Investigating The Effects of Influenza Virus on Alternative Cleavage and Polyadenylation

Alastair Louey



Pembroke College, University of Oxford

A Thesis submitted for the degree of Doctor of Philosophy
Trinity Term 2018

Abstract

Investigating the Effects of Influenza Virus on Alternative Cleavage and Polyadenylation

Alastair Louey
Pembroke College
Doctor of Philosophy in Medical Sciences
Trinity 2018

Influenza viruses are a major threat for global health causing both seasonal epidemics and occasional global pandemic outbreaks. While the host has several ways to combat viral infections by activating innate immunity and inducing apoptosis, influenza has evolved many strategies to counter antiviral defences of the host cell. One relevant strategy requires non-structural protein NS1, which is dynamic and interacts with a diverse range of viral and cellular factors to antagonise host antiviral defence and facilitate viral replication. Amongst the cellular factors, NS1 interacts and inhibits cleavage and polyadenylation specificity factor 30 (CPSF30). CPSF30 plays a central role in polyA (pA) site recognition and, interestingly in plants, pA site selection. Therefore, inhibition of CPSF30 compromises host 3'end processing and may affect alternative cleavage and polyadenylation (APA). Furthermore, while APA has been implicated in antiviral responses in herpes simplex virus (HSV-1) and vesicular stomatitis virus (VSV) infection, the potential impact on APA of host transcripts during influenza infection has not been explored. APA plays a fundamental role in regulating transcriptome complexity and gene expression in eukaryotes. Over 70% of human genes undergo alternative polyadenylation, which is resulted from alternative usage of multiple cleavage and polyadenylation (pA) sites found in the 3'UTR (3'UTR-APA) or in the upstream coding region (CR-APA). 3'UTR length modulation through APA affects gene expression regulation by altering the incorporation of cis-regulatory elements that may mediate messenger RNA stability, localisation and translatability. Global shifts in UTR length have been identified in different cell types and cell conditions; however many aspects of gene expression regulation remain unclear.

This thesis develops a consistent infection time course experiment, where the necessary controls for robust efficient infections and post-infection controls to determine biologically

relevant periods were considered. Combining deep sequencing, subcellular fractionation and molecular biological techniques, the focus of this project assesses the dynamics of infection, the impact of Influenza A/WSN/33 infection on shifting APA profiles in A549 lung epithelial cells, their contributions to gene expression regulation, and physiological impact on host shut-off and host response. Additionally, in order to delineate the mechanism involved in inducing APA, CPSF30 knockdown and transient NS1 over expression experiments were established to mimic infection.

Overall this thesis demonstrates influenza A has the ability to hijack the host nuclear and cytoplasmic mRNA pool, but fails to efficiently export viral mRNA to the cytoplasm late in infection. Efficient infection of A549 cells induces significant APA profile changes in both nuclear and cytoplasmic fractions with the majority being independent from the effects of other stresses such as osmotic shock. Physiologically relevant genes that exhibit APA during infection can contribute to gene and protein expression regulation. Furthermore, NS1 mediated abrogation of CPSF30 during infection has been demonstrated to be responsible for APA changes in a small but physiological relevant subset of genes that regulate host immune responses and promote viral propagation.

Acknowledgements

I would like to give special thanks to my supervisor Andre Furger for all of his invaluable support, his passion in guiding me to become a better and more confident scientist and person, and his encouragements and inspiration during the course of my studies. I would also thank everyone in the Furger Lab, especially Zhiqiao Wang, Radhika Patel, Harry Fischl and Cilian Brophy, for their help in the lab and on a personal level, and for creating such a pleasant workplace. I would also like to thank my collaborators Ervin Fodor, Benjamin Nilsson and the rest of the Fodor Lab for their generous guidance, support and work.

Most importantly, I am extremely grateful to my parents and my sister who continue to support me unconditionally and thus encouraging me to always seek for improvement and reach for higher ambitions in life. I would also like to thank my grandparents, aunty Carol, uncle Richard, the rest of my family and my closest friends who are always there for me. I also could not have done this without the support and love of Candice Chan. Special mentions are given to Professor Tak Lee, Professor Yuen Kwok-Yung and Professor Lan Hui Yao for their mentorship and support. Finally, life in Oxford has been magical and eye-opening. This experience was certainly enhanced by my friends from the Biochemistry department (David, Nadiya, Alberto, Javi and Florian) and Pembroke College.

Abbreviations

Units

In addition to SI units, the following additional units are used:

bp	Base pair
g	standard gravity: 9.8m/s^2
kDa	Kilodalton
MOI	Multiplicity of Infection
PFU	Plaque forming unit
pH	-$\log_{10}$c, c is [H^+] in mol/l
rpm	Revolutions per minute

Other Abbreviations

A549	Adenocarcinomic human alveolar basal epithelial cell
APA	Alternative cleavage and polyadenylation
ATP	Adenosine-5’-triphosphate
aUTR	alternative untranslated region
cDNA	Complementary DNA
CFIm	Mammalian cleavage factors
CPA	Cleavage and polyadenylation
CPSF	Cleavage and polyadenylation specificity factors
CR-APA	Coding region APA
cRNA	Complementary RNA
cRNP	Complementary ribonucleoprotein
CstF	Cleavage stimulation factor
CTD	C-terminal domain
DAPI	4′,6-diamidino-2-phenylindole
DDX5	DEAD-Box Helicase 5
DMEM	Dulbecco’s modified eagle medium
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide-5’-triphosphate
DSE	Downstream element
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence

ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EJC	Exon junction complex
FC	Fold change
FCS	Fetal calf serum
GO	Gene ontology
GTP	Guanosine-5’-trisphosphate
HA	Haemagglutinin
HEK293T	Human embryonic kidney cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEXIM1	Hexamethylene Bisacetamide Inducible 1
hnRNP	Heterogenous nuclear ribonucleoproteins
hpi	Hours post-infection
HRP	Horseradish peroxidase
HSV	Herpes simplex virus
IAV	Influenza virus
IFN	Interferon
KCl	Potassium chloride
M	Matrix protein
MAVS	Mitochondrial antiviral-signalling protein
MDBK	Madin-Darby Bovine Kidney Epithelial cells
MEM	Minimal essential medium eagle
miRISC	miRNA induced silencing complex
miRNA	MicroRNA
NA	Neuraminidase
NEP	nuclear export protein
NES	Nuclear export signal
NMD	Nonsense mediated decay
NP	nucleoprotein
NPC	Nuclear pore complex
NS	Non-structural protein
NTP	Nucleoside trisphosphate
ORF	Open reading frame
PA	polymerase acid
PABPC1	Poly(A) binding protein cytoplasmic 1

PABPN1	nuclear poly(A) binding protein
PAGE	Polyacrylamide gel electrophoresis
PAMP	Pathogen associated molecules patterns
PAP	poly(A) polymerase
PB	Polymerase basic
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
Pol II	RNA polymerase II
pre-mRNA	Pre-messenger RNA
PRR	Pathogen recognition receptors
qRT-PCR	Quantitive real time PCR
RdRP	Heterotrimeric viral RNA-dependent RNA polymerase
RNA	Ribonucleic acid
RPM	Reads per million
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription Polymerase chain reaction
RTP	Room temperature
SDS	Sodium dodecyl sulphate
SEM	Standard error of the mean
siRNA	small interfering RNA
SMAD3	Mothers against decapentaplegic homolog 3
snoRNA	Host small nucleolar RNA
snRNA	Small nuclear RNA
TBE	Tris-borate-EDTA
TBS-T	Tris buffered saline with Tween20
TEMED	N,N,N′,N′-Tetramethylethylenediamine
TRIM25	Tripartite motif-containing protein 25
Tris	Tris (hydroxymethyl) aminomethane
USE	Upstream element
UTR	Untranslated region
vRNA	Viral RNA
vRNP	Viral ribonucleoprotein complexes
VSV	Vesicular stomatitis virus
WSN	Influenza A/WSN/33 (H1N1) virus
WT	Wild type

List of Figures

Figure 1.1 Schematic Overview of Cleavage and Polyadenylation 6
Figure 1.2 Schematic Overview of Alternative Cleavage and Polyadenylation 10
Figure 1.3 Active and Passive APA 12
Figure 1.4 The Influenza A Virion 26
Figure 1.5 The Influenza A virus Life Cycle 28
Figure 1.6 The Influenza Virus RNA Synthesis 32
Figure 1.7 Models of Viral Transcription and Replication 33
Figure 1.8 Innate Immune Response against influenza A Virus Infection 46

Figure 3.1 Quality Controls to Determine Viral Titre, Purity of Ultracentrifuged Viral Stock and Dynamics of Viral Protein Expression during the Course of Infection 71
Figure 3.2 Infection Efficiency Compared between Cells with Aspirated Virus Media vs Cells with Non-Aspirated Virus Media after 1 hour Room Temperature (RTP) Incubation 73
Figure 3.3 Significant Differences in Relative Amount of Viral mRNA when Aspirating vs not Aspirating Viral Media after 1 hour in RTP 75
Figure 3.4 Effect of Aspirating Virus after 1 hour in RTP on total 3'UTR APA events 77
Figure 3.5 Subcellular Fractionation Controls for Infection Time Course 80
Figure 3.6 Schematic Workflow of Established Infection Time Course Experiment 82

Figure 4.1 Immunofluorescence Microscopy Staining for NP Performed After Every Infection Time Course Experiment Ensuring Efficient Infection 88
Figure 4.2 Infection Time-Course Subcellular Fractionation Control 90
Figure 4.3 Dynamics of Infection 94
Figure 4.4 Global Analysis of WSN/33 Infection 3'UTR APA 97
Figure 4.5 Isolation of Genes that Exhibit 3'UTR APA in a Minimum of 3 of Time points 99
Figure 4.6 Gene Ontology Analysis of Significant 3'UTR APA genes in 4 Different Categories 100
Figure 4.7 Differentially Expressed Genes Overlapped with 3'UTR APA Genes Significant at a Minimum 3 Time Points 103
Figure 4.8 Gene Ontology Analysis of Overlapped Genes in APA cohort and Differential Gene Expression Cohort 104/105
Figure 4.9 Selected Examples of Physiologically Relevant Genes Displaying Differential Gene Expression and 3'UTR APA 111
Figure 4.10 Viral-induced 3'UTR APA are Independent of those Induced by Osmotic Stress 114
Figure 4.11 Selected Examples of Physiologically Relevant Genes 3'UTR APA Induced Uniquely by Infection as well as those that are Induced by Both Osmotic Stress and Infection. 115
Figure 4.12 Real Time qPCR to Verify 3'UTR APA of HEXIM1 and DDX5 118
Figure 4.13 Real time qPCR to Verify 3'UTR APA or Failure of Termination of MAVS 119
Figure 4.14 3'UTR APA resulted in Downstream Effects in Gene and Protein Expression 121
Figure 4.15 Functional Consequences of Viral infection when Expressing Selected Genes 124
Figure 4.16 Global WSN/33 Induced CR-APA Events 127
Figure 4.17 Differentially Expressed Genes Overlapped with CR-APA Genes Significant at a Minimum of 3 Time Points 128

Figure 4.18 Selected Examples of Physiologically Relevant Genes Displaying Differential Gene Expression and CR-APA 132
Figure 4.19 Viral-induced CR-APA are Independent of those Induced by Osmotic Shock 134

Figure 5.1 Overview of Influenza A Virus non-structural 1 (NS1) protein interactions 148
Figure 5.2 Subcellular Fraction and Western Blot Controls of CPSF30 knockdown 151
Figure 5.3 Global analysis of CPSF30 Knock-down 3'UTR APA Followed by Comparisons with WSN Infection 3'UTR APA Events 154
Figure 5.4 Gene Ontology Analysis of Overlapped Genes from A/WSN/33 Infection and CPSF30KD 155
Figure 5.5 Examples of Stringently Selected Genes which Display Differential Regulation 3'UTR APA during CPSF30KD Experiments. 156
Figure 5.6 Establishing NS1-G184R mutant control 160
Figure 5.7 Subcellular Fraction and Western Blot Controls of GFP, NS1, NS1 Mutant Transient Expression in HEK293T Cells 162
Figure 5.8 Global Analysis of NS1 and NS1 Mutant Expression 3'UTR APA Followed by Comparisons with Infection 3'UTR APA Events 166
Figure 5.9 Gene Ontology Analysis of Overlapped Genes from WSN Infection, NS1 and NS1 Mutant Expression 167
Figure 5.10 Examples of Stringently Selected Genes which Display Differential Regulation 3'UTR APA during NS1, NS1 Mutant and GFP Expression Experiments 168
Figure 5.11 Verifying Whether NS1 abrogation of CPSF30 is the Mechanism Responsible for Regulating 3'UTR APA 170

Figure 6.1 DDX5 as Prime Example of a Gene with Differential APA, Gene and Protein Regulation Caused by NS1 Abrogation of CPSF30 During IAV WSN/33 Infection 181

List of Tables

Table 1.1 Core factors of Cleavage and Polyadenylation Machinery 7
Table 1.2 Trans-factors known to influence CPA site choice at the point of cleavage 14
Table 1.3 Influenza A virus genome segments, encoded proteins and function 27
Table 1.4. Specific residues of NS! Affecting CPSF30 binding 42

Table 2.1 Sequence of primers used in this study 64
Table 2.2 Antibodies 66

Table 4.1 Physiologically relevant genes selected from GO analysis 106
Table 4.2 Overlaps between gene with differential gene expression and CR-APA 129

Table 5.1 Overlapping genes that are regulated after CPSF30KD and NS1 overexpression and upon infection, but NOT after osmotic shock/ NS1-G184R 171

Table of Contents

Abstract i
Acknowledgements iii
Abbreviations iv
List of Figures vii
List of Tables ix

Chapter 1: Introduction 1

1.1 Eukaryotic Pre-mRNA processing
1.1 Pre-mRNA processing
1.1.1 5' end capping 1
1.1.2 Splicing 2
1.1.3 Cleavage and Polyadenylation 4
1.1.3a Core Cleavage and Polyadenylation Sequence Elements 4
i) Cleavage and Polyadenylation Site
ii) The Cleavage site
iii) Auxiliary Elements
1.1.3b Core Cleavage and Polyadenylation Machinery 5
1.2 Alternative Cleavage and Polyadenylation 9

1.3 Active and Passive APA 11
1.3.1 Regulation of Active APA 13
1.1.3a Strength of the CPA site 13
1.1.3b Trans-factors mediating CPA site choice 13

1.3.2 Regulation of Passive APA 16
1.3.2a 3'UTR-APA and stability 16
1.3.2b 3'UTR-APA and translation 17
1.3.2c 3'UTR-APA and localization 18

1.4. APA Profiles In Context of Different Physiological Environments 19
1.4.1 APA in Proliferating and activated cells 19
1.4.2 APA in Development, Tissue and Cell Types 20
1.4.3 APA in Diseases 21
1.4.4 APA in Stress, Virus and Immunity 22
1.5 Influenza 24

1.6 Influenza A Virion 25

1.7 Viral Life Cycle 29
1.7.1 Virus Attachment and Entry 29
1.7.2 Fusion, Uncoating and Nuclear import 30
1.7.3 Viral Transcription and Replication Cycle 31
1.7.3a Viral transcription 34
i) Cap-snatching 34
ii) Polyadenylation 35
iii) Splicing 35
1.7.3b Viral gene expression 36

1.7.3c Viral Replication 37
i) cRNA synthesis 37
ii) vRNA synthesis 38
1.7.4 Nuclear Export and Trafficking of vRNP Complexes 39
1.7.5 Genome Packaging 40
1.7.6 Virus Assembly, Budding and Release 40

1.8 Influenza Host Shut Off Strategies 41
1.8.1 NS1-mediated Host Shut Off 41
1.8.2 Pol II Degradation by Viral RdRP 43
1.8.3 Host Transcripts Degradation by PA-X 44

1.9 Host Immunity against Influenza A Virus Infection 45

1.10 Aim of Study 48

Chapter 2: Materials and Methods 27

2.1 Cell Culture 49
2.2 Influenza Virus 49
2.2.1 Influenza A/WSN/33 Harvest 49
2.2.Plaque Assay 49
2.2.3 Gradient Ultracentrifugation Purification 50
2.2.4 SDS-PAGE and Silver Stain 50
2.2.5 Immunofluorescence Microscopy 51
2.3 Infection Time Course Experiment 51
2.4 Subcellular RNA fraction 51
2.5 Reverse Transcription PCR 52
2.6 Denaturing Agarose RNA Gel Electrophoresis 53
2.7 Protein Isolation 53
2.8 SDS-PAGE and Western Blotting 53
2.9 Primer Extension Analysis 54
2.9.1 Primers 54
2.9.2 Radiolabelling of Oligodeoxyribonucleotides 54
2.9.3 Primer extension 55
2.10 QuantSeq 3'mRNA Sequencing 55
2.10.1 QuantSeq 3'mRNA-Seq Library Preparation 55
2.10.2 hsDNA Bioanlayzer Assay 56
2.10.3 RNA-seq on Ion Proton Platform 56
2.10.4 Bioinformatics Pipeline to Assess APA and Gene Expression 56
2.10.5 Bioinformatics Pipeline to Assess Infection Dynamics 57
2.10.6 Enrichr Gene Ontology Analysis 57
2.11 Quantitative Real-Time PCR 57
2.12 RNAi Transfection 58
2.13 Cloning 59
2.13.1 Genomic DNA isolation 59
2.13.2 Polymerase Chain Reaction (PCR) 59
2.13.3 Restriction Enzyme Digestion 59
2.13.4 Kinase Reaction 59
2.13.5 Ligation 60
2.13.6 Site-Directed PCR Mutagenesis 60

2.13.7 DPNI Parental DNA Digestion 60
2.13.8 Transformation of Plasmids to Competent Cells 61
2.13.9 Plasmid DNA Isolation 61
2.14 Transient Overexpression of NS1, Polymerase Subunits and Mutants in HEK293T 62
2.15 Luciferase Reporter Gene Assay 62
2.16 Osmotic Shock Treatment 63

Chapter 3: Establishing Controls and Development of the Infection Time Course Experiment 67

3.1 Introduction 67
3.2 Results 68
3.2.1 Determining Biologically Relevant Infection Time Points 68
3.2.2 Establishing an Efficient Viral Infection Protocol for Steady State APA Analysis 72
3.2.3 Rationale for Subcellular Fractionation 78
3.3 Discussion 81

Chapter 4: Investigating Role Influenza Viral Stress on APA and Gene Expression 83

4.1 Introduction 83
4.2 Results 87
4.2.1 Dynamics of Infection 87

4.2.2 Influenza Viral Infection Induces 3'UTR APA and Differential Gene Expression of Host Transcripts 95
4.2.2a Significant APA Changes Induced During Influenza Infection 96
4.2.2b Relating Significant APA Changes to Changes in Differential Gene Expression 101
4.2.2c Viral-Induced UTR APA is Independent of Other Stress Responses 112
4.2.2d RT-qPCR verified APA events 116
4.2.2e 3'UTR APA Results in Downstream Effects in Gene and Protein Expression 120

4.2.3 Impact of Influenza Infection on Coding Region APA of Host Transcripts 125
4.2.3a Influenza Infection Induces Coding Region APA of Host Transcripts 125
4.2.3b Viral-induced CR-APA is Independent of Osmotic Shock Responses 133

4.3 Discussion 135
4.4 Summary 144

Chapter 5: Deciphering Mechanisms of 3'UTR APA Regulation During Influenza Viral Infection 146

5.1 Introduction 146
5.1a The Relationship between the Viral Non-Structural Protein 1 and the Host Cleavage and Polyadenylation Machinery 146
5.1b Aims 149
5.2 Results 150
5.2.1 3'UTR APA Profile Changes After CPSF30 Knockdown 150
5.2.1a Ensure Efficient Knockdown of CPSF30 150
5.2.1b Investigating CPSF30 Knockdown effects on 3'UTR-APA in HEK293T cells 152

5.2.2 Investigating 3'UTR APA Effects of Transient NS1 Expression in HEK293T cells 159
5.2.2a Establishing transient expression of GFP, NS1-G184R and NS1 in HEK293T cells 159
5.2.2b Investigating NS1 and NS1-G184R mutant expression effects on 3'UTR-APA in HEK293T cells 163
5.2.2c Verifying Whether NS1 Abrogation of CPSF30 is the Mechanism Responsible for Regulating 3'UTR APA During Infection 169

5.3 Discussion 171

5.4 Summary 175

Chapter 6: Discussion 177

6.1 Thesis Summary 178
6.2 Advantages and Limitations of 3'end Sequencing for Global APA Analysis 179
6.3 Alternative Mechanisms Responsible for Regulating APA During Infection 179
6.3.1 NS1 Inhibition of Splicing as an Alternative Atrategy for APA Regulation During Infection 182
6.3.2 Enhancement of NS1-CPSF30 Interaction by Viral Polymerase to Regulate APA 183
6.3.3 Influenza Viral Strain Potentially Effects APA? 183
6.4 Conclusion 184

Chapter 7: Supplementary Data 186

References 197

Chapter 1: Introduction

Eukaroytic Pre-mRNA Processing

1.1 Pre-mRNA processing

Nascent protein-coding genes are transcribed by RNA Polymerase II (Pol II) to produce precursor messenger RNAs (pre-mRNAs). In eukaryotes, pre-mRNAs require processing to produce mature mRNA for nuclear export to cytoplasm and translation.

Three essential processing mechanisms are required: 5'end capping, splicing and 3'end cleavage and polyadenylation. While initially, these processes were thought to occur after transcription, instead it is widely accepted that mRNAs can be processed co-transcriptionally (Proudfoot et al., 2002), and the processing machinery has been shown to associate and interact with Pol II (Bentley 2014; Saldi et al. 2016), therefore making mRNA processing more efficient.

1.1.1 5'end capping

5'cap addition is the first processing step of nascent pre-mRNA. First, an RNA 5' triphosphatase hydrolyses the triphosphate of the first nucleotide to a diphosphate. Subsequently, a guanosine cap structure is added to the first nucleotide via a 5'-5' triphosphate linkage catalysed by Gyanylyltransferase. To finalise the cap structure, the terminal guanosine residue is methylated by Guanylyl-7-methyltransferase. The 5'cap is essential for nuclear mRNA export and protection against 5'-3' exoribonucleases (Beelman and Parker 1995). In addition, the 5'cap interacts with translation initiation factor eIF4E and other translation initiation factors to enhance

translation by facilitating the binding of ribosomes to the transcript (Lewis and Izaurflde 1997).

1.1.2 Splicing

Most eukaryotic genes encode transcripts containing alternating exons and introns. Since only exons encode the amino acid sequence of the protein in protein coding genes, mRNA splicing must occur to remove the non-coding introns from pre-mRNA. A large multi-protein-RNA complex (spliceosome) is vital for this two-step reaction which directs precise excision of introns, and fuses the exons together to form the continuous open reading frame (ORF), which is the template for protein synthesis. Proteins are deposited between two exons forming an exon-junction complex (EJC), which can influence the translation and localisation of the transcript.

With a few exceptions, splicing occurs in almost all eukaryotes. In humans, approximately 97% of genes contain introns and undergo splicing (Grzybowska 2012). Alternative splicing, occurring in 95% of mammalian genes and being tissue specific, contributes to increased transcriptome and proteome complexity by the inclusion of different subsets of introns and exons in the mature mRNA.

1.1.3 Cleavage and Polyadenylation

Cleavage and Polyadenylation (CPA) is the final step of pre-mRNA processing at the 3'end. With the exception of replication dependent histone mRNAs in metazoans, all protein coding pre-mRNAs are matured by 2-step CPA into function mRNAs containing a uniform 3'end (Marzluff et al., 2008). First, an endonucleoytic cleavage reaction occurs at a specific location in the transcript. This is proceeded by the synthesis of the poly(A) tail (up to a chain of 250 adenosine residues in humans) in a

template independent fashion. The latter step is catalysed by an evolutionarily conserved enzyme poly (A) polymerase (PAP) (Proudfoot 2011). CPA is an essential prelude to subsequent transcription termination, nuclear export to cytoplasm of mRNA and protection from RNase degradation (Guhaniyogi and Brewer. 2001).

1.1.3a Core Cleavage and Polyadenylation Sequence Elements

i) Cleavage and Polyadenylation Site

The cleavage and polyadenylation (CPA) site, found in 3'UTR of the pre-mRNA, consists of *cis* sequences that flank either side of cleavage site (Figure 1.1A).

The core pA signal is a bipartite sequence element that constitutes a conserved A-rich hexamer sequence, located 21 nucleotides upstream of the cleavage site and a less-conserved GU or U-Rich downstream element (DSE), located 10-30 nucleotides downstream of cleavage site. Although the most common (70-75% of all types of pA signals) pA signal hexamer sequence is AAUAAA/ AUUAAAA, termed canonical poly(A) signal, more than 10 variations of non-canonical hexamer sequences were identified, where more are yet to be discovered (Tian et al., 2005). Mutations or removals of the hexamer sequence can significantly reduce CPA (Fitzgerald and Shank., 1981; Montell et al., 1983; Wickens and Stephenson., 1984). These laboratory studies were further strengthened by naturally occurring polymorphisms found in the human global genes, which are associated with the thalassemia phenotype (Orkin et al., 1985; Chen et al., 2006; Danckwardt et al., 2008; Higgs et al., 1983). Contrary to the hexamer, the DSE has been considered to much more tolerant to single nucleotide changes (Zhao et al., 1999). With the exception of a subset of genes with degenerated hexamer sequences, base changes in the DSE reduced 3'end processing significantly (Nunes et al., 2010).

Numerous factors contribute to the strength of a poly(A) site including the presence of canonical hexamer sequence, the composition and distance of the DSE from the hexamer and presence of additional upstream aux elements (which can enhance recruitment of CPA factors).

ii) The Cleavage Site

The Point of Cleavage occurs between hexamer and DSE. In humans, cleavage usually occurs after a CA or UA dinucleotide (Li and Du., 2013). However, actual site where cleavage occurs is known to be heterogenous.

iii) Auxiliary Elements

In addition to the core sequence elements, the strength of CPA sites can be influenced by the presence of additional auxiliary sequences. These enhancers can be found both downstream and upstream of the cleavage sites. For instance, downstream G-rich quadruplex structures, or downstream and upstream U-rich regions can have an impact on enhancing cleavage efficiency (Neve et al., 2017).

1.1.3b Core Cleavage and Polyadenylation Machinery

A multi-protein machinery assembles onto the CPA site to enable the 2-step cleavage and polyadenylation reaction. More than 80 proteins are associated with the CPA machinery (Shi et al., 2009), although around 20 are considered to be core factors (functional details displayed in Table1.1). Part of this core machinery includes 4 highly conserved core factors: Cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), mammalian cleavage factors I and II (CFIm and CFIIm). Additional core factors comprises symplekin, PAP, nuclear poly(A)

binding protein (PABPN1) and CTD of the largest subunit of RNA polymerase II (Pol II).

3’end formation is initiated by the recognition of the hexamer and downstream GU rich region by Cleavage and Polyadenylation Specificity Factor (CPSF) and Cleavage Stimulation Factor (CstF). CPSF and CstF have multiple subunits, and they complex in a cooperative manner (Murthy and Manley., 1995) as outlined in figure 1.1B. Endonuclease cleavage reaction is subsequently mediated by CPSF-73/100 (Mandel et al., 2006).

The resulting 3’end product of cleavage reaction is degraded by hPcf11, while the 5’ end remains bound by the machinery. hFip1 and CPSF-160 recruits PAP to catalyse poly A synthesis. Although the addition of adenosine is initially slow, but it is accelerated upon recruitment of nuclear poly(A) binding protein (PABPN1) (Wahle 1991) with increasing PAP affinity to the RNA by 80 fold (Kerwitz et al., 2003)). Additionally, the speed of assembly of a functional CPA complex is dependent on the strength of the CPA site but is generally achieved within about 10 seconds (Chao et al., 1999).

As with the co-transcriptional relationship of CPA , 5’capping and splicing, the CPA machinery interacts with the transcriptional machinery, specifically the C-terminal domain of Pol II, which increases efficiency of pre-mRNA processing (Bentley 2014).

The CPA reaction is summarised in Figure 1.1C.

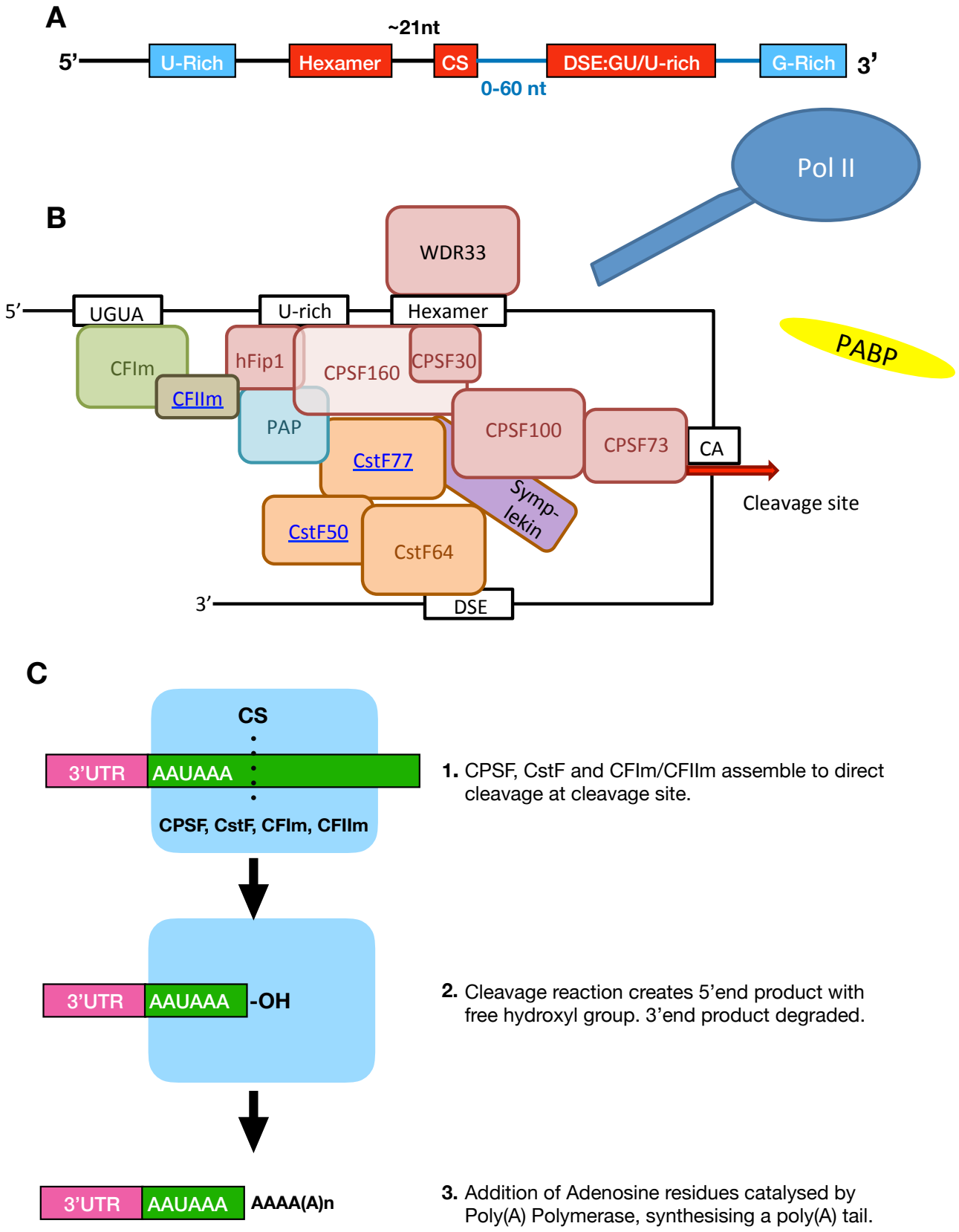


Figure 1.1 Schematic Overview of Cleavage and Polyadenylation

Details of figures described in Chapter 1.1.3

A. Cis-elements present at the 3'end of mRNA which mediates cleavage and polyadenylation

B. The core factors of cleavage and polyadenylation machinery.

C. Schematic outline of cleavage and polyadenylation reaction in 3'end formation. CS represents cleavage site. AAUAAA represents canonical hexamer signal.

Figure adapted from Neve et al., 2017.

Table 1.1. Core factors of Cleavage and Polyadenylation Machinery (Neve et al., 2017)

Factor	Function
CPSF	**Recognises the hexamer poly(A) signal**
CPSF30	Recognises and interacts with hexamer sequence (Sun et al., 2018) . Central role in pA site choice in plants (Thomas et al., 2012).
WDR33	Large protein of 146kDa. Binds directly to AAUAAA hexamer (Schöemann et al., 2014). Possess WD40 domains to mediate protein-protein interactions with other CPSF subunits.
CPSF73/100	Have endonuclease activity to perform cleavage reaction
CPSF160	Meidates interaction with hexamer sequence. Important for CstF interaction. Involved with poly(A) polymerase interaction.
hFip1	Interacts with poly(A) polymerase. Binds RNA at upstream or downstream U-rich regions.
CstF	**Recruited to DSE. Essential for cleavage but not polyadenylation**
CstF64	Mediates interaction between CstF and pre-mRNA downstream of poly(A) signal via RNA recognition Motif (RRM). Biinds to U-rich region strongly.
CstF77	Connecting CstF64 and CstF50 subunits forming CstF complex. Facilitates interaction between CstF and other proteins such as CPSF160.
CstF50	Contains 7 repeats of WD40 motifs which mediate protein-protein interactions with Pol II CTD.
CFIm	**Stabilises the binding of CPSF complex to the pre-mRNA. Crucial factor in recognising non-canonical poly(A) signals. 4 subunits: CFIm25/59/68/72. CFIm25/68 interact with RNA via RRM.**
CFIIm	**Required for cleavage but not polyadenylation.**
CFIIAm	Contains cleavage reaction components.
CFIIBm	Stimulates cleavage reaction.
hCIP1	Interacts with CPSF and CFIm, may bridge these two proteins during cleavage reaction.
hPcf11	Degradation of RNA associated with Pol II after cleavage.
Symplekin	**Provides Scaffolding for the CPA machinery. Role in histone 3'end formation. Required for cytoplasmic polyadenylation.**
Poly(A) Polymerase (PAP)	**Catalyses addition of adenosine residues to 3'end of cleaved pre-mRNA. Recruited to machinery by hFip1 and CPSF. Uses ATP as substrate.**
PABPN1	**Recruited to CPSF and PAP to accelerate addition of adenosine to poly(A) tail. May modulate this function by interacting with pA signal, CPSF and poly(A) polymerase.**
RPB1	**Large subunit of Pol II CTD. Provides platform that couples pre-mRNA processing to transcription. Required for efficient cleavage.**

1.2 Alternative Cleavage and Polyadenylation

Alternative cleavage and polyadenylation (APA) plays a fundamental role in regulating transcriptome complexity and gene expression in eukaryotes. With ~70% of all human genes containing multiple CPA sites, these genes can undergo APA, whereby the alternative usage of multiple CPA sites can generate multiple distinct RNA transcripts from a single gene (Derti et al., 2012). Depending on the relative positions of these CPA sites in a gene, APA events can be distinguished into 2 separate categories: untranslated region (3'UTR) APA and coding region APA (CR-APA) (Figure 1.2). With the exception of some genes containing only one CPA site in the 3'UTR that is used for cleavage and polyadenylation, they produce mature transcripts with the same UTR length which translate into a single protein product. This is called constitutive polyadenylation (Figure 1.2A).

3'UTR APA occurs when alternative CPA sites are found in the 3'UTR, downstream of terminal exon (Figure 1.2B). Since they are positioned outside of the coding region, differential usage of CPA sites does not result in alteration of transcript coding capacity, but instead may influence the transcript stability, localisation, translational efficiency and gene expression. The fates of 3'UTR APA transcript isoforms can be influenced by the presence or absence of *cis* regulatory elements present in the alternative UTR. It is not always clear what proportion of UTR-APA events have a physiological impact, but reports have suggested highly expressed genes tend to utilise the proximal CPA sites and thus have shorter 3'UTRs, whereas lowly expressed genes are more likely to use distal CPA sites (Ji et al., 2009; Ji et al., 2011). The majority of studies in the APA field have focussed on investigating 3'UTR APA, however CR-APA has also been shown to be relevant, whereby 40% of APA events in mice were identified to be CR-APA.

Coding region APA involves multiple CPA sites present at positions upstream of the 3'UTR in the coding region (either in exonic or intronic region) (Figure 1.2C). APA events can produce APA mRNA isoforms with different coding capacity and encode different proteins that potentially have altered functions. Additionally, CR-APA generated isoforms do not introduce a premature translational termination codon, therefore are not targeted by Nonsense Mediated Decay (NMD) degradation pathway, facilitating translation of a truncated protein, with potentially altered function (Lejeune and Maquat. 2005). The classic example of CR-APA occurs in the gene encoding IgM antibody. Upon B cell activation CstF-64 enables CPA site switching from using the poly A site in the terminal exon, which in resting B cells produce a longer membrane bound form, to using a poly A site in an upstream exon producing a truncated protein (Takagaki et al., 1996).

Figure 1.2 Alternative Cleavage and Polyadenylation.

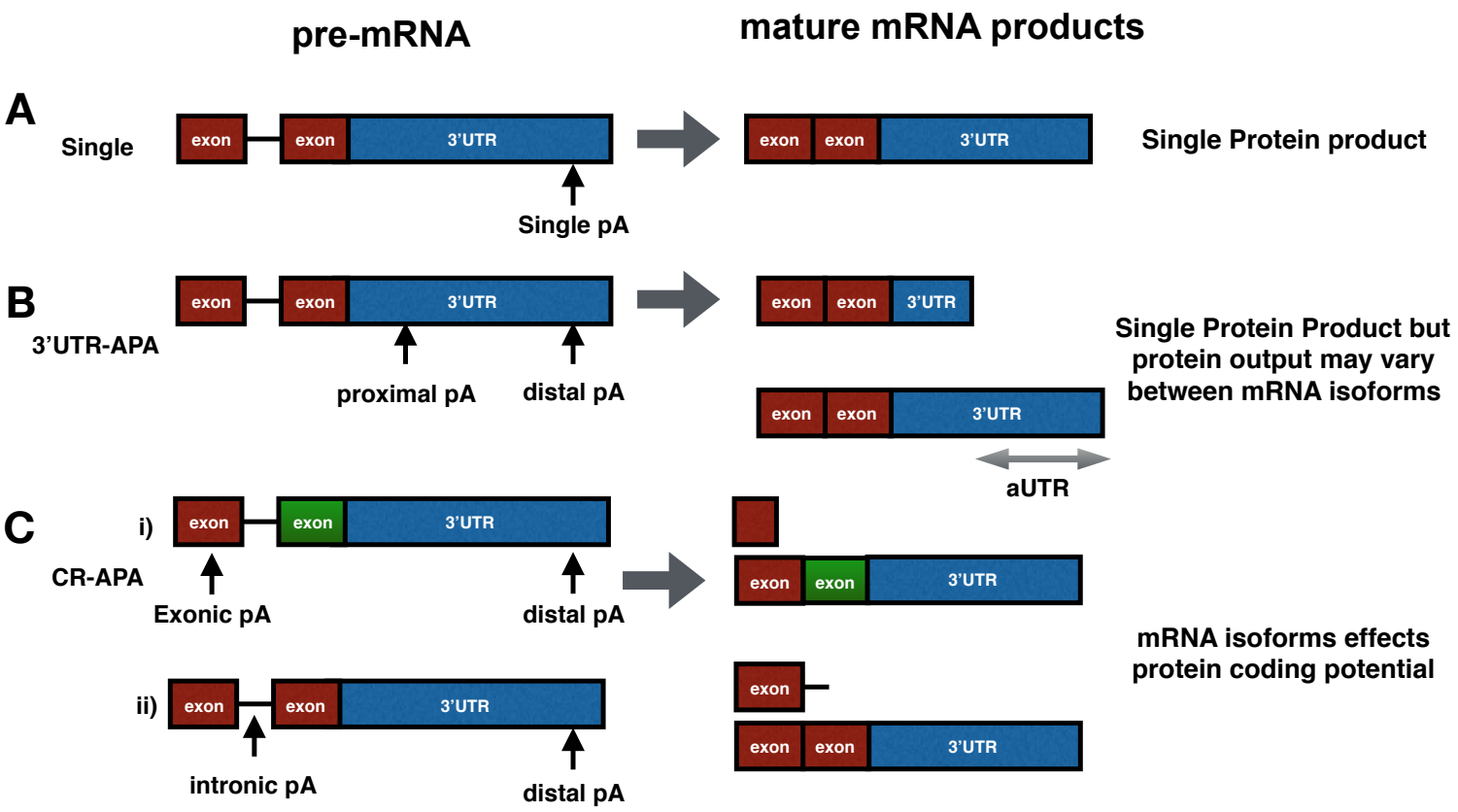


Figure 1.2 Schematic Overview of Alternative Cleavage and Polyadenylation

Details of figures reviewed in Chapter 1.2

A. Constitutive polyadenylation: some genes containing only one CPA site in the 3'UTR that is used for cleavage and polyadenylation, and therefore produces mature transcripts with the same UTR length and single protein product.

B. 3'UTR APA: Usage of an alternative CPA site in the UTR will not affect encoded protein, but will affect UTR length which may harbour regulatory cis-elements on alternative UTR (aUTR) such as microRNA target sites, RNA protein binding sites and destabilising elements.

C. CR-APA: Upstream CPA sites can be found in upstream i) exon or ii) intron. Usage of these upstream CPA sites affects protein coding potential and product of the transcript.

1.3 Active and Passive APA

In order to unravel the physiological role of global APA changes, it is vital to determine the mechanisms that are responsible for shaping these profiles. Thus, a distinction has to be drawn between alterations of APA profiles that are due to post-transcriptional mechanisms, or due to specific selection of one CPA site over another at the point of cleavage.

In the case of “active” APA, APA profiles changes are resulted from cleavage site selection at a co-transcriptional level. Active APA can be accomplished by changing the availability of factors from the CPA machinery that either enhance or suppress the usage of a particular CPA site over another in response to specific cues or change in state of a cell. These changes will be seen in nucleus (Figure 1.3a).

Conversely, “passive” APA results in APA profile changes that are modulated from post-transcriptional events such as miRNA targeting, nuclear export, and stability of specific APA mRNA isoforms. Effects will be seen in cytoplasm (Figure 1.3b).

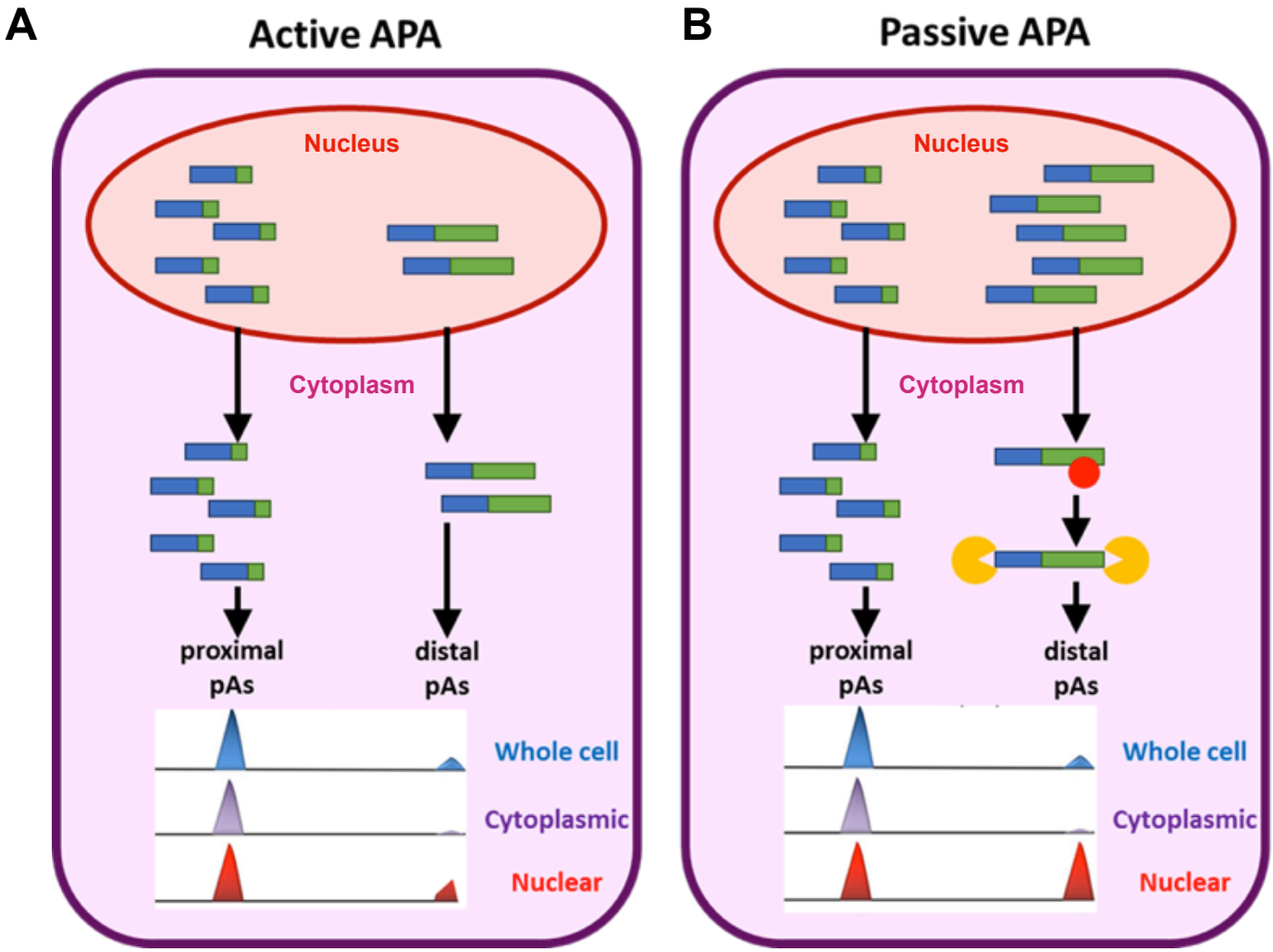


Figure 1.3 Active and Passive APA

Details of figures reviewed in Chapter 1.2

A. Active APA occurs when the abundance of short and long 3'UTRs is determined at the point of cleavage in the nucleus. The effects will influence whole-cell APA profiles.

B. Passive APA occurs when the abundance of short and long 3'UTRs is determined by post-transcriptional regulatory mechanisms. The effects will be seen in the cytoplasm. As an example, the binding of an RBP (red circle) targets the long isoform for exonuclease degradation (yellow Pacman). If analyses were done with whole cell RNA, APA profiles will look that same as with active APA. Using subcellular fractionation approach allows differentiation of how the transcriptome is regulated. Figures adapted from Neve et al., 2017.

1.3.1 Regulation of Active APA

1.3.1a Strength of the CPA site

The strength of the CPA site can be modulated by the presence of canonical cis-elements. Usually stronger CPA sites result in increased usage compared to weaker ones. The distal CPA sites are more likely to display the canonical hexamer poly(A) signal, and have a GU or U rich DSE. Proximal CPA sites are more likely to lack such signals but instead have non-canonical structure including G-rich and CU-rich elements (Martin et al., 2012; Shepard et al., 2011; Beaudoing and Gatheret. 2001), making it less likely for CPA machinery to recognise them efficiently. Thus, distal CPA sites are more likely to be stronger and are generally conserved from yeast to humans (Smibert et al., 2012). Furthermore, many evidence has suggested distal CPA sites are used by default, however proximal poly (A sites) are used to exclude regulatory elements found in the UTRs (Di Giammartino et al., 2001).

1.3.1 b *Trans*-factors mediating CPA site choice

Since cis-elements influencing polyadenylation efficiency do not change, the fluctuation of several *trans*-factors such as core CPA site factors and RNA binding proteins have been shown to influence CPA site choice. Additionally, transcriptional regulatory mechanisms such as RNA pol II elongation rate can also regulate CPA site choice.

Altogether, as summarised from table 1.2 that is adapted from Neve et al., 2017, there are ubiquitous amounts of *trans*-factors, when depleted or compromised, that will mediate CPA site usage during 3'end processing of the pre-mRNA.

Table 1.2 Trans-factors known to influence CPA site choice at the point of cleavage.

Factor	Function	Reference
CPA factors		
CFIm	Promotes usage of distal CPA sites by binding to and suppressing the usage of proximal suboptimal poly(A) sites. Specifically binding and activating enhancer-containing CPA sites.	(Martin et al., 2012) (Masamha et al., 2014) (Zhu et al., 2018)
CFIm25	Down-regulation of CFIm25 in non-small cell lung cancer causes 3'UTR shortening and regulation of oncogenes APA (IGF1R) and promote their protein expression. Downregulation of NUDT21 (responsible for encoding CFIm25) leading to depletion of CFIm25 also observed in hepatocellulr carcinoma.	(Huang et al., 2018) (Tan S et al., 2018) (Sun et al., 2017)
CstF64	Binds to U-Rich regions. Depletion of CstF64 leads to shifts of usage to distal CPA sites in a small number of genes. Elevated CstF64 levels increases usage of weaker proximal poly(A) site.	(Takagaki et al. 1996) (Hwang et al. 2016)
βCstF64	A neuronal splice variant of Cstf64 that associates with the Cstf complex and stimulates and activate weaker proximal CPA sites.	(Shankarling and MacDonald, 2013)
τCstF64	Promotes usage of proximal CPA site specifically in the brain and testis. Depletion leads to usage of distal CPA site.	(MacDonald and McMahon, 2010)
CstF77	Regulates its own expression by associating itself in a negative feedback loop. Increased levels of CstF77 activates intron 3 CPA site to produce truncated transcript.	(Luo et al., 2013)
hFip1	Component of CPSF complex. Regulation of APA by Fip1 is dependent on the distance between CPA sites. When CPA sites are close together, FIP1 blocks CstF binding at the proximal site and therefore results in distal CPA site usage. Conversely when far apart, Fip1 promotes usage of weaker proximal CPA sites. Negatively regulated by Small nuclear RNAs (SnoRNA) SNORD50A.	(Lackford et al., 2014) (Li et al., 2015) (Huang et al., 2017)
hPcf11	Enhances the use of proximal CPA sites by binding directly to pre-mRNA.	(Li et al., 2015) (West et al., 2008)
PABPC1	Similar to PABPN1, it promotes distal CPA site usage. Although, no obvious cis-elements are enriched around CPA sites regulated by PABPC1.	(Li et al., 2015)
PABPN1	Promotes usage of distal CPA sites by competing with CPSF for CPA site binding and by inhibiting usage of proximal weaker CPA sites. Depletion of functional PABPN1 in OPMD causes widespread 3'UTR shortening.	(Jenal et al., 2012) (De Klerk et al., 2012) (Simonelig et al., 2012) (Beaulieu et al., 2012)
other RNA-binding factors		
αCP	Enhances usage of CPA sites that are close to C-rich motifs	(Ji et al., 2013)
CIRBP/ RBM3	Binds to GNNGNNG motif in 3'UTR. Inhibition of proximal CPA site usage caused by increase expression in cold shock.	(Liu et al., 2013)

CPEB1	By shuttling into nucleus, mediates usage of CPA sites close to CPEs and recruits CPA machinery to use proximal CPA site during proliferation.	(Bava et al., 2013)
ELAVL1	Binds AU-rich elements in 3'UTR. Can regulate both APA and splicing.	(Mansfield & Keene 2012; Zhu et al. 2007).
FUS	when bound downstream of CPA site, it enhances CPA efficiency via CPSF160 recruitment. However, when binding site is upstream of a CPA site, premature RNA pol II termination occurs resulting in down regulation of that CPA site.	(Masuda et al. 2015)
hnRNP A2/ B1	Binds to UAGG motif in 3'UTR. 3'UTR lengthening demonstrated when depletion of hnRNP A2/B1 occurs in pluripotent stem cells.	(Martinez et al., 2016)
hnRNP C	Binds U-rich sequences and make nearly CPA sites and suppress usage.	(Gruber et al., 2016)
hnRNP F	Binds G-rich DSE and competes with CstF64 and suppress usage of nearby CPA site.	(Alkan et al., 2006)
hnRNP H	H1 binds DSE to promote proximal CPA site usage. H2 binds G-rich regions to enhance CstF64 binding. High expression of H2 in proliferating cells enhances proximal CPA site usage. Opposite seen if H2 depletion occurs.	(Katz et al., 2010) (Bagga et al., 1998) (Shepard et al., 2011)
hnRNP K	Binds UCCCUU motif in 3'UTR and suppress usage of nearby CPA site.	(Naganuma et al., 2012)
Mbnl	Binds to R/YGCY motif. Binding nearby CPA site inhibits its usage due to steric hindrance. Conversely, binding upstream of CPA sites enhances its usage. Binding can alter localisation of transcript.	(Batra et al,. 2014) (Taliaferro et al., 2016)
Nova	Neural specific factor that binds YCAY motifs in the 3'UTR and can influence CPA site choice depending on the location of these motifs.	(Licatalosi et al., 2008) (Hu et al., 2014)
PABPC1	Promotes distal CPA site usage.	(Li et al., 2015)
U1 snRNP	Suppresses the usage of cryptic CPA sites and enable full length transcription.	(Kaida et al., 2010)
Other Features and conditions that can influence CPA site choice		
DICER	In the nucleus, Dicer can modify the chromatin environment to alter pol II elongation rate and thus affect CPA site usage. In closed chromatin conditions, pol II transcribes at a slower rate and enhance proximal CPA site usage.	(Neve et al., 2016)
Nucleosome positioning	Depletion of nucleosomes can increase exposed CPA site usage.	(Spies et al. 2009) (Huang et al. 2013)
Transcription rate	Slow transcription rate results in longer duration between proximal and distal CPA sites thereby causing increased frequency of proximal CPA site usage.	(Neve et al., 2016) (Pinto et al., 2011) (Wood et al., 2008)
Pol II elongation rate	Promotes proximal CPA site usage. Transcription pausing at G-rich regions can also enhance nearby CPA site usage.	(Pinto et al. 2011) (Liu et al., 2017) (Saldi et al., 2018)
Chromatin state	Open chromatin state favours proximal CPA site usage resulting in active APA regulation of 3'UTR shortening	(Li et al. 2016)

1.3.2 Regulation of Passive APA

Cis-elements present in the 3'UTR have a potential for post-transcriptional regulation of APA isoforms. Binding of *trans*-factors, for instance RNA binding proteins (RBP) and microRNAs (miRNAs), to cis-elements can influence transcript stability, protein output and localisation.

1.3.2a 3'UTR-APA and stability

Global studies have shown evidence of significant miRNA mediated regulation of mRNA destabilisation and translational repression (Guo et al., 2010; Hendrickson et al., 2009), where >60% of human genes contain miRNA target sites (Friedman et al., 2009). miRNAs bind to complementary regions in the UTR. Once bound, miRNA induced silencing complex (miRISC) is assembled to either inhibit translation or trigger mRNA degradation (Bartel et al., 2004). mRNA degradation occurs primarily through mRNA deadenylation, mediated by recruiting CAF1-CCR4-NOTI deadenylase complex through binding of GW182 to the polyA-binding protein (PABP) (Eulalio et al., 2009).

The presence of destabilising elements such as miRNA target sites and motifs for RBP in the UTR, the binding of which recruits RNA degradation machinery causing RNA decay, therefore 3'UTRs are thought to have a repressive role. Truncated 3'UTRs are assumed to result in stabilisation, promotion of translation and increased gene expression of such transcripts due to the removal of potentially repressive cis-elements in the 3'UTR (Morris et al., 2012). However, studies have discovered that 40% longer UTR isoforms were more stable than their shorter counterparts due to binding of several RBPs, and a multitude of stabilising cis-elements have now been identified. Despite evidence of miRNA and RBP regulation of specific genes and

significant differences in half-lives of up to 35% of alternative 3'UTRs (Mayr 2016), global impact of APA on mRNA stability (Spies et al., 2013) and mRNA abundance (Jia et al., 2017; Lianoglou et al., 2013) is fairly modest. A recent analysis reinforced this claim, discovering ~10% of all APA events in HEK293 cells are found to be regulated by post-transcriptional passive APA and around 3% of those are destabilised by miRNAs. Furthermore, this study showed that miRNAs can target both short and long UTR-APA isoforms for degradation (Neve et al., 2016). The modest global impact of APA on mRNA stability in mammalian cells however is not universal. In yeast, single nucleotide changes in 3'UTR length could affect transcript stability (Geisberg et al., 2014; Gupta et al., 2014).

1.3.2b 3'UTR-APA and translation

APA events have been suggested in many studies to alter translational efficiency, where 3'UTR length is correlated with protein output (Yoon et al., 2013; Akman et al., 2012; Pinto et al., 2011; Tranter et al., 2011; Mayr and Bartel, 2009). 3'UTR shortening was found, due to proposed lack of negative regulatory elements, to be associated with increased stability and consequently increased translational efficiency (Mayr 2016; Mayr and Bartel, 2009). However, UTR length has a minimal effect on protein output (Gruber et al., 2014), and opposing effects have been seen where the transcript isoforms with extended UTRs are translated with higher efficiency (Yoon et al., 2013; Spies et al., 2013; Pinto et al., 2011). Ultimately, despite individual examples of APA mediated shortening of UTRs of genes (CDC6, HSPA2, ECE-1, CCND2 and DICER1) leading to higher translational output, the correlation between UTR length and translation efficiency is defined as complex and not necessarily linear.

1.3.2c 3’UTR-APA and localisation

UTRs can play an important role in directing the subcellular localisation of transcripts and enable localised protein translation. 3’UTR lengthening in the brain has been shown to be important in controlling the localisation of APA mRNA isoforms (Miura et al., 2013). For example, α-synuclein transcripts with extended 3’UTR (associated with Parkison disease pathology), not only increase translational output but also affects their localisation away from the synaptic-terminals to the mitochondria (Rhinn et al., 2012). Global studies comparing RNA isolated from the stoma in neurons and dendrites found that distal APA isoforms induced during neuronal differentiation are more likely to be localised in neurite projections than their proximal counterparts (Taliaferro et al., 2016).

3’UTRs can also mediate mRNA localisation to organelles such as the endoplasmic reticulum (ER) (Loya et al., 2008), mitochondria (Marc et al., 2002), and to cytoplasmic RNA granules to isolate transcripts from the translation machinery (Han et al., 2012). Recently, 3’UTR-directed localisation to organelles has been demonstrated with distal isoform of CD47, which allows binding of SET and HuR resulting in localisation to the plasma membrane, where it is translated. Conversely, protein produced by the shorter UTR isoform was retained in the endoplasmic reticulum (Berkovits and Mayr, 2015).

Furthermore 3’UTR can also influence nuclear export. Splicing factors SRSF3 and SRSF7 shown regulate CPA site selection and selective nuclear cytoplasmic transport by recruiting export factor NXF1. SRSF3 promotes distal CPA site usage and controls export of long 3’UTR APA isoforms. Conversely, SRSF7 promotes proximal pA site usage thereby controlling export of short 3’UTR APA isoforms

(Müller-McNicoll et al., 2016). Additionally, a recent global study identified significant number of nuclear were retained APA mRNA isoforms when comparing nuclear and cytoplasmic APA profiles in HEK293 cells. Some of these APA isoforms possess intron retention resulted from incomplete splicing, which is proposed to prevent nuclear export into the cytoplasm (Neve et al., 2016).

1.4 APA Profiles In Context of Different Physiological Environments

Advances in next generation sequencing have allowed global transcriptome analyses of cellular APA profiles. This propagated a huge number of studies looking at general shifts in APA profiles upon a change in cellular state, differences in tissue type, and exposure to stress and diseases. Interestingly, after performing 3'end sequencing of 1.2 billion reads from more than 360 human, mouse and rat samples, a recent study has shown that approximately 80% of mammalian genes contain at least one conserved CPA site and approximately 50% have conserved APA events (Wang et al., 2018). Conservation of APA correlates with gene expression, function and age, in particular with genes associated with cell morphology ,cell proliferation, and mRNA metabolism. Whereas tissue specific genes with low APA rate such as brain specific genes tend to evolve APA (Wang et al., 2018). An overview of these studies and findings will be covered in this section.

1.4.1 APA in Proliferating and activated cells

3'UTR shortening (pattern of shifting usage of the proximal CPA sites) has been seen upon increased cellular proliferation (Elkon et al., 2012; Sandberg et al., 2008). Studies of murine T-lymphocyte activation demonstrated 3'UTR shortening, as opposed to CR-APA, where equal movements towards both proximal and distal CPA sites has been seen upon cell activation (Sandberg et al., 2008). In general, studies

comparing an extensive range of both murine and human cell lines affirmed that proliferation rate was inversely correlated with 3'UTR length (Sandberg et al., 2008).

Moreover, a study investigating membrane depolarisation in synapse development of hippocampal rat neurons results in 3'UTR shortening in a specific subset of genes associated with synaptic control (Flavell et al., 2008). A majority of the cases are CR-APA, which results in truncation of transcripts and protein with altered functions.

1.4.2 APA in Development, Tissue and Cell Types

3'UTR lengthening during development and differentiation is generally conserved amongst species (Wang et al., 2018; Hu et al., 2017; Hoque et al., 2013l; Ulitsky et al., 2012; Ji et al., 2009). The initial discovery of this phenomenon in murine C2C12 myoblast cells observed increased usage of distal CPA sites in differentiation and development (Ji et al.,2009). 3'UTR lengthening presented an opportunity for gene expression fine-tuning due to the presence of regulatory elements within the transcript 3'UTR. Conversely, 3'UTR shortening was observed in induced pluripotent stem cells from fully differentiated somatic cells (Ji and Tian, 2009). Further evidence of significant 3'UTR shortening was seen in a recent study that demonstrated, through Nudt21 RNA-processing factor suppression, enhanced generation of induced pluripotent stem cells and shift in proximal CPA site usage in genes that potently influence cell fate transitions (Brumbaugh et al., 2018). The same has been observed when studying spermatogenesis in mouse, where shortening of 3'UTRs avoids destabilising and transposable elements that are located in the aUTR of these genes (Li et al., 2016).

When comparing specific tissues and cell types, they appear to have characteristic APA profiles (Neve et al., 2016). Striking differences in APA profiles has been observed between mammalian and inveterate brain and testes. In the brain usage of the distal CPA site is favoured, whilst the opposite is seen in the testes (Miura et al., 2013; Smibert et al., 2012; Wang et al., 2008). In the testis, in addition to simple switches between common CPA sites, unique CPA sites are activated. These unique CPA sites tend to have a lower incidence of the AAUAAA canonical hexamer, and have unique upstream and downstream elements (Liu et al., 2007). Since these sites are inefficiently used in somatic cells, it has been proposed that testis-specific CPA factors may be responsible for CPA site switching (MacDonald and McMahon, 2010), such as CstF64 variant (τCstF64) (Dass et al., 2002).

1.4.3 APA in Diseases

Global shifts to proximal CPA sites leading to 3'UTR shortening has been well observed in cancers (Huang et al., 2018; Xia et al., 2014; Masamha et al., 2014; Lin et al., 2012; Fu et al., 2011; Mayr and Bartel 2009; Sandberg et al., 2008). It is thought that 3'UTR shortening in cancer functions to exclude instability cis-elements or microRNA target sites present in the alternative UTR, therefore consequently increasing mRNA and protein expression of specific oncogenes including CCND1 (Mayr and Bartel, 2009) and IGF1R by depleting CFIm25 (Sun et al., 2017; Huang et al., 2018; Tan et al., 2018). However, this is not always the case in all cancerous cell lines in which significant number of genes possess 3'UTR lengthening (Fu et al., 2011; Lin et al., 2012; Singh et al., 2009), proving the complexity of APA regulation in cancer. Although, APA profiles may prove valuable as biomarkers for cancer progression and prognosis (Lembo et al., 2012; Morris et al., 2012).

Many APA events have been linked with specific disease phenotypes such as immunodysregulation polyendocrinopathy enteropathy X-linked (IPEX) syndrome and thalassemia (Bennett et al., 2001; Whitelaw and Proudfoot, 1986; Orkin et al., 1985). In the case of these diseases, they are generally resulted from mutations of CPA site cis-elements of one specific gene. Mutations in genes encoding the cleavage and polyadenylation machinery have also been impacted in disease (Curinha et al., 2014; Tian and Manley 2013). Additionally, mutation of PABPN1, known regulator of APA, leads to loss of function and general shift to proximal CPA sites in a subset of genes that leads to oculopharyngeal muscular dystrophy (Jenal et al., 2012; de Klerk and Venema, 2012; Brais et al., 1998).

1.4.4 APA in Stress, Virus and Immunity

Few studies have explored the contribution of APA to stress response. Global 3'UTR lengthening has been observed in cold temperature stress (Liu et al., 2013) as well as in UV induced DNA damage coupled with increased usage of the terminal exon CPA site (CR-APA) (Graber et al., 2013). On the other hand, heat shock stress causes 3'UTR shortening in individual heat shock protein genes (Kraynik et al., 2015; Tranter et al., 2011). Virus infections are also known to induce APA. Transcriptome wide studies of Herpes Simplex virus (HSV) and Vesicular stomatitis virus (VSV) viral infection have shown to cause global shifts to 3'UTR shortening thus contributing to regulation in innate immunity and host shutoff (Jia et al.,2017; Hu et al., 2016; Tang et al., 2016). In the case of HSV, while HSV infected cell culture polypeptide 27 (ICP27) has been shown to inhibit splicing of certain introns in less than 1 % of cellular genes, to promote alternative downstream 5' splicing sites usage and to retain specific introns, it is also responsible for inducing expression of prematurely cleaved and polyadenylated pre-mRNAs at cryptic intronic CPA sites. Thus this leads

to reduced expression while increasing protein coding diversity of targeted genes know for regulating innate immunity pathways (Tang et al., 2016).

Ultimately, there are no consensus trends regarding altered APA profiles in stress.

1.5 Influenza

Influenza viruses are accountable for seasonal epidemics and occasional worldwide pandemics that cause significant mortality and morbidity. From December 2017, the World Health Organisation (WHO) newly estimates influenza epidemics result in between 290,000 to 650,000 of ever-increasing annual deaths worldwide (WHO, 2017). This is a significant increase from last year's estimation, which came up to be between 250,000 to 500,000 deaths worldwide. Seasonal epidemics arise from antigenic drift, a mechanism by which slow accumulation of mutations driven by the error-prone viral RNA polymerase allows mutants to evade the host adaptive immune response. On the other hand, worldwide pandemics occur from antigenic shift events, where the transmission or reassortment of influenza virus strains from non-human species to humans (zoonotic virus) can result in a global spread accompanying increased mortality rate due to a lack of cross protection. Influenza is a highly contagious disease of the respiratory tract, displaying respiratory symptoms such as fever, cough, throat pain, nasal congestion, sneezing, headache, muscle ache, and malaise.

Influenza viruses belong to the Orthomyxoviridae family which is divided into 7 genera: *Influenzavirus A, B, C, D, Thogotovirus, Isavrius and Quaranjavirus* (Palese and Shaw, 2013). Orthomyxoviridae are categorised as negative-sense single stranded RNA viruses containing a segmented RNA genome, which replicate in host nucleus. Having segmented RNA genomes allows genetic reassortment and genetic information exchange between viruses of the same genus, but not across types (Palese and Shaw., 2013). Different genera of influenza viruses infect different species. While influenza B (humans, seals and ferrets), C (human, dogs and swine) and D (cattle) only infect a relatively low range of hosts, influenza A viruses infect

more diverse range of species including aquatic birds as their primary reservoir, humans, swine, bats, horses, whales, seals and non-aquatic birds (Neumann and Kawaoka., 2015).

Influenza A viruses (IAVs) can be further divided into subtypes on the basis of 2 glycoproteins, haemagglutinin (HA) and neuraminidase (NA), on the surface of the virus. Currently, there are 19 known HA (H1-H19) and 11 know NA (N1-N11) subtypes. Many different combinations of HA and NA proteins are possible.

1.6 Influenza A Virion

Influenza virion comes in different shapes including spherical, filamentous or bacilliform morphologies (Wright et al., 2013). Spherical virions are roughly 100nm in diameter, whereas filamentous particles can be in excess of 300nm in length.

Enveloped with a lipid bilayer, which is acquired during budding from the host cell membrane, trimeric haemagglutinin (HA), tetrameric neuraminidase (NA) membrane glycoproteins and tetrameric M2 ion channels are found embedded within the lipid envelope. The approximate ratio of HA to NA is around 4 to 1, with M2 tetramers being far less abundant (Palese and Shaw, 2013; Harris et al., 2006). Within the envelope, the inner shell is a scaffold formed by matrix protein M1.

The viral genome of Influenza A is packaged in eight individual viral ribonucleoprotein (vRNP) complexes (Figure 1.4). Each of the vRNPs is composed of a single viral RNA (vRNA) segment that is coated by viral nucleoprotein (NP) and bound by a heterotrimeric viral RNA-dependent RNA polymerase (RdRP) at the 5’ and 3’ ends , which consists of polymerase basic 1 (PB1), polymerase basic 2 (PB2) and

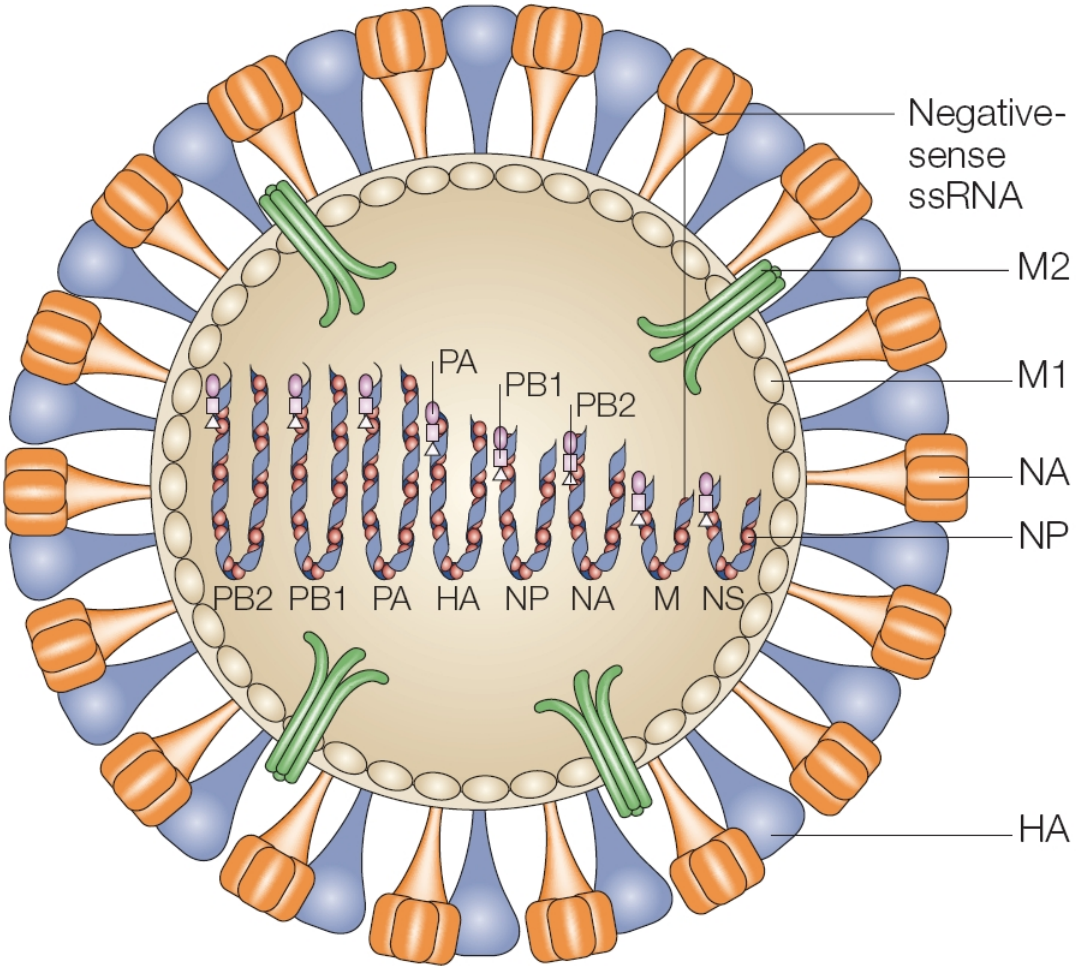


Figure 1.4 The influenza A virion
This is a schematic of the influenza A virion showing all structural proteins and vRNP complexes of the 8 individual viral gene segments. Details are described in Chapter 1.6. (Figure is adapted from Nelson and Holmes, 2007)

polymerase acidic (PA) (Arranz et al., 2012; Moeller et al., 2012). Non-structural protein 1 (NS1) and nuclear export protein (NEP) have also been found inside influenza A virions, but it remains a mystery as to whether their incorporation is required or an accidental by-product (Richardson and Akkina, 1991; Hutchinson et al., 2014). Additionally, influenza viruses encode several auxiliary proteins such as PA-X and PB1-F2 that are responsible for a wide range of functions during infection but are not known to be incorporated into influenza virions (Vasin et al., 2014).

Interestingly, some host factors are also integrated into influenza virions (Shaw et al., 2008).

A summary of each influenza A virus genome segments, encoded proteins and function is presented in Table 1.3.

