Host Gene Expression Change During Viral Stress



Cillian Brophy

Pembroke College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

July 2022

Abstract

Viral infection induces extensive changes in the host transcriptome. One crucial way in which the transcriptome is remodelled is via changes to host mRNA 3'-end processing, but how virus infection induces these changes, and to what extent, is poorly understood. Over 70% of human genes have the potential to be cleaved and polyadenylated at multiple sites, and differential usage of these sites can regulate gene expression by affecting transcript properties and by altering transcript coding capacity.

In this investigation, we used two strains of influenza A virus as models to study host alternative cleavage and polyadenylation (APA) changes induced by viral infection. Influenza A is an RNA virus that replicates in the nuclei of host cells, and has previously been reported to interact with host 3'-end processing factors. We utilised a powerful experimental approach, combining 3'RNA-sequencing with a custom-built bioinformatic pipeline, to analyse APA changes in the nuclei and cytoplasms of host cells during viral infections.

We identified widespread APA changes in the 3'-untranslated and coding regions of mRNAs in viral-infected cells. We also extracted many physiologically relevant APA changes which have the potential to influence viral replication, suggesting that APA is another layer of gene regulation defined by competition between the host and virus. Additionally, we examined the mechanistic basis for APA during viral infection by analysing cis-element sequences surrounding poly(A) sites and uncovered highly significant frequency biases, suggesting that cis-element differences play a critical role in viral-induced differential poly(A) site selection.

Overall, this thesis characterises host gene expression changes during influenza

infection, with further exploration of differences in host poly(A) site usage. It identifies APA changes which are likely to be of physiological importance for viral propagation, and it examines the underlying mechanisms directing APA in an infection context.

Declaration

I declare that this thesis is entirely my own work, and except where otherwise stated, describes my own research.

Cillian Brophy, July 2022.

Acknowledgements

I would first like to thank Andre Furger for all of his guidance through the years. His enthusiasm for science is contagious and energised my own passion whenever I needed it. I am also grateful to Ervin Fodor, who always offered valuable wisdom and support. I would also like to thank all members of the Furger and Fodor groups who I have had the pleasure to work with. In particular, I would like to single out Mark Hickling and Harriet Lester - they have not only been immense academic supports but have also become great friends. Thank you also to the late Percival Stanion and his family, and to the University of Oxford, whose financial support enabled my study.

I have greatly valued my time in Pembroke College and Trinity College Dublin, and am grateful to all of the friends I have made along the way. Special mentions go to Constantin Schneider and Allie Hexley who are inspirational people and true friends. Patrick and Brittany Duma too, who I feel very lucky to have met. Linea and Matt, I'm so glad to be able to count you as great friends.

Thank you to my family who have always politely pretended to follow along when they have asked what it is I'm doing anyway. I particularly have deep gratitude for my grandparents - Kakas and Tony have helped me get to where I am today, and Jimmy was one of my biggest fans.

Finally, Hannah has been a massive positive force in my life and I am so glad we have been able to support each other. You are my best friend and I would not have wanted to be cooped up during a pandemic and while writing a thesis with anyone else.

