODBOURN, S. & RANDALL, R. E. 2002. Differences in interferon sensitivity and biological properties of two related isolates of simian virus 5: a model for virus persistence. *Virology,* 293**,** 234-42.

CHILDS, K., RANDALL, R. & GOODBOURN, S. 2012. Paramyxovirus V proteins interact with the RNA Helicase LGP2 to inhibit RIG-I-dependent interferon induction. *J Virol,* 86**,**3411-21.

CHILDS, K., STOCK, N., ROSS, C., ANDREJEVA, J., HILTON, L., SKINNER, M., RANDALL, R. & GOODBOURN, S. 2007. mda-5, but not RIG-I, is a common target for paramyxovirus V proteins. *Virology,* 359**,** 190-200.

CHILDS, K. S., ANDREJEVA, J., RANDALL, R. E. & GOODBOURN, S. 2009. Mechanism of mda-5 Inhibition by paramyxovirus V proteins. *J Virol,* 83**,** 1465-73.

CHU, Y. & YANG, X. 2011. SUMO E3 ligase activity of TRIM proteins. *Oncogene,* 30**,** 1108-16.

COLLINS, P. L., HILL, M. G., CRISTINA, J. & GROSFELD, H. 1996. Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus. *Proc Natl Acad Sci U S A,* 93**,** 81-5.

COLLINS, P. L. & WERTZ, G. W. 1983a. cDNA cloning and transcriptional mapping of nine polyadenylylated RNAs encoded by the genome of human respiratory syncytial virus. *Proc Natl Acad Sci U S A,* 80**,** 3208-12.

COLLINS, P. L. & WERTZ, G. W. 1983b. cDNA cloning and transcriptional mapping of nine polyadenylylated RNAs encoded by the genome of

human respiratory syncytial virus. *Proc Natl Acad Sci U S A,* 80**,** 3208-12.

CORDEY, S. & ROUX, L. 2007. Further characterization of a paramyxovirus transcription initiation signal: search for required nucleotides upstream and importance of the N phase context. *J Gen Virol,* 88**,** 1555-64.

CUI, X. F., IMAIZUMI, T., YOSHIDA, H., BORDEN, E. C. & SATOH, K. 2004. Retinoic acid-inducible gene-I is induced by interferon-gamma and regulates the expression of interferon-gamma stimulated gene 15 in MCF-7 cells. *Biochem Cell Biol,* 82**,** 401-5.

CURRAN, J., BOECK, R. & KOLAKOFSKY, D. 1991. The Sendai virus P gene expresses both an essential protein and an inhibitor of RNA synthesis by shuffling modules via mRNA editing. *EMBO J,* 10**,** 3079-85.

CURRAN, J. & KOLAKOFSKY, D. 1999. Replication of paramyxoviruses. *Adv Virus Res,* 54**,** 403-22.

CURRAN, J., MARQ, J. B. & KOLAKOFSKY, D. 1992. The Sendai virus nonstructural C proteins specifically inhibit viral mRNA synthesis. *Virology,* 189**,** 647-56.

CURRAN, J., MARQ, J. B. & KOLAKOFSKY, D. 1995a. An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J Virol,* 69**,** 849-55.

CURRAN, J., MARQ, J. B. & KOLAKOFSKY, D. 1995b. An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J Virol,* 69**,** 849-55.

DE, B. P., HOFFMAN, M. A., CHOUDHARY, S., HUNTLEY, C. C. & BANERJEE, A. K. 2000. Role of NH(2)- and COOH-terminal domains of the P protein of human parainfluenza virus type 3 in transcription and replication. *J Virol,* 74**,** 5886-95.

DELENDA, C., HAUSMANN, S., GARCIN, D. & KOLAKOFSKY, D. 1997a. Normal cellular replication of Sendai virus without the trans-frame, nonstructural V protein. *Virology,*228**,** 55-62.

DELENDA, C., HAUSMANN, S., GARCIN, D. & KOLAKOFSKY, D. 1997b. Normal cellular replication of Sendai virus without the trans-frame, nonstructural V protein. *Virology,*228**,** 55-62.

DEVAUX, P., PRINISKI, L. & CATTANEO, R. 2013. The measles virus phosphoprotein interacts with the linker domain of STAT1. *Virology,* 444**,** 250-6.

DIDCOCK, L., YOUNG, D. F., GOODBOURN, S. & RANDALL, R. E. 1999a. Sendai virus and simian virus 5 block activation of interferon-responsive genes: importance for virus pathogenesis. *J Virol,* 73**,** 3125-33.

DIDCOCK, L., YOUNG, D. F., GOODBOURN, S. & RANDALL, R. E. 1999b. The V protein of simian virus 5 inhibits interferon signalling by targeting STAT1 for proteasome-mediated degradation. *J Virol,* 73**,** 9928-33.

DOHM, J. C., LOTTAZ, C., BORODINA, T. & HIMMELBAUER, H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res,*36**,** e105.

DREYFUS, M. & RÉGNIER, P. 2002. The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell,* 111**,** 611-3.

DURBIN, A. P., MCAULIFFE, J. M., COLLINS, P. L. & MURPHY, B. R. 1999. Mutations in the C, D, and V open reading frames of human parainfluenza virus type 3 attenuate replication in rodents and primates. *Virology,* 261**,** 319-30.

ECKERDT, F. & STREBHARDT, K. 2006. Polo-like kinase 1: target and regulator of anaphase-promoting complex/cyclosome-dependent proteolysis. *Cancer Res,* 66**,** 6895-8.

EGELMAN, E. H., WU, S. S., AMREIN, M., PORTNER, A. & MURTI, G. 1989. The Sendai virus nucleocapsid exists in at least four different helical states. *J Virol,* 63**,** 2233-43.

FEARNS, R., COLLINS, P. L. & PEEPLES, M. E. 2000. Functional analysis of the genomic and antigenomic promoters of human respiratory syncytial virus. *J Virol,* 74**,** 6006-14.

FEARNS, R., PEEPLES, M. E. & COLLINS, P. L. 1997. Increased expression of the N protein of respiratory syncytial virus stimulates minigenome replication but does not alter the balance between the synthesis of mRNA and antigenome. *Virology,* 236**,** 188-201.

FEARNS, R., PEEPLES, M. E. & COLLINS, P. L. 2002. Mapping the transcription and replication promoters of respiratory syncytial virus. *J Virol,* 76**,** 1663-72.

FEARNS, R. & PLEMPER, R. K. 2017. Polymerases of paramyxoviruses and pneumoviruses. *Virus Res,* 234**,** 87-102.

FEARNS, R., YOUNG, D. F. & RANDALL, R. E. 1994. Evidence that the paramyxovirus simian virus 5 can establish quiescent infections by remaining inactive in cytoplasmic inclusion bodies. *J Gen Virol,* 75 (Pt 12)**,** 3525-39.

FELLER, J. A., SMALLWOOD, S., HORIKAMI, S. M. & MOYER, S. A. 2000. Mutations in conserved domains IV and VI of the large (L) subunit of the

sendai virus RNA polymerase give a spectrum of defective RNA synthesis phenotypes. *Virology,* 269**,** 426-39.

FINKE, S., COX, J. H. & CONZELMANN, K. K. 2000. Differential transcription attenuation of rabies virus genes by intergenic regions: generation of recombinant viruses overexpressing the polymerase gene. *J Virol,* 74**,** 7261-9.

FUENTES, S. M., SUN, D., SCHMITT, A. P. & HE, B. 2010. Phosphorylation of paramyxovirus phosphoprotein and its role in viral gene expression. *Future Microbiol,* 5**,** 9-13.

GAINEY, M. D., DILLON, P. J., CLARK, K. M., MANUSE, M. J. & PARKS, G. D. 2008. Paramyxovirus-induced shutoff of host and viral protein synthesis: role of the P and V proteins in limiting PKR activation. *J Virol,* 82**,** 828-39.