Table 1.3 influenza A virus genome segments, encoded proteins and function

Segment	Length (nucleotides)	Encoded protein(s)	Function	Reference
1	2341	PB2	viral polymerase (RdRP) subunit: cap-binding, innate immunity antagonist	Field and Winter, 1982
		PB2-S1	PB2-related protein, localised in mitochondrial, interferon antagonist	Yamayosh et al., 2016
2	2341	PB1	RdRP subunit with catalytic domain	Winter and Fields, 1982
		PB1-N40	PB1-related with unknown functions	Wise et al., 2009
		PB1-F2	Apoptosis regulator	Chen et al., 2001
3	2233	PA	RdRP subunit responsible for endonucleolytic cap cleavage	Field and Winter, 1982
		PA-X	Modulates host immune responses with endonuclease activity	Jagger et al., 2012
		PA-N155	PA-related, unknown function	Muramoto et al., 2013
		PA-N128	PA-related, unknown function	
4	1778	HA	Binds Sialic acid on host cell surface glycoprotein receptors and facilitates virus binding	Winter et al., 1981a
5	1565	NP	Encapsidates vRNA and cRNA, major component of viral ribonucleoprotein complexes involved in transcription and replication	Winter and Fields, 1981
6	1413	NA	Cleaves sialic acid to facilitate virus release	Fields et al., 1981
7	1027	M1	Structural virion protein, forms scaffold matrix inside virion. Essential for virus assembly and budding. Export of vRNPs. Binds vRNP complexes and HA and NA	Winter and Fields, 1980
		M2	Act as proton ion channel that acidifies the virion and facilitates vRNP release.	
		M3	Unknown function	Lamb et al., 1981
		M4	Unknown function	Shih et al., 1998

		M42	M2-related. Unknown function	Wise et al., 2012
8	890	NS1	Interferon antagonist, induce host shut off via many mechanisms	Winter et al., 1981b
		NEP	Mediates vRNP nuclear export, regulate vRNA transcription and replication	
		NS3	NS1-related and potential role in host adaptation	Selman et al. 2012
		NEG8	Unknown function	Zhirnov et al., 2007

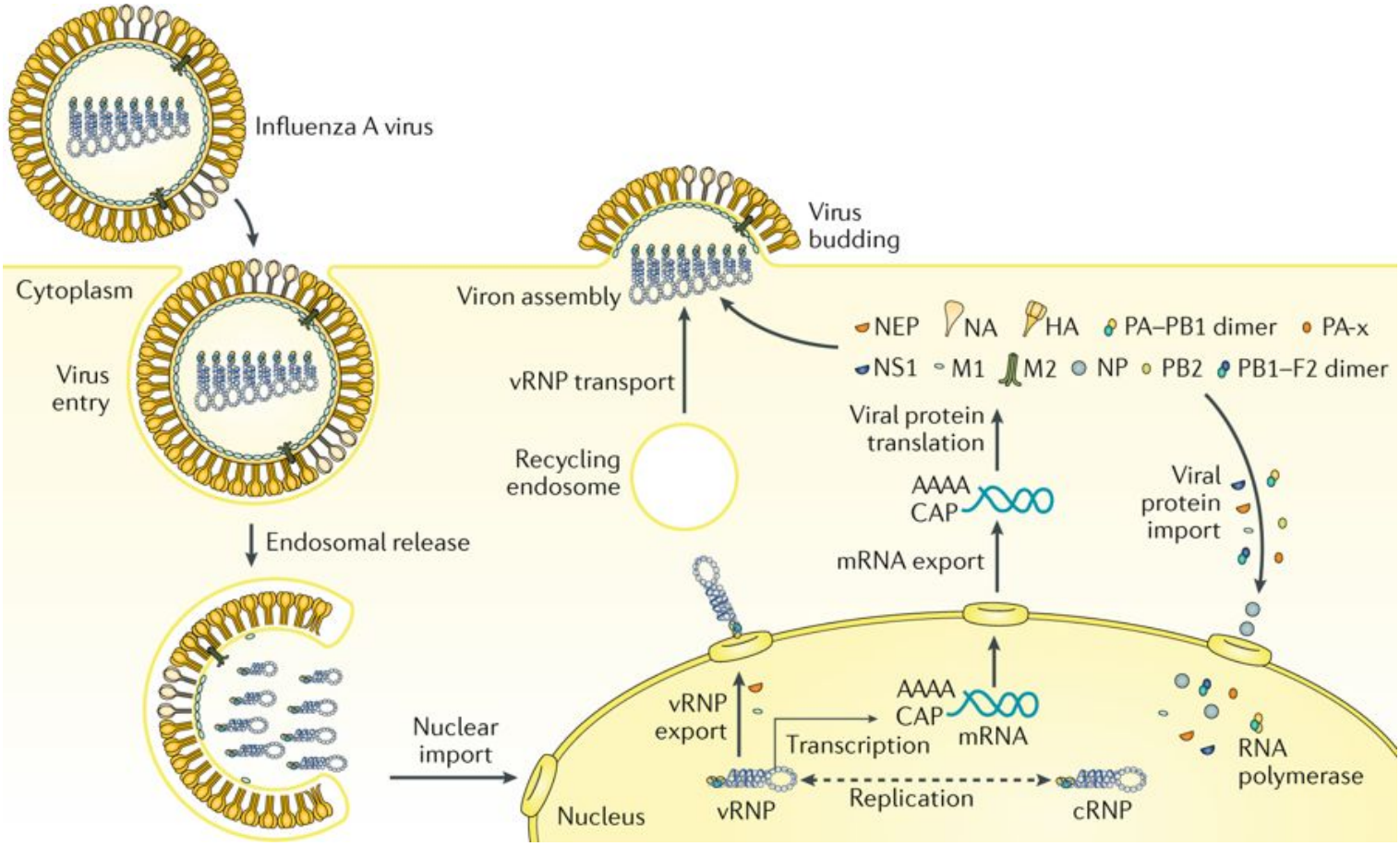


Figure 1.5 The influenza A Virus Life Cycle

Influenza virus infection begins with attachment to host cellular receipts that contain terminal α-2,3-linked or α-2,6-linked sialic acids via viral HA glycoprotein. Virus entry occurs by receptor-mediated endocytosis. Endosomal acidification results in conformation change in viral HA protein, exposing a fusion peptide that facilitates viral-endosome fusion, thereby releasing vRNP complexes into the cytoplasm. vRNP complexes are subsequently imported into the nucleus, where vRNA genome is transcribed and replicated by the viral polymerase complex (RdRP). Viral mRNAs are export to cytoplasm for translation of viral proteins. Replication of nascent negative sense vRNA is performed via complementary positive-sense cRNA. Progeny vRNP complexes are export to the cytoplasm and packaged into virions at the plasma membrane along with M1, M2, NEP and NS1. Progeny virions are released by budding. (Image adapted from Krammer et al., 2018)

1.7 Viral Life Cycle

Influenza A viruses infect and replicate in respiratory and gastrointestinal epithelial cells, as well as alveolar macrophages and dendritic cells (Hoffmann et al., 1997; Seo et al., 2004; Smed-Sörensen et al., 2012). Typically, the influenza virus life cycle takes 8-10 hours in tissue culture (Wright et al., 2013). IAVs interact and usurp the host cell during infection in order to replicate. The life cycle can be divided into the following stages; attachment and entry; fusion; uncoating and viral ribonucleoprotein (vRNP) nuclear entry; transcription and replication of viral genome; export of vRNPs from nucleus; assembly and budding at host cell plasma membrane. A schematic diagram of IAV life cycle is summarised in Figure 1.5.

1.7.1 Virus Attachment and Entry

The initial step in the IAV infection life cycle requires virus binding to host cells in order for host cell entry to happen. Glycoprotein HA is essential for this step since it binds to sialic acids present on cellular glycoproteins on the cell membrane surface.

There are 2 major linkages between sialic acids and the carbohydrates they are bound to in glycoproteins: Sialic acids linked from C2 to C3 of galactose (α-2,3) and silica acids linked from C2 to C6 of galactose (α-2,6). IAVs have evolved to utilise and recognise these linkages specific to their host species. α-2,3 linkages are preferentially bound by avian and equine IAVs, and are found in bird gastrointestinal tract and in mammalian lower respiratory tract. Whereas, human and mammalian IAVs recognises α-2,6 linkages. Interestingly, swine IAVs recognise both (Skehel and Wiley 2000). Thus, swine are considered to be good ‘mixing vessels’ for genetic reassortment as they can be infected with avian and human influenza viruses simultaneously, allowing the production of dangerous pathogenic viruses. Upon viral

attachment via sialic acid binding, the virion is internalised through receptor-mediated endocytosis and enters the host cell in an endosome.

1.7.2 Fusion, Uncoating and Nuclear import

The next step, once internalised into the endosome, involves fusion of the IAV virion with the endosomal membrane. This process is promoted by the low pH (5-6) environment provided by the endosome, which facilitates the structural rearrangement of the viral HA protein. Viral fusion requires cleavage of the viral HA protein from the HA0 precursor into HA1 (contains receptor binding domain) and HA2 (contains fusion peptide subunits) subunits by host-specific proteases either in the preceding round of viral infection (Chen et al., 1999; Boulay et al., 1987) or within the endosome (Zhirnov et al., 2002). The decrease in pH facilitates the conformational changes of the cleaved HA protein, therefore exposes the hydrophobic HA2 fusion peptide and permits fusion of viral membrane to endosomal membrane (Bullough et al., 1994). At the same time, the acidic environment allows viral M2 ion channels to open up and pump H+ ions from endoscope into the virion, acidifying the virion interior (Pinto et al., 1992). In turn, weak interactions between M1 and vRNP complexes gets disrupted and subsequently releases vRNP from the virion structure into the cytoplasm (Pintoand Lamb.,2006; Skehel and Wiley., 2000).

Following the release of vRNPs into the cytoplasm, they enter the nucleus by the importin-α /β pathway (Hutchinson and Fodor.,2013). NP contains NLS signal that is responsible for nuclear import of vRNP complexes. Unlike newly synthesised RNPs from the cell whereby NLS are masked by its coat of M1, RNPs entering a cell from a virion are not coated with M1, thus competent for nuclear import. The NLS of NP is bound by importin-α (1/5), which is proceeded by the binding of importin-β,

subsequently allowing the protein complex to dock at the nuclear pore complex and to translocate into the nucleus.

1.7.3 Viral Transcription and Replication Cycle

After vRNPs are transported to the nucleus, viral transcription of viral messenger RNA (mRNA) occurs with the help of viral RdRP in vRNPs, and viral replication of the negative-sense viral RNA (vRNA) occurs via a positive-sense complementary RNA (cRNA) as an intermediate to produce progeny vRNPs (te Velthuis and Fodor, 2016) (Figure 1.6). It is also important to note that the template for both transcription and replication of vRNA is not the vRNA by itself but rather the vRNP complex. Since NP is part of the vRNP complex, it is therefore essential for vRNA synthesis *in vivo*. In fact, these four proteins (NP, PB1, PB2 and PA) are the minimal set of viral proteins sufficient to drive transcription and replication (Huang et al., 1990). Ultimately, viral RNA synthesis plays an important role not only in ensuring progeny virus production, but also linked with viral pathogenicity (Engelhardt and Fodor, 2006). A more detailed schematic diagram (Figure 1.7) summarised the events of viral transcription and replication.

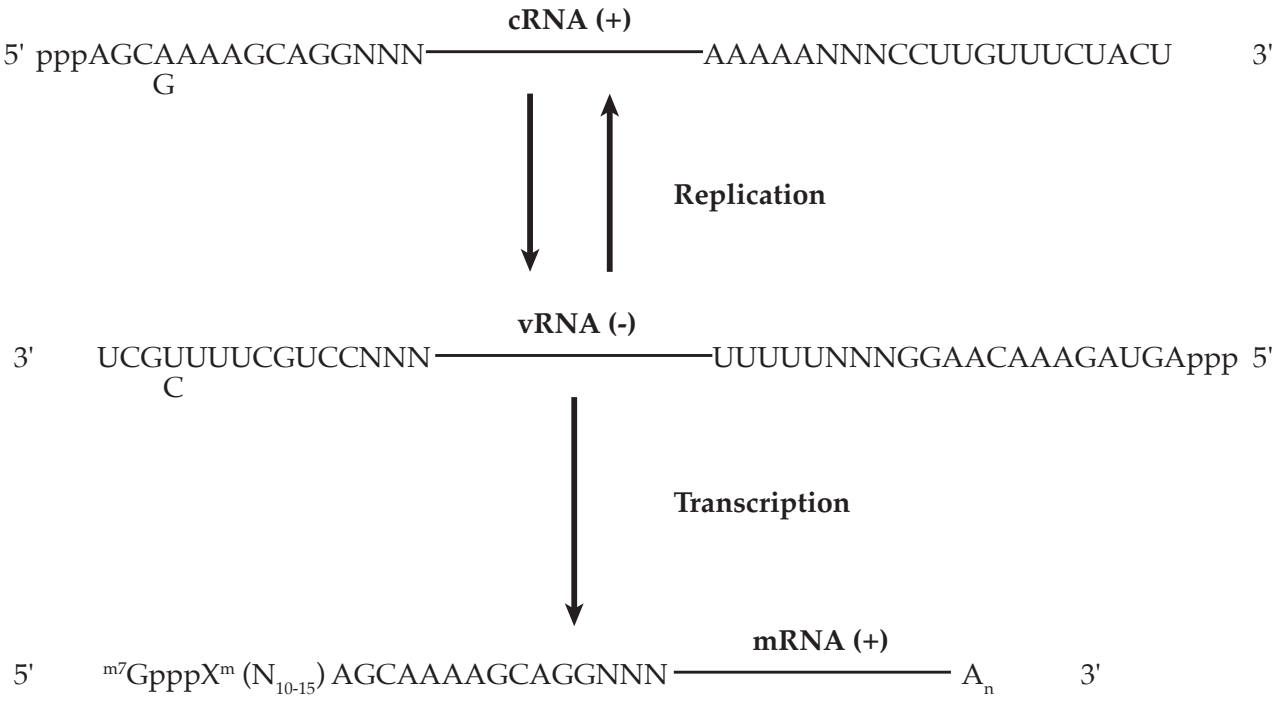


Figure 1.6 Influenza Virus RNA Synthesis

This a schematic representation of how influenza virus RNA species are synthesised by the viral RdRP. The incoming negative sense RNA viral RNA (vRNA) shown in the middle with conserved noncoding sequences at either end. The polyadenylation signal (poly-uridine stretch) is at the 5'end of the vRNA. The messenger RNA (mRNA) derives its cap and 10-13 5' nucleotides from host mRNAs and has a 3'poly(A) tail. The complementary RNA (cRNA) is an exact positive sense copy of the parental virion vRNA, which serves as a template for nascent vRNA replication. The variation of at the position 4 in the 3'vRNA non-coding region is shown. Figure adapted from Palese and Shaw, 2013).

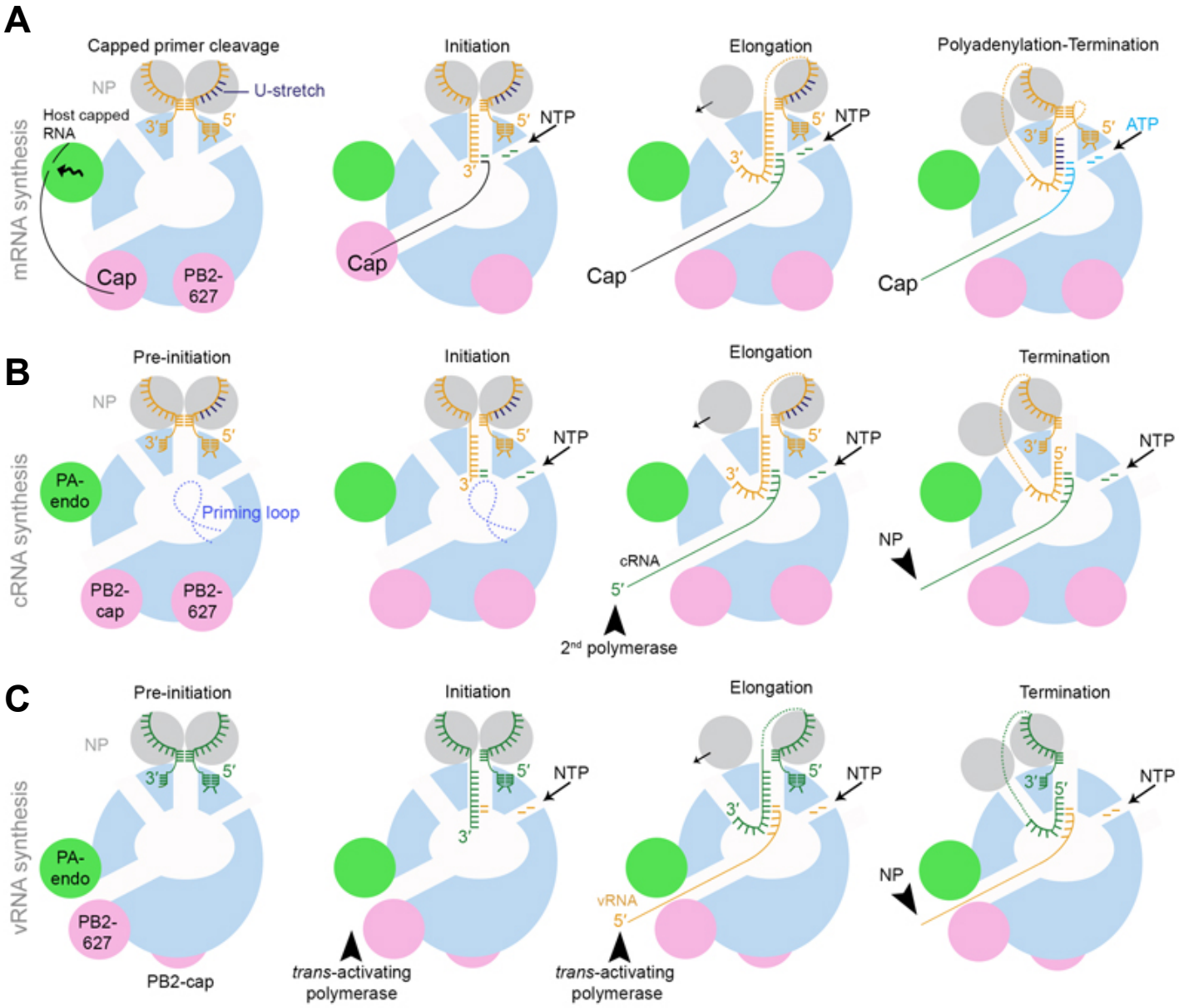


Figure 1.7 Models of Viral Transcription and Replication

These are schematics of **A.** mRNA transcription **B.** cRNA synthesis and **C.** vRNA synthesis. These figures are described in detail in Chapter 1.7.3 (Figure adapted from te Velthuis and Fodor, 2016).

1.7.3a Viral transcription

Viral mRNAs have similar features to cellular mRNAs such as the 5' cap structure and poly(A) tail and can be subjected to splicing (Engelhardt and Fodor, 2006). Thus these features are essential in order to facilitate successful transcription. Viral transcription is summarised in figure 1.7A.

i) Cap-snatching

Transcription initiation begins with a process called cap-snatching. Since influenza viruses do not encode an mRNA capping enzyme, therefore they rely on the host transcription machinery to provide a 5' 7-methylguanosine (m7G) cap structure. Cap-snatching occurs when the viral RdRP is in its pre-initiation state, where PB2 cap-binding and the PA endonuclease domains are aligned to allow PB2-binding of 5'cap (Cianci et al., 1995; Li et al., 1998) and PA cleavage 10-15 nucleotides downstream of 5'capped RNA (creating a short capped RNA primer) (Beaton and Krug et al., 1981; Plotch et al., 1981; Dias et al., 2009; Yuan et al., 2009). Subsequently, the 3'end of the capped primer is transferred from the endonuclease active site into the RdRP active site through the product exit channel, which is likely assisted by PB2 cap-binding domain rotation (Reich et al., 2014). Interestingly, host small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs) and promoter- associated capped small RNAs (generated by RNA polymerase II pausing during transcription initiation) are preferred over pre-mRNAs or mRNAs in cap-snatching (Sikora et al., 2014; Gu et al., 2015; Koppstein et al., 2015). In tandem, the 3'end of the vRNA templates needs to migrate from its binding on the RdRP surface through the template entry channel in the RdRP active site. The 3'end of vRNA is stabilised in the active site either by base pairing of the initiating nucleotide or interaction with the capped primer. Transcription begins by addition of a G (at position C2) or a C (at position G3) of the

3’ vRNA template to the 3’ of the capped primer (Beaton and Krug, 1981; Shaw and Lamb, 1984). Subsequently transcription elongation is initiated in a vRNA template-dependent manner.

ii) Polyadenylation

Polyadenylation of pre-mRNA transcripts are required for successful transcription. Unlike host transcription, influenza virus mRNA transcripts do not possess a poly(A) signal to direct cleavage and polyadenylation. In fact, polyadenylation of viral mRNA is mediated in a host-independent manner through the viral RdRP. Transcription elongation progresses in a vRNA template-dependent manner up until a 5-7 nucleotide poly(U) stretch reaches the RdRP active site, located 16 nucleotides from the 5’ terminus of the vRNA template (Robertson et al., 1981). Subsequently, the incorporation of multiple AMPs and the formation of a poly(A) tail is facilitated by the stuttering of the RdRP (Fodor et al., 1994).

iii) Splicing

Splicing of viral mRNA relies on the cellular spliceosome, but the mechanisms involved are still under consideration (Lamb and Horvath, 1991). Viral mRNA is spliced inefficiently, in contrast to cellular mRNA, where only 10% of viral transcripts are spliced (Lamb et al., 1980; 1981). NS1 has been shown to be involved along with host factor in regulating splicing of M1 mRNA (Robb and Fodor, 2012). Other factors such as viral NS1-binding protein and heterogenous nuclear ribonucleoprotein K (hnRNPK) (Tsai et al., 2013), viral RdRP, cellular splicing factor SF2/ ASF (Shih and Krug, 1996) and more host factors (Dubiois et al., 2014; Fournier et al., 2014) have been implicated in viral mRNA splicing. Originally, NS1 was thought to also regulate

its own mRNA splicing (Smith and Inglis, 1985; Garaigorta and Ortín, 2007) but the notion was recently rejected (Robb et al., 2010).

Only three viral segments produce transcripts through alternative splicing: Segment 1 (PB2), segment 7 (M) and segment 8 (NS). Segment 1 produces unspliced PB2 which encodes PB2 RdRP subunit, and a spliced PB2-S1 mRNA which encodes the auxiliary PB2-S1 protein (Yamayoshi et al., 2016). Segment 7 produces unspliced M1 that encodes the matrix protein M1. Additionally, segment 7 can produce alternatively spliced transcripts that are responsible for translating M2 ion channel proteins (Lamb and Choppin, 1981; Lamb et al., 1981) as well as M3 (unknown function) (Lamb et al., 1981) and M4 that is similar to M1 but with an internal deletion (Shih et al., 1998). Segment 8 produces the unspliced NS1 mRNA that translates into the multifunctional NS1 protein (Winter et al., 1981b; Hale et al., 2008). It can also create a spliced NS2 mRNA that encodes the nuclear export protein NEP (Lamb and Lai, 1980) and NS3 encoding mRNA with unknown function (Selman et al., 2012).

1.7.3b Viral gene expression

In order for viral gene expression to occur, viral mRNA associated ribonucleoprotein complex (mRNP) has to be assembled and exported from the nucleus to the cytoplasm. This process requires numerous cellular factors and ongoing host RNA polymerase II transcription (Tokunaga et al., 2006; Amorim et al., 2007; Read and Digard, 2010; York and Fodor, 2013), as well as viral-induced hijacking of cellular mRNA binding proteins and nuclear export machinery of mRNA (Bier et al., 2011). Viral mRNPs are thought be exported through the NXF1/p15 pathway (Hao et al., 2006; Wang et al., 2008; Read and Digard, 2010). Interestingly, multifunctional viral

NS1 is heavily involved in inhibiting viral mRNP export by forming an inhibitory complex with NXF1/ p15 (Satterly et al., 2007; Wang et al., 2008).

Once exported into the cytoplasm, viral gene expression can occur. Interestingly viral gene expression of different segments appears to be differentially regulated as different viral proteins are expressed in unequal amounts and at different times (Enami et al., 1985; Shapiro et al., 1987; Hatada et al., 1989). During early infection, viral NS1 and NP are expressed in high amounts. This is beneficial for the virus since NS1 plays an important role in inhibiting host innate immune response (Hale et al., 2008) and NP is required for efficient viral genome replication (Turrell et al., 2013). By contrast M1, NA and HA are expressed later in infection in the interest of the virus (Shapiro et al., 1987; Smith and Hay, 1982; Yángüez and Nieto, 2011) as M1 is necessary for viral genome replication inhibition (Watanabe et al., 1996; Perez and Donis, 1998) and nuclear export of vRNP complexes (Martin and Helenius, 1991a; Bui et al., 2000; Baudin et al., 2001); while HA and NA are required for virion assembly in the final stages of the viral cycle.

The switch between transcription or replication is still not well understood, but viral RdRP-mediated degradation of RNA Pol II might play a role in this switch (Vreede and Fodor, 2010). However generally during the viral life cycle, transcription of vRNA into mRNA is observed first, while replication products (cRNA and vRNA) appear after viral gene and protein expression (Barrett et al., 1979; Hay et al., 1982).

1.7.3c Viral Replication

i) cRNA synthesis

Viral RNA genome replication is a two-stage primer-independent process. Replication

begins with complementary RNA (cRNA) synthesis (Figure 1.7B). Similar to transcription initiation, translocation of the 3' terminus of vRNA into the RdRP active site is required for replication initiation. The first nucleotide to be integrated by the RdRP is GTP through base-pairing to the penultimate C2 residue in the 3' terminus of the vRNA, thus allowing the formation of a phosphodiester bond with an ATP nucleotide that base-pairs with the terminal U1 residue of the 3'vRNA and resulting in a pppApG dinucleotide (Deng et al., 2006b; Vreede et al., 2008). The formation of a dinucleotide is essential for the support of the priming loop, and this process is otherwise known as terminal initiation (Deng et al., 2006b; Vreede et al., 2008; te Velthuis et al., 2016). After initiation, elongation of nascent cRNA occurs as the vRNA template is thread through the active site, similar to mRNA transcription, with little disruption of the vRNP structure. In contrast to transcription, the 5'end of vRNA needs to be released from its binding pocket to be copied in full for replication terminal by a mechanism yet to be understood. The newly synthesised cRNA then associates with a newly synthesised viral RdRP and NP as soon as the 5'end emerges, and is ready to act as a template for vRNA synthesis.

ii) vRNA synthesis

The second stage of replication requires cRNA template for nascent vRNA synthesis (Figure 1.7C). It begins with the RdRP polymerase being in an 'inactive' state with the same conformation as RdRP when bound to the 5' end of cRNA (Thierry et al., 2016). Replication initiation occurs when the RdRP interacts with a trans-activating RdRP polymerase and subsequently induces conformational rearrangement of cRNA-bound complex (cRNP) , which allows translocation of the 3' cRNA terminus into the active site (York et al., 2013) and allows NTPs to enter via the NTP entry channel. *De novo* initiation occurs at the 4th and 5th residue of the cRNA template,

otherwise known as internal initiation, without the participation of the priming loop (Deng et al., 2006; Reich et al., 2014; te Velthuis et al., 2016) . The resulting pppApG dinucleotide relocates to the 3'end of the cRNA template and is elongated by the polymerase. Elongation and termination proceeds just like in cRNA synthesis. Subsequently the 5'end of the vRNA product is bound by the trans-activating polymerase, whilst the rest of the vRNA associates with NP to assemble into vRNPs (reviewed in te Velthuis et al., 2016).

1.7.4 Nuclear Export and Trafficking of vRNP Complexes

After viral genome replication and vRNP assembly in the nucleus, vRNP complexes must be exported through the nuclear membrane into the cytoplasm and to the cell membrane for packaging into progeny virions. The proteins responsible for this are viral M1 and NEP (Paterson and Fodor, 2012). M1 has been shown to associate with vRNP, which is involved in both vRNP assembly and essential for nuclear export of vRNP (Bui et al., 2000; Huang et al., 2001; Martin and Helenius, 1991b). However M1 alone is not sufficient for nuclear export of vRNP. NEP has been proposed to act as an adaptor between vRNA-M1 complexes and the importin-β protein CRM1 to facilitate nuclear vRNP export (O'Neill et al., 1998; Neumann et al., 2000). The interactions, through the C terminal domain of M1 and NP on vRNPS, as well as the M1 N-terminal binding of NEP C-terminus, is conjectured to block the nuclear localisation signal (NLS) on M1 N-terminal domain. Therefore, this prevents the re-import of M1-vRNP complexes after nuclear export. Nuclear export of vRNP-M1-NEP complex is mediated by CRM1, which when bound by Ran-GTP, recognises the nuclear export signal (NES) on the N-terminus of NEP and facilitates trafficking via the nuclear pore complex (NPC) (Brunotte et al., 2014). However, nuclear export is not solely achieved through this mechanism. NP and M1, or M1 alone, have also

been proposed to mediate this CRM1-dependent nuclear export (Bui et al., 2000; Elton et al., 2001; Yu et al., 2012).

After nuclear export, vRNPs accumulate at the perinuclear microtubule organising centre (MTOC), where they interact with RAB11 through PB2. RAB11 then directs vRNP to the cell membrane via vesicular anterograde microtubule transport (Bruce et al., 2010; Amorim et al., 2011; Momose et al., 2011).

1.7.5 Genome Packaging

Influenza A viruses need to ensure that the full genome of exactly eight vRNPs are packaged into the progeny virions in order for successful viral particles production. Genome packaging was thought to be random, which has been evidenced by the fact that most influenza virus particles are non-infectious because they are missing one or more genome segments (Hutchinson et al., 2010; Brooke et al., 2013). However, recently, it has been proposed that packaging is strictly controlled by RNA-RNA interactions to incorporate a single copy of each genome segment into progeny viral particles (Gerber et al., 2014).

1.7.6 Virus assembly, Budding and Release

To complete the viral replication cycle, progeny virions must be assembled and exit the host cell. Egress of virions is initiated by budding from the lipid raft domains on the apical surface of polarised cell cell membrane (Rodriguez Boulan and Sabatini, 1978; Simons and Ikonen, 1997; Scheiffele et al., 1999). This initiation is caused by the accumulation of HA and NA clusters at the lipid raft domains of the cell membrane, and cause membrane deformation (Chen et al., 2007a). M1 is then directed to the budding virion by interacting with the cytoplasmic tails of HA and NA

and subsequently forms the interior matrix of the virion (Ali et al., 2000; Zhang et al., 2000; Noton et al., 2007). The M1 matrix then act as a scaffold for the incorporation of the eight vRNP complexes (Noda et al., 2006). Finally, M2 incorporation, which is mediated by M1 (Chen et al., 2008), has been proposed to cause cleavage of the membrane and facilitate release of nascent virion into the extracellular space (Rossman et al., 2010; Rossman and Lamb, 2011). The exit of the virion is then facilitated by NA glycoprotein on the surface of nascent virion by hydrolysing sialic acids on the cell surface to prevent reattachment and reinfection of the same cell (Air and Laver, 1989; Huang et al., 2008).

1.8 Influenza Host Shut Off Strategies

1.8.1 NS1-mediated Host Shut Off

NS1 is a critical multifunctional protein that allows the virus to antagonise host immune responses (typically Interferon) and manipulate cellular homeostatic mechanisms (Hale et al., 2008; Krug, 2015). Its structure is formed by a RNA-binding N-terminal domain and an effector C-terminal domain (Hale et al., 2008). Type 1 IFN-mediated response inhibition by NS1 is mediated by several different molecular mechanisms, which are not shared by all influenza A variants of NS1 (Salvatore et al., 2002; Kochs the al., 2007; Hale et al., 2010; Kuo et al., 2010; Kainov et al., 2011; Ramos et al., 2013). One of the mechanisms to regulate type I IFN and host shut off is accomplished through inhibition of host mRNA 3'end processing (Nemeroff et al., 1998), although this mechanism is not conserved in all variants. This inhibition is mediated by the interactions of cleavage and polyadenylation factor CPSF30 and the effector domain of NS1 (Nemeroff et al., 1998). Therefore, since CPSF30 is responsible for recognising poly(A) signals at the 3' end of mRNA (Sun et al., 2018), abrogating CPSF30 may result in blockage of cleavage and polyadenylation of RNA

and ultimately preventing protein expression. Interestingly, NS1 shutoff has no effect on viral mRNAs since their poly(A) tails are generated by RdRP stuttering rather than by cleavage and polyadenylation of the RNA (Poon et al., 1999). Furthermore, Bauer et al. has demonstrated NS1 can inhibit transcription termination, but termination defects were also seen in cells infected with NS1-deficient Udorn or WSN (Bauer et al., 2018). More recently, SUMOlylation of an intrinsically disordered region (IDR) of NS1 expressed by the 1918 pandemic IAV strain can cause global termination defects and host shut off (Zhao et al., 2018). Interestingly, this small ubiquitin-like modifier (SUMO) site is only unique to 1918 NS1 C-terminal domain and not present in other variants (Zhao et al., 2018). Therefore host shut off effects induced by SUMOlyation may be limited to 1918 H1N1 strain and not conserved in other strains such as WSN/33.

Previous structural, sequence and functional analysis studies have identified conserved binding sites for CPSF30 on NS1. Table 1.4, adapted from Levene and Gaglia, 2018, summarises how specific residues on NS1 can affect host shut-off activity of NS1. Strong/weak activity for NS1 variants refers to naturally occurring variants with either strong or weak CPSF30 binding. na refers to loss of activity when residue is experimentally mutated.

Table 1.4 Specific Residues of NS1 affecting CPSF30 binding

NS1 residue (Amino Acid)	Strong Activity	Weak Activity	Function	Reference
103	F	L	F103/M106 is required for CPSF30 binding in many strains (not sufficient in canine H3N8, pH1N1)	Twu et al., 2007; Das et al., 2008
106	M	I,V		
144	L	na	Required for CPSF30 binding (Udorn,WSN)	Twu et al., 2006

184-188	GLEWN	na	Required for CPSF30 binding in vitro and in cells (Udorn) (WSN)	Noah et al., 2003; Das et al., 2008; Steidl et al., 2010
223-237	ARTARSKVRRDK MAD	na	Required for PABPN binding (Udorn)	Li et al., 2001
55	K	E	When residue changes to 'strong', it restores strong CPSF30 binding in pH1N1 strains (with F103, M106); D189N alters host shutoff activity in H3N2 and 1918 H1N1	Hale et al., 2010; Clark et al., 2017; Nogales et al., 2017;
90	I	L		
108	K	R		
123	V	I		
125	D	E,G		
131	E	K		
189	D	G,N		
205	S	N		
64	I	T	Changes to 'weak' activity residues reduce CPSF30 binding and IFN antagonisms in H3N2 strains; V194I also reduces CPSF30 binding and host shutoff in 1918 H1N1 NS1	Dediego et al., 2016; Nagoles et al., 2017
194	V	I		
98	L	S	Mutations abolish CPSF30 binding during H3N2 mouse adaptation	Forbes et al., 2012, 2013
180	V	A		
96	E	na	Mutation causes temperature sensitive mRNA cleavage (Udorn)	Shimizu et al., 1999
191-195	EALQR	deleted	Deletion in H5N1 NS1 reduces CPSF30 binding and IFN antagonism	Zhu et al., 2008

1.8.2 Pol II Degradation by Viral RdRP

As mentioned in 1.7.3, viral transcription is dependent on the host RNA polymerase II to supply 5' capped transcripts, which are bound and cleaved by the cap-binding and endonuclease domains of viral RdRP. Therefore, RdRP has to maintain a coordinated association with host pol II. The C- terminal domain (CTD) of the large subunit of host pol II harbours multiple repeats of amino acid heptad (Tyr1-Ser2-

Pro3-Thr4-Ser5-Pro6-Ser7), which are post-translationally modified to regulate transcription and RNA processing, and act as a platform for various enzymes (Zaborowaska et al., 2016a). Viral RdRP specifically interacts with phosphorylated serine 5 form of Pol II (Engelhardt et al., 2005; Martínez-Alonzo et al., 2016; Lukarska et al., 2017), which is compatible with capping of cellular mRNAs, however does not interact with the elongation form of Pol II (Ser2-phosphorylated). Although this interaction cannot interfere with transcription initiation of host transcripts, it is still beneficial for viral cap-snatching.

Apart from cap-snatching, this interaction also leads to ubiquitination and proteasomal degradation of the Pol II large subunit (Rodriguez et al., 2007; Vreede et al., 2010). Furthermore, the reduction in pol II loading on the host genome was recently explored in more detail by Bauer et al. (Bauer et al., 2018), whereby low levels of pol II occupancy was observed after the 3'transcription termination and polyadenylation sites indicating that transcription termination is impaired (Bauer et al., 2018). Subsequently, this results in inhibition of IFN-stimulate genes and thus host antiviral response (Vreede et al., 2010).

1.8.3 Host Transcripts Degradation by PA-X

PA-X is responsible for host RNA degradation and is produced by a ribosomal frameshift when translating the PA mRNA transcript (Firth et al., 2012; Jagger et al., 2012). Structurally, PA-X has a similar N-terminal endonuclease to PA, but terminates in a short 41 or 61 amino acid long C terminal domain (Jagger et al., 2012; Oishi et al., 2015; Khaperskyy et al., 2016). Similar to PA, PA-X preferentially degrades Pol II transcripts, as opposed to polymerase I or III transcripts, without effecting viral mRNA. Furthermore, PA-X has the ability cleave both ssRNA and dsRNA, which fits

its role in host shut off and cap snatching (Bavagnoli et al., 2015). PA-X's ability to degrade host mRNAs by cleaving RNAs internally and subsequently leads to cellular Xrn1 exonuclease degradation (Khaperskyy et al., 2016). Additionally, PA-X also mediates nuclear accumulation of cytoplasmic poly (A) binding protein (PABPC1), resulting in nuclear retention and hyperadenylation of host mRNAs (Khaperskyy et al., 2014) and reduce the amount of translatable host mRNAs. Altogether, PA-X has the ability to reduce gene expression and facilitating host immune response inhibition.

1.9 Host Immunity against Influenza A Virus Infection

Host immune systems have evolved multiple ways to defend against and exterminate foreign viral infection. Immune defence is divided into 2 major systems, innate and adaptive immunity.

While adaptive immunity plays a crucial role in the clearance of viral pathogens at later infection stages, innate immunity provides the first line of defence against IAV infection. Firstly, recognition of the virus is required to initiate innate immune signalling, whereby host pathogen recognition receptors (PRRs) such as retinoic acid inducible gene I (RIG I) and toll-like receptors (TLRs) identify pathogen associated molecules patterns (PAMPs) (Cao, 2016). Additionally, melanoma differentiation associated gene 5 can also detect cytoplasmic viral transcription products (Pichlmair et al., 2006). RIG-I is critical for detecting single stranded RNA IAV, thus following recognition, RIG-I is active and its cascade activation and recruitment domains (CARDs) are exposed. CARD is then modulated by dephosphorylation or ubiquitination by E3 ligases such as TRIM25 (Carthagena et al., 2009; Nguyen et al., 2016). CARD-dependent association of RIG-I and MAVS subsequently trigger

downstream transduction signalling at the outer mitochondrial membrane (Katze et al., 2008; Andrejeva et al., 2004; Pichlmair et al ., 2006; Reikine et al., 2014), which leads to activation of transcription factors interferon regulatory factor 3 (IRF3), 7 (IRF7) and nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kB), and consequently expressing a variety of interferons (IFN) (type 1 and 3) and pro inflammatory cytokines (such as TNF, IL6 and IL1β) (Akira et al., 2006). Additionally, TLR 3, 7 and 8 are responsible for sensing IAV in cytoplasmic endosomes during virus replication (Kawai and Akira, 2010). IFN and cytokines stimulation is similar to

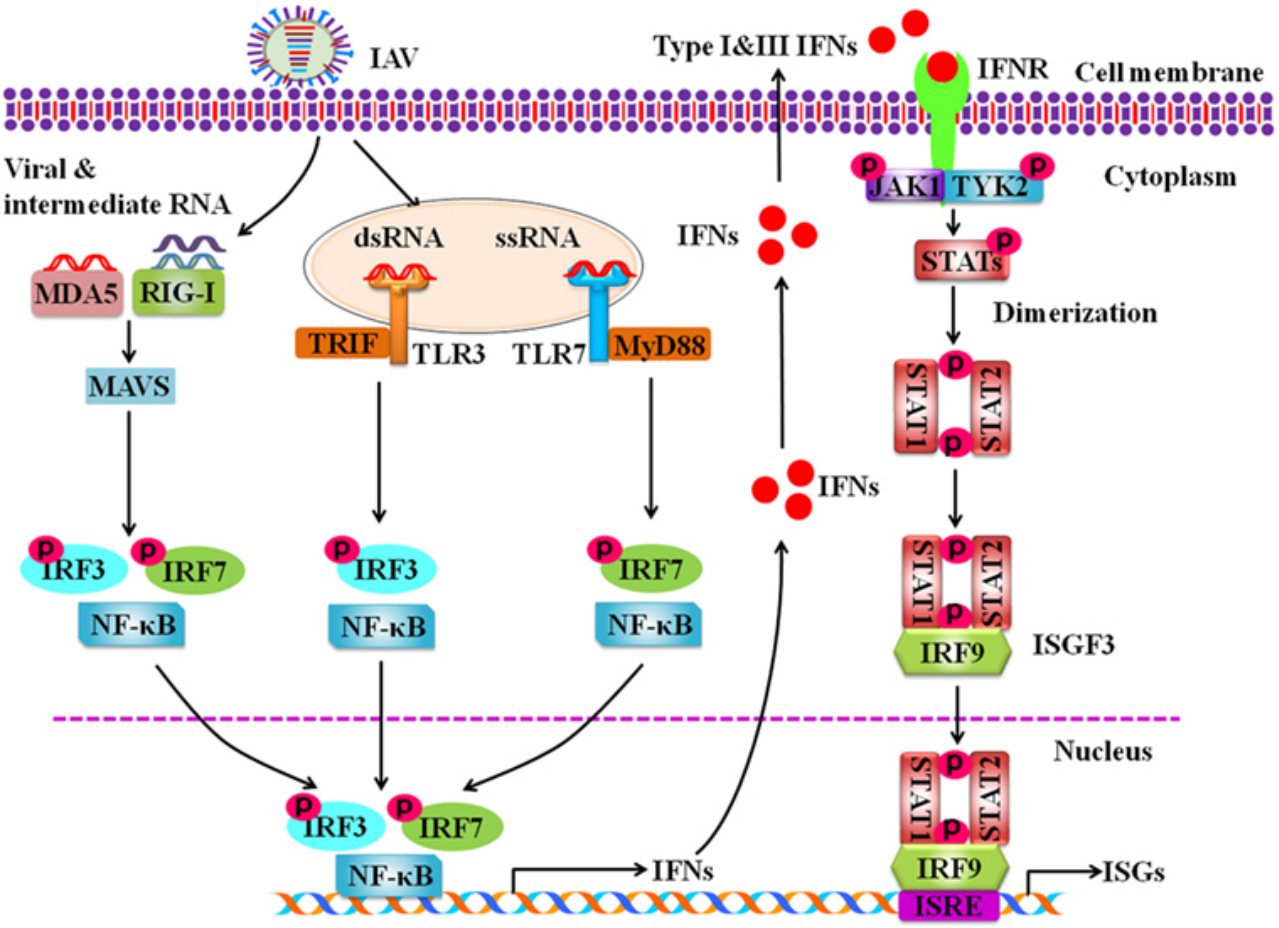


Figure 1.8 Innate Immune Response against influenza A Virus Infection (adapted from Chen et al., 2018)
Intracellular detection of IAV infection by PRRs activates transcription factors IRF3,7 and NF-kB. The PRRs displayed are retinoic acid- inducible gene- I protein (RIG-I), melanoma differentiation-associated gene 5, and toll-like receptors (TLR). These activated transcription factors trigger the expression of type 1 and 3 interferons (IFNs). IFNs secreted by infected cells interact with their specific receipts, thus activating JAK/STAT signalling pathway that triggers expression of IFN-stimulated genes.

RIG-I pathway, however specific TLR such as TLR 3 activates specific IRFs like IRF3 whereas TLR 7 activates IRF7 (Reviewed in Chen et al., 2018). After transcription of genes encoding IFNs, type 1 IFN interacts with IFN-α/β receptor, while type 3 IFN interacts with IFNL receptors in an autocrine or paracrine manner, which then activates the Janus kinase signal transducer and activator of transcription (JAK-STAT) pathway. Subsequently, it leads to STAT 1 and 2 phosphorylation and binding with IRF9 to form the ISG factor 3 (ISGF3) complex. Finally, once ISGF3 is translocated to the nucleus and binds to IFN-stimulated response element, transcription of numerous IFN-stimulated genes (ISGs) (MX, interferon induced transmembrane protein family (IFITMs), cholesterol 25-hydroxylase (CH25H) and TRIM proteins) occurs (Schneider et al., 2014). A simplified schematic diagram of innate immunity pathways is displayed in figure 1.8.

Adaptive immunity also plays an important role in IAV infection defence. T cells, known as CD4+ and CD8+ T cells, and B cells are key for adaptive immune response. After infection, CD8+ T cells are activated by dendritic cells (DCs) (migrated from lungs to T-cell zone of draining lymph nodes) leading to T-cell proliferation and differentiation into cytotoxic T lymphocytes thus restricting viral replication and killing viral-infected cells by producing cytokines and cytotoxic granules containing perforin and granzymes (Ho et al., 2011; Chen et al., 2018). CD8+ T- cell differentiation were also shown to be enhanced by type 1 and 3 IFN, IL2 and IL12 (Whitmire et al., 2005; Pipkin et al., 2010). CD4+ T cells are equally important for adaptive defence against IAV viruses. Similar to CD8+ T cells, CD4+ T cells are activated by DCs migrated from lung to draining lymph nodes during infection (Lukens et al., 2009). Subsequently, they can differentiate into Th1 cells, express antiviral cytokines such as type 3 IFN, TNF and IL2, activate alveolar

macrophages and activate B cells through CD40 ligand (Szabo et al., 2000; Liu et al., 2012; Swain et al., 2012). Additionally, CD4+ T cells can differentiate into Th2, Th17, regulatory T cells (Treg), follicular helper T cells, and occasionally killer cells (Zhu et al., 2010). Even though CD4+ T cells can direct CD8+ T cell responses by secreting various cytokines, the precise role of CD4+ T cells to facilitate and regulate CD8+ T cell repression of IAV infection remains elusive since CD8+ T cell can be activated independently of CD4+ T cells in mice (Gruta and Turner, 2014).

1.10 Aim of Study

The aim of this project was to investigate the changes of APA profile regulation during influenza A/WSN/33 infection in A549 lung epithelial cells and to decipher the mechanism involved in causing such changes using a subcellular fractionation approach. This project aimed to combine the use of deep sequencing, bioinformatic analysis and classical biochemistry techniques to give a comprehensive outlook of transcriptome changes during infection. Many global scale studies investigating APA fail to explore the physiological impact of these APA changes at an individual gene level, instead they mostly focus on general trends identified by bioinformatics. Therefore, this thesis aimed to identify physiologically impactful APA events during infection and understand how this may be consequential to host shut-off.

Chapter 2: Materials and Methods

2.1 Cell Culture

Human embryonic kidney (HEK-293T) and adenocarcinomic human alveolar basal epithelial (A549) cells were maintained in DMEM (10% FCS). MDBK cells were maintained in MEM (10% FCS) and 2mM L-glutamine. Reaching 60-80% confluency, cells were passaged by detaching the cell monolayer from the cell culture flask with TrypLE Express trypsin (Gibco) and diluted appropriately in cell culture media.

2.2 Influenza Virus

Influenza A/WSN/33 (H1N1) virus (WSN) was generously provided by the Fodor Lab.

2.2.1 Influenza A/WSN/33 Harvest

Wild-type Influenza A/WSN/33 virus stocks (gift from Ervin Fodor, Dunn School of Pathology, University of Oxford) were amplified by applying virus-containing medium to 80% confluent Madin-Darby Bovine Kidney Epithelial (MDBK) cells in 175cm^2 cell culture flasks for 1 hour at room temperature in MEM (0.5% FCS) at low MOI (0.001-0.01), and subsequent incubation at 37°C and 5% CO_2 for 72 hours. Media containing virus was harvested and cleared from cellular debris by centrifugation at 4000x g for 10 minutes at 4°C. Supernatants stored at -80°C.

2.2.2 Plaque Assay

Influenza A viruses were serially diluted tenfold in a total volume of 250µl MEM (0.5% FCS) and 2mM L-glutamine. Monolayers of MDBK cells on 6-well plates were washed with PBS and infected with 200µL of serially diluted virus for 1 hour at room temperature. After

infection, 2ml of agarose solution (2% agarose dissolved in PBS diluted 1:1 in MEM (0.5% FCS) and 2mM L-glutamine) was overlaid to polymerise at room temperature. The plates were incubated at 37°C and 5% CO_2 for 72 hours, proceeded by removal of agarose layer and staining of plaques using 2mL of 0.2% Coomassie Brilliant Blue dye (7.5% glacial acetic acid and 50% methanol).

2.2.3 Gradient Ultracentrifugation Purification

Influenza A/WSN/33 virus was purified as described (Hutchinson et al., 2015). Harvested virus was centrifuged at 4000x g for 10 minutes at 4°C to separate away from cellular debris. This was followed by transfer to thick-walled ultracentrifuge tubes and ultracentrifuged at 17,000x g for 15 minutes at 4°C in a SW32 Ti Rotor and Optima XPN-80 Ultracentrifuge (Beckman Coulter). Supernatants were layered meticulously onto 30% sucrose cushion in virus suspension buffer (VSB) (5mM $CaCl_2$, 100mM NaCl and 20mM Tris-HCl [pH 7.4]) and pelleted by ultracentrifugation at 112,000x g for 90 minutes at 4°C. Supernatants are carefully aspirated and viral pellets were resuspended in 1ml VSB and incubated overnight at 4°C. Resuspended purified virus was harvested and stored at -80°C. Plaque assays were done to determine viral titres.

2.2.4 SDS-Page and Silver Stain

SDS-PAGE gels were transferred to clean plastic dish and fixed in 50 % methanol and 10% acetic acid for 10 minutes. The fixed gel was sensitised in 50% methanol and 2.5% sensitizer, followed by three washes in dH_2O. Gels were stained and developed per manufacturer's instructions. Gels were silver stained using SilverXpress kit (Invitrogen) according to the manufacturer's instructions.

2.2.5 Immunofluorescence Microscopy

A549 cells were grown on sterilised 13mm glass coverslips in 24-well plates. Cells were washed with PBS prior to infection with influenza A/ WSN /33 at an MOI of 5. At mock, 3, 6, 9, 12 and 18 hours post-infection cells were washed with PBS twice and fixed at room temperature for 15 minutes with 4% paraformaldehyde in 250 mM HEPES (pH 7.9) prior to three washes with PBS. Cells were then permeabilised at room temperature for 15 minutes in 0.25% Triton X-100 in PBS and washed three times with PBS. Cells were blocked in blocking PBS (0.5% Triton X-100, 3% bovine serum albumin, 10% normal goat serum) overnight at 4°C. Cells stained using rabbit polyclonal anti-NP antibody at a 1:500 dilution in blocking PBS for 1 hour at room temperature. Cells were washed three times with blocking PBS. Subsequently, cells were stained using a goat anti-rabbit IgG antibody conjugated to an Alexa532 dye at a dilution of 1:500 in blocking PBS for 1 hour at room temperature. Cells were washed twice with blocking PBS and washed twice with PBS prior to mounting in Mowiol (Calbiochem) containing 1 µg/ mL 4’,6-diamidino-2phenylindole (DAPI; Sigma-Aldrich). Images were obtained on a EVOS Flood Cell Imaging Station (Life Technologies). ImageJ software was used for image analysis.

2.3 Infection Time Course Experiment

A confluent monolayer of human A549 cells 15cm dishes were infected with gradient ultracentrifugation purified influenza A/WSN/33 (H1N1) virus at an MOI of 5 in DMEM (0.5% FCS) for 1 h at room temperature. Infected cells were incubated with time-points 3h, 6h , 9h, 12 h and 18h at 37°C.

2.4 Subcellular RNA Fractionation

A549 cells grown in 2 x 15 cm dishes were detached by scrapping and washed twice with ice-cold PBS. Centrifugation at 1000 x g for 5 minutes to pellet the cells. The cell pellet

was re-suspended in 1 mL of Lysis Buffer (10mM Tris pH8, 140mM NaCl, 1.5mM $MgCl_2$, 0.5% NP-40). 150ul of the lysate was extracted for protein expression analysis. The remaining lysate was centrifuged at 1,000 x g for 3 minutes at 4°C and the supernatant (500µl) was removed as the cytoplasmic fraction. Cytoplasmic fraction was further purified by centrifuging at 12,000 x g for 1 minute at 4°C. The nuclear pellet was re-suspended in 1ml Lysis Buffer and 100 µl Detergent Stock (3.3% w.v. sodium deoxycholate, 6.6% v/v Tween40) was added in a drop-by-drop manner under slow vortexing. Having left the sample in ice for 5 minutes, it was centrifuged at 1000 x g for 3 minutes at 4°C. The supernatant was discarded and the pellet was washed in 1mL Lysis Buffer. The remaining pellet was re-suspended in 1mL Trizol® (Thermo Fisher) and fully dissolved. 500µl Trizol® was added to cytoplasmic lysate. Fractions were left at room temperature for 5 minutes. Chloroform (20% of volume) was added and the mixture was subsequently vigorously vortexed before being centrifuged at 12,000g for 15 minutes at 4°C. The upper aqueous phase containing RNA was allowed to precipitate in isopropanol overnight at -20°C .

The RNA pellets were DNase 1-treated (Roche) as per manufacturers instructions. RNA was subsequently purified through phenol-chloroform extraction and was precipitated by adding 2.5x volume of ethanol and 10% sodium acetate after incubation for >1hour at -20°C. RNA was finally collected by centrifuging at 13,000 rpm for 20-30minutes at 4°C and the pellet was resuspended in RNase-free water. RNA concentrations were quantified using a NanoDrop ND1000.

2.5 Reverse Transcription PCR

1µg RNA was reverse transcribed into cDNA using a oligod(T) primer with RevertAid Reverse Transcriptase (Thermo Fisher)/ SuperScript III Reverse Transcriptase (Thermo Fisher) as per the manufacturer's instructions. PCR was performed with 1µl of cDNA

template with Taq Polymerase (NEB) as per manufacturer's instructions. The annealing temperature was altered depending on the melting temperature of the 2 primers used in the reaction (Table 2.1). Primers were designed with online tool Primer Blast (www.ncbi.nlm.nih.gov/tools/primer-blast) to amplify targeted region. MWG Eurofins was used to synthesise primers.

2.6 Denaturing Agarose RNA Gel Electrophoresis

A denaturing 1% formaldehyde/agarose gel (1x formaldehyde gel runner buffer, 3.3% formaldehyde, 1% agarose) was used to assess the quality of RNA and fractionation. 2µg RNA were mixed with formamide, formaldehyde gel running buffer and formaldehyde (Neve et al., 2016). RNA samples mixtures were denatured at 65°C for 15 minutes, and subsequently mixed with formaldehyde gel loading buffer and ethidium bromide. The gel was run for 2 hours at 65V in 1x formaldehyde gel running buffer. RNA gels were visualised under $UV_{320\ nm}$ light and pictured on Kodak Gel Camera System.

2.7 Protein Isolation

For whole cell isolation, cell pellets were re-suspended in 100/150 µl lysis buffer. Protein lysates were incubated with 1 µl Benzonase® (Merck) for 30 minutes at 37°C to degrade all forms of nucleic acids. Protein Concentration were obtained through Bradford Test and Qubit® Protein Assay Kit as per manufacturer's instructions.

2.8 SDS-PAGE and Western Blotting

Protein Samples were mixed with 4x Laemmli Buffer and boiled for 10 minutes at 90°C, before loading onto an SDS-PAGE gel (5% acrylamide stacking gel at pH 6.8; 8%,10% or 12% acrylamide resolving gel at pH 8.8) in parallel to a protein ladder. The gel was run in 1x SDS running buffer for 20 minutes at 65V and 1-2 hours at 120V.

Proteins were transferred from SDS-PAGE gels onto polyvinylidene difluoride membranes (PDVF) pre-soaked in 100% methanol and equilibrated in 1x transfer buffer (25mM Tris-HCl, 192mM glycine, 20% methanol), via wet-transfer in 1x transfer buffer and at 4°C for 1-2 hours (constant current of 250mM).

Membranes were pre-blocked in 3% or 5% non-fat milk powder in TBS-T (50mM Tris, 150mM NaCl, 0.05% Tween 20) under agitation for 1 hour at room temperature. Primary antibody incubations were done overnight at 4°C. Following 3x 5 minute washes in TBS-T, Horseradish peroxidase (HRP)- conjugated secondary antibody incubations were done for 1 hour at room temperature. Antibodies and corresponding dilutions used for each protein target are stated in table 2.2. Proteins were detected using an enhanced chemo luminescence ECL kit (LifeTech) and exposed on X-ray film (Armesham).

2.9 Primer Extension Analysis

Primer extensions were performed by Ben Nilsson (Dunn School of Pathology, University of Oxford).

2.9.1 Primers

Influenza A virus segment 6 derived RNAs and cellular 5S rRNA were detected using specific oligodeoxyribonucleotides as present in Table 2.1. Primer NA+ was used to detect mRNA and cRNA and primer NA- was used to detect vRNA derived from viral segment 6.

2.9.2 Radiolabelling of Oligodeoxyribonucleotides

Oligodeoxyribonucleotides were labelled in a 10µl reaction containing 10 pmol of relevant primer, 10 µCi [γ-32P] ATP and 2.5 U of PNK in 1x kinase buffer A. The reaction was performed for 1h at 37°C. Excess unlabelled [γ-32P] ATP and reaction constituents were

removed using the QIAquick Nucleotide Removal Kit (QIAGEN) according to the manufacturer's specifications. Radiolabelled oligodeoxyribonucleotides were eluted in 30 µl ddH_2O and stored at -20°C.

2.9.3 Primer Extension

Primer extension assays were performed as previously described (Robb et al., 2009). Briefly, in a 5µl reaction, RNA was mixed with an excess of radiolabelled primers complementary to positive and negative sense viral segment 6 RNA and complementary to cellular 5S rRNA which is used as an internal loading control. The mixture was denatured by heating to 95°C for 3 minutes and immediately cooled on ice. 5µl of transcription mix (2x First Strand Buffer [Invitrogen], 20mM dithiothreitol [DTT], 1mM dNTP mix [Thermo Scientific} and 50U Superscript III RT [Invitrogen]) was added and the reaction was heated to 50°C for 1 hour, before heat inactivating the reaction at 70°C for 10minutes. Reactions were terminated by addition of 10µl of 80% formamide, 1 mM EDTA, bromophenol blue and xylene cyan and incubation at 95°C for 3 minutes. Reaction products were resolved by 6% (for full-length NA segment template) or 12% (for 47-nucleotide NA segment template) denaturing PAGE containing 7M urea in TBE buffer and visualised by autoradiography.

2.10 QuantSeq 3' mRNA Sequencing

2.10.1 QuantSeq 3' mRNA-Seq Library Preparation

Libraries for RNA sequencing were prepared with the Lexogen QuantSeq 3' mRNA-Seq kit as per manufacturer's instructions. Input for nuclear fractions was 500ng and 3-5ug for cytoplasmic fractions. 13 PCR cycles were used for amplification.

2.10.2 hsDNA Bioanalyzer Assay

The quality of the libraries generated from Quant-Seq kit were assessed using a high sensitivity DNA (hsDNA) chip on the Agilent 2100 Bioanalyzer as per manufacturer's instructions. The library concentrations were quantified by selecting the region between 50bp and 1000bp from the Bioanalyzer output profile.

2.10.3 RNA-seq on Ion Proton Platform

3'QuantSeq Libraries were diluted to a final concentration of 350pM. The Ion PI™ Hi-Q™ Chef Kit was used with the ion Chef™ System for sequencing as per the manufacturer's instructions. FASTQ file were aligned to the human genome (hg19) using STAR (Spliced Transcripts Alignment to a Reference). Resulting BAM files for further processing were required to have a minimum of 1 million aligned reads.

2.10.4 Bioinformatics Pipeline to Assess APA and Gene Expression

All RNA-seq experiments were done with true biological replicates of samples processed several weeks apart and with different cell batches. The filtered reads from each replicate were combined and processed through an established bioinformatics pipeline to extract statistically significant APA events and differential gene expression analysis between respective samples (Hoque et al., 2013; Neve et al., 2016). Significant APA changes were defined as $P<0.01$ Fisher's exact Test, with a minimum of 5% change in APA isoform abundance. Gene expression/ mRNA abundance blues were calculated from the combined reads of poly(A) sites mapping to a single gene, and normalised to reads per million (RPM). Significant Gene expression changes were interpreted by $p<0.01$ Fisher's exact Test and a fold change greater than 2.

2.10.5 Bioinformatics Pipeline to Assess Infection Dynamics

Raw unmapped read sequences from uBAM files from each sample were uploaded onto Galaxy (https://useglaxy.org/). uBAM files were converted to fastq files via Galaxy Picard SamtoFastq and aligned to A/WSN/33 (H1N1) nucleotide sequences (Appendix A) via Galaxy BWA-MEM mapping tool to generate BAM files. Subsequently, SAMtools Flagstat was performed to tabulate total number of reads aligned to the WSN Genome.

To assess infection dynamics in each sample, the % of viral reads was calculated by dividing the number of viral aligned reads over the total reads (viral & human total reads)*100.

Evaluation of reads aligned to the 8 individual viral genome segment was done with the help of Harry Fischl (Department of Biochemistry, University of Oxford). The script used was developed on R studios.

2.10.6 Enrichr Gene Ontology Analysis

Gene ontology analysis was done with Enrichr (http://amp.pharm.mssm.edu/Enrichr/) (Chen et al., 2013; Kuleshov et al., 2016). Lists of genes showing differential expression and APA regulation were submitted and analysed for enrichment in 'Biological Processes'. A significant enrichment was defined by $p<0.05$.

2.11 Quantitive Real-Time PCR

1ug RNA was reverse transcribed into cDNA using a oligod(T) primer with SuperScript III Reverse Transcriptase (replaced with H2O in control reactions) as per the manufacturer's instructions. cDNA was diluted (1:100), analysed by Real Time quantitive PCR (qPCR)

using a RotorGene (Corbett), SYBR Green mix (Bioline) and primers. Primer pairs were used to detect both isoforms generated by proximal and distal poly(A) site usage (representing total mRNA isoforms), and another primer pair were used to detect isoforms generated by cleavage at the distal poly(A) site (representing long 3'UTR isoforms). The ratio of long 3'UTR isoforms to total mRNA isoforms at the different time points were subsequently log2 transformed (log2). The different ratios were then compared between infected (3, 9 and 12 hpi) and mock nuclear and cytoplasmic samples using qRT-PCR results of the 3 independent biological repeats for each time point.

.

2.12 RNAi Transfection

Transient protein knockdown was achieved by RNAi transfection. CPSF30 (5'-ACGCCAACAGAUGCACCAAAGTT, 5'-CUUUGGUGCAUCUGUUGGCGUTT) siRNA from MWG Eurofins was ordered according to the sequence obtained from Tian Lab. WDR33 and SAM68 silencer select siRNAs were ordered from Life Technologies. HEXIM1 siRNA (5'-CGAUGACGACUUCAUGGAATT) were ordered from Ambion Life Technologies. The siRNA was transfected in a reaction mix containing Opti-MEM medium and Lipofectamine RNAiMax (Invitrogen) when cells reached 30% confluence. After 24 hours of the first transfection, siRNA containing media was replaced by "normal" DMEM (10%FCS) media. Cells were transfected again 24 hours after the recovery step using the same conditions as described in the first step. RNA fractions and protein were extracted 24 hours after the second transfection. The knockdown efficiency was measured by RT-PCR or Western Blotting. Negative control siRNA was transfected by the same method for optimal and fair comparisons.

2.13 Cloning

2.13.1 Genomic DNA isolation

Approximately 1.5x10^7 A549 or HEK293T cells were used for genomic DNA (gDNA) isolation. Cells were washed in PBS, trypsinised, and resuspended in 200µl PBS. gDNA was isolated using Genomic DNA Isolation Kit (Sigma-Aldrich) as per manufacturer's instructions.

2.13.2 Polymerase Chain Reaction (PCR)

Phusion® High-Fidelity DNA polymerase (NEB) was used for high fidelity amplification. PCR was performed according to manufacturer's instructions. Target genes were amplified from A549 and HEK293T gDNA. The annealing temperature was adjusted for each reaction according to the melting temperature of the primer pairs, and extension time was adjusted according to product size. Reactions were executed by a thermocycler.

2.13.3 Restriction Enzyme Digestion

Restriction enzyme digestion reactions were performed in a 50µl reaction containing DNA, appropriate 10x enzyme buffer, 1µl restriction enzymes and water. Reaction were incubated at 37°C for 2 hours and inactivated at 65°C or 80°C for 20 minutes. Some restriction enzymes cannot be heat inactivated, therefore digestion products were purified by running an agarose gel and using QuiaQuick Gel Extraction Kit (Qiagen) as per manufacturer's instructions.

2.13.4 Kinase Reaction

The kinase enzyme catalyses phosphorylation by transferring a phosphate group from a nucleoside triphosphate (ATP) to the 5'-hydroxyl terminus of polynucleotides (double and

single stranded DNA and RNA) and nucleoside 3-monophosphates. The kinase reaction of primers/ PCR products was done in 1x T4 Ligase buffer (NEB) and 1 unit of T4 PNK (NEB). The mixture was incubated at 37°C for 30 minutes and inactivated at 65°C for 20 minutes.

2.13.5 Ligation

DNA ligation was performed in a 20µl mixture of digested DNA insert and DNA plasmid, 2µl T4 DNA ligation buffer, 1µl T4 DNA ligase (NEB) and water. The mixture was incubated at 16°C overnight and inactivated at 65°C for 20 minutes.

2.13.6 Site-Directed PCR Mutagenesis

Mutations were introduced to plasmids pcDNA -NS1 (Robb et al., 2010), and (luciferase expression plasmids, DDX5 Long). PCR was performed using ~100ng template plasmid, with 0.5µM overlapping primers containing the altered sequences (Table 2.1:g550a NS1 Mut, DDX5 PA Mut). The mutated plasmids were obtained using these primers to amplify the entire template plasmid using Q5 PCR polymerase (NEB) or Phusion® High-Fidelity DNA polymerase (NEB) (as per manufacturers instructions) in 1x Q5 reaction buffer or 1x Phusion HF buffer. The reaction mixture was incubated in thermocycler for 30 cycles as per manufacturer's instruction with specific annealing temperature and extension time that is dependent on melting temperature of primer pairs and product size.

2.13.7 DpnI Parental DNA Digestion

DpnI digestion, for 1-2 hours at 37°C, was performed to target and digest methylated parent DNA (without mutation) and not the PCR product. PCR purification was subsequently performed using QIAquick PCR Purification Kit (Qiagen) to remove primers, nucleotides, enzymes and other impurities from the product DNA samples.

2.13.8 Transformation of Plasmids into Competent Cells

Chemically competent E.Coli DH5α (Life Technologies) was thawed on ice. 2-2.5µl of plasmid DNA was added to 50 µl of DH5α competent cells. The mixture was incubated for 15 minutes on ice, then heated shocked at 42°C for 30s before another 15 minutes incubation on ice. 190µl LB medium (10g/L Trypton, 5g/L Yeast extract, 10g/L NaCl, pH7), was added to the mix and incubated at 37°C in shaker for 1 hour. The mixture was spread on LB agar plates containing 100ug/ml ampicillin and incubated at 37°C overnight.

2.13.9 Plasmid DNA Isolation

A single bacterial colony was picked from LB agar plates and incubated in a suitable volume (5ml for mini-prep, 60ml for midi-prep, 100ml for maxi-prep) of 1X LB buffer overnight for plasmid DNA isolation. Cells were pelleted by centrifugation at 4°C for 15 minutes at 6000g. For small scale plasmid DNA isolation, the pellet was resuspended in 200µl P1 Buffer. 250µl P2 Buffer was added and mixed gently by inverting 4-6 times and incubated at room temperature for 5 minutes. 350µl P3 Buffer was added and mixed thoroughly by inverting 4-6 times and incubated on ice for 15 minutes. The mixture was then centrifuged at 18000g for 10 minutes at 4°C. The resulting supernatant was transferred to 1ml 100% isopropanol for precipitation. The mixture was spun immediately at 16000g for 20 minutes at 4°C. The DNA pellet was washed with 1ml 80% ethanol without disturbing the pellet. The pellet was air-dried for 10 minutes and dissolved 30 µl of water. For larger scale plasmid DNA preparation, all volumes of buffers are adjusted according to manufacturer’s instructions (Qiagen plasmid Midi/Maxi Kit).

2.14 Transient Overexpression of NS1, Polymerase Subunits and Mutants in HEK293T

In experiments where cells were subjected to subcellular fractionation, approximately 3 × 10^7 HEK-293T cells were transiently transfected with 30 µg of the protein expression plasmids pcDNA-NS1 (Robb et al., 2010), pcDNA-NS1-G184R (Das et al., 2008), pcDNA-PB1, pcDNA-PB2, pcDNA-PA (Fodor et al., 2002) or pcDNA-GFP (Fodor and Smith, 2004) and Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions. 48 h post-transfection cells were harvested and fractionated and processed as described above.

For whole cell protein analysis, approximately 5.9 × 10^5 HEK-293T cells were transiently transfected in 6-well plates with 1 µg of pcDNA-NS1, pcDNA-NS1-G184R, pcDNA-PB1, pcDNA-PB2, pcDNA-PA or pcDNA-GFP and Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions. 48 h post-transfection cells were harvested and lysed in 100 µl lysis buffer (10 mM Tris pH8, 140 mM NaCl, 1.5 mM $MgCl_2$, 0.5% NP-40) and analysed by SDS-PAGE and Western blots.

2.15 Luciferase Reporter Gene Assay

The short and long 3'UTRs of human target gene of DDX5 (using primers from Table 2.1: DDX5 F1 AgeI, short R1, Long R1 XbaI (luciferase)) were amplified from the A549 gDNA by Phusion® High-Fidelity DNA polymerase (NEB). 3'UTRs of differing lengths termed short (from usage of proximal pA site) and long (from usage of distal pA site), were cloned into piS2 plasmids (Addgene; gift from David Bartel). piS2 plasmids are deigned to provide a rapid quantitative approach for monitoring changes in expression of target gene fused to a luciferase reporter gene. piS2 vector contains a SV40 promoter, Renilla luciferase

reporter gene, multiple cloning region (downstream of the Renilla translational stop codon), SV40 pA signal and Ampicilin resistance gene.

Site-directed PCR mutagenesis on the proximal pA site, as described in section 2.13.6, was only performed on plasmids containing the long 3'UTR. Subsequently, DPNI treatment (2.13.7) was applied to digest parent plasmids with wild-type pA site. PCR purification was performed, which led to transformation and plasmid DNA isolation as stated above.

For reporter gene assays, 293T cells were plated on 24-well plates. When the cells reached 60-70% confluence, they were transfected using Lipofectamine 3000 (Invitrogen) with 250ng of plasmid DNA containing either the short UTR (usage of proximal CPA site) or the long UTR (usage of distal CPA site) that is fused to Renilla luciferase reporter gene, and 250ng of control plasmid pGL4.20 vector containing firefly luc2 luciferase reporter gene and BMAL1 promoter. Transfected cells were incubated for 48 hours. Subsequently, cells were washed with PBS, 100µl Promega Luciferase Assay Systems kit 1x PLB lysis buffer was applied and rocked for 15 minutes at RTP. The lysate was transferred onto Eppendorf tube and centrifuged at 16,000g for 30 seconds. 20µl of the supernatant was used for the reaction and added to 100µl of Luciferase Assay Reagent II, and control Firefly Luciferase expression was measured. After which 100µl of Stop & Glo® reagent was added to the mix, and a second measurement was performed for Renilla Luciferase expression (short or long expression). Luciferase expression was measured on the PHERAstar Luminometer.

2.16 Osmotic Shock Treatment

A549 cells were treated with 80mM KCl (osmotic stress) in DMEM containing 0.5%FCS at 37˚C for 1 h (refer to Vilborg et al., 2015; Vilborg & Steitz., 2017).

Table 2.1 Sequence of primers used in this study

Primer	5'-3' Sequence
AP oligo DT	GCCCACGCGTCGACTAGTACTTTTTTTTTTTTTTT
MALAT1 F	AGAGAATTCCCTGCAAATTGTTAACAGAA
MALAT1 R	TCAGCTTCCGCTAAGATGCTAGCTT
GAPDH F	CTGAGAACGGGAAGCTTGTC
GAPDH R	CGTCCACCACTAGACACGTTG
MAVS Total F (QPCR)	TGTCCATCCACCAACCCAAATGTCC
MAVS Total R (QPCR)	CTTATTTCTGTACTTTTTGTAGGAA
MAVS Long F (QPCR)	GAAGAAGCTGGCAGCCACTA
MAVS Long R (QPCR)	GTACTGTGTGTCCCCCATTC
HEXIM1 Total F (QPCR)	GAGACTAGACTGAAACTTTTT
HEXIM1 Total R (QPCR)	ACTCTTCAGTTTCACATAATTAAAAG
HEXIM1 Long F (QPCR)	CAATTTTTTCCTTGAATGGGTTTAAG
HEXIM1 Long R (QPCR)	CTTTTGAATGGCCGAGGCAGAGTTAA
TRIM25 Total F (QPCR)	GTTAGGAGTATTTGAGCTGTTATTTC
TRIM25 Total R (QPCR)	GCTGACAACAGTATGACGAA
TRIM25 Long F (QPCR)	CCTTGTTTCAATTGGGGTAG
TRIM25 Long R (QPCR)	CCTATTACCCTGGAAATGTC
SMAD3 Total F (QPCR)	TGACAGCTTTCTGGGATGTC
SMAD3 Total R (QPCR)	GGGCTCTTCCTACCTAGAAC
SMAD3 Long F (QPCR)	GTGCTGGAACATCATCTCAG
SMAD3 Long R (QPCR)	CTAAGAGTCAAAGTCCCTGC
DDX5 Total F (QPCR)	ACATGGGAATTGCAGAAATGAC
DDX5 Total R (QPCR)	TTGTGAAAACACTGCCTGCAT
DDX5 Long F (QPCR)	ACTTCGGCTCACCAACGTAA
DDX5 Long R (QPCR)	TGCTCCAACCTACCCTCTCC
HEXIM F1 SPEI (Luciferase)	AGTACTAGTACTGAAACTTTTTGGGGAGG
HEXIM Short R1 NOTI (Luciferase)	ACTGCGGCCGCTGTTCTCAGATTTGGCCAC
HEXIM1 Long R1 NOT1 (Luciferase)	ACTGCGGCCGCATAATTTTTAGACCAAAGAAAAACT
HEXIM1 PA Mut F1 (Luciferase)	TGAGAACACGCAAATTCATTTTAG
HEXIM1 PA Mut R1 (Luciferase)	CTAAAATGAATTTGCGTGTTCTCA
SMAD3 F1 ECORV (Luciferase)	AGTGATATCAGACATCAAGTATGGTAGGGGAG
SMAD3 Short R1 NOT1 (Luciferase)	ACTGCGGCCGCGAATGCAACTGACTACATAAACCA

Primer	5’-3’ Sequence
SMAD3 Long R1 NOT1 (Luciferase)	ACTGCGGCCGCAGAAGGGCACTTACAAGAGGAAG
SMAD3 PA Mut F1 (Luciferase)	GTTGCATTCCGCAAATCAACTTTATC
SMAD3 PA Mut R1 (Luciferase)	GATAAAGTTGATTTGCGGAATGCAAC
MAVS F1 AgeI (Luciferase)	TAAACCGGTTGAAGCCCTGGGCTCTTCCCAC
MAVS Short R1 XbaI (Luciferase)	ACTTCTAGATCCACATCTCCAACAGGGAG
MAVS Long R1 AgeI (Luciferase)	ATCTCTAGATTTATTTTTTGAGTCTGG
MAVS PA Mut F1 (Luciferase)	GGAGATGTGGAACGCAAAACCACC
MAVS PA Mut R1 (Luciferase)	GGTGGTTTTGCGTTCCACATCTCC
TRIM25 F1 AgeI (Luciferase)	TAAACCGGTGCAGGCTGTAGGCACTTG
TRIM25 Short R1 XbaI (Luciferase)	TAAACCGGTGCAGGCTGTAGGCACTTG
TRIM25 Long R1 XbaI (Luciferase)	ATCTCTAGAATAAAGGATACAAATCAG
TRIM25 PA Mut F (Luciferase)	GTTTGAAAATGTCTATACGCAAAAGCTTT
TRIM25 PA Mut R (Luciferase)	AAAGCTTTTGCGTATAGACATTTTCAAAC
DDX5 F1 AgeI (Luciferase)	AGTACCGGTGACTTTAGAAGTATATGTAAATGTCT GT
DDX5 Short R1 NotI (Luciferase)	TAAGCGGCCGCGGAAGGCCATGCATTGCCTT
DDX5 Long R1 NotI (Luciferase)	TAAGCGGCCGCAGTTAAGACGACCACAGGC
DDX5 PA Mut F	GCATGGCCTTCCACTAGTAAATATTTGAAG
DDX5 PA Mut R	CTTCAAATATTTACTAGTGGAAGGCCATGC
NA+ (Primer Extension)	TCCAGTATGGTTTTGATTTCCG
NA- (Primer Extension)	TGGACTAGTGGGAGCATCAT
5 rRNA (Primer Extension)	TCCCAGGCGGTCTCCCATCC
g550a NS1 Mut F (NS1 Mutation Control)	TATTATTCCATTCAAGTCTTCCGATGAGGACCCCA AC
g550a NS1 Mut R (NS1 Mutation Control)	GTTGGGGTCCTCATCGGAAGACTTGAATGGAATA ATA
MAVS readthrough F	GATCTCACCACTGTACTCCA
MAVS readthrough R	CATTTTGCCTGCCCTGCCACTT

Table 2.2 Antibodies

Antibody	Host	Concentration	Supplier	Supplier Number
HSP60	Rabbit	1:1000	Bethyl	A302-846A
HEXIM1	Rabbit	1:500	Abcam	ab25388
DDX5	Rabbit	1:1000	Bethyl	A300-522A-T
MAVS	Rabbit	1:500	Abcam	ab31334
SMAD3	Rabbit	1:500	GeneTex	GTX108638
NS1	Rabbit	1:1000	GeneTex	GTX125990
PA	Rabbit	1:1000	GeneTex	GTX1188991
PB1	Rabbit	1:1000	GeneTex	GTX125923
PB2	Rabbit	1:1000	GeneTex	GTX125925
NP (immunofluorescence)	Rabbit	1:500	Kind Gift from Otto Haller (University of Freiburg, Germany)	
M1	Rabbit	1:1000	GeneTex	GTX127356

Chapter 3: Establishing Controls and Development of the Infection Time Course Experiment

3.1 Introduction

The aim of this thesis is to investigate APA during influenza A/WSN/33 infection. Influenza A viruses have evolved many strategies to interact with host factors in order to facilitate replication, survival and also to abrogate immune response, as mentioned in chapter 1. One of the strategies involves nonstructural protein 1 (NS1) which has been shown to interact with cleavage and polyadenylation factor (CPSF30) thus restricting cleavage and polyadenylation of host transcripts (Nemeroff et al., 1998). Since CPSF30 has been shown to regulate CPA site choice in plants (Thomas et al., 2012), this provides a potential link between infection and APA.