Contents

1 Introduction **24**

1.1 General Introduction 24

1.2 Eurkaryotic pre-mRNA processing 25

1.2.1 5' Capping 26

1.2.2 Splicing 27

1.2.3 Cleavage and Polyadenylation 30

1.3 Alternative Cleavage and Polyadenylation 40

1.3.1 UTR-APA 41

1.3.2 CR-APA 49

1.3.3 APA regulation 51

1.3.4 APA in Health and Disease 62

1.3.5 Techniques used to study APA 67

1.4 Influenza 72

1.4.1 Introduction to Influenza A 72

1.4.2 Life Cycle of Influenza A 77

1.4.3 Interplay between IAV and host systems 83

1.5 Thesis objective and outline 88

2 Materials and Methods **90**

2.1 Cells . . . 90

2.2 Cell culture . . . 90

2.3 Influenza virus . . . 91

2.3.1 Harvest of influenza A/Udorn/72 (Udorn) and influenza A/WSN/33 (WSN) . . . 91

2.3.2 Purification of influenza A virus . . . 91

2.3.3 Plaque assay . . . 92

2.3.4 Infection of mammalian cells with influenza A . . . 92

2.3.5 Immunofluorescent microscopy of virus-infected cells . . . 93

2.4 Subcellular RNA fractionation . . . 94

2.5 Denaturing agarose gel electrophoresis (RNA) . . . 95

2.6 RNA Sequencing . . . 95

2.6.1 QuantSeq 3'mRNA-Seq Library Preparation . . . 95

2.6.2 High-sensitivity DNA (hsDNA) Bioanalyser Assay . . . 96

2.6.3 RNA-seq on the Ion Proton Platform . . . 96

2.7 Bioinformatics . . . 97

2.7.1 APA and gene expression pipeline . . . 97

2.7.2 PAS signal and PAS type analyses . . . 102

2.7.3 K-mer analyses . . . 102

2.7.4 aUTR motif discovery . . . 104

2.7.5 Gene ontology analyses . . . 104

2.7.6 Venn diagrams . . . 105

3 Characterising the host alternative cleavage and polyadenylation

landscape in influenza A H3N2 Udorn-infected cells 106

3.1 Introduction 106

3.1.1 Aims 108

3.2 Results 108

3.2.1 Design of the Udorn infection experiment 108

3.2.2 Establishing a bioinformatic pipeline to analyse alternative cleavage and polyadenylation and differential gene expression 115

3.2.3 3'UTR APA changes are induced in Udorn-infected cells . . 115

3.2.4 Genes regulated by 3' UTR APA are highly enriched for mRNA processing in Udorn-infected cells 121

3.2.5 Differential gene expression changes are induced in Udorn-infected cells 124

3.2.6 Coding-region APA changes are induced in Udorn-infected cells 131

3.3 Discussion 140

3.4 Summary 146

4 Characterising the host alternative cleavage and polyadenylation landscape in influenza A H1N1 WSN-infected cells 147

4.1 Introduction 147

4.2 Results 150

4.2.1 WSN infection induces widespread 3'UTR APA changes . . 150

4.2.2 Genes regulated by 3'UTR APA are highly enriched for mRNA processing in WSN-infected cells 158

4.2.3 aUTR motif analysis of lengthening nuclear transcripts . . . 162

4.2.4 Differential gene expression changes are induced in WSN-infected cells . . . 166
4.2.5 Interplay between host 3'UTR APA and differential gene expression in WSN-infected cells . . . 170
4.2.6 Coding-region APA changes in WSN-infected cells . . . 175
4.3 Discussion . . . 188

5 Investigating the characteristics of differentially regulated poly(A) sites in influenza-infected cells 197
5.1 Introduction . . . 197
5.1.1 Aims . . . 198
5.2 Results . . . 198
5.2.1 A-rich poly(A) sites are enriched in upregulated PASs during flu infection . . . 198
5.2.2 Distal-most PASs are less frequently used in genes regulated by APA during flu infection . . . 200
5.2.3 Differences in cis-elements underlie differential PAS usage during flu infection . . . 207
5.3 Discussion . . . 221
5.4 Summary . . . 224

6 Discussion and Future Directions 225
6.1 Final discussions, future directions and conclusions . . . 225
6.1.1 Thesis summary . . . 225
6.1.2 Widespread host gene expression changes are induced during IAV infection . . . 226

6.1.3 Mechanisms underpinning APA during IAV infection 228
6.1.4 Concluding remarks 230

A Appendix **232**

A.1 Full lists of genes experiencing a significant PAS switch during Udorn infection (Nucleus) 232
A.2 Full lists of genes experiencing a significant PAS switch during Udorn infection (Cytoplasm) 238
A.3 Gene ontology analyses of genes experiencing a 3'UTR PAS switch in in Udorn-infected cells in the "biological process" category (Nuclear) 244
A.4 Gene ontology analyses of genes experiencing a 3'UTR PAS switch in in Udorn-infected cells in the "biological process" category (Cytoplasm) 246
A.5 Top 100 differentially expressed genes in Udorn-infected cells 248
A.6 Gene ontology analyses of differentially expressed genes in Udorn-infected cells 252
A.7 Top 50 most significant genes undergoing 3'UTR APA switches (Nucleus, WSN) 257
A.8 Top 50 most significant genes undergoing 3'UTR APA switches (Cytoplasm, WSN) 262
A.9 Overlaps of genes regulated by APA in the Udorn and WSN datasets 267
A.10 Gene ontology analyses of genes experiencing a 3'UTR PAS switch in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Nuclear) 270