GALINSKI, M. S., TROY, R. M. & BANERJEE, A. K. 1992. RNA editing in the phosphoprotein gene of the human parainfluenza virus type 3. *Virology,* 186**,** 543-50.

GARG, R. K. 2002. Subacute sclerosing panencephalitis. *Postgrad Med J,* 78**,** 63-70.

GREEN, T. J., ZHANG, X., WERTZ, G. W. & LUO, M. 2006. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science,* 313**,** 357-60.

GRIFFIN, D. E., LIN, W. H. & PAN, C. H. 2012. Measles virus, immune control, and persistence. *FEMS Microbiol Rev,* 36**,** 649-62.

GROGAN, C. C. & MOYER, S. A. 2001. Sendai virus wild-type and mutant C proteins show a direct correlation between L polymerase binding and inhibition of viral RNA synthesis. *Virology,* 288**,** 96-108.

GUO, J., PETERS, K. L. & SEN, G. C. 2000. Induction of the human protein P56 by interferon, double-stranded RNA, or virus infection. *Virology,* 267**,** 209-19.

GUPTA, K. C. & KINGSBURY, D. W. 1984. Complete sequences of the intergenic and mRNA start signals in the Sendai virus genome: homologies with the genome of vesicular stomatitis virus. *Nucleic Acids Res,* 12**,** 3829-41.

GUPTA, K. C. & KINGSBURY, D. W. 1985. Polytranscripts of Sendai virus do not contain intervening polyadenylate sequences. *Virology,* 141**,** 102-9.

GUTSCHE, I., DESFOSSES, A., EFFANTIN, G., LING, W. L., HAUPT, M., RUIGROK, R. W., SACHSE, C. & SCHOEHN, G. 2015. Structural virology. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science,* 348**,** 704-7.

HABJAN, M., ANDERSSON, I., KLINGSTRÖM, J., SCHÜMANN, M., MARTIN, A., ZIMMERMANN, P., WAGNER, V., PICHLMAIR, A., SCHNEIDER, U., MÜHLBERGER, E., MIRAZIMI, A. & WEBER, F. 2008. Processing of genome 5' termini as a strategy of negative-strand RNA viruses to avoid RIG-I-dependent interferon induction. *PLoS One,* 3**,** e2032.

HAMAGUCHI, M., YOSHIDA, T., NISHIKAWA, K., NARUSE, H. & NAGAI, Y. 1983. Transcriptive complex of Newcastle disease virus. I. Both L and P proteins are required to constitute an active complex. *Virology,* 128**,** 105-17.

HARDY, R. W. & WERTZ, G. W. 1998. The product of the respiratory syncytial virus M2 gene ORF1 enhances readthrough of intergenic junctions during viral transcription. *J Virol,* 72**,** 520-6.

HARROWER, J., KIEDRZYNSKI, T., BAKER, S., UPTON, A., RAHNAMA, F., SHERWOOD, J., HUANG, Q. S., TODD, A. & PULFORD, D. 2016. Sexual Transmission of Zika Virus and Persistence in Semen, New Zealand, 2016. *Emerg Infect Dis,* 22**,** 1855-7.

HATEM, A., BOZDAĞ, D., TOLAND, A. E. & ÇATALYÜREK, Ü. 2013. Benchmarking short sequence mapping tools. *BMC Bioinformatics,* 14**,** 184.

HAUSMANN, S., GARCIN, D., DELENDA, C. & KOLAKOFSKY, D. 1999a. The versatility of paramyxovirus RNA polymerase stuttering. *J Virol,* 73**,** 5568-76.

HAUSMANN, S., GARCIN, D., MOREL, A. S. & KOLAKOFSKY, D. 1999b. Two nucleotides immediately upstream of the essential A6G3 slippery sequence modulate the pattern of G insertions during Sendai virus mRNA editing. *J Virol,* 73**,** 343-51.

HE, B. & LAMB, R. A. 1999. Effect of inserting paramyxovirus simian virus 5 gene junctions at the HN/L gene junction: analysis of accumulation of mRNAs transcribed from rescued viable viruses. *J Virol,* 73**,** 6228-34.

HE, B., PATERSON, R. G., STOCK, N., DURBIN, J. E., DURBIN, R. K., GOODBOURN, S., RANDALL, R. E. & LAMB, R. A. 2002. Recovery of paramyxovirus simian virus 5 with a V protein lacking the conserved cysteine-rich domain: the multifunctional V protein blocks both interferon-beta induction and interferon signaling. *Virology,* 303**,** 15-32.

HEANEY, C. D., KMUSH, B., NAVAS-ACIEN, A., FRANCESCONI, K., GÖSSLER, W., SCHULZE, K., FAIRWEATHER, D., MEHRA, S., NELSON, K. E., KLEIN, S. L., LI, W., ALI, H., SHAIKH, S., MERRILL, R. D., WU, L., WEST, K. P., CHRISTIAN, P. & LABRIQUE, A. B. 2015. Arsenic exposure and hepatitis E virus infection during pregnancy. *Environ Res,* 142**,** 273-80.

HEGGENESS, M. H., SCHEID, A. & CHOPPIN, P. W. 1980. Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proc Natl Acad Sci U S A,* 77**,** 2631-5.

HIEBERT, S. W., PATERSON, R. G. & LAMB, R. A. 1985. Identification and predicted sequence of a previously unrecognized small hydrophobic protein, SH, of the paramyxovirus simian virus 5. *J Virol,* 55**,** 744-51.

HOFFMAN, M. A. & BANERJEE, A. K. 2000. Precise mapping of the replication and transcription promoters of human parainfluenza virus type 3. *Virology,* 269**,** 201-11.

HORIKAMI, S. M., CURRAN, J., KOLAKOFSKY, D. & MOYER, S. A. 1992. Complexes of Sendai virus NP-P and P-L proteins are required for defective interfering particle genome replication in vitro. *J Virol,* 66**,** 4901-8.

HORIKAMI, S. M., HECTOR, R. E., SMALLWOOD, S. & MOYER, S. A. 1997a. The Sendai virus C protein binds the L polymerase protein to inhibit viral RNA synthesis. *Virology,*235**,** 261-70.

HORIKAMI, S. M., HECTOR, R. E., SMALLWOOD, S. & MOYER, S. A. 1997b. The Sendai virus C protein binds the L polymerase protein to inhibit viral RNA synthesis. *Virology,*235**,** 261-70.

HORIKAMI, S. M., SMALLWOOD, S. & MOYER, S. A. 1996a. The Sendai virus V protein interacts with the NP protein to regulate viral genome RNA replication. *Virology,* 222**,**383-90.

HORIKAMI, S. M., SMALLWOOD, S. & MOYER, S. A. 1996b. The Sendai virus V protein interacts with the NP protein to regulate viral genome RNA replication. *Virology,* 222**,**383-90.

HORNUNG, V., ELLEGAST, J., KIM, S., BRZÓZKA, K., JUNG, A., KATO, H., POECK, H., AKIRA, S., CONZELMANN, K. K., SCHLEE, M., ENDRES, S. & HARTMANN, G. 2006. 5'-Triphosphate RNA is the ligand for RIG-I. *Science,* 314**,** 994-7.

HORVATH, C. M. 2004a. The Jak-STAT pathway stimulated by interferon alpha or interferon beta. *Sci STKE,* 2004**,** tr10.

HORVATH, C. M. 2004b. The Jak-STAT pathway stimulated by interferon gamma. *Sci STKE,* 2004**,** tr8.

HORVATH, C. M. & LAMB, R. A. 1992. Studies on the fusion peptide of a paramyxovirus fusion glycoprotein: roles of conserved residues in cell fusion. *J Virol,* 66**,** 2443-55.