In order to investigate the link between infection and APA, a consistent and well-controlled protocol will have to be established, which accounts for several critical factors. A well developed infection time course will have to consider the following: determining the biologically relevant infection time points; the amount of virus needed to establish an infection; investigating ways to purify virus media to limit effects to the virus rather than contaminating factors stemming from the amplifier host cell; ensuring efficient infection of high proportion of cells in every independent biological repeat; and finally utilise a clean subcellular fraction protocol to isolate nuclear and cytoplasm fractions for further downstream analyses of APA such as 3' deep sequencing and qRT-PCR. To examine the 'global' effects of influenza A virus infection on host alternative polyadenylation in human lung epithelial cells, biologically relevant infection periods and the dynamics of influenza infection need to be understood.

3.2 Results

3.2.1 Determining Biologically Relevant Infection Time Points

In order to evaluate the effects of influenza virus infection on APA, large scale growth of A/WSN/33 viral stocks were necessary for subsequent use for the infection time course. A/WSN/33 virus was grown in bovine MDBK cells and viral supernatants were harvested 2 days after infection, and decontaminated from cellular debris by low-speed centrifugation. Importantly, to exclude the effects of any contaminating co-purified host factors that may be present in the growth media of the viral supernatant, viral preparations were purified by ultracentrifugation through a sucrose gradient (Hutchinson et al., 2014).

Next, plaque assays were performed to determine the viral titres as “plaque forming units per ml” (pfu/ml) therefore known amounts of virus can be used for subsequent infections (Figure 3.1A). MDBK cells were infected by 6 serial 10-fold dilutions of the viral stock and were subsequently overlaid by 2% agarose media for 3 days at 37°C. Plaques are formed when cells within the fixed agarose monolayer are infected, and the spread of the infection from new viruses are restricted to neighbouring cells by the gel. Consequently, each infectious particle produces a circular zone of infected cells, which are described as plaques. In order to calculate the viral concentration in pfu/ml, using the plaque assay in figure 3.1A as an example, plaques were counted from each well that were infected from serial 10-fold dilutions of virus stocks (e.g. 160 plaques in 10^{-5} dilution).The concentration of the viral stock is calculated by counting the number plaques and factoring in the dilution factor and the volume of virus used to infect each well. Since 160 plaques were counted in10^{-5} dilution, at 10^{0} dilution the equivalent plaque forming units would be $1.6x10^{7}$. As 0.4ml of viral media was used to infect the cells, $1.6x10^{7}$ PFU had to be divided by 0.4ml to get the viral concentration of $4x10^{8}$ PFU/ml.

The viral stocks with known concentration were then further analysed by subjecting equal concentrations of viral stocks that were purified (second lane) and viral stocks that were not purified (third lane) to SDS-PAGE and silver staining (Figure 3.1B). This shows a significantly reduced “smearing” of proteins present in the purified viral stock (Figure 3.2B, unpurified lane). In contrast, clear individual protein bands corresponding to viral proteins (Hutchinson et al., 2014; Shaw et al., 2008) that are largely free of contaminating MBDK host cell proteins are apparent when purified viral stocks are loaded (Figure 3.1B, purified lane). However, it is important to note that although ultracentrifugation will eliminate most contaminating factors, low level of “virion-sized” material such as exosomes will not be removed (Hutchinson and Fodor, 2014). This problem can be solved by performing a haemadsorption/elution step, which first selects sialic acid binding materials such as viral haemagglutinin (HA), and then selects materials with sialic acid cleaving properties such as viral neuraminidase (NA). However, it has been shown that this method introduces alternative contaminants derived from chicken red bloods cells (Hutchinson et al., 2014). I thus decided that, whilst being aware of the possible issues associated with potential contaminating exosomes particles, not to further purify the viral stocks used for the experiments described in this thesis.

Once the harvested viral stocks were purified, viral infection dynamics of the stocks needed to be established to determine the optimal biologically relevant infection time points for the subsequent APA analysis. For this, A549 cells were infected with A/WSN/33 at a high multiplicity of infection (MOI) of 5 by using 5 times the amount of virus to the number of cells. According to the Poisson distribution theory, a high multiplicity of infection will ensure that nearly every cell is infected. To gauge the viral protein expression dynamics, whole cell protein was then extracted from cells at 5 time points (3,6,9,12,18 hours post infection (hpi)). As a control, whole cell protein from uninfected cells (mock)

were harvested at the equivalent time point of the 3h post-infection samples. The protein samples from the mock and infected cells from the 1st replicate at different time points were subsequently analysed by western blotting using antibodies against viral protein NS1. As can be seen in Figure 3.1C, NS1 protein began to accumulate at 6 hpi and peak at 9 hpi with a slight decrease at 18 hpi (Figure 3.1C). Given the established link between NS1 and host cleavage and polyadenylation (Nemeroff et al., 1998) and the above established NS1 protein expression dynamics, the time points including and after 6 hpi are reasonable to address how influenza infection may affect host APA.

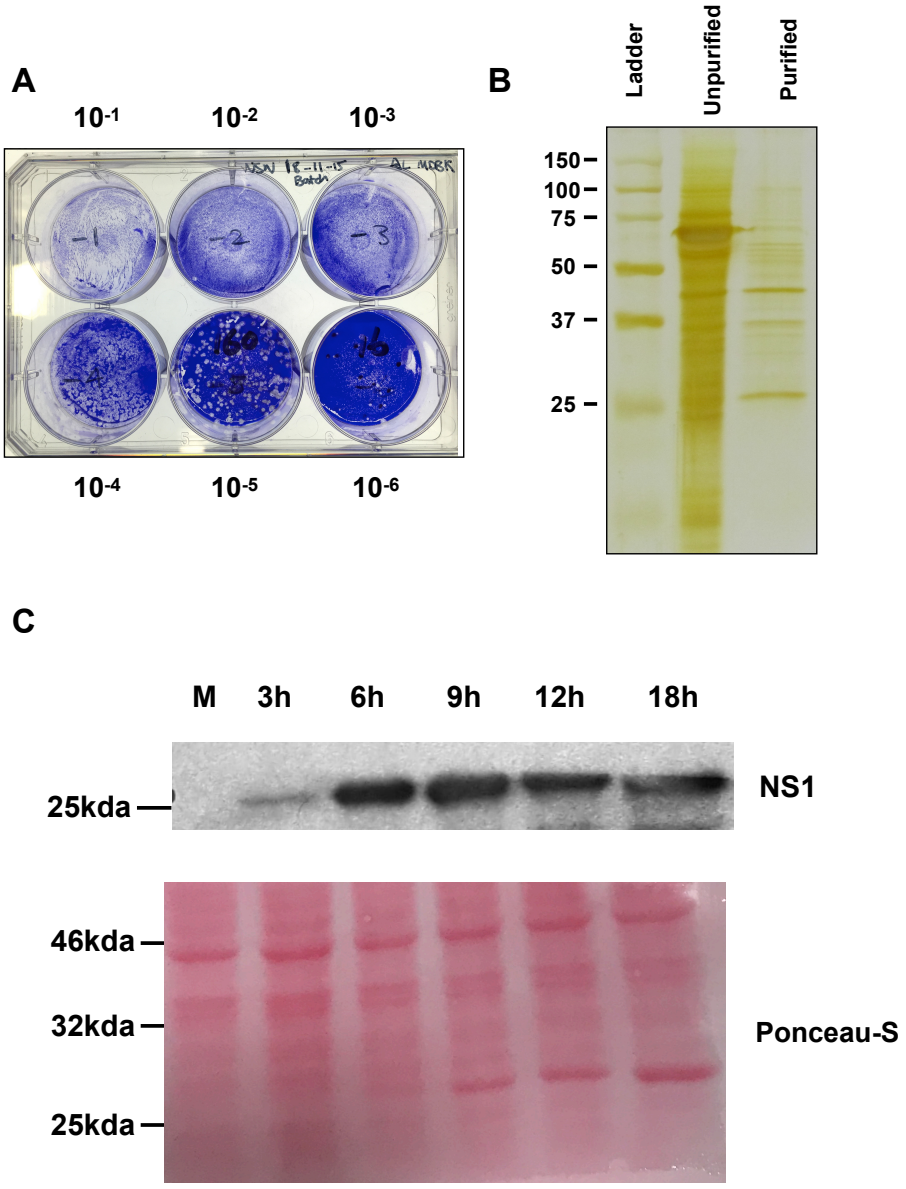


Figure 3.1 Quality Controls to Determine Viral Titre, Purity of Ultracentrifuged Viral Stock and Dynamics of Viral Protein Expression during the Course of Infection

A. Plaque Assays were performed to determine the concentration of viral stock known as viral titre: Bovine MDBK cells were plated on 6-wells. At 90-100% confluency, cells were infected with 400µl of serially diluted virus. 2ml of 2% agarose overlay was applied and plates were stained after 72 h incubation at 37°C. Viral titre (plaque forming units/ml) was determined by dividing the average number of plaques by dilution factor multiplied by volume of virus used.

B. Influenza A/WSN/33 (H1N1) virus harvested from viral supernatant was subjected to purification by ultracentrifugation through a 30% sucrose gradient. Equal amounts of unpurified (2nd lane) and purified (3rd lane) viral stocks were analysed by SDS-PAGE and silver staining. Molecular masses, indicated in the ladder lane (1st lane), are represented in kilodaltons. Purified virus displayed significantly reduced “smearing” of proteins, thus less contaminants compared to unpurified virus.

C. Human A549 Lung cells were infected with A/WSN/33 virus at MOI 5. Whole cell lysates containing protein were extracted at the displayed time points (3, 6, 9, 12, 18 hours post infection), and mock. Viral protein expression dynamics were analysed by western blotting using antibodies against viral protein NS1. NS1 protein expression began to accumulate at 6 hpi, whilst NS1 expression peak at 9 hpi, its expression begins decrease at 18hpi. Ponceau-S stains of the western membranes are shown as a loading control.

3.2.2 Establishing an Efficient Viral Infection Protocol for Steady State APA Analysis

Since analysis of differential APA and gene expression after viral infection takes a global approach, infection at a high efficiency whereby a high proportion of cells are successfully infected is required. Therefore, cells were infected at a high MOI of 5 and fixed at indicated time points (mock, 3, 6, 9, 12, 18 hpi) for immunofluorescent microscopy in order to control for infection efficiency.

In order for A549 cells to be initially infected, room temperature incubation with the virus for 1 hour prior to incubation at 37°C was required. The purpose of room temperature incubation is to allow virions to effectively attach to sialic acids present on cellular surface membrane glycoproteins via haemagglutinin. At room temperature, neuraminidase activity remains inactive whereby its ability to catalyse the hydrolysis of sialic acid residues on host cell surface membranes is inhibited. Thus, it allows virions to attach to cells and initiate viral entry. The initial strategy for infection was to aspirate the viral media after room temperature incubation in order to 'synchronise' all of the infections. However, infection efficiency was vastly reduced in cells with removed virus media after 1 hour room temperature incubation compared to cells with virus not removed after 1 hour room temperature incubation. As clearly shown at 6 hours infection at 37°C (Figure 3.2A, 3.2B-Middle Panels), staining of infected cells with an antibody targeting the viral NP protein already detected signals of NP in the majority of the cells without aspirating viral media contrary to cells with aspirated virus. At 18 hpi, both conditions displayed efficient infection. Hence, this suggests that aspirating viral media after 1 hour reduces infection efficiency in comparison to cells infected without viral aspiration.

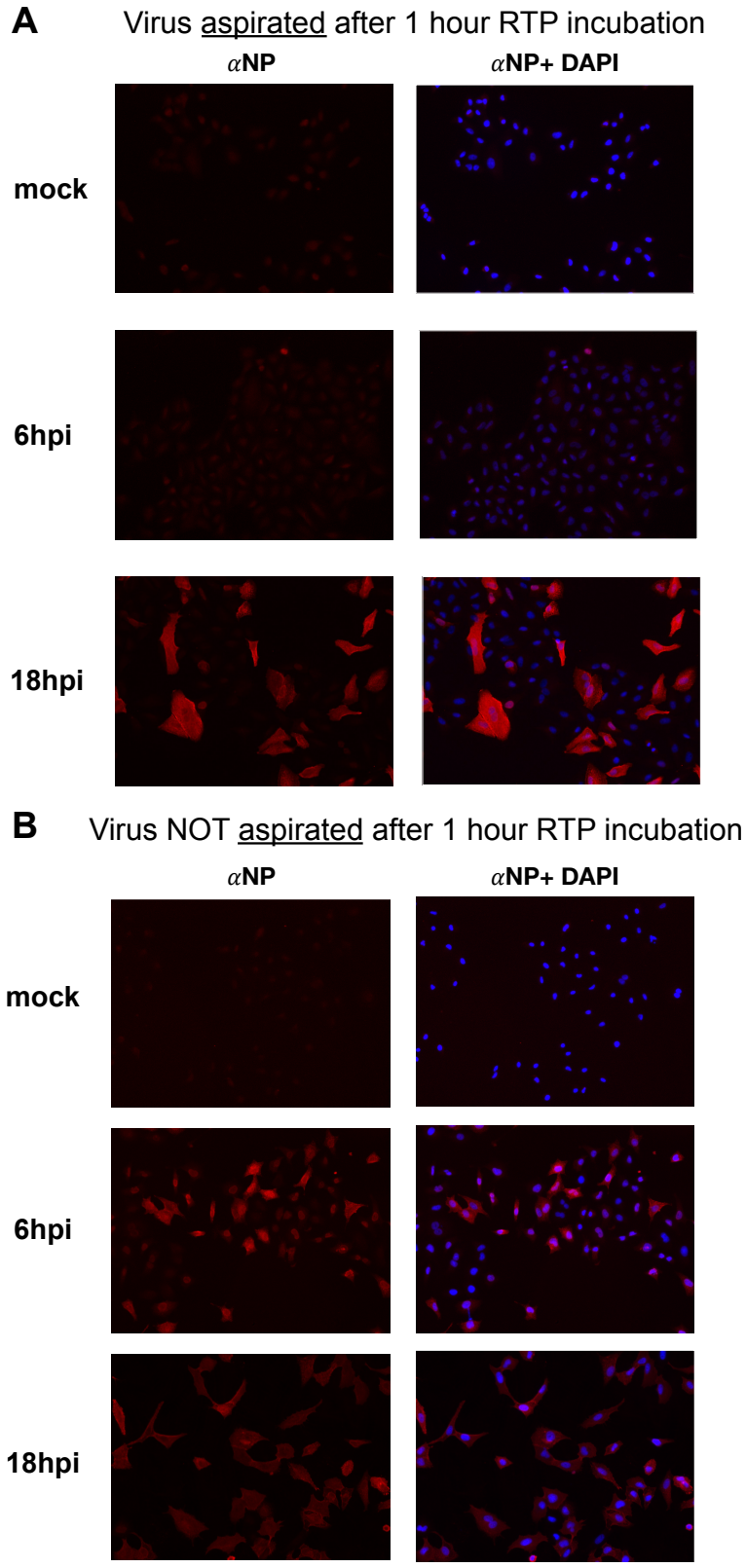


Figure 3.2 Infection Efficiency Compared Between Cells with Aspirated Virus Media vs Cells with Non-Aspirated Virus Media after 1 hour Room Temperature (RTP) incubation.

Human A549 cells were infected with influenza A/WSN/33 (H1N1) virus at an MOI of 5. At time points post-infection cells mock, 6 and 18 hours were fixed and viral nucleoprotein (NP) (Red) and cell nuclei (Blue) were stained.

A. After 1 hour incubation at RTP prior to 37°C incubation, viral media was aspirated.

B. After 1 hour incubation at RTP prior to 37°C incubation, viral media was not aspirated.

When comparing 6 hpi (middle panels, **A** and **B**), cells with viral media aspirated after 1 hour RTP incubation displayed significantly less NP expression when compared to cells with non aspirated viral media. Efficient infection observed in both strategies at 18 hpi (bottom panel, **A** and **B**)

In fact, an initial trial experiment was then conducted, whereby further comparisons of both strategies of either aspirating viral media or not aspirating the viral media after 1 hour in RTP was initiated by utilising 3'sequencing with subsequent bioinformatic analyses. A549 cells were infected at a high multiplicity of infection (MOI=5) with influenza A/WSN/33 (H1N1) virus. Cells were subjected to subcellular fractionation to isolate nuclear and cytoplasmic RNA. For RNA analysis, the isolates of the different fractions were subsequently used to create 3'RNAseq libraries using the QuantSeq library kit from Lexogen. Briefly, an oligo dT primer is used to reverse transcribe polyadenylated RNA and second strand synthesis is performed using barcoded random primers which results in the dsDNA that are enriched in sequences close to the polyA site. The DNA fragments are subsequently purified and the library is PCR amplified compatible for sequencing on the IonProton platform. The amplified libraries were subsequently sequenced and host reads were mapped against the hg19 human genome using the inbuilt Torrent Mapping alignment Software (TMAP). In tandem, the viral reads were mapped to the WSN/33 genome using the Galaxy (https://usegalaxy.org/) platform as described in the Material and Methods. BAM files obtained from sequence alignment were then subjected to downstream bioinformatics analysis using Samtools Flagstat on Galaxy to evaluate the % of reads aligned to A/WSN/33 viral genome at the indicated time points.

By comparing the relative amount of viral mRNA between 'aspirated' and 'non aspirated' infection strategies, infection efficiency demonstrated significant differences between them. Infected cells with 'aspirated' strategy from 1 individual biological repeat displayed less relative amount of viral mRNA when comparing to infected cells with 'non-aspirated' strategy which is averaged between 2 individual biological repeats in all time points (Figure 3.3). For 'aspirated' samples, the peak relative amount is at a much later time point (18hpi) with only 10% of all polyadenylated transcripts in the nucleus are mapped to the viral

genome (Figure 3.3A, pink). By contrast, 'non-aspirated' samples display peak amount of viral mRNA at an earlier time point (9hpi) with 29% of all polyadenylated transcripts in the nucleus are produced by the virus (Figure 3.3B, pink). With the exception at 9 hpi for 'aspirated virus' cytoplasmic fraction, where the peak reaches 8.5% (Figure 3.3A, blue), the rest of the time points display lower relative amount of viral mRNA in comparison to the 'non-aspirated' cytoplasmic fractions (Figure 3.4B, blue). Thus, by aspirating the viral media after 1 hour in RTP incubation, not only is the infection dynamics different, but also infection efficiency is significantly reduced when virus media is aspirated.

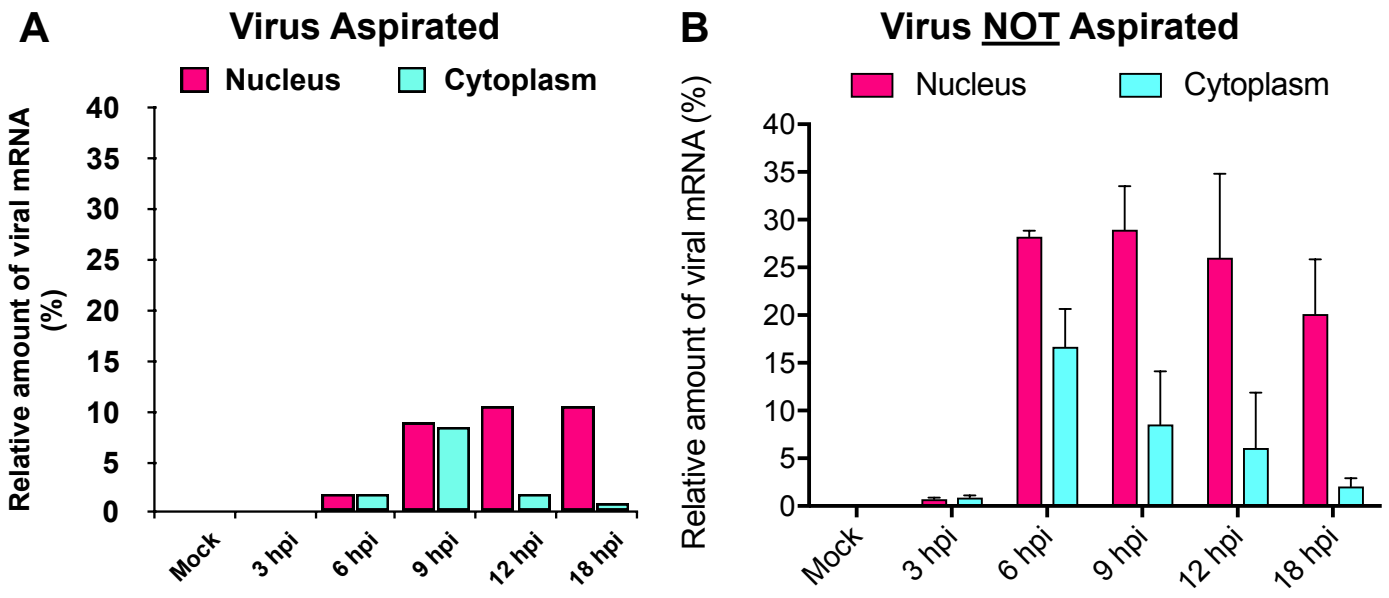


Figure 3.3 Significant Differences in Relative Amount of Viral mRNA when Aspirating vs Not Aspirating Viral Media after 1 hour in RTP

Bioinformatics analysis done by mapping sequencing reads to A/WSN/33 genome by Galaxy. Samtools FlagStat generated number of reads aligned to viral genome. Relative amount of viral mRNA (%) was calculated by dividing number of reads aligned to viral genome over total (viral and human) reads per time point.

A. Bar graph represents relative amount of viral mRNA from 'viral media aspirated' nuclear (pink) and cytoplasmic (blue) fractions at indicated time points.

B. Bar graph represents relative amount of viral mRNA from 'viral media non-aspirated' nuclear (pink) and cytoplasmic (blue) fractions at indicated time points.

Infected cells with 'aspirated' strategy from 1 individual biological repeat displayed less relative amount of viral mRNA when comparing to infected cells with 'non-aspirated' strategy which is averaged between 2 individual biological repeats in all time points.

Furthermore, the analysis of total 3'UTR APA events (presented in scatterplots in figure 3.4) was done to further emphasise the detrimental effects of infection efficiency by aspirating viral media after 1 hour in RTP. When comparing between nuclear fractions of time points 6 (figure 3.4 A, B) and 18 hpi (Figure 3.4 C, D), it clearly shows significantly less, with near 9-fold difference, of total 3'UTR APA events in the 'virus aspirated' fractions than 'non-aspirated'. Similar fold differences can be seen in other time points and fractions (cytoplasmic). Scatterplots of 'aspirated' (supplementary figure 1) and 'non-aspirated' (supplementary figure 2) are shown in supplementary figures.

These methods, serving as an alternative to immunofluorescence microscopy, indicate by aspirating the virus after 1 hour, the infection dynamics and efficiency can be reduced significantly. This is in agreement with the immunofluorescence experiment (figure 3.3). Viral media was therefore not removed prior to 37°C incubation, and immunofluorescence microscopy was performed after every time course experiment to ensure high infection efficiency.

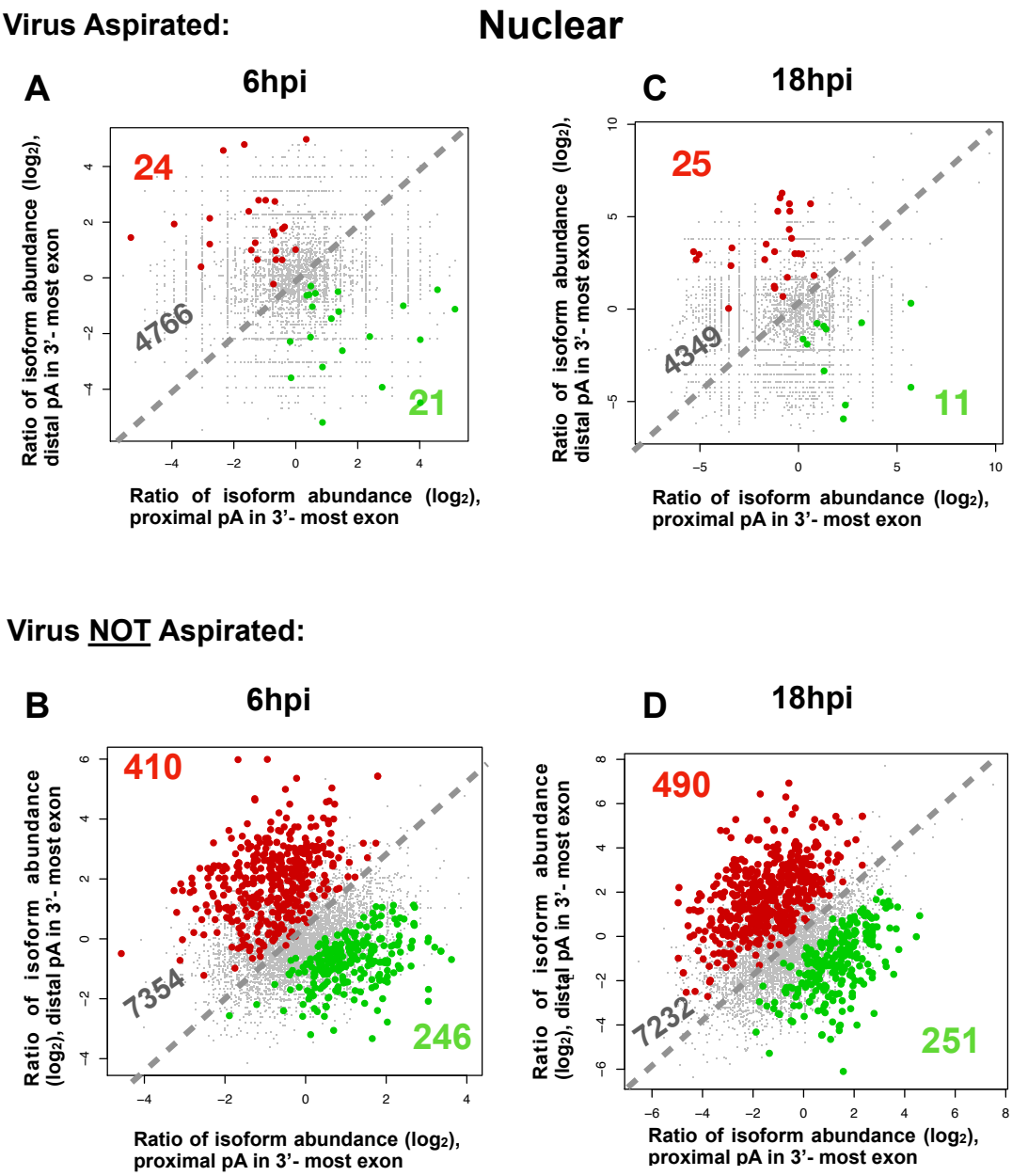


Figure 3.4 Effect of Aspirating Virus after 1 hour in RTP on total 3'UTR APA events

A. A total of 4766 expressed genes (grey) were identified to be regulated by APA in nuclear RNA isolated from 6hpi infected A549 cells compared to mock cells; viral media was aspirated after 1 hour incubation in RTP. 45 (0.94%) of these APA events were statistically significant (Fisher's exact test, $p<0.01$, with a minimum 5% change in APA isoform abundance). Red dots represent 3'UTR lengthening, while green dots represent 3'UTR shortening events.

B. As in panel A, comparing 6hpi to mock in nuclear RNA, but isolated from infected cells with non-aspirated viral media. A total of 7354 (grey) expressed genes identified, with 656 (8.9%) of these APA events were statistically significant.

C. A total of 4349 expressed genes identified to be regulated by APA in nuclear RNA isolated from 18hpi cells compared to mock cells; viral media aspirated. 36 (0.83%) of these APA events were statistically significant.

D. As in panel C, comparing 18hpi to mock in nuclear RNA, but isolated from infected cells with non-aspirated viral media. A total of 7232 expressed genes identified, with 741 (10.2%) of these APA events were statistically significant.

9-fold less difference of total 3'UTR APA events in the 'virus aspirated' fractions than in 'virus non-aspirated' fractions (only nuclear displayed, whilst it is also true for cytoplasmic fractions). Rest of the Scatterplots are displayed in supplementary figures 1 and 2.

3.2.3 Rationale for Subcellular Fractionation

Dynamic changes in APA profiles in response to specific diseases and physiological cues has been observed in many studies. It is imperative to elucidate the physiological roles of global APA changes and determine the mechanisms that are responsible for shaping these profiles. Since most studies focus on changes in whole cell RNA APA profiles, it is therefore difficult to distinguish between whether APA is regulated through active or passive APA mechanisms. Therefore, misinterpretations of the underlying mechanisms can occur when determining the cause and effect of differential APA profiles and in turn misconstrue the physiological relevance of APA. For example, in the context of cancer, shortening events are often attributed to the selection of proximal sites at the point of cleavage to avoid microRNA targeting post transcriptionally. However, this interpretation overlooked the fact that different cancer subtypes have distinctive miRNA signatures (Calin and Croce., 2006) and that microRNA levels are subjected to dramatic changes during oncogenic transformation (Zhou et al., 2015). Hence differential APA profiles in the context of cancer may be due to passive APA acting at the post-transcriptional level. A study highlighted this possibility in MCF-7 breast cancer cell lines by showing genes with targets for up-regulated microRNAs in their alternative UTR led to increased possibility of 3'UTR shortening relative to those without targets. The abundance and enrichment of microRNA target sites in the alternative UTR has been shown to influence selective degradation of longer transcripts isoforms by unregulated microRNAs and heavily impact APA profiles in MCF-cells (Liaw et al., 2013).

To tackle the aforementioned issues of misinterpreting differential APA profiles, subcellular fractionation approach permits the comparisons of APA profiles in nuclear and cytoplasmic fractions of the same cell population, and was utilised in this project to distinguish whether changes in APA profiles are due to differential selection of CPA site choice described as

‘Active’ APA where the changes will seen in the nucleus, or due to changes at a post-transcriptional level known as ‘Passive’ APA effects of which would be seen in the cytoplasm.

The advantages of subcellular fractionation are: firstly RNA used for APA analysis is from cells that are completely unperturbed, unlike previous studies that attempted to quantify the decay rates of individual APA isoforms after blocking transcription either by actinomycin D in mouse cells (Gupta et al., 2014) or by using temperature-sensitive strain of Saccharomyces cerevisiae containing a modified *rpb1-1* allele that arrests RNA Pol II transcription at 37°C (Spies et al., 2013) therefore inducing additional stress to cells. Secondly, differentially regulated APA isoforms can be identified simply through comparing two samples, whereas transcriptional arrest experiments rely on multiple time points (Spies et al., 2013). Finally, subcellular fraction protocol can be applied to any cell population, from cultured cells to tissue samples.

However, the subcellular fractionation approach does have its disadvantages. This approach cannot distinguish between two mechanisms that alter APA isoform ratios in the cytoplasm compared to nucleus: either through differences in nuclear export rates or through differences in stability. It can only highlight that there is a difference in the regulation of these APA isoforms between the fractions. Individual gene studies are required to determine exactly how APA isoforms are differentially regulated.

Inevitably, for this approach to be successful, highly pure subcellular RNA fractions are required and a reproducible protocol with rigorous quality control steps is essential. Mock cells and cells exposed to influenza viral stress at 5 different time points were therefore fractionated and the purity of the different fractions were assessed. Reverse transcription

PCR (RT-PCR) of long non-coding RNA *MALAT1* was used as a sensitive method to confirm its presence in the nucleus and absence in the cytoplasm (Wilusz et al., 2008) relative to the amplification of *GAPDH,* which served as a loading control. The sole amplification of *MALAT1* in nucleus confirms successful fractionation (Figure 3.5A).

The quality of cell fractions was further assessed by observing the distribution of ribosomal RNA precursors (pre-rRNAs). In humans, 45S and 32S pre-rRNAs are processed into 28S, 18S and 5.8S rRNAs in the nucleus prior export to cytoplasm (Henras et al., 2015; Lafontaine, 2015). To observe pre-rRNAs, 2µg of RNA from each fraction was run on a denaturing 1% agarose gel. The restricted appearance of pre-rRNAs in the nuclear fractions and the absence of precursors in the cytoplasmic fractions further confirmed the purity with minimal contamination of nuclear RNA (Figure 3.5B). The RNA gel, in addition to assessing fractionation, allowed confirmation of RNA quality.

Despite the robustness of these controls, they cannot distinguish whether any cytoplasmic contaminants are within the nuclear fractions. Altogether, these data show that this is a highly efficient fractionation protocol, yielding high purity and quality of RNA for downstream experiments.

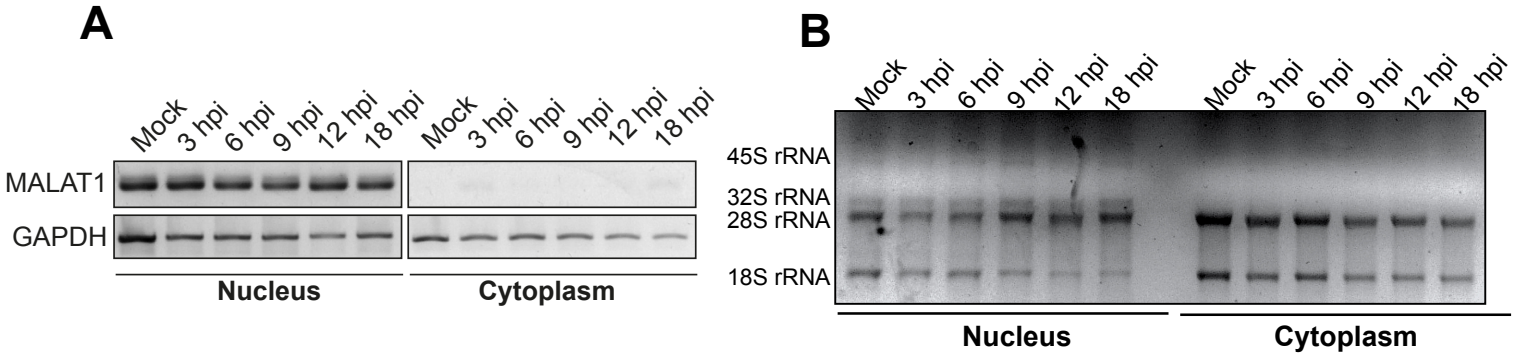


Figure 3.5 Subcellular Fractionation Controls for Infection Time Course (Previous Page)

A. Reverse Transcription PCR (RT-PCR) with primers targeting nuclear retained *MALAT1* (upper panel) and loading control *GAPDH* (lower panel). 1µg of RNA from each subcellular fraction was used. Only nuclear fraction (top left panel) amplified for *MALAT1*.

B. Denaturing RNA gel loaded with 2µg of RNA from each subcellular RNA fraction displaying the sole presence of precursor rRNA in the nucleus. The identity of each ribosomal RNA band is shown on the left hand side of the gel.

3.3 Discussion

Altogether, the combination of aforementioned controls facilitated the development of a consistent infection time course experiment. Pre-infection controls ensuring harvest of good titre viral stocks (determined by plaque assays) and subsequent purification of viral stocks allowed clean and robust infections with minimal contaminants. Post-infection controls from western blots and immunofluorescence microscopy allowed the determination of biologically relevant periods where potential changes in APA profiles are expected to be seen. Immunofluorescence microscopy detecting for viral NP indicated the differences in infection efficiencies when comparing infected cells with aspirated viral media after 1 hour RTP incubation to infected cells without aspirated viral media. Thus, in all future infection time courses, viral media was left on the cells after 1 hour RTP incubation to ensure highly efficient coverage and infection. A high MOI of 5 was used to ensure good infection of the majority of cells, as evidenced by immunofluorescence microscopy experiments that were performed after every time course experiment. Subcellular fractionation approach was applied after infection in order to isolate for high quality nuclear and cytoplasmic RNA. The quality was assessed by controls: RT-PCR of MALAT1 and denaturing RNA gels detecting the localisation of precursor rRNAs. Finally, once the purity of RNAs from each sample was determined, 3’ region extraction and deep sequencing on the Ion Torrent platform can be performed to assess the global changes in APA profiles between influenza viral infection stressed cells and mock cells. Furthermore, individual gene analyses such as quantitive real-time PCR (qRT PCR), luciferase expression assays, gene over expression and knockdown experiments analyses effects to

viral RNA production can also be executed. As displayed in figure 3.6, the established infection time course experiment was summarised and thus carried out in the next chapter.

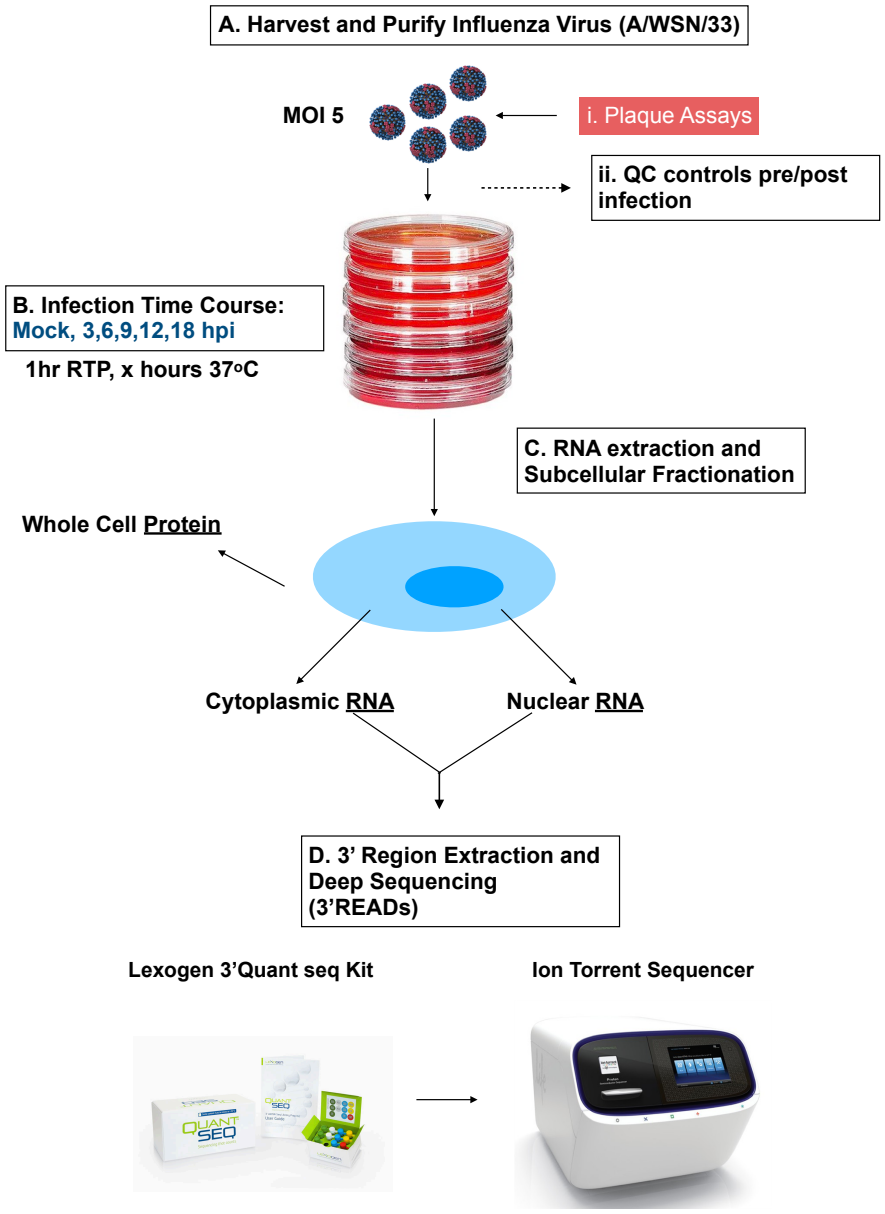


Figure 3.6 Schematic Workflow of Established Infection Time Course Experiment

A. Influenza A/WSN/33 (H1N1) was grown in bovine MDBK cells and viral supernatants were harvested 2 days after and centrifuged at low speeds to decontaminate any cellular debris. Ultracentrifugation through a sucrose gradient was performed to purify and concentrate viral stock. i) Plaque assays were performed to determine viral titre ii) Pre/post infection controls were performed to ensure clean and optimal infection.

B. Infection time points were determined as follows. A549 cells were subjected to infection at MOI of 5, initially incubated with viral media at RTP for 1 hour, subsequently incubated at 37°C until indicated time points.

C. RNA extraction and subcellular fractionation performed after indicated infection time points to isolate nuclear and cytoplasmic RNA.

D. 3' region extraction using Lexogen 3'Quant seq kit and deep sequencing were performed using Ion Torrent platform after fractionation.

Chapter 4: Investigating Role of Influenza Viral Stress on APA and Gene Expression

4.1 Introduction

Influenza viruses are a major threat for global health causing both seasonal epidemics as well as occasional global pandemic outbreaks, resulting in a significant burden on economies and health care systems world wide. Influenza viruses have a segmented single-stranded negative-sense RNA genome, each of which is coated by viral nucleoprotein (NP) and bound by a heterotrimeric viral RNA-dependent RNA polymerase forming a viral ribonucleoprotein (vRNP) complexes (Arranz et al., 2012; Moeller et al., 2012). The viral polymerase is responsible for copying the viral genome and transcribing viral mRNA (te Velthuis & Fodor, 2016). In order to copy the negative-sense viral RNA (vRNA) genome, the viral polymerase needs to produce an intermediate positive-sense complementary RNA (cRNA), which then is used as a template for vRNA synthesis. Viral transcription is mediated when the cap-binding domain of the PB2 subunit of the viral polymerase binds to the 7-methylguanosine caps at the 5' termini of host RNA transcripts and the endonuclease domain of the PA polymerase subunit subsequently cleaves the host transcript 8 to 14 nucleotides downstream of the cap. The resulting 3' end of the capped RNA fragment is fed into the active site of the viral polymerase and used as a primer to initiate viral transcription (Dias et al., 2009; Guilligay et al., 2008; Yuan et al., 2009; Krug et al., 1979). Therefore, viral transcription is dependent on active transcription by the host RNA polymerase II (Pol II) to supply these capped 5' RNA ends. In contrast to host transcription which relies on numerous enzymes for cleavage and polyadenylation (CPA) of host mRNA transcripts (Mandel et al., 2008; Shi et al., 2015), the viral

polymerase acts independently of the host CPA machinery and is able to generate a polyA tail by stuttering on a polyU stretch near the 5' vRNA end (Poon et al., 1999).

The effects of influenza virus infection on the host cell have been extensively studied. While the cell has several ways to combat viral infections by activating the innate immune response and inducing apoptosis to contain the viral infection, the virus has evolved multiple mechanisms to prevent the cell from responding to the infection. It is known that in the late stages of viral infection, protein synthesis is dominated by viral proteins (Yángüez et al., 2011). However, recent studies have shown that this observation is due to a takeover of the mRNA pool by the virus and not preferential translation of viral transcripts (Bercovich-Kinori et al., 2016). Influenza A viruses possess several mechanisms by which they are able to inhibit production of functional host mRNAs without blocking Pol II activity. The viral PA-X protein has been demonstrated to specifically degrade Pol II transcripts (Bavagnoli et al., 2015), thereby increasing the relative abundance of viral transcripts in the mRNA pool, resulting in enhanced translation of viral transcripts (la Luna et al., 1995). PA-X has also been shown to induce accumulation of PABPC1, thus stimulating nuclear retention and hyperadenylation of host transcripts (Khapersky et al., 2014). Viral infection also causes inhibition of Pol II elongation (Chan et al., 2006) and eventually degradation of Pol II by the viral polymerase (Vreede et al., 2010; Vreede and Fodor, 2010). The viral NS1 protein has been shown to interact with CPSF30, a crucial component of the host CPA machinery, to restrict cleavage and polyadenylation of host transcripts (Nemeroff et al., 1998). NS1 has also been shown to suppress U6 snRNA activity thus inhibiting splicing (Qiu et al., 1995). Additionally, interaction of NS1 with PABPII restricts polyadenylation of host transcripts by PAP, resulting in host transcripts with short polyA tails that are unable to be be exported from the nucleus (Chen et al., 1999). NS1 also interferes with nuclear export of host transcripts directly by interacting with and blocking the function of

components of the mRNA nuclear export machinery (Kuss et al., 2013; Satterly et al., 2007). Nuclear export of viral mRNA has been shown to also use NXF1-independent pathways, thus making it immune to NS1-mediated inhibition of nuclear export.

Cleavage and polyadenylation is an essential co-transcriptional pre-mRNA processing reaction that matures 3'end of primary protein encoding transcripts into functional mRNAs. This process involves the cleavage and polyadenlyation complex that recognises the poly(A) signals present in pre-mRNAs. The canonical poly(A) signal is a bipartite sequence element that composes an A-rich hexameric sequence [(A(A/U)UAAA)] and a U/GU rich downstream element (DSE). The direct interactions of components of the CPA complex cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CSTF) at these sequence elements allow assembly, catalysis of cleavage of the pre-mRNA, and subsequent polyadenylation at the 3'end (Proudfoot et al., 2002). With more than half of human genes containing multiple poly(A) signals, alternative cleavage and polyadenylation (APA) plays a fundamental role in regulating transcriptome complexity and gene expression (Tian et al., 2005; Neve and Furger, 2014). Depending on where the different poly(A) signals are located in genes, APA events can be distinguished into 2 separate categories: the alternative poly(A) site usage results in transcripts that can differ in length and their coding potential, thus referred as coding region APA (CR-APA). The second type of transcripts generated by APA code for identical proteins and only differ in the length of their 3' untranslated regions (3'UTRs), which are alluded to as 3'UTR-APA. While CR-APA elicits a physiological impact from generating different proteins, 3'UTR-APA results in mRNA isoforms that encode the same protein but differ in the length of their 3'UTRs. These alternative UTRs may include or exclude cis-regulatory elements that may regulate gene expression at the level of mRNA stability, mRNA localization, and translation efficiency. Furthermore, differential 3'UTRs can not only affect protein output but they can

also have the capacity to regulate final protein destination (Berkovits and Mayr, 2015). Advances in next generation sequencing in the last decade has allowed the exploration of APA's regulatory potential in cells responding to specific cues including proliferation, differentiation, stress and disease progression (Neve et al., 2017; Tian and Manley, 2016). However to date, the underlying mechanisms that regulate APA in response to these cues and in particular what the physiological impact of these APA changes is, remains largely ill understood.

Recently, APA has been implicated in antiviral responses in herpes simplex virus 1 (HSV-1) and vesicular stomatitis virus (VSV) infection. In VSV infection of macrophages, genes exhibiting APA and differential expression were enriched for immune-related and signalling functions (Ji et al., 2017), whereas infection with HSV-1 mainly resulted in APA of genes involved in stress response and nuclear transport (Hu et al., 2016). Together, these studies indicate that APA could potentially be used to regulate the innate immune response to viral infections.

However, while the effects of influenza virus infection on CPA and nuclear export of mRNA has been widely studied, the impact on APA of host transcripts during and influenza infection have not been explored. Given that other viruses have been shown to affect host mRNA APA profiles during infection, combined with the knowledge that the flu NS1 protein inactivates a critical cleavage and polyadenylation factor (CPSF30), I hypothesised that influenza infection may cause widespread APA in host cells and this may affect the host viral interactions. To test this hypothesis, I infected human lung epithelial cells with influenza A/WSN/33 (H1N1) virus and determined the nuclear and cytoplasmic APA profiles. Nuclear and cytoplasmic profiles are studied to have a better resolution of the mechanisms that force these changes. To that end I address the following questions:

1. Are host nuclear and cytoplasmic APA profiles changing during influenza virus infection?
2. If influenza infection causes host gene APA, do these changes have any physiological consequence and to what extent do they change the host virus relationship?
3. If influenza infection causes host gene APA, to what extent are these events triggered by viral factors or are they simply part of a generic host response to stress?

4.2 Results:

4.2.1 Dynamics of Infection

To examine the global effects of influenza A virus infection on host alternative polyadenylation in human lung epithelial cells, A549 cells were infected at a high multiplicity of infection (MOI=5) with influenza A/WSN/33 (H1N1) virus. For all experiments described forthwith, I used at least two independent biological repeats and I strictly adhered to the protocol and quality control steps as outlined in figure 3.7 (Schematic workflow of established infection time course experiment).

To assess host mRNA APA dynamics during influenza infection, I infected A549 cells with ultracentrifuge purified viral stocks (free of contaminating factors from the amplification host cells) (Hutchinson et al., 2014). The virus was not aspirated and immune fluorescence using an antibody targeting the viral nucleoprotein (NP) was used to monitor the degree of infection in all experiments for the all time points and for each individual experiment (Figure 4.1). NP was expressed in all the time points post 6h

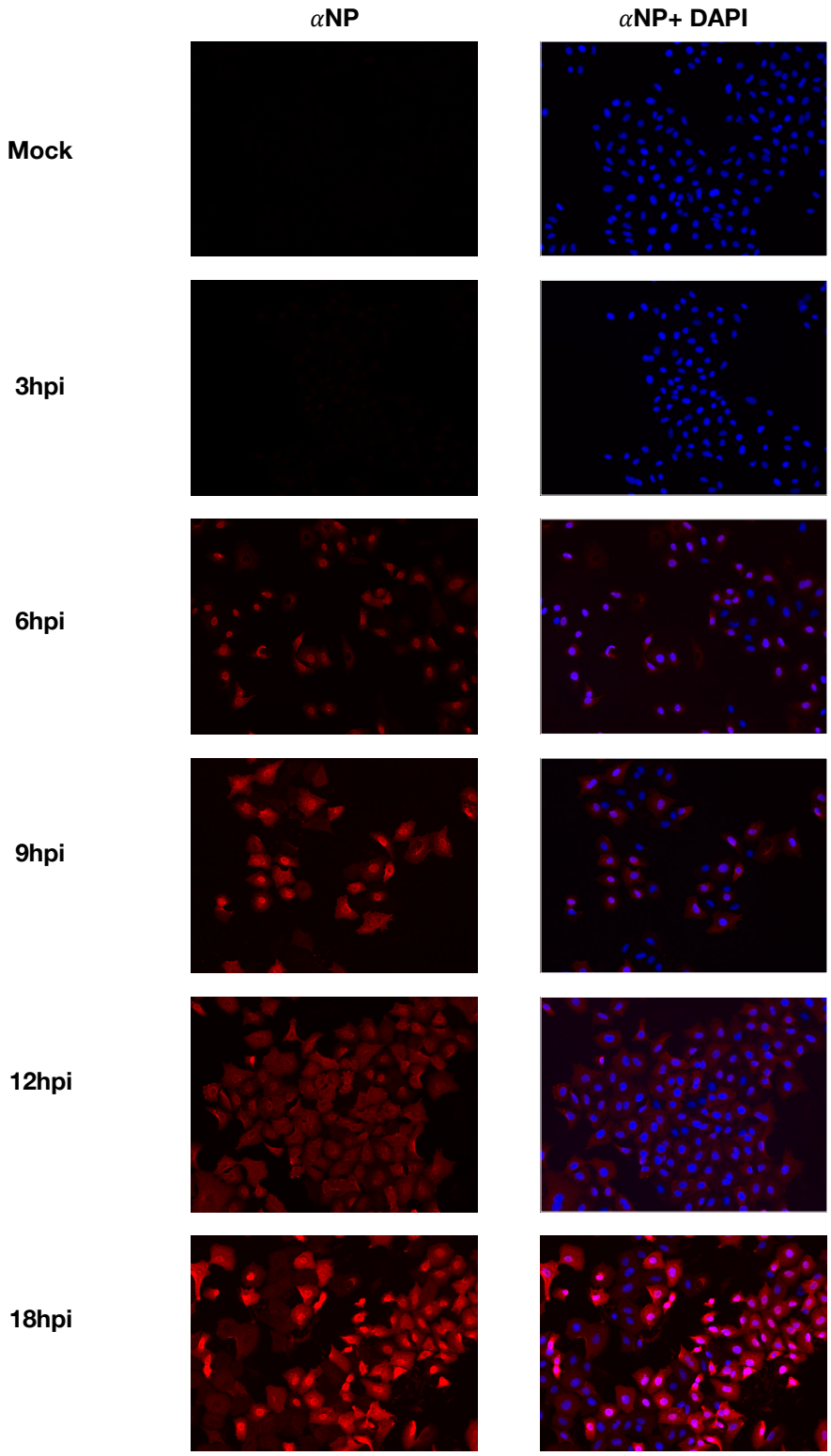


Figure 4.1 Immunofluorescence Microscopy Staining for NP Performed After Every Infection Time Course Experiment Ensuring Efficient Infection

Human cells were infected with influenza A/WSN/33 (H1N1) virus at an MOI of 5, viral media was not removed after 1 hour RTP incubation before incubation at the indicated infection time points at 37°C. Infection efficiency of every indicated time point were determined after every infection time course experiment by immunofluorescence microscopy staining for viral NP (red) and cell nuclei (blue). This serves as a control to show that the majority of the cells were infected. This is an immunofluorescence microscopy experiment done in tandem with the first biological repeat of the infection time course.

in more that 75% of cells (Figure 4.1 see time points 6h, 9h, 12h and 18h) thus reaching the threshold for further analysis.

Subsequently, nuclear and cytoplasmic RNA were extracted as described (Neve et al., 2016) from 3h, 6h, 9h, 12h,18h post infection and mock. For each individual replicate, the RNA isolates were subjected to vigorous quality control, assessing both the fractionation purity by monitoring MALAT1 and GAPDH by RT-PCR and RNA integrity by visual inspection of the cellular ribosomal RNAs in the nuclear and cytoplasmic fractions. As can be seen in Figure 4.2A, the strictly nuclear RNA MALAT1 was only detected in the nuclear fractions and was absent in the cytoplasmic fractions at all time points indicating that the fractions were indeed enriched in the respective RNAs. The “purity” of the fractions was further highlighted by the restriction of the ribosomal precursors RNAs to the nuclear fractions (see 45S and 32S lanes in Figure 4.2B). In addition, rRNA bands were present at all time points indicating that the integrity of the RNA was not significantly compromised during the isolation procedure.

Furthermore, for each replicate, a small sample of cells representing each time point was collected prior to the fractionation procedure for whole cell protein isolation to monitor viral protein expression dynamics during the infection (see chapter 3 (NS1 expression) and below Figure 4.3).

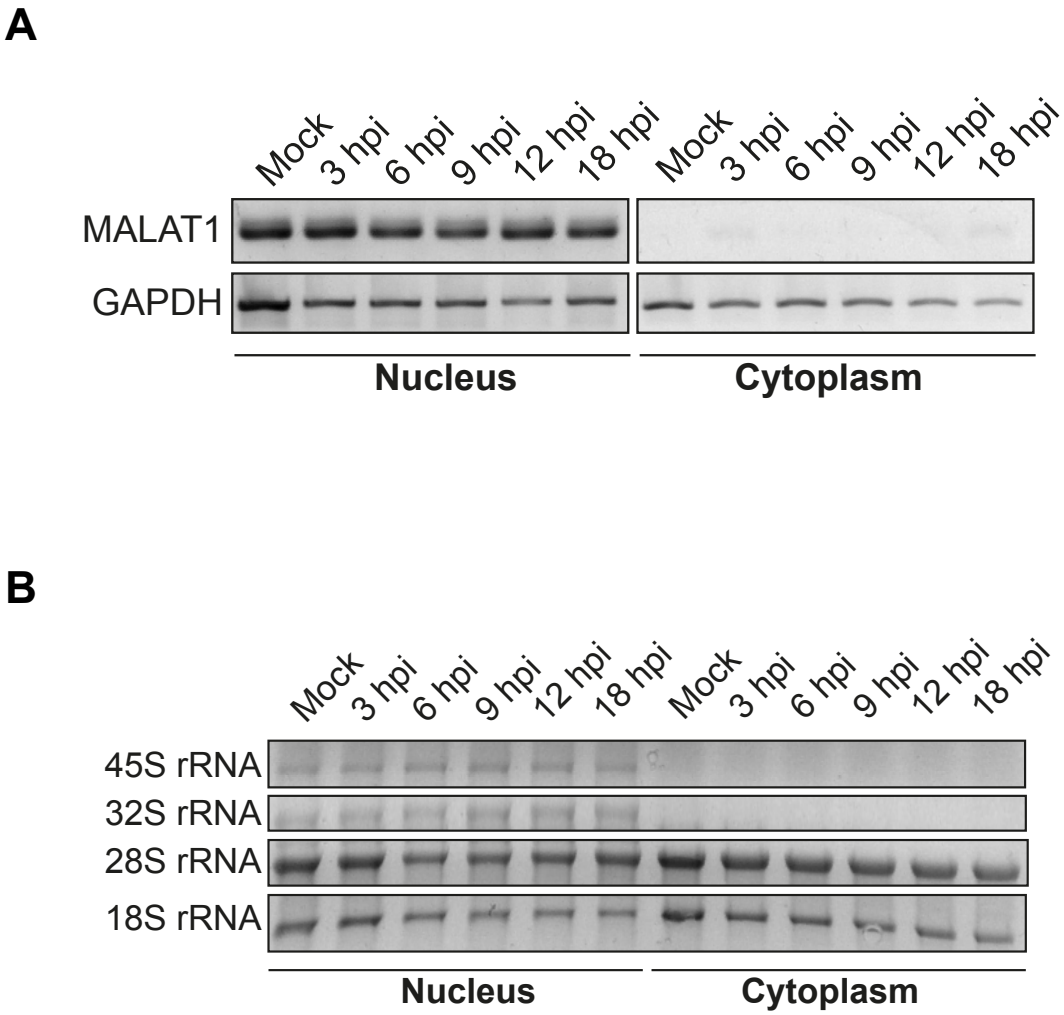


Figure 4.2 Infection Time-Course Subcellular Fractionation Control.

A. Reverse Transcription PCR (RT-PCR) with primers targeting nuclear retained MALAT1 (upper panel) and loading control GAPDH (lower panel). 1µg of RNA from each subcellular fraction was used. Only nuclear fraction (top left panel) amplified for MALAT1.

B. Denaturing RNA gel loaded with 2µg of RNA from each subcellular RNA fraction displaying the sole presence of precursor rRNA in the nucleus. The identity of each ribosomal RNA band is shown on the left hand side of the gel.

After establishing successful infection of the majority of cells and confirming the integrity and purity of the nuclear and cytoplasmic RNA fractions, I next focused to characterise the infection further by monitoring both viral RNA and viral protein expression at the different time points. To that end, I used a western blot approach as describe in chapter three to visualise the expression of critical proteins such as NS1, M1, NP and PB2 during the chosen infection times. As can be seen in figure 4.3A, viral proteins and importantly NS1 began to accumulate at 6 hpi. Whilst NS1, NP and PB2 peak at 9 hpi, M1 expression was continuously increasing. The observed 6hpi accumulation of viral proteins was in agreement with previous studies that have shown a strong preference in viral transcripts translation after 6 hours of infection (Yángüez et al., 2011) and similar to the previous chapter for NS1 expression. The expression of the loading control heat shock protein 60 (HSP60) remained constant. Since, viral protein expression dynamics were established, the time points including and after 6 hpi are reasonable to address how influenza infection may affect host APA.

For RNA analysis, the isolates of the different fractions were subsequently used to create RNAseq libraries using the QuantSeq library kit from Lexogen. Briefly, an oligo dT primer was used to reverse transcribe polyadenylated RNA and second strand synthesis was performed using barcoded random primers which results in the dsDNA that are enriched in sequences close to the polyA site. The DNA fragments were subsequently purified and the library was PCR amplified to be compatible for sequencing on the IonProton platform. The amplified libraries were subsequently sequenced and host reads were mapped against the hg19 genome using the inbuilt Torrent Mapping alignment Software (TMAP). The resulting BAM files were then processed as described below. The viral reads were mapped to the WSN genome using the Galaxy (https://useglaxy.org/) platform as described in the Material and Methods. For the subsequent bioinformatics analysis the BAM files from each

time point and for each fraction were pooled. An initial comparison of the pooled host and viral mapped reads of infected cells revealed that influenza viruses generated significant amounts of viral adenylated mRNAs. Similarly to protein, viral adenylated mRNAs peak at 6h to 9h post infection and at these time points up to 29% of all nuclear polyadenylated mRNAs are of viral origin (Figure 4.3B). Interestingly, the relative amount of adenylated viral mRNA in the cytoplasm was much lower compared to the nuclear peaking at 6h and representing around 18% of total reads. Intriguingly, the proportion of viral cytoplasmic reads gradually decreases after the 6h time point to less than 5% at 18hpi (Figure 4.3B).

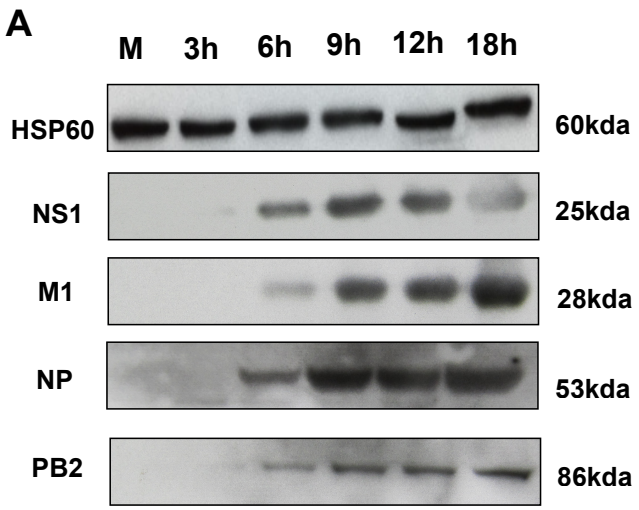
To determine the dynamics of the 8 individual viral mRNA segments, we observed the relative proportion of reads mapped to the individual A/WSN/33 mRNA segments encoding the PB2, PB1, PA, HA, NP, NA, M and NS proteins at each individual time point (analysis done with the help of Harry Fischl, Department of Biochemistry, University of Oxford). As can be seen in figure 4.3C, the level of expression of the individual segments in the nucleus was very different but the overall trend was as observed with the total mRNA in that levels sharply increased from 6hpi and remained high throughout the 18h period. Similarly, in the cytoplasmic fraction, all 8 mRNAs appeared to peak at 6hpi and then decay gradually thereafter as has been observed looking at total mRNA reads (Figure 4.3B). Interestingly, in both the nuclear and cytoplasmic fractions, the polymerase segments PB2, PB1, and PA were transcribed at significantly lower levels compared to the remaining segments (Figure 4.3 C,D). However, it is important to note that the number of reads per segment does not give any information about gene expression of alternatively spliced segment 1 (PB1), 7(M) and 8(NS) transcripts.

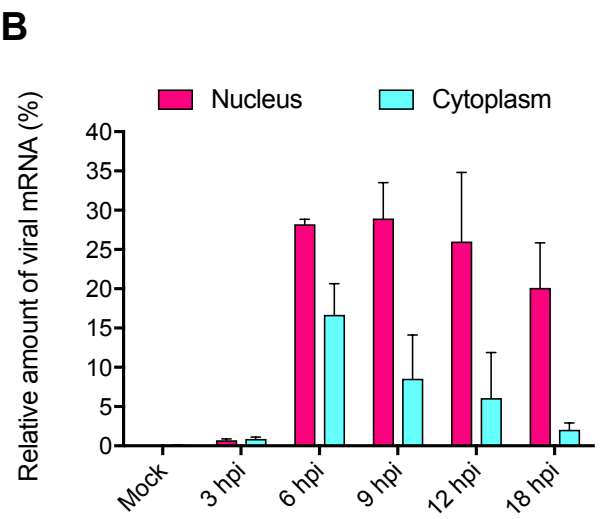
Collectively this analysis showed that viral mRNA expression peaks at 6hpi and whilst the expression levels remained high in the nucleus, the mRNA levels for all segments relative

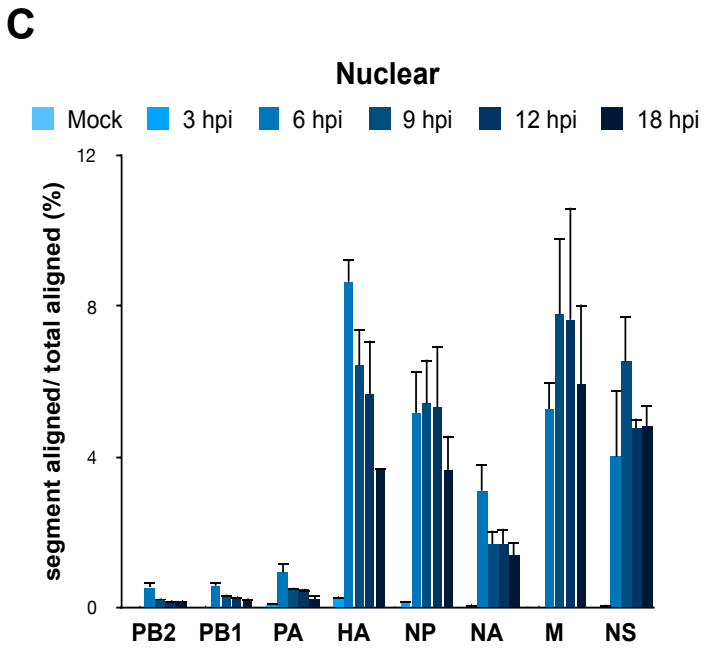
to the total host and viral mRNA reads in the cytoplasm decrease significantly after 6 hpi, and constituted less than 5% of total mRNA reads at 18hpi. This difference in dynamics between the nuclear viral mRNA and their cytoplasmic counterparts may reflect a constraint of the nuclear cytoplasmic export at later infection time points.

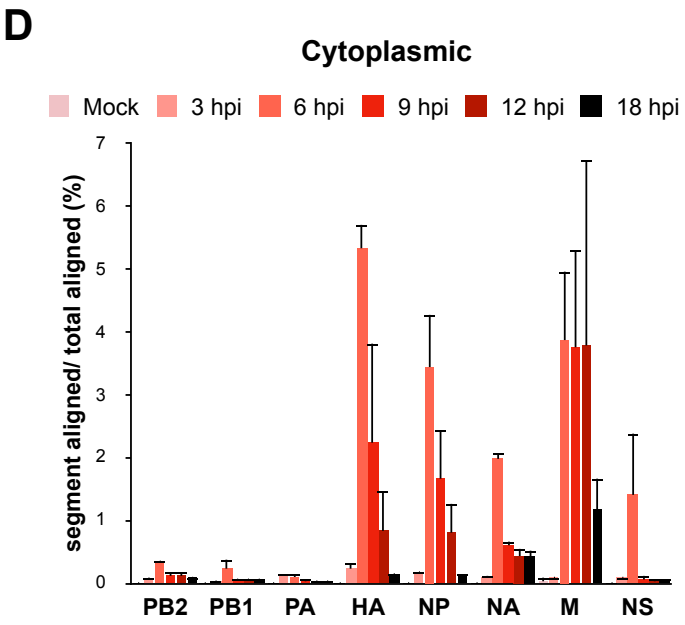
The bioinformatics data was then verified by primer extension analysis (performed by Ben Nilsson, Sir William Dunn School of Pathology, University of Oxford) of the accumulation of segment 6 viral RNA (encodes Neuraminadase (NA)) in the nucleus and cytoplasm, revealing the same trend for segment 6 mRNA as for the total and individual segment viral mRNA population (Figure 4.3E, F). Quantification of the band intensities was done on ImageJ (Figure 4.3F). Very low levels of mRNA, cRNA and vRNA were detected at 3 hpi in both nuclear and cytoplasmic fractions, while a rapid increase in all viral RNA species was observed from 6 hpi. In the nucleus, both vRNA and cRNA remained stable after 9 hpi, while a significant decrease in mRNA was observed later in infection, especially when observing the cytoplasmic samples. Interestingly though, at 6 hpi significant amounts of cRNA, which is not thought to be exported from the nucleus, are detected in the cytoplasm. Therefore, the primer extension largely confirmed our observation regarding the expression dynamics of the viral mRNAs described above and the dynamics of the v and c viral RNA was similar to the reduction in viral mRNA levels observed previously (Shapiro et al., 1987; Robb et al., 2009; Vreede et al., 2011).

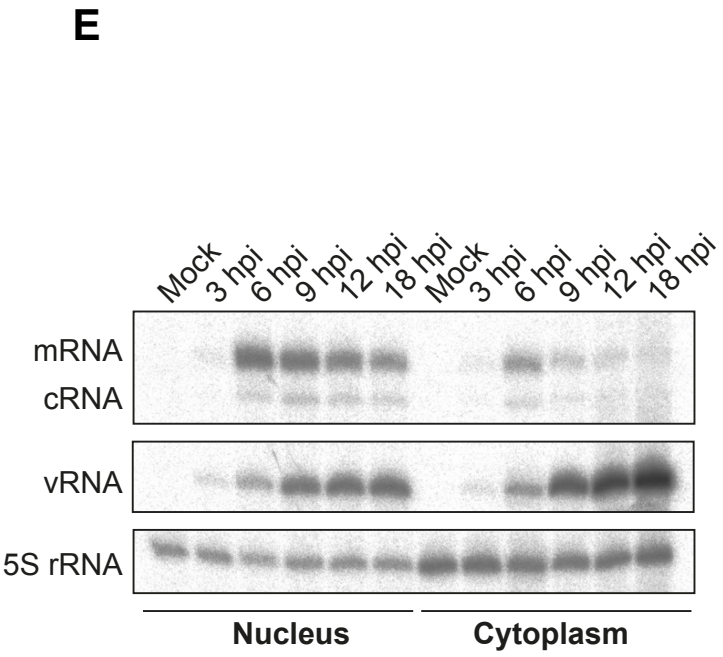
Together, these results demonstrated that the majority of host cells are infected at the 6hpi time point and that the 6h-18h period can capture early to late infection events. In addition, we found a rather intriguing discrepancy between nuclear and cytoplasmic viral mRNA levels at 9h and later time points potentially indicating that late in infection, nuclear cytoplasmic export of viral mRNAs may be compromised.


A
M
3h
6h
9h
12h
18h
HSP60
60kda
NS1
25kda
M1
28kda
NP
53kda
PB2
86kda


B
Nucleus
Cytoplasm
Relative amount of viral mRNA (%)
40
35
30
25
20
15
10
5
0
Mock
3 hpi
6 hpi
9 hpi
12 hpi
18 hpi


C
Nuclear
Mock
3 hpi
6 hpi
9 hpi
12 hpi
18 hpi
segment aligned/ total aligned (%)
12
8
4
0
PB2
PB1
PA
HA
NP
NA
M
NS


D
Cytoplasmic
Mock
3 hpi
6 hpi
9 hpi
12 hpi
18 hpi
segment aligned/ total aligned (%)
7
6
5
4
3
2
1
0
PB2
PB1
PA
HA
NP
NA
M
NS


E
Mock
3 hpi
6 hpi
9 hpi
12 hpi
18 hpi
Mock
3 hpi
6 hpi
9 hpi
12 hpi
18 hpi
mRNA
cRNA
vRNA
5S rRNA
Nucleus
Cytoplasm

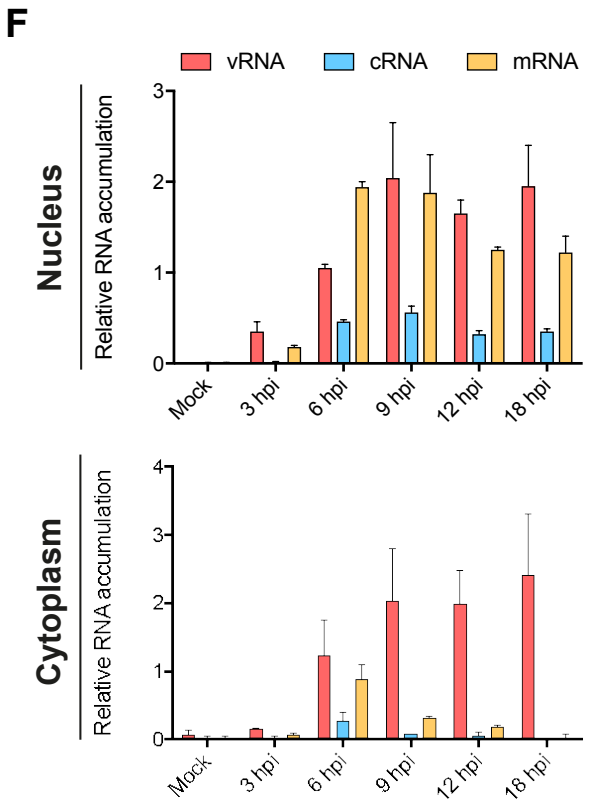

F
vRNA
cRNA
mRNA
Nucleus
Relative RNA accumulation
3
2
1
0
Mock
3 hpi
6 hpi
9 hpi
12 hpi
18 hpi
Cytoplasm
Relative RNA accumulation
4
3
2
1
0
Mock
3 hpi
6 hpi
9 hpi
12 hpi
18 hpi

Figure 4.3 Dynamics of Infection (previous page)

A. Human A549 Lung cells were infected with A/WSN/33 virus at MOI 5. Whole cell lysates containing protein were extracted at the displayed time points (3, 6, 9, 12, 18 hours post infection), and mock (non-infected). Viral protein expression dynamics were analysed by western blotting using antibodies against viral proteins including NS1, NP, PB2 and M1. All viral proteins began to accumulate at 6 hpi ,whilst NS1, NP and PB2 peak at 9 hpi, M1 expression was continuously increasing. HSP60 loading control remained constant.