A.11 Gene ontology analyses of genes experiencing a 3'UTR PAS switch in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Cytoplasm) 273
A.12 Venn diagrams depicting differential gene expression in WSN-infected cells. 275
A.13 Gene ontology analyses of differentially expressed genes in WSN-infected cells 276
A.14 Gene ontology and BioPlanet pathway analyses of genes experiencing a CR-UTR PAS switch in WSN-infected cells (Cytoplasmic, any timepoint) 280
A.15 Usage of each PAS signal for PASs regulated by APA in the nucleus during WSN infection: 6hpi, 9hpi and 12hpi 284
A.16 Location of PAS within gene body for PASs regulated by APA in the nucleus during WSN infection: 6hpi, 9hpi and 12hpi 287
A.17 K-mers (4- and 6-mers) enriched or depleted in PASs upregulated and downregulated during WSN infection (Nucleus): 6hpi, 9hpi and 12hpi 290

List of Tables

1.1 Core cleavage and polyadenylation factors 39

1.2 Classical influenza A proteins and key functions 76

3.1 Table of the top 20 most significant genes undergoing 3'UTR APA switches (Nucleus, Udorn) 119

3.2 Table of the top 20 most significant genes undergoing 3'UTR APA switches (Cytoplasm, Udorn) 120

3.3 Table of GO biological processes for upregulated genes in cytoplasm which were also undergoing APA in Udorn-infected cells 130

3.4 Gene ontology analysis of lengthening CR-UTR APA changes in Udorn-infected cells (Cytoplasmic) 136

3.5 BioPlanet pathway analysis of lengthening CR-UTR APA changes in Udorn-infected cells (Cytoplasmic) 137

3.6 Gene ontology analysis of shortening CR-UTR APA changes in Udorn-infected cells (Cytoplasmic) 138

3.7 BioPlanet pathway analysis of shortening CR-UTR APA changes in Udorn-infected cells (Cytoplasmic) 139

4.1 Table exemplifying relationship between 3'UTR APA and differential transcript levels (WSN 18hpi) 174

4.2 Table listing genes included in the influenza-related enriched pathways for coding-region APA in WSN-infected cells (cytoplasmic shortening) 181

4.3 Table showing deltaRA values for ZC3HAV1, DHX30, AIMP2 and IFNAR2 184

A.1 Table of all genes undergoing significant 3'UTR APA switches (Nuclear, Udorn) 237

A.2 Table of all genes undergoing significant 3'UTR APA switches (Cytoplasm, Udorn) 243

A.3 Gene ontology analysis of lengthening 3'UTR APA events detected in Udorn-infected cells in the "biological process" category (Nuclear) 244

A.4 Gene ontology analysis of shortening 3'UTR APA events detected in Udorn-infected cells in the "biological process" category (Nuclear) 245

A.5 Gene ontology analysis of lengthening 3'UTR APA events detected in Udorn-infected cells in the "biological process" category (Cytoplasmic) 246

A.6 Gene ontology analysis of shortening 3'UTR APA events detected in Udorn-infected cells in the "biological process" category (Cytoplasmic) 247

A.7 Top 100 most significant differentially regulated genes in Udorn-infected cells (Nuclear) 250

A.8 Top 100 most significant differentially regulated genes in Udorn-infected cells (Cytoplasmic) 252

A.9 Gene ontology analysis of significantly upregulated genes in Udorn-infected cells in the "biological process" category (Nuclear) 253

A.10 Gene ontology analysis of significantly downregulated genes in Udorn-infected cells in the "biological process" category (Nuclear) 254

A.11 Gene ontology analysis of significantly upregulated genes in Udorn-infected cells in the "biological process" category (Cytoplasmic) . . 256

A.12 Gene ontology analysis of significantly downregulated genes in Udorn-infected cells in the "biological process" category (Cytoplasmic) . . 257

A.13 Table of the top 50 most significant genes undergoing 3'UTR APA switches (Nuclear, WSN, 6hpi) 259

A.14 Table of the top 50 most significant genes undergoing 3'UTR APA switches (Nuclear, WSN, 9hpi) 260