HORVATH, C. M., PATERSON, R. G., SHAUGHNESSY, M. A., WOOD, R. & LAMB, R. A. 1992. Biological activity of paramyxovirus fusion proteins: factors influencing formation of syncytia. *J Virol,* 66**,** 4564-9.

HOWARD, M. & WERTZ, G. 1989. Vesicular stomatitis virus RNA replication: a role for the NS protein. *J Gen Virol,* 70 (Pt 10)**,** 2683-94.

HU, C. & GUPTA, K. C. 2000. Functional significance of alternate phosphorylation in Sendai virus P protein. *Virology,* 268**,** 517-32.

HU, C. J., KATO, A., BOWMAN, M. C., KIYOTANI, K., YOSHIDA, T., MOYER, S. A., NAGAI, Y. & GUPTA, K. C. 1999. Role of primary constitutive phosphorylation of Sendai virus P and V proteins in viral replication and pathogenesis. *Virology,* 263**,** 195-208.

HUANG, A. S. & BALTIMORE, D. 1970. Defective viral particles and viral disease processes. *Nature,* 226**,** 325-7.

HUANG, C., KIYOTANI, K., FUJII, Y., FUKUHARA, N., KATO, A., NAGAI, Y., YOSHIDA, T. & SAKAGUCHI, T. 2000. Involvement of the zinc-binding capacity of Sendai virus V protein in viral pathogenesis. *J Virol,* 74**,** 7834-41.

HUANG, I. C., BAILEY, C. C., WEYER, J. L., RADOSHITZKY, S. R., BECKER, M. M., CHIANG, J. J., BRASS, A. L., AHMED, A. A., CHI, X., DONG, L., LONGOBARDI, L. E., BOLTZ, D., KUHN, J. H., ELLEDGE, S. J., BAVARI, S., DENISON, M. R., CHOE, H. & FARZAN, M. 2011. Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. *PLoS Pathog,* 7**,** e1001258.

HUANG, Y. T. & WERTZ, G. W. 1982. The genome of respiratory syncytial virus is a negative-stranded RNA that codes for at least seven mRNA species. *J Virol,* 43**,** 150-7.

HUANG, Y. T. & WERTZ, G. W. 1983. Respiratory syncytial virus mRNA coding assignments. *J Virol,* 46**,** 667-72.

HWANG, L. N., ENGLUND, N. & PATTNAIK, A. K. 1998. Polyadenylation of vesicular stomatitis virus mRNA dictates efficient transcription termination at the intercistronic gene junctions. *J Virol,* 72**,** 1805-13.

INOUE, M., TOKUSUMI, Y., BAN, H., KANAYA, T., SHIRAKURA, M., TOKUSUMI, T., HIRATA, T., NAGAI, Y., IIDA, A. & HASEGAWA, M. 2003. A new Sendai virus vector deficient in the matrix gene does not form virus particles and shows extensive cell-to-cell spreading. *J Virol,* 77**,** 6419-29.

IRIE, T., OKAMOTO, I., YOSHIDA, A., NAGAI, Y. & SAKAGUCHI, T. 2014. Sendai virus C proteins regulate viral genome and antigenome synthesis to dictate the negative genome polarity. *J Virol,* 88**,** 690-8.

ISENI, F., BAUDIN, F., GARCIN, D., MARQ, J. B., RUIGROK, R. W. & KOLAKOFSKY, D. 2002. Chemical modification of nucleotide bases and mRNA editing depend on hexamer or nucleoprotein phase in Sendai virus nucleocapsids. *RNA,* 8**,** 1056-67.

ISHIHARA, K. & HIRANO, T. 2002. IL-6 in autoimmune disease and chronic inflammatory proliferative disease. *Cytokine Growth Factor Rev,* 13**,** 357-68.

IVERSON, L. E. & ROSE, J. K. 1981. Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. *Cell,* 23**,** 477-84.

JOHNSON, R. T., LAZZARINI, R. A. & WAKSMAN, B. H. 1981. Mechanisms of virus persistence. *Ann Neurol,* 9**,** 616-7.

KATO, A., KIYOTANI, K., HASAN, M. K., SHIODA, T., SAKAI, Y., YOSHIDA, T. & NAGAI, Y. 1999. Sendai virus gene start signals are not equivalent in reinitiation capacity: moderation at the fusion protein gene. *J Virol,* 73**,** 9237-46.

KATO, A., KIYOTANI, K., SAKAI, Y., YOSHIDA, T. & NAGAI, Y. 1997a. The paramyxovirus, Sendai virus, V protein encodes a luxury function required for viral pathogenesis. *EMBO J,* 16**,** 578-87.

KATO, A., KIYOTANI, K., SAKAI, Y., YOSHIDA, T., SHIODA, T. & NAGAI, Y. 1997b. Importance of the cysteine-rich carboxyl-terminal half of V protein for Sendai virus pathogenesis. *J Virol,* 71**,** 7266-72.

KATO, H., TAKEUCHI, O., MIKAMO-SATOH, E., HIRAI, R., KAWAI, T., MATSUSHITA, K., HIIRAGI, A., DERMODY, T. S., FUJITA, T. & AKIRA, S. 2008. Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. *J Exp Med,* 205**,** 1601-10.

KATO, H., TAKEUCHI, O., SATO, S., YONEYAMA, M., YAMAMOTO, M., MATSUI, K., UEMATSU, S., JUNG, A., KAWAI, T., ISHII, K. J., YAMAGUCHI, O., OTSU, K., TSUJIMURA, T., KOH, C. S., REIS E SOUSA, C., MATSUURA, Y., FUJITA, T. & AKIRA, S. 2006. Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature,* 441**,** 101-5.

KAWAI, T., TAKAHASHI, K., SATO, S., COBAN, C., KUMAR, H., KATO, H., ISHII, K. J., TAKEUCHI, O. & AKIRA, S. 2005. IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat Immunol,* 6**,** 981-8.

KAWANO, M., KAITO, M., KOZUKA, Y., KOMADA, H., NODA, N., NANBA, K., TSURUDOME, M., ITO, M., NISHIO, M. & ITO, Y. 2001. Recovery of infectious human parainfluenza type 2 virus from cDNA clones and properties of the defective virus without V-specific cysteine-rich domain. *Virology,* 284**,** 99-112.

KEENE, J. D., THORNTON, B. J. & EMERSON, S. U. 1981. Sequence-specific contacts between the RNA polymerase of vesicular stomatitis virus and the leader RNA gene. *Proc Natl Acad Sci U S A,* 78**,** 6191-5.

KELLER, M. A., MURPHY, S. K. & PARKS, G. D. 2001. RNA replication from the simian virus 5 antigenomic promoter requires three sequence-dependent elements separated by sequence-independent spacer regions. *J Virol,* 75**,** 3993-8.

KELLER, M. A. & PARKS, G. D. 2003. Positive- and negative-acting signals combine to determine differential RNA replication from the paramyxovirus simian virus 5 genomic and antigenomic promoters. *Virology,* 306**,** 347-58.

KHRAMEEVA, E. E. & GELFAND, M. S. 2012. Biases in read coverage demonstrated by interlaboratory and interplatform comparison of 117 mRNA and genome sequencing experiments. *BMC Bioinformatics,* 13 Suppl 6**,** S4.

KILLIP, M. J., YOUNG, D. F., GATHERER, D., ROSS, C. S., SHORT, J. A., DAVISON, A. J., GOODBOURN, S. & RANDALL, R. E. 2013. Deep sequencing analysis of defective genomes of parainfluenza virus 5 and their role in interferon induction. *J Virol,* 87**,** 4798-807.

KIMURA, Y., ITO, Y., SHIMOKATA, K., NISHIYAMA, Y. & NAGATA, I. 1975. Temperature-sensitive virus derived from BHK cells persistently infected with HVJ (Sendai virus). *J Virol,* 15**,** 55-63.