B. Bioinformatics analysis done by mapping sequencing reads to A/WSN/33 genome by Galaxy. Samtools FlagStat generated number of reads aligned to viral genome. Relative amount of viral mRNA (%) was calculated by dividing number of reads aligned to viral WSN/33 genome over total (viral and human (hg19)) reads per time point. Nuclear (pink) and cytoplasmic (blue) fractions represented at indicated time points. The graph shows the mean accumulation of viral RNA from 2 independent biological replicates with error bars representing the standard error of the mean (SEM).

C. Nuclear reads mapped to individual WSN/33 viral gene segments (listed in the order above on the x-axis) at the indicated time points post-infection. Colour scheme is described as follows: As infection goes from mock to 18 hpi later in infection, light blue turns to darker blue. The graph shows the mean relative amount of individual segment RNA from 2 biological replicates with error bars representing SEM.

D. Similar to **C**., cytoplasmic reads were mapped to individual WSN/33 viral gene segments at indicated time points post-infection. Colour red goes darker as infection reaches later time points from mock to 18 hpi. The graph shows the mean relative amount of individual segment RNA from 2 biological replicates with error bars representing SEM.

E. The accumulation of viral NA encoded by segment 6 RNA (mRNA, cRNA, vRNA) was analysed by primer extension and 6% PAGE in nuclear and cytoplasmic fractions of each individual time points. 5S rRNA was used as an internal loading control.

F. The bar graph is a quantitative representative of the primer extension shown in E. Quantification of band intensities were done on ImageJ over two biological repeats of primer extension showing mean accumulation of viral RNA with error bars representing SEM. Top panel (nuclear) and Bottom panel (cytoplasmic).

4.2.2 Influenza Viral Infection Induces 3'UTR APA and Differential Gene Expression of Host Transcripts

After confirmation of the quality control of the fractionation, RNA integrity and infection for each time point and subcellular fraction I next focused on our core questions and established the APA profiles of the host mRNAs.

The APA profiles of nuclear and cytoplasmic RNA fractions from the two biological replicates of the infection time course experiments were subjected to an APA analysis. As described above, the BAM reads from each time point and subcellular fractions of the two independent biological replicates were combined and subsequently processed with an established bioinformatics pipeline (Hoque et al., 2013; Neve et al., 2016). This pipeline extracts statistically significant UTR- and CR-APA events. Additionally, it also identifies statistically significant gene expression changes of all genes including each APA mRNA

isoform based on a Fisher's exact test, P<0.01 (for more details see material and methods).

4.2.2a Significant APA Changes Induced During Influenza Infection

An established bioinformatics pipeline (Hoque et al., 2013; Neve et al., 2016) was utilised to assess APA changes during A/WSN/33 infection. As stated earlier, for this analysis the reads of two biological repeats were combined, and significantly changing APA events (p<0.01 Fisher's exact Test, with a minimum 5% change in APA isoform abundance) in the nucleus and the cytoplasm were identified. To analyse 3'UTR APA, A/WSN/33 infected time points (3-18 hpi) were compared to mock (non-infected) samples between nuclear samples and alternatively between cytoplasmic samples. When multiple CPA sites were present, the two most abundant CPA sites were selected for analysis in the pipeline. The significant relative decrease in usage of the proximal CPA site is represented as a lengthening event. Conversely, the significant increase in usage of the proximal CPA site is characterised as a shortening event.

The statistical analysis of the relative usage of the two most used 3'UTR polyadenylation sites for each gene identified around 7000 number of polyadenylation events in the nuclear and approximately 6000 in the cytoplasmic fractions for each time point (Supplementary Figure 2; see scatter plots grey dots). The comparison of the infected to non-infected nuclear fractions revealed approximately 10% or 700 individual APA events where alternatively polyadenylated transcripts showed a statistically significant change in relative usage between the proximal and distal sites (Figure 4.4A). For all these comparisons both shortening and lengthening events were extracted. However, twice as many lengthening compared to shortening events occur in the nucleus for all the time points past 3h of infection, suggesting that influenza infection causes lengthening in around 500 genes by

suppression of the proximal polyA sites (Figure 4.4). Whilst a tendency towards lengthening in infected cells was evident in the nuclear fraction, there was no clear trend towards shortening or lengthening in the cytoplasm. Furthermore, whilst in the nuclear fractions there appeared to have a clear increase in statistically significant APA events after 6hpi compared to the 3h time point, in the cytoplasm similar numbers of APA events occur at the earlier 3h time point. As I have shown in chapter 3 and above, no significant

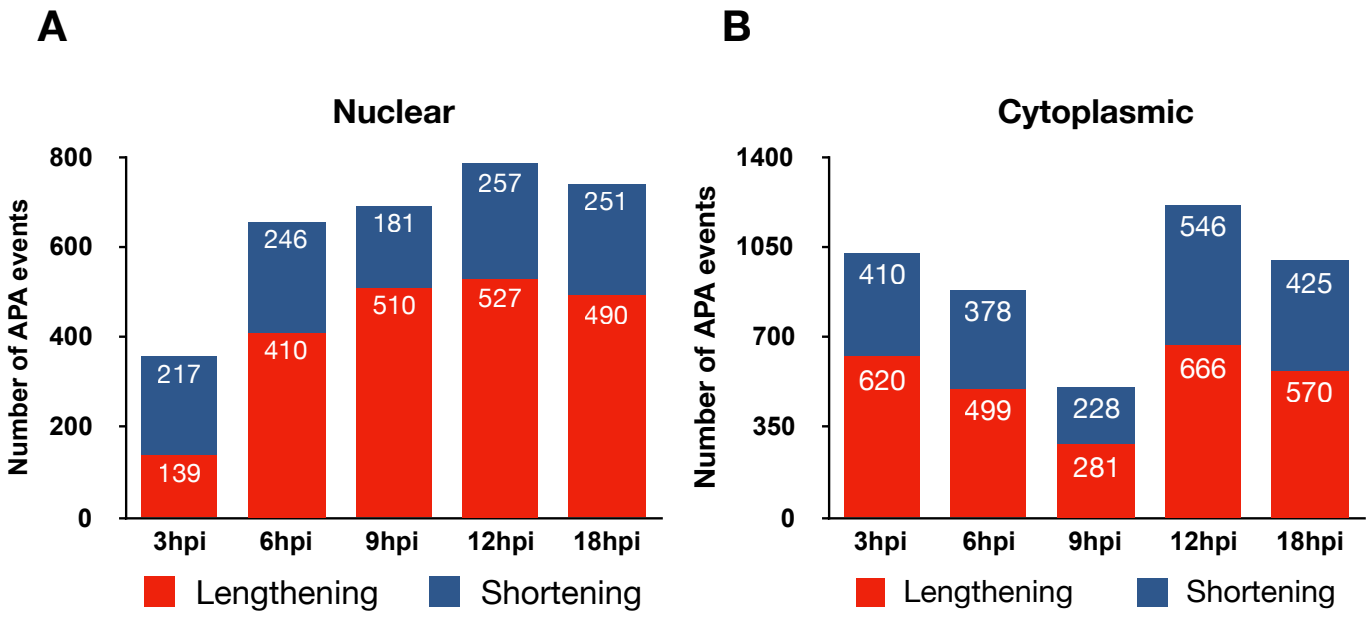


Figure 4.4 Global Analysis of WSN/33 infection 3'UTR APA Events
Bar graph to visualise number of significant 3'UTR APA events (Fisher's exact Test, $p<0.01$, with a minimum 5% change in APA isoform abundance) after WSN/33 infection in both nuclear over 2 independent biological repeats and cytoplasmic fractions over 2 biological repeats of sequencing. The number of lengthening events are presented in red, whereas the number of shortening events are presented in blue. The total detectable genes in the nucleus (Panel A) are around 7000. The total detectable genes in cytoplasmic fractions (Panel B) are around 6000 (Supplementary figure 2, grey dots on scatterplots). % of total 3'UTR APA events were calculated by dividing total number of significant APA events over total genes detected for each fraction. Proportion (%) of APA events exhibiting lengthening or shortening were also calculated.

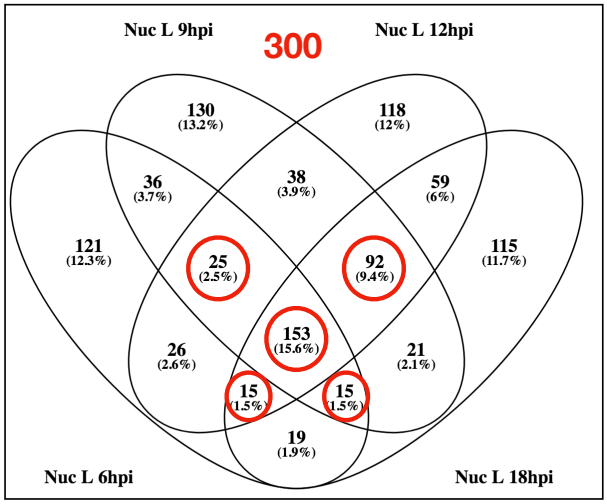
viral protein expression is detectable at this time point suggesting that the cytoplasmic changes at the 3h time point are unlikely to be mediated by viral proteins and may be due to a stress response caused by the infection.

Despite extracting statistically significant APA changes when comparing RNA from infected and non-infected cells as above, this analysis however does not indicate whether the

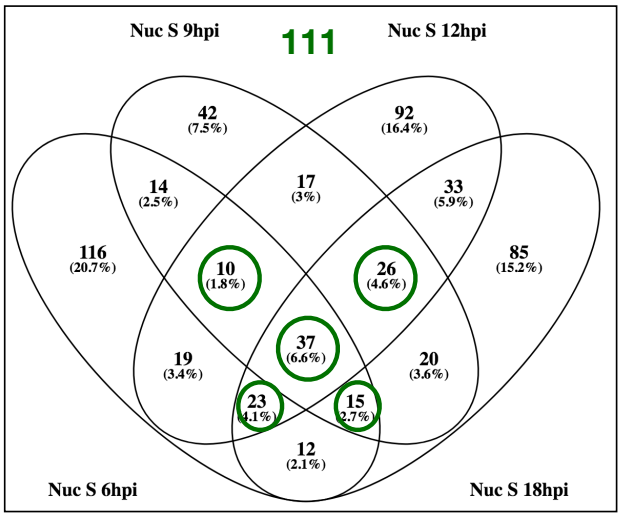
detected APA events were specific for each time point or if they were the same events that persist throughout the infection period. I first aimed to extract APA events that may be triggered by viral proteins and can be detected at multiple infection time points, especially from 6-18 hpi as it was when viral proteins were previously shown to accumulate. To that end, I implemented a stringent selection criterion by focusing on genes that undergo statistically significant APA in at least three of the sampled time points between 6-18 hours post infection (hpi). I extracted these overlapping genes for each for the four possible categories: Nuclear Lengthening, Nuclear Shortening, Cytoplasmic Lengthening and Cytoplasmic Shortening (Figure 4.5).

As can be seen in figure 4.5, 300 genes or around 30% of all the nuclear lengthening events occur in at least three times between the 6h and 18h post infection time points (Figure 4.5 A). In the case of the nuclear shortening cohort, 111 genes or around 20% of events are reoccurring in at least three time points (Figure 4.5 B). 296 genes (~30%) of genes undergo lengthening or 235 genes (~30%) undergo shortening in the cytoplasm do overlap and occur in at least three different time points (Figure 4.5 C and D). From this I conclude that a significant number of genes undergo APA upon influenza infection, and once occurred these APA events would persist throughout the observed infection period (6h-18hpi).

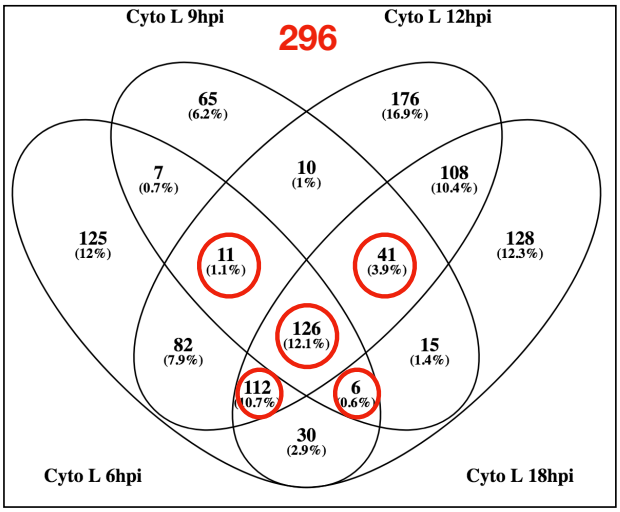
A



B



C



D

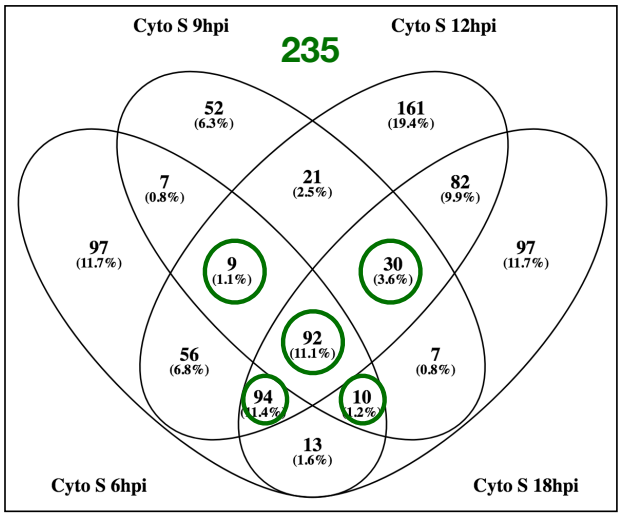


Figure 4.5 Isolation of Genes that Exhibit 3'UTR APA in a Minimum of Three of Time Points

Venn diagrams were generated between 4 biologically relevant time points (6-18hpi) of genes that exhibit significant APA ($p<0.01$ Fisher's exact Test, with a minimum 5% change in APA isoform abundance). Specifically, the circled overlapped section of the diagrams represent significant APA genes in these possible combinations of time points: (6, 9, 12 hpi; 9,12,18hpi; 6,9,18hpi; 9, 12, 18hpi; 6-18 hpi). 4 categories have been identified: A) Nuclear Lengthening: 300 genes B) Nuclear Shortening: 111 genes C) Cytoplasmic Lengthening: 296 genes D) Cytoplasmic Shortening: 235 genes.

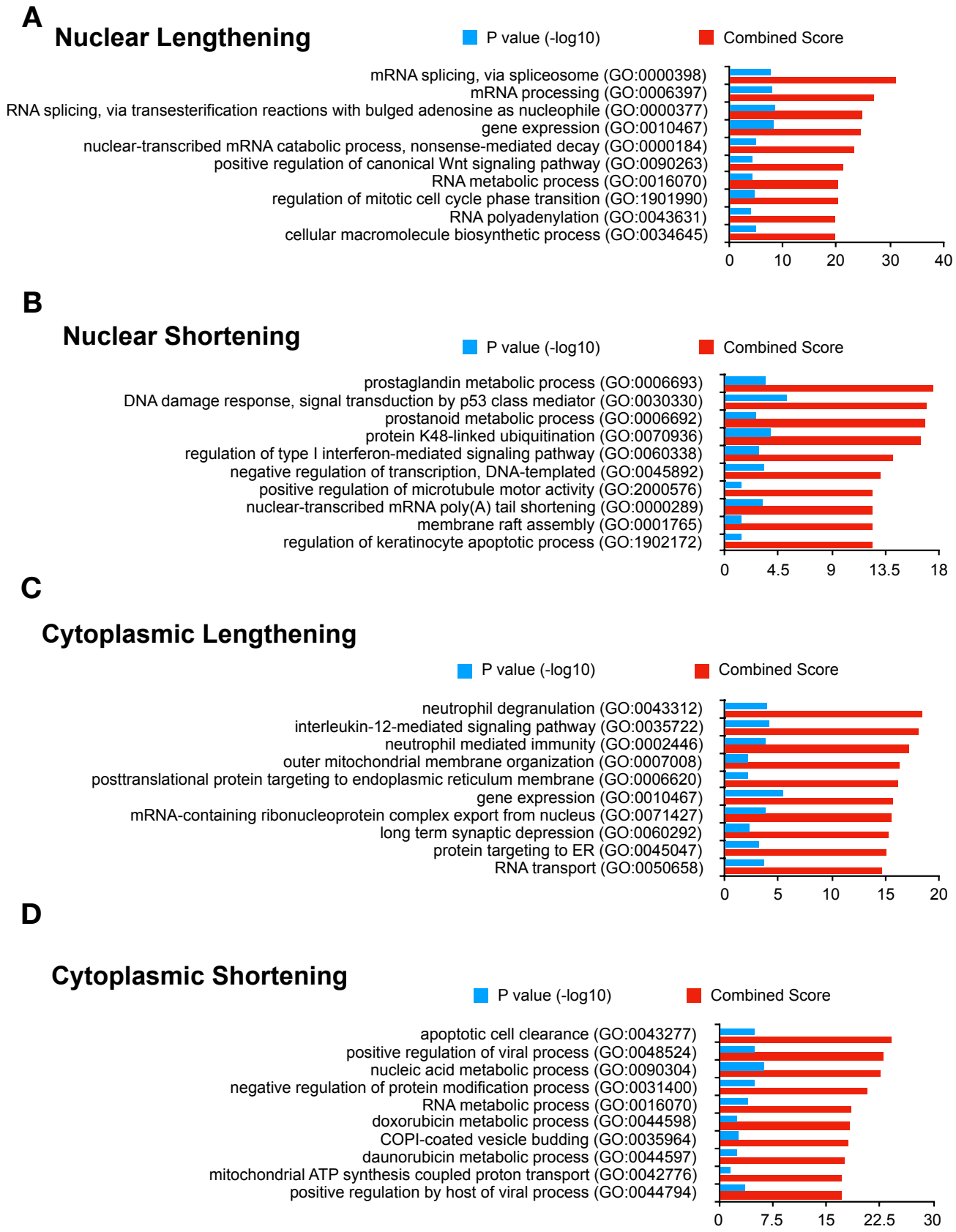


Figure 4.6 Gene Ontology Analysis of Significant 3'UTR APA genes in 4 Different Categories

The ten most enriched GO terms for biological processes and their statistical significance ($p<0.05$) of the genes were associated with A) nuclear lengthening B) nuclear shortening C) cytoplasmic lengthening D) cytoplasmic shortening genes. Gene ontology analysis was performed using Enrichr.

To address whether these events could have any physiological relevance for the infection process, the virus or the host, I performed a Gene Ontology (GO) analysis using the Enrichr (http://amp.pharm.mssm.edu/Enrichr/) platform and focused on the ontology “biological process”. As can be seen in figure 4.6A, the genes that undergo lengthening in the nucleus upon influenza infection are associated with pre-mRNA processing processes including polyadenylation. In contrast, nuclear shortening of UTRs occurs in genes that are linked to innate immune response and negative regulation of transcription (Figure 4.6 B). Cytoplasmic lengthening genes were significantly enriched for immune response and regulation of gene expression (Figure 4.6 C). Lastly, cytoplasmic shortening cluster were significantly associated with RNA metabolic processes and positive regulation of viral process (Figure 4.6 D). Altogether, I identified a significant number of genes that are subjected to APA during influenza infection and these genes have been linked with processes that are highly relevant for host defence mechanisms and viral propagation.

4.2.2b Relating Significant APA Changes to Changes in Differential Gene Expression

In the previous section I identified genes that are undergoing APA during an influenza infection period. I have also shown that the affected genes belong to categories that can be linked to both host defence and viral propagation. Whilst I have detected APA events in these potentially highly relevant genes, I next needed to establish whether the respective shortening or lengthening events do have a physiological impact. One way by which APA can have a physiological impact is when the lengthening or shortening results in mRNA isoforms that differ in their stability (reviewed in Neve et al., 2017). This is of particular importance for APA events that penetrate into the cytoplasm. Thus I wanted to extract genes that exhibited APA in at least three infection time points and are also either up (UP)-

or down regulated (DN) in at least three time points from 6h-18 hpi versus non infected cells according to a Fisher's exact Test, $p<0.01$. The Fisher exact test revealed that there are 1934 genes that are significantly down-regulated and 1450 genes that are significantly up-regulated . By overlapping the APA cohort and differential gene expression cohort, 8 individual categories depending on whether they are lengthening or shortening and are up or down regulated can be formed: 1: DN nuclear lengthening, 2: DN nuclear shortening, 3: DN cytoplasmic lengthening, 4: DN cytoplasmic shortening (Figure 4.7A); 5: UP nuclear lengthening, 6: UP nuclear shortening, 7: UP cytoplasmic lengthening, 8: UP cytoplasmic shortening (Figure 4.7B)). Importantly, no clear trend between decreased or increased expression can be made in the cohort 1-6 and 8 (Figure 4.7C). There appears to be a tendency for genes that have undergone lengthening in the cytoplasm to be expressed at an elevated level. However, this analysis shows almost 50% of the genes undergoing APA upon viral infection change their expression levels. This is rather surprising given that only around 10% of APA isoforms have exhibited differential expression levels in resting HEK293 cells (Neve et al. 2016) and indicates that viral induced APA can have a significant impact on host mRNA levels. One of the major challenges in the APA field is to determine whether the observed changing APA profiles do have a physiological consequence. The complexity of the results described above suggests that to evaluate whether the identified APA events during influenza viral infection have a physiological impact, individual genes that undergo viral induced APA need to be selected and investigated.

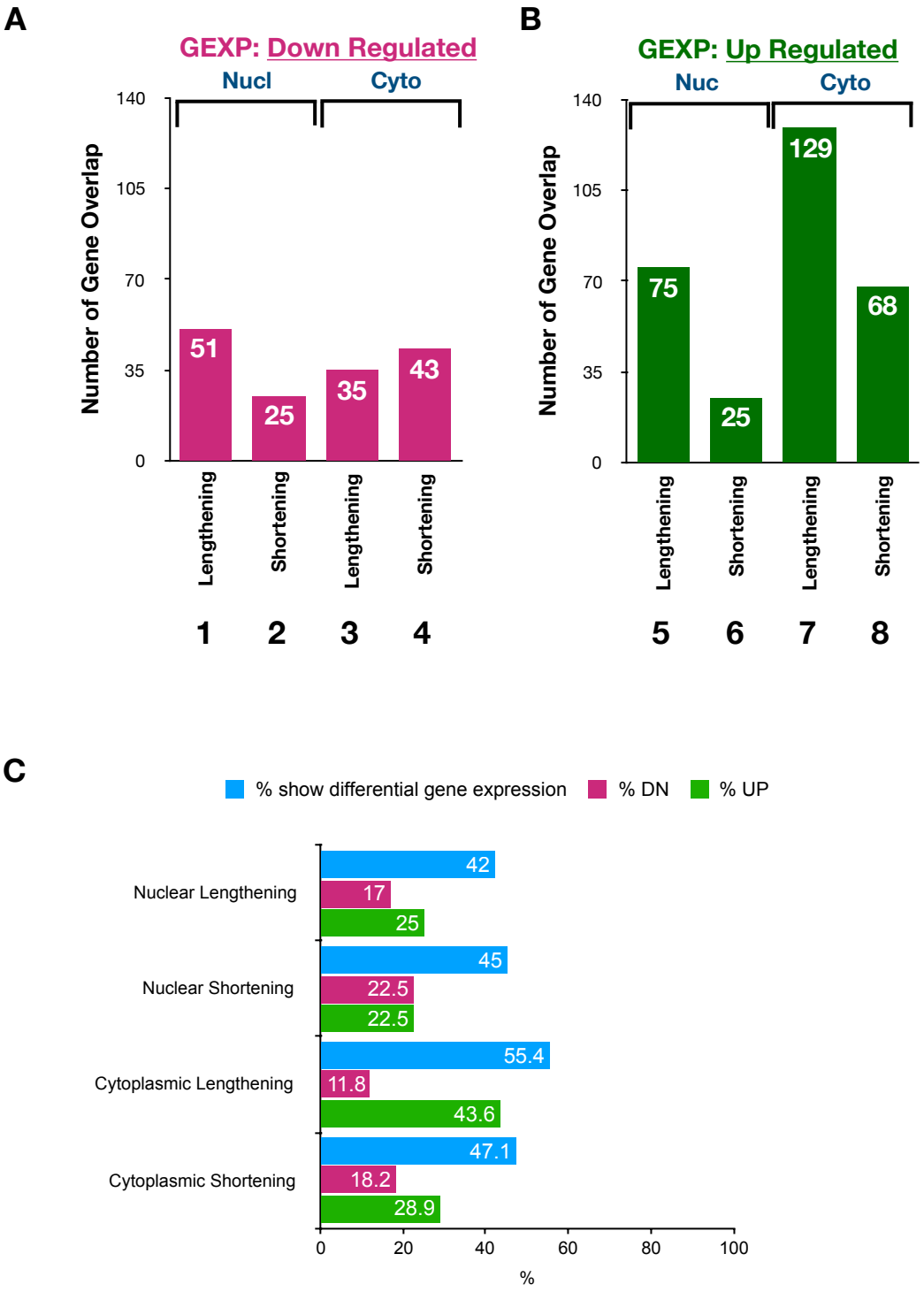


Figure 4.7 Differentially Expressed Genes Overlapped with 3'UTR APA Genes Significant at a Minimum 3 Time Points

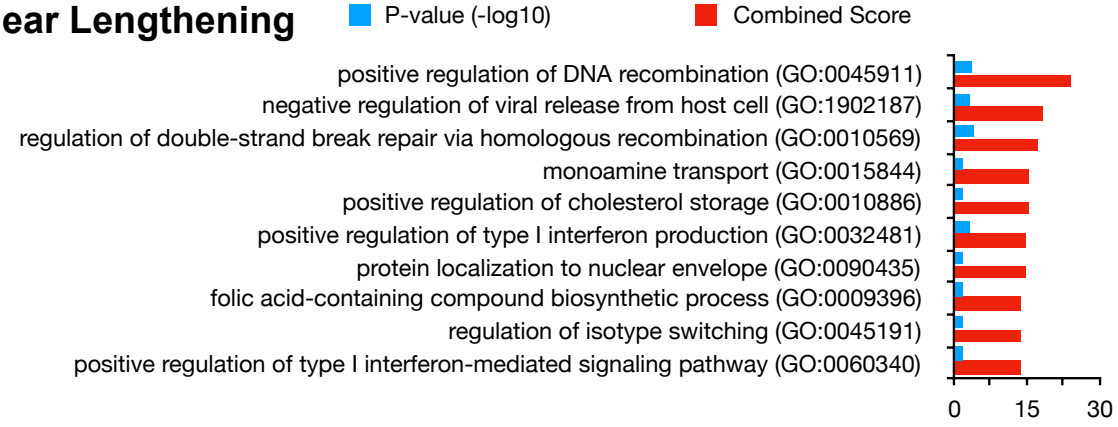
A) Down regulated genes (total of 1934 genes) were overlapped with 4 main selected APA categories in both nucleus and cytoplasm. The bar graph represents number of gene overlaps of Gene clusters 1: DN nuclear lengthening, 2: DN nuclear shortening, 3: DN cytoplasmic lengthening, 4: DN cytoplasmic shortening.

B) Up regulated genes (total of 1450 genes) were overlapped with 4 main selected APA categories in both nucleus and cytoplasm. The bar graph represents number of gene overlaps of Gene clusters 5: UP nuclear lengthening, 6: UP nuclear shortening, 7: UP cytoplasmic lengthening, 8: UP cytoplasmic shortening

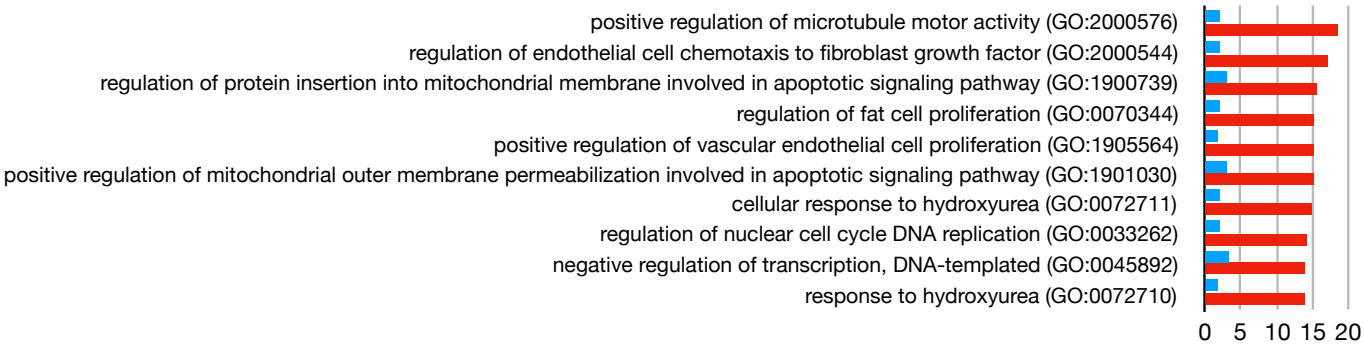
C) Proportion of APA events exhibited differential gene expression summarised in blue (%). The purple bar represents the % of the total APA events, what % are DN and what % are UP (green bar).

Category: **Gexp: DN**

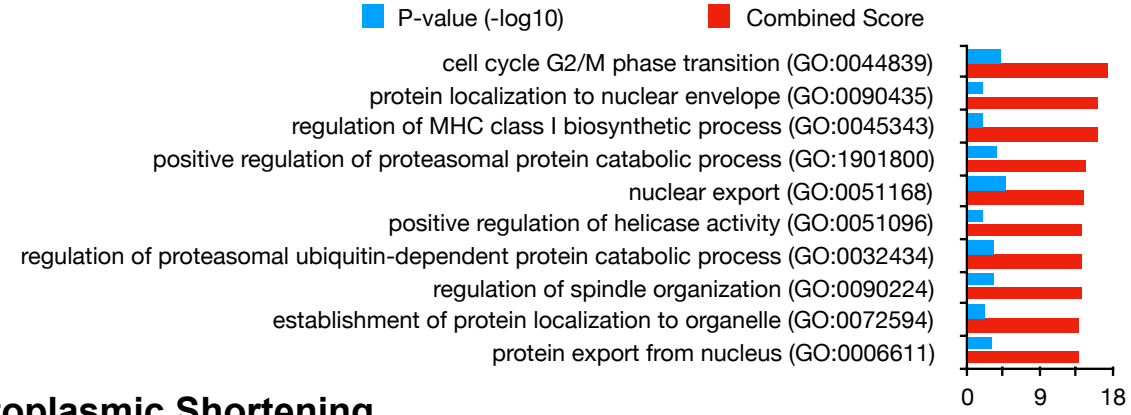
1 Nuclear Lengthening



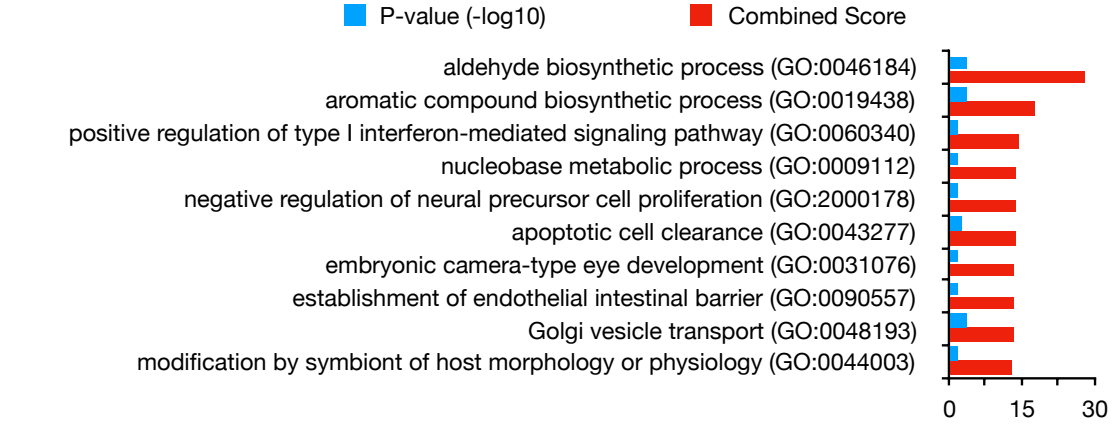
2 Nuclear Shortening



3 Cytoplasmic Lengthening



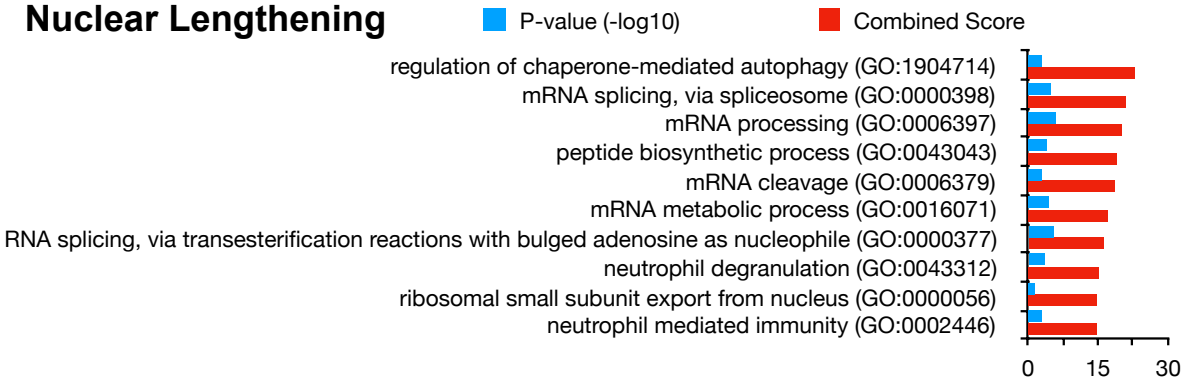
4 Cytoplasmic Shortening



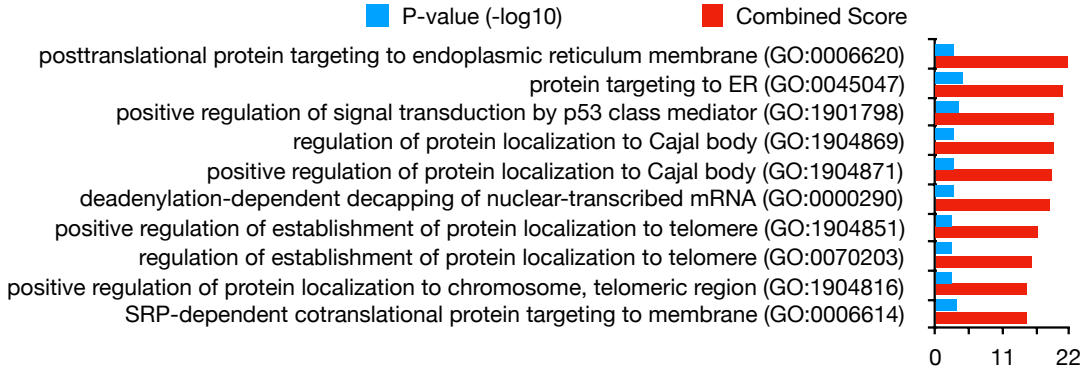
Category:

Gexp: UP

5 Nuclear Lengthening



7 Cytoplasmic Lengthening



8 Cytoplasmic Shortening

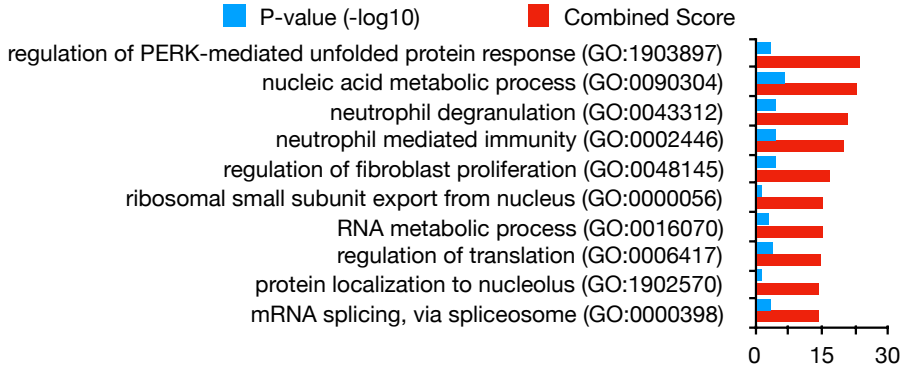


Figure 4.8 Gene Ontology Analysis of Overlapped Genes in APA cohort and Differential Gene Expression Cohort (over 2 pages)

The ten most enriched GO terms for biological processes and their statistical significance ($p<0.05$) of the genes were associated with gene cohorts 1: DN nuclear lengthening, 2: DN nuclear shortening, 3: DN cytoplasmic lengthening, 4: DN cytoplasmic shortening (Figure 4.7A); 5: UP nuclear lengthening, 7: UP cytoplasmic lengthening, 8: UP cytoplasmic shortening

6: UP nuclear shortening cohort showed no significant enrichment for any categories.

Table 4.1 Physiologically Relevant Genes Selected From GO Analysis

Genes	Categories	Gexp	APA	Fraction
TRIM8, TRIM11	Negative regulation of viral release from host cell	DN	Lengthening	Nucear
MAVS, CRCP, POLR2E	Positive Type 1 Interferon production regulation			
NutF2	Protein localisation to nuclear envelope	DN	Lengthening	Cytoplasm
NUTF2, NUP2, EIF4E	Nuclear export			
MAVS	Positive Type 1 interferon production regulation	DN	Shortening	Nuclear
MNT, CBX6,SHC1, WWC1, E2F1, ZNF217	Negative regulation of transcription			
EEFIA1, PSMA5, HSP90AA1, DDX3X, PSMC2, PRDX6	Neutrophil mediated immunity	UP	Lengthening	Nuclear
DDX5, PCF11, WDR3	mRNA splicing, mRNA processing, metabolism, cleavage			
HEXIM1, DDX5, RPL23	Positive regulation of signal transduction by p53 class mediator	UP	Lengthening	Cytoplasm
RPL23, RPL13, RPL35A, RPL15, USF2	Viral Transcription			
DDX3X, PABPC1, CHMP4B	Positive regulation of viral life cycle			
FTH1, COTL1, STOM, IQGAP1, MIF, RAC1, B2M, CTSB, FTL	Neutrophil mediated immunity	UP	Shortening	Cytoplasm
HNRNPL, HNRNPA3, HNRNPF, POLR2G, POLR2H, RAN	Nucleic acid metabolic process			
RPL13A, ETF1n POLR2G, POLR2H, RAN, EIF1	Regulation of translation			

To identify suitable gene candidates for further analysis, gene lists within the 8 categories above were subjected to a GO analysis. As can be seen in figure 4.8, with the exception of category 6 gene cluster where it did not show enrichment for biological processes, the top 10 hits for each gene list revealed categories that are highly relevant in the context of a host viral relationship as summarised in Table 4.1. Interestingly, down-regulated gene lists (1-4) were associated with negative regulation of viral release from host cell, positive regulation of type I IFN production, negative regulation of transcription, protein localisation to nuclear envelope and nuclear export. Up-regulated genes on the other hand were associated with mRNA splicing, processing, metabolism, cleavage, positive regulation signal transduction by p53 class mediator, viral transcription and life cycle and translation. Based on this analysis and previously reported links to host viral interaction processes, I selected DDX5, HEXIM1 and MAVS for further analysis. In addition, I also included TRIM25 and SMAD3. Whilst they have not been part of a particular category, they have nevertheless been strongly implicated in regulating host viral interactions.

HEXIM1, in recent literature, has been described to have multiple functions that may be relevant to influenza. It contains an extended coiled-coiled C-terminal domain responsible for specifically binding some non-coding RNAs (Schonichen et al., 2010). Interestingly, depending on the RNA species that binds, HEXIM1 will determine the complex that assembles on it, thus determining the physiological function (Michels and Bensaude 2018). Initially, it was described as a non-coding RNA-dependent transcription repressor. It requires the binding of 7SK RNA to inhibit positive transcription elongation factor (P-TEFb) (Yik et al., 2003; Michels et al., 2004; Dibarne and Bensaude, 2009). P-TEFb is required to release RNA polymerase II (pol II) pausing to achieve transcription (Zhou and Price, 2012; Jonkers et al., 2015; Liu et al., 2017) by masking CDK9 (one of the P-TEFb subunits)

catalytic site and occupying the substrate binding site via HEXIM1 PYNT sequence (Kobbi et al., 2016), thus preventing phosphorylation of serine 2 of the Pol II CTD (Marshall et al., 1996). Increased pol II pausing during transcription initiation may promote influenza cap-snatching which is beneficial for influenza transcription. Conversely, HEXIM1 has also been shown to partner with NEAT1, which is a long abundant nuclear RNA pol II transcript required for formation of nuclear paraspeckles (Adriaens and Marine, 2017). Their interaction contributes to innate immune response against foreign DNA (Morchikh et al., 2017)). Although influenza is a RNA virus, the increase in HEXIM1 expression may effect viral growth by activating immune response (has been shown by Holm et al., 2016 IAV targets STING dependent but cGAS-independent pathway).

DDX5 has been shown to have emerging roles in viral infection. It is a member of the large ASp-Glu-Ala-Asp (DEAD) box ATP dependent RNA helicase family (Lane and Hoeffler., 1980). DDX5 is known to participate in all aspects of RNA metabolism from transcription, translation, RNA decay to miRNA processing (Watanabe et al., 2001; Liu, 2002; Wilson et al., 2004; Lin et al., 2005; Jalal et al., 2007; Zonta et al., 2013; Dardenne et al., 2014). Moreover, it is involved with cancer development, apoptosis, adipogenesis and nuclear translocation of Wnt-β-catenin protein through phosphorylation (Jost et al., 1999; Yang et al., 2005; Saporita et al., 2011). Given DDX5's crucial role in RNA biology, several RNA viruses were discovered to interact with the protein to promote viral replication including influenza virus via functional genomics and proteomics techniques (Bortz et al., 2011).

MAVS and SMAD3 were examples that also exhibited lengthening, however only SMAD3 is significantly lengthening in both nucleus and cytoplasm from 6-18 hpi as MAVS is only significantly lengthening in the nucleus from 6-12 hpi (suggesting a passive APA event).

Contrary to the aforementioned genes, MAVS and SMAD3 are both down-regulated in gene expression. As shown in GO analysis and the literature, MAVS is required for innate immunity regulation. In simple terms, upon RNA viral species recognition, RIG-I (major sensor of viral RNA) interacts with MAVS in the mitochondrial membrane and subsequently phosphorylate and activate both IRF3 and 7 by IKKe and TBK (Kawai and Akira, 2010). In turn, interferon (IFN) is secreted and binds to specific IFN receptors in an autocrine or paracrine manner and activate the JAK/STAT pathway. Thus leading to formation of the ISG factor 3 transcription complex and drives expression of antiviral genes such as Mx GTPases and protein kinase R (PKR).

Although SMAD3 was not enriched in any of the gene clusters during GO analysis, it is a physiologically relevant gene that displays significant lengthening from 6 to 18 hpi in both nuclear and cytoplasmic fractions, and down-regulated in gene expression. Evidence has been provided on SMAD3's role as a main mediator downstream of TGF-β in T cells (Li et al., 2006; Park et al., 2016). SMAD3 mediates the TGF-β-dependent inhibition of CD4 T-cell proliferation, maintaining tolerance to the host by restraining CD4 activation, causes Th1 skewing in a cell-intrinsic manner, and limits cytotoxic activity of myeloid effectors (Delisle et al., 2013).

TRIM25 is an example of a gene that exhibits significant 3'UTR APA shortening from 6hpi until 18 hpi in the nucleus but only significant at 18 hpi in the cytoplasm, and down-regulated in gene expression after infection. TRIM25 is involved with multiple cellular processes (cancer, development and innate immunity) (Heikel et al., 2016). Recently, TRIM25 has been shown to be a type 1 and type 2 IFN-inducible E3 ligase (Carthagena et al., 2009), thus it has a role in antiviral defence. Studies have elucidated its role in mediating ubiquitination of RIG-I. With K63-linked ubiquitin chains on RIG-I, it promotes its

interaction with MAVS and subsequently activate downstream intracellular antiviral signalling. It has been shown to positively regulate melanoma differentiation-associated protein 5 (MDA5)-MAVS-TRAF6 antiviral axis and activates NF-kB (Lee et al., 2015), and acetylates p53 which is an indispensable modification for p53 antiviral activity (Zhang et al., 2015). Additionally, TRIM25 enhances zinc-finger antiviral protein (ZAP) which promotes inhibition of viral mRNA translation. Conversely, TRIM25 may have a dual role in RIG-I regulation, whereby it has been shown to negatively regulate RIG-I through stabilisation of FAT10 (Nguyen et al., 2016).

In a first instance I inspected the APA profiles of the 5 selected genes on the IGV genome browser. As can be seen in figure 4.9, the genome browser screenshots displaying the APA events of the 5 selected genes during viral infection highlight the significant changes in the APA profiles occurring at the 6hpi and later time points both in the nuclear and cytoplasmic fractions (Figure 4.9 A,B top and bottom panels respectively). Importantly and particularly evident for HEXIM1, DDX5 and SMAD3, APA profile changes in the nucleus mirror those in the cytoplasm indicating the likelihood that these APA changes are due to a change in poly(A) site selection at the point of cleavage, perhaps mediated by viral proteins such as NS1. Unlike HEXIM1 and DDX5, MAVS shows extensive lengthening from the 6hpi time point onwards in the nucleus, but this lengthening phenomenon does not penetrate into the cytoplasm. This indicates that MAVS 3'UTR long isoforms are somehow not competent for nuclear and cytoplasmic export.

In summary, the selection of these five genes provided us with two examples that undergo lengthening and a reduction in overall gene expression (DDX5 and HEXIM1) and two examples that are lengthening during viral infection but show the opposite trend regarding expression levels (MAVS and SMAD3) and one example that showed significant

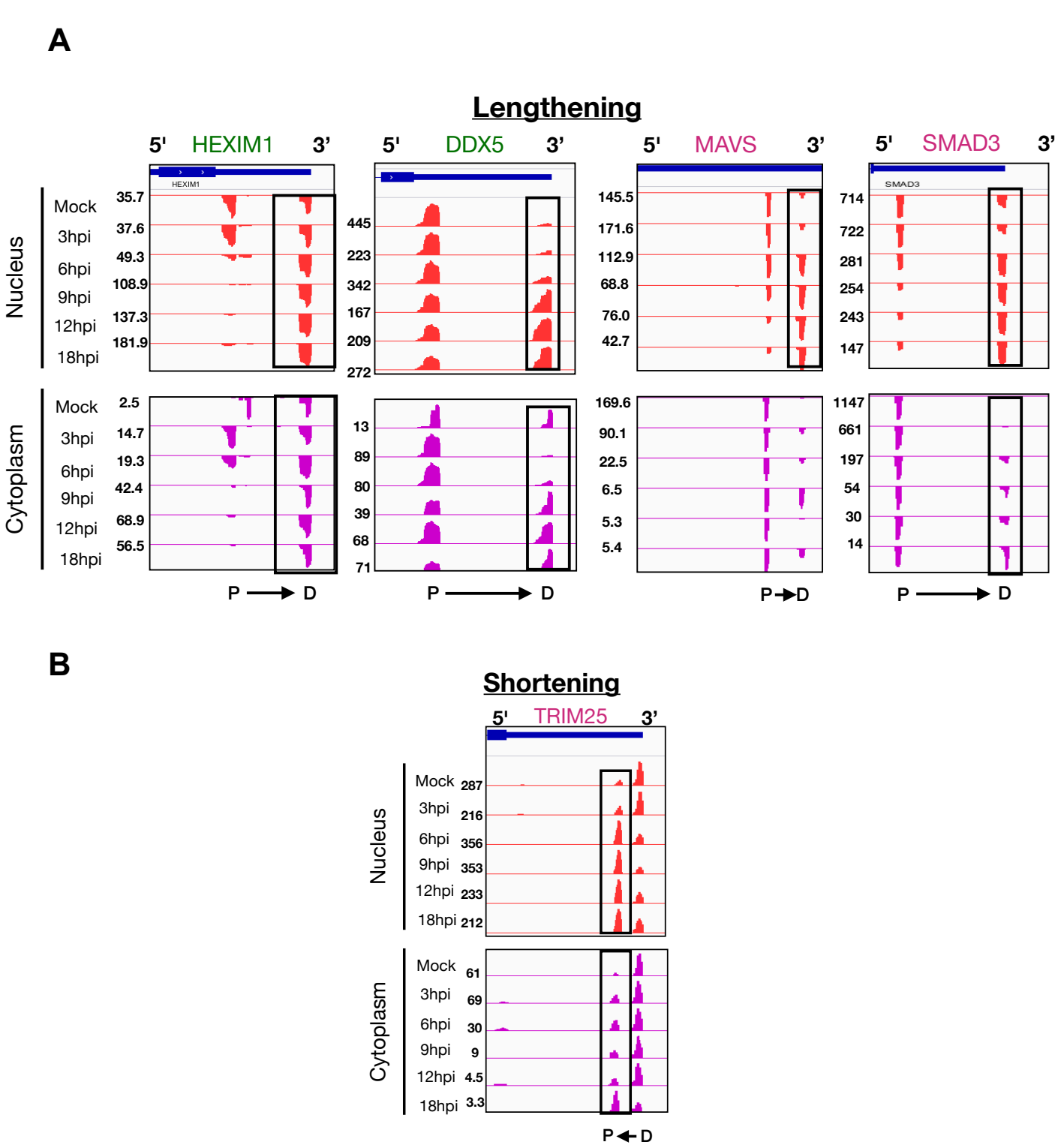


Figure 4.9 Selected Examples of Physiologically Relevant Genes Displaying Differential Gene Expression and 3'UTR APA.

A. From Left to Right: Screenshots of HEXIM1 showed 3'UTR lengthening from 6 hpi onwards in the nucleus (red) and from 9 hpi in the cytoplasm (purple) and is up-regulated in gene expression. DDX5 displayed significant 3'UTR lengthening from 6 hpi in the nucleus (red) and from 9 hpi in the cytoplasm (purple), and is up-regulated in gene expression . MAVS showed 3'UTR lengthening only in the nucleus (red) from 6-12 hpi and down regulated in gene expression. SMAD3 is significantly lengthening in both nucleus (red) and cytoplasm (purple) from 6-18 hpi and exhibited down regulation of gene expression.

B. Screenshots of TRIM25 as an example of a gene that exhibits significant 3'UTR APA shortening from 6hpi till18 hpi in the nucleus (red) but only significant at 18 hpi in the cytoplasm (purple), and down regulation of gene expression after infection. NOTE: Black boxes represent significant APA shift to either P (proximal) or D (distal) CPA site .

shortening and lower gene expression during infection (TRIM25). Importantly, five genes have previously been shown to be associated with pathways that are highly critical during viral infection.

4.2.2c Viral-induced UTR APA is Independent of Other Stress Responses

It has recently been shown by our collaborators that some transcriptional effects mediated by influenza infection are identical to those observed when cells are subjected to osmotic stress (Bauer et al., 2018). This raises the possibility that some effects on host gene expression may not necessarily be due to the expression of viral proteins but rather part of a generic cellular stress response. In order to determine whether osmotic shock causes similar effects on alternative polyadenylation as influenza infection, A549 cells were exposed to high salt concentrations (80mM KCl) for 1 hour in 37°C. The cells from three independent biological repeats were subsequently fractionated to extract nuclear and cytoplasmic RNA. QuantSeq Libraries were generated and subjected to 3'Sequencing and analysed through the bioinformatics pipeline as described above for the IAV infection. The APA analysis showed that osmotic shock triggers extensive APA revealed a significant induction of 3'UTR APA caused by osmotic shock (Figure 4.10A). Similar to the influenza mediated APA, the nuclear APA profiles changes are predominantly associated with lengthening of the 3'UTR after osmotic shock (Figure 4.10A left column). In the cytoplasm, no clear trend towards either lengthening or shortening is evident (Figure 4.10A right column). I next assessed the overlap between osmotic shock and influenza mediated APA events. As can be seen in Figure 4.10 B, less than 8% of genes that undergo lengthening under osmotic stress are also showing the same trend during a viral infection in the nucleus. For the rest of the nuclear and cytoplasmic APA profile changes, less than 4% of APA events do overlap. This would suggest that overall there is little overlap between the

APA events triggered by a generic stress and the majority of APA events that occur in cells are due to influenza infection. Furthermore, I next compared the stress induced and viral induced profile changes of the five genes previously selected. This comparison showed that APA of HEXIM1, DDX5 and MAVS was unique to influenza infection and osmotic shock did not trigger similar APA profile changes in any of these genes. (Figure 4.11 A). Conversely, the APA profile changes observed during influenza infection in TRIM25 and SMAD3 were identical to those observed during osmotic stress. (Figure 4.11B).

This analysis forced us to reassess our selection of five genes and I decided to prioritise HEXIM1, DDX5 and MAVS for further analysis as the APA profile changes observed in SMAD3 and TRIM25 may not be specific to influenza virus infection.

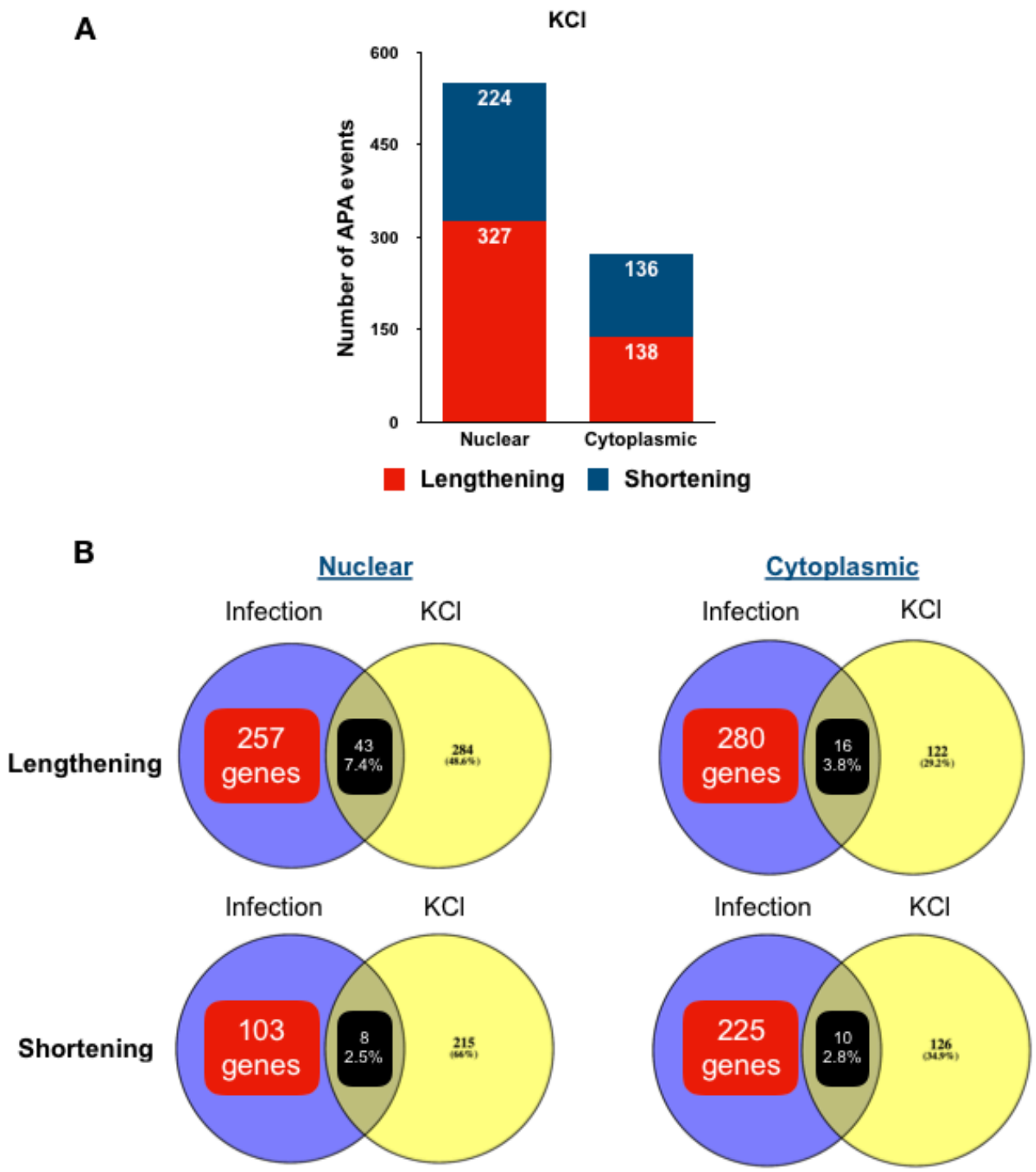


Figure 4.10 Viral-induced 3'UTR APA are Independent of those Induced by Osmotic Stress

A. Global 3'UTR APA events during osmotic stress in both nuclear in and cytoplasmic fractions: Bar charts display the number of events showing significant differential 3'UTR APA regulation in nuclear and cytoplasmic fractions over 3 independent biological repeats; number of lengthening events represented by red, number of shortening events represent by blue. Significant differentially represented APA isoforms are those in which $p<0.01$ (Fisher's Exact Test) and with a minimal change in isoform abundance change of ≧5% relative to total APA isoform associated with that gene. The total detectable genes were 6685 in the nucleus and 6210 in the cytoplasm (Supplementary Figure 3, see scatterplots grey dots).

B. Venn diagrams showing number of overlaps between infection and osmotic stress induced APA genes in categories of nuclear lengthening, shortening, and cytoplasmic lengthening, shortening. Unique infection induced APA is on the left (red box). Overlaps of genes are shown in the middle of the Venn diagram (black box).

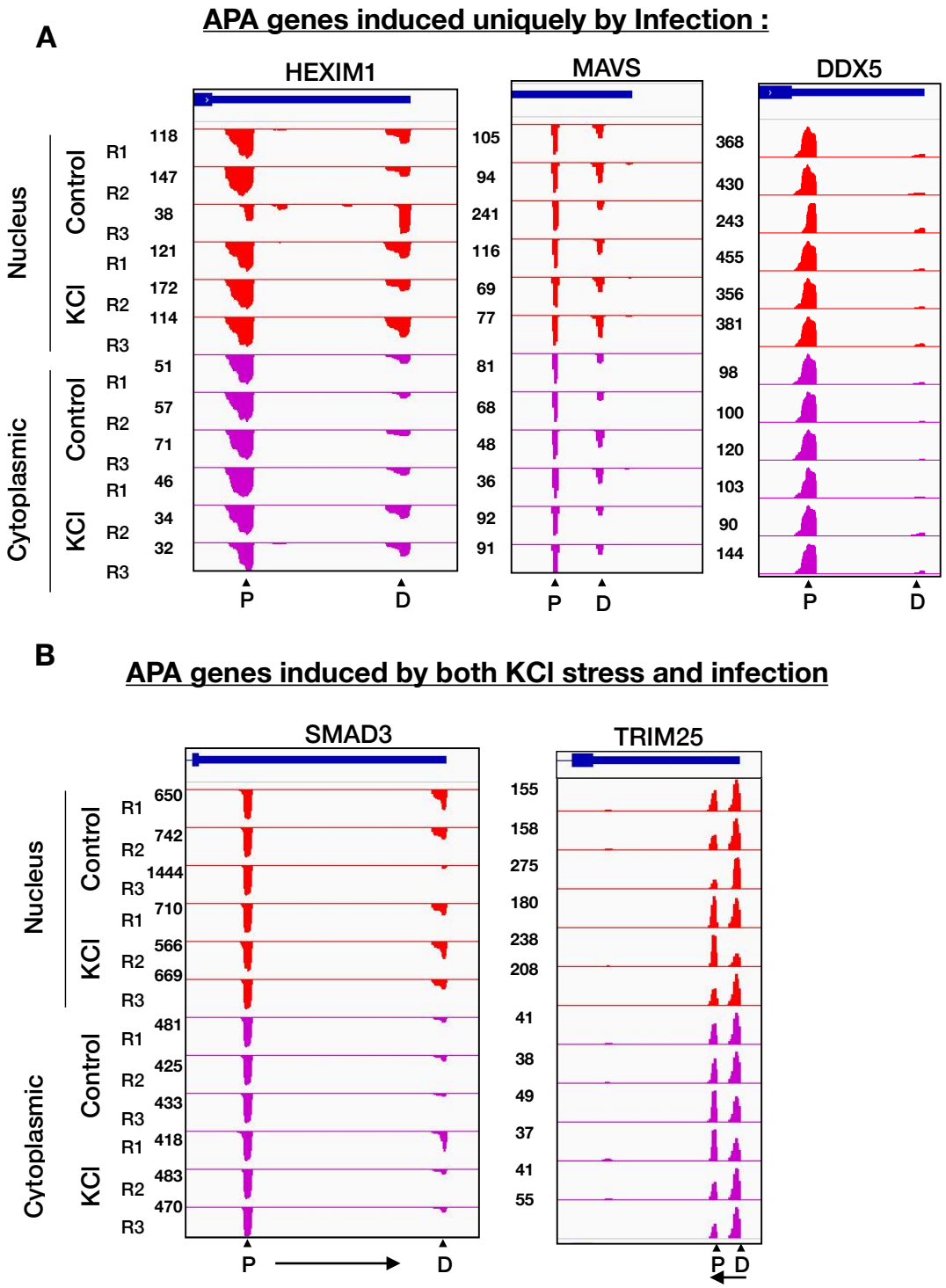


Figure 4.11 Selected Examples of Physiologically Relevant Genes 3'UTR APA Induced Uniquely by Infection as well as those that are Induced by Both Osmotic Stress and Infection.

A. From Left to Right: Screenshots of HEXIM1, MAVS and DDX5, taken from all 3 biological replicates of osmotic stress experiments, showed no signs of significant 3'UTR lengthening both in the nucleus (red) and in the cytoplasm (purple). SMAD3 is significantly lengthening in both nucleus (red) and cytoplasm (purple) from 6-18 hpi and exhibited down regulation of gene expression. This comparison showed that APA of HEXIM1, DDX5 and MAVS was unique to influenza infection and osmotic shock did not trigger similar APA profile changes in any of these genes.

B. Screenshots of SMAD3 and TRIM25 taken from all 3 biological replicates of osmotic stress experiments. The APA profile changes observed during influenza infection (Figure 4.9) in TRIM25 and SMAD3 were identical to those observed during osmotic stress. As shown in the black boxes, SMAD3 exhibit the shift of CPA site usage from proximal (P) to distal (D) (lengthening) CPA site. TRIM25 exhibited the shift of CPA site usage from distal to proximal CPA site.

4.2.2d RT-qPCR verified APA events

So far, all the data were obtained using an RNAseq based approach. In order to verify the APA events, a RT-qPCR approach was used as an alternative method. Gene specific primers were designed to amplify a specific section (section 1, Figure 4.12A) of the mRNAs that captures all 3'UTR isoforms and thus reflects total output. A second primer pair was designed to specifically amplify the alternative UTR (Section 2, Figure 4.12A) An oligodT primer (Table 2.1) was used to reverse transcribe the polyadenylated transcripts of the nuclear and cytoplasmic RNA isolated from cells after 3hpi, 9hpi, 12hpi and mock cells. The primer sets described above are then employed to perform the real time PCR amplifications. The real time PCR values obtained were then used to calculate the ratio between the total mRNA isoforms (primer pair 1) and the corresponding long 3'UTR isoform (primer pair 2) . For this, ratios were calculated by dividing the relative concentration of long over the relative concentration of the total 3'UTR isoform output and subsequently log2 transformed (log2). The different ratios were then compared between infected (3, 9 and 12 hpi) and mock nuclear and cytoplasmic samples, using qRT-PCR results of the 3 independent biological repeats for each time point. As can be seen in figure 4.12B and C, the lengthening of HEXIM1 (left panel) and DDX5 (right panel) observed in the nuclear (top panels) and cytoplasmic (bottom panels) fraction by RNASeq was confirmed by the RT-qPCR approach. In both cases a distinct significant (two-tailed, paired t-test, $p<0.05$) shift in CPA site usage from the proximal site in uninfected cells to the distal CPA site in cells was demonstrated, but as expected there was no indication of lengthening at 3 hpi when comparing to mock. Therefore, the RT-PCR analysis mirrored the RNAseq data described above confirming that influenza viral infection causes a distinct 3'UTR lengthening in both HEXIM1 and DDX5 in the cytoplasm as well as in the nucleus.

In the case of MAVS, RT-qPCR could not show significant lengthening at 9hpi and 12 hpi when compared to mock in three independent nuclear repeats (Supplementary Figure 4). Interestingly, as shown in figure 4.9A, lengthening of MAVS was limited to the nucleus thus suggesting a possibility that the nuclear RNA is not exported into the cytoplasm. I hypothesised that nuclear retention was due to MAVS undergoing termination defect, which was caused by the suppression of the first CPA site. Also there was a possibility that the 'distal' CPA site may be detected due to internal priming despite the fact that the bioinformatics pipeline should exclude most these events, but some could penetrate. Closer inspection near the 'distal' CPA site where the RNAseq reads end of MAVS did reveal the presence of an A-rich region. Further clues from a study elucidated MAVS, amongst other antiviral immune signalling genes, might not possess two distinct CPA sites and in fact contains a single poly(A) site located in its 3'UTR (Jia et al.,2017). I therefore tested whether it was an actual bona fide PAS, whether the AU-rich region was falsely identified as a CPA site, and whether this was simply a termination defect in host transcription. Random primers were used for reverse transcription to avoid the same strategy as 3'Seq. Using primers both upstream and downstream of the hypothesised alternative CPA site (Figure 4.13A), I was able to demonstrate that viral infection induces potential failure of termination thus categorised as a 'read-through' event rather than APA in the case of MAVS. If it was a real CPA site one would expect for primer pair 2 to not amplify for any product, since cleavage would occur at the distal CPA site, therefore the ratio between 'upstream'/ 'downstream' would be more than 1. Evidently since the ratio of upstream/ downstream is close to 1 after 9 hpi in two independent biological repeats, it confirmed that the 'lengthening' event was indeed due to termination defect (Figure 4.13B).

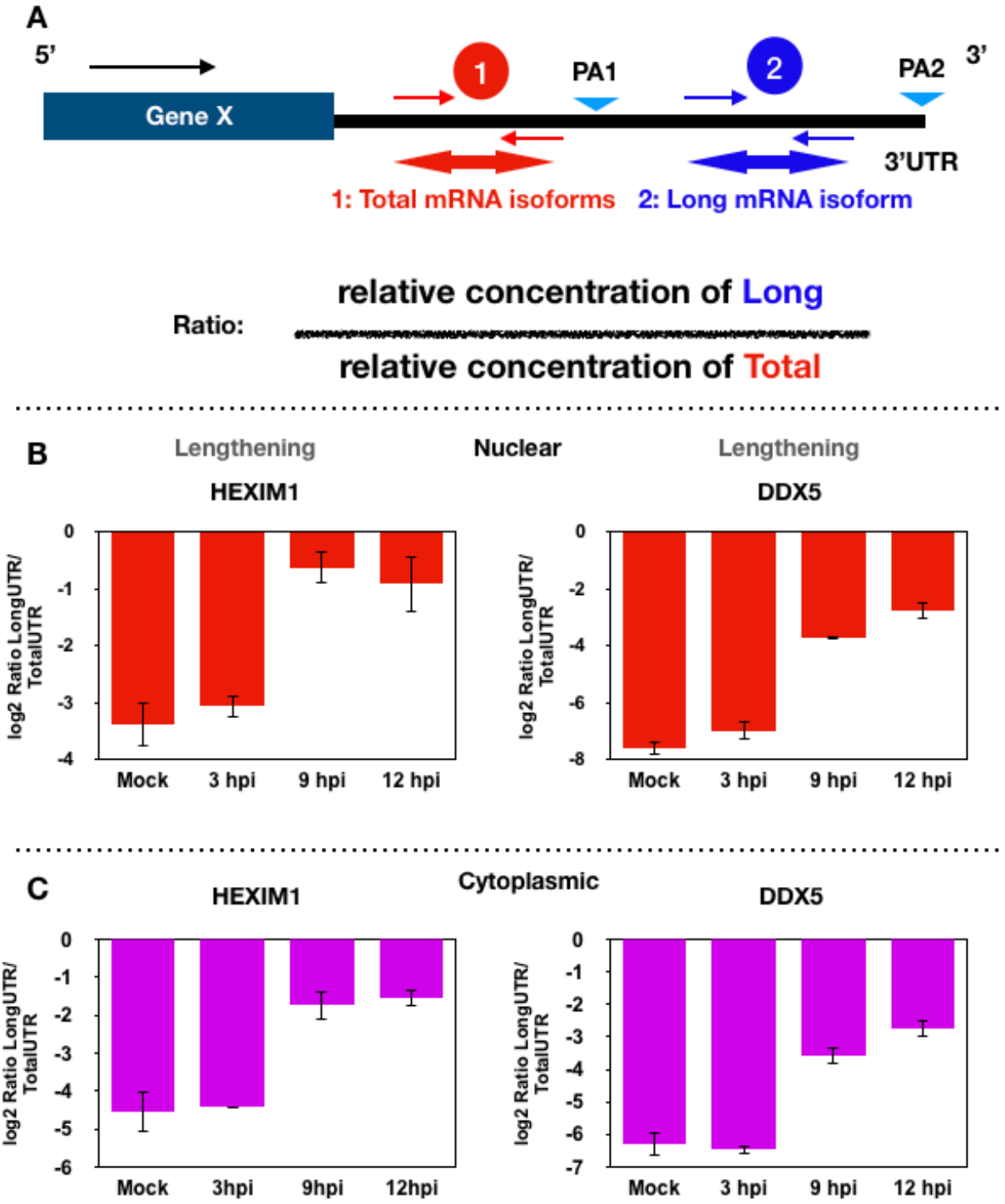


Figure 4.12 Real time qPCR to verify 3'UTR APA of HEXIM1 and DDX5

A. Schematic demonstration of designing primers for detecting total mRNA isoforms (primer pair 1) and long mRNA informs (primer pair 2).

B. Real-time q-PCR analysis of selected APA genes (HEXIM1, DDX5) upon IAV infection in A549 cells in nucleus (red). The x-axis denotes 3 time points post infection (3,9,12 hpi) and mock (non-infected), and the y-axis denotes the $\log_2$ ratio of longer 3' UTR isoforms (L) to total mRNA isoforms (T). Data are presented as mean log2 ratio L/T with error bars representing SEM of 3 biological repeats. Paired, 2 tailed t-test performed to compare groups, with $P<0.05$ considered significant.

C. Same analysis was done as panel B, selected APA genes upon IAV infection in A549 cells in cytoplasm (purple). $p<0.05$ was considered significant

HEXIM1 and DDX5, in both nucleus and cytoplasm, show significant 3'UTR lengthening after infection (9 and 12hpi).

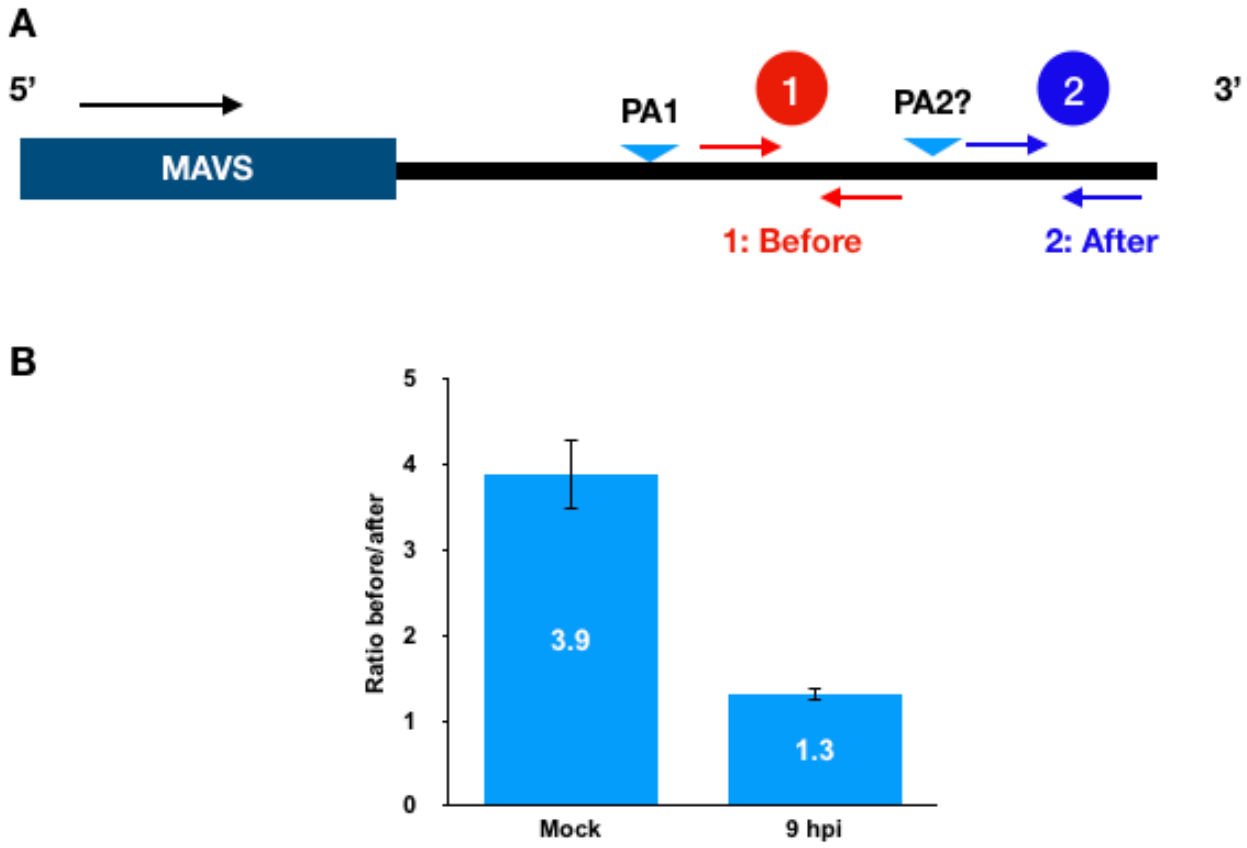


Figure 4.13 Real time qPCR to verify 3'UTR APA or failure of termination of MAVS

A. Random primers were used for reverse transcription to avoid the same strategy as 3'Seq. Schematic diagram demonstrating the use of primers both upstream (primer pair 1) and downstream (primer pair 2) of the hypothesised alternative CPA site.