A.15 Table of the top 50 most significant genes undergoing 3'UTR APA switches (Nuclear, WSN, 12hpi) 261

A.16 Table of the top 50 most significant genes undergoing 3'UTR APA switches (Nuclear, WSN, 18hpi) 262

A.17 Table of the most significant genes undergoing 3'UTR APA switches (Cytoplasmic, WSN, 6hpi) 263

A.18 Table of the top 50 most significant genes undergoing 3'UTR APA switches (Cytoplasmic, WSN, 9hpi) 264

A.19 Table of the top 50 most significant genes undergoing 3'UTR APA switches (Cytoplasmic, WSN, 12hpi) 265

A.20 Table of the top 50 most significant genes undergoing 3'UTR APA switches (Cytoplasmic, WSN, 18hpi) 267

A.21 Gene ontology analysis of genes lengthened by 3'UTR APA in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Nuclear) 271

A.22 Gene ontology analysis of genes shortened by 3'UTR APA in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Nuclear) 272

A.23 Gene ontology analysis of genes lengthened by 3'UTR APA in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Cytoplasm) 273

A.24 Gene ontology analysis of genes shortened by 3'UTR APA in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Cytoplasm) 274

A.25 Gene ontology analysis of significantly upregulated genes in WSN-infected cells in the "biological process" category (Nuclear) 276

A.26 Gene ontology analysis of significantly downregulated genes in WSN-infected cells in the "biological process" category (Nuclear) 277

A.27 Gene ontology analysis of significantly upregulated genes in WSN-infected cells in the "biological process" category (Cytoplasmic) . . 278

A.28 Gene ontology analysis of significantly downregulated genes in WSN-infected cells in the "biological process" category (Cytoplasmic) . . 279

A.29 Gene ontology analysis of lengthening CR-UTR APA changes in WSN-infected cells (Cytoplasmic) 280

A.30 BioPlanet pathway analysis of shortening CR-UTR APA changes in WSN-infected cells (Cytoplasmic) 281

A.31 Gene ontology analysis of shortening CR-UTR APA changes in WSN-infected cells (Cytoplasmic) 282

A.32 BioPlanet pathway nalysis of shortening CR-UTR APA changes in WSN-infected cells (Cytoplasmic) 283

List of Figures

1.1 Co-transcriptional pre-mRNA processing 26
1.2 Core cleavage and polyadenylation machinery 32
1.3 Diagrammatic overview of UTR-APA and CR-APA 42
1.4 Coding region APA 50
1.5 Schematic of an influenza virion 75
1.6 Influenza viral life cycle 83

2.1 Bioinformatic pipeline used to detect APA and differential gene expression 99
2.2 PAS groupings for APA analysis 101
2.3 Schematic showing parameters of the K-mer analysis of sequences surrounding PASs 103

3.1 Experimental approach used for the IAV Udorn infection 112
3.2 Udorn plaque assay and RNA gel 113
3.3 IAV (Udorn) infection efficiency of A549 cells 114
3.4 Significant 3'UTR APA events seen across the nuclei and cytoplasms of Udorn-infected cells 118

3.5 Gene ontology analyses of 3'UTR APA events detected in Udorn-infected cells in the "biological process" category 123

3.6 Differential gene expression changes observed in Udorn-infected cells 125

3.7 Gene ontology analyses of genes differentially expressed in Udorn-infected cells ("biological process" category) 128

3.8 Coding-region APA in Udorn-infected cells 133

4.1 Overview of 3'UTR APA changes within nuclear and cytoplasmic fractions during the WSN infection 153

4.2 All significant 3'UTR APA events across the WSN infection, in both cytoplasmic and nuclear samples 155

4.3 Overlaps of genes experiencing a 3'UTR APA change in WSN-infected cells 157

4.4 Gene ontology analyses of 3'UTR APA events detected in WSN-infected cells in the "biological process" category 161

4.5 Enriched aUTR motifs in distal APA isoforms in WSN-infected cells, nucleus, 12hpi 165

4.6 Barcharts of differential gene expression changes in WSN-infected cells. 167

4.7 Gene ontology analyses of genes differentially expressed in WSN-infected cells ("biological process" category) 169

4.8 Coding-region APA events in WSN-infected cells. 177

4.9 Gene ontology analyses of genes that switch PASs between coding region and the 3'UTR in WSN-infected cells ("biological process" and "pathways" category) 179