KIYOTANI, K., TAKAO, S., SAKAGUCHI, T. & YOSHIDA, T. 1990. Immediate protection of mice from lethal wild-type Sendai virus (HVJ) infections by a temperature-sensitive mutant, HVJpi, possessing homologous interfering capacity. *Virology,* 177**,** 65-74.

KOLAKOFSKY, D. 2016. Paramyxovirus RNA synthesis, mRNA editing, and genome hexamer phase: A review. *Virology,* 498**,** 94-98.

KOLAKOFSKY, D., LE MERCIER, P., ISENI, F. & GARCIN, D. 2004. Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology,* 318**,** 463-73.

KOLAKOFSKY, D., PELET, T., GARCIN, D., HAUSMANN, S., CURRAN, J. & ROUX, L. 1998a. Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited. *J Virol,* 72**,** 891-9.

KOLAKOFSKY, D., PELET, T., GARCIN, D., HAUSMANN, S., CURRAN, J. & ROUX, L. 1998b. Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited. *J Virol,* 72**,** 891-9.

KOLAKOFSKY, D., ROUX, L., GARCIN, D. & RUIGROK, R. W. 2005. Paramyxovirus mRNA editing, the "rule of six" and error catastrophe: a hypothesis. *J Gen Virol,* 86**,** 1869-77.

KOMATSU, T., TAKEUCHI, K. & GOTOH, B. 2007. Bovine parainfluenza virus type 3 accessory proteins that suppress beta interferon production. *Microbes Infect,* 9**,** 954-62.

KOMATSU, T., TAKEUCHI, K., YOKOO, J. & GOTOH, B. 2004a. C and V proteins of Sendai virus target signaling pathways leading to IRF-3 activation for the negative regulation of interferon-beta production. *Virology,* 325**,** 137-48.

KOMATSU, T., TAKEUCHI, K., YOKOO, J. & GOTOH, B. 2004b. C and V proteins of Sendai virus target signaling pathways leading to IRF-3 activation for the negative regulation of interferon-beta production. *Virology,* 325**,** 137-48.

KONDO, T., YOSHIDA, T., MIURA, N. & NAKANISHI, M. 1993. Temperature-sensitive phenotype of a mutant Sendai virus strain is caused by its insufficient accumulation of the M protein. *J Biol Chem,* 268**,** 21924-30.

KOTWAL, G. J., HATCH, S. & MARSHALL, W. L. 2012. Viral infection: an evolving insight into the signal transduction pathways responsible for the innate immune response. *Adv Virol,* 2012**,** 131457.

KOZAREWA, I., NING, Z., QUAIL, M. A., SANDERS, M. J., BERRIMAN, M. & TURNER, D. J. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods,* 6**,** 291-5.

KRISTENSSON, K. & NORRBY, E. 1986. Persistence of RNA viruses in the central nervous system. *Annu Rev Microbiol,* 40**,** 159-84.

KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol,* 33**,** 1870-4.

KUO, L., FEARNS, R. & COLLINS, P. L. 1997. Analysis of the gene start and gene end signals of human respiratory syncytial virus: quasi-templated initiation at position 1 of the encoded mRNA. *J Virol,* 71**,** 4944-53.

KUROTANI, A., KIYOTANI, K., KATO, A., SHIODA, T., SAKAI, Y., MIZUMOTO, K., YOSHIDA, T. & NAGAI, Y. 1998. Sendai virus C proteins are categorically nonessential gene products but silencing their expression severely impairs viral replication and pathogenesis. *Genes Cells,* 3**,** 111-24.

LAHAYE, X., VIDY, A., POMIER, C., OBIANG, L., HARPER, F., GAUDIN, Y. & BLONDEL, D. 2009. Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: Evidence that NBs are sites of viral transcription and replication. *J Virol,* 83**,** 7948-58.

LAMB, R. A. & PARKS, G. D. 2013. *Fields Virology,* Philadelphia, Lippincott Williams & Wilkins.

LANDER, E. S. & WATERMAN, M. S. 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics,* 2**,** 231-9.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol,* 10**,**R25.

LATORRE, P., KOLAKOFSKY, D. & CURRAN, J. 1998. Sendai virus Y proteins are initiated by a ribosomal shunt. *Mol Cell Biol,* 18**,** 5021-31.

LESER, G. P., ECTOR, K. J. & LAMB, R. A. 1996. The paramyxovirus simian virus 5 hemagglutinin-neuraminidase glycoprotein, but not the fusion glycoprotein, is internalized via coated pits and enters the endocytic pathway. *Mol Biol Cell,* 7**,** 155-72.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LI, J., RAHMEH, A., MORELLI, M. & WHELAN, S. P. 2008. A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping. *J Virol,* 82**,** 775-84.

LI, K., MARKOSYAN, R. M., ZHENG, Y. M., GOLFETTO, O., BUNGART, B., LI, M., DING, S., HE, Y., LIANG, C., LEE, J. C., GRATTON, E., COHEN, F. S. & LIU, S. L. 2013. IFITM proteins restrict viral membrane hemifusion. *PLoS Pathog,* 9**,** e1003124.

LIANG, B., LI, Z., JENNI, S., RAHMEH, A. A., MORIN, B. M., GRANT, T., GRIGORIEFF, N., HARRISON, S. C. & WHELAN, S. P. J. 2015. Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. *Cell,* 162**,** 314-327.

LIN, G. Y., PATERSON, R. G. & LAMB, R. A. 1997. The RNA binding region of the paramyxovirus SV5 V and P proteins. *Virology,* 238**,** 460-9.

LIN, G. Y., PATERSON, R. G., RICHARDSON, C. D. & LAMB, R. A. 1998. The V protein of the paramyxovirus SV5 interacts with damage-specific DNA binding protein. *Virology,*249**,** 189-200.

LIN, Y., HORVATH, F., ALIGO, J. A., WILSON, R. & HE, B. 2005a. The role of simian virus 5 V protein on viral RNA synthesis. *Virology,* 338**,** 270-80.

LIN, Y., HORVATH, F., ALIGO, J. A., WILSON, R. & HE, B. 2005b. The role of simian virus 5 V protein on viral RNA synthesis. *Virology,* 338**,** 270-80.

LINGEMANN, M., SURMAN, S., AMARO-CARAMBOT, E., SCHAAP-NUTT, A., COLLINS, P. L. & MUNIR, S. 2015. The aberrant gene-end transcription signal of the matrix M gene of human parainfluenza virus

type 3 downregulates fusion F protein expression and the F-specific antibody response in vivo. *J Virol,* 89**,** 3318-31.

LIUZZI, M., MASON, S. W., CARTIER, M., LAWETZ, C., MCCOLLUM, R. S., DANSEREAU, N., BOLGER, G., LAPEYRE, N., GAUDETTE, Y., LAGACÉ, L., MASSARIOL, M. J., DÔ, F., WHITEHEAD, P., LAMARRE, L., SCOUTEN, E., BORDELEAU, J., LANDRY, S., RANCOURT, J., FAZAL, G. & SIMONEAU, B. 2005. Inhibitors of respiratory syncytial virus replication target cotranscriptional mRNA guanylylation by viral RNA-dependent RNA polymerase. *J Virol,* 79**,** 13105-15.

LU, L. L., PURI, M., HORVATH, C. M. & SEN, G. C. 2008. Select paramyxoviral V proteins inhibit IRF3 activation by acting as alternative substrates for inhibitor of kappaB kinase epsilon (IKKe)/TBK1. *J Biol Chem,* 283**,** 14269-76.

MACLACHLAN, N. J. 1994. The pathogenesis and immunology of bluetongue virus infection of ruminants. *Comp Immunol Microbiol Infect Dis,* 17**,** 197-206.

MACLELLAN, K., LONEY, C., YEO, R. P. & BHELLA, D. 2007. The 24-angstrom structure of respiratory syncytial virus nucleocapsid protein-RNA decameric rings. *J Virol,* 81**,**9519-24.