B. This is a graph comparing the ratio upstream /downstream of 9 hpi to mock in the nucleus. Ratio of before/after of mock and 9 hpi are displayed as mean ratio with error bars representing SEM in two independent biological repeats. Since the ratio of before/after is close to 1 after 9 hpi in two independent biological repeats, it confirmed that the 'lengthening' event was indeed due to termination defect

4.2.2e 3'UTR APA Results in Downstream Effects in Gene and Protein Expression

In the previous section I have established that influenza viral infection has a dramatic and specific effect on the host APA profiles. I have been able to further identify a number of genes that undergo APA and that this is concomitant with a change in RNA expression levels. Most importantly, I have shown that this includes genes that have previously been linked with processes that are critical for the host viral relationship. There was a focus on three specific examples HEXIM1, MAVS and DDX5 all of which show virus specific APA and change their mRNA expression levels.To clarify whether these changes do have a physiological consequence, I next assessed the expression of the three genes at the protein level. To that end, I isolated total cellular protein of uninfected (Figure 4.14A, Mock) and influenza infected cells at 3h, 6h, 9h, 12h and 18 hpi. The expression levels were then assessed in a western blot using commercially available antibodies recognising the three proteins in question and two different loading controls (HSP60 and GAPDH) that do not undergo APA and show no changes in expression levels during the infection. As can be seen in Figure 4.14A, MAVS and HEXIM1 protein levels are reduced at the 12h and 18h late transfection time points. For MAVS, suppression of the proximal polyA site in at the point of cleavage and the consequential retention of mRNAs in the nucleus are likely to account for this reduction. In the case of HEXIM1, despite the increase in overall transcripts at later infection time points (see figure 4.7) it appears that this is not affecting protein levels in a similar way. HEXIM1 protein levels actually drop below the levels seen in uninfected cells at 12 and 18hpi (Figure 14A, HEXIM1). The protein levels of DDX5 in contrast, are similar to the corresponding mRNA levels (Figure 4.7) and progressively increase from the 6hpi point onwards in infected cells (Figure 14A, DDX5).

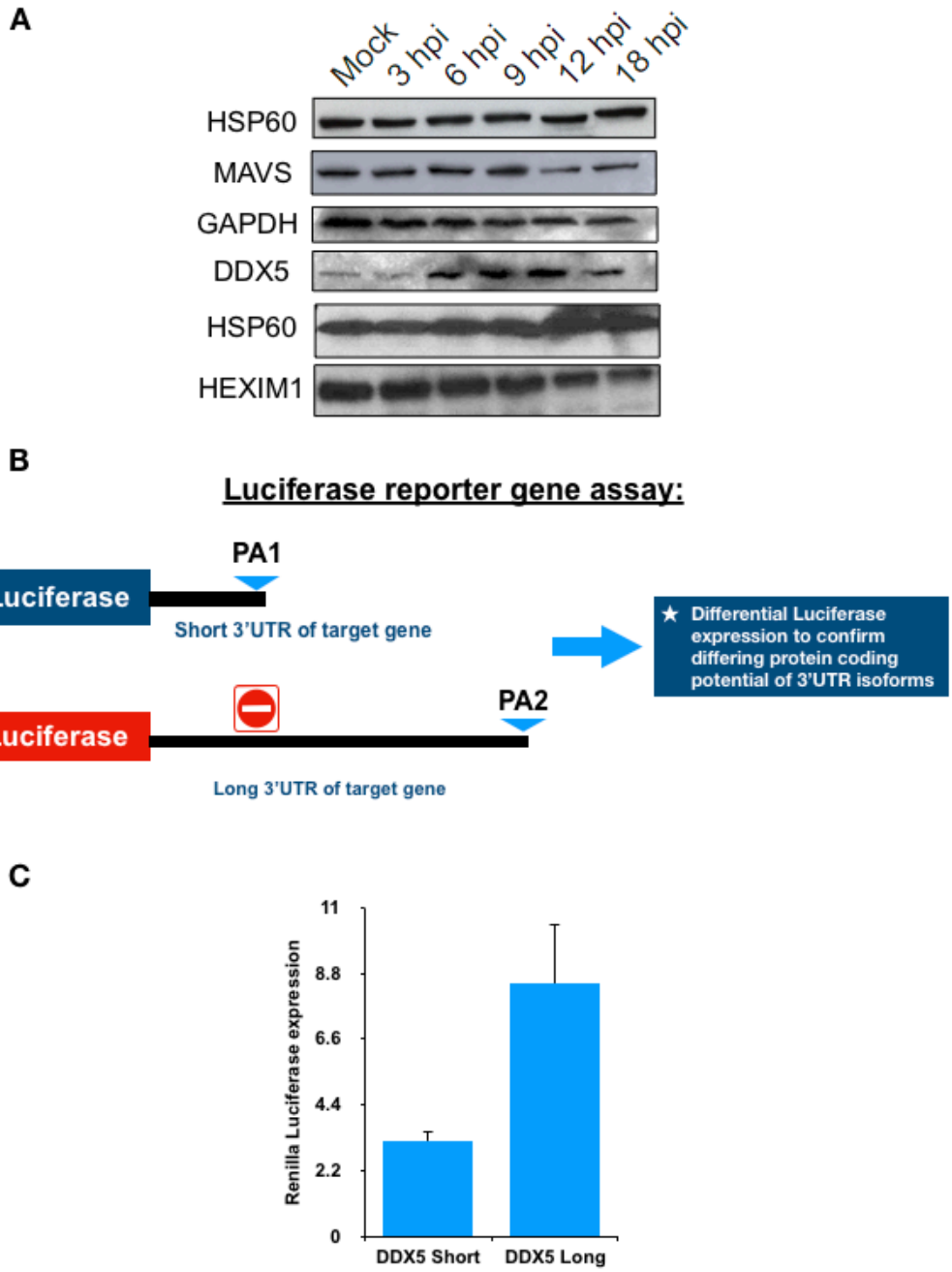


Figure 4.14 3'UTR APA resulted in downstream effects in Gene and Protein Expression

A. Human A549 cells were infected with A/WSN/33 virus at an MOI of 5. Whole cell lysates were analysed by SDS-PAGE and western blotting using commercially available antibodies targeting the three proteins in question (MAVS, DDX5, HEXIM1) and two different loading controls (HSP60 and GAPDH). From top to bottom: HSP60 is a loading control for MAVS western analysis. MAVS showed decrease in protein expression at 12 hpi. GAPDH serves as a loading control for DDX5 western analysis. DDX5 showed increase in protein expression at 6 hpi. HSP60 was used as a loading control for HEXIM1 western blots. HEXIM1 showed decrease in protein expression after 12 hpi.

B. Schematic diagram of piS2-3'UTR constructs: top panel: short 3'UTR Bottom panel: long 3'UTR with proximal CPA site mutation. For the full length reporter plasmid the proximal polyA site was inactivated by site directed mutagenesis that changed the polyA site from AATAAA to ACTAGT (SpeI restriction site)

C. DDX5 piS2-3'UTR constructs were transfected into 293T cells. The luciferase expression of reporter possessing 3'UTR of targets was normalised to firefly luc2 reporter. The bar chart shows the mean reporter activity values with error bars indicating SEM (3 independent experiments). The reporter plasmid containing the long DDX5 UTR resulted in significantly more (>3 fold, $p<0.05$, one tailed, 2-sample equal t-test) luciferase activity compared to the short UTR containing plasmid.

These results suggest that the viral induced switch to the longer DDX5 mRNA isoform increases transcript levels thus resulting in an increased protein output. To verify whether the alternative UTR not only impacts on the mRNA level but also whether it affects protein output of target genes, I used a luciferase reporter gene based assay. To test the effect of the constitutive and alternative UTRs on protein expression, the corresponding sequences were cloned into the piS2 *Renilla* luciferase reporter plasmid downstream of the *Renilla* luciferase translational stop codon (Figure 4.12B). For the full length reporter plasmid, the proximal polyA site was inactivated by site directed mutagenesis that changed the polyA site from AATAAA to ACTAGT (SpeI restriction site). A control plasmid pGL4.20 vector containing firefly luc2 luciferase reporter gene and BMAL1 promoter along either of the two plasmids with the 'short' or 'long' UTR were then transfected into HEK293T cells and isolated protein was subsequently analysed for luciferase activity on a PHERAstar Luminometer. As can be seen in Figure 4.12C, the reporter plasmid containing the long DDX5 UTR resulted in significantly more (>3 fold, $p<0.05$, one tailed, 2-sample equal t-test) luciferase activity compared to the short UTR containing plasmid. This suggests that the viral induced switch from the proximal to the distal polyA site in DDX5 results in a boost of DDX5 protein expression by increasing both the transcript levels and perhaps translation efficiency. This is highly significant as DDX5 expression has been shown to be beneficial for influenza viral replication (Bortz et al., 2011). To further determine whether alternative 3'UTRs could influence translation efficiency, a polysome profiling assay will be required.

Furthermore, in order to confirm the physiological relevance of APA and the concomitant differentially expressed genes such as DDX5, HEXIM1 and MAVS, the effects of viral RNA expression were observed after over expressing these genes (done by Benjamin Nilsson) in infected HEK293T cells. HEK293T transfection of expression vectors (pcDNA3)

containing HEXIM1, DDX5 and (pCAGGS) containing MAVS were done over 48 hours. Over expression was verified by western blot (Supplementary figure 5A) along with loading control GAPDH (Supplementary figure 5B). Subsequently, cells over expressing target genes were infected with A/ WSN/33 at an MOI 0.001 for 48 hours. Total RNA was extracted from cells with the help of Trizol and viral replication was assessed using the primer extension approach described in chapter 2.9. The effects of over expressing DDX5, in contrast to previous reports, and HEXIM1 did not yield an effect on viral RNA expression. However, in cells that over-express MAVS, overall viral RNA expression was significantly decreased (Figure 4.15). This is indicates that over expressing MAVS certainly has an anti-viral phenotype. Whereas, it is still unclear regarding DDX5 and HEXIM1 whether they have any effect on viral growth, thus siRNA knockdown experiments are currently performed to resolve this.

Altogether, I have shown significant APA profile changes after influenza A/WSN/33 infection in both host nuclear and cytoplasmic fractions. The changes seen in genes were highly relevant for viral host relationship. I have also shown that there was a link between APA and differential gene expression levels for selected examples. The majority of viral induced APA were also distinct from generic stress response induced by osmotic shock, but from the 5 selected genes, APA profile changes observed during influenza infection in TRIM25 and SMAD3 were identical to those observed during osmotic stress. Furthermore, my findings also suggested that APA could have a physiological impact for host and virus. With examples of genes such as DDX5, 3’UTR APA lengthening of DDX5 upon infection has an impact on mRNA levels and combined with the fact that the alternative UTR containing isoform can increase translational output thus impacting on the overall protein levels. Thus overall these results confirmed and agree with the original hypotheses.

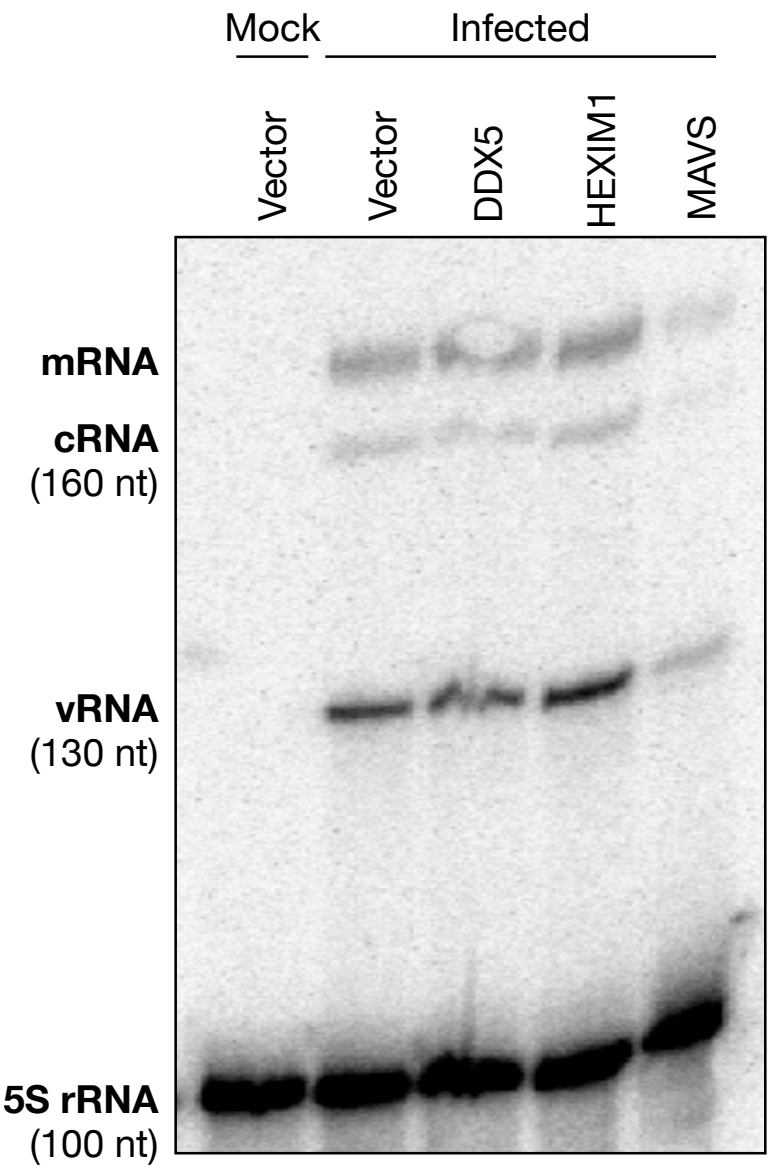


Figure 4.15 Functional Consequences of Viral Infection when Expressing Selected Genes

Constitutive over expression of DDX5, HEXIM1 and MAVS proteins was performed by transfecting expression vectors into HEK293T cells over 48 hours. Proteins of interest were flag tagged and expression was verified by using anti-FLAG-M2 antibody as displayed in supplementary figure 5. Primer extensions and 6% PAGE performed by Ben Nilsson analysed accumulation of viral RNA after expressing DDX5, HEXIM1 and MAVS in 293T cells. Expressing DDX5 and HEXIM1 did not alter viral RNA expression, although as expected MAVS over expression depleted mRNA, cRNA and vRNA. 5S rRNA was used a loading control.

4.2.3 Impact of influenza infection on Coding Region APA of Host Transcripts

4.2.3a Influenza Infection Induces Coding Region APA of Host Transcripts

Whilst 3'UTR APA has been more commonly studied, changes in CR-APA, where an upstream CPA site in an upstream exon or intron is utilised, have not been extensively studied. In the case of CR-APA, usage of an upstream CPA site alters the coding message and therefore impacts the protein produced. This event is described as CR-APA shortening. Whereas the shift in usage of CPA in the 3'most exon is described as CR-APA lengthening. Due to the nature of CR-APA events, it is easier to establish a physiological consequence since the shift in CPA site choice in CR-APA would result in the production of a different protein.

To isolate events that shift between CR-APA isoforms, the equivalent subcellular fractions of influenza infected and non-infected were compared. This was analysed by combining all the reads upstream of the terminal exon and comparing these to the combined reads within the 3'UTR. Using the same established bioinformatics pipeline stated in 4.2.2, reads from two biological repeats were combined, and significantly changing APA events ($p<0.01$ Fisher's exact Test, with a minimum 5% change in APA isoform abundance) in the nucleus and the cytoplasm were identified separately. 14000 of polyadenylation events were observed in both the nucleus and cytoplasm for each time point (Supplementary Figure 6). The comparison of the infected to non-infected nuclear fractions revealed a tendency of shortening in nuclear fraction of infected cells, and interestingly was different to 3'UTR APA where a lengthening trend was observed in the nucleus. However, similar to 3'UTR APA,

there was no clear trend towards shorting or lengthening in the cytoplasm. Furthermore, fewer CR-APA events were observed in both nuclear and cytoplasmic fractions upon infection than 3'UTR APA events (Figure 4.16A).

The same selection process as chapter 4.2.2b was implemented to identify genes that exhibit CR-APA after a minimum of 3 indicated infection time points (6-18 hpi). This selection process yielded 4 categories: Nuclear lengthening (44 genes), nuclear shortening (96 genes), cytoplasmic lengthening (88 genes) and cytoplasmic shortening (84 genes). GO analysis was subsequently performed using the Enrichr platform in order to address whether these events could have any physiological relevance for the infection process for the virus or the host. The analysis revealed that genes in nuclear lengthening cluster were strongly associated with mRNA processing (Figure 4.16B). Nuclear shortening were strongly enriched for RNA splicing and cell cycle regulation (Figure 4.16C). Whilst both cytoplasmic lengthening and shortening did not show any enrichment of biological processes, it is certainly interesting as these biological processes also are overlapped with 3'UTR-APA events in the nucleus, suggesting that APA events of genes associated with these categories are regulated by both 3'UTR- and CR-APA.

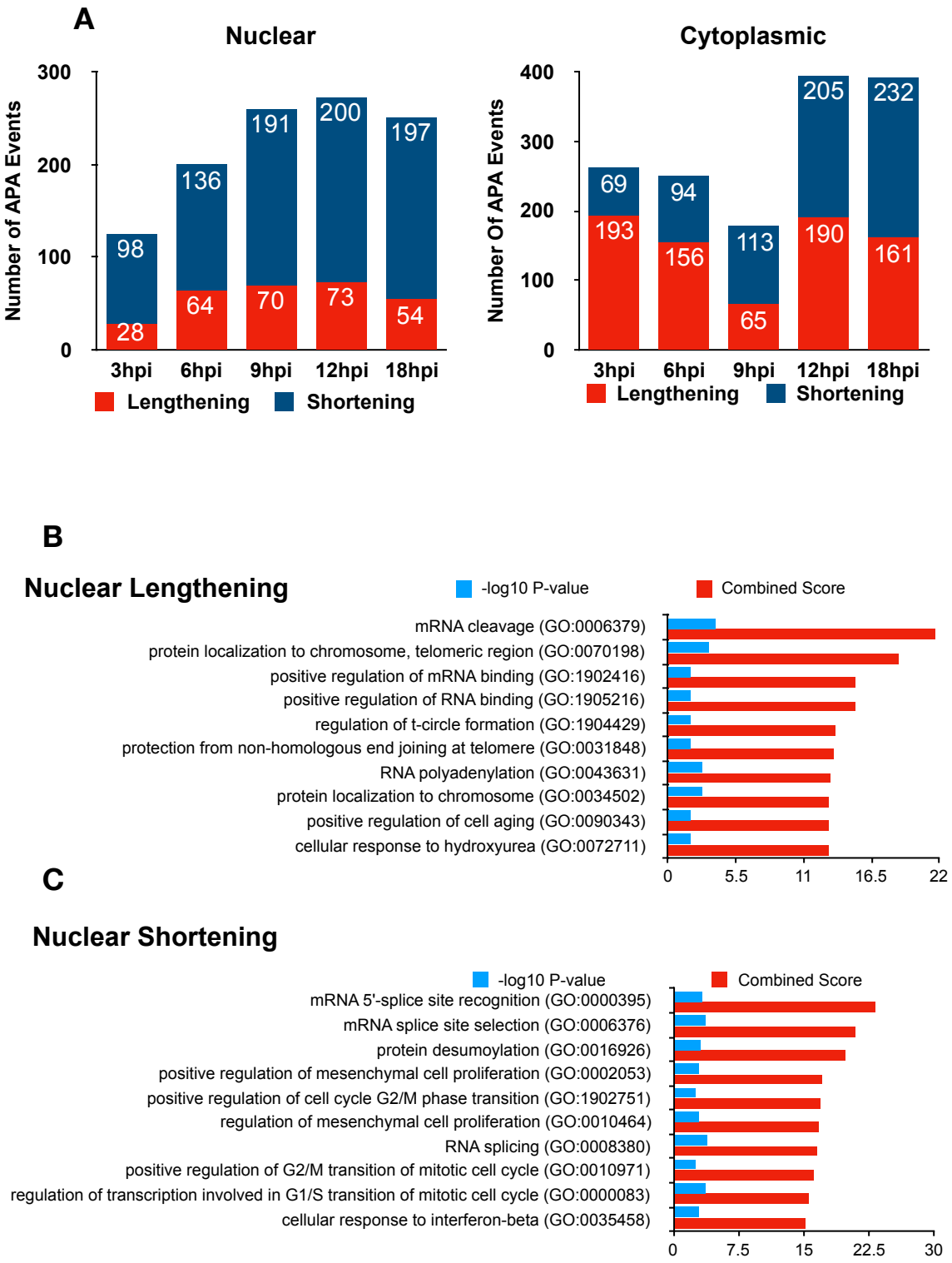


Figure 4.16 Global WSN/33 Induced CR-APA Events

A. Bar graphs to visualise number of significant CR-APA events (Fisher's exact Test, $p<0.01$, with a minimum 5% change in APA isoform abundance) after A/WSN/33 infection in both nuclear fraction over 2 independent biological repeats and cytoplasmic fractions over 2 biological repeats of sequencing. The number of lengthening events are presented in red, whereas the number of shortening events are presented in blue. The total detectable genes in the nucleus (Panel A) and cytoplasm (Panel B) are around 14000 (Supplementary figure 6, grey dots on scatterplots).

The ten most enriched GO terms for biological processes and their statistical significance ($p<0.05$) of the genes were associated with **B.** nuclear lengthening **C.** nuclear shortening. Genes associated with cytoplasmic lengthening and shortening were not significantly enriched for any biological processes. Gene ontology analysis was performed using Enrichr.

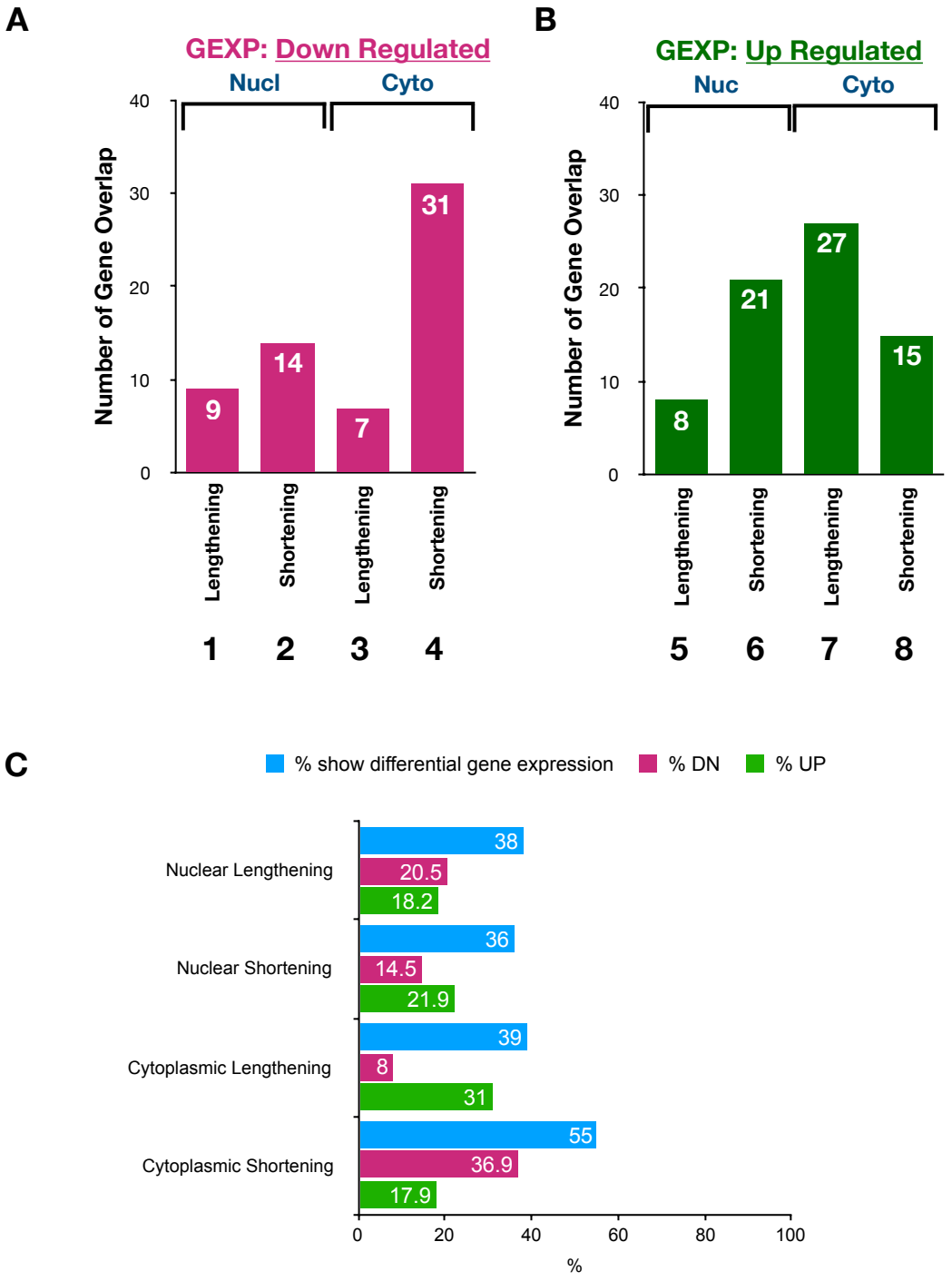


Figure 4.17 Differentially Expressed Genes Overlapped with CR-APA Genes Significant at a Minimum of 3 Time Points

A) Down regulated genes (total of 1934 genes) were overlapped with 4 main selected CR-APA categories in both nucleus and cytoplasm. Bar graph represents number of gene overlaps of Gene clusters 1: DN nuclear lengthening, 2: DN nuclear shortening, 3: DN cytoplasmic lengthening, 4: DN cytoplasmic shortening.

B) Up regulated genes (total of 1450 genes) were overlapped with 4 main selected APA categories in both nucleus and cytoplasm. Bar graph represents number of gene overlaps of Gene clusters 5: UP nuclear lengthening, 6: UP nuclear shortening, 7: UP cytoplasmic lengthening, 8: UP cytoplasmic shortening

C) Proportion of CR-APA events exhibited differential gene expression summarised in blue (%). The purple bar represents the % of the total APA events, what % are DN and what % are UP (green bar).

Table 4.2- Overlaps between Genes with differential Gene Expression and CR-APA

NL x DN (9)	NS x DN (14)	C L x DN (7)	CS x DN (31)	NL x UP (8)	NS x UP (21)	CL x UP (27)	CS x UP (15)
ATRX	AXL	ATRX	B4GALNT1	ANKRD16	ANP32B	ANKRD16	ASXL1
CYHR1	BRD4	EGFR	B9D2	FAM81A	AP3B1	ANP32B	CYB5D2
GIT1	CPSF3L	GNB1L	C10orf54	MSL1	CYB5D2	ANTXR1	EIF4G3
GLS	ELOVL5	INPP5E	C15orf38-AP3S2	PCBP2	DYNLL1	CAPZB	JOSD1
LMNA	FDFT1	PPAN-P2RY11	C16orf55	PCF11	HEXA	CAST	NACA
MRPL43	LRP5	PPHLN1	CBX2	TCFL5	HSP90AA1	CDKN1B	PCBP2
PPAN-P2RY11	MPDU1	SQSTM1	CCDC71	THAP2	IQGAP1	CSNK2A2	PCMT1
TKT	MUM1		CYBASC3	WDR33	KIFC3	CSTF3	PFDN2
TMPO	MVP		DENND1A		MATR3	ERGIC1	PTPN9
	PKM2		DGKD		NDUFA8	HSP90AA1	RPL34
	RBCK1		DHX30		PCMT1	HSP90B1	RPS23
	SQSTM1		EXOSC4		PSIP1	LBR	SLC25A37
	TNPO2		HS1BP3		PTGR1	LOC729082	TP53
	USP31		IFNAR2		PTPN9	MDH2	ZNF652
			INF2		SH3BP4	METTL5	
			MAP2K5		SLC25A37	NRP2	
			MLPH		SMARCE1	PABPC1	
			MTG1		STAT1	PPA2	
			MUM1		TFDP1	PSIP1	
			NDST1		ZNF280D	PTGR1	
			NPLOC4			RBX1	
			NUP210			SET	
			PDCD2			TAF1	
			PSKH1			TBCA	
			PUS1			TMED2	
			RAD1			VDAC1	
			RNF220			ZNHIT1	
			SH2D4A				
			SNAPC4				
			SRC				
			TSPAN14				

Next, similar to the selection process mentioned in chapter 4.2.2b, the aim was to address what proportion of CR-APA genes overlap with differentially expressed genes that are either up-regulated (1450 genes) or down-regulated (1934 genes) at a minimum of 3 infection time-points from 6-18 hpi, and thus identifying those that may have a potential physiological impact. By overlapping the CR-APA cohort and gene expression cohort, 8 individual categories depending on whether they are lengthening or shortening and are up or down regulated can be formed (shown in Figure 4.17A, B). 38% of nuclear lengthening genes exhibited in significant change in gene expression. 36% of nuclear shortening genes resulted in significant gene expression changes. For the cytoplasmic categories, 39% of the lengthening and 55% of the shortening was accompanied by differential gene expression (Figure 4.17C). These results again suggest that viral induced CR-APA can have a significant impact on host mRNA levels. All the overlapped genes in each of the 8 individual categories were listed in table 4.2.

Next, to identify CR-APA gene candidates that can have physiological impact, research was done on each individual gene listed in table 4.2 whereby some CR-APA events were physiologically relevant. Since transcripts will be translated in the cytoplasm, a focus on cytoplasmic changes will likely detect a great physiological impact (red boxes in Table 4.2). Selected genes (highlighted in yellow in Table 4.2) were physiologically relevant to immune response. DHX30 exhibited the shift of usage to the proximal upstream CPA either in the upstream exon or intron (shortening), as well as down regulation in gene expression (Figure 4.18A). RNA helicase DEAH (Asp-Glu-Ala-His) box polypeptide 30 (DHX30) has been shown to bind ZAP to facilitate its antiviral action (Ye et al., 2010). Similar to DHX30, IFNAR2 demonstrated CR-APA shortening and down regulation in gene expression (Figure 4.18B). It encodes Interferon Alpha and Beta Receptor Subunit 2. Its association with IFNAR1 forms type 1 interferon receptor, which has been shown to have an

indispensable role for IFN-α signifying pathway and antiviral immunity (Zhang et al., 2017). PCBP2 exhibited cytoplasmic shortening and up-regulation in gene expression (Figure 4.18C). PCBP2 is a member of the hnRNP E family of RNA binding proteins that interact with single stranded polyC tract motifs. In addition to maintaining RNA stability and regulating mRNA translation, it is involved with regulation of viral defence and negative regulation of MAVS signalling pathway (Makeyev and Liebhaber, 2002). Altogether, infection induced APA and change in expression of these 3 genes. The fact that these effects are only seen in the cytoplasm suggests that CR-APA of these genes are regulated at the post-transcriptional level. Additionally, as these effects penetrate through to the cytoplasm, this creates an opportunity in elucidating how host response and shut-off occurs during viral infection.

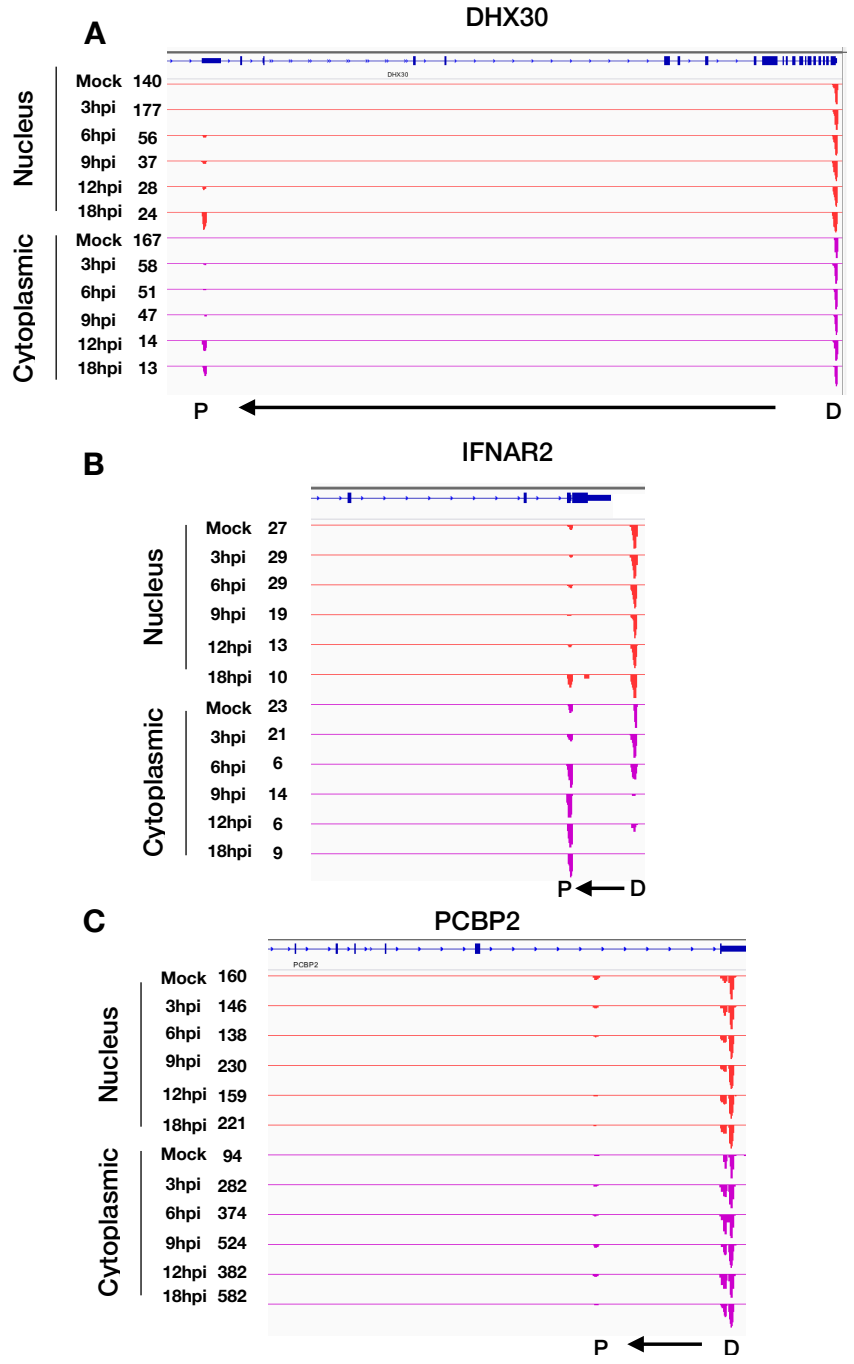


Figure 4.18 Selected Examples of Physiologically Relevant Genes Displaying Differential Gene Expression and CR-APA.

A. Screenshot of DHX30 as an example of a gene that displays significant shifts in CPA site usage to using the proximal (p) CPA site and down regulation of gene expression. Shortening is only limited in cytoplasm resulting in truncated protein product.

B. Screenshot of IFNAR2 displaying CR-APA shortening and down regulation in gene expression only in cytoplasmic (purple) fractions.

C. Screenshot of PCBP2 displaying CR-APA shortening and up-regulation of gene expression only in cytoplasmic (purple) fractions.

All IGV images are displayed with a 5' to 3' direction. P represents proximal CPA site, D represents distal CPA site.

4.2.3b Viral-induced CR-APA is Independent of Osmotic Shock Responses

Similar to chapter 4.2.2c, the aim was to determine whether the effects of alternative polyadenylation caused by influenza were unique from a generic cellular stress such as osmotic shock. Analysis of CPA site usage revealed significant induction of CR-APA after osmotic shock, similar to viral infection (Figure 4.19A). Similar to influenza-mediated APA, after osmotic shock the nuclear APA profiles changes are predominantly associated with shortening of the 3'UTR, and no clear trend towards either lengthening or shortening is evident in the cytoplasm. I next compared genes exhibiting APA in viral infection and osmotic shock, which revealed only a small proportion of genes where APA was triggered in both cases (ranging from 3.5% to 4.2% overlap) (Figure 4.19B). I next compared viral induced CR-APA profile changes of the 3 genes previously selected. The analysis of the selected genes revealed that CR-APA of DHX30 and PCBP2 was unique to influenza infection and osmotic shock did not trigger similar APA profiles. On the other hand, CR-APA of IFNAR2 observed during influenza infection were identical to those observed during osmotic stress.

This analysis forced us to reassess our selection of the three genes and, thus DHX30 and PCBP2 should be the focus for future experiments.

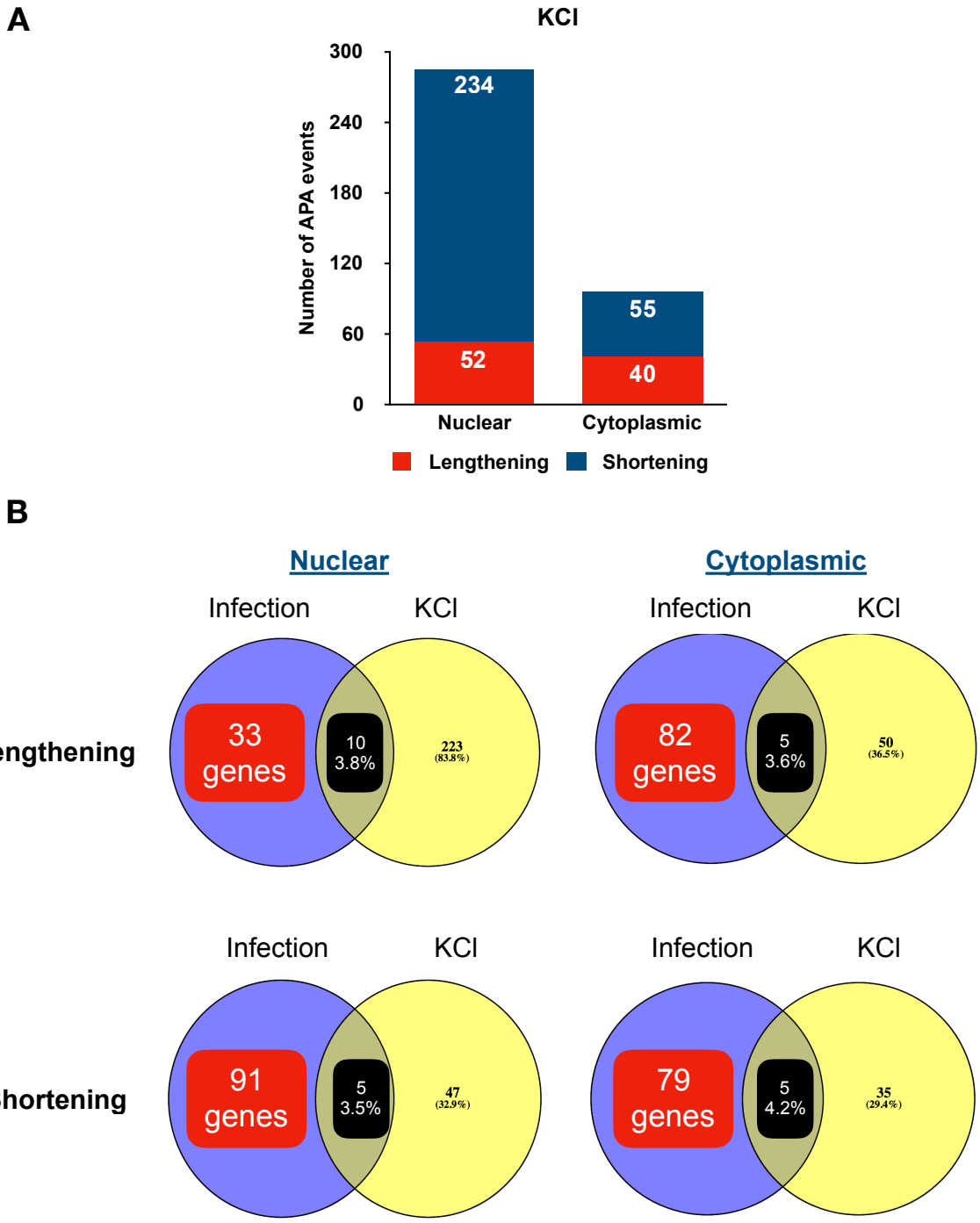


Figure 4.19 Viral-induced CR-APA are Independent of those Induced by Osmotic Stress

A. Global 3'UTR APA events during osmotic stress in both nuclear in and cytoplasmic fractions: Bar charts display the number of events showing significant differential CR- APA regulation in nuclear and cytoplasmic fractions over 3 independent biological repeats; number of lengthening events represented by red, number of shortening events represent by blue. Significant differentially represented APA isoforms are those in which $p<0.01$ (Fisher's Exact Test) with a minimal change in isoform abundance change of ≧5% relative to total APA isoform associated with that gene. The total detectable genes were 13706 in nucleus and 13481 in cytoplasm (Supplementary Figure 7, see scatterplots grey dots).

B. Venn diagrams showing number of overlaps between infection and osmotic stressed induced CR-APA genes in categories of Nuclear lengthening, shortening, and cytoplasmic lengthening, shortening. Unique infection induced APA is on the left (red box). Overlaps of genes are shown in the middle of the Venn diagram (black box).

To summarise 4.2.3, significant CR-APA profile changes are observed after influenza infection in both host nuclear and cytoplasm fractions. The changes seen in genes were highly relevant for viral host relationship, similar to 3'UTR APA events. Research done on each individual genes listed in table 4.2 revealed 3 genes DHX30, PCBP2 and IFNAR2 that potentially have a physiological impact during infection especially on immune response. Despite the majority of viral induced CR-APA were distinct from osmotic stress response overlapping, yet CR-APA changes observed during infection in IFNAR2 were identical to those observed during osmotic stress. Finally since the project was limited by time, further experiments will need to address the question of whether truncation of DHX30 and PCBP2 proteins will impact protein expression and what impact will this have on viral replication.

4.3 Discussion

The aim of this study was to investigate the global changes in APA in response to influenza virus infection in A549 cells. Thus, a global approach with sequencing coupling with traditional biochemistry techniques was implemented. Subcellular fractionation used to study APA was utilised to distinguish between active and passive APA, and also to provide interesting insights on infection dynamics.

The kinetics of influenza virus infection and composition of the host transcriptome were captured during infection time course and analysed. First, analysis by 3'Seq observing the dynamics of viral RNA synthesis revealed viral mRNA abundance peaked at 9 hpi in the nucleus and 6 hpi in the cytoplasm. Previously studies in human HEK293T cells showed similar viral mRNA, cRNA and vRNA synthesis profiles, however the mRNA levels peaked earlier in infection (4 to 6 hpi) compared to A549 cells (Robb et al., 2009; Vreede et al., 2011). Thus, this implies that dynamics of influenza A/WSN/33 (H1N1) virus RNA synthesis may be cell line dependent. Further analysis revealed that a significant proportion of total

mRNA in the cell was viral especially in the nucleus (peaking at 29%), and in the cytoplasm (peaking at 17%). This is in agreement with a previous study done in A/Puerto Rico/8/1934 (H1N1) infected A549 cells, whereby an even higher fraction (57.3%) of mRNA in the cells are viral (Bercovich-Kinori et al., 2016). Since 3'Seq could only sequence polyadenylated transcripts, which therefore cannot determine cRNA and vRNA expression, primer extensions specific to viral NA were performed to address this issue. Whilst the levels of vRNA remained stable in the nucleus and cytoplasm after 9 hpi, a sharp decline in viral mRNA levels after 6 hpi in the cytoplasm was observed, but not in the nucleus. This phenomenon was also observed when sequence alignment to 8 specific viral gene segments was performed. Similar to the findings of this chapter, a reduction in viral mRNA levels was also initially observed during pulse-labelling in BHK21 cells (Shapiro et al., 1987) and by primer extension in HEK293T cells (Robb et al., 2009; Vreede et al., 2011). The reduction in viral mRNA can be explained by a host shut off which results from reduction of host transcription and therefore also viral transcription. However, whole cell RNA was used and analysed in these studies. By utilising subcellular fractionation to observe both nuclear and cytoplasmic mRNA levels, this chapter provided another dimension of insight on the possibilities of viral mRNA nuclear export failure, which the previous studies were not able to. Nuclear export of IAV mRNAs is dependent on ongoing RNA polymerase II activity (Amorim et al., 2006). Since RNA Polymerase II degradation has been shown to occur later in infection (Rodriguez et al., 2007; Vreede and Fodor, 2010; Vreede et al., 2010; Bauer et al., 2018), therefore this can affect viral mRNA export in addition to inducing host shut-off. Furthermore, NS1 is heavily involved in inhibiting viral mRNP export by forming an inhibitory complex with NXF1/ p15 (Satterly et al., 2007; Wang et al., 2008). Therefore, as viral gene expression progresses, the failure of export of viral mRNAs could be explained by increasing NS1 levels (Figure 4.3A). Additionally, future experiments will be required to address whether the failure of export is limited to viral RNA

or also includes host transcripts, and ultimately determine whether this is a host mediated response mechanism or a collateral consequence of the viral induced host shut-off. Additionally, nascent RNA sequencing analysis could be used to provide further evidence of whether the host transcriptional rate is reduced at a similar rate as viral transcription providing more information on host shut-off. Viral protein expression was also analysed, whereby expression begins from 6 hpi and peaking at 9 hpi. Similar to RNA synthesis profiles, protein levels declines at later stages of infection with the exception of M1. Altogether regarding the results of infection dynamics together with immunofluorescence microscopy allowed a sound prediction that any effects of influenza infection should be seen from 6 hpi onwards.

Next in this chapter I also elucidated influenza A virus infection induced APA of host transcripts. Overall, a preference for lengthening of 3'UTRs was seen, however it is important not to discount the fact that there were many shortening events. Interestingly, shortening of 3'UTR has been reported in cells infected by vesicular stomatitis virus (VSV) (Jia et al., 2017). Since whole cell RNA was used as oppose to the subcellular fractionation strategy, this may be a contributing factor for different 3'UTR APA trends when compared with our study. Despite extracting significant APA changes, this analysis does not indicate whether the detected APA events were specific for each time point or if they were events that persist throughout the infection period. Therefore, I applied a highly stringent selection that allowed isolation of genes that exhibited 3'UTR APA in at least 3 infection time points between 6-18 hpi, thus facilitating downstream comparisons. Since the emphasis of this chapter was to discover the physiological effect of APA, a focus on genes that exhibited changes in CPA site usage and differential gene expression was essential. Differentially expressed genes, also stringently selected for being significant at a minimum of 3 infection time points, were overlapped with the aforementioned APA gene

clusters (nuclear lengthening, shortening and cytoplasmic lengthening and shortening). A significant proportion of 3'UTR APA genes exhibited differential gene expression. Interestingly, down-regulated genes that exhibited APA changes were significantly associated with nuclear export, positive type1 interferon production regulation, negative regulation of viral release and protein localisation. Up-regulated APA genes were enriched for neutrophil mediate immunity, mRNA splicing, processing, metabolism and positive regulation of viral life cycle. Evidently, influenza virus infection via APA enhanced gene expression can therefore potentially alter immune response and cellular processes to favour viral replication and host shut off. Furthermore, there were commonalities of GO enrichment with other virus infections such as from VSV, which at 8-16 hpi APA genes were enriched for protein transport, localisation, regulation of cellular immune response (Jia et al.,2017); similar to HSV APA genes were associated with chromatin modification, intracellular transport and regulation of transcription (Hu et al., 2016). Genes (HEXIM1, DDX5, MAVS, SMAD3 and TRIM25) were then chosen as examples for further analysis. The selection was based on the GO analysis, their role in viral host interactions, and most importantly the fact that they undergo APA during viral infection which appears to affect their expression levels.

HEXIM1 has multiple functions depending on the binding partner that may exhibit either a pro-viral or and anti-viral function toward influenza infection. When HEXIM1 is partnered with 7SK RNA, which has been shown to inhibit P-TEFb (Yik et al., 2003; Michels et al., 2004; Dibarne and Bensaude, 2009), this increases the chance of RNA Polymerase II pausing during transcription initiation (Zhou et al., 2012; Jonkers et al., 2015; Li et al., 2017). Thus, this creates an opportunity for IAV to cap snatch and subsequently help IAV transcription. On the other hand, HEXIM1 when partnered with NEAT1 regulates innate immunity and therefore can have a negative effect on viral transcription and replication.

Since it is up-regulated in gene expression and have 3'UTR lengthening after infection, it will be interesting to see whether its up-regulation will affect viral transcription and replication efficiency. However, our western blots (Figure 4.12A) indicate that despite the increase in transcript levels HEXIM1 proteins are not elevated in infected cells and expression of HEXIM1 (Figure 4.13) had no detectable effect on viral replication. Thus, the above mentioned hypothesis that elevated levels of HEXIM1 may create a favourable host environment for the virus has to be discarded. However, since HEXIM1 levels are reduced it will be necessary to evaluate whether reduced levels have an impact in viral replication. Unfortunately these experiments could not be completed due to time constraints.

DDX5 has been shown to be up-regulated and lengthening in 3'UTR APA after infection. Although it has an antiviral function during hepatitis B virus and myxoma virus infection (DNA viruses) (Zhang et al., 2016; Rahman et al., 2017), its been shown promote viral replication in RNA viruses including severe acute respiratory syndrome (SARS) coronavirus (Chen et al., 2009), human immunodeficiency virus 1 (HIV-1) (Zhou et al., 2013), hepatitis C virus (Goh et al., 2009), Japanese encephalitis virus (Li et al., 2013), porcine reproductive and respiratory syndrome virus (PRRSV) (Zhao et al., 2015), and influenza virus (Bortz et al., 2011). It has been shown that knockdown of DDX5 and DDX17 has led to a significant reduction in viral titre (Bortz et al., 2011). Additionally, it has been speculated that influenza viruses have evolved to hijack host DDX5 in order to facilitate nuclear export of viral proteins (Cheng et al., 2018). Since DDX5 expression is up-regulated during infection, an experiment was designed to over-express DDX5 in HEK293T cells. Since DDX5 is reported to have a pro-viral function, I was expecting to see an increase in viral RNA expression after over expressing DDX5. However, this was not the case where primer extensions performed by Ben Nilsson did not yield a change in viral RNA expression (Figure 4.13). Perhaps, one possible assumption I can make is that

influenza infection increase DDX5 expression in a way to maintain DDX5 protein expression at a critical level, but no further impact on viral replication and transcription can be seen once this critical level of DDX5 expression is reached. Alternatively, DDX5 over expression was not able to alter viral RNA expression as flag tagging proteins can potentially abrogate protein function and localisation.

MAVS is widely known to be involved with innate immunity especially in the context of influenza. Both retinoid acid-inducible gene 1 (RIG-I) and melanoma differentiation-associated gene 5 (MDA-5) can recognise IAV RNA and bind to MAVS (Pichlmair et al ., 2006; Reikine et al., 2014), consequently activating inhibitor of nuclear factor kappa-B kinase (IKK) and TANK-binding kinase 1 (TBK-1). This allows for activation of interferon regulatory factor 3/7, and initiate the transcription of IFN and IFN-induced genes (Akira et al., 2006). Nuclear factor-kappaB (NF-kB) is activated too. During infection, MAVS was down-regulated in gene expression. In vivo studies have shown viral PB2 interacts with MAVS and is able to inhibit IFN-β expression (Graef et al., 2010; Varga et al., 2012). However, since this chapter demonstrated MAVS 3'UTR changes were caused by failure of transcriptional termination via RT-qPCR, this may provide another way for the virus to shut down innate immunity. Although, this was not an official APA event, but influenza infection still induced a shift from the usage of the CPA site and decreased gene and protein expression. Perhaps another clue that points towards termination defect of MAVS was the fact that the 'long' 3'UTR isoform seem to be retained in the nucleus and did not penetrate though to the cytoplasm. Whole transcriptome RNA sequencing is needed to resolve this. Furthermore, I have confirmed observations from the study where MAVS had been shown to have a single poly(A) site and exhibited decrease in mRNA abundance upon VSV viral infection (Jia et al., 2017). Thus, this provides further evidence that viruses may use APA as a conserved mechanism to shut down immune responses.

SMAD3 acts as a mediator of the signals initiated by the transforming growth factor beta (TGF-β) superfamily of cytokines, which regulate cell proliferation, differentiation and death (Moustakas et al., 2001). SMAD3 plays a major role in TGF-β mediated growth inhibition (Li et al., 2006; Park et al., 2016). Down-regulation of gene expression was observed, thus as shown by a study involving RNAi depletion of SMAD3 but not SMAD2 sufficiently blocked TGF-β mediated cell cycle arrest and growth inhibition (Kim et al., 2005). This is interesting as the literature has shown that influenza virus infection induces cell death (Takizawa et al., 1993; Kash et al., 2006b). Conversely, a recent study revealed that pandemic H1N1 influenza A suppresses immunogenic RIPK3-driven dendritic cell death (Hartmann et al., 2017).

TRIM25 has been shown to be a type 1 and type 2 IFN-inducible E3 ligase (Carthagena et al., 2009), thus it has a role in antiviral defence. However, a study has stated TRIM25 has a dual role in RIG-I regulation, whereby it has been shown to negatively regulate RIG-I through stabilisation of FAT10 (Nguyen et al., 2016). 3'UTR shortening and down-regulation of gene expression was observed. Contrary to findings from a VSV study, where a luciferase reporter gene assay demonstrated the use of the shorter 3'UTR isoform resulted in increased protein coding potential (Jia et al., 2017), this chapter revealed a down-regulation of gene expression. The effects of 3'UTR shortening of TRIM25 on protein expression will need to be clarified by westerns or luciferase reporter assay experiments. Furthermore, functional tests such as knockdowns or over expression experiments will be need to confirm TRIM25 effects on influenza infection.

Whilst these are interesting genes that display significant APA and differential gene expression during infection, a recent study done by Bauer and colleagues have shown influenza infection viral stress correlates highly with osmotic shock (Vilborg et al., 2015),

heat shock (Mahat et al., 2016) and oxidative stress (Giannakakis et al., 2015) in terms of termination defects of pol II transcribed genes. Out of the 5 genes selected above, SMAD3 and TRIM25 APA profiles from infection were identical with those observed in osmotic shock. Thus, downstream analyses were therefore focused on DDX5, HEXIM1 and MAVS since APA of these genes were shown to be unique to influenza infection.

Furthermore, the link between APA events and differential gene and protein expression needed to be established. Thus western blots were performed to ensure that genes effected by APA resulted in change in protein expression, which was the case for DDX5, HEXIM1 and MAVS. Additionally, I ensured that differential protein expression of these genes were due to infection but not by protein half-life turnover by utilising protein pulse-labelling data from the Schwanhäusser paper. Since the infection time course spans 18 hours, DDX5 with protein half life of 19 hours, HEXIM1 with protein half life of 28 hours, and MAVS with protein half life of 78 hours (Schwanhäusser et al., 2011) signifies that changes were likely to be caused by infection. Since the alternative 3'UTR can harbour cis-regulatory elements such as miRNA target sites, destabilising elements and RNA binding protein sites, the shift in usage of the alternative 3'UTR can potentially affect stability, translational efficiency, nuclear export and localisation (Neve et al., 2017). To assess the translational output of the different isoforms, a luciferase reporter assay strategy was performed confirming that the shift in APA isoforms in DDX5 increases protein output (only data for DDX5 was shown). Altogether with 3'seq, RT-qPCR, western blots and luciferase assays, DDX5 was shown to shift the usage of the distal CPA in both nucleus and cytoplasm, signifying this is an active APA event whereby APA of DDX5 is resulted from change in CPA site choice and not from post-transcriptional events. Luciferase assays have shown lengthening of DDX5 led to an increase in protein output which is concomitant to evidence of increased protein expression shown in western blots.

Thus, these further supports our original hypothesis that some influenza induced APA events can lead to changes in gene and protein expression and have a potential physiological impact on host-shut off and viral replication.

Since influenza has been shown to affect 3'UTR APA, it would be logical to see if influenza infection could cause CR-APA changes. Observations of global CR-APA during infection, although fewer when compared to 3'UTR APA, identified a significant number of events. The comparison of the infected to non-infected nuclear fractions revealed the majority of the events have a tendency of CR-APA shortening. Similar to 3'UTR APA analysis, selection was performed to isolate genes that exhibit CR-APA at a minimum of 3 indicated infection times (6-18 hpi). GO analysis revealed that biological processes associated with mRNA splicing, mRNA processing and cell cycle regulation were also overlapped with 3'UTR-APA, suggesting that influenza infection induces both 3'UTR and CR-APA to target similar biological processes.

Next in order to address whether CR-APA has a physiological impact, CR-APA genes were compared with differentially expressed genes at a minimum of 3 time points. Significant proportion of genes that exhibited CR-APA also resulted in differential gene expression. Overlapped genes, likely relevant to viral infection and response, were selected from the list (table 4.2). The 3 genes were DHX30, IFNAR2 and PCBP2. They all have commonalities in regulating immune response. Both DHX30 and IFNAR2 have positive roles in innate immunity against viral infection (Ye et al., 2010; Zhang et al., 2017). Since both DHX30 and IFNAR2 have exhibited CR-APA shortening and down regulation of gene expression, the shorter isoforms would encode for truncated proteins, and potentially alter their function in regulating innate immunity against viral infection. Interestingly, truncation mutations of DHX30 were performed to map the binding domains of DHX30 to zinc-finger

antiviral protein (ZAP), whereby the N- terminal is required for binding ZAP while the core domain and C- terminal domain alone failed to interact with ZAP (Ye et al., 2010). Conversely PCBP2, which negatively regulates MAVS antiviral pathway, has been shown to also exhibit CR-APA shortening and is up-regulated. Perhaps, by regulating CR-APA during infection, this can potentially provide another way for the virus to shut down IFN responses. Further evidence will be needed in the future to address the following questions: 1. Does CR-APA cause differential gene and protein expression in these genes? 2. What are the effects of CR-APA induced truncation of protein isoforms on function? 3. Do these events affect viral replication? Western blots, luciferase reporter gene assays, and truncation mutations of these genes will need to be performed to elicit a clear functional effect of CR-APA during infection.

Finally, overlaps between CR-APA genes induced from infection were compared with CR-APA genes induced from osmotic shock. Osmotic shock induces many CR-APA events, but very little overlap has been observed. Out of the 3 genes, only CR-APA of IFNAR2 was significant in both osmotic shocked and infected cells. Thus, future experiments mentioned above will be focused on the PCBP2 and DHX30.

4.4 Summary

Overall this chapter has addressed the 3 questions:

1. Are host nuclear and cytoplasmic APA profiles changing during influenza virus infection?
2. If influenza infection causes host gene APA, do these changes have any physiological consequence and to what extend do they change the host virus relationship?
3. If influenza infection causes host gene APA, to what extent are these events triggered by viral factors or are they simply part of a generic host response to stress?

The first question was answered by demonstrating that influenza infection has been found to influence APA nuclear and cytoplasmic profiles through 3'seq, and using RT-qPCR to verify that some of these APA events are true. Next I addressed the second question by showing almost half of 3'UTR-APA events caused by infection were differentially regulated in gene expression. Indeed through GO and individual gene analysis, these APA and differentially expressed genes are strongly implicated in regulating host viral interactions and may be a general means by which viruses induce a state of transcriptional shut-off in infected cells. From this, I can also speculate that perhaps APA can act as an intermediary state to cause host shut-off. DDX5 is a prime example of a gene that addressed the 3 questions whereby DDX5 3'UTR APA is induced by influenza infection, which leads to differential gene and protein expression. Even though expressing DDX5 did not result in increased viral RNA expression, I cannot ignore that fact that DDX5 knockdown can lead to reduced viral titres (Bortz et al., 2011). Although progress had been made through luciferase reporter gene assays, western blots and target gene over-expression, further functional analysis such as knockdown experiments, luciferase report gene assays and truncation mutations of target genes of physiologically relevant genes including those that are regulated by CR-APA will be performed to provide more insight on whether they will have an effect on infection. Finally, I have demonstrated that the majority of APA events induced by influenza infection are more likely to triggered by viral factors than simply being a part of a generic host response to stress.

In the next chapter, I would like to investigate the mechanism behind viral-induced APA.

Chapter 5: Deciphering Mechanisms of 3'UTR APA Regulation During Influenza Viral Infection

5.1 Introduction

As demonstrated in the previous chapter, influenza viral infection of A549 cells with a high multiplicity of infection exhibited significant APA events, which was accompanied by the change in protein output of a number of genes that are implicated in host defence to infection and viral propagation. The next step involves establishing the molecular mechanisms that trigger APA in these genes upon influenza infection.

An obvious starting point is the well-known influenza non-structural protein 1 (NS1). NS1 has been shown to interact with the cleavage and polyadenylation specificity factor 30 (CPSF30) and compromise 3'end processing as summarised in detail below. NS1 thus provides an obvious link between the viral infection and APA in particular since in the previous chapter I established that host APA is most prolific at infection time points where high levels of NS1 are present in the cell.

5.1a The Relationship between the Viral Non-Structural Protein 1 and the Host Cleavage and Polyadenylation Machinery

During cleavage and polyadenylation, the cleavage and polyadenylation complex is involved in recognising the poly(A) signals present in a pre-mRNA including the A(A/U)UAAA hexamer, U/GU-rich downstream element, cleavage site and other auxiliary sequences as summarised in chapter 1. The position of 3'end formation depends on specific cis-elements that are present within the pre-mRNA. Arguably the key cis element is the conserved canonical AAUAAA located 10-30 nucleotides upstream of the cleavage site. Previous studies have shown 70% of all CPA sites contain this canonical hexamer

(Neve et al., 2016). Additionally, mutations of any nucleotide of AAUAAA could reduce or abolish mRNA 3'end cleavage and processing, with the exception for AUUAAA whereby (16% of all CPA sites possess this hexamer (Tian et al., 2005)) 66% cleavage activity and 77% polyadenylation activity was retained despite mutations (Sheets et al., 1990; Proudfoot et al.,2011). The cleavage and polyadenylation specificity factor (CPSF) protein complex, containing the six subunits (CPSF160,WDR33, CPSF100, CPSF73, Fip1 and CPSF30), is responsible for recognising the conserved hexanucleotide, however CPSF160, CPSF30, WDR33 and hFIP1 proved to be sufficient to direct polyadenylation in vitro (Schönemann et al., 2014). Moreover, CPSF30 and WDR33 directly and synergistically associate with the AAUAAA hexamer and are required for 3' end processing (Shi et al., 2014; Sun et al., 2018). CPSF30 is the smallest subunit of CPSF complex with a molecular weight of 30 KDa. It has 5 zinc finger domains, and a zinc knuckle domain which is responsible for enhancing RNA binding activity (Barabino et al., 1997). Cryo-electron microscopy structure of the quaternary complex of human CPSF-160, WDR33, CPSF30 and an AAUAAA RNA revealed how CPSF30 specifically recognises A1, A2, A4, and A5 of the poly(A) signal where zinc finger 2 (ZF2) interacts with A1 and A2 and zinc finger 3 (ZF3) recognises A4 and A5 bases. Interestingly, U3 and A6 of the hexamer directly contacts WDR33 by forming an intramolecular Hoogsteen base pair, whilst CPSF160 acts as a scaffold (Sun et al., 2018).

NS1 is a multifunctional protein that allows the virus to antagonise host immune responses and replicate effectively (Krug, 2015). Structurally, it has a RNA binding N-terminal domain and a C-terminal effector domain (Figure 5.1).The N-terminal domain's common functions are related to counteracting host antiviral by inhibiting IFN type 1 production (reviewed in Hale et al., 2008). For instance, it has been shown to inhibit activation of interferon regulatory factor 3 (Talon et al., 2000), NF-kB (Wang et al., 2000), IFN 1 induced protein kinase RNA activated (PKR) (Bergmann et al., 2000), and avoid recognition and activation of 2', 5'-oligo(A) synthetase- RNAse L pathway. Additionally, NS1 has been shown to avoid RIG-I mediated induction of IFN 1 by interacting with RIG-I (Mibayashi et al., 2007) and TRIM25 (Koliopoulos et al., 2018). Perhaps more importantly and relevant to this chapter, the C-terminal effector domain of NS1 is responsible for inhibiting 3'end processing of cellular pre-mRNA by interacting with CPSF30 as one mechanism to suppress host antiviral gene expression (Nemeroff et al., 1998; Noah et al., 2003).

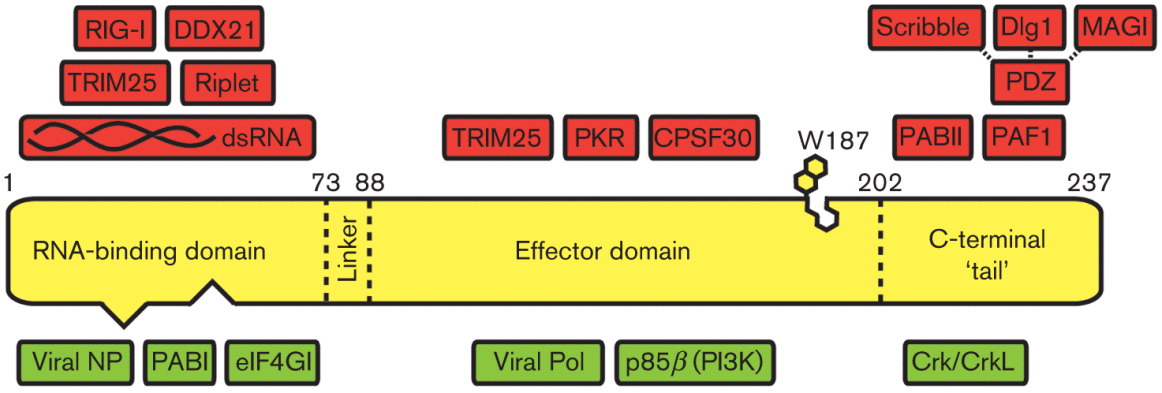


Figure 5.1 Overview of influenza A Virus non-structural 1 (NS1) protein and Interactions.
Linear schematic of NS1 domain layout and interactions. NS1 protein (yellow) split into four distinct regions: the N-terminal RNA binding domain (RBD), the linker region (LR), the effector domain (ED) and the C-terminal 'tail'. Both LR and C-terminal tail vary in length among strains: longest total NS1 length is 237aa, but 230aa is also common. NS1 interacts with many viral and cellular factors indicated in red to inhibit cellular antiviral defences, while NS1 stimulates the 'green' factors in order to enhance viral replication. (RIG-I, retinoic acid inducible gene I; DDX21, DEAD box helicase 21; TRIM25, tripartite motif containing 25; PKR, double stranded RNA activated protein kinase; PABI, poly(A) binding protein I; eIF4GI, eukaryotic translation initiation factor 4 gamma, 1; Pol, Polymerase). Figure was adapted from Hale et al., 2008.

Recently, NS1 alone was shown to be responsible for host transcription termination defects caused by influenza virus infection through its interaction with CPSF30 (Bauer et al., 2018). Additionally, termination defects can be enhanced by SUMOylation of 1918 H1N1 NS1 and suppress host transcriptional response during infection (Zhao et al., 2018). However more notably, since ZF2 and ZF3 of CPSF30 interacts with AAUAAA hexamer, NS1A has been shown to block ZF2 and ZF3 and thus hinders RNA host 3'end processing (Twu et al., 2006). Furthermore, CPSF30 plays an important role for regulation of CPA site usage in mammalian cells and has been shown to be involved in regulating APA in plants (Delaney et al., 2006; Thomas et al., 2012). Thus, it is logical to hypothesise NS1A's potential ability to affect APA via CPSF30 abrogation. Linear schematic of the NS1 domain layout and interactions were summarised in figure 5.1.

5.1b Aims

The aim of this chapter is to decipher the mechanism of host gene APA regulation in influenza viral infection. The main question is whether NS1 expression and abrogation of CPSF30 is a key contributor for host gene APA. To achieve this, I decided to evaluate APA in host cells that have been transfected with siRNAs targeting CPSF30 and so mimicking the NS1 mediated reduction of CPSF30 in infected cells. Furthermore, to separate the potential impact of NS1 on APA from any other viral factors, I also tracked APA profile changes in cells that have been transfected with a plasmid expressing NS1. To achieve maximum transfection efficiencies, these experiments were conducted using HEK293 cells. The main objective of this set of experiments was to determine the role that NS1 plays in modulating host APA profiles. The outcome of these approaches should clarify whether the interactions between NS1 and CPSF30 are involved in the regulation of host gene APA and unravel some of the strategies employed by influenza A/WSN/ 33 virus that contribute to host shut-off and viral propagation.

5.2 Results

5.2.1 3'UTR APA Profile Changes After CPSF30 Knockdown

5.2.1a Ensuring Efficient Knockdown of CPSF30

As outlined above, CPSF30 is a key component of the host cleavage and polyadenylation machinery and has been reported to be involved in regulating APA in plants (Thomas et al., 2012). Therefore, I first aimed to address whether CPSF30 also plays an important role for regulation of poly (A) site usage in mammalian cells. For this a siRNA-approach was utilised as described to silence CPSF30 gene expression in HEK293T cells. The rationale behind using HEK293 cells was due to the low transfection efficiencies associated with A549 epithelial cells. To maximise the knock down, I used two "hits" of CPSF30 or scrambled siRNAs. Cells were then subjected to subcellular fractionation to isolate nuclear and cytoplasm RNA fractions. Efficiency of the subcellular fractionation was confirmed by RNA gel detecting for nuclear precursor rRNAs (45S, 32S) which were only observed in the nuclear fractions, while processed rRNAs (28S,18S, 5S) were also found in the cytoplasmic fractions (Figure 5.2A). Protein from whole cell lysate were also extracted and analysed by western blots, verifying that a significant decrease in CPSF30 knock-down protein expression compared to the scrambled control was achieved (Fig. 5.2B). Quantification of western blots was performed on ImageJ, where the band intensity values of CPSF30 knock-down or scrambled control were normalised to those of their respective HSP60 loading control intensity to determine the CPSF30 knock-down and control ratios. Subsequently, the normalised CPSF30 knock-down ratios were compared to the normalised scrambled showing a 61% reduction was achieved. Thereafter, analysis was performed on RNA from three independent knock down experiments.

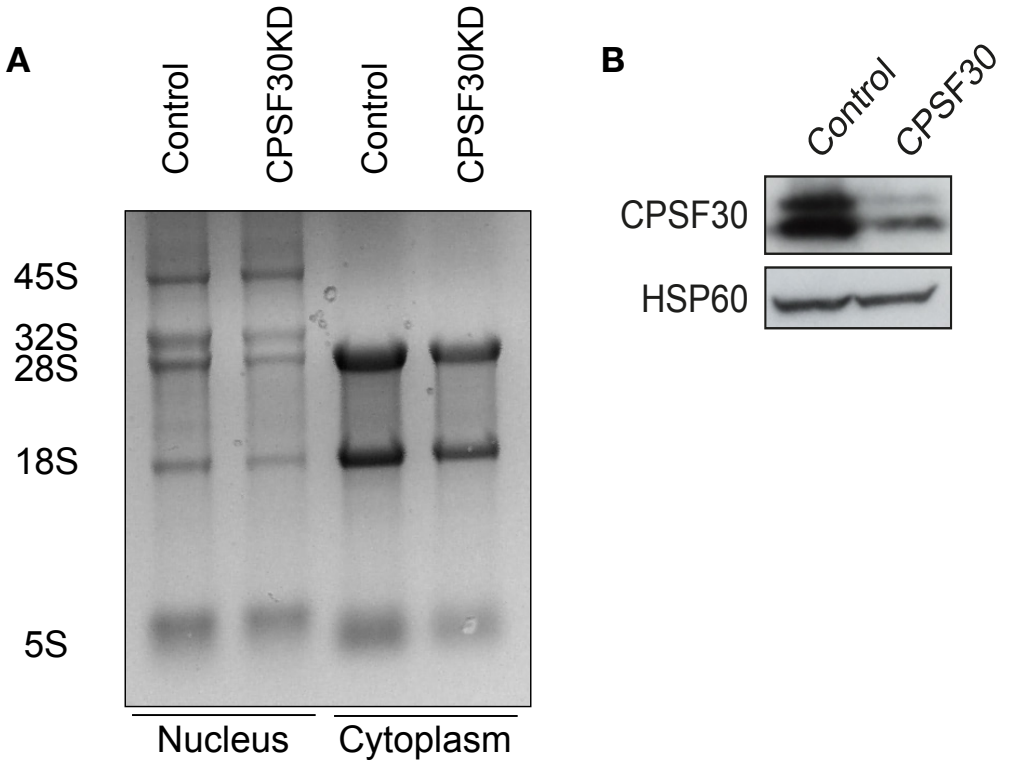


Figure 5.2 Subcellular Fraction and Western Blot Controls of CPSF30 Knockdown

A. Subcellular fractionation control of scrambled control and CPSF30 knock-down experiment. Denaturing RNA gel with 2 µg of RNA from nuclear and cytoplasmic fractions. The identity of each ribosomal RNA band is displayed on the left-hand side of the gel.

B. Efficient knockdown of CPSF30 was verified by performing western blot. The same membrane was probed with anti CPSF30 and HSP60 (loading control). The control refers to scrambled control siRNA treated HEK293T cells. Band intensity values, quantified from ImageJ, of CPSF30 knock-down or scrambled control were normalised to those of their respective HSP60 loading control intensity values to obtain normalised CPSF30 knock-down and control ratios.

5.2.1b Investigating CPSF30 Knockdown Effects on 3'UTR-APA in HEK293T cells

To that end, libraries for sequencing from the different fractions were prepared using the 3'Quant Lexogen kit, library quality and quantity were verified using a High Sensitivity dsDNA Bioanalyzer kit, and sequencing of libraries was performed on the Ion Proton platform. Utilising the developed bioinformatic pipeline and combining reads from 3 independent nuclear biological repeats and 2 independent cytoplasmic biological repeats, the APA profiles were established. The comparison between wild type and CPSF30 knock down cells captured a total of 6582 3' UTR APA events in the nucleus and 1096 of them (16.7%) showed statistically significant changes in poly(A) site usage. Similarly, for the cytoplasmic fraction, 5886 APA events were identified and 921 (15.6%) of them showed statistically significant relative changes in poly(A) site usage. Interestingly for both the nuclear and cytoplasmic fractions, CPSF30 knock-down resulted in profiles shifts that showed a clear tendency for preferential usage of the distal over the proximal poly(A) site thus causing overall lengthening of UTRs relative to wild type cells. The lengthening was particularly evident in the cytoplasmic fraction, where 85% of all statistically significant APA changes were associated with UTR lengthening (Figure 5.3A). These results interestingly contradict with other knockdown studies and studies that worked with Arabidopsis thaliana CPSF30 mutants. It has been shown that knock-down of CPSF30 had a modest effect on APA (Li et al., 2015), and the mutant CPSF30 induces a large number of APA events without a clear preference for proximal or distal CPA site usage during response to oxidative stress (Thomas et al., 2012).

Next, I wanted to verify whether there are any correlations between infection and CPSF30 KD. Thus, comparison of genes exhibiting APA in CPSF30 knock-down between APA genes during viral infection at a minimum of three time points were analysed and yielded 4

main categories: Nuclear lengthening, nuclear shortening, cytoplasmic lengthening, and cytoplasmic shortening (Figure 5.3). Although there were a significant amount of common genes, the majority of APA events were affected uniquely in one or the other conditions. Starting on the top left panel of figure 5.3B, 112 genes (11.5%) overlapped between infection and CPSF30 knock-down in the nuclear lengthening gene cluster. 20 (5%) genes were commonly induced in nuclear shortening category shown in the bottom left panel. In the cytoplasmic fractions, 75 (7.5%) of genes overlapped in the lengthening category (top right panel), and 17 (4.7%) genes overlapped in the shortening category (bottom right panel) (Figure 5.3B). Conclusions from these results suggest that abrogation of CPSF30 certainly induces many APA events, but only small subsets of APA genes in infection are regulated by CPSF30 abrogation. Therefore, it would be interesting to interrogate whether these subset of genes are involved in any viral relevant biological processes.