4.10 Plots showing differential CR-UTR PAS usage in WSN-infected cells for ZC3HAV1, DHX30, AIMP2 and IFNAR2 187

5.1 Relative amounts in the nucleus of each signal sequence in PASs regulated by APA (Top) and those not (Bottom), Udorn infection . 201

5.2 Usage of each PAS signal for PASs regulated by APA in the nucleus (Top) and those not (Bottom), 18hpi (WSN infection) 202

5.3 Location of PAS within the gene body of those regulated by APA (Top) and those not (Bottom) in the nucleus, Udorn infection . . . 205

5.4 Location of PAS within gene body for PASs regulated by APA in the nucleus (Top) and those not (Bottom), 18hpi (WSN infection) . 206

5.5 Numbers of significantly enriched or depleted 4-mers in PASs regulated by APA during Udorn infection (Nucleus) 209

5.6 Unique 4-mers or 6-mers with significantly biased frequency at each timepoint throughout the WSN-infection (Nucleus), "All" APA . . 210

5.7 Numbers of significantly enriched or depleted 4-mers in PASs regulated by APA during WSN infection, 18hpi (Nucleus) 212

5.8 Top 50 most common enriched and depleted 4-mers (Top 50) in PASs regulated by APA during Udorn infection (Nucleus) 217

5.9 Top 50 most common enriched and depleted 6-mers (Top 50) in PASs regulated by APA during Udorn infection (Nucleus) 218

5.10 Top 50 most common enriched and depleted 4-mers (Top 50) in PASs regulated by APA during WSN infection, 18 hours post-infection (Nucleus) 219

5.11 Top 50 most common enriched and depleted 6-mers (Top 50) in PASs regulated by APA during WSN infection, 18 hours post-infection (Nucleus) 220

A.1 Overlaps of genes experiencing a 3'UTR PAS switch in the Udorn and WSN datasets 268

A.2 Overlaps of genes regulated by CR-UTR APA in the Udorn and WSN datasets 269

A.3 Overlaps of differentially expressed genes in WSN-infected cells . . . 275

A.4 Usage of each PAS signal for PASs regulated by APA in the nucleus (Top) and those not (Bottom), 6hpi (WSN infection) 284

A.5 Usage of each PAS signal for PASs regulated by APA in the nucleus (Top) and those not (Bottom), 9hpi (WSN infection) 285

A.6 Usage of each PAS signal for PASs regulated by APA in the nucleus (Top) and those not (Bottom), 12hpi (WSN infection) 286

A.7 Location of PAS within gene body for PASs regulated by APA in the nucleus (Top) and those not (Bottom), 6hpi (WSN infection) . . 287

A.8 Location of PAS within gene body for PASs regulated by APA in the nucleus (Top) and those not (Bottom), 9hpi (WSN infection) . . 288

A.9 Location of PAS within gene body for PASs regulated by APA in the nucleus (Top) and those not (Bottom), 12hpi (WSN infection) . 289

A.10 Top 50 most common enriched and depleted 4-mers (Top 50) in PASs regulated by APA during WSN infection, 6 hours post-infection (Nucleus) 291

A.11 Top 50 most common enriched and depleted 6-mers (Top 50) in PASs regulated by APA during WSN infection, 6 hours post-infection (Nucleus) 292

A.12 Top 50 most common enriched and depleted 4-mers (Top 50) in PASs regulated by APA during WSN infection, 9 hours post-infection (Nucleus) 293

A.13 Top 50 most common enriched and depleted 6-mers (Top 50) in PASs regulated by APA during WSN infection, 9 hours post-infection (Nucleus) 294

A.14 Top 50 most common enriched and depleted 4-mers (Top 50) in PASs regulated by APA during WSN infection, 12 hours post-infection (Nucleus) 295

A.15 Top 50 most common enriched and depleted 6-mers (Top 50) in PASs regulated by APA during WSN infection, 12 hours post-infection (Nucleus) 296