MALUR, A. G., CHOUDHARY, S. K., DE, B. P. & BANERJEE, A. K. 2002a. Role of a highly conserved NH(2)-terminal domain of the human parainfluenza virus type 3 RNA polymerase. *J Virol,* 76**,** 8101-9.

MALUR, A. G., GUPTA, N. K., DE BISHNU, P. & BANERJEE, A. K. 2002b. Analysis of the mutations in the active site of the RNA-dependent RNA polymerase of human parainfluenza virus type 3 (HPIV3). *Gene Expr,* 10**,** 93-100.

MALUR, A. G., GUPTA, N. K., DE BISHNU, P. & BANERJEE, A. K. 2002c. Analysis of the mutations in the active site of the RNA-dependent RNA polymerase of human parainfluenza virus type 3 (HPIV3). *Gene Expr,* 10**,** 93-100.

MALUR, A. G., HOFFMAN, M. A. & BANERJEE, A. K. 2004. The human parainfluenza virus type 3 (HPIV 3) C protein inhibits viral transcription. *Virus Res,* 99**,** 199-204.

MANUSE, M. J. & PARKS, G. D. 2009. Role for the paramyxovirus genomic promoter in limiting host cell antiviral responses and cell killing. *J Virol,* 83**,** 9057-67.

MCGIVERN, D. R., COLLINS, P. L. & FEARNS, R. 2005. Identification of internal sequences in the 3' leader region of human respiratory syncytial virus that enhance transcription and confer replication processivity. *J Virol,* 79**,** 2449-60.

MELCHJORSEN, J., JENSEN, S. B., MALMGAARD, L., RASMUSSEN, S. B., WEBER, F., BOWIE, A. G., MATIKAINEN, S. & PALUDAN, S. R. 2005. Activation of innate defense against a paramyxovirus is mediated by RIG-I and TLR7 and TLR8 in a cell-type-specific manner. *J Virol,* 79**,** 12944-51.

MORIN, B., RAHMEH, A. A. & WHELAN, S. P. 2012. Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase. *EMBO J,* 31**,** 1320-9.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods,* 5**,** 621-8.

MOSCONA, A. 2005. Entry of parainfluenza virus into cells as a target for interrupting childhood respiratory disease. *J Clin Invest,* 115**,** 1688-98.

MOUNTCASTLE, W. E., COMPANS, R. W., LACKLAND, H. & CHOPPIN, P. W. 1974. Proteolytic cleavage of subunits of the nucleocapsid of the paramyxovirus simian virus 5. *J Virol,* 14**,** 1253-61.

MURPHY, S. K. & PARKS, G. D. 1999. RNA replication for the paramyxovirus simian virus 5 requires an internal repeated (CGNNNN) sequence motif. *J Virol,* 73**,** 805-9.

NAGAI, Y. & YOSHIDA, T. 1984. Viral pathogenesis: mechanism of acute and persistent infections with paramyxoviruses. *Nagoya J Med Sci,* 46**,** 1-17.

NAGATA, I., KIMURA, Y., ITO, Y. & TANAKA, T. 1972. Temperature-sensitive phenomenon of viral maturation observed in BHK cells persistently infected with HVJ. *Virology,*49**,** 453-61.

NAKATSU, Y., TAKEDA, M., OHNO, S., KOGA, R. & YANAGI, Y. 2006. Translational inhibition and increased interferon induction in cells infected with C protein-deficient measles virus. *J Virol,* 80**,** 11861-7.

NIKOLIC, J., LE BARS, R., LAMA, Z., SCRIMA, N., LAGAUDRIÈRE-GESBERT, C., GAUDIN, Y. & BLONDEL, D. 2017. Negri bodies are viral factories with properties of liquid organelles. *Nat Commun,* 8**,** 58.

NISHIO, M., NAGATA, A., TSURUDOME, M., ITO, M., KAWANO, M., KOMADA, H. & ITO, Y. 2004. Recombinant Sendai viruses with L1618V mutation in their L polymerase protein establish persistent infection, but not temperature sensitivity. *Virology,* 329**,** 289-301.

NISHIO, M., OHTSUKA, J., TSURUDOME, M., NOSAKA, T. & KOLAKOFSKY, D. 2008. Human parainfluenza virus type 2 V protein inhibits genome replication by binding to the L protein: possible role in promoting viral fitness. *J Virol,* 82**,** 6130-8.

NOTON, S. L., DEFLUBÉ, L. R., TREMAGLIO, C. Z. & FEARNS, R. 2012. The respiratory syncytial virus polymerase has multiple RNA synthesis activities at the promoter. *PLoS Pathog,* 8**,** e1002980.

NOTON, S. L. & FEARNS, R. 2015. Initiation and regulation of paramyxovirus transcription and replication. *Virology,* 479-480**,** 545-54.

OGINO, T. & BANERJEE, A. K. 2007. Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol Cell,* 25**,** 85-97.

OGINO, T. & BANERJEE, A. K. 2010. The HR motif in the RNA-dependent RNA polymerase L protein of Chandipura virus is required for unconventional mRNA-capping activity. *J Gen Virol,* 91**,** 1311-4.

OGINO, T., KOBAYASHI, M., IWAMA, M. & MIZUMOTO, K. 2005. Sendai virus RNA-dependent RNA polymerase L protein catalyzes cap methylation of virus-specific mRNA. *J Biol Chem,* 280**,** 4429-35.

OGINO, T., YADAV, S. P. & BANERJEE, A. K. 2010. Histidine-mediated RNA transfer to GDP for unique mRNA capping by vesicular stomatitis virus RNA polymerase. *Proc Natl Acad Sci U S A,* 107**,** 3463-8.

ORTON, R. J., GU, Q., HUGHES, J., MAABAR, M., MODHA, S., VATTIPALLY, S. B., WILKIE, G. S. & DAVISON, A. J. 2016. Bioinformatics tools for analysing viral genomic data. *Rev Sci Tech,* 35**,** 271-85.

OYOLA, S. O., OTTO, T. D., GU, Y., MASLEN, G., MANSKE, M., CAMPINO, S., TURNER, D. J., MACINNIS, B., KWIATKOWSKI, D. P., SWERDLOW, H. P. & QUAIL, M. A. 2012. Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genomics,* 13**,** 1.

OZATO, K., SHIN, D. M., CHANG, T. H. & MORSE, H. C. 2008. TRIM family proteins and their emerging roles in innate immunity. *Nat Rev Immunol,* 8**,** 849-60.

PALOSAARI, H., PARISIEN, J. P., RODRIGUEZ, J. J., ULANE, C. M. & HORVATH, C. M. 2003. STAT protein interference and suppression of cytokine signal transduction by measles virus V protein. *J Virol,* 77**,** 7635-44.

PARKS, C. L., WITKO, S. E., KOTASH, C., LIN, S. L., SIDHU, M. S. & UDEM, S. A. 2006. Role of V protein RNA binding in inhibition of measles virus minigenome replication. *Virology,* 348**,** 96-106.

PARKS, G. D. & ALEXANDER-MILLER, M. A. 2013. Paramyxovirus activation and inhibition of innate immune responses. *J Mol Biol,* 425**,** 4872-92.

PARKS, G. D., WARD, C. D. & LAMB, R. A. 1992. Molecular cloning of the NP and L genes of simian virus 5: identification of highly conserved domains in paramyxovirus NP and L proteins. *Virus Res,* 22**,** 259-79.

PATERSON, R. G., HARRIS, T. J. & LAMB, R. A. 1984. Analysis and gene assignment of mRNAs of a paramyxovirus, simian virus 5. *Virology,* 138**,** 310-23.

PATTON, J. T., DAVIS, N. L. & WERTZ, G. W. 1984. N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus. *J Virol,*49**,** 303-9.