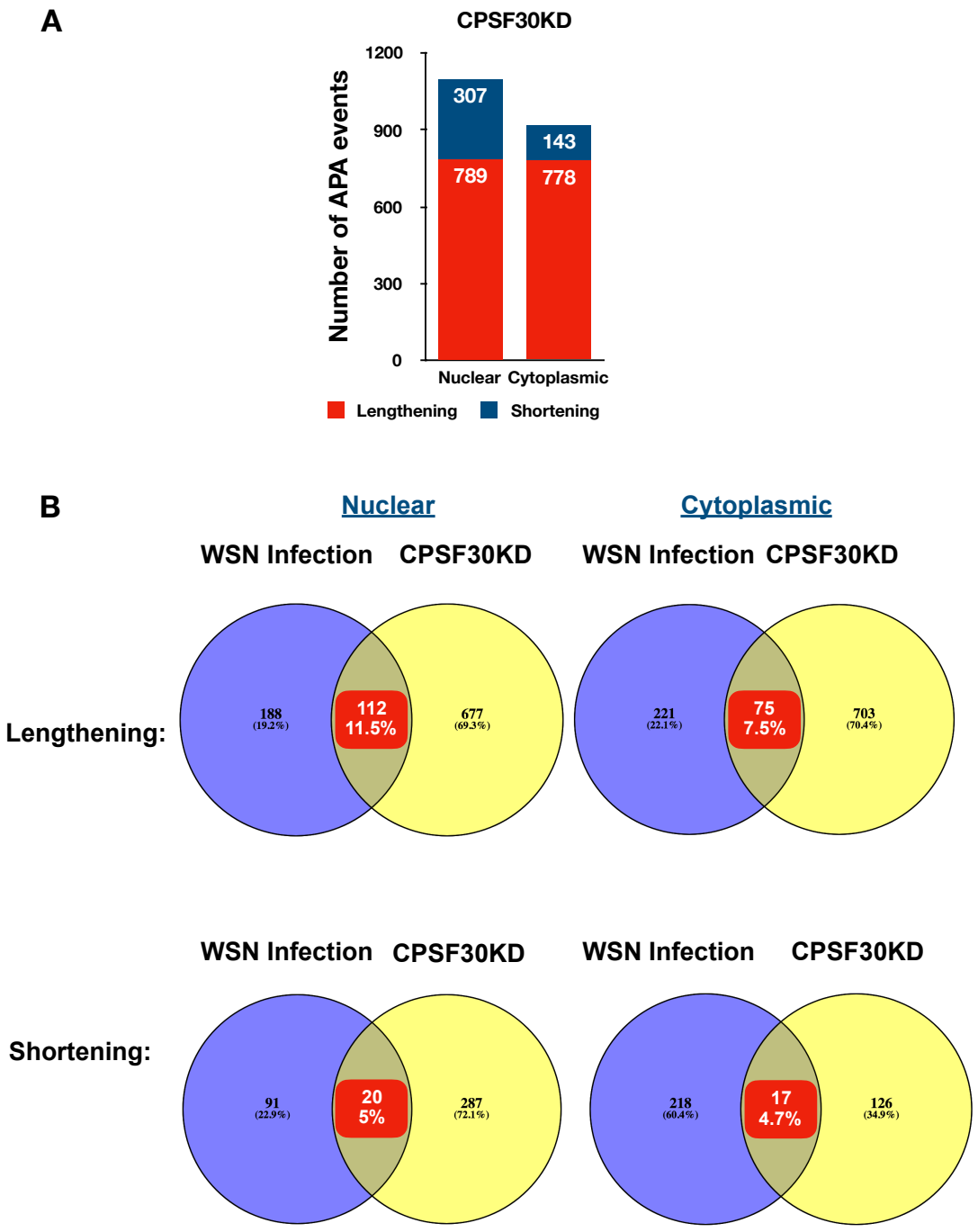


Figure 5.3 Global Analysis of CPSF30 Knock-down 3'UTR APA Followed by Comparisons with A/WSN/33 Infection 3'UTR APA Events.

A. Bar graph to visualise number of significant 3'UTR APA events (Fisher's exact Test, $p<0.01$, with a minimum 5% change in APA isoform abundance) after CPSF30 knock-down in both nuclear over 3 independent biological repeats and cytoplasmic fractions over 2 biological repeats of sequencing. The number lengthening events are presented in red, whereas the number of shortening events are presented in blue. The total detectable genes in the nucleus are 6582. The total detectable genes in cytoplasmic are 5886 (**Supplementary figure 8**). % of total 3'UTR APA events were calculated by dividing total number of significant APA events over total genes detected for each fraction. Proportion (%) of APA events exhibiting lengthening or shortening were also calculated.

B. Venn diagrams displaying number of overlaps between identified 3'UTR APA events between A/WSN/33 infection and CPSF30KD that belong to 4 categories: Nuclear lengthening (top left panel), nuclear shortening (bottom left panel); cytoplasmic lengthening (top right panel), cytoplasmic shortening (bottom right panel). Number of and % of genes

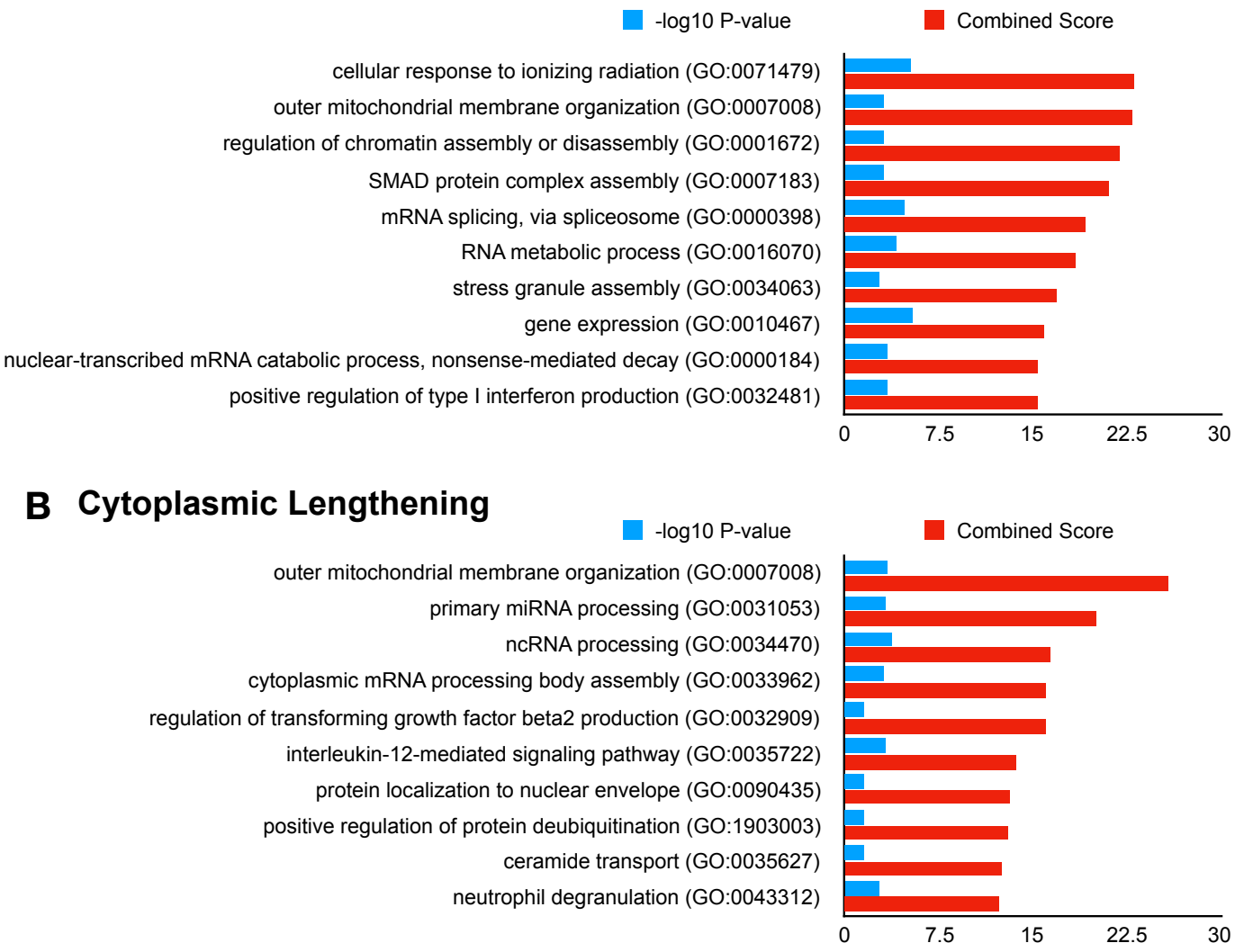


Figure 5.4 Gene Ontology Analysis of Overlapped Genes from WSN Infection and CPSF30KD

The top 10 biological processes gene ontology categories with significant enrichment ($p<0.05$) of the genes were associated with nuclear lengthening **(A)** and cytoplasmic lengthening **(B)**. Those groups that showed an overrepresentation of the proximal APA isoform both in the nucleus and cytoplasm were also analysed but showed no significant enrichment for any categories. Gene ontology analysis was performed using Enrichr.

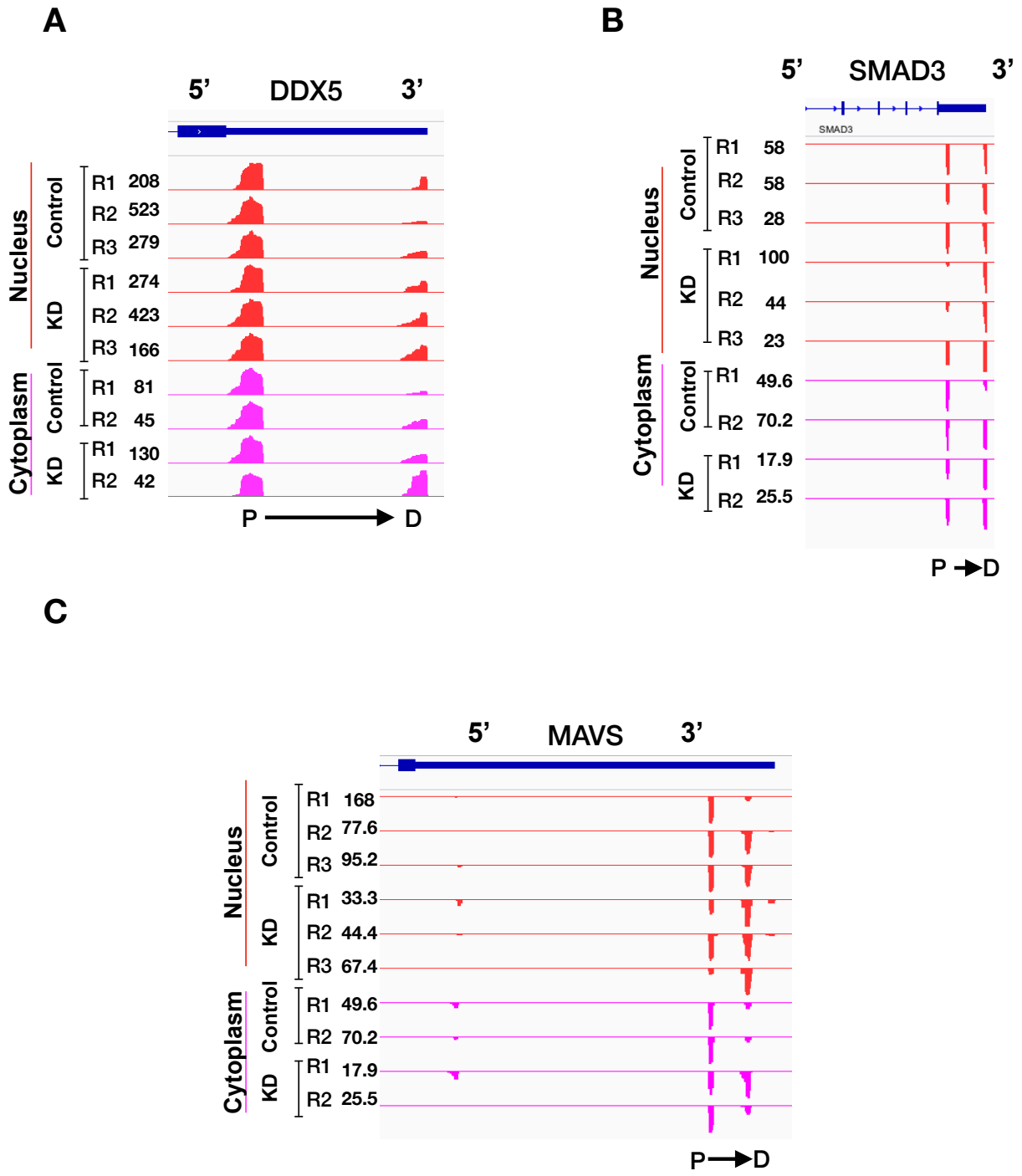


Figure 5.5 Examples of Stringently Selected Genes which Display Differential Regulation 3'UTR APA during CPSF30KD Experiments.

A. Screenshot of DDX5 as an example of a gene that displays significant shifts in CPA site usage to using the distal CPA site in the 3'UTR, when comparing nuclear (red) or cytoplasmic (purple) fractions of CPSF30KD to scrambled control in HEK293T cells.

B. Screenshot of SMAD3 displaying 3'UTR lengthening when comparing nuclear (red) or cytoplasmic (purple) fractions of CPSF30KD to scrambled control in HEK293T cells.

C. Screenshot of MAVS displaying 3'UTR lengthening only when comparing nuclear (red) fractions of CPSF30 knock-down to scrambled control in HEK293T cells

All IGV images are displayed with a 5' to 3' direction. P represents proximal CPA site, D represents distal CPA site. NOTE:

Overlapped genes in each of the 4 categories were subsequently analysed by GO analysis of biological processes. Only nuclear and cytoplasmic lengthening gene clusters showed significant enrichment for specific categories. Genes in the “nuclear lengthening category” after infection and CPSF30 knock-down were strongly associated with mitochondrial membrane organisation, regulation of chromatin assembly or disassembly, SMAD protein complex assembly, mRNA splicing, mRNA metabolism, gene expression and positive regulation of type I interferon production (Figure 5.4A). Cytoplasmic lengthening genes from both infection and CPSF30 knock-down were significantly enriched for similar biological processes apart from miRNA processing, ncRNA processing and protein localisation to nuclear envelope (Figure 5.4B). From this, I were able to determine both infection and CPSF30 knock-down were responsible for inducing APA in a common subset of genes involved in suppressing host response, gene expression and mRNA regulation. Furthermore when comparing to the GO analysis performed in chapter 4, these significantly enriched GO biological processes terms are also homologous to those observed in infection. Therefore I conclude that perhaps these physiological processes are regulated by CPSF30 abrogation during infection.

Referring back to chapter 4, physiologically relevant genes exhibiting differential CPA site usage ($p<0.01$ Fisher’s exact Test, with a minimum 5% change in APA isoform abundance) and gene expression ($p<0.01$, Fisher’s exact Test, Fold Change (FC)> 2) from influenza A infection at a minimum of three time points were selected through GO and individual gene analyses. 5 stringently selected genes with functions relating to immune response and viral interactions: DDX5, HEXIM1, MAVS, SMAD3 and TRIM25 were picked for further analyses. Although HEXIM1 did not exhibit significant APA lengthening in CPSF30 knock-down cells like during infection, the sequencing data indicated that CPSF30 knock-down caused a shift towards the usage of HEXIM1 distal CPA site. APA statistical significance

was determined by -log10(p-value)*s (s represents the direction of significance of up or down regulation of the proximal CPA site), thus a decrease in usage of proximal CPA site will determine it as a minus sign. HEXIM1 statistical significance (SS) value was -1.23, p-value: 0.0589 in the nucleus; SS value was -0.43, p- value: 0.38 in the cytoplasm, thus signifying that it is lengthening but above the cut off I set ($p<0.01$). Similar with TRIM25, whereby 3'UTR shortening was seen during infection, its SS values for nuclear (1.24, $p=0.0575$) and for cytoplasm (0.49, $p=0.32$) signifies it is shortening but just above the cut off during CPSF30 knock-down. However more importantly, DDX5, MAVS and SMAD3 of those genes indeed exhibited significant APA changes during CPSF30 knock-down (Figure 5.5). DDX5 (Figure 5.5A) and SMAD3 (Figure 5.5B) exhibited APA lengthening in both nuclear and cytoplasm fractions in all three replicates. Similar to the trend observed in infection, MAVS only exhibited APA lengthening in the nucleus (Figure 5.5C). Altogether, these results are hugely valuable since this may indicate that inhibition of the core poly(A) machinery CPSF30, possibly via NS1 interaction, will lead to significant APA of the selected genes that are physiologically relevant during viral infection. Therefore the next series of experiments involve NS1 transient over-expression in HEK293T cells to determine whether APA genes, commonly induced by CSPF30 knock-down and infection, are induced via NS1 activity.

5.2.2 Investigating 3'UTR APA Effects of Transient NS1 Expression in HEK293T cells

5.2.2a Establishing Transient Expression of GFP, NS1-G184R and NS1 in HEK293T cells

CPSF30 knock-down has been demonstrated to induce changes in APA regulation. Moreover, particular subsets of physiologically relevant genes (immune response, gene expression and mRNA regulation) that result in significant APA changes are homologous to those that are induced by infection, thus signifying CPSF30's role in modulating APA during infection.

Amongst the numerous interactions as summarized in figure 5.1, the multifunctional viral protein NS1 is known to interact with and abrogate CPSF30 during infection (Nemeroff et al., 1998) I therefore attempted to emulate infection and mock by over expressing NS1 transfecting expression plasmids containing NS1, control GFP (provided by Fodor Lab, Dunn School of Pathology, Univeristy of Oxford) or mutant NS1 into HEK293T cells, and subsequently analysed the effect of NS1 expression on APA and whether these APA events coincide with infection and CPSF30KD experiments. Two controls were used, GFP pcDNA and pcDNA containing NS1 mutant (NS1-G184R), which has been shown to be defective in CPSF30-binding by disrupting F2 F3 binding hydrophobic pocket in the effector domain and thus mimicking the GFP control (Das et al., 2008; Steidle et al., 2010). Firstly, to establish the NS1 mutant expression plasmid, single nucleotide mutagenesis within the NS1 sequence was initiated. Single nucleotide mutation was introduced into the NS1 cDNA sequences in pcDNA-NS1 (Robb et al., 2010) with designed PCR primers (g550a NS1 mut primers shown in table 2.1 *). PCR was performed as described in chapter 2.13.6 with overlapping primers containing the altered sequences (Figure 5.6A).

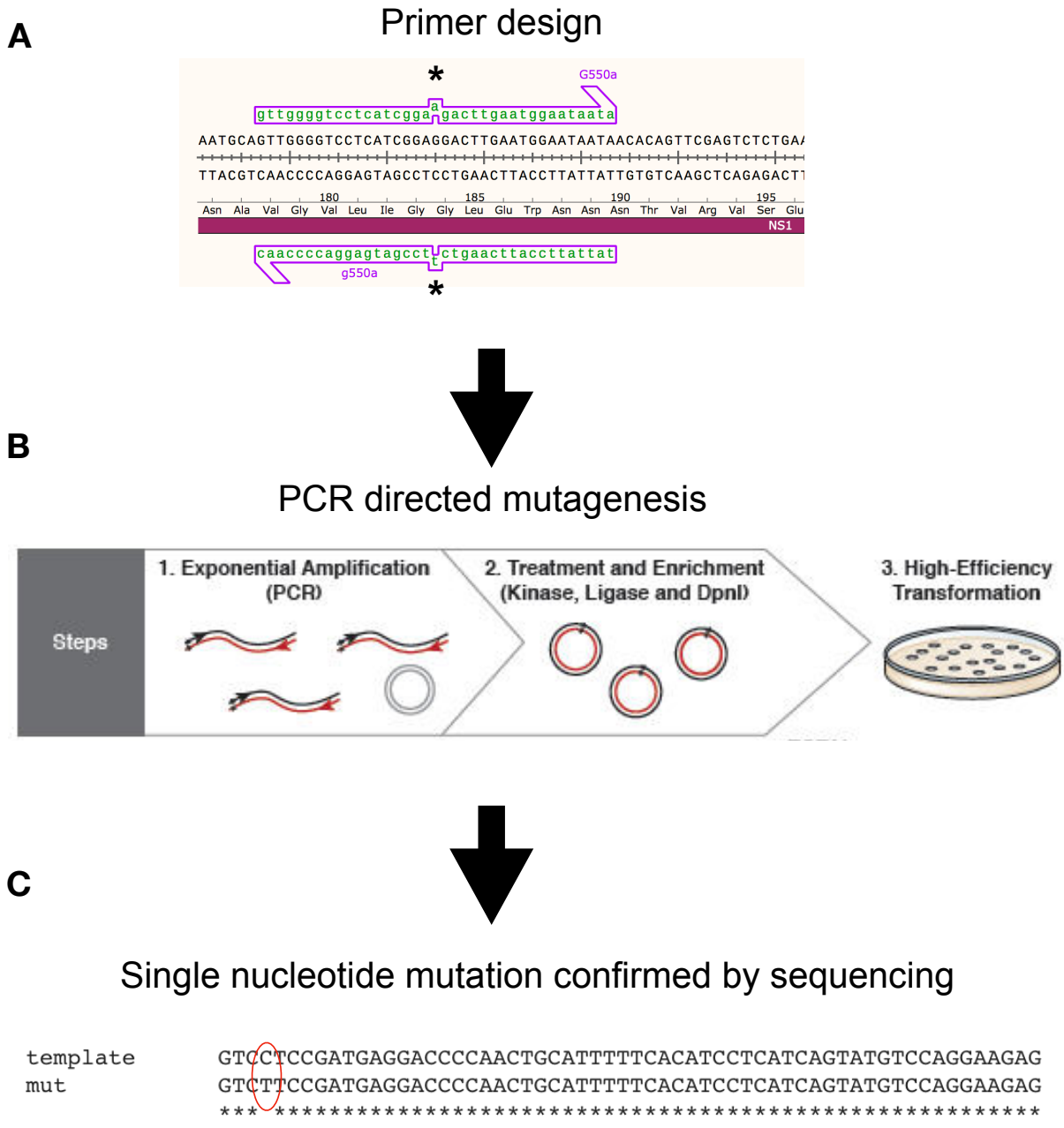


Figure 5.6 Establishing NS1-G184R mutant control

A. NS1 mutant (NS1-G184R) has been shown to be defective in CPSF30-binding (Steidle et al., 2010). Single nucleotide mutation represented by * was introduced into the NS1 cDNA sequences in pcDNA-NS1 (Robb et al., 2010) with PCR primers (g550a NS1 mut primers shown in table 2.1).

B. PCR was performed as described in chapter 2.13.6 with overlapping primers containing the altered sequences. The single nucleotide mutation from G to A results in a change from glycine to arginine at amino acid position 184, hence the name NS1-G184R or MUT. DPNI digestion was performed to target and digest methylated parent DNA without mutation and not the PCR product with the mutation. Subsequently, the plasmid DNA were transformed into chemically competent E.Coli DH5α cells, colonies were selected for mini-prep plasmid isolation.

C. Sequencing performed by Source Bioscience to verify the mutation displayed in the red circle (sequenced 3' to 5' strand).

The single nucleotide mutation from G to A results in a change from glycine to arginine at amino acid position 184, hence the name NS1-G184R or MUT. DpnI digestion was then performed to target and digest methylated parent DNA without mutation and not the PCR product with the mutation. Subsequently, the plasmid DNA were transformed into chemically competent E.coli DH5α cells, colonies were selected to grow and plasmid was isolated by mini-prep (Figure 5.6B), then sequencing to verify the mutation (Figure 5.6C). After NS1 mutant plasmids were established, pcDNA-GFP/NS1-G184R/NS1 plasmids were then transfected into HEK293T cells for 48 hours over 3 independent biological replicates. As per usual, transfected cells were subjected to fractionation, and quality control of fractionation quality was applied and assessed by RNA gel to detect for nuclear retained 45S and 32S pre-rRNA (Figure 5.7A), and protein was isolated from whole-cell lysate. Before 3'sequencing, western blots verified successful transfections, where no NS1 protein can be detected after transfecting with pcDNA-GFP, conversely NS1 protein was detectable in lanes NS1 and NS1-G184R (Figure 5.7B , bottom panel).

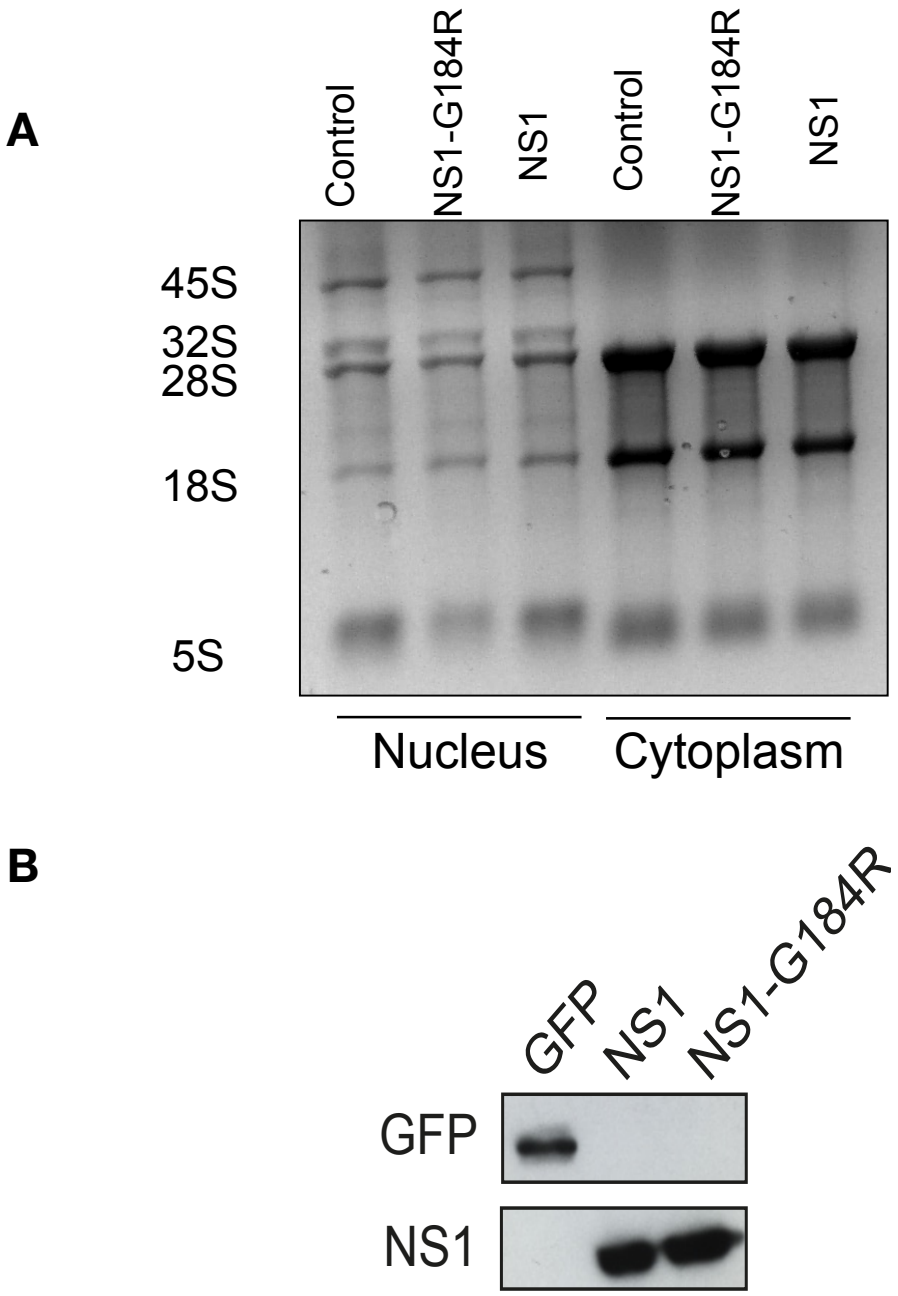


Figure 5.7 Subcellular Fraction and Western Blot Controls of GFP, NS1, NS1 Mutant Transient Expression in HEK293T Cells

A. Subcellular fractionation GFP control, NS1-G184R, NS1 expression in HEK293T cells. Denaturing RNA gel with 2 µg of RNA from nuclear (left half) and cytoplasmic (right half) fractions. The identity of each ribosomal RNA band is displayed on the left-hand side of the gel.

B. NS1 over expression was confirmed by performing western blotting. The same membrane was probed with anti GFP (top panel) and NS1 (bottom panel).

5.2.2b Investigating NS1 and NS1-G184R mutant expression effects on 3'UTR-APA in HEK293T cells

In order to investigate NS1's role in regulating APA, 3'Seq was performed after library preps of nuclear and cytoplasmic fractions. Reads of 3 independent biological replicates were processed and combined by the established APA analysis pipeline. Subsequent comparisons were made between NS1 or NS1-G184R mutant with GFP control (Figure 5.8A,B). The comparison between GFP control and NS1 expressing cells captured a total of 6976 3' UTR APA events in the nucleus and 239 of them (3.4%) showed statistically significant changes in poly(A) site usage. For the cytoplasmic fraction, 6837 APA events were identified and 627 (9.1%) of them showed statistically significant relative changes in poly(A) site usage. Similarly with NS1 mutant, of the 6956 total detected genes in the nucleus 311 (4.4%) genes resulted in APA. In the cytoplasm, of the 6817 total genes, 568 (8.5%) genes resulted in APA (total genes detected displayed in scatterplots in supplementary figure 9). This led us to speculate that NS1 alone may not be sufficient to cause a significant amount of APA as observed in infection.

Regarding APA trends, a predominantly lengthening trend was seen in NS1 over expressed nuclear (72% of total APA events) and cytoplasmic (61% of total APA events) fractions (Figure 5.8A). This is very interesting as 3'UTR lengthening was also observed during A/WSN/33 infection and CPSF30 knock-down, therefore this led us to believe that somehow NS1 abrogation of CPSF30 can potentially exhibit a lengthening trend during infection. Conversely, differences in trends were seen in NS1-G184R mutant expression. In the nuclear fractions, it was predominantly shortening showing 66% were shortening events of the total APA events. However, there were relatively more lengthening events (55%) out of the total APA events (Figure 5.8B). This provides further evidence that NS1 could be a contributing factor involved in inducing a lengthening trend in influenza

infection. Next, I wanted to verify whether there are any correlations between infection and NS1 expression. Thus, genes exhibiting APA in NS1, NS1 mutant expression were compared to infection induced APA genes that were significant at a minimum of three time points, and were analysed yielding 4 main categories: Nuclear lengthening, nuclear shortening, cytoplasmic lengthening, and cytoplasmic shortening (Figure 5.8C). Since the NS1 mutant served as a control, any APA genes that overlap in all 3 conditions were excluded from further analysis (centre of Venn diagram). Analysis from the the 4 main categories demonstrated a 9.8% overlap of genes in the nuclear lengthening cluster; 1.1% overlap in the nuclear shortening cluster; 4.3% in the cytoplasmic lengthening cluster; 2.9% overlap in the cytoplasmic shortening cluster (Figure 5.8C). This suggests that only a very low proportion of overlapped APA genes between infection and NS1 were observed. I speculate that NS1 alone is only responsible for inducing a small proportion of the APA events during infection and therefore perhaps require the aid other potential factors such as other viral factors to induce APA in full force during infection.

Even though a small overlap of APA genes was observed between NS1 transfection and influenza infection, the ultimate goal was not only to link NS1's role in APA regulation during infection but also was to determine whether these subset of genes are biologically relevant to host shut-off and viral interactions. GO analysis, where the top ten enriched biological processes listed, was therefore performed for all genes that exhibited APA in both infection and NS1 over expression. For genes that appear to be significantly lengthening in the nucleus, genes were enriched for mRNA capping, processing, transcription initiation, protein localisation and cell cycle. Genes within the nuclear shortening category revealed enrichment with cell cycle and immune response. For cytoplasmic lengthening, they were significantly associated with p53 singling pathway regulation, protein localisation and modulation by virus of host morphology and physiology.

Genes within the cytoplasmic shortening category were significantly associated with positive regulation of viral process and regulation of viral protein levels. Interestingly, from the 4 categories shown in figure 5.9A-D, significantly enriched GO biological processes terms turned out to be homologous to what was seen in both infection and CPSF30 knock-down experiments. As a result, gene ontology analysis was able to determine infection, CPSF30 knock-down and NS1 were likely responsible for inducing APA in these common subset of genes involved in mRNA processing, mRNA metabolism, immune response and positive regulation of virus.

Furthermore I referred back to the 5 stringently selected genes with functions relating to immune response and viral interactions: DDX5, HEXIM1, MAVS, SMAD3 and TRIM25. Interestingly, only DDX5 and SMAD3 exhibited similar significant 3'UTR APA changes after NS1 expression as in infection and CPSF30 knock-down. However, DDX5 exhibited APA lengthening only in the cytoplasm after NS1 expression (Figure 5.10A), where as in both viral infection and CPSF30 knock-down, DDX5 were seen to exhibit statistically significant APA lengthening in both nuclear and cytoplasmic fractions. On the other hand, SMAD3 exhibited significant lengthening in both nuclear and cytoplasmic fractions (Figure 5.10B) identical to infection, CPSF30 knock-down, and osmotic shock experiments. This is indeed a significant result as these examples further supports our hypothesis regarding NS1-mediated APA in infection via CPSF30 abrogation.

Moreover for MAVS, where a 3'UTR termination defect was seen during infection and CPSF30 knock-down, its statistical significant values for nuclear (-1.25, p=0.0562) signifies termination defect in the nucleus has occurred but was just above the statistical cut-off during NS1 over expression. Thus, an assumption can be made that NS1 over expression alone may not be enough to induce MAVS termination defect. Interestingly, MAVS

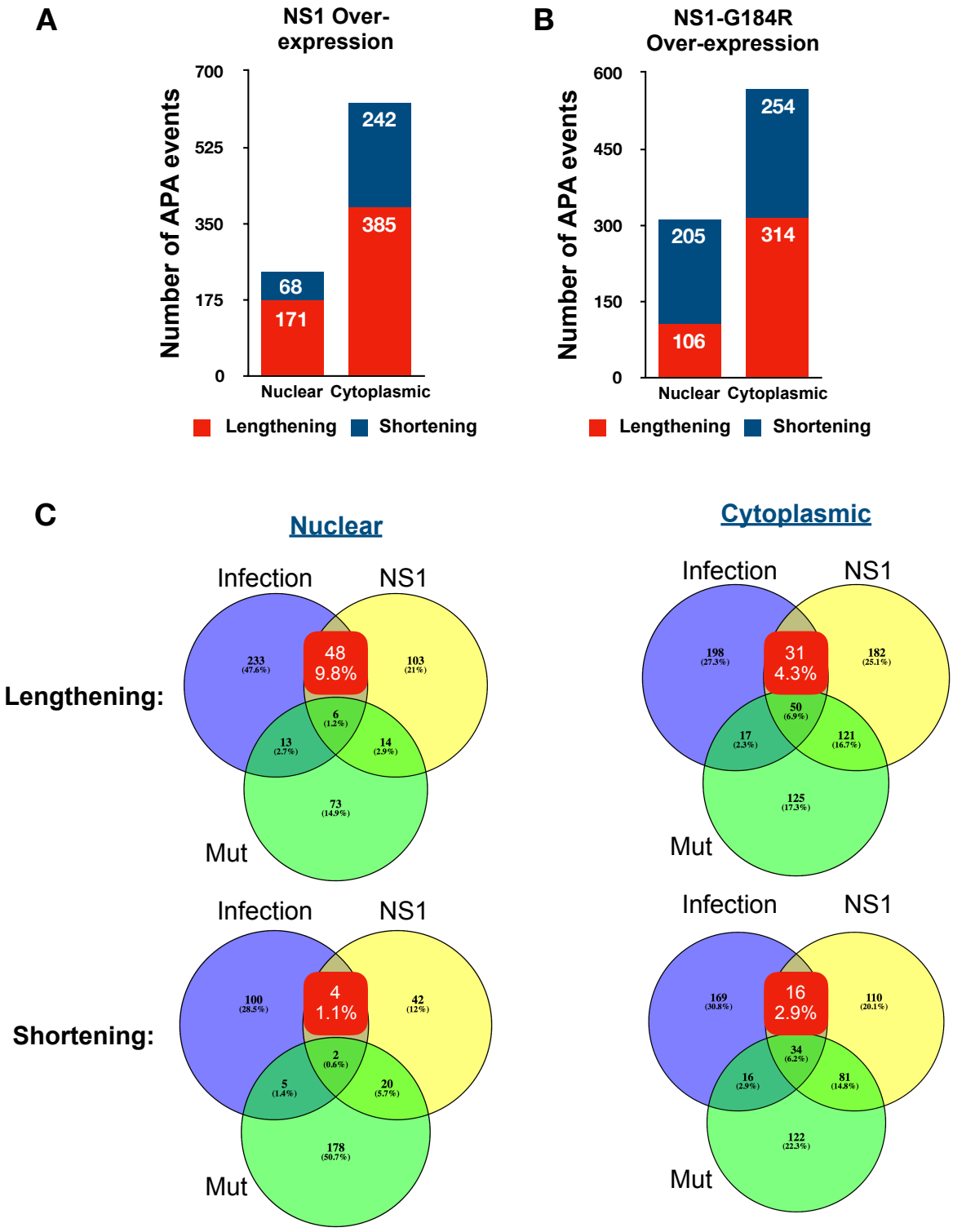


Figure 5.8 Global Analysis of NS1 and NS1 Mutant Expression 3'UTR APA Followed by Comparisons with Infection 3'UTR APA Events.

A. Bar graph to visualise number of significant 3'UTR APA events (Fisher's exact Test, $p<0.01$, with a minimum 5% change in APA isoform abundance) after NS1 (Left panel) and NS1- G184R mutant (right panel) in both nuclear and cytoplasmic fractions over 3 independent biological repeats of sequencing. The numbers of lengthening events are presented in red, whereas the number of shortening events are presented in blue. The total detectable genes in the nucleus after NS1 over expression are 6976. The total detectable genes after NS1 expression in the cytoplasm are 6837 (**Supplementary figure 9**). NS1 mutant total nuclear detectable genes are 6956; NS1 mutant total cytoplasmic detectable genes are 6817.

B. Venn diagrams displaying number of overlaps between identified 3'UTR APA events between A/WSN/33 infection, NS1 and NS1 mutant (Mut) that belong to 4 categories: Nuclear lengthening (top left panel), nuclear shortening (bottom left panel); cytoplasmic lengthening (top right panel), cytoplasmic shortening (bottom right panel). Number of and % of genes overlapped are displayed in the red boxes.

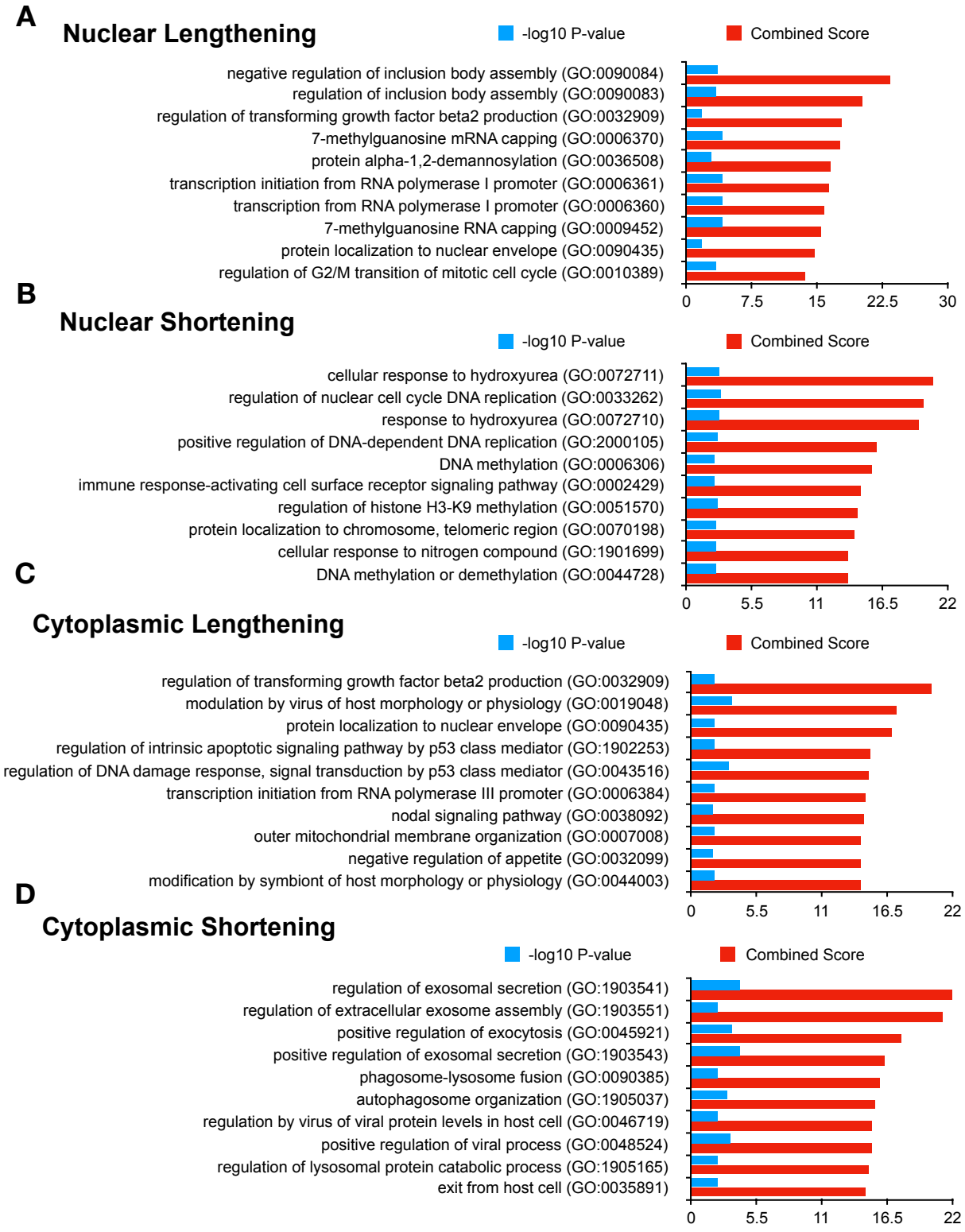


Figure 5.9 Gene Ontology Analysis of Overlapped Genes from WSN Infection, NS1 and NS1 Mutant Expression
The top 10 biological processes gene ontology categories with significant enrichment ($p<0.05$) of the genes were associated with nuclear lengthening **(A),** nuclear shortening **(B),** cytoplasmic lengthening **(C)** and cytoplasmic shortening **(D)**. Gene ontology analysis was performed using Enrichr.

exhibited significant ‘shortening’ in the cytoplasm after NS1-G184R mutant expression, which perhaps signifies once the interaction between CPSF30 is abrogated, the effects of termination defect will be reversed.

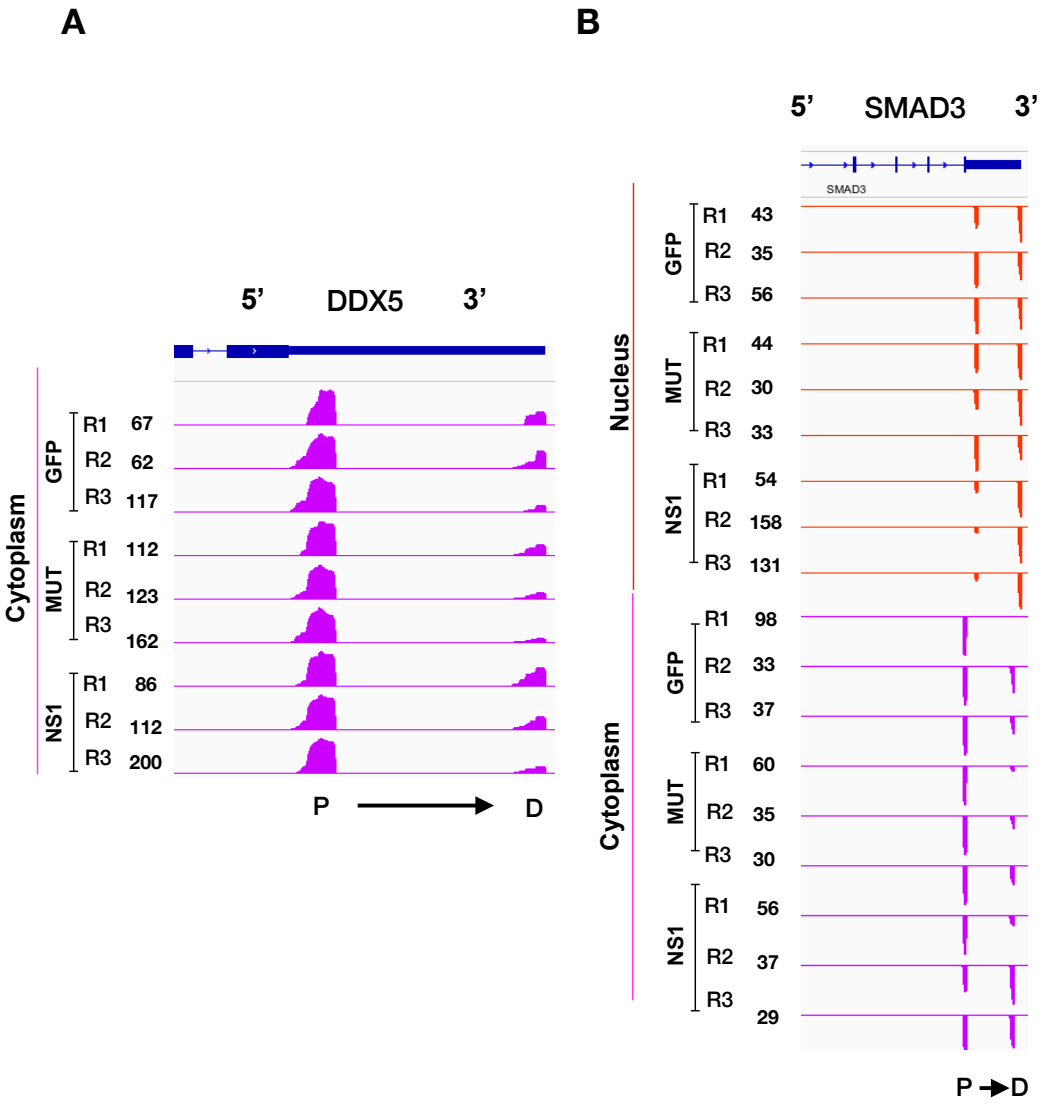


Figure 5.10 Examples of Stringently Selected Genes which Display Differential Regulation 3’UTR APA during NS1, NS1 Mutant and GFP Expression Experiments.

A. Screenshot of DDX5 as an example of a gene that displays significant shifts in CPA site usage to using the distal CPA site in the 3’UTR, when comparing cytoplasmic (purple) fractions of NS1 expression to GFP and NS1 mutant (MUT) control in HEK293T cells.

B. Screenshot of SMAD3 displaying 3’UTR lengthening when comparing nuclear (red) or cytoplasmic (purple) fractions of NS1 over expression to GFP and NS1 mutant (MUT) control in HEK293T cells.

All IGV images are displayed with a 5’ to 3’ direction. P represents proximal CPA site, D represents distal CPA site. NOTE:

5.2.2c Verifying Whether NS1 Abrogation of CPSF30 is the Mechanism Responsible for Regulating 3'UTR APA During Infection

Since the ultimate goal of the chapter is to discover whether NS1's abrogation of CPSF30 is the mechanism required for APA during infection, I therefore needed to isolate APA genes that were induced by NS1 over expression, CPSF30 knock-down and infection, but excluding those that exhibit differential APA regulation from osmotic shock experiments and NS1 mutant expression. Venn diagrams were utilised to yield the following 4 categories: nuclear lengthening and shortening; cytoplasmic lengthening and shortening (Figure 5.11). Venn diagrams labelled with CPSF30 knock-down represents APA genes that were commonly induced in both CPSF30 knock-down and infection. NS1 labelled represents APA genes that were induced in both NS1 over expression and infection, but not after NS1-G184R mutant expression. Based on the analysis of the genes where APA is induced in influenza A/WSN/33 infection, CPSF30 knock-down and in NS1 expression, the number of overlaps between APA genes induced by infection, CPSF30 knock-down and NS1 expression, and % overlap are displayed in the red box within the Venn diagrams in figure 5.11. As shown in the Venn diagrams, 18 (12.2%) genes, from nuclear lengthening category exhibited significant differential APA regulation in all 3 conditions. Nuclear shortening displayed 1 (3.6%) APA gene that was significant in all 3 conditions. In the cytoplasmic categories, 15 (15%) lengthening genes and 3 (7.9%) shortening genes were significant in all 3 conditions. These results suggest that only a small proportion of genes were observed where poly (A) site usage was affected in all three conditions. I therefore focused on those that are physiologically relevant and result in APA changes via the NS1 mediated abrogation of CPSF30 during infection (listed in Table 5.1). DDX5 is a prime example, whereby it is significantly lengthening in the cytoplasmic fractions in all three conditions (Figure 5.11, top right panel). DDX5 gene expression was also significantly up regulated (Fisher's exact Test, $p<0.01$) in both NS1 expression and CSPF30 knock-down

similar to what was observed during infection. This signifies that during infection, DDX5 differential APA regulation is likely mediated by NS1 and CPSF30 abrogation, followed by an increase in gene and protein expression. Furthermore, since DDX5 has been shown to promote viral growth (Bortz et al., 2011; Cheng et al., 2018), I can conclude that influenza A/WSN/33 infection can induce DDX5 APA by NS1 expression and CPSF30 abrogation as a mechanism to promote its replication. However, altogether the evidence does not suggest that NS1 and CPSF30 alone are responsible for regulating and facilitating all APA changes during viral infection.

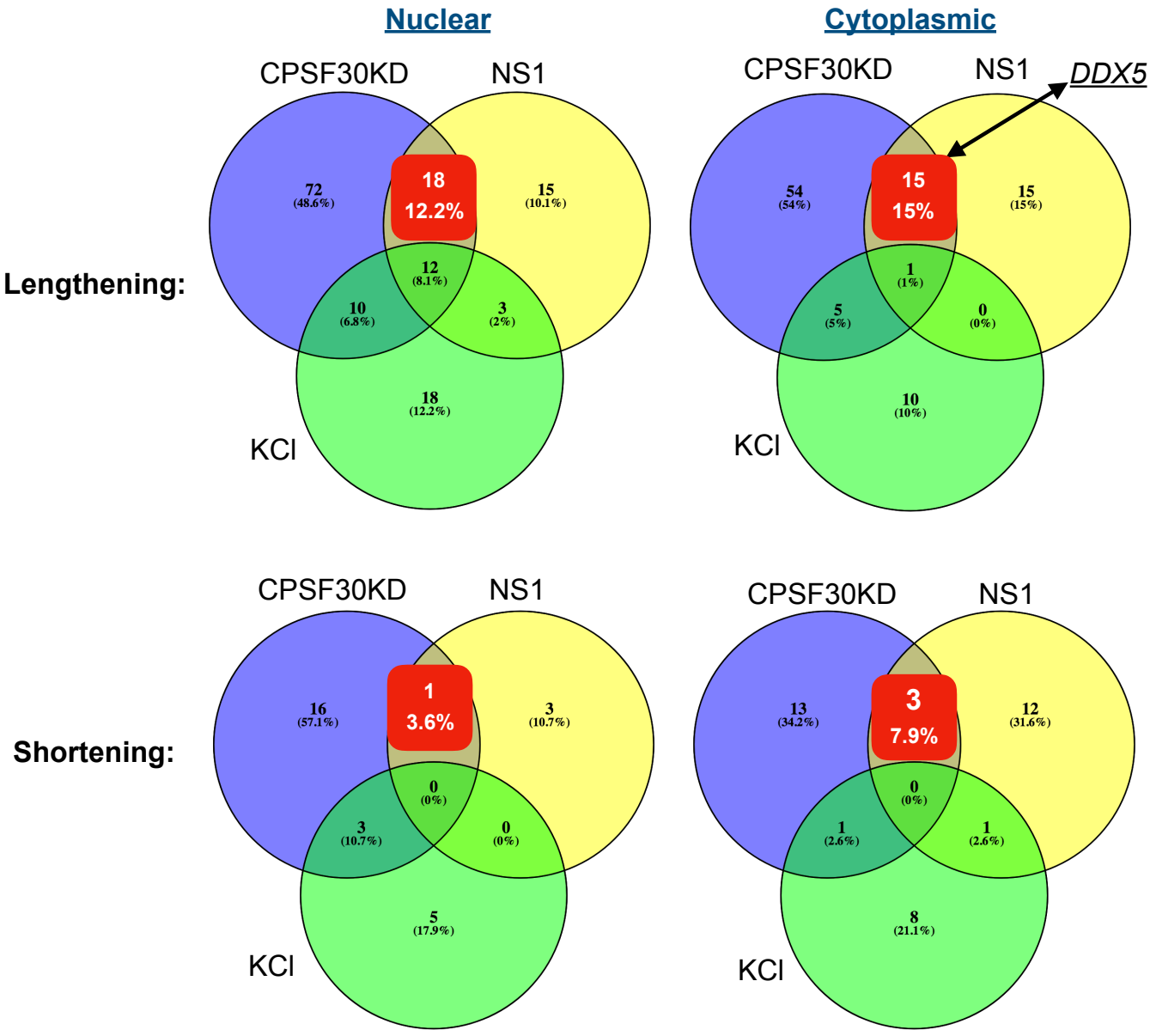


Figure 5.11 Verifying Whether NS1 abrogation of CPSF30 is the Mechanism Responsible for Regulating 3'UTR APA (previous page)

Venn Diagrams displaying overlaps between 3'UTR APA events induced from WSN infection, CPSF30 knock-down, NS1 over expression, but excluding those induced by osmotic shock (KCl) or NS1 mutant over expression. 4 categories: Nuclear lengthening (top left panel), nuclear shortening (bottom left panel); cytoplasmic lengthening (top right panel), cytoplasmic shortening (bottom right panel). Number of and % of genes overlapped are displayed in the red boxes.

Table 5.1 Overlapping genes that are regulated after CPSF30KD and NS1 expression and upon infection, but NOT after osmotic shock/ NS1-G184R)

Nuclear Lengthening	Nuclear Shortening	Cytoplasmic Lengthening	Cytoplasmic Shortening
C11orf84	APPL1	FLNB	DSTN
DNAJB4		GPI	LAPTM4B
DYNC1H1		HN1	RAB7A
ECT2		HSPA4	
FAM160B2		NUTF2	
FUS		C11orf84	
HNRPDL		DDX5	
KIAA0195		MDM2	
POLR2E		AGPAT6	
POLR2F		GTF3C4	
RRN3		HCCS	
RSL1D1		HES4	
HN1		HOTAIRM1	
KIF1B		SOX12	
NUTF2		TCEB2	
SOX12			
UBA52			
VPS52			

5.3 Discussion

The focus of this chapter was to decipher the mechanism responsible for APA regulation during influenza A/WSN/33 viral infection. The interactions between CPSF30 and NS1 make it plausible to inhibit CPSF30 function, and potentially lead to alterations in APA regulation. Therefore, CPSF30 knockdown and transient over expression of NS1 were performed to HEK293T cells to best mimic one of the strategies used by influenza A viral infection to avoid host defence.

In this study, CPSF30 knock-down induced significant number of relative changes in CPA site usage, of which only less than 20% of total genes detected from both nuclear and cytoplasmic fractions exhibited APA. This is contrary to a study, where specific cleavage and polyadenylation factors were knocked down including CPSF30, which had a modest effect on APA (Li et al., 2015). Knockdown of CPSF30 resulted in increased relative usage of distal CPA sites, perhaps more exaggerated than during influenza infection. Interestingly, in recent studies, knockdowns of 3'processing factors were performed. PABPN1, CFIm68, CFIm25 depletions (Jenal et al., 2012; Masamha et al., 2014; Martin et al., 2012) resulted in increased relative usage of the proximal CPA site. Conversely, knock-down of the CstF64 paralog CstF64τ led to increased relative usage of distal CPA sites (Yao et al., 2012). Additionally, a CPSF30 mutant was shown to alter CPA site choice in a large majority of *Arabidopsis* plant genes (Thomas et al., 2012), which concurred with the data in this study. However, the difference is that CPSF30 mutant of *Arabidopsis thaliana* (At) regulates a large number of APA events without a clear preference for proximal or distal CPA site usage during response to oxidative stress (Thomas et al., 2012), contrary to the lengthening pattern seen in this chapter. This led us to believe that the differences in APA trends between plants and human may be due to structural differences between plant and mammalian CPSF30. The majority of CPSF30 core 3 zinc finger (ZnF) motifs between mammals, yeast and plants are conserved where CPSF30 proteins possess distinct CCCH zinc finger motifs. However, instead of 5 ZNFs seen in animals and yeast, plant CPSF30 only harbours 3 motifs specifically lacking the first and fifth and the C-terminal zinc knuckle motifs present in *Drosophila*, *C. elegans*, zebrafish and others (Bai and Tolias. 1998; Delaney et al., 2006). Additionally, the 3 ZnF motifs of *Arabidopsis thaliana* CPSF30 have different biochemical functions, where the first and third ZnFs are involved lnRNA binding and endonuclease activity and the second ZnF function remains unknown

(Delaney et al., 2006; Adepalli and Hunt. 2007). Thus the culmination of different structures and functions of CPSF30 between mammalian and plants may induce different APA trends.

Next, I wanted to investigate the relationship between infection and CPSF30 knock-down. The results demonstrated little overlap between APA genes induced by CPSF30 knock-down and infection. This suggests that knock-down or abrogating CPSF30 during infection will only contribute to APA regulation of a small subset of genes. Subsequently, these overlapped genes were interrogated for any similarities in biological processes relating to viral interactions or immune response. GO analyses revealed similar significant enrichment of biological processes between infection and CPSF30KD, including positive regulation of IFN 1 production, mRNA splicing, mRNA metabolism, gene expression regulation, protein localisation and miRNA processing, thus proposed a possibility that CPSF30 abrogation during infection may effect APA of these subset of genes. Additionally, out of the stringently selected genes from chapter 4, DDX5, SMAD3 and MAVS displayed identical shift in CPA site usage in CPSF30KD as in infection. Therefore, the APA profile changes in these genes seen during influenza infection could indeed be the result of NS1 mediated CPSF inhibition and potentially effect expression of subsets of physiologically relevant genes relating to immune response, mRNA processing, splicing and protein localisation, thus uncovering novel strategies for the virus to avoid host defence.

After I discovered that CPSF30 knock-down results in APA regulation of physiologically relevant genes, I then explored the possible role of NS1 in regulating these APA changes. NS1 has been shown to interact with zinc finger 2 and 3 of CPSF30, interestingly in plants the mammalian CPSF30 zinc fingers corresponds to the zinc 1 and 2 of *Arabidopsis thaliana* CPSF30 (Chakrabarti and Hunt. 2015). In order to connect NS1's involvement

with CPSF30 abrogation and subsequent induction of APA of host genes, NS1 transient over expression in HEK293T cells were established. This study demonstrated that only a small % of total genes resulted in APA after NS1 and NS1-G184R mutant over expression in comparison to the GFP control in both nuclear and cytoplasmic fractions. However more encouragingly, 3'UTR APA lengthening trend was observed, which was similar to the trends observed in influenza A/WSN/33 infection and CPSF30 knock-down. APA genes induced by NS1 over expression (excluding those induced by mutant NS1) was subsequently overlapped with those induced by infection, but a small proportion of APA genes induced by NS1 over expression (excluding those induced by mutant NS1) was overlapped with those influenced by infection. GO analysis comparison between NS1 over expression, CPSF30KD and influenza infection experiments revealed that the significantly enriched GO biological processes were nearly homologous including those associated with mRNA processing, mRNA metabolism, immune response and positive regulator of virus. It has been shown that altering activity of CPSF30 in different cellular settings such as cancer (Chen et al., 2013; Chen et al., 2014) may lead to large scale reprogramming of gene expression. This is particularly relevant under conditions whereby NS1 can alter the activity and function of CPSF30 in vivo. Interestingly as stated in chapter 4, gene expression down-regulated APA genes after infection were associated with nuclear export, positive type 1 interferon production regulation, negative regulation of viral release and protein localisation, where as up-regulated APA genes from infection were enriched for neutrophil mediated immunity, mRNA splicing, processing, metabolism and positive regulation of viral life cycle. Most of these GO biological processes overlap with those enriched in NS1 over expression and depletion of CPSF30. Thus this provided additional evidence that NS1 abrogation of CPSF30 may regulate APA and alter gene expression of these particular subset of genes in order to facilitate viral transcription, replication and survival by inducing host shut-off of genes relating to immune regulation.

Finally, the ultimate comparison between influenza infection, CPSF30 knock-down, NS1 expression (excluding those that are induced during osmotic shock and NS1-G184R mutant) were done to determine the mechanism for 3'UTR APA regulation in A/WSN/33 infection. Particular genes (listed in table 5.1) were found to exhibit significant APA regulation in all 3 conditions such as DDX5 and SMAD3. Although in the case of SMAD3, 3'UTR APA lengthening was not unique to infection, CPSF30 knock-down and NS1 but also was induced by osmotic stress therefore this could be a result of stress and perhaps not viral mediated. Altogether, evidence from both chapter 4 and 5 elucidated the fact that influenza virus can certainly utilise NS1-mediated abrogation of CPSF30 as a mechanism to regulate differential usage of CPA sites in physiological relevant genes such as 3'UTR lengthening in DDX5, thereby increasing its gene and protein expression, and maintain viral replication in influenza RNA viruses (Bortz et al., 2011). While there are some common APA genes between the 3 conditions, the overall evidence does not suggest that NS1-mediated abrogation of CPSF30 alone is responsible for regulating and facilitating all APA during viral infection, and therefore perhaps require the cooperation of other viral factors to induce APA during infection. Despite this, this project was able to elucidate another way for influenza A virus to manipulate host transcriptome to facilitate viral transcription and shutting down immune response for its own survival.

5.4 Summary

To conclude this chapter, both CPSF30 knock-down and NS1 transient expression as an individual experiment induced significant APA events. However very little overlap has been seen when compared with influenza A/WSN/33 infection. This signifies that NS1 abrogation of CPSF30 alone may not be the only and major mechanism for regulating 3'UTR APA. However, one cannot discount the fact that relevant genes such as DDX5 is

regulated by this mechanism and thus may be beneficial for the virus in accordance to its function.

There are certainly limitations to the CPSF30KD and NS1 over expression experiments. One major limitation of both experiments is there is uncertainty of what proportions of the cells exhibited CPSF30 depletion or NS1 over expression. This is a problem since this is a global transcriptome study and thus requiring a majority of cells to exhibit either of these phenotypes. Perhaps in order to solve this, a stable expression cell line can be generated for NS1 over expression instead of transient expression. Alternatively, an inducible knockdown system can be developed for CPSF30 as a more powerful tool than RNA interference. Finally, for future aims and experiments, it would be interesting to further explore whether the other genes (common for infection, NS1 and CPSF30 experiments) in table 5.1 have any physiological relevance to host shut-off or response. Ultimately, other mechanisms and viral factors that are responsible for APA regulation in infection will need to be further explored.

Chapter 6: Disussion and Future Directions

6.1 Thesis Summary

This is the first ever study to deeply investigate the changes of APA regulation during influenza A/WSN/33 infection in A549 lung epithelial cells and to decipher the mechanism involved in causing such changes. This project combined the use of deep sequencing, bioinformatic analysis and classical biochemistry techniques to give a comprehensive outlook of transcriptome changes during infection.

Chapter 3 primarily focused on the development of a consistent infection time course experiment by considering the necessary controls for robust efficient infections with minimal contaminants, along with post-infection controls to determine biologically relevant periods where potential APA profiles changes are expected to occur. Additionally, this chapter demonstrated established controls to verify the quality of subcellular fractionation technique, which was then applied throughout the other chapters in order to distinguish between active and passive APA, as well as to provide interesting insights on infection dynamics. Applying all the controls and strategies mentioned in chapter 3, chapter 4 uncovered unique global APA changes in response to influenza virus infection in A549 cells by 3'sequencing, after which these changes were further verified by traditional techniques. Moreover, this chapter elucidated the possibilities of viral mRNA nuclear export failure late in infection. Subsequently, chapter 5 aimed to decipher whether viral NS1 and abrogation of cleavage and polyadenylation factor CPSF30 is the mechanism responsible for the global APA changes during infection. CPSF30 knock-down and NS1 expression in HEK293T cells were therefore established to best mimic one of the strategies used by influenza viral infection to avoid host defence and to induce host shut

off. The same strategies were applied from chapter 3 and 4 to investigate the global APA changes and subsequently overlapped with those effected by infection.

The key outcomes and findings of this thesis are:

1. Influenza A virus hijacks host nuclear and cytoplasmic mRNA pool, but fails to efficiently export viral mRNA to the cytoplasm in later stages of infection.
2. Influenza A/WSN/33 infection in A549 cells induced host RNA APA profile changes in both nuclear and cytoplasmic fractions.
3. Majority of viral-induced APA is independent of stress response from osmotic shock.
4. Identification of physiologically relevant genes that exhibited APA during infection can contribute to regulation of gene and protein expression.
5. NS1 mediated abrogation of CPSF30 during infection reproduces APA profile changes in a small proportion but physiologically relevant subset of genes that regulate host immune responses and promote viral propagation.

This final chapter will discuss the importance of these findings and outline future objectives to tackle remaining questions regarding the importance of APA in influenza infection.

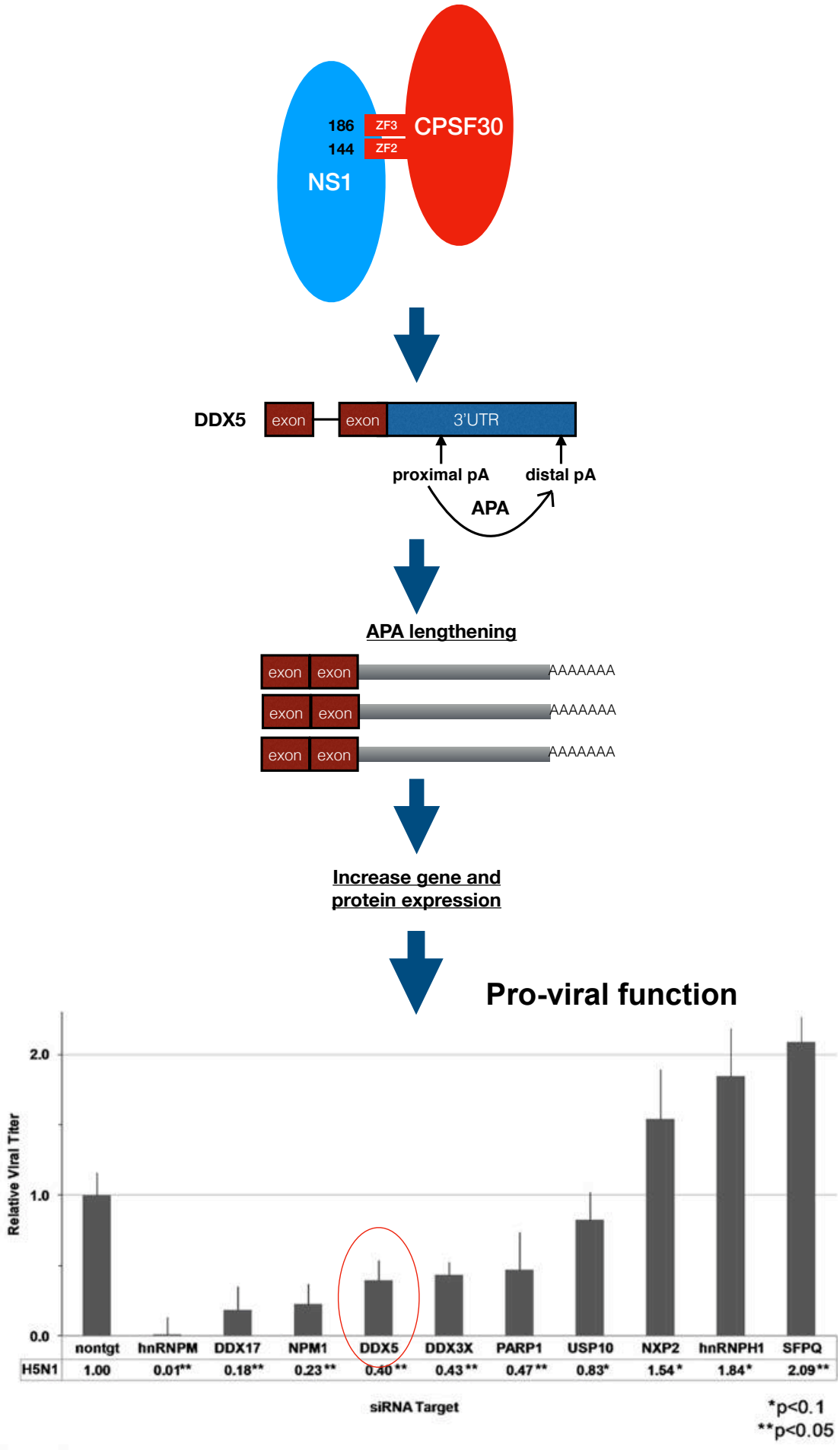
6.2 Advantages and Limitations of 3'end Sequencing for Global APA Analysis

Global analysis of APA were performed by 3'sequencing through out the project. This facilitated a higher resolution analysis of poly(A) site usage by sequencing only the terminal 200 bases when compared with whole transcriptome sequencing, where reads are often skewed towards the coding region of the transcript. Additionally, 3'sequencing allows accurate assessment of gene expression as each transcript generates one sequencing fragment, and at the same time being highly cost-effective. However there are limitations to all 3'end sequencing techniques. 3'end sequencing techniques fail to identify sequences upstream of transcripts, due to the technical limitation of short read lengths, and thus are unable to describe an entire RNA molecule from 5' to 3' end. Moreover, this creates an obstacle for identifying common events such as alternative splicing and distal pA site selection (Neve et al., 2016). Therefore the only way to truly identify the complete transcript structure on a transcriptome scale will be to sequence the full length RNA as a single molecule, which is becoming more attainable (Sharon et al., 2013).

6.3 Alternative Mechanisms Responsible for Regulating APA During Infection

This is the first study to extensively investigate APA profile changes under influenza A viral infection and to uncover the possible mechanism for such regulation. Differential regulation of APA isoforms can have profound effects on post-transcriptional gene regulation by affecting mRNA stability, cellular localisation, nuclear export and translation from biological stimuli such as influenza infection (Neve et al., 2017). As shown in chapter 4, gene expression down-regulated APA genes after infection were associated with nuclear export, positive type 1 interferon production regulation, negative regulation of viral release and protein localisation. Where as up-regulated APA genes from infection were enriched for

neutrophil mediated immunity, mRNA splicing, processing, metabolism and positive regulation of viral life cycle. Most of these GO biological processes overlap with those enriched in NS1 over expression and depletion of CPSF30. Thus, this signifies influenza virus has evolved to utilise NS1 abrogation of CPSF30 to regulate APA and alter gene expression of these particular subset of genes to facilitate viral transcription, replication and survival by inducing host shut-off of genes relating to immune regulation along with many other viral strategies to antagonise innate immunity (summarised in section 1.8 ;Hale et al., 2010). One clear model example is DDX5 (Figure 6.1), through NS1 mediated abrogation of CPSF30, differential regulation of APA occurs leading to increased gene and protein expression to promote RNA viral growth. In order to assess why APA lengthening or shortening may cause changes in gene and protein expression in these subset of genes, individual gene analysis (table 5.1) will need to be further explored.



Bortz et al., 2011

Figure 6.1 DDX5 as Prime Example of a Gene with Differential APA, Gene and Protein Regulation Caused by NS1 Abrogation of CPSF30 During IAV WSN/33 Infection (previous page)

DDX5, through NS1 mediated (via amino acid residues 144-186) abrogation of CPSF30 via ZF2 and ZF3 results in differential regulation of APA (3'UTR lengthening), leading to increased gene and protein expression to promote RNA viral growth. DDX5 pro-viral function was proposed by a study done by Bortz and colleagues. Bar graph demonstrates host factor RNAi silencing alters progression of the H5N1 virus life cycle, especially DDX5 circled in red (adapted from Bortz et al., 2011). 36 hours siRNA transfected A549 cells were then infected by influenza A/VN/1203/04 (H5N1) HALo virus (MOI = 0.1). Subsequently, average relative titres of virus were then taken at 20 hpi. Values shown are averages of plaque and limiting dilution assay ratios for each host target siRNA relative to nontarget (nontgt) siRNA values, with significance estimated by an unpaired, 2-tailed t test. Knockdown of DDX5 has shown a significant decrease in relative viral titre, thereby signifying DDX5's potential as a pro-viral factor.

6.3.1 NS1 Inhibition of Splicing as an Alternative Atrategy for APA Regulation During Infection

NS1 has the ability to induce APA via CPSF30 abrogation, however focusing on CPSF30 knock-down may limit our view of other possible mechanisms that can induce APA. NS1 has the ability to inhibit numerous other cellular components through its RNA binding and effector domains (Figure 5.1). NS1 has been shown to interact with U2 and U6 snRNA splicesome subunits in the classic complex, but also with a U6 counterpart U6atac (Qiu et al., 1995; Wang and Krug, 1998). Influenza A viruses have developed strategies to replace host RNA pol II for transcription, as well as 5' capping and 3' end polyadenylation, but not splicing. Influenza virus through NS1 is able to hijack the host splicing machinery to process IAV NS and M segments (reviewed in Dubois et al., 2014). More importantly, it has been shown that NS1 interactions with U2 and U6 snRNAs will block cellular gene expression by inhibiting complete spliceosome recruitment and cause host splicing defects. Recently, Neve and colleagues have demonstrated incomplete splicing can result in nuclear retention as well as influenced CPA site selection. Interestingly, through incomplete splicing, some genes exhibited intron retention, which has been linked to distal CPA site usage (Neve et al., 2016). Apparently, this was potentially due to the perturbation of a well-known interconnection between splicing and 3'end processing (Proudfoot et al.,

2002). Therefore, through interruption of splicing, this may potentially provide another opportunity for influenza NS1 to regulate APA profile changes. Future considerations of full length RNA sequencing will be required to detect these events.

6.3.2 Enhancement of NS1-CPSF30 Interaction by Viral Polymerase to Regulate APA

APA regulation may also be affected by the strength by which NS1 binds to CPSF30. It has been shown that NS1 proteins from different strains bind CPSF30 with different strengths *in vitro*. In fact recent evidence suggests that Influenza A viral polymerase protein PA, not PB1 or PB2, is an integral part of the CPSF30-NS1 protein complex required for increased interaction strength, stabilisation of the complex, and subsequent inhibitory activity during infection (Kuo and Krug., 2009). Since we only over express NS1 in chapter 5, the strength of CPSF30 binding might be effected without the presence of viral PA and therefore may contribute to differences of APA regulation when compared to infection. This may also explain why APA overlaps between CPSF30 knock-down, NS1 over-expression and infection are low, because perhaps APA regulation requires all viral factors such as PA and NS1 to synergistically interact with each other to induce APA in full force as demonstrated in infection. For future experiments, it would be interesting to analyse global APA in transient over expression of PA and NS1 and subsequently compare to the current infection APA data to see if they exhibit better correlation.

6.3.3 Influenza Viral Strain Potentially Effects APA?

As mentioned in the 6.3.2, NS1 proteins from different strains bind CPSF30 with different strengths *in vitro*. Several residues of NS1 are required for CPSF30 interaction, including F103 and M106 (Das et al., 2008) which are conserved among most human IAVs, but are mutated in some strains, such as IAV PR8 (Das et al., 2008). The NS1 of PR8 IAV

consistently fails to interact with CPSF30 (Twu et al., 2007; Wang et al., 2017). Additionally, H5N1 influenza A/Hong Kong/483/97 (HK97) virus has been shown to have an intrinsic defect in binding CPSF30. However in strains isolated after 1998, with L103F and I106M mutations, NS1 of H5N1 viruses increased NS1 interaction with CPSF30, thereby able to suppress type I interferon expression and facilitate viral replication (Twu et al., 2007). Therefore, we can assume that strength in APA regulation can vary between strains. APA profile changes should be compared in the future between WSN/33 and other strains to see if APA changes are conserved.

To conclude, this project has elucidated an unexplored strategy for influenza A to induce APA and host shut off by NS1 mediated abrogation of CPSF30. Although, solely analysing NS1 over expression and CPSF30 knock-down as a strategy employed by influenza virus to induce APA is not enough.

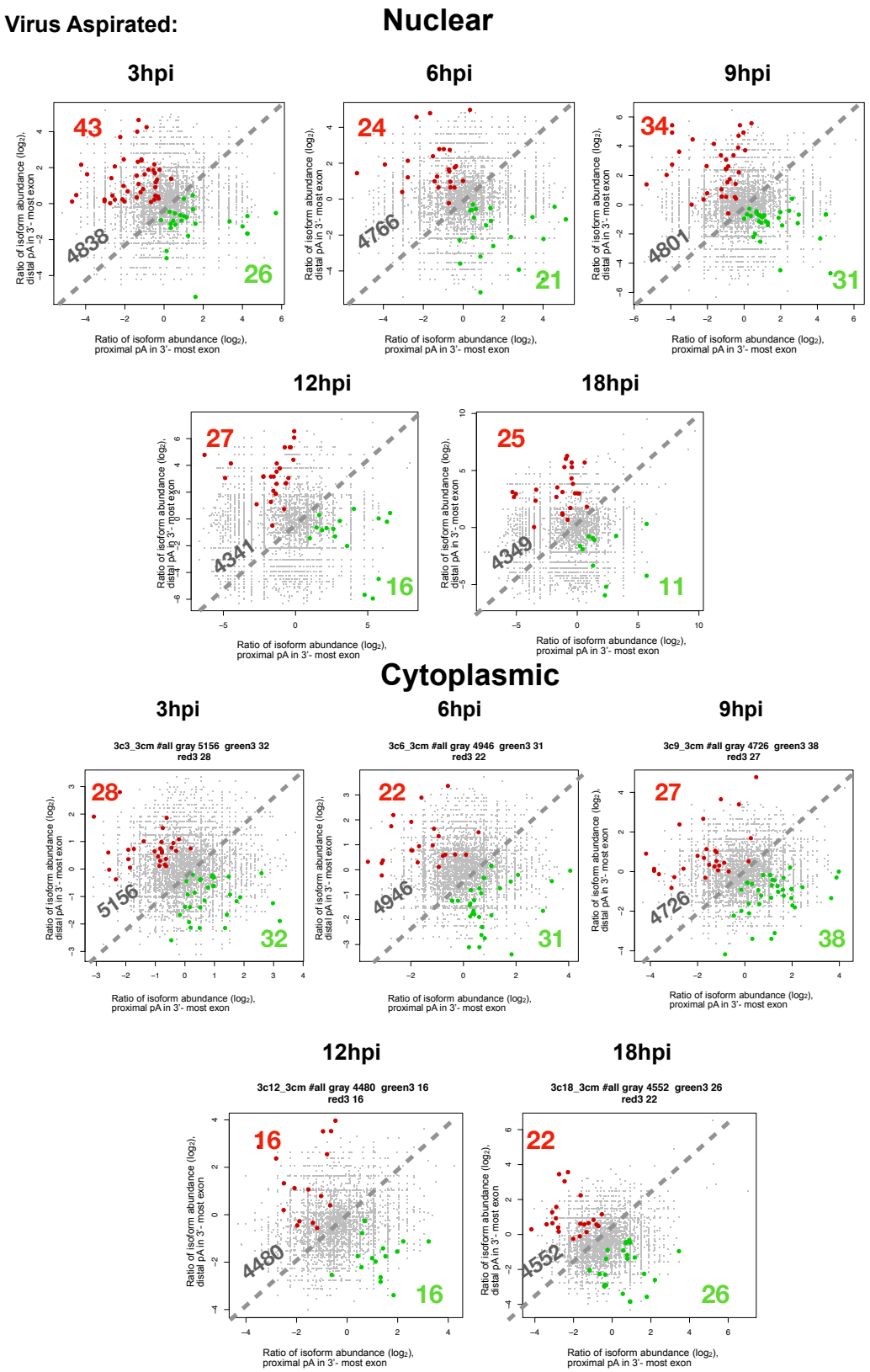
6.4 Conclusion

This work has provided novel understanding into influenza A infection dynamics and most importantly an alternative strategy used by the virus for host shut off via APA regulation. The infection time course pipeline developed here has not only applied extra layers of controls but also facilitated a consistent experimental approach to study global APA changes during infection. With the ever increasing development of high-throughput sequencing, this has no doubt made huge leaps to our knowledge of APA, as well as molecular genetics. However, the global studies can fall into a trap that leads to a focus of identifying general trends that apply to large subsets of genes rather than considering the unique functions, regulation and physiological consequences at an individual gene level. Therefore this project has attempted to combine the transcriptome approach and direct experimentation at an individual gene level to highlight the true value of this global study.

This approach allowed isolation of genes of which APA is likely to play a role in their regulation and thus impact both the host and the virus depending on their function.

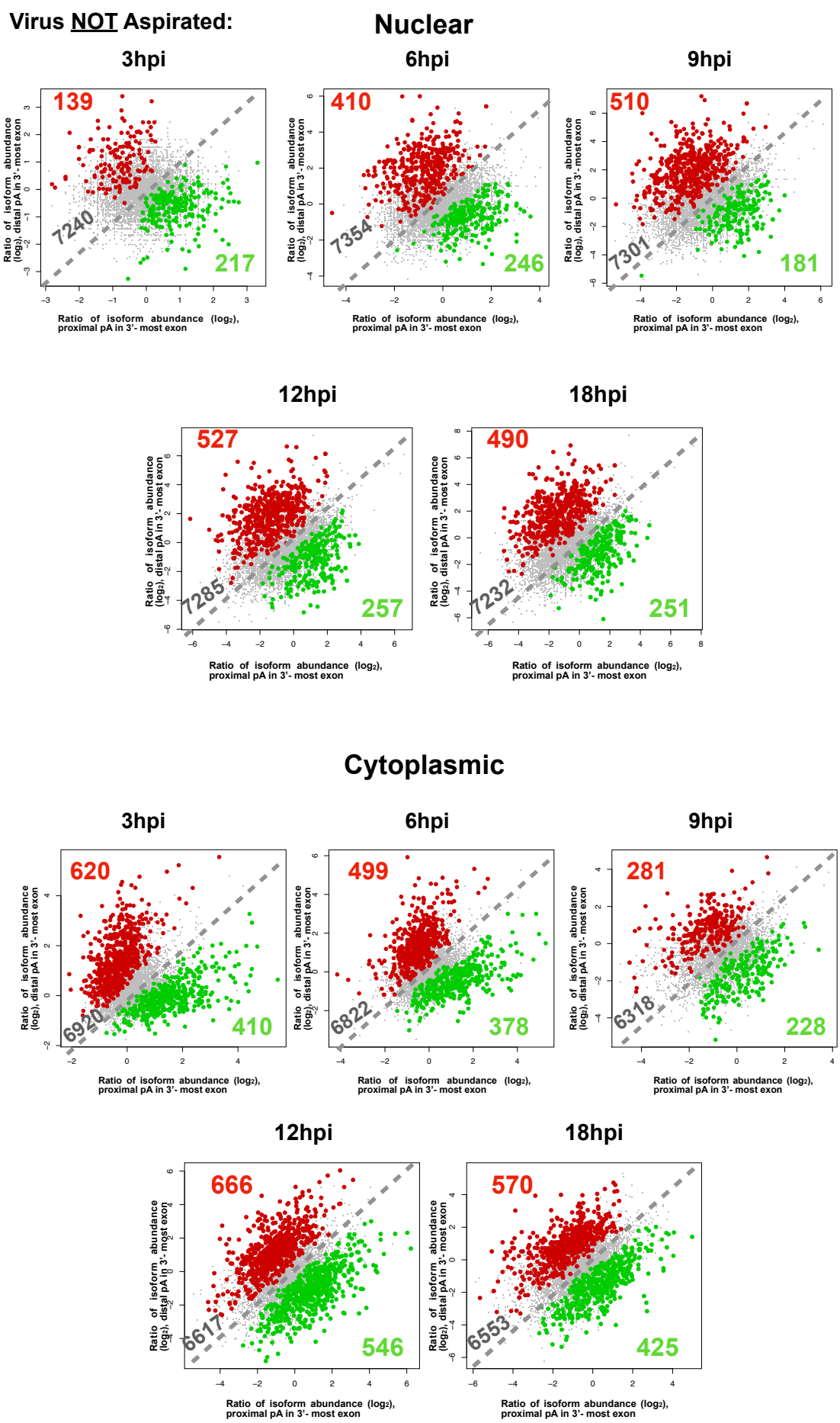
Despite having added knowledge to both the APA and viral field, there are still many unanswered questions. A future challenge is to discover exactly how the Influenza A virus distinguishes between different cellular mRNAs in their regulation for its own benefit. Additionally, it would be interesting to discover whether APA events are conserved between viral strains. Finally, to further highlight APA's importance, assessing other mechanisms involved in differential CPA site usage will give further insight to any unexplored viral host shut off strategies.

Chapter 7: Supplementary Data



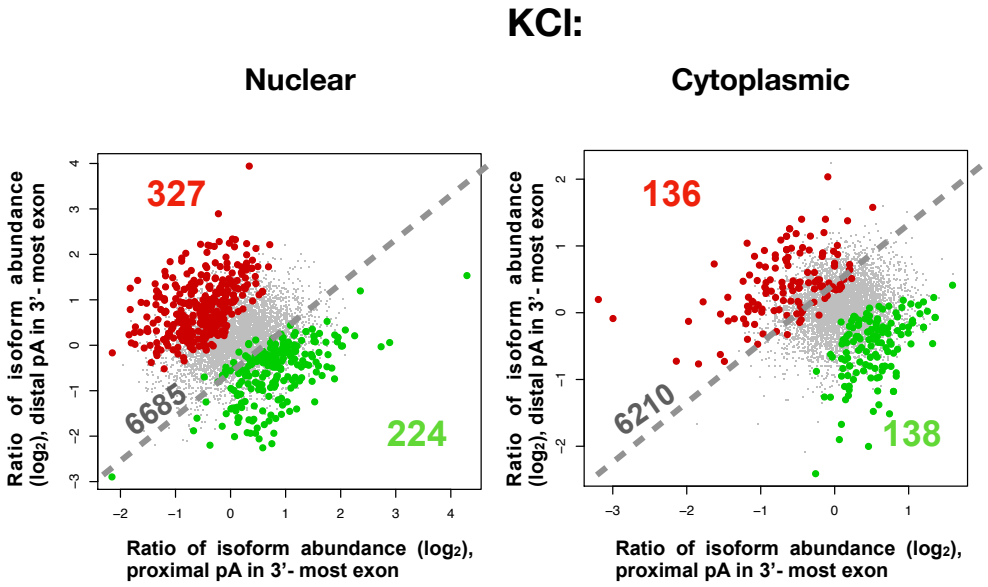
Supplementary Figure 1: 3'UTR APA in 'Aspirated' Infection strategy

Statistically significant 3'UTR APA events (p<0.01 Fisher's exact Test, with a minimum 5% change in APA isoform abundance) identified in WSN/33 infection (virus aspirated after 1 hour RTP) in 5 time points compared to mock in the nucleus and cytoplasm . Red dots: 3'UTR lengthening events; Green dots: 3'UTR shortening events. Grey dots: Total events



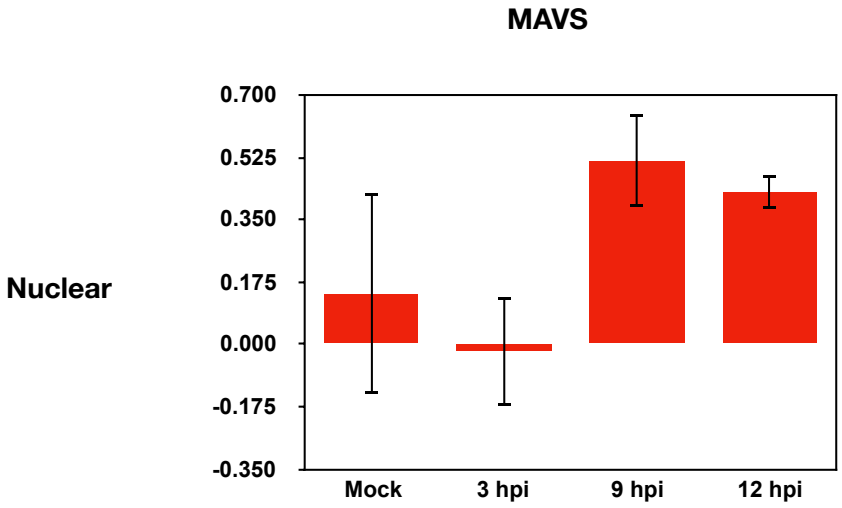
Supplementary Figure 2: 3'UTR APA in 'non-aspirated' Infection strategy (combined)

Statistically significant 3'UTR APA events (p<0.01 Fisher's exact Test, with a minimum 5% change in APA isoform abundance) identified in WSN/33 infection (virus non-aspirated after 1 hour RTP) in 5 time points compared to mock in the nucleus and cytoplasm over 2 independent biological repeats. Red dots: 3'UTR lengthening events; Green dots: 3'UTR shortening events. Grey dots: Total events



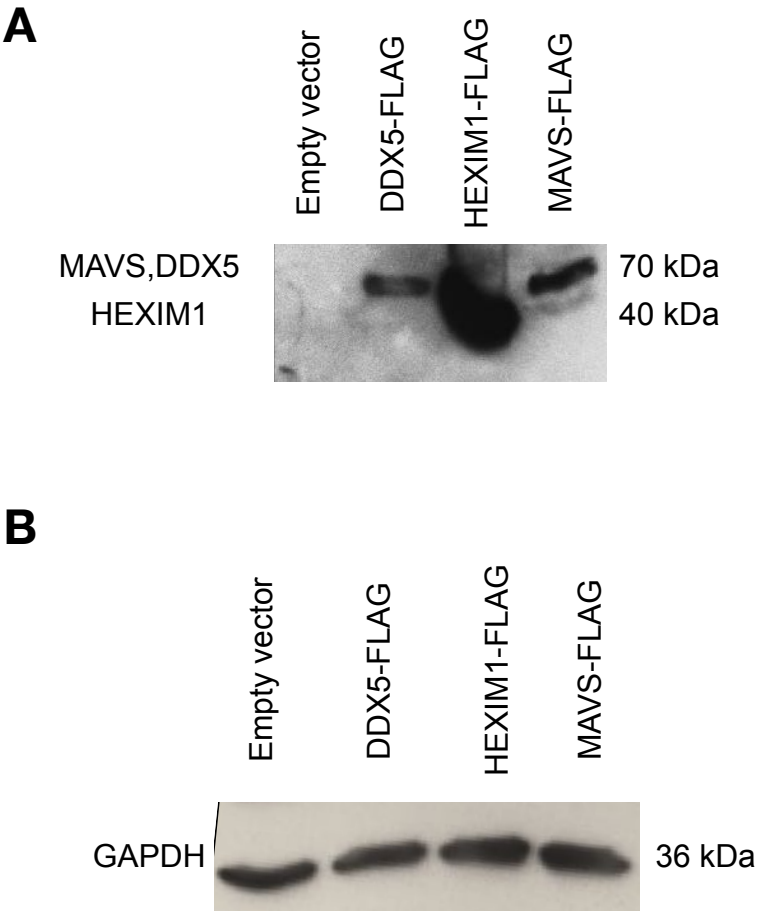
Supplementary Figure 3: 3'UTR APA in Osmotic Shock

Statistically significant 3'UTR APA events (p<0.01 Fisher's exact Test, with a minimum 5% change in APA isoform abundance) identified after osmotic shock compared to control in the nucleus and cytoplasm . Red dots: 3'UTR lengthening events; Green dots: 3'UTR shortening events. Grey dots: Total events



Supplementary Figure 4: Real time qPCR to verify 3'UTR APA of MAVS in nucleus, however Results are Insignificant

Real-time q-PCR analysis of MAVS upon IAV infection in A549 cells in nucleus (red). The x-axis denotes 3 time points post infection (3,9,12 hpi) and mock, and the y-axis denotes the $\log_2$ ratio of longer 3' UTR isoforms (L) to total mRNA isoforms (T). Data are presented as mean log2 ratio L/T with error bars representing SEM of 3 biological repeats. Paired, 2 tailed t-test performed to compare groups, with P<0.05 considered significant. MAVS qPCR could not show significant lengthening at 9hpi and 12 hpi when compared to mock in three independent nuclear repeats.

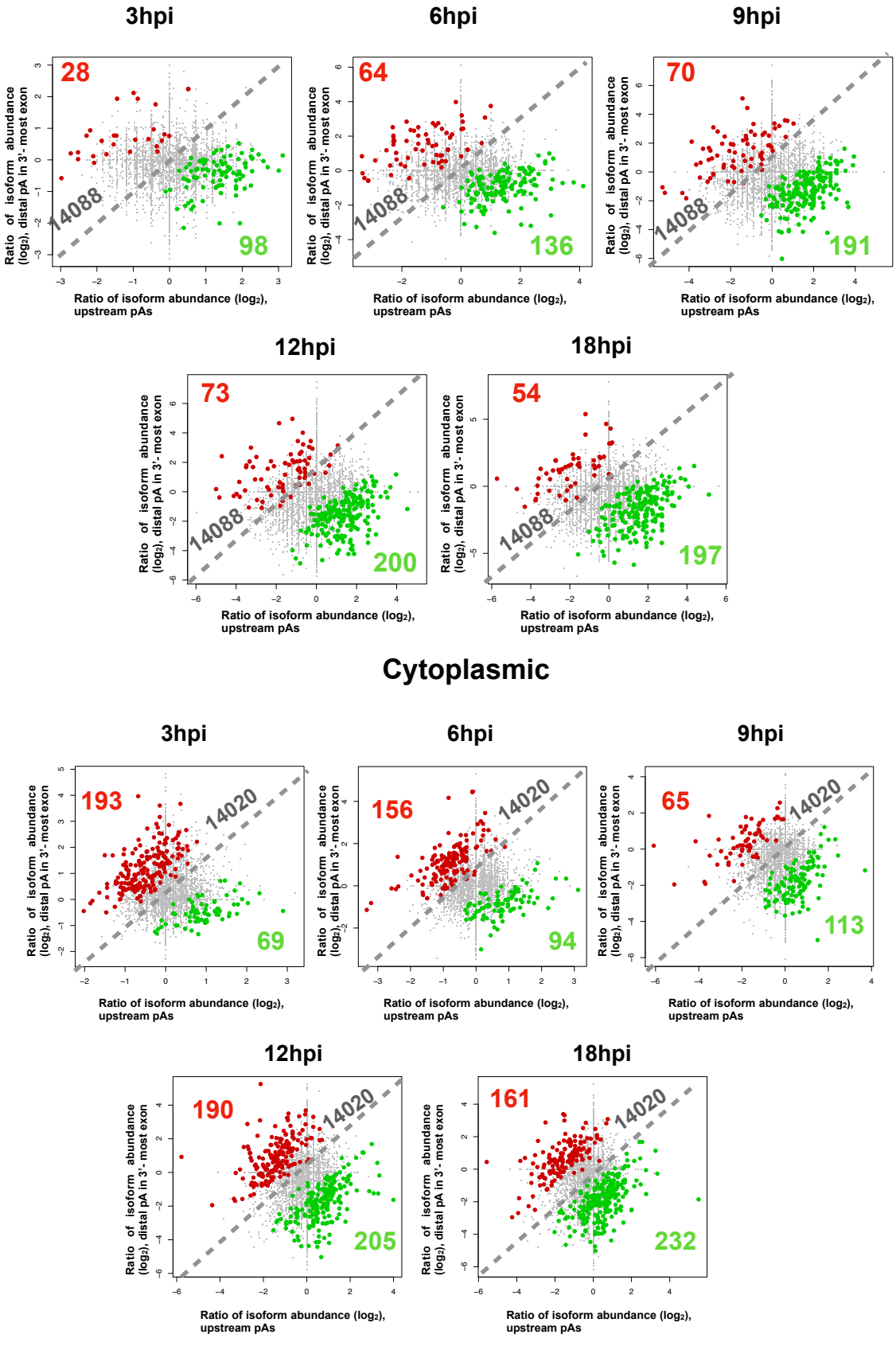


Supplementary Figure 5: Western Blot Verifying Constitutive Over-Expression of DDX5, HEXIM1 and MAVS (performed by Ben Nilsson)

Western blotting of FLAG-tagged proteins (DDX5, HEXIM1 and MAVS) in order to verify constitutive over expression.

A. Detection of FLAG-tag incorporated in each protein were performed with mouse monoclonal anti-GLAG-M2 (Sigma, F3165) and HRP-conjugated donkey polyclonal anti-mouse IgG (1:5000) (NA931V) secondary antibody. Sizes of MAVS and DDX5 are reported to be around 70kDA, HEXIM1 size is around 40kDa.

B. GAPDH loading control was verified using rabbit polyclonal anti-GAPDH (Sigma, G9545) with HRP-conjugated donkey polyclonal anti-rabbit IgG (1:5000) secondary antibody (NA934V). Size of GAPDH is 36kDA. Equal loading has been shown.

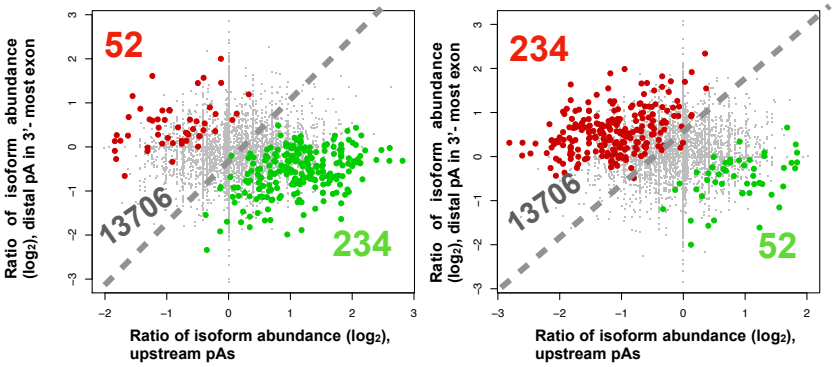


Supplementary Figure 6: CR-APA in infection

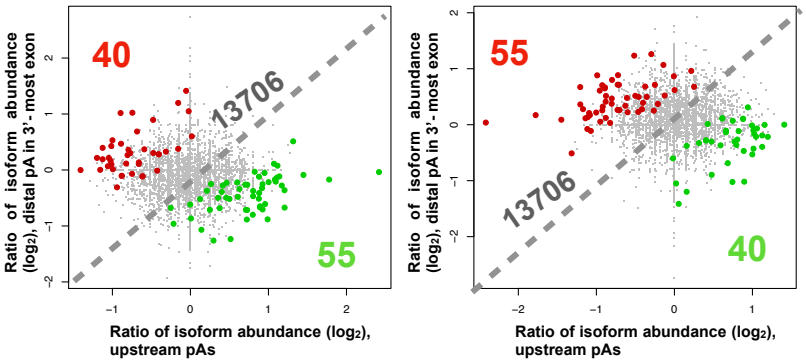
Different trends to using upstream (green dots) or terminal exon (red dots) CPA sites in the nucleus (top half) and the cytoplasm (bottom half) during WSN/33 infection. Total events depicted as grey dots.

KCl CR-APA :

Nuclear

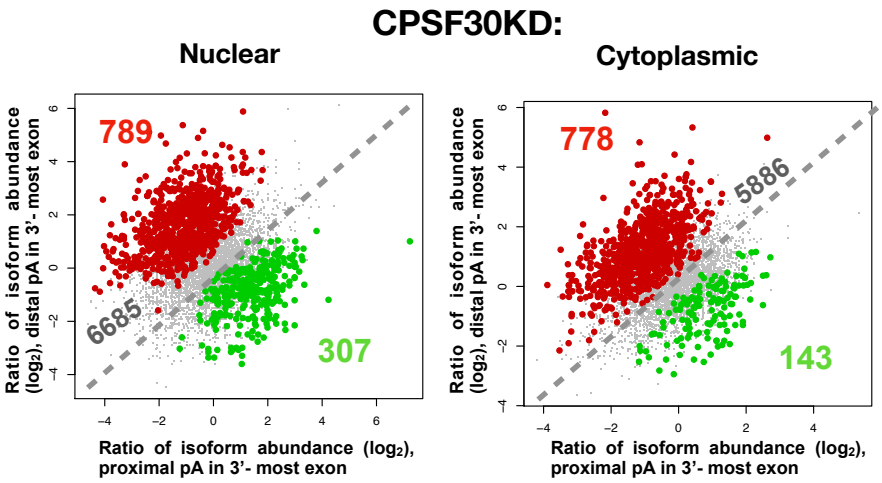


Cytoplasmic



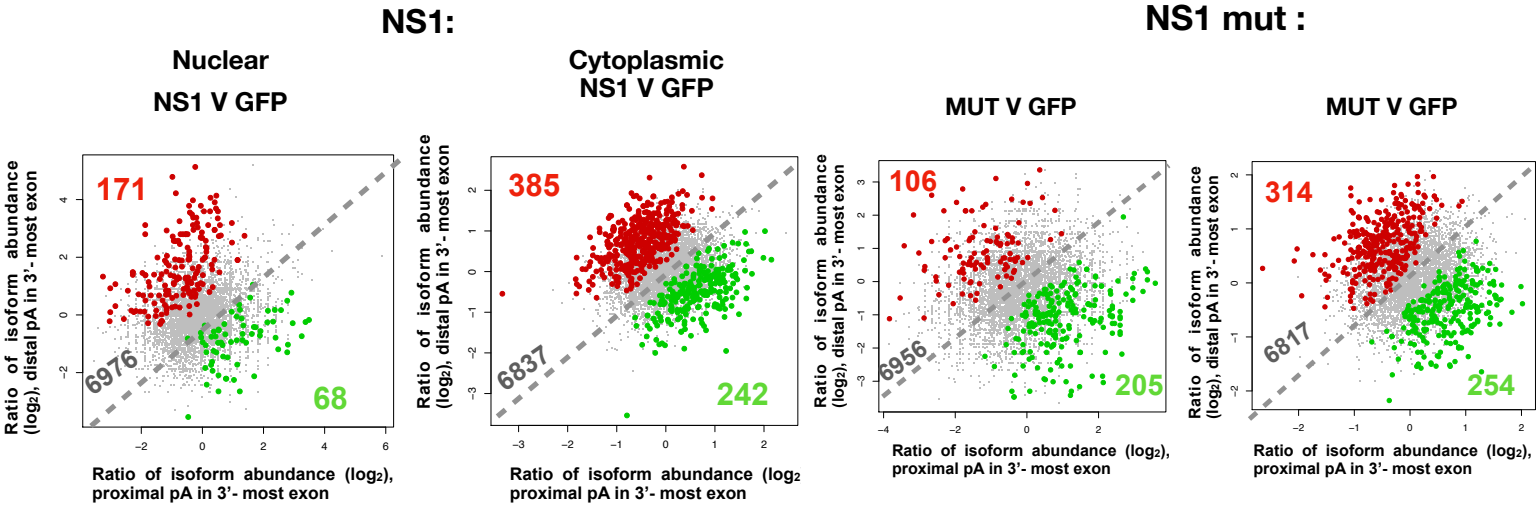
Supplementary Figure 7 CR-APA in Osmotic Shock

Different trends to using upstream (green dots) or terminal exon (red dots) CPA sites in the nucleus (Top half) and the cytoplasm (Bottom half) during osmotic shock.



Supplementary Figure 8: 3'UTR APA in CPSF30 Knock Down

Statistically significant 3'UTR APA events (p<0.01 Fisher's exact Test, with a minimum 5% change in APA isoform abundance) identified in CPSF30KD compared to scrambled control in the nucleus (left) and cytoplasm (right) . Red dots: 3'UTR lengthening events; Green dots: 3'UTR shortening events. Grey dots: Total events



Supplementary Figure 9: 3'UTR APA in Transient Over-Expression of WT WSN/33 NS1 and NS1 G184R mut

Statistically significant 3'UTR APA events (p<0.01 Fisher's exact Test, with a minimum 5% change in APA isoform abundance) identified in WT NS1 and NS1 mut over expression compared to GFP control in the nucleus (left) and cytoplasm (right) . Red dots: 3'UTR lengthening events; Green dots: 3'UTR shortening events. Grey dots: Total events

Appendix A: WSN33 genome

>PB1
AGCGAAAGCAGGCAAACCATTTGAATGGATGTCAATCCGACTTTACTTTTCTTAAAAGTGCCAGCACAAA
ATGCTATAAGCACAACTTTCCCTTATACTGGAGACCCTCCTTACAGCCATGGGACAGGAACAGGATACAC
CATGGATACTGTCAACAGGACACATCAGTACTCAGAAAGGGGAAGATGGACAACAAACACCGAAACTGGA
GCACCGCAACTCAACCCGATTGATGGGCCACTGCCAGAAGACAATGAACCAAGTGGTTATGCCCAAACAG
ATTGTGTATTGGAAGCAATGGCCTTCCTTGAGGAATCCCATCCTGGTATCTTTGAGACCTCGTGTCTTGA
AACGATGGAGGTTGTTCAGCAAACACGAGTGGACAAGCTGACACAAGGCCGACAGACCTATGACTGGACT
CTAAATAGGAACCAGCCTGCTGCAACAGCATTGGCCAACACAATAGAAGTGTTCAGATCAAATGGCCTCA
CGGCCAATGAATCtGGAAGGCTCATAGACTTCCTTAAGGATGTAATGGAGTCAATGAACAAAGAAGAAAT
GGAGATCACAACTCATTTTCAGAGAAAGAGACGAGTGAGAGACAATATGACTAAGAAAATGGTGACACAG
AGAACAATAGGTAAAAGGAAGCAGAGATTGAACAAAAGGAGTTATCTAATTAGGGCATTaACCCTGAACA
CAATGACCAAAGATGCTGAGAGAGGGAAGCTAAAACGGAGAGCAATTGCAACCCCAGGGATGCAAATAAG
GGGGTTTGTATACTTTGTTGAGACACTAGCAAGGAGTATATGTGAGAAACTTGAACAATCAGGATTGCCA
GTTGGAGGCAATGAGAAGAAAGCAAAGTTGGCAAATGTTGTAAGGAAGATGATGACCAATTCTCAGGACA
CTGAAATTTCTTTCACCATCACTGGAGATAACACCAAATGGAACGAAAATCAGAACCCTCGGATGTTTTT
GGCCATGATCACATATATAACCAGAAATCAGCCCGAATGGTTCAGAAATGTTCTAAGTATTGCTCCAATA
ATGTTCTCAAACAAAATGGCGAGACTGGGAAAGGGGTACATGTTTGAGAGCAAGAGTATgAAAcTTAGAA
CTCAAATACCTGCAGAAATGCTAGCAAGCATCGATTTGAAATACTTCAATGATTCAACTAGAAAGAAGAT
TGAAAAAATCCGGCCGCTCTTAATAGATGGGACTGCATCATTGAGCCCTGGAATGATGATGGGCATGTTC
AATATGTTAAGTACTGTATTAGGCGTCTCCATCCTGAATCTTGGACAAAAGAGACACACCAAGACTACTT
ACTGGTGGGATGGTCTTCAATCTTCTGATGATTTTGCTCTGATTGTGAATGCACCCAATCATGAAGGGAT
TCAAGCCGGAGTCAACAGGTTTTATCGAACCTGTAAGCTACTTGGAATTAATATGAGCAAGAAAAAGTCT
TACATAAACAGAACAGGTACATTTGAATTCACAAGTTTTTTCTATCGTTATGGGTTTGTTGCCAATTTCA
GCATGGAGCTTCCCAGCTTTGGGGTGTCTGGGATCAACGAGTCTGCGGACATGAGTATTGGAGTTACTGT
CATCAAAAACAATATGATAAACAATGATCTTGGTCCAGCAACCGCTCAAATGGCCCTTCAGCTGTTCATC
AAAGATTACAGGTACACGTACCGGTGCCATAGAGGTGACACACAAATACAAACCCGAAGATCATTTGAAA
TAAAGAAACTGTGGGAGCAAACCCATTCCAAAGCTGGACTGCTGGTCTCCGACGGAGGCCCAAATTTATA
CAACATTAGAAATCTCCACATTCCTGAAGTCTGCTTGAAATGGGAATTAATGGATGAGGATTACCAGGGG
CGTTTATGCAACCCACTGAACCCATTTGTCAACCATAAAGACATTGAATCAGTGAACAATGCAGTGATAA
TGCCAGCACATGGTCCAGCCAAAAACATGGAGTATGATGCTGTTGCAACAACACAtTCCTGGATCCCCAA
AAGAAATCGATCCATCTTGAATACAAGCCAAAGAGGAATACTTGAAGATGAACAAATGTACCAAAAGTGC
TGCAACTTATTTGAAAAATTCTTCCCCAGCAGTTCATACAGAAGACCAGTCGGGATATCCAGTATGGTGG
AGGCTATGGTTTCCAGAGCCCGAATTGATGCACGAATTGATTTCGAATCTGGAAGGATAAAGAAAGAGGA
GTTCACTGAGATCATGAAGATCTGTTCCACCATTGAAGAGCTCAGACGGCAAAAATAGTGAATTTAGCTT
GTCCTTCATGAAAAAATGCCTTGTTTCTACT
>PB2
AGCGAAAGCAGGTCAATTATATTCAATATGGAAAGAATAAAAGAACTAAGGAATCTAATGTCGCAGTCTC
GCACTCGCGAGATACTCACAAAAACCACCGTGGACCATATGGCCATAATCAAGAAGTACACATCAGGAAG
ACAGGAGAAGAACCCAGCACTTAGGATGAAATGGATGATGGCAATGAAATATCCAATTACAGCAGACAAG
AGGATAACGGAAATGATTCCTGAGAGAAATGAGCAGGGACAAACTTTATGGAGTAAAATGAATGACGCCG
GATCAGACCGAGTGATGGTATCACCTCTGGCTGTGACATGGTGGAATAGGAATGGACCAGTGACAAGTAC
AGTTCATTATCCAAAAATCTACAAAACTTATTTTGAAAAAGTCGAAAGGTTAAAACATGGAACCTTTGGC
CCTGTCCATTTTAGAAACCAAGTCAAAATACGTCGAAGAGTTGACATAAATCCTGGTCATGCAGATCTCA
GTGCCAAAGAGGCACAGGATGTAATCATGGAAGTTGTTTTCCCTAACGAAGTGGGAGCCAGGATACTAAC
ATCGGAATCGCAACTAACGACAACCAAAGAGAAGAAAGAAGAACTCCAGGGTTGCAAAATTTCTCCTCTG
ATGGTGGCATACATGTTGGAGAGAGAACTGGTCCGCAAAACGAGATTCCTCCCAGTGGCTGGTGGAACAA
GCAGTGTGTACATTGAAGTGTTGCATTTGACCCAAGGAACATGCTGGGAACAGATGTACACTCCAGGAGG
GGAGGCGAGGAATGATGATGTTGATCAAAGCTTAATTATTGCTGCTAGAAACATAGTAAGAAGAGCCACA
GTaTCAGCAGATCCACTAGCATCTTTATTGGAGATGTGCCACAGCACGCAGATTGGTGGAATAAGGATGG
TAAACATCCTTAGGCAGAACCCAACAGAAGAGCAAGCCGTGGATATTTGCAAGGCTGCAATGGGACTGAG
AATTAGCTCATCCTTCAGTTTTGGTGGATTCACATTTAAGAGAACAAGCGGATCATCAGTCAAGAGAGAG
GAAGAGGTGCTTACGGGCAATCTTCAGACATTGAAGATAAGAGTGCATGAGGGATATGAAGAGTTCACAA
TGGTTGGGAGAAGAGCAACAGCTATACTCAGAAAAGCAACCAGGAGATTGATTCAGCTGATAGTGAGTGG
GAGaGACGAACAGTCGATTGCCGAAGCAATAATTGTGGCCATGGTATTTTCACAAGAGGATTGTATGATA
AAAGCAGTTAGAGGTGACCTGAATTTCGTCAATAGGGCGAATCAGCGATTGAAtCCCATGCACCAACTTT
TGAGACATTTTCAGAAGGATGCAAAGGtGCTCTTTCAAAATTGGGGAATTGAATCCATCGACAATGTGAT
GGGAATGATCGGGATATTGCCCGACATGACTCCAAGCACCGAGATGTCAATGAGAGGAGTGAGAATCAGC
AAAATGGGGGTAGATGAGTATTCCAGCGCGGAGAAGATAGTGGTGAGCATTGACCGTTTTTTGAGAGTTA
GGGACCAACGTGGGAATGTACTACTGTCTCCCGAGGAGGTCAGTGAAACACAGGGAACAGAGAAACTGAC
AATAACTTACTCATCGTCAATGATGTGGGAGATTAATGGTCCTGAATCAGTGTTGGTCAATACCTATCAG
TGGATtATCAGAAACTGGGAAACTGTTAAAATTCAGTGGTCCCAGAATCCTACAATGCTGTACAATAAAA
TGGAATTTGAGCCATTTCAGTCTTTAGTTCCAAAGGCCGTTAGAGGCCAATACAGTGGGTTTGTGAGAAC
TCTGTTCCAACAAATGAGGGATGTGCTTGGGACATTTGATACCGCTCAGATAATAAAACTTCTTCCCTTC
GCAGCCGCTCCACCAAAGCAAAGTGGAATGCAGTTCTCCTCATTGACTATAAATGTGAGGGGATCAGGAA

TGAGAATACTTGTAAGGGGCAATTCTCCAgTATTCAACTACAACAAGACCACTAAAAGACTCACAGTTCT
CGGAAAGGATGCTGGCCCTTTAACTGAAGACCCAGATGAAGGCACAGCTGGAGTTGAGTCCGCAGTTCTG
AGAGGATTCCTCATTCTGGGCAAAGAAGACAGGAGATATGGACCAGCATTAAGCATAAATGAACTGAGCA
ACCTTGCGAAAGGAGAGAAGGCTAATGTGCTAATTGGGCAAGGAGACGTGGTGTTGGTAATGAAACGGAA
ACGGAACTCTAGCATACTTACTGACAGCCAGACAGCGACCAAAAGAATTCGGATGGCCATCAATTAGTGT
CGAATAGTTTAAAAACGACCTTGTTTCTACT
>PA
AGCGAAAGCAGGTACTGATTCAAAATGGAAGATTTTGTGCGACAATGCTTCAATCCGATGATTGTCGAGC
TTGCGGAAAAGGCAATGAAAGAGTATGGAGAGGACCTGAAAATCGAAACAAACAAATTTGCAGCAATATG
CACTCACTTGGAAGTGTGCTTCATGTATTCAGATTTTCACTTCATCGATGAGCAAGGCGAGTCAATAGTC
GTAGAACTTGGCGATCCAAATGCACTTTTGAAGCACAGATTTGAAATAATCGAGGGAAGAGATCGCACAA
TAGCCTGGACAGTAATAAACAGTATTTGCAACACTACAGGGGCTGAGAAACCAAAGTTTCTACCAGATTT
GTATGATTACAAGAAGAATAGATTCATCGAAATTGGAGTAACAAGGAGAGAAGTTCACATATACTATCTG
GAAAAGGCCAATAAAATTAAATCTGAGAAGACACACATCCACATTTTCTCATTCACTGGGGAGGAAATGG
CCACAAAGGCCGACTACACTCTCGATGAAGAAAGCAGGGCTAGGATCAAAACCAGGCTATTCACCATAAG
ACAAGAAATGGCTAGCAGAGGCCTCTGGGATTCCTTTCGTCAGTCCGAGAGAGGCGAAGAGACAATTGAA
GAAAGATTTGAAATCACAGGAACAATGCGCAAGCTTGCCGACCAAAGTCTCCCGCCAAACTTCTCCAGCC
TTGAAAATTTTAGAGCCTATGTGGATGGATTCGAACCGAACGGCTACATTGAGGGCAAGCTTTCTCAAAT
GTCCAAAGAAGTAAATGCTAGAATTGAACCTTTTTTGAAATCAACACCACGACCACTTAGACTTCCGGAT
GGGCCTCCCTGTTCTCAGCGGTCCAAATTCCTGCTGATGGATGCCTTAAAATTAAGCATTGAGGACCCAA
GTCATGAGGGAGAGGGGATACCGCTATATGATGCAATCAAATGCATGAGAACATTCTTTGGATGGAAGGA
ACCCAATGTTGTTAAACCACACGAAAAGGGAATAAATCCAAATTATCTTCTGTCATGGAAGCAAGTACTG
GCAGAACTGCAGGACATTGAGAATGAGGAGAAAATTCCAAGGACTAAAAATATGAAGAAAACGAGTCAGT
TAAAGTGGGCACTTGGTGAGAACATGGCACCAGAAAAGGTAGACTTTGACGATTGTAAAGATGTAGGCGA
TTTGAAGCAATATGATAGTGATGAACCAGAATTGAGGTCGCTTGCAAGTTGGATTCAGAATGAGTTCAAC
AAGGCATGTGAACTGACCGATTCAAGCTGGATAGAGCTCGATGAGATTGGAGAAGATGCGGCTCCAATTG
AACACATTGCAAGCATGAGAAGGAATTATTTCACAGCAGAGGTGTCTCATTGCAGAGCCACAGAATACAT
AATGAAGGGGGTGTACATCAATACTGCCTTGCTTAATGCATCCTGTGCAGCAATGGATGATTTCCAATTg
ATTCCAATGATAAGCAAGTGTAGAACTAAGGAGGGAAGGCGAAAGACCAATTTGTACGGTTTCATCATAA
AAGGAAGATCCCACTTAAGGAATGACACCGATGTGGTAAACTTTGTGAGCATGGAGTTTTCCCTCACTGA
CCCAAGACTTGAACCACACAAATGGGAGAAGTACTGTGTTCTTGAGGTAGGAGATATGCTTCTAAGAAGT
GCCATAGGCCATGTGTCAAGGCCTATGTTCTTGTATGTGAGGACAAATGGAACCTCAAAAATTAAAATGA
AATGGGGGATGGAAATGAGGCGTTGCCTCCTTCAGTCACTTCAACAAATCGAGAGTATGATTGAAGCTGA
GTCCTCTGTCAAGGAGAAAGACATGACCAAAGAGTTCTTTGAAAACAAATCAGAAACATGGCCCGTTGGA
GAGTCCCCCAAAGGAGTGGAGGAAGGTTCCATTGGGAAGGTCTGCAGAACTTTATTGGCAAAGTCGGTAT
TCAACAGCTTGTATGCATCTCCACAACTGGAAGGATTTTCAGCTGAATCAAGAAAACTGCTTCTTATCGT
TCAGGCTCTTAGGGACAACCTGGAACCTGGGACCTTTGATCTTGGGGGGCTATATGAAGCAATTGAGGAG
TGCCTGATTAATGATCCCTGGGTTTTGCTTAATGCTTCTTGGTTCAACTCCTTCCTCACACATGCATTGA
GATAGTTGTGGCAATGCTACTATTTGCTATCCATACTGTCCAAAAAAGTACCTTGTTTCTACT
>HA
AGCAAAAGCAGGGGAAAATAAAAACAACCAAAATGAAGGCttttgTACTgGTCCTGTTATATGCATTTGT
AGCTACAGATGCAGACACAATATGTATAGGCTACCATGCGAACAACTCAACCGACACTGTTGACACAATA
TTCGAGAAGAATGTGGCAGTGACACATTCTGTTAACCTGCTCGAAGACAGACACAACGGGAAACTATGTA
AATTAAAAGGAATAGCCCCACTACAATTGGGGAAATGTAACATCACCGGATGGCTCTTGGGAAATCCAGA
ATGCGACTCACTGCTTCCAGCGAGATCATGGTCCTACATTGTAGAAACACCAAACTCTGAGAATGGAGCA
TGTTATCCAGGAGATTTCATCGACTATGAGGAACTGAGGGAGCAATTGAGCTCAGTATCATCATTAGAAA
GATTCGAAATATTTCCCAAGGAAAGTTCATGGCCCAACCACACATTCAACGGAGTAACAGTATCATGCTC
CCATAGGGGAAAAAGCAGTTTTTACAGAAATTTGCTATGGCTGACGAAGAAGGGGGATTCATACCCAAAG
CTGACCAATTCCTATGTGAACAATAAAGGGAAAGAAGTCCTTGTACTATGGGGTGTTCATCACCCGTCtA
GCAGTGATGAGCAACAGAGTCTCTATAGTAATGGAAATGCTTATGTCTCTGTAGCGTCTTCAAATTATAA
CAGGAGATTCACCCCGGAAATAGCTGCAAGGCCCAAAGTAAAAGATCAACATGGGAGGATGAACTATTAC
TGGACCTTGCTAGAACCCGGAGACACAATAATATTTGAGGCAACTGGTAATCTAATAGCACCATGGTATG
CTTTCGCACTGAGTAGAGGGTTTGAGTCCGGCATCATCACCTCAAACGCGTCAATGCATGAGTGTAACAC
GAAGTGTCAAACACCCCAGGGATCTATAAACAGCAATCTCCCTTTCCAGAATATACACCCAGTCACAATA
GGAGAGTGCCCAAAATATGTCAGGAGTACCAAATTGAGGATGGTTACAGGACTAAGAAACATCCCATCCA
TTCAATACAGAGGTCTATTTGGAGCCATTGCTGGTTTTATTGAGGGGGGATGGACTGGAATGATAGATGG
ATGGTATGGTTATCATCATCAGAATGAACAGGGATCAGGCTATGCAGCGGATCAAAAAAGCACACAGAAT
GCCATTAACGGGATTACAAACAAGGTGAACTCTGTTATCGAGAAAATGAACACTCAATTCACAGCTGTGG
GTAAAGAATTCAACAACTTAGAAAAAAGGATGGAAAATTTAAATAAAAAAGTTGATGATGGGTTTCTGGA
CATTTGGACATATAATGCAGAATTGTTAGTTCTACTGGAAAATGAAAGAACTTTGGATTTCCATGACTTA
AATGTGAAGAATCTGTACGAGAAAGTAAAAAGCCAATTAAAGAATAATGCCAAAGAAATCGGAAATGGGT
GTTTTGAGTTCTACCACAAGTGTGACAATGAATGCATGGAAAGTGTAAGAAATGGGACTTATGATTATCC
AAAATATTCAGAAGAATCAAAGTTGAACAGGGAAAAGATAGATGGAGTGAAATTGGAATCAATGGGGGTG
TATCAGATTCTGGCGATCTACTCAACTGTCGCCAGTTCACTGGTGCTTTTGGTCTCCCTGGGGGCAATCA
GTTTCTGGATGTGTTCTAATGGGTCTTTGCAGTGCAGAATATGCATCTGAGATTAGGATTTCAGAAATAT
AAGGAAAAACACCCTTGTTTCTACT

>NP
AGCAAAAGCAGGGTAGATAATCACTCACAGAGTGACATCGAAATCATGGCGACCAAAGGCACCAAACGAT
CTTACGAACAGATGGAGACTGATGGAGAACGCCAGAATGCCACTGAAATCAGAGCATCTGTCGGAAAAAT
GATTGaTGGAATTGGACGATTCTACATCCAAATGTGCACCGAACTTAAACTCAGTGATTATGAGGGACGG
CTGATTCAGAACAGCTTAACAATAGAGAGAATGGTGCTCTCTGCTTTTGACGAGAGGAGGAATAAATATC
TAGAAGAACATCCCAGTGCGGGGAAAGATCCTAAGAAAACTGGAGGACCTATATACAGGAGAGTAGATGG
AAAGTGGAgGAGAGAACTCATCCTTTATGACAAAGAAGAAATAAGACGAATCTGGCGCCAAGCTAATAAT
GGTGACGATGCAACGGCTGGTCTGACTCACATGATGATCTGGCACTCCAATTTGAATGATGCAACTTACC
AGAGGACAAGAGCTCTTGTTCGCACAGGAATGGATCCCAGGATGTGCTCACTGATGCAGGGTTCAACCCT
CCCTAGGAGGTCTGGGGCCGCAGGTGCTGCAGTCAAAGGAGTTGGAACAATGGTGATGGAATTGATCAGA
ATGATCAAACGTGGGATCAATGATCGGAACTTCTGGAGGGGTGAGAATGGACGGAGAACAAGGATTGCTT
ATGAAAGAATGTGCAACATTCTCAAAGGGAAATTTCAAACAGCTGCACAAAGAaCAATGGTGGATCAAGT
GAGAGAGAGCCGGAATCCAGGAAATGCTGAGTTCGAAGATCTCATCTTTTtTAGCACGGTCTGCACTCATA
TTGAGAGGGTCAGTTGCTCACAAGTCCTGCCTGCCTGCCTGTGTGTATGGAtCTGCCGTAGCCAGTGGAT
ACGACTTTGAAAGAGAGGGATACTCTCTAGTCGGAATAGACCCTTTCAGACTGCTTCAAAACAGCCAAGT
ATACAGCCTAATCAGACCAAATGAGAATCCAGCACACAAGAGTCAACTGGTGTGGATGGCATGCCATTCT
GCTGCATTTGAAGATCTAAGAGTATCAAGCTTCATCAGAGGGACGAAAGTGGTCCCAAGAGGGAAGCTTT
CCACTAGAGGAGTTCAAATTGCTTCCAATGAAAACATGGAGACTATGGAATCAAGTACCCTTGAACTGAG
AAGCAGATACTGGGCCATAAGGACCAGAAGTGGAGGGAACACCAATCAACAGAGGGCTTCCTCGGGCCAA
ATCAGCATACAACCTACGTTCTCAGTACAGAGAAATCTCCCTTTTGACAGACCAACCATTATGGCAGCAT
TCACTGGGAATACAGAGGGGAGAACATCTGACATGAGAACCGAAATCATAAGGCTGATGGAAAGTGCAAG
ACCAGAAGATGTGTCTTTCCAGGGGCGGGGAGTCTTCGAGCTCTCGGACGAAAAGGCAaCGAGCCCGATC
GTGCCCTCCTTTGACATGAGTAATGAAGGATCTTATTTCTTCGGAGACAATGCAGAGGAGTACGACAATT
AAAGAAAAATACCCTTGTTTCTACT
>NA
AGCGAAAGCAGGAGTTTAAATGAATCCAAACCAGAAAATAATAACCATTGGaTCgATCTGTATGGTAGTC
GGAATAATTAGCCTAATATTGCAAATAGGAAATATAATCTCAATATGGATTAGCCATTCAATTCAAACCG
GAAATCAAAACCATACTGGAATATGCAACCAAGGCAGCATTACCTATAAAGTTGTTGCTGGGCAGGACTC
AACTTCAGTGATATTAACCGGCAATTCATCTCTTTGTCCCATCCGTGGGTGGGCTATACACAGCAAAGAC
AATGGCATAAGAATTGGTTCCAAAGGAGACGTTTTTGTCATAAGAGAGCCTTTTATTTCATGTTCTCACT
TGGAATGCAGGACCTTTTTTCTGACTCAAGGCGCCTTACTGAATGACAAGCATTCAAGGGGGACCTTTAA
GGACAGAAGCCCTTATAGGGCCTTAATGAGCTGCCCTGTCGGTGAAGCTCCGTCCCCGTACAATTCAAGG
TTTGAATCGGTTGCTTGGTCAGCAAGTGCATGTCATGATGGAaTGGGCTGGCTAACAATCGGAATTTCTG
GTCCAGATGATGGAGCAGTGGCTGTATTAAAATACAACgGCATAATAACTGAAACCATAAAAAGTTGGAG
GAAGAATATATTGAGAACACAAGAGTCTGAATGTACCTGTGTAAATGGTTCATGTTTTACCATAATGACC
GATGGCCCAAGTGATGGGCTGGCCTCGTACAAAATTTTCAAGATCGAGAAGGGGAAGGTTACTAAATCAA
TAGAGTTGAATGCACCTAATTCTCACTACGAGGAATGTTCCTGTTACCCTGATACCGGCAAAGTGATGTG
TGTGTGCAGAGACAATTGGCACGGTTCGAACCGACCATGGGTGTCCTTCGACCAAAACCTAGATTATAAA
ATAGGATACATCTGCAGTGGGGTTTTCGGTGACAACCCGCGTCCCAAAGATGGAACAGGCAGCTGTGGCC
CAGTGTCTGCTGATGGAGCAAACGGAGTAAAGGGATTTTCATATAAGTATGGtAATGGTGTTTGGATAGG
AAGGACTAAAAGTGACAGTTCCAGACATGGGTTTGAGATGATTTGGGATCCTAATGGATGGACAGAGACT
GATAGTAGGTTCTCTATGAGACAAGATGTTGTGGCAATgACTgATCGGTCAGGGTACAGCGGAAGTTTCG
TTCAACATCCTGAGCTAACAGGGCTAGACTGTATGAGGCCTTGCTTCTGGGTTGAATTAATCAGGGGGCT
ACCTGAGGAGGACGCAATCTGGACTAGTGGGAGCATCATTTCTTTTTGTGGTGTGAATgGTGATACTGTA
GATTGGTCTTGGCCAGACGGTGCTGAGcTcCCGTTCACCATTGACAAGTAGTTTGTTCAAAAAACTCCTT
GTTTCTACT
>M
AGCAAAAGCAGGTAGATATTGAAAGATGAGTCTTCTAACCGAGGTCGAAACGTACGTTCTCTCTATCGTC
CCGTCAGGCCCCCTCAAAGCCGAGATCGCACAGAGACTTGAAGATGTCTTTGCAGGGAAGAACACCGATC
TTGAGGTTCTCATGGAATGGCTAAAGACAAGACCAATCCTGTCACCTCTGACTAAGGGGATTTTAGGATT
TGTGTTCACGCTCACCGTGCCCAGTGAGCGGGGACTGCAGCGTAGACGCTTTGTCCAAAATGCTCTTAAT
GGGAACGGAGATCCAAATAACATGGACAAAGCAGTTAAACTGTATAGGAAGCTTAAGAGGGAGATAACAT
TCCATGGGGCCAAAGAAATAGCACTCAGTTATTCTGCTGGTGCACTTGCCAGTTGTATGGGCCTCATATA
CAACAGGATGGGGGCTGTGACCACTGAAGTGGCATTTGGCCTGGTATGCGCAACCTGTGAACAGATTGCT
GACTCCCAGCATCGGTCTCATAGGCAAATGGTGACAACAACCAATCCACTAATCAGACATGAGAACAGAA
TGGTTCTAGCCAGCACTACAGCTAAGGCTATGGAGCAAATGGCTGGATCGAGTGAGCAAGCAGCAGAGGC
CATGGATATTGCTAGTCAGGCCAGGCAAATGGTGCAGGCGATGAGAACCATTGGGACTCATCCTAGCTCC
AGTGCTGGTCTAAAAGATGATCTTCTTGAAAATTTGCAGGCCTATCAGAAACGAATGGGGGTGCAGATGC
AACGATTCAAGTGATCCTCTCGTCATTGCAGCAAATATCATTGAGATCTTGCACTTGATATTGTGGATTC
TTGATCGTCTTTTTTTCAAATGCATTTATCGTCGCTTTAAATACGGTTTGAAAAGAGGGCCTTCTACGGA
AGGAGTGCCAGAGTCTATGAGGGAAGAATATCGAAAGGAACAGCAGAATGCTGTGGATGTTGACGATGGT
CATTTTGTCAACATAGAGCTGGAGTAAAAAACTACCTTGTTTCTACT
>NS
AGCAAAAGCAGGGTGACAAAGACATAATGGATCCAAACACTGTGTCAAGCTTTCAGGTAGATTGCTTTCT
TTGGCATGTCCGCAAAAGAGTTGCAGACCAAGAACTAGGTGATGCCCCATTCCTTGATCGGCTTCGCCGA
GATCAGAAGTCCCTAAGAGGAAGAGGCAGCACTCTCGGTCTGGACATCGAAACAGCCACCCGTGCTGGAA

AGCAAATAGTGGAGCGGATTCTGAAGGAAGAATCTGATGAGGCACTCAAAATGACCATGGCCTCTGTACC
TGCATCGCGCTACCTAACTGACATGACTCTTGAGGAAATGTCAAGGCACTGGTTCATGCTCATGCCCAAG
CAGAAAGTGGCAGGCCCTCTTTGTATCAGAATGGACCAGGCGATCATGGATAAGAACATCATACTGAAAG
CGAACTTCAGTGTGATTTTTGACCGGCTGGAGACTCTAATATTACTAAGGGCCTTCACCGAAGAGGGGAC
AATTGTTGGCGAAATTTCACCACTGCCCTCTCTTCCAGGACATACTGATGAGGATGTCAAAAATGCAGTT
GGGGTCCTCATCGGAGGACTTGAATGGAATAATAACACAGTTCGAGTCTCTGAAACTCTACAGAGATTCG
CTTGGAGAAGCAGTAATGAGAATGGGAGACCTCCACTCACTCCAAAACAGAAACGGAAAATGGCGGGAAC
AATTAGGTCAGAAGTTTGAAGAAATAAGATGGTTGATTGAAGAAGTGAGACACAGACTGAAGATAACAGA
GAATAGTTTTGAGCAAATAACATTTATGCAAGCCTTACAACTATTGCTTGAAGTGGAGCAAGAGATAAGA
ACTTTCTCGTTTCAGCTTATTTAATAATAAAAAACACCCTTGTTTCTACT

References

Ablasser, A., Poeck, H., Anz, D., Berger, M., Schlee, M., Kim, S., … Hornung, V. (2009). Selection of Molecular Structure and Delivery of RNA Oligonucleotides to Activate TLR7 versus TLR8 and to Induce High Amounts of IL-12p70 in Primary Human Monocytes. *The Journal of Immunology*. http://doi.org/10.4049/jimmunol.0803001

Addepalli, B., & Hunt, A. G. (2007). A novel endonuclease activity associated with the Arabidopsis ortholog of the 30-kDa subunit of cleavage and polyadenylation specificity factor. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkm457

Adriaens, C., & Marine, J. C. (2017). NEAT1-containing paraspeckles: Central hubs in stress response and tumor formation. *Cell Cycle*. http://doi.org/10.1080/15384101.2016.1235847

Air, G. M., & Laver, W. G. (1989). The neuraminidase of influenza virus. *Proteins: Structure, Function, and Bioinformatics*. http://doi.org/10.1002/prot.340060402

Akira, S., Uematsu, S., & Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell*. http://doi.org/10.1016/j.cell.2006.02.015

Akman, B. H., Can, T., & Elif Erson-Bensan, A. (2012). Estrogen-induced upregulation and 3′-UTR shortening of CDC6. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gks855

Ali, A., Avalos, R. T., Ponimaskin, E., & Nayak, D. P. (2000). Influenza Virus Assembly: Effect of Influenza Virus Glycoproteins on the Membrane Association of M1 Protein. *Journal of Virology*. http://doi.org/10.1128/JVI.74.18.8709-8719.2000

Alkan, S. A., Martincic, K., & Milcarek, C. (2006). The hnRNPs F and H2 bind to similar sequences to influence gene expression. *Biochemical Journal*. http://doi.org/10.1042/BJ20050538

Amorim, M. J., Bruce, E. A., Read, E. K. C., Foeglein, A., Mahen, R., Stuart, A. D., & Digard, P. (2011). A Rab11- and Microtubule-Dependent Mechanism for Cytoplasmic Transport of Influenza A Virus Viral RNA. *Journal of Virology*. http://doi.org/10.1128/JVI.02606-10

Amorim, M. J., Read, E. K., Dalton, R. M., Medcalf, L., & Digard, P. (2007). Nuclear export of influenza A virus mRNAs requires ongoing RNA polymerase II activity. *Traffic*. http://doi.org/10.1111/j.1600-0854.2006.00507.x

Arranz, R., Coloma, R., Chichón, F. J., Conesa, J. J., Carrascosa, J. L., Valpuesta, J. M., … Martín-Benito, J. (2012). The structure of native influenza virion ribonucleoproteins. *Science*. http://doi.org/10.1126/science.1228172

Bagga, P. S., Arhin, G. K., & Wilusz, J. (1998). DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation in vitro. *Nucleic Acids Research*. http://doi.org/10.1093/nar/26.23.5343

Bai, C., & Tolias, P. P. (1998). Drosophila clipper/CPSF 30K is a post-transcriptionally regulated nuclear protein that binds RNA containing GC clusters. *Nucleic Acids Research*, *26*(7), 1597–1604. Retrieved from http://dx.doi.org/10.1093/nar/26.7.1597

Barabino, S. M., Hübner, W., Jenny, A., Minvielle-Sebastia, L., & Keller, W. (1997). The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes & Development* , *11*(13), 1703–1716. Retrieved from http://genesdev.cshlp.org/content/11/13/1703.abstract

Barrett, T., Wolstenholme, A. J., & Mahy, B. W. J. (1979). Transcription and replication of influenza virus RNA. *Virology*, *98*(1), 211–225. http://doi.org/https://doi.org/10.1016/0042-6822(79)90539-7

Bartel, D. P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*. http://doi.org/10.1016/S0092-8674(04)00045-5

Batra, R., Charizanis, K., Manchanda, M., Mohan, A., Li, M., Finn, D. J., … Swanson, M. S. (2014). Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Molecular Cell*. http://doi.org/10.1016/j.molcel.2014.08.027

Baudin, F., Petit, I., Weissenhorn, W., & Ruigrok, R. W. H. (2001). In vitro dissection of the membrane and RNP binding activities of influenza virus M1 protein. *Virology*. http://doi.org/10.1006/viro.2000.0804

Bauer, D. L. V., Tellier, M., Martínez-Alonso, M., Nojima, T., Proudfoot, N. J., Murphy, S., & Fodor, E. (2018). Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. *Cell Reports*. http://doi.org/10.1016/j.celrep.2018.04.047

Bava, F. A., Eliscovich, C., Ferreira, P. G., Miñana, B., Ben-Dov, C., Guigó, R., … Méndez, R. (2013). CPEB1 coordinates alternative 3′-UTR formation with translational regulation. *Nature*. http://doi.org/10.1038/nature11901

Bavagnoli, L., Cucuzza, S., Campanini, G., Rovida, F., Paolucci, S., Baldanti, F., & Maga, G. (2015). The novel influenza A virus protein PA-X and its naturally deleted variant show different enzymatic properties in comparison to the viral endonuclease PA. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkv926

Beaton, A. R., & Krug, R. M. (1981). Selected host cell capped RNA fragments prime influenza viral RNA transcription in vivo. *Nucleic Acids Research*. http://doi.org/10.1093/nar/9.17.4423

Beaudoing, E., & Gautheret, D. (2001). Identification of alternate polyadenylation sites and analysis of their tissue distribution using EST Data. *Genome Research*. http://doi.org/10.1101/gr.190501

Beaulieu, Y. B., Kleinman, C. L., Landry-Voyer, A.-M., Majewski, J., & Bachand, F. (2012). Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. *PLoS Genetics*. http://doi.org/10.1371/journal.pgen.1003078

Beelman, C. A., & Parker, R. (1995). Degradation of mRNA in eukaryotes. *Cell*. http://doi.org/10.1016/0092-8674(95)90326-7

Bennett, C. L., Brunkow, M. E., Ramsdell, F., O'Briant, K. C., Zhu, Q., Fuleihan, R. L., … Chance, P. F. (2001). A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA→AAUGAA) leads to the IPEX syndrome. *Immunogenetics*, *53*(6), 435–439. http://doi.org/10.1007/s002510100358

Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics*. http://doi.org/10.1038/nrg3662

Bercovich-Kinori, A., Tai, J., Gelbart, I. A., Shitrit, A., Ben-Moshe, S., Drori, Y., … Stern-Ginossar, N. (2016). A systematic view on influenza induced host shutoff. *ELife*. http://doi.org/10.7554/eLife.18311

Bergmann, M., Garcia-Sastre, A., Carnero, E., Pehamberger, H., Wolff, K., Palese, P., & Muster, T. (2000). Influenza virus NS1 protein counteracts PKR-mediated inhibition of replication. *Journal of Virology*. http://doi.org/10.1128/JVI.74.13.6203-6206.2000

Berkovits, B. D., & Mayr, C. (2015). Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature*. http://doi.org/10.1038/nature14321

Bier, K., York, A., & Fodor, E. (2011). Cellular cap-binding proteins associate with influenza virus mRNAs. *Journal of General Virology*. http://doi.org/10.1099/vir.0.029231-0

Bortz, E., Westera, L., Maamary, J., Steel, J., Albrecht, R. A., Manicassamy, B., … García-Sastrea, A. (2011). Host- and strain-specific regulation of influenza virus polymerase activity by interacting cellular proteins. *MBio*. http://doi.org/10.1128/mBio.00151-11

Boulan, E. R., & Sabatini, D. D. (1978). Asymmetric budding of viruses in epithelial monlayers: a model system for study of epithelial polarity. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.75.10.5071

Boulay, F., Doms, R. W., Wilson, I., & Helenius, A. (1987). The influenza hemagglutinin precursor as an acid-sensitive probe of the biosynthetic pathway. *EMBO J*.

Brais, B., Bouchard, J. P., Xie, Y. G., Rochefort, D. L., Chrétien, N., Tomé, F. M. S., … Rouleau, G. A. (1998). Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nature Genetics*. http://doi.org/10.1038/ng0298-164

Brooke, C. B., Ince, W. L., Wrammert, J., Ahmed, R., Wilson, P. C., Bennink, J. R., & Yewdell, J. W. (2013). Most Influenza A Virions Fail To Express at Least One Essential Viral Protein. *Journal of Virology*. http://doi.org/10.1128/JVI.02284-12

Bruce, E. A., Digard, P., & Stuart, A. D. (2010). The Rab11 Pathway Is Required for Influenza A Virus Budding and Filament Formation. *Journal of Virology*. http://doi.org/10.1128/JVI.00307-10

Brumbaugh, J., Di Stefano, B., Wang, X., Borkent, M., Forouzmand, E., Clowers, K. J., … Hochedlinger, K. (2018). Erratum: Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling (Cell (2018) (172)(1-2)(106–120.e21)(S0092867417313703)(10.1016/j.cell.2017.11.023)). *Cell*. http://doi.org/10.1016/j.cell.2017.12.035

Brunotte, L., Flies, J., Bolte, H., Reuther, P., Vreede, F., & Schwemmle, M. (2014). The nuclear export protein of H5N1 influenza A viruses recruits Matrix 1 (M1) protein to the viral ribonucleoprotein to mediate nuclear export. *Journal of Biological Chemistry*. http://doi.org/10.1074/jbc.M114.569178

Bui, M., Wills, E. G., Helenius, A., & Whittaker, G. R. (2000). Role of the influenza virus M1 protein in nuclear export of viral ribonucleoproteins. *Journal of Virology*. http://doi.org/10.1128/JVI.74.4.1781-1786.2000

Bullough, P. A., Hughson, F. M., Skehel, J. J., & Wiley, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature*. http://doi.org/10.1038/371037a0

Calin, G. A., & Croce, C. M. (2006). MicroRNA signatures in human cancers. *Nature Reviews Cancer*. http://doi.org/10.1038/nrc1997

Cao, X. (2016). Self-regulation and cross-regulation of pattern-recognition receptor signalling in health and disease. *Nature Reviews Immunology*. http://doi.org/10.1038/nri.2015.8

Carthagena, L., Bergamaschi, A., Luna, J. M., David, A., Uchil, P. D., Margottin-Goguet, F., … Nisole, S. (2009). Human TRIM Gene Expression in Response to Interferons. *PLoS ONE*, *4*(3), e4894. http://doi.org/10.1371/journal.pone.0004894

Chakrabarti, M., & Hunt, A. G. (2015). CPSF30 at the interface of alternative polyadenylation and cellular signaling in plants. *Biomolecules*. http://doi.org/10.3390/biom5021151

Chan, A. Y., Vreede, F. T., Smith, M., Engelhardt, O. G., & Fodor, E. (2006). Influenza virus inhibits RNA polymerase II elongation. *Virology*. http://doi.org/10.1016/j.virol.2006.03.005

Chao, L. C., Jamil, A., Kim, S. J., Huang, L., & Martinson, H. G. (1999). Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. *Molecular and Cellular Biology*. http://doi.org/10.1128/MCB.19.8.5588

Chen, B. J., Leser, G. P., Jackson, D., & Lamb, R. A. (2008). The Influenza Virus M2 Protein Cytoplasmic Tail Interacts with the M1 Protein and Influences Virus Assembly at the Site of Virus Budding. *Journal of Virology*. http://doi.org/10.1128/JVI.01184-08

Chen, B. J., Leser, G. P., Morita, E., & Lamb, R. a. (2007). Influenza virus hemagglutinin and neuraminidase, but not the matrix protein, are required for assembly and budding of plasmid-derived virus-like particles. *Journal of Virology*. http://doi.org/10.1128/JVI.00361-07

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., … Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. http://doi.org/10.1186/1471-2105-14-128

Chen, J. M., Férec, C., & Cooper, D. N. (2006). A systematic analysis of disease-associated variants in the 3′ regulatory regions of human protein-coding genes I: General principles and overview. *Human Genetics*. http://doi.org/10.1007/s00439-006-0180-7

Chen, W., Guo, W., Li, M., Shi, D., Tian, Y., Li, Z., … Deng, W. (2013). Upregulation of cleavage and polyadenylation specific factor 4 in lung adenocarcinoma and its critical role for cancer cell survival and proliferation. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0082728

Chen, W., Calvo, P. A. A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., … Yewdell, J. W. W. (2001). A novel influenza A virus mitochondrial protein that induces cell death. *Nature Medicine*. http://doi.org/10.1038/nm1201-1306

Chen, X., Liu, S., Goraya, M. U., Maarouf, M., Huang, S., & Chen, J.-L. (2018). Host Immune Response to Influenza A Virus Infection. *Frontiers in Immunology*, *9*, 320. http://doi.org/10.3389/fimmu.2018.00320

Chen, Y., Azman, S. N., Kerishnan, J. P., Zain, R. B., Chen, Y. N., Wong, Y. L., & Gopinath, S. C. B. (2014). Identification of host-immune response protein candidates in the sera of human oral squamous cell carcinoma patients. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0109012

Chen, Z., Li, Y., & Krug, R. M. (1999). Influenza A virus NS1 protein targets poly(A)-binding protein II of the cellular 3'-end processing machinery. *The EMBO Journal*. http://doi.org/10.1093/emboj/18.8.2273

Cheng, W., Chen, G., Jia, H., He, X., & Jing, Z. (2018). DDX5 RNA helicases: Emerging roles in viral infection. *International Journal of Molecular Sciences*. http://doi.org/10.3390/ijms19041122

Cianci, C., Tiley, L., & Krystal, M. (1995). Differential activation of the influenza virus polymerase via template RNA binding. *Journal of Virology*.