PELET, T., CURRAN, J. & KOLAKOFSKY, D. 1991. The P gene of bovine parainfluenza virus 3 expresses all three reading frames from a single mRNA editing site. *EMBO J,* 10**,**443-8.

PFALLER, C. K. & CONZELMANN, K. K. 2008. Measles virus V protein is a decoy substrate for IkappaB kinase alpha and prevents Toll-like receptor 7/9-mediated interferon induction. *J Virol,* 82**,** 12365-73.

PICHLMAIR, A., SCHULZ, O., TAN, C. P., NÄSLUND, T. I., LILJESTRÖM, P., WEBER, F. & REIS E SOUSA, C. 2006. RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science,* 314**,** 997-1001.

PICKAR, A., XU, P., ELSON, A., LI, Z., ZENGEL, J. & HE, B. 2014. Roles of serine and threonine residues of mumps virus P protein in viral transcription and replication. *J Virol,*88**,** 4414-22.

PLUMET, S., HERSCHKE, F., BOURHIS, J. M., VALENTIN, H., LONGHI, S. & GERLIER, D. 2007. Cytosolic 5'-triphosphate ended viral leader transcript of measles virus as activator of the RIG I-mediated interferon response. *PLoS One,* 2**,** e279.

POOLE, E., HE, B., LAMB, R. A., RANDALL, R. E. & GOODBOURN, S. 2002. The V proteins of simian virus 5 and other paramyxoviruses inhibit induction of interferon-beta. *Virology,* 303**,** 33-46.

POROTTO, M., MURRELL, M., GREENGARD, O., DOCTOR, L. & MOSCONA, A. 2005. Influence of the human parainfluenza virus 3 attachment protein's neuraminidase activity on its capacity to activate the fusion protein. *J Virol,* 79**,** 2383-92.

PRECIOUS, B., CHILDS, K., FITZPATRICK-SWALLOW, V., GOODBOURN, S. & RANDALL, R. E. 2005. Simian virus 5 V protein acts as an adaptor, linking DDB1 to STAT2, to facilitate the ubiquitination of STAT1. *J Virol,* 79**,** 13434-41.

PRECIOUS, B., YOUNG, D. F., BERMINGHAM, A., FEARNS, R., RYAN, M. & RANDALL, R. E. 1995. Inducible expression of the P, V, and NP genes

of the paramyxovirus simian virus 5 in cell lines and an examination of NP-P and NP-V interactions. *J Virol,* 69**,** 8001-10.

PRECIOUS, B. L., CARLOS, T. S., GOODBOURN, S. & RANDALL, R. E. 2007. Catalytic turnover of STAT1 allows PIV5 to dismantle the interferon-induced anti-viral state of cells. *Virology,* 368**,** 114-21.

QANUNGO, K. R., SHAJI, D., MATHUR, M. & BANERJEE, A. K. 2004. Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. *Proc Natl Acad Sci U S A,* 101**,** 5952-7.

QIU, X., FU, Q., MENG, C., YU, S., ZHAN, Y., DONG, L., REN, T., SUN, Y., TAN, L., SONG, C., HAN, X. & DING, C. 2016. Kinetic analysis of RNA editing of Newcastle disease virus P gene in the early period of infection. *Acta Virol,* 60**,** 71-7.

RA, K. & PL, C. 2013. *Fields Virology,* Philadelphia, Lippincott Williams & Wilkins.

RADFORD, A. D., CHAPMAN, D., DIXON, L., CHANTREY, J., DARBY, A. C. & HALL, N. 2012. Application of next-generation sequencing technologies in virology. *J Gen Virol,* 93**,**1853-68.

RAFTERY, N. & STEVENSON, N. J. 2017. Advances in anti-viral immune defence: revealing the importance of the IFN JAK/STAT pathway. *Cell Mol Life Sci,* 74**,** 2525-2535.

RANDALL, R. E. & GOODBOURN, S. 2008. Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *J Gen Virol,* 89**,**1-47.

RANDALL, R. E. & GRIFFIN, D. E. 2017. Within host RNA virus persistence: mechanisms and consequences. *Curr Opin Virol,* 23**,** 35-42.

RASSA, J. C. & PARKS, G. D. 1998. Molecular basis for naturally occurring elevated readthrough transcription across the M-F junction of the paramyxovirus SV5. *Virology,*247**,** 274-86.

RASSA, J. C. & PARKS, G. D. 1999. Highly diverse intergenic regions of the paramyxovirus simian virus 5 cooperate with the gene end U tract in viral transcription termination and can influence reinitiation at a downstream gene. *J Virol,* 73**,** 3904-12.

RASSA, J. C., WILSON, G. M., BREWER, G. A. & PARKS, G. D. 2000. Spacing constraints on reinitiation of paramyxovirus transcription: the gene end U tract acts as a spacer to separate gene end from gene start sites. *Virology,* 274**,** 438-49.

REUTTER, G. L., CORTESE-GROGAN, C., WILSON, J. & MOYER, S. A. 2001. Mutations in the measles virus C protein that up regulate viral RNA synthesis. *Virology,* 285**,** 100-9.

ROBINSON, M. D. & OSHLACK, A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol,* 11**,** R25.

ROUTH, A. & JOHNSON, J. E. 2014. Discovery of functional genomic motifs in viruses with ViReMa-a Virus Recombination Mapper-for analysis of next-generation sequencing data. *Nucleic Acids Res,* 42**,** e11.

ROUX, L. & HOLLAND, J. J. 1979. Role of defective interfering particles of Sendai virus in persistent infections. *Virology,* 93**,** 91-103.

ROUX, L. & HOLLAND, J. J. 1980. Viral genome synthesis in BHK 21 cells persistently infected with Sendai virus. *Virology,* 100**,** 53-64.

ROUX, L., SIMON, A. E. & HOLLAND, J. J. 1991. Effects of defective interfering viruses on virus replication and pathogenesis in vitro and in vivo. *Adv Virus Res,* 40**,** 181-211.

ROZENBLATT, S., GESANG, C., LAVIE, V. & NEUMANN, F. S. 1982. Cloning and characterization of DNA complementary to the measles virus mRNA encoding hemagglutinin and matrix protein. *J Virol,* 42**,** 790-7.

RUIGROK, R. W., CRÉPIN, T. & KOLAKOFSKY, D. 2011. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol,* 14**,** 504-10.

SANDERSON, C. M., AVALOS, R., KUNDU, A. & NAYAK, D. P. 1995. Interaction of Sendai viral F, HN, and M proteins with host cytoskeletal and lipid components in Sendai virus-infected BHK cells. *Virology,* 209**,** 701-7.

SCHNEIDER, H., KAELIN, K. & BILLETER, M. A. 1997. Recombinant measles viruses defective for RNA editing and V protein synthesis are viable in cultured cells. *Virology,*227**,** 314-22.

SCHNEIDER, W. M., CHEVILLOTTE, M. D. & RICE, C. M. 2014. Interferon-stimulated genes: a complex web of host defenses. *Annu Rev Immunol,* 32**,** 513-45.

SCHWARTZ-CORNIL, I., MERTENS, P. P., CONTRERAS, V., HEMATI, B., PASCALE, F., BRÉARD, E., MELLOR, P. S., MACLACHLAN, N. J. & ZIENTARA, S. 2008. Bluetongue virus: virology, pathogenesis and immunity. *Vet Res,* 39**,** 46.

SHABMAN, R. S., HOENEN, T., GROSETH, A., JABADO, O., BINNING, J. M., AMARASINGHE, G. K., FELDMANN, H. & BASLER, C. F. 2013. An upstream open reading frame modulates ebola virus polymerase translation and virus replication. *PLoS Pathog,* 9**,** e1003147.

SHIMAZU, Y., TAKAO, S. I., IRIE, T., KIYOTANI, K., YOSHIDA, T. & SAKAGUCHI, T. 2008. Contribution of the leader sequence to homologous viral interference among Sendai virus strains. *Virology,* 372**,** 64-71.