Clark, A. M., Nogales, A., Martinez-Sobrido, L., Topham, D. J., & DeDiego, M. L. (2017). Functional Evolution of Influenza Virus NS1 Protein in Currently Circulating Human 2009 Pandemic H1N1 Viruses. *Journal of Virology*, *91*(17). Retrieved from http://jvi.asm.org/content/91/17/e00721-17.abstract

Curinha, A., Braz, S. O., Pereira-Castro, I., Cruz, A., & Moreira, A. (2014). Implications of polyadenylation in health and disease. *Nucleus*. http://doi.org/10.4161/nucl.36360

Danckwardt, S., Hentze, M. W., & Kulozik, A. E. (2008). 3′ end mRNA processing: molecular mechanisms and implications for health and disease. *The EMBO Journal*. http://doi.org/10.1038/sj.emboj.7601932

Dardenne, E., PolayEspinoza, M., Fattet, L., Germann, S., Lambert, M. P., Neil, H., … Auboeuf, D. (2014). RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. *Cell Reports*. http://doi.org/10.1016/j.celrep.2014.05.010

Das, K., Ma, L.-C., Xiao, R., Radvansky, B., Aramini, J., Zhao, L., … Montelione, G. T. (2008). Structural basis for suppression of a host antiviral response by influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America*. http://doi.org/10.1073/pnas.0805213105

Dass, B., McDaniel, L., Schultz, R. A., Attaya, E., & MacDonald, C. C. (2002). The gene CSTF2T, encoding the human variant CSTF-64 polyadenylation protein τCstF-64, lacks introns and may be associated with male sterility. *Genomics*. http://doi.org/10.1016/S0888-7543(02)96862-X

De Klerk, E., Venema, A., Anvar, S. Y., Goeman, J. J., Hu, O. H., Trollet, C., … 'T Hoen, P. A. C. (2012). Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gks655

DeDiego, M. L., Nogales, A., Lambert-Emo, K., Martinez-Sobrido, L., & Topham, D. J. (2016). NS1 Protein Mutation I64T Affects Interferon Responses and Virulence of Circulating H3N2 Human Influenza A Viruses. *Journal of Virology*. http://doi.org/10.1128/JVI.01039-16

Delaney, K. J., Xu, R., Zhang, J., Li, Q. Q., Yun, K.-Y., Falcone, D. L., & Hunt, A. G. (2006). Calmodulin interacts with and regulates the RNA-binding activity of an Arabidopsis polyadenylation factor subunit. *Plant Physiology*. http://doi.org/10.1104/pp.105.070672

Delisle, J.-S., Giroux, M., Boucher, G., Landry, J.-R., Hardy, M.-P., Lemieux, S., … Perreault, C. (2013). The TGF-β-Smad3 pathway inhibits CD28-dependent cell growth and proliferation of CD4 T cells. *Genes and Immunity*. http://doi.org/10.1038/gene.2013.4

Deng, T., Vreede, F. T., & Brownlee, G. G. (2006). Different de novo initiation strategies are used by influenza virus RNA polymerase on its cRNA and viral RNA promoters during viral RNA replication. *Journal of Virology*. http://doi.org/10.1128/JVI.80.5.2337-2348.2006

Derti, A., Garrett-Engele, P., MacIsaac, K. D., Stevens, R. C., Sriram, S., Chen, R., … Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. *Genome Research*. http://doi.org/10.1101/gr.132563.111

Di Giammartino, D. C., Nishida, K., & Manley, J. L. (2011). Mechanisms and Consequences of Alternative Polyadenylation. *Molecular Cell*. http://doi.org/10.1016/j.molcel.2011.08.017

Dias, A., Bouvier, D., Crépin, T., McCarthy, A. A., Hart, D. J., Baudin, F., … Ruigrok, R. W. H. (2009). The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*. http://doi.org/10.1038/nature07745

Diribarne, G., & Bensaude, O. (2009). 7SK RNA, a non-coding RNA regulating P-TEFb, a general transcription factor. *RNA Biology*. http://doi.org/10.4161/rna.6.2.8115

Dubois, J., Terrier, O., & Rosa-Calatrava, M. (2014). Influenza viruses and mRNA splicing: Doing more with less. *MBio*. http://doi.org/10.1128/mBio.00070-14

Elton, D., Simpson-Holley, M., Archer, K., Medcalf, L., Hallam, R., McCauley, J., & Digard, P. (2001). Interaction of the Influenza Virus Nucleoprotein with the Cellular CRM1-Mediated Nuclear Export Pathway. *Journal of Virology*. http://doi.org/10.1128/JVI.75.1.408-419.2001

Enami, M., Fukuda, R., & Ishihama, A. (1985). Transcription and replication of eight RNA segments of influenza virus. *Virology*, *142*(1), 68–77. http://doi.org/https://doi.org/10.1016/0042-6822(85)90423-4

Engelhardt, O. G., & Fodor, E. (2006). Functional association between viral and cellular transcription during influenza virus infection. *Reviews in Medical Virology*. http://doi.org/10.1002/rmv.512

Eulalio, A., Tritschler, F., & Izaurralde, E. (2009). The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. *RNA (New York, N.Y.)*. http://doi.org/10.1261/rna.1703809

Fields, S., & Winter, G. (1982). Nucleotide sequences of influenza virus segments 1 and 3 reveal mosaic structure of a small viral RNA segment. *Cell*. http://doi.org/10.1016/0092-8674(82)90348-8

Fields, S., Winter, G., & Brownlee, G. G. (1981). Structure of the neuraminidase gene in human influenza virus A/PR/8/34. *Nature*. http://doi.org/10.1038/290213a0

Firth. (2012). Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC_UUU_CGU and is in the +1 direction. *Open Biology*. http://doi.org/10.1098/rsob.120109

Fitzgerald, M., & Shenk, T. (1981). The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell*. http://doi.org/10.1016/0092-8674(81)90521-3

Flavell, S. W., Kim, T. K., Gray, J. M., Harmin, D. A., Hemberg, M., Hong, E. J., … Greenberg, M. E. (2008). Genome-Wide Analysis of MEF2 Transcriptional Program Reveals Synaptic Target Genes and Neuronal Activity-Dependent Polyadenylation Site Selection. *Neuron*. http://doi.org/10.1016/j.neuron.2008.11.029

Fodor, E., Crow, M., Mingay, L. J., Deng, T., Sharps, J., Fechter, P., & Brownlee, G. G. (2002). A Single Amino Acid Mutation in the PA Subunit of the Influenza Virus RNA

Polymerase Inhibits Endonucleolytic Cleavage of Capped RNAs. *Journal of Virology*. http://doi.org/10.1128/JVI.76.18.8989-9001.2002

Fodor, E., Pritlove, D. C., & Brownlee, G. G. (1994). The Influenza-Virus Panhandle Is Involved in the Initiation of Transcription. *J Virol*.

Fodor, E., & Smith, M. (2004). The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex. *Journal of Virology*. http://doi.org/10.1128/JVI.78.17.9144-9153.2004

Forbes, N. E., Ping, J., Dankar, S. K., Jia, J. J., Selman, M., Keleta, L., … Brown, E. G. (2012). Multifunctional adaptive NS1 mutations are selected upon human influenza virus evolution in the mouse. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0031839

Forbes, N., Selman, M., Pelchat, M., Jia, J. J., Stintzi, A., & Brown, E. G. (2013). Identification of adaptive mutations in the influenza a virus non-structural 1 gene that increase cytoplasmic localization and differentially regulate host gene expression. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0084673

Fournier, G., Chiang, C., Munier, S., Tomoiu, A., Demeret, C., Vidalain, P. O., … Naffakh, N. (2014). Recruitment of RED-SMU1 Complex by Influenza A Virus RNA Polymerase to Control Ciral mRNA Splicing. *PLoS Pathogens*. http://doi.org/10.1371/journal.ppat.1004164

Friedman, R. C., Farh, K. K. H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*. http://doi.org/10.1101/gr.082701.108

Fu, Y., Sun, Y., Li, Y., Li, J., Rao, X., Chen, C., & Xu, A. (2011). Differential genome-wide profiling of tandem 3'UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Research*. http://doi.org/10.1101/gr.115295.110.21

Garaigorta, U., & Ortín, J. (2007). Mutation analysis of a recombinant NS replicon shows that influenza virus NS1 protein blocks the splicing and nucleo-cytoplasmic transport of its own viral mRNA. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkm230

Geisberg, J. V., Moqtaderi, Z., Fan, X., Ozsolak, F., & Struhl, K. (2014). Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell*. http://doi.org/10.1016/j.cell.2013.12.026

Gerber, M., Isel, C., Moules, V., & Marquet, R. (2014). Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends in Microbiology*. http://doi.org/10.1016/j.tim.2014.04.001

Giannakakis, A., Zhang, J., Jenjaroenpun, P., Nama, S., Zainolabidin, N., Aau, M. Y., … Guccione, E. (2015). Contrasting expression patterns of coding and noncoding parts of the human genome upon oxidative stress. *Scientific Reports*. http://doi.org/10.1038/srep09737

Goubau, D., Schlee, M., Deddouche, S., Pruijssers, A. J., Zillinger, T., Goldeck, M., … Reis E Sousa, C. (2014). Antiviral immunity via RIG-I-mediated recognition of RNA bearing 59-diphosphates. *Nature*. http://doi.org/10.1038/nature13590

Graber, J. H., Nazeer, F. I., Yeh, P. C., Kuehner, J. N., Borikar, S., Hoskinson, D., & Moore, C. L. (2013). DNA damage induces targeted, genome-wide variation of poly(A) sites in budding yeast. *Genome Research*. http://doi.org/10.1101/gr.144964.112

Graef, K. M., Vreede, F. T., Lau, Y.-F., McCall, A. W., Carr, S. M., Subbarao, K., & Fodor, E. (2010). The PB2 Subunit of the Influenza Virus RNA Polymerase Affects Virulence by Interacting with the Mitochondrial Antiviral Signaling Protein and Inhibiting Expression of Beta Interferon. *Journal of Virology*. http://doi.org/10.1128/JVI.00879-10

Gruber, A. J., Schmidt, R., Gruber, A. R., Martin, G., Ghosh, S., Belmadani, M., … Zavolan, M. (2016). A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Research*. http://doi.org/10.1101/gr.202432.115

Gruber, A. R., Martin, G., Müller, P., Schmidt, A., Gruber, A. J., Gumienny, R., … Zavolan, M. (2014). Global 3′ UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nature Communications*. http://doi.org/10.1038/ncomms6465

Grzybowska, E. A. (2012). Human intronless genes: Functional groups, associated diseases, evolution, and mRNA processing in absence of splicing. *Biochemical and Biophysical Research Communications*. http://doi.org/10.1016/j.bbrc.2012.06.092

Gu, W., Gallagher, G. R., Dai, W., Liu, P., Li, R., Trombly, M. I., … Finberg, R. W. (2015). Influenza A virus preferentially snatches noncoding RNA caps. *RNA*. http://doi.org/10.1261/rna.054221.115

Guhaniyogi, J., & Brewer, G. (2001). Regulation of mRNA stability in mammalian cells. *Gene*. http://doi.org/S037811190100350X [pii]

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., … Cusack, S. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Structural and Molecular Biology*. http://doi.org/10.1038/nsmb.1421

Guo, H., Ingolia, N. T., Weissman, J. S., & Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*. http://doi.org/10.1038/nature09267

Gupta, I., Clauder-Münster, S., Klaus, B., Järvelin, A. I., Aiyar, R. S., Benes, V., … Steinmetz, L. M. (2014). Alternative polyadenylation diversifies post-transcriptional regulation by selective RNA-protein interactions. *Molecular Systems Biology*, *10*, 719. http://doi.org/10.1002/msb.135068

Hale, B. G., Steel, J., Medina, R. A., Manicassamy, B., Ye, J., Hickman, D., … García-Sastre, A. (2010). Inefficient control of host gene expression by the 2009 pandemic H1N1 influenza A virus NS1 protein. *Journal of Virology*. http://doi.org/10.1128/JVI.00081-10

Hale, B. G., Albrecht, R. A., & García-Sastre, A. (2010). Innate immune evasion strategies of influenza viruses. *Future Microbiology*. http://doi.org/10.2217/fmb.09.108

Hale, B. G., Randall, R. E., Ortin, J., & Jackson, D. (2008). The multifunctional NS1 protein of influenza A viruses. *Journal of General Virology*. http://doi.org/10.1099/vir.0.2008/004606-0

Han, T. W., Kato, M., Xie, S., Wu, L. C., Mirzaei, H., Pei, J., … McKnight, S. L. (2012). Cell-free formation of RNA granules: Bound RNAs identify features and components of cellular assemblies. *Cell*. http://doi.org/10.1016/j.cell.2012.04.016

Hao, L., Sakurai, A., Watanabe, T., Sorensen, E., Nidom, C. A., Newton, M. A., … Kawaoka, Y. (2008). Drosophila RNAi screen identifies host genes important for influenza virus replication. *Nature*. http://doi.org/10.1038/nature07151

Harris, A., Cardone, G., Winkler, D. C., Heymann, J. B., Brecher, M., White, J. M., & Steven, A. C. (2006). Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.0607614103

Hartmann, B. M., Albrecht, R. A., Zaslavsky, E., Nudelman, G., Pincas, H., Marjanovic, N., … Sealfon, S. C. (2017). Pandemic H1N1 influenza A viruses suppress immunogenic RIPK3-driven dendritic cell death. *Nature Communications*. http://doi.org/10.1038/s41467-017-02035-9

Hatada, E., Hasegawa, M., Mukaigawa, J., Shimizu, K., & Fukuda, R. (1989). Control of influenza virus gene expression: quantitative analysis of each viral RNA species in infected cells. *Journal of Biochemistry*.

Heikel, G., Choudhury, N. R., & Michlewski, G. (2016). The role of Trim25 in development, disease and RNA metabolism. *Biochemical Society Transactions*. http://doi.org/10.1042/BST20160077

Hendrickson, D. G., Hogan, D. J., McCullough, H. L., Myers, J. W., Herschlag, D., Ferrell, J. E., & Brown, P. O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biology*. http://doi.org/10.1371/journal.pbio.1000238

Henras, A. K., Plisson-Chastang, C., O'Donohue, M. F., Chakraborty, A., & Gleizes, P. E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdisciplinary Reviews: RNA*. http://doi.org/10.1002/wrna.1269

Higgs, D. R., Goodbourn, S. E. Y., Lamb, J., Clegg, J. B., Weatherall, D. J., & Proudfoot, N. J. (1983). α-Thalassaemia caused by a polyadenylation signal mutation. *Nature*. http://doi.org/10.1038/306398a0

Ho, A. W. S., Prabhu, N., Betts, R. J., Ge, M. Q., Dai, X., Hutchinson, P. E., … Kemeny, D. M. (2011). Lung CD103+ Dendritic Cells Efficiently Transport Influenza Virus to the Lymph Node and Load Viral Antigen onto MHC Class I for Presentation to CD8 T Cells. *The Journal of Immunology*. http://doi.org/10.4049/jimmunol.1100987

Hofmann, P., Sprenger, H., Kaufmann, A., Bender, A., Hasse, C., Nain, M., & Gemsa, D. (1997). Susceptibility of mononuclear phagocytes to influenza A virus infection and possible role in the antiviral response. *J Leukoc Biol*.

Holm, C. K., Rahbek, S. H., Gad, H. H., Bak, R. O., Jakobsen, M. R., Jiang, Z., … Paludan, S. R. (2016). Influenza A virus targets a cGAS-independent STING pathway that controls enveloped RNA viruses. *Nature Communications*. http://doi.org/10.1038/ncomms10680

Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., … Tian, B. (2013). Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nature Methods*. http://doi.org/10.1038/nmeth.2288

Hu, B., Li, X., Huo, Y., Yu, Y., Zhang, Q., Chen, G., … Zhou, J. (2016). Cellular responses to HSV-1 infection are linked to specific types of alterations in the host transcriptome. *Scientific Reports*. http://doi.org/10.1038/srep28075

Hu, W., Liu, Y., & Yan, J. (2014). Microarray meta-analysis of RNA-binding protein functions in alternative polyadenylation. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0090774

Hu, W., Li, S., Park, J. Y., Boppana, S., Ni, T., Li, M., … Xiang, M. (2017). Dynamic landscape of alternative polyadenylation during retinal development. *Cellular and Molecular Life Sciences*, *74*(9), 1721–1739. http://doi.org/10.1007/s00018-016-2429-1

Huang, C., Shi, J., Guo, Y., Huang, W., Huang, S., Ming, S., … Yao, C. (2017). A snoRNA modulates mRNA 3′ end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkx651

Huang, H., Chen, J., Liu, H., & Sun, X. (2013). The nucleosome regulates the usage of polyadenylation sites in the human genome. *BMC Genomics*. http://doi.org/10.1186/1471-2164-14-912

Huang, I.-C., Li, W., Sui, J., Marasco, W., Choe, H., & Farzan, M. (2008). Influenza A Virus Neuraminidase Limits Viral Superinfection. *Journal of Virology*. http://doi.org/10.1128/JVI.00079-08

Huang, J., Weng, T., Ko, J., Chen, N. yuan, Xiang, Y., Volcik, K., … Lu, X. (2018). Suppression of cleavage factor Im 25 promotes the proliferation of lung cancer cells through alternative polyadenylation. *Biochemical and Biophysical Research Communications*. http://doi.org/10.1016/j.bbrc.2018.06.087

Huang, T. S., Palese, P., & Krystal, M. (1990). Determination of influenza virus proteins required for genome replication. *Journal of Virology*.

Huang, X., Liu, T., Muller, J., Levandowski, R. A., & Ye, Z. (2001). Effect of influenza virus matrix protein and viral RNA on ribonucleoprotein formation and nuclear export. *Virology*. http://doi.org/10.1006/viro.2001.1067

Hutchinson, E. C., Charles, P. D., Hester, S. S., Thomas, B., Trudgian, D., Martï¿½nez-Alonso, M., & Fodor, E. (2015). Conserved and host-specific features of influenza virion architecture. *Nature Communications*, *5*. http://doi.org/10.1038/ncomms5816

Hutchinson, E. C., & Fodor, E. (2013). Transport of the influenza virus genome from nucleus to nucleus. *Viruses*. http://doi.org/10.3390/v5102424

Hutchinson, E. C., von Kirchbach, J. C., Gog, J. R., & Digard, P. (2010). Genome packaging in influenza A virus. *Journal of General Virology*. http://doi.org/10.1099/vir.0.017608-0

Hutchinson, E., & Fodor, E. (2014). Purification of influenza virions by haemadsorption and ultracentrifugation. Retrieved from http://dx.doi.org/10.1038/protex.2014.027

Jagger, B. W., Wise, H. M., Kash, J. C., Walters, K. A., Wills, N. M., Xiao, Y. L., … Digard, P. (2012). An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science*. http://doi.org/10.1126/science.1222213

Jalal, C., Uhlmann-Schiffler, H., & Stahl, H. (2007). Redundant role of DEAD box proteins p68 (Ddx5) and p72/p82 (Ddx17) in ribosome biogenesis and cell proliferation. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkm058

Jenal, M., Elkon, R., Loayza-Puch, F., Van Haaften, G., Kühn, U., Menzies, F. M., … Agami, R. (2012). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell*. http://doi.org/10.1016/j.cell.2012.03.022

Ji, Z., Lee, J. Y., Pan, Z., Jiang, B., & Tian, B. (2009). Progressive lengthening of 3’ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.0900028106

Ji, Z., Luo, W., Li, W., Hoque, M., Pan, Z., Zhao, Y., & Tian, B. (2011). Transcriptional activity regulates alternative cleavage and polyadenylation. *Molecular Systems Biology*. http://doi.org/10.1038/msb.2011.69

Ji, Z., & Tian, B. (2009). Reprogramming of 3′ Untranslated Regions of mRNAs by Alternative Polyadenylation in Generation of Pluripotent Stem Cells from Different Cell Types. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0008419

Jia, X., Yuan, S., Wang, Y., Fu, Y., Ge, Y., Ge, Y., … Xu, A. (2017). The role of alternative polyadenylation in the antiviral innate immune response. *Nature Communications*. http://doi.org/10.1038/ncomms14605

Jonkers, I., & Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*. http://doi.org/10.1038/nrm3953

Jost, J. P., Schwarz, S., Hess, D., Angliker, H., Fuller-Pace, F. V., Stahl, H., … Siegmann, M. (1999). A chicken embryo protein related to the mammalian DEAD box protein p68 is tightly associated with the highly purified protein-RNA complex of 5-MeC-DNA glycosylase. *Nucleic Acids Research*. http://doi.org/10.1093/nar/27.16.3245

Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., & Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*. http://doi.org/10.1038/nature09479

Kainov, D. E., Müller, K. H., Theisen, L. L., Anastasina, M., Kaloinen, M., & Muller, C. P. (2011). Differential Effects of NS1 Proteins of Human Pandemic H1N1/2009, Avian Highly Pathogenic H5N1, and Low Pathogenic H5N2 Influenza A Viruses on Cellular Pre-mRNA Polyadenylation and mRNA Translation. *Journal of Biological Chemistry* , *286*(9), 7239–7247. http://doi.org/10.1074/jbc.M110.203489

Kash, J. C., Tumpey, T. M., Proll, S. C., Carter, V., Perwitasari, O., Thomas, M. J., … Katze, M. G. (2006). Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. *Nature*. http://doi.org/10.1038/nature05181

Katz, Y., Wang, E. T., Airoldi, E. M., & Burge, C. B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*. http://doi.org/10.1038/nmeth.1528

Kawai, T., & Akira, S. (2010). The role of pattern-recognition receptors in innate immunity: Update on toll-like receptors. *Nature Immunology*. http://doi.org/10.1038/ni.1863

Kerwitz, Y., Kühn, U., Lilie, H., Knoth, A., Scheuermann, T., Friedrich, H., … Wahle, E. (2003). Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. *EMBO Journal*. http://doi.org/10.1093/emboj/cdg347

Khaperskyy, D. A., Emara, M. M., Johnston, B. P., Anderson, P., Hatchette, T. F., & McCormick, C. (2014). Influenza A Virus Host Shutoff Disables Antiviral Stress-Induced Translation Arrest. *PLoS Pathogens*. http://doi.org/10.1371/journal.ppat.1004217

Khaperskyy, D. A., Schmaling, S., Larkins-Ford, J., McCormick, C., & Gaglia, M. M. (2016). Selective Degradation of Host RNA Polymerase II Transcripts by Influenza A Virus PA-X Host Shutoff Protein. *PLoS Pathogens*. http://doi.org/10.1371/journal.ppat.1005427

Kim, H. S., Luo, L., Pflugfelder, S. C., & Li, D. Q. (2005). Doxycycline inhibits TGF-β1-induced MMP-9 via Smad and MAPK pathways in human corneal epithelial cells. *Investigative Ophthalmology and Visual Science*. http://doi.org/10.1167/iovs.04-0929

Kobbi, L., Demey-Thomas, E., Braye, F., Proux, F., Kolesnikova, O., Vinh, J., … Bensaude, O. (2016). An evolutionary conserved Hexim1 peptide binds to the Cdk9 catalytic site to inhibit P-TEFb. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.1612331113

Kochs, G., Garcia-Sastre, A., & Martinez-Sobrido, L. (2007). Multiple Anti-Interferon Actions of the Influenza A Virus NS1 Protein. *Journal of Virology*. http://doi.org/10.1128/JVI.02581-06

Koliopoulos, M. G., Lethier, M., Van Der Veen, A. G., Haubrich, K., Hennig, J., Kowalinski, E., … Rittinger, K. (2018). Molecular mechanism of influenza A NS1-mediated TRIM25 recognition and inhibition. *Nature Communications*. http://doi.org/10.1038/s41467-018-04214-8

Koppstein, D., Ashour, J., & Bartel, D. P. (2015). Sequencing the cap-snatching repertoire of H1N1 influenza provides insight into the mechanism of viral transcription initiation. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkv333

Krammer, F., Smith, G. J. D., Fouchier, R. A. M., Peiris, M., Kedzierska, K., Doherty, P. C., … García-Sastre, A. (2018). Influenza. *Nature Reviews Disease Primers*, *4*(1), 3. http://doi.org/10.1038/s41572-018-0002-y

Kraynik, S. M., Gabanic, A., Anthony, S. R., Kelley, M., Paulding, W. R., Roessler, A., … Tranter, M. (2015). The stress-induced heat shock protein 70.3 expression is regulated by a dual-component mechanism involving alternative polyadenylation and HuR. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. http://doi.org/10.1016/j.bbagrm.2015.02.004

Krug, R. M. (2015). Functions of the influenza A virus NS1 protein in antiviral defense. *Current Opinion in Virology*. http://doi.org/10.1016/j.coviro.2015.01.007

Krug, R. M., Broni, B. A., & Bouloy, M. (1979). Are the 5′ ends of influenza viral mrnas synthesized in vivo donated by host mRNAs? *Cell*. http://doi.org/10.1016/0092-8674(79)90052-7

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., … Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkw377

Kuo, R. L., & Krug, R. M. (2009). Influenza a virus polymerase is an integral component of the CPSF30-NS1A protein complex in infected cells. *J Virol*. http://doi.org/10.1128/JVI.01491-08

Kuo, R. L., Zhao, C., Malur, M., & Krug, R. M. (2010). Influenza A virus strains that circulate in humans differ in the ability of their NS1 proteins to block the activation of IRF3 and interferon-β transcription. *Virology*. http://doi.org/10.1016/j.virol.2010.09.012

Kuss, S. K., Mata, M. A., Zhang, L., & Fontoura, B. M. A. (2013). Nuclear imprisonment: Viral strategies to arrest host mRNA nuclear export. *Viruses*. http://doi.org/10.3390/v5071824

La Gruta, N. L., & Turner, S. J. (2014). T cell mediated immunity to influenza: Mechanisms of viral control. *Trends in Immunology*. http://doi.org/10.1016/j.it.2014.06.004

Lackford, B., Yao, C., Charles, G. M., Weng, L., Zheng, X., Choi, E. A., … Shi, Y. (2014). Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO Journal*. http://doi.org/10.1002/embj.201386537

Lafontaine, D. L. J. (2015). Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nature Structural & Molecular Biology*. http://doi.org/10.1038/nsmb.2939

Lamb, R. A., Lai, C. J., & Choppin, P. W. (1981). Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.78.7.4170

Lamb, R. A., & Horvath, C. M. (1991). Diversity of coding strategies in influenza viruses. *Trends in Genetics*. http://doi.org/10.1016/0168-9525(91)90326-L

Lamb, R. A., & Lai, C. J. (1980). Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping nonstructural proteins of influenza virus. *Cell*. http://doi.org/10.1016/0092-8674(80)90484-5

Lane, D. P., & Hoeffler, W. K. (1980). SV40 large T shares an antigenic determinant with a cellular protein of molecular weight 68,000. *Nature*. http://doi.org/10.1038/288167a0

Lee, N.-R., Kim, H.-I., Choi, M.-S., Yi, C.-M., & Inn, K.-S. (2015). Regulation of MDA5-MAVS Antiviral Signaling Axis by TRIM25 through TRAF6-Mediated NF-κB Activation. *Mol. Cells*. http://doi.org/10.14348/molcells.2015.0047

Lejeune, F., & Maquat, L. E. (2005). Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Current Opinion in Cell Biology*. http://doi.org/10.1016/j.ceb.2005.03.002

Lembo, A., Di Cunto, F., & Provero, P. (2012). Shortening of 3′UTRs correlates with poor prognosis in breast and lung cancer. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0031129

Levene, R. E., & Gaglia, M. M. (2018). Host Shutoff in Influenza A Virus: Many Means to an End. *Viruses*, *10*(9). http://doi.org/10.3390/v10090475

Lewis, J. D., & Izaurflde, E. (1997). The Role of the Cap Structure in RNA Processing and Nuclear Export. *European Journal of Biochemistry*. http://doi.org/10.1111/j.1432-1033.1997.00461.x

Li, M. L., Ramirez, B. C., & Krug, R. M. (1998). RNA-dependent activation of primer RNA production by influenza virus polymerase: Different regions of the same protein subunit constitute the two required RNA-binding sites. *EMBO Journal*. http://doi.org/10.1093/emboj/17.19.5844

Li, Q., Ye, F., Shi, Y., Zhang, L., Wang, W., Tu, Z., … Li, Y. (2006). Nuclear Translocation of SMAD3 May Enhance the TGF-β/SMADS Pathway in High Glucose Circumstances. *Transplantation Proceedings*. http://doi.org/10.1016/j.transproceed.2006.06.092

Li, W., Park, J. Y., Zheng, D., Hoque, M., Yehia, G., & Tian, B. (2016). Alternative cleavage and polyadenylation in spermatogenesis connects chromatin regulation with post-transcriptional control. *BMC Biology*. http://doi.org/10.1186/s12915-016-0229-6

Li, W., You, B., Hoque, M., Zheng, D., Luo, W., Ji, Z., … Tian, B. (2015). Systematic Profiling of Poly(A)+ Transcripts Modulated by Core 3' End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation. *PLoS Genetics*. http://doi.org/10.1371/journal.pgen.1005166

Li, X. Q., & Du, D. (2013). RNA polyadenylation sites on the genomes of microorganisms, animals, and plants. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0079511

LI, Y., CHEN, Z.-Y., WANG, W., BAKER, C. C., & KRUG, R. M. (2001). The 3′-end-processing factor CPSF is required for the splicing of single-intron pre-mRNAs in vivo. *RNA*, *7*(6), 920–931. http://doi.org/DOI: undefined

Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S., & Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes and Development*. http://doi.org/10.1101/gad.229328.113

Liaw, H. H., Lin, C. C., Juan, H. F., & Huang, H. C. (2013). Differential MicroRNA Regulation Correlates with Alternative Polyadenylation Pattern between Breast Cancer and Normal Cells. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0056958

Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., … Darnell, R. B. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*. http://doi.org/10.1038/nature07488

Lin, C., Yang, L., Yang, J. J., Huang, Y., & Liu, Z.-R. (2005). ATPase/Helicase Activities of p68 RNA Helicase Are Required for Pre-mRNA Splicing but Not for Assembly of the Spliceosome. *Molecular and Cellular Biology*. http://doi.org/10.1128/MCB.25.17.7484-7493.2005

Lin, Y., Li, Z., Ozsolak, F., Kim, S. W., Arango-Argoty, G., Liu, T. T., … John, B. (2012). An in-depth map of polyadenylation sites in cancer. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gks637

Liu, D., Brockman, J. M., Dass, B., Hutchins, L. N., Singh, P., McCarrey, J. R., … Graber, J. H. (2007). Systematic variation in mRNA 3′-processing signals during mouse spermatogenesis. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkl919

Liu, J., Wu, X., Zhang, H., Pfeifer, G. P., & Lu, Q. (2017). Dynamics of RNA Polymerase II Pausing and Bivalent Histone H3 Methylation during Neuronal Differentiation in Brain Development. *Cell Reports*. http://doi.org/10.1016/j.celrep.2017.07.046

Liu, S.-Y., Sanchez, D. J., Aliyari, R., Lu, S., & Cheng, G. (2012). Systematic identification of type I and type II interferon-induced antiviral factors. *Proceedings of the National Academy of Sciences of the United States of America*. http://doi.org/10.1073/pnas.1114981109

Liu, X., Freitas, J., Zheng, D., Oliveira, M. S., Hoque, M., Martins, T., … Moreira, A. (2017). Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in Drosophila melanogaster. *RNA*. http://doi.org/10.1261/rna.062661.117

Liu, Y., Hu, W., Murakawa, Y., Yin, J., Wang, G., Landthaler, M., & Yan, J. (2013). Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. *Scientific Reports*. http://doi.org/10.1038/srep02054

Liu, Z.-R. (2002). p68 RNA Helicase Is an Essential Human Splicing Factor That Acts at the U1 snRNA-5' Splice Site Duplex. *Molecular and Cellular Biology*. http://doi.org/10.1128/MCB.22.15.5443-5450.2002

Loya, A., Pnueli, L., Yosefzon, Y., Wexler, Y., Ziv-Ukelson, M., & Arava, Y. (2008). The 3???-UTR mediates the cellular localization of an mRNA encoding a short plasma membrane protein. *RNA*. http://doi.org/10.1261/rna.867208

Lukarska, M., Fournier, G., Pflug, A., Resa-Infante, P., Reich, S., Naffakh, N., & Cusack, S. (2017). Structural basis of an essential interaction between influenza polymerase and Pol II CTD. *Nature*. http://doi.org/10.1038/nature20594

Lukens, M. V, Kruijsen, D., Coenjaerts, F. E. J., Kimpen, J. L. L., & van Bleek, G. M. (2009). Respiratory syncytial virus-induced activation and migration of respiratory dendritic cells and subsequent antigen presentation in the lung-draining lymph node. *Journal of Virology*. http://doi.org/10.1128/JVI.00452-09

Lund, J. M., Alexopoulou, L., Sato, a, Karow, M., Adams, N. C., Gale, N. W., … Flavell, R. a. (2004). Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc Natl Acad Sci U S A*. http://doi.org/10.1073/pnas.0400937101 [doi]\r0400937101 [pii]

Luo, W., Ji, Z., Pan, Z., You, B., Hoque, M., Li, W., … Tian, B. (2013). The Conserved Intronic Cleavage and Polyadenylation Site of CstF-77 Gene Imparts Control of 3′ End Processing Activity through Feedback Autoregulation and by U1 snRNP. *PLoS Genetics*. http://doi.org/10.1371/journal.pgen.1003613

MacDonald, C. C., & McMahon, K. W. (2010). Tissue-specific mechanisms of alternative polyadenylation: Testis, brain, and beyond. *Wiley Interdisciplinary Reviews: RNA*. http://doi.org/10.1002/wrna.29

Mahat, D. B., Salamanca, H. H., Duarte, F. M., Danko, C. G., & Lis, J. T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Molecular Cell*. http://doi.org/10.1016/j.molcel.2016.02.025

Makeyev, A. V., & Liebhaber, S. A. (2002). The poly(C)-binding proteins: A multiplicity of functions and a search for mechanisms. *RNA*. http://doi.org/10.1017/S1355838202024627

Mandel, C. R., Bai, Y., & Tong, L. (2008). Protein factors in pre-mRNA 3′-end processing. *Cellular and Molecular Life Sciences*. http://doi.org/10.1007/s00018-007-7474-3

Mandel, C. R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J. L., & Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. *Nature*. http://doi.org/10.1038/nature05363

Mansfield, K. D., & Keene, J. D. (2012). Neuron-specific ELAV/Hu proteins suppress HuR mRNA during neuronal differentiation by alternative polyadenylation. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkr1114

Marc, P., Margeot, A., Devaux, F., Blugeon, C., Corral-Debrinski, M., & Jacq, C. (2002). Genome-wide analysis of mRNAs targeted to yeast mitochondria. *EMBO Reports*. http://doi.org/10.1093/embo-reports/kvf025

Marshall, N. F., Peng, J., Xie, Z., & Price, D. H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *The Journal of Biological Chemistry*. http://doi.org/10.1074/jbc.271.43.27176

Martin, G., Gruber, A. R., Keller, W., & Zavolan, M. (2012). Genome-wide Analysis of Pre-mRNA 3' End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3' UTR Length. *Cell Reports*. http://doi.org/10.1016/j.celrep.2012.05.003

Martin, K., & Helenius, A. (1992). Transport of incoming influenza virus nucleocapsids into the nucleus. *Trends in Cell Biology*. http://doi.org/10.1016/0962-8924(92)90130-F

Martin, K., & Heleniust, A. (1991). Nuclear transport of influenza virus ribonucleoproteins: The viral matrix protein (M1) promotes export and inhibits import. *Cell*. http://doi.org/10.1016/0092-8674(91)90576-K

Martínez-Alonso, M., Hengrung, N., & Fodor, E. (2016). RNA-Free and Ribonucleoprotein-Associated Influenza Virus Polymerases Directly Bind the Serine-5-Phosphorylated Carboxyl-Terminal Domain of Host RNA Polymerase II. *Journal of Virology*. http://doi.org/10.1128/JVI.00494-16

Martinez, F. J., Pratt, G. A., Van Nostrand, E. L., Batra, R., Huelga, S. C., Kapeli, K., … Yeo, G. W. (2016). Protein-RNA Networks Regulated by Normal and ALS-Associated Mutant HNRNPA2B1 in the Nervous System. *Neuron*. http://doi.org/10.1016/j.neuron.2016.09.050

Marzluff, W. F., Wagner, E. J., & Duronio, R. J. (2008). Metabolism and regulation of canonical histone mRNAs: Life without a poly(A) tail. *Nature Reviews Genetics*. http://doi.org/10.1038/nrg2438

Masamha, C. P., Xia, Z., Yang, J., Albrecht, T. R., Li, M., Shyu, A. Bin, … Wagner, E. J. (2014). CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature*. http://doi.org/10.1038/nature13261

Masuda, A., Takeda, J. I., Okuno, T., Okamoto, T., Ohkawara, B., Ito, M., … Ohno, K. (2015). Position-specific binding of FUS to nascent RNA regulates mRNA length. *Genes and Development*. http://doi.org/10.1101/gad.255737.114

Mayr, C. (2016). Evolution and Biological Roles of Alternative 3'UTRs. *Trends in Cell Biology*. http://doi.org/10.1016/j.tcb.2015.10.012

Mayr, C., & Bartel, D. P. (2009). Widespread Shortening of 3′UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell*. http://doi.org/10.1016/j.cell.2009.06.016

Mibayashi, M., Martinez-Sobrido, L., Loo, Y.-M., Cardenas, W. B., Gale, M., & Garcia-Sastre, A. (2007). Inhibition of Retinoic Acid-Inducible Gene I-Mediated Induction of Beta Interferon by the NS1 Protein of Influenza A Virus. *Journal of Virology*. http://doi.org/10.1128/JVI.01265-06

Michels, A. A., & Bensaude, O. (2018). Hexim1, an RNA-controlled protein hub. *Transcription*. http://doi.org/10.1080/21541264.2018.1429836

Michels, A. A., Fraldi, A., Li, Q., Adamson, T. E., Bonnet, F., Nguyen, V. T., … Bensaude, O. (2004). Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. *EMBO Journal*. http://doi.org/10.1038/sj.emboj.7600275

Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J. O., & Lai, E. C. (2013). Widespread and extensive lengthening of 39 UTRs in the mammalian brain. *Genome Research*. http://doi.org/10.1101/gr.146886.112

Moeller, A., Kirchdoerfer, R. N., Potter, C. S., Carragher, B., & Wilson, I. A. (2012). Organization of the influenza virus replication machinery. *Science*. http://doi.org/10.1126/science.1227270

Momose, F., Sekimoto, T., Ohkura, T., Jo, S., Kawaguchi, A., Nagata, K., & Morikawa, Y. (2011). Apical transport of influenza A virus ribonucleoprotein requires Rab11-positive recycling endosome. *PLoS ONE*. http://doi.org/10.1371/journal.pone.0021123

Montell, C., Fisher, E. F., Caruthers, M. H., & Berk, A. J. (1983). Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3′ consensus sequence AAUAAA. *Nature*. http://doi.org/10.1038/305600a0

Morchikh, M., Cribier, A., Raffel, R., Amraoui, S., Cau, J., Severac, D., … Benkirane, M. (2017). HEXIM1 and NEAT1 Long Non-coding RNA Form a Multi-subunit Complex that Regulates DNA-Mediated Innate Immune Response. *Molecular Cell*. http://doi.org/10.1016/j.molcel.2017.06.020

Morris, A. R., Bos, A., Diosdado, B., Rooijers, K., Elkon, R., Bolijn, A. S., … Agami, R. (2012). Alternative cleavage and polyadenylation during colorectal cancer development. *Clinical Cancer Research*. http://doi.org/10.1158/1078-0432.CCR-12-0543

Moustakas, a, Souchelnytskyi, S., & Heldin, C. H. (2001). Smad regulation in TGF-beta signal transduction. *Journal of Cell Science*. http://doi.org/10.1152/ajpgi.00208.2007.

Müller-McNicoll, M., Botti, V., de Jesus Domingues, A. M., Brandl, H., Schwich, O. D., Steiner, M. C., … Neugebauer, K. M. (2016). SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes and Development*. http://doi.org/10.1101/gad.276477.115

Muramoto, Y., Noda, T., Kawakami, E., Akkina, R., & Kawaoka, Y. (2013). Identification of novel influenza A virus proteins translated from PA mRNA. *Journal of Virology*. http://doi.org/10.1128/JVI.02656-12

Murthy, K. G. K., & Manley, J. L. (1995). The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes and Development*. http://doi.org/10.1101/gad.9.21.2672

Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y. F., Goshima, N., & Hirose, T. (2012). Alternative 3???-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO Journal*. http://doi.org/10.1038/emboj.2012.251

Nelson, M. I., & Holmes, E. C. (2007). The evolution of epidemic influenza. *Nature Reviews Genetics*. http://doi.org/10.1038/nrg2053

Nemeroff, M. E., Barabino, S. M. L., Li, Y., Keller, W., & Krug, R. M. (1998). Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3′ end formation of cellular pre-mRNAs. *Molecular Cell*. http://doi.org/10.1016/S1097-2765(00)80099-4

Neumann, G., Hughes, M. T., & Kawaoka, Y. (2000). Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *EMBO Journal*. http://doi.org/10.1093/emboj/19.24.6751

Neumann, G., & Kawaoka, Y. (2015). Transmission of influenza A viruses. *Virology*. http://doi.org/10.1016/j.virol.2015.03.009

Neve, J., Burger, K., Li, W., Hoque, M., Patel, R., Tian, B., … Furger, A. (2016). Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome Research*. http://doi.org/10.1101/gr.193995.115

Neve, J., & Furger, A. (2014). Alternative polyadenylation: less than meets the eye? *Biochemical Society Transactions*. http://doi.org/10.1042/BST20140054

Neve, J., Patel, R., Wang, Z., Louey, A., & Furger, A. M. (2017). New dawn for the poly(A) site: ending the message expands gene regulation. *RNA Biology*. http://doi.org/10.1080/15476286.2017.1306171

Nguyen, N. T. H., Now, H., Kim, W. J., Kim, N., & Yoo, J. Y. (2016). Ubiquitin-like modifier FAT10 attenuates RIG-I mediated antiviral signaling by segregating activated RIG-I from its signaling platform. *Scientific Reports*. http://doi.org/10.1038/srep23377

Noah, D. L., Twu, K. Y., & Krug, R. M. (2003). Cellular antiviral responses against influenza A virus are countered at the posttranscriptional level by the viral NS1A protein via its binding to a cellular protein required for the 3′ end processing of cellular pre-mRNAS. *Virology*. http://doi.org/10.1016/S0042-6822(02)00127-7

Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R. H., & Kawaoka, Y. (2006). Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature*. http://doi.org/10.1038/nature04378

Nogales, A., Martinez-Sobrido, L., Topham, D. J., & DeDiego, M. L. (2017). NS1 Protein Amino Acid Changes D189N and V194I Affect Interferon Responses, Thermosensitivity, and Virulence of Circulating H3N2 Human Influenza A Viruses. *Journal of Virology*. http://doi.org/10.1128/JVI.01930-16

Noton, S. L., Medcalf, E., Fisher, D., Mullin, A. E., Elton, D., & Digard, P. (2007). Identification of the domains of the influenza A virus M1 matrix protein required for NP binding, oligomerization and incorporation into virions. *Journal of General Virology*. http://doi.org/10.1099/vir.0.82809-0

Nunes, N. M., Li, W., Tian, B., & Furger, A. (2010). A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. *EMBO Journal*. http://doi.org/10.1038/emboj.2010.42

O'Neill, R. E., Talon, J., & Palese, P. (1998). The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *EMBO Journal*. http://doi.org/10.1093/emboj/17.1.288

Oishi, K., Yamayoshi, S., & Kawaoka, Y. (2015). Mapping of a Region of the PA-X Protein of Influenza A Virus That Is Important for Its Shutoff Activity. *Journal of Virology*. http://doi.org/10.1128/JVI.01132-15

Orkin, S. H., Cheng, T. C., Antonarakis, S. E., & Kazazian, H. H. (1985). Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. *The EMBO Journal*. http://doi.org/10.1002/J.1460-2075.1985.TB03650.X

Palese, P. & Shaw, M.L. (2013). Orthomyxoviridae. In: D. M. Knipe & P. M. Howley (eds.). *Fields Virology*. Philadelphia, pp. 1151–1185.

Park, B. V., Freeman, Z. T., Ghasemzadeh, A., Chattergoon, M. A., Rutebemberwa, A., Steigner, J., … Cox, A. L. (2016). TGFβ1-mediated SMAD3 enhances PD-1 expression on antigen-specific T cells in cancer. *Cancer Discovery*. http://doi.org/10.1158/2159-8290.CD-15-1347

Paterson, D., & Fodor, E. (2012). Emerging Roles for the Influenza A Virus Nuclear Export Protein (NEP). *PLoS Pathogens*. http://doi.org/10.1371/journal.ppat.1003019

Perez, D. R., & Donis, R. O. (1998). The matrix 1 protein of influenza A virus inhibits the transcriptase activity of a model influenza reporter genome in Vivo. *Virology*. http://doi.org/10.1006/viro.1998.9318

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*. http://doi.org/10.1126/science.1132998

Pinto, L. H., Holsinger, L. J., & Lamb, R. A. (1992). Influenza virus M2 protein has ion channel activity. *Cell*. http://doi.org/10.1016/0092-8674(92)90452-I

Pinto, L. H., & Lamb, R. A. (2006). The M2 proton channels of influenza A and B viruses. *Journal of Biological Chemistry*. http://doi.org/10.1074/jbc.R500020200

Pinto, P. A. B., Henriques, T., Freitas, M. O., Martins, T., Domingues, R. G., Wyrzykowska, P. S., … Moreira, A. (2011). RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO Journal*. http://doi.org/10.1038/emboj.2011.156

Pipkin, M. E., Sacks, J. A., Cruz-Guilloty, F., Lichtenheld, M. G., Bevan, M. J., & Rao, A. (2010). Interleukin-2 and Inflammation Induce Distinct Transcriptional Programs that Promote the Differentiation of Effector Cytolytic T Cells. *Immunity*. http://doi.org/10.1016/j.immuni.2009.11.012

Plotch, S. J., Bouloy, M., Ulmanen, I., & Krug, R. M. (1981). A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell*. http://doi.org/10.1016/0092-8674(81)90449-9

Poon, L. L., Pritlove, D. C., Fodor, E., & Brownlee, G. G. (1999). Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. *Journal of Virology*. http://doi.org/10074205

Proudfoot, N. J. (2011). Ending the message: Poly(A) signals then and now. *Genes and Development*. http://doi.org/10.1101/gad.17268411

Proudfoot, N. J., Furger, A., & Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell*. http://doi.org/10.1016/S0092-8674(02)00617-7

Qui, Y., Nemeroff, M., & Krug, R. (1995). The influenza virus NS1 protein binds to a specific region in human U6 snRNA and inhibits U6-U2 and U6-U4 snRNA interactions during splicing. *RNA*. http://doi.org/D - NLM: PMC1369083 EDAT- 1995/05/01 MHDA- 1995/05/01 00:01 CRDT- 1995/05/01 00:00 PST - ppublish

Rahman, M. M., Bagdassarian, E., Ali, M. A. M., & McFadden, G. (2017). Identification of host DEAD-box RNA helicases that regulate cellular tropism of oncolytic Myxoma virus in human cancer cells. *Scientific Reports*. http://doi.org/10.1038/s41598-017-15941-1

Ramos, I., Carnero, E., Bernal-Rubio, D., Seibert, C. W., Westera, L., García-Sastre, A., & Fernandez-Sesma, A. (2013). Contribution of double-stranded RNA and CPSF30 binding domains of influenza virus NS1 to the inhibition of type I interferon production and activation of human dendritic cells. *Journal of Virology*. http://doi.org/10.1128/JVI.02247-12

Read, E. K. C., & Digard, P. (2010). Individual influenza A virus mRNAs show differential dependence on cellular NXF1/TAP for their nuclear export. *Journal of General Virology*. http://doi.org/10.1099/vir.0.018564-0

Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crepin, T., … Cusack, S. (2014). Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature*. http://doi.org/10.1038/nature14009

Reikine, S., Nguyen, J. B., & Modis, Y. (2014). Pattern recognition and signaling mechanisms of RIG-I and MDA5. *Frontiers in Immunology*. http://doi.org/10.3389/fimmu.2014.00342

Rhinn, H., Qiang, L., Yamashita, T., Rhee, D., Zolin, A., Vanti, W., & Abeliovich, A. (2012). Alternative α-synuclein transcript usage as a convergent mechanism in Parkinson's disease pathology. *Nature Communications*. http://doi.org/10.1038/ncomms2032

Richardson, J. C., & Akkina, R. K. (1991). NS2protein of influenza virus is found in purified virus and phosphorylated in infected cells. *Archives of Virology*. http://doi.org/10.1007/BF01319232

Robb, N. C., Smith, M., Vreede, F. T., & Fodor, E. (2009). NS2/NEP protein regulates transcription and replication of the influenza virus RNA genome. *The Journal of General Virology*. http://doi.org/10.1099/vir.0.009639-0

Robb, N. C., & Fodor, E. (2012). The accumulation of influenza A virus segment 7 spliced mRNAs is regulated by the NS1 protein. *Journal of General Virology*. http://doi.org/10.1099/vir.0.035485-0

Robb, N. C., Jackson, D., Vreede, F. T., & Fodor, E. (2010). Splicing of influenza A virus NS1 mRNA is independent of the viral NS1 protein. *Journal of General Virology*. http://doi.org/10.1099/vir.0.022004-0

Robertson, J. S., Schubert, M., & Lazzarini, R. A. (1981). Polyadenylation sites for influenza virus mRNA. *Journal of Virology*.

Rodriguez, A., Perez-Gonzalez, A., & Nieto, A. (2007). Influenza Virus Infection Causes Specific Degradation of the Largest Subunit of Cellular RNA Polymerase II. *Journal of Virology*. http://doi.org/10.1128/JVI.02129-06

Rossman, J. S., Jing, X., Leser, G. P., Balannik, V., Pinto, L. H., & Lamb, R. A. (2010). Influenza Virus M2 Ion Channel Protein Is Necessary for Filamentous Virion Formation. *Journal of Virology*. http://doi.org/10.1128/JVI.00119-10

Rossman, J. S., & Lamb, R. A. (2011). Influenza virus assembly and budding. *Virology*. http://doi.org/10.1016/j.virol.2010.12.003

Saldi, T., Cortazar, M. A., Sheridan, R. M., & Bentley, D. L. (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *Journal of Molecular Biology*. http://doi.org/10.1016/j.jmb.2016.04.017

Saldi, T., Fong, N., & Bentley, D. L. (2018). Transcription elongation rate affects nascent histone pre-mRNA folding and 3′ end processing. *Genes and Development*. http://doi.org/10.1101/gad.310896.117

Salvatore, M., Basler, C. F., Parisien, J.-P., Horvath, C. M., Bourmakina, S., Zheng, H., … García-Sastre, A. (2002). Effects of influenza A virus NS1 protein on protein expression: the NS1 protein enhances translation and is not required for shutoff of host protein synthesis. *Journal of Virology*. http://doi.org/10.1128/JVI.76.3.1206

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. a, & Burge, C. B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science (New York, N.Y.)*. http://doi.org/10.1126/science.1155390

Saporita, A. J., Chang, H. C., Winkeler, C. L., Apicelli, A. J., Kladney, R. D., Wang, J., … Weber, J. D. (2011). RNA helicase DDX5 is a p53-independent target of ARF that

participates in ribosome biogenesis. *Cancer Research*. http://doi.org/10.1158/0008-5472.CAN-11-1472

Satterly, N., Tsai, P.-L., van Deursen, J., Nussenzveig, D. R., Wang, Y., Faria, P. A., … Fontoura, B. M. A. (2007). Influenza virus targets the mRNA export machinery and the nuclear pore complex. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.0610977104

Scheiffele, P., Rietveld, A., Wilk, T., & Simons, K. (1999). Influenza viruses select ordered lipid domains during budding from the plasma membrane. *Journal of Biological Chemistry*. http://doi.org/10.1074/jbc.274.4.2038

Schneider, W. M., Chevillotte, M. D., & Rice, C. M. (2014). Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annual Review of Immunology*. http://doi.org/10.1146/annurev-immunol-032713-120231

Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A. R., Keller, W., … Wahle, E. (2014). Reconstitution of CPSF active in polyadenylation: Recognition of the polyadenylation signal by WDR33. *Genes and Development*. http://doi.org/10.1101/gad.250985.114

Schönichen, A., Bigalke, J. M., Urbanke, C., Grzesiek, S., Dames, S. A., & Geyer, M. (2010). A flexible bipartite coiled coil structure is required for the interaction of hexim1 with the P-TEFb subunit cyclin T1. *Biochemistry*. http://doi.org/10.1021/bi902072f

Schulz, O., Diebold, S. S., Chen, M., Näslund, T. I., Nolte, M. A., Alexopoulou, L., … Reis E Sousa, C. (2005). Toll-like receptor 3 promotes cross-priming to virus-infected cells. *Nature*. http://doi.org/10.1038/nature03326

Schwanhüusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., … Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*. http://doi.org/10.1038/nature10098

Selman, M., Dankar, S. K., Forbes, N. E., Jia, J. J., & Brown, E. G. (2012). Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. *Emerging Microbes and Infections*. http://doi.org/10.1038/emi.2012.38

Seo, S. H., Webby, R., & Webster, R. G. (2004). No apoptotic deaths and different levels of inductions of inflammatory cytokines in alveolar macrophages infected with influenza viruses. *Virology*. http://doi.org/10.1016/j.virol.2004.08.019

Shankarling, G. S., & MacDonald, C. C. (2013). Polyadenylation site-specific differences in the activity of the neuronal βCstF-64 protein in PC-12 cells. *Gene*. http://doi.org/10.1016/j.gene.2013.08.007

Shapiro, G. I., Gurney Jr., T., & Krug, R. M. (1987). Influenza virus gene expression: control mechanisms at early and late times of infection and nuclear-cytoplasmic transport of virus-specific RNAs. *Journal of Virology*.

Sharon, D., Tilgner, H., Grubert, F., & Snyder, M. (2013). A single-molecule long-read survey of the human transcriptome. *Nature Biotechnology*. http://doi.org/10.1038/nbt.2705

Shaw, M. L., Stone, K. L., Colangelo, C. M., Gulcicek, E. E., & Palese, P. (2008). Cellular proteins in influenza virus particles. *PLoS Pathogens*. http://doi.org/10.1371/journal.ppat.1000085

Shaw, M. W., & Lamb, R. A. (1984). A specific sub-set of host-cell mRNAs prime influenza virus mRNA synthesis. *Virus Research*. http://doi.org/10.1016/0168-1702(84)90003-0

Sheets, M. D., Ogg, S. C., & Wickens, M. P. (1990). Point mutations in AAUAAA and the poly (A) addition site: Effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Research*. http://doi.org/10.1093/nar/18.19.5799

Shepard, P. J., Choi, E. A., Lu, J., Flanagan, L. A., Hertel, K. J., & Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA*. http://doi.org/10.1261/rna.2581711

Shi, Y. (2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes & Development*. http://doi.org/10.1101/gad.250993.114

Shi, Y., Di Giammartino, D. C., Taylor, D., Sarkeshik, A., Rice, W. J., Yates, J. R., … Manley, J. L. (2009). Molecular Architecture of the Human Pre-mRNA 3′ Processing Complex. *Molecular Cell*. http://doi.org/10.1016/j.molcel.2008.12.028

Shi, Y., & Manley, J. L. (2015). The end of the message: Multiple protein–RNA interactions define the mRNA polyadenylation site. *Genes and Development*. http://doi.org/10.1101/gad.261974.115

Shih, S. R., & Krug, R. M. (1996). Novel exploitation of a nuclear function by influenza virus: the cellular SF2/ASF splicing factor controls the amount of the essential viral M2 ion channel protein in infected cells. *EMBO J*. http://doi.org/10.1002/j.1460-2075.1996.tb00925.x

Shih, S. R., Suen, P. C., Chen, Y. S., & Chang, S. C. (1998). Novel spliced transcript of influenza A/WSN/33 virus. *Virus Genes*. http://doi.org/10.1023/A:1008024909222

Shimizu, K., Iguchi, A., Gomyou, R., & Yasushi, O. (1999). Influenza virus inhibits cleavage of the HSP70 pre-mRNAs at the polyadenylation site. *Virology*. http://doi.org/10.1006/viro.1998.9555

Sikora, D., Rocheleau, L., Brown, E. G., & Pelchat, M. (2014). Deep sequencing reveals the eight facets of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process. *Scientific Reports*. http://doi.org/10.1038/srep06181

Simonelig, M. (2012). PABPN1 shuts down alternative poly(A) sites. *Cell Research*. http://doi.org/10.1038/cr.2012.86

Simons, K., & Ikonen, E. (1997). Functional rafts in cell membranes. *Nature*. http://doi.org/10.1038/42408

Singh, P., Alley, T. L., Wright, S. M., Kamdar, S., Schott, W., Wilpan, R. Y., … Graber, J. H. (2009). Global changes in processing of mRNA 3′ untranslated regions characterize clinically distinct cancer subtypes. *Cancer Research*. http://doi.org/10.1158/0008-5472.CAN-09-2236

Skehel, J. J., & Wiley, D. C. (2000). Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annual Review of Biochemistry*. http://doi.org/10.1146/annurev.biochem.69.1.531

Smed-Sörensen, A., Chalouni, C., Chatterjee, B., Cohn, L., Blattmann, P., Nakamura, N., … Mellman, I. (2012). Influenza a virus infection of human primary dendritic cells impairs their ability to cross-present antigen to CD8 T cells. *PLoS Pathogens*. http://doi.org/10.1371/journal.ppat.1002572

Smibert, P., Miura, P., Westholm, J. O., Shenker, S., May, G., Duff, M. O., … Lai, E. C. (2012). Global Patterns of Tissue-Specific Alternative Polyadenylation in Drosophila. *Cell Reports*. http://doi.org/10.1016/j.celrep.2012.01.001

Smith ', D. B., & Inglis, S. C. (1985). Regulated production of an influenza virus spliced mRNA mediated by virus-specific products. *The EMBO Journal*.

Smith, G. L., & Hay, A. J. (1982). Replication of the influenza virus genome. *Virology*, *118*(1), 96–108. http://doi.org/https://doi.org/10.1016/0042-6822(82)90323-3

Spies, N., Burge, C. B., & Bartel, D. P. (2013). 3′ UTR-Isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Research*. http://doi.org/10.1101/gr.156919.113

Spies, N., Nielsen, C. B., Padgett, R. A., & Burge, C. B. (2009). Biased Chromatin Signatures around Polyadenylation Sites and Exons. *Molecular Cell*. http://doi.org/10.1016/j.molcel.2009.10.008

Steidle, S., Martínez-Sobrido, L., Mordstein, M., Lienenklaus, S., García-Sastre, A., Stäheli, P., & Kochs, G. (2010). Glycine 184 in nonstructural protein NS1 determines the virulence of influenza A virus strain PR8 without affecting the host interferon response. *Journal of Virology*. http://doi.org/10.1128/JVI.00701-10

Sun, M., Ding, J., Li, D., Yang, G., Cheng, Z., & Zhu, Q. (2017). NUDT21 regulates 3′-UTR length and microRNA-mediated gene silencing in hepatocellular carcinoma. *Cancer Letters*. http://doi.org/10.1016/j.canlet.2017.09.026

Sun, Y., Zhang, Y., Hamilton, K., Manley, J. L., Shi, Y., Walz, T., & Tong, L. (2018). Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *Proceedings of the National Academy of Sciences*, *115*(7), E1419 LP-E1428. Retrieved from http://www.pnas.org/content/115/7/E1419.abstract

Swain, S. L., McKinstry, K. K., & Strutt, T. M. (2012). Expanding roles for CD4 + T cells in immunity to viruses. *Nature Reviews Immunology*. http://doi.org/10.1038/nri3152

Szabo, S. J., Kim, S. T., Costa, G. L., Zhang, X., Fathman, C. G., & Glimcher, L. H. (2000). A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell*. http://doi.org/10.1016/S0092-8674(00)80702-3

Takagaki, Y., Seipelt, R. L., Peterson, M. L., & Manley, J. L. (1996). The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell*. http://doi.org/10.1016/S0092-8674(00)82000-0

Takizawa, T., Matsukawa, S., Higuchi, Y., Nakamura, S., Nakanishi, Y., & Fukuda, R. (1993). Induction of programmed cell death (apoptosis) by influenza virus infection in tissue culture cells. *The Journal of General Virology*. http://doi.org/10.1099/0022-1317-74-11-2347

Taliaferro, J. M., Vidaki, M., Oliveira, R., Olson, S., Zhan, L., Saxena, T., … Burge, C. B. (2016). Distal Alternative Last Exons Localize mRNAs to Neural Projections. *Molecular Cell*. http://doi.org/10.1016/j.molcel.2016.01.020

Talon, J., Horvath, C. M., Polley, R., Basler, C. F., Muster, T., Palese, P., & Garcia-Sastre, A. (2000). Activation of Interferon Regulatory Factor 3 Is Inhibited by the Influenza A Virus NS1 Protein. *Journal of Virology*. http://doi.org/10.1128/JVI.74.17.7989-7996.2000

Tan, S., Li, H., Zhang, W., Shao, Y., Liu, Y., Guan, H., … Zhu, T. (2018). NUDT21 negatively regulates PSMB2 and CXXC5 by alternative polyadenylation and contributes to hepatocellular carcinoma suppression. *Oncogene*. http://doi.org/10.1038/s41388-018-0280-6

Tang, S., Patel, A., & Krause, P. R. (2016). Herpes simplex virus ICP27 regulates alternative pre-mRNA polyadenylation and splicing in a sequence-dependent manner. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.1609695113

te Velthuis, A. J. W., & Fodor, E. (2016). Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nature Reviews Microbiology*. http://doi.org/10.1038/nrmicro.2016.87

Thierry, E., Guilligay, D., Kosinski, J., Bock, T., Gaudon, S., Round, A., … Cusack, S. (2016). Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2 Domains. *Molecular Cell*. http://doi.org/10.1016/j.molcel.2015.11.016

Thomas, P. E., Wu, X., Liu, M., Gaffney, B., Ji, G., Li, Q. Q., & Hunt, A. G. (2012). Genome-Wide Control of Polyadenylation Site Choice by CPSF30 in Arabidopsis. *The Plant Cell*. http://doi.org/10.1105/tpc.112.096107

Tian, B., Hu, J., Zhang, H., & Lutz, C. S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gki158

Tian, B., & Manley, J. L. (2013). Alternative cleavage and polyadenylation: The long and short of it. *Trends in Biochemical Sciences*. http://doi.org/10.1016/j.tibs.2013.03.005

Tokunaga, K., Shibuya, T., Ishihama, Y., Tadakuma, H., Ide, M., Yoshida, M., … Tani, T. (2006). Nucleocytoplasmic transport of fluorescent mRNA in living mammalian cells: Nuclear mRNA export is coupled to ongoing gene transcription. *Genes to Cells*. http://doi.org/10.1111/j.1365-2443.2006.00936.x

Tranter, M., Helsley, R. N., Paulding, W. R., McGuinness, M., Brokamp, C., Haar, L., … Jones, W. K. (2011). Coordinated post-transcriptional regulation of Hsp70.3 gene expression by microRNA and alternative polyadenylation. *Journal of Biological Chemistry*. http://doi.org/10.1074/jbc.M111.221796

Tsai, P. L., Chiou, N. T., Kuss, S., García-Sastre, A., Lynch, K. W., & Fontoura, B. M. A. (2013). Cellular RNA Binding Proteins NS1-BP and hnRNP K Regulate Influenza A Virus RNA Splicing. *PLoS Pathogens*. http://doi.org/10.1371/journal.ppat.1003460

Turrell, L., Lyall, J. W., Tiley, L. S., Fodor, E., & Vreede, F. T. (2013). The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. *Nature Communications*. http://doi.org/10.1038/ncomms2589

Twu, K. Y., Kuo, R.-L., Marklund, J., & Krug, R. M. (2007). The H5N1 Influenza Virus NS Genes Selected after 1998 Enhance Virus Replication in Mammalian Cells. *Journal of Virology*. http://doi.org/10.1128/JVI.00006-07

Twu, K. Y., Noah, D. L., Rao, P., Kuo, R.-L., & Krug, R. M. (2006). The CPSF30 Binding Site on the NS1A Protein of Influenza A Virus Is a Potential Antiviral Target. *Journal of Virology*. http://doi.org/10.1128/JVI.80.8.3957-3965.2006

Ulitsky, I., Shkumatava, A., Jan, C. H., Subtelny, A. O., Koppstein, D., Bell, G. W., … Bartel, D. P. (2012). Extensive alternative polyadenylation during zebrafish development. *Genome Research*. http://doi.org/10.1101/gr.139733.112

Varga, Z. T., Ramos, I., Hai, R., Schmolke, M., García-Sastre, A., Fernandez-Sesma, A., & Palese, P. (2011). The influenza virus protein PB1-F2 inhibits the induction of type i interferon at the level of the MAVS adaptor protein. *PLoS Pathogens*. http://doi.org/10.1371/journal.ppat.1002067

Vasin, A. V., Temkina, O. A., Egorov, V. V., Klotchenko, S. A., Plotnikova, M. A., & Kiselev, O. I. (2014). Molecular mechanisms enhancing the proteome of influenza A viruses: An overview of recently discovered proteins. *Virus Research*. http://doi.org/10.1016/j.virusres.2014.03.015

Vilborg, A., Passarelli, M. C., Yario, T. A., Tycowski, K. T., & Steitz, J. A. (2015). Widespread Inducible Transcription Downstream of Human Genes. *Molecular Cell*. http://doi.org/10.1016/j.molcel.2015.06.016

Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T. A., … Shalgi, R. (2017). Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.1711120114

Vreede, F. T., Gifford, H., & Brownlee, G. G. (2008). Role of Initiating Nucleoside Triphosphate Concentrations in the Regulation of Influenza Virus Replication and Transcription. *Journal of Virology*. http://doi.org/10.1128/JVI.00627-08

Vreede, F. T., Ng, A. K.-L., Shaw, P.-C., & Fodor, E. (2011). Stabilization of Influenza Virus Replication Intermediates Is Dependent on the RNA-Binding but Not the Homo-Oligomerization Activity of the Viral Nucleoprotein. *Journal of Virology*. http://doi.org/10.1128/JVI.00695-11

Vreede, F. T., Chan, A. Y., Sharps, J., & Fodor, E. (2010). Mechanisms and functional implications of the degradation of host RNA polymerase II in influenza virus infected cells. *Virology*. http://doi.org/10.1016/j.virol.2009.10.003

Wahle, E. (1991). Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors. *Journal of Biological Chemistry*.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., … Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*. http://doi.org/10.1038/nature07509

Wang, J., Oberley-Deegan, R., Wang, S., Nikrad, M., Funk, C. J., Hartshorn, K. L., & Mason, R. J. (2009). Differentiated human alveolar type II cells secrete antiviral IL-29 (IFN-lambda 1) in response to influenza A infection. *Journal of Immunology (Baltimore, Md. : 1950)*. http://doi.org/10.4049/jimmunol.182.3.1296

Wang, L., Fu, B., Li, W., Patil, G., Liu, L., Dorf, M. E., & Li, S. (2017). Comparative influenza protein interactomes identify the role of plakophilin 2 in virus restriction. *Nature Communications*. http://doi.org/10.1038/ncomms13876

Wang, R., Zheng, D., Yehia, G., & Tian, B. (2018). A compendium of conserved cleavage and polyadenylation events in mammalian genes. *Genome Research* . http://doi.org/10.1101/gr.237826.118

Wang, W., & Krug, R. M. (1998). U6atac snRNA, the highly divergent counterpart of U6 snRNA, is the specific target that mediates inhibition of AT-AC splicing by the influenza virus NS1 protein. *RNA (New York, N.Y.)*.

Wang, W., Cui, Z. Q., Han, H., Zhang, Z. P., Wei, H. P., Zhou, Y. F., … Zhang, X. E. (2008). Imaging and characterizing influenza A virus mRNA transport in living cells. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkn475

Wang, X., Li, M., Zheng, H., Muster, T., Palese, P., Beg, a a, & García-Sastre, A. (2000). Influenza A virus NS1 protein prevents activation of NF-kappaB and induction of alpha/beta interferon. *Journal of Virology*. http://doi.org/10.1128/JVI.74.24.11566-11573.2000.Updated

Watanabe, K., Handa, H., Mizumoto, K., & Nagata, K. (1996). Mechanism for inhibition of influenza virus RNA polymerase activity by matrix protein. *Journal of Virology*.

Watanabe, M., Yanagisawa, J., Kitagawa, H., Takeyama, K. I., Ogawa, S., Arao, Y., … Kato, S. (2001). A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor α coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. *EMBO Journal*. http://doi.org/10.1093/emboj/20.6.1341

West, S., & Proudfoot, N. J. (2008). Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gkm1112

Whitelaw, E., & Proudfoot, N. (1986). Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *The EMBO Journal*. http://doi.org/10.1002/j.1460-2075.1986.tb04587.x

Whitmire, J. K., Tan, J. T., & Whitton, J. L. (2005). Interferon-gamma acts directly on CD8+ T cells to increase their abundance during virus infection. *The Journal of Experimental Medicine*. http://doi.org/10.1084/jem.20041463

Wickens, M., & Stephenson, P. (1984). Role of the Conserved AAUAAA Sequence: Four AAUAAA Point Mutants Prevent Messenger RNA 3 End Formation. *Science*, *226*(4678), 1045–1051.

Wilusz, J. E., Freier, S. M., & Spector, D. L. (2008). 3′ End Processing of a Long Nuclear-Retained Noncoding RNA Yields a tRNA-like Cytoplasmic RNA. *Cell*. http://doi.org/10.1016/j.cell.2008.10.012

Winter, G., & Fields, S. (1980). Cloning of influenza cDNA into M13: The sequence of the RNA segment encoding the A/PR/8/34 matrix protein. *Nucleic Acids Research*. http://doi.org/10.1093/nar/8.9.1965

Winter, G., & Fields, S. (1982). Nucteotide sequence of human influenza A/PR/8/34 segment 2. *Nucleic Acids Research*. http://doi.org/10.1093/nar/10.6.2135

Winter, G., Fields, S., & Brownlee, G. G. (1981). Nucleotide sequence of the haemagglutinin gene of a human influenza virus H1 subtype. *Nature*. http://doi.org/10.1038/292072a0

Winter, G., Fields, S., & Ratti, G. (1981). The structure of two subgenomic RNAs from human infhienza virus A/PR/8/34. *Nucleic Acids Research*. http://doi.org/10.1093/nar/9.24.6907

Wise, H. M., Foeglein, A., Sun, J., Dalton, R. M., Patel, S., Howard, W., … Digard, P. (2009). A Complicated Message: Identification of a Novel PB1-Related Protein Translated from Influenza A Virus Segment 2 mRNA. *Journal of Virology*. http://doi.org/10.1128/JVI.00826-09

Wise, H. M., Hutchinson, E. C., Jagger, B. W., Stuart, A. D., Kang, Z. H., Robb, N., … Digard, P. (2012). Identification of a Novel Splice Variant Form of the Influenza A Virus M2 Ion

Channel with an Antigenically Distinct Ectodomain. *PLoS Pathogens*. http://doi.org/10.1371/journal.ppat.1002998

Wood, A. J., Schulz, R., Woodfine, K., Koltowska, K., Beechey, C. V, Peters, J., … Oakey, R. J. (2008). Regulation of alternative polyadenylation by genomic imprinting. *Genes & Development*. http://doi.org/10.1101/gad.473408

Wright, P.F., Neumann, G. & Kawaoka, Y. (2013). Orthomyxoviridae. In: D. M. Knipe & P. M. Howley (eds.). *Fields Virology*. Philadelphia: Lippincott Williams and Wilkins, pp. 1186–1243.

Xia, Z., Donehower, L. A., Cooper, T. A., Neilson, J. R., Wheeler, D. A., Wagner, E. J., & Li, W. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'2-UTR landscape across seven tumour types. *Nature Communications*. http://doi.org/10.1038/ncomms6274

Yamayoshi, S., Watanabe, M., Goto, H., & Kawaoka, Y. (2016). Identification of a Novel Viral Protein Expressed from the PB2 Segment of Influenza A Virus. *Journal of Virology*. http://doi.org/10.1128/JVI.02175-15

Yang, L., Lin, C., & Liu, Z.-R. (2005). Phosphorylations of DEAD box p68 RNA helicase are associated with cancer development and cell proliferation. *Molecular Cancer Research : MCR*. http://doi.org/10.1158/1541-7786.MCR-05-0022

Yángüez, E., Castello, A., Welnowska, E., Carrasco, L., Goodfellow, I., & Nieto, A. (2011). Functional impairment of eIF4A and eIF4G factors correlates with inhibition of influenza virus mRNA translation. *Virology*. http://doi.org/10.1016/j.virol.2011.02.012

Ye, P., Liu, S., Zhu, Y., Chen, G., & Gao, G. (2010). DEXH-Box protein DHX30 is required for optimal function of the zinc-finger antiviral protein. *Protein and Cell*. http://doi.org/10.1007/s13238-010-0117-8

Yik, J. H. N., Chen, R., Nishimura, R., Jennings, J. L., Link, A. J., & Zhou, Q. (2003). Inhibition of P-TEFb (CDK9/cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Molecular Cell*. http://doi.org/10.1016/S1097-2765(03)00388-5

Yoon, Y., McKenna, M. C., Rollins, D. A., Song, M., Nuriel, T., Gross, S. S., … Glatt, C. E. (2013). Anxiety-associated alternative polyadenylation of the serotonin transporter mRNA confers translational regulation by hnRNPK. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.1301485110

York, A., Hengrung, N., Vreede, F. T., Huiskonen, J. T., & Fodor, E. (2013). Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.1315068110

York, A., & Fodor, E. (2013). Biogenesis, assembly and export of viral messenger ribonucleoproteins in the influenza A virus infected cell. *RNA Biology*. http://doi.org/10.4161/rna.25356

Yu, M., Liu, X., Cao, S., Zhao, Z., Zhang, K., Xie, Q., … Liu, W. (2012). Identification and Characterization of Three Novel Nuclear Export Signals in the Influenza A Virus Nucleoprotein. *Journal of Virology*. http://doi.org/10.1128/JVI.06159-11

Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., … Liu, Y. (2009). Crystal structure of an avian influenza polymerase PAN reveals an endonuclease active site. *Nature*. http://doi.org/10.1038/nature07720

Zaborowska, J., Egloff, S., & Murphy, S. (2016). The pol II CTD: New twists in the tail. *Nature Structural and Molecular Biology*. http://doi.org/10.1038/nsmb.3285

Zhang, J., Leser, G. P., Pekosz, A., & Lamb, R. A. (2000). The cytoplasmic tails of the influenza virus spike glycoproteins are required for normal genome packaging. *Virology*. http://doi.org/10.1006/viro.2000.0228

Zhang, P., Elabd, S., Hammer, S., Solozobova, V., Yan, H., Bartel, F., … Blattner, C. (2015). TRIM25 has a dual function in the p53/Mdm2 circuit. *Oncogene*. http://doi.org/10.1038/onc.2015.21

Zhang, Q., Zeng, L. P., Zhou, P., Irving, A. T., Li, S., Shi, Z. L., & Wang, L. F. (2017). IFNAR2-dependent gene expression profile induced by IFN-α in Pteropus alecto bat cells and impact of IFNAR2 knockout on virus infection. *PloS One*. http://doi.org/10.1371/journal.pone.0182866

Zhao, J., Hyman, L., & Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and Molecular Biology Reviews : MMBR*. http://doi.org/10.1016/J.CELL.2011.01.032

Zhao, N., Sebastiano, V., Moshkina, N., Mena, N., Hultquist, J., Jimenez-Morales, D., … Marazzi, I. (2018). Influenza virus infection causes global RNAPII termination defects. *Nature Structural & Molecular Biology*, *25*(9), 885–893. http://doi.org/10.1038/s41594-018-0124-7

Zhirnov, O. P., Vorobjeva, I. V., Saphonova, O. A., Poyarkov, S. V., Ovcharenko, A. V., Anhlan, D., & Malyshev, N. A. (2009). Structural and evolutionary characteristics of HA, NA, NS and M genes of clinical influenza A/H3N2 viruses passaged in human and canine cells. *Journal of Clinical Virology*. http://doi.org/10.1016/j.jcv.2009.05.030

Zhou, L., Liu, F., Wang, X., & Ouyang, G. (2015). The roles of microRNAs in the regulation of tumor metastasis. *Cell & Bioscience*, *5*, 32. http://doi.org/10.1186/s13578-015-0028-8

Zhou, Q., Li, T., & Price, D. H. (2012). RNA Polymerase II Elongation Control. *Annual Review of Biochemistry*. http://doi.org/10.1146/annurev-biochem-052610-095910

Zhu, H., Zhou, H. L., Hasman, R. A., & Lou, H. (2007). Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences. *Journal of Biological Chemistry*. http://doi.org/10.1074/jbc.M609349200

Zhu, J., Yamane, H., & Paul, W. E. (2010). Differentiation of Effector CD4 T Cell Populations. *Annual Review of Immunology*. http://doi.org/10.1146/annurev-immunol-030409-101212

Zhu, Q., Yang, H., Chen, W., Cao, W., Zhong, G., Jiao, P., … Chen, H. (2008). A Naturally Occurring Deletion in Its NS Gene Contributes to the Attenuation of an H5N1 Swine Influenza Virus in Chickens. *Journal of Virology*. http://doi.org/10.1128/JVI.00978-07

Zhu, Y., Wang, X., Forouzmand, E., Jeong, J., Qiao, F., Sowd, G. A., … Shi, Y. (2018). Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. *Molecular Cell*. http://doi.org/10.1016/j.molcel.2017.11.031

Zonta, E., Bittencourt, D., Samaan, S., Germann, S., Dutertre, M., & Auboeuf, D. (2013). The RNA helicase DDX5/p68 is a key factor promoting c-fos expression at different levels from transcription to mRNA export. *Nucleic Acids Research*. http://doi.org/10.1093/nar/gks1046