SHINGAI, M., EBIHARA, T., BEGUM, N. A., KATO, A., HONMA, T., MATSUMOTO, K., SAITO, H., OGURA, H., MATSUMOTO, M. & SEYA, T. 2007. Differential type I IFN-inducing abilities of wild-type versus vaccine strains of measles virus. *J Immunol,* 179**,** 6123-33.

SIDHU, M. S., CROWLEY, J., LOWENTHAL, A., KARCHER, D., MENONNA, J., COOK, S., UDEM, S. & DOWLING, P. 1994. Defective measles virus in human subacute sclerosing panencephalitis brain. *Virology,* 202**,** 631-41.

SLEAT, D. E. & BANERJEE, A. K. 1993. Transcriptional activity and mutational analysis of recombinant vesicular stomatitis virus RNA polymerase. *J Virol,* 67**,** 1334-9.

SMALLWOOD, S., EASSON, C. D., FELLER, J. A., HORIKAMI, S. M. & MOYER, S. A. 1999. Mutations in conserved domain II of the large (L) subunit of the Sendai virus RNA polymerase abolish RNA synthesis. *Virology,* 262**,** 375-83.

SMALLWOOD, S., HÖVEL, T., NEUBERT, W. J. & MOYER, S. A. 2002. Different substitutions at conserved amino acids in domains II and III in the Sendai L RNA polymerase protein inactivate viral RNA synthesis. *Virology,* 304**,** 135-45.

SMALLWOOD, S. & MOYER, S. A. 2004. The L polymerase protein of parainfluenza virus 3 forms an oligomer and can interact with the heterologous Sendai virus L, P and C proteins. *Virology,* 318**,** 439-50.

SPEHNER, D., DRILLIEN, R. & HOWLEY, P. M. 1997. The assembly of the measles virus nucleoprotein into nucleocapsid-like particles is modulated by the phosphoprotein. *Virology,* 232**,** 260-8.

SPRIGGS, M. K. & COLLINS, P. L. 1986. Human parainfluenza virus type 3: messenger RNAs, polypeptide coding assignments, intergenic sequences, and genetic map. *J Virol,*59**,** 646-54.

SPRIGGS, M. K., JOHNSON, P. R. & COLLINS, P. L. 1987. Sequence analysis of the matrix protein gene of human parainfluenza virus type 3: extensive sequence homology among paramyxoviruses. *J Gen Virol,* 68 (Pt 5)**,** 1491-7.

STETSON, D. B. & MEDZHITOV, R. 2006. Type I interferons in host defense. *Immunity,* 25**,** 373-81.

STILLMAN, E. A. & WHITT, M. A. 1997. Mutational analyses of the intergenic dinucleotide and the transcriptional start sequence of vesicular

stomatitis virus (VSV) define sequences required for efficient termination and initiation of VSV transcripts. *J Virol,* 71**,** 2127-37.

STILLMAN, E. A. & WHITT, M. A. 1999. Transcript initiation and 5'-end modifications are separable events during vesicular stomatitis virus transcription. *J Virol,* 73**,** 7199-209.

STRAHLE, L., GARCIN, D. & KOLAKOFSKY, D. 2006. Sendai virus defective-interfering genomes and the activation of interferon-beta. *Virology,* 351**,** 101-11.

STRICKER, R., MOTTET, G. & ROUX, L. 1994. The Sendai virus matrix protein appears to be recruited in the cytoplasm by the viral nucleocapsid to function in viral assembly and budding. *J Gen Virol,* 75 (Pt 5)**,** 1031-42.

STRÄHLE, L., GARCIN, D., LE MERCIER, P., SCHLAAK, J. F. & KOLAKOFSKY, D. 2003. Sendai virus targets inflammatory responses, as well as the interferon-induced antiviral state, in a multifaceted manner. *J Virol,* 77**,** 7903-13.

STRÄHLE, L., MARQ, J. B., BRINI, A., HAUSMANN, S., KOLAKOFSKY, D. & GARCIN, D. 2007. Activation of the beta interferon promoter by unnatural Sendai virus infection requires RIG-I and is inhibited by viral C proteins. *J Virol,* 81**,** 12227-37.

SUN, D., LUTHRA, P., LI, Z. & HE, B. 2009. PLK1 down-regulates parainfluenza virus 5 gene expression. *PLoS Pathog,* 5**,** e1000525.

SUN, M., FUENTES, S. M., TIMANI, K., SUN, D., MURPHY, C., LIN, Y., AUGUST, A., TENG, M. N. & HE, B. 2008. Akt plays a critical role in replication of nonsegmented negative-stranded RNA viruses. *J Virol,* 82**,** 105-14.

SÁNCHEZ-APARICIO, M. T., GARCIN, D., RICE, C. M., KOLAKOFSKY, D., GARCÍA-SASTRE, A. & BAUM, A. 2017. Loss of Sendai virus C protein leads to accumulation of RIG-I immunostimulatory defective interfering RNA. *J Gen Virol,* 98**,** 1282-1293.

TAKAHASI, K., YONEYAMA, M., NISHIHORI, T., HIRAI, R., KUMETA, H., NARITA, R., GALE, M., INAGAKI, F. & FUJITA, T. 2008. Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Mol Cell,* 29**,** 428-40.

TAKEUCHI, K., TANABAYASHI, K., HISHIYAMA, M., YAMADA, A. & SUGIURA, A. 1991. Variations of nucleotide sequences and transcription of the SH gene among mumps virus strains. *Virology,* 181**,** 364-6.

TAKIMOTO, T. & PORTNER, A. 2004. Molecular mechanism of paramyxovirus budding. *Virus Res,* 106**,** 133-45.

TAPPAREL, C., HAUSMANN, S., PELET, T., CURRAN, J., KOLAKOFSKY, D. & ROUX, L. 1997. Inhibition of Sendai virus genome replication due to promoter-increased selectivity: a possible role for the accessory C proteins. *J Virol,* 71**,** 9588-99.

TAPPAREL, C., MAURICE, D. & ROUX, L. 1998. The activity of Sendai virus genomic and antigenomic promoters requires a second element past the leader template regions: a motif (GNNNNN)3 is essential for replication. *J Virol,* 72**,** 3117-28.

TAWAR, R. G., DUQUERROY, S., VONRHEIN, C., VARELA, P. F., DAMIER-PIOLLE, L., CASTAGNÉ, N., MACLELLAN, K., BEDOUELLE, H., BRICOGNE, G., BHELLA, D., ELÉOUËT, J. F. & REY, F. A. 2009. Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science,* 326**,** 1279-83.

TERRIER, O., DURUPT, F., CARTET, G., THOMAS, L., LINA, B. & ROSA-CALATRAVA, M. 2009. Engineering of a parainfluenza virus type 5 fusion protein (PIV-5 F): development of an autonomous and hyperfusogenic protein by a combinational mutagenesis approach. *Virus Res,* 146**,** 115-24.

THOMAS, S. M., LAMB, R. A. & PATERSON, R. G. 1988. Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell,* 54**,** 891-902.

TILAK, M. K., BOTERO-CASTRO, F., GALTIER, N. & NABHOLZ, B. 2018. Illumina Library Preparation for Sequencing the GC-Rich Fraction of Heterogeneous Genomic DNA. *Genome Biol Evol,* 10**,** 616-622.

TIMANI, K. A., SUN, D., SUN, M., KEIM, C., LIN, Y., SCHMITT, P. T., SCHMITT, A. P. & HE, B. 2008. A single amino acid residue change in the P protein of parainfluenza virus 5 elevates viral gene expression. *J Virol,* 82**,** 9123-33.

TOBER, C., SEUFERT, M., SCHNEIDER, H., BILLETER, M. A., JOHNSTON, I. C., NIEWIESK, S., TER MEULEN, V. & SCHNEIDER-SCHAULIES, S. 1998. Expression of measles virus V protein is associated with pathogenicity and control of viral RNA synthesis. *J Virol,* 72**,** 8124-32.

TRAPNELL, C., PACHTER, L. & SALZBERG, S. L. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* 25**,** 1105-11.

TREMAGLIO, C. Z., NOTON, S. L., DEFLUBÉ, L. R. & FEARNS, R. 2013. Respiratory syncytial virus polymerase can initiate transcription from position 3 of the leader promoter. *J Virol,* 87**,** 3196-207.

TSURUDOME, M., BANDO, H., KAWANO, M., MATSUMURA, H., KOMADA, H., NISHIO, M. & ITO, Y. 1991. Transcripts of simian virus 41 (SV41) matrix gene are exclusively dicistronic with the fusion gene which is also transcribed as a monocistron. *Virology,* 184**,** 93-100.

ULANE, C. M., KENTSIS, A., CRUZ, C. D., PARISIEN, J. P., SCHNEIDER, K. L. & HORVATH, C. M. 2005. Composition and assembly of STAT-targeting ubiquitin ligase complexes: paramyxovirus V protein carboxyl terminus is an oligomerization domain. *J Virol,* 79**,** 10180-9.

VAN CLEVE, W., AMARO-CARAMBOT, E., SURMAN, S. R., BEKISZ, J., COLLINS, P. L., ZOON, K. C., MURPHY, B. R., SKIADOPOULOS, M. H. & BARTLETT, E. J. 2006. Attenuating mutations in the P/C gene of human parainfluenza virus type 1 (HPIV1) vaccine candidates abrogate the inhibition of both induction and signaling of type I interferon (IFN) by wild-type HPIV1. *Virology,* 352**,** 61-73.

VIDAL, S., CURRAN, J. & KOLAKOFSKY, D. 1990. A stuttering model for paramyxovirus P mRNA editing. *EMBO J,* 9**,** 2017-22.

VIDAL, S. & KOLAKOFSKY, D. 1989a. Modified model for the switch from Sendai virus transcription to replication. *J Virol,* 63**,** 1951-8.

VIDAL, S. & KOLAKOFSKY, D. 1989b. Modified model for the switch from Sendai virus transcription to replication. *J Virol,* 63**,** 1951-8.

WANG, J. T., MCELVAIN, L. E. & WHELAN, S. P. 2007. Vesicular stomatitis virus mRNA capping machinery requires specific cis-acting signals in the RNA. *J Virol,* 81**,** 11499-506.

WANSLEY, E. K. & PARKS, G. D. 2002. Naturally occurring substitutions in the P/V gene convert the noncytopathic paramyxovirus simian virus 5 into a virus that induces alpha/beta interferon synthesis and cell death. *J Virol,* 76**,** 10109-21.

WHELAN, S. P. 2008. Response to "Non-segmented negative-strand RNA virus RNA synthesis in vivo". *Virology,* 371**,** 234-7.

WHELAN, S. P. & WERTZ, G. W. 1999. Regulation of RNA synthesis by the genomic termini of vesicular stomatitis virus: identification of distinct sequences essential for transcription but not replication. *J Virol,* 73**,** 297-306.

WHELAN, S. P. & WERTZ, G. W. 2002. Transcription and replication initiate at separate sites on the vesicular stomatitis virus genome. *Proc Natl Acad Sci U S A,* 99**,** 9178-83.

WHETTER, L. E., MACLACHLAN, N. J., GEBHARD, D. H., HEIDNER, H. W. & MOORE, P. F. 1989. Bluetongue virus infection of bovine monocytes. *J Gen Virol,* 70 (Pt 7)**,** 1663-76.

WICKENS, M., ANDERSON, P. & JACKSON, R. J. 1997. Life and death in the cytoplasm: messages from the 3' end. *Curr Opin Genet Dev,* 7**,** 220-32.

WILDE, A. & MORRISON, T. 1984. Structural and functional characterization of Newcastle disease virus polycistronic RNA species. *J Virol,* 51**,** 71-6.

WILHELM, B. T. & LANDRY, J. R. 2009. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods,* 48**,** 249-57.

WITKO, S. E., KOTASH, C., SIDHU, M. S., UDEM, S. A. & PARKS, C. L. 2006. Inhibition of measles virus minireplicon-encoded reporter gene expression by V protein. *Virology,*348**,** 107-19.

WONG, T. C. & HIRANO, A. 1986. Functional cDNA library for efficient expression of measles virus-specific gene products in primate cells. *J Virol,* 57**,** 343-8.

XU, J., SUN, Y., LI, Y., RUTHEL, G., WEISS, S. R., RAJ, A., BEITING, D. & LÓPEZ, C. B. 2017. Replication defective viral genomes exploit a cellular pro-survival mechanism to establish paramyxovirus persistence. *Nat Commun,* 8**,** 799.

YONEYAMA, M., KIKUCHI, M., MATSUMOTO, K., IMAIZUMI, T., MIYAGISHI, M., TAIRA, K., FOY, E., LOO, Y. M., GALE, M., AKIRA, S., YONEHARA, S., KATO, A. & FUJITA, T. 2005. Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J Immunol,* 175**,** 2851-8.

YONEYAMA, M., KIKUCHI, M., NATSUKAWA, T., SHINOBU, N., IMAIZUMI, T., MIYAGISHI, M., TAIRA, K., AKIRA, S. & FUJITA, T. 2004. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol,* 5**,** 730-7.

YOSHIDA, T., HAMAGUCHI, M., NARUSE, H. & NAGAI, Y. 1982. Persistent infection by a temperature-sensitive mutant isolated from a Sendai virus (HVJ) carrier culture: its initiation and maintenance without aid of defective interfering particles. *Virology,* 120**,** 329-39.

YOUNG, D. F., CARLOS, T. S., HAGMAIER, K., FAN, L. & RANDALL, R. E. 2007. AGS and other tissue culture cells can unknowingly be persistently infected with PIV5; a virus that blocks interferon signalling by degrading STAT1. *Virology,* 365**,** 238-40.

YOUNG, D. F., DIDCOCK, L., GOODBOURN, S. & RANDALL, R. E. 2000. Paramyxoviridae use distinct virus-specific mechanisms to circumvent the interferon response. *Virology,* 269**,** 383-90.

ZHANG, D., HE, F., BI, S., GUO, H., ZHANG, B., WU, F., LIANG, J., YANG, Y., TIAN, Q., JU, C., FAN, H., CHEN, J., GUO, X. & LUO, Y. 2016. Genome-Wide Transcriptional Profiling Reveals Two Distinct Outcomes in Central Nervous System Infections of Rabies Virus. *Front Microbiol,* 7**,** 751.

ZHANG, S., CHEN, L., ZHANG, G., YAN, Q., YANG, X., DING, B., TANG, Q., SUN, S., HU, Z. & CHEN, M. 2013. An amino acid of human

parainfluenza virus type 3 nucleoprotein is critical for template function and cytoplasmic inclusion body formation. *J Virol,* 87**,** 12457-70.

ZHANG, S., JIANG, Y., CHENG, Q., ZHONG, Y., QIN, Y. & CHEN, M. 2017. Inclusion Body Fusion of Human Parainfluenza Virus Type 3 Regulated by Acetylated α-Tubulin Enhances Viral Replication. *J Virol,* 91.

ZHAO, H. & BANERJEE, A. K. 1995. Interaction between the nucleocapsid protein and the phosphoprotein of human parainfluenza virus 3. Mapping of the interacting domains using a two-hybrid system. *J Biol Chem,* 270**,** 12485-90.

ZHAO, S., FUNG-LEUNG, W. P., BITTNER, A., NGO, K. & LIU, X. 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One,* 9**,**e78644.