Abbreviations

3'UTR	3' untranslated region
4sU-seq	4-thiouridine RNA sequencing
APA	Alternative cleavage and polyadenylation
ARE	AU-rich element
ATP	Adenosine triphosphate
aUTR	Alternative UTR
CBC	Cap-binding complex
cDNA	Complementary DNA
CDS	Coding sequence
CFI	Cleavage factor I
CFII	Cleavage factor II
CPA	Cleavage and polyadenylation
CPSF	Cleavage and polyadenylation factor
CR-APA	Coding region APA
cRNA	Complementary RNA
CSTF	Cleavage stimulation factor
CTD	Carboxy-terminal domain
cUTR	Constitutive UTR
DAPI	4',6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
DSE	Downstream elements
EJC	Exon junction complex

ER	Endoplasmic reticulum
EST	Expressed sequenced tags
FluPol	Influenza A polymerase
GO	Gene ontology
GTP	Guanosine triphosphate
HPI	Hours post-infection
HPV	Human papillomavirus
HSV-1	Herpes simplex virus 1
IAV	Influenza A virus
IgM	Immunoglobulin M
miRNA	MicroRNA
MOI	Multiplicity of infection
mRNA	Messenger RNA
ncRNA	Non-coding RNA
NP	Nucleoprotein
NS1	Non-structural protein 1
PABPN1	Nuclear poly(A) binding protein I
PAF	PolII-associated factor
PAP	Poly(A) polymerase
PAS	Polyadenylation site
PEDV	Porcine epidemic diarrhea virus
PolII	RNA polymerase II
Pre-mRNA	Pre-cursor mRNA

RBP	RNA-binding protein
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
rRNA	Ribosomal RNA
SC	Satellite cell
scRNA-seq	Single-cell RNA sequencing
siRNA	Small interfering RNA
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein
TLI	Tandem UTR length index
tRNA	Transfer RNA
Udorn	H3N2 influenza A/Udorn/72
UTR	Untranslated region
UTR-APA	Untranslated region APA
vRNA	Viral RNA
vRNP	Viral ribonucleoprotein
VSV	Vesicular stomatitis virus
WSN	H1N1 influenza A/WSN/33

Chapter 1

Introduction

1.1 General Introduction

The human genome consists of about 20,000 protein-coding genes, and thousands more non-coding RNA genes (Salzberg 2018). Expression of these genes is tightly regulated and modulated at several different points along the journey from genotype to phenotype. Whether or not this multifaceted process is suitably controlled can ultimately decide the fate of an organism.

Robert Koch discovered over a hundred years ago that pathogens cause infections, and it has become clear with decades of research since, that host-pathogen interactions modulate gene expression - although the precise effects of this remain unclear in many cases. Viruses are classic examples of pathogenic agents that interact with and manipulate host gene expression (Glaunsinger and Ganem 2004; López-Denman et al. 2021; Lyles 2000). They rely on the host for their ability to propagate and have evolved numerous ways to hijack host cellular machinery to ensure that their viral proteins are synthesised and packaged. Virus infection is a

highly dynamic process, and some viruses have genomes or proteins that enter the nucleus at some stage during their life cycle (Cohen, Au, and Panté 2011). One such virus, the influenza virus, replicates entirely within the nucleus, where it can directly interfere with delicately balanced host nuclear gene expression processes.

Today, viruses and their public health implications have been starkly brought to the forefront in the wake of the COVID-19 pandemic. Future pandemics, including an influenza pandemic, are a certainty, and it is therefore crucial that we understand the mechanisms behind virus-host interactions so that targeted therapies can be appropriately designed.

1.2 Eurkaryotic pre-mRNA processing

One of the major steps along the pathway from gene activation to protein production is the generation of a nascent pre-mRNA molecule from a DNA template by RNA polymerase II, a process known as transcription. In eukaryotes, this transcript must undergo additional processing before it is considered to be a mature mRNA that is ready to be exported to the cytoplasm and decoded into a functional protein. Three principle co-transcriptional modifications must be made to the pre-mRNA molecule before it is ready to be exported out of the nucleus and decoded by ribosomes, and this involves the spatial and temporal co-ordination of multiple complex reactions (Bentley 2014). These consist of the addition of a 5' cap, the splicing of introns, and the generation of a poly(A) tail at the 3' end (Proudfoot, Furger, and Dye 2002) (Figure 1.1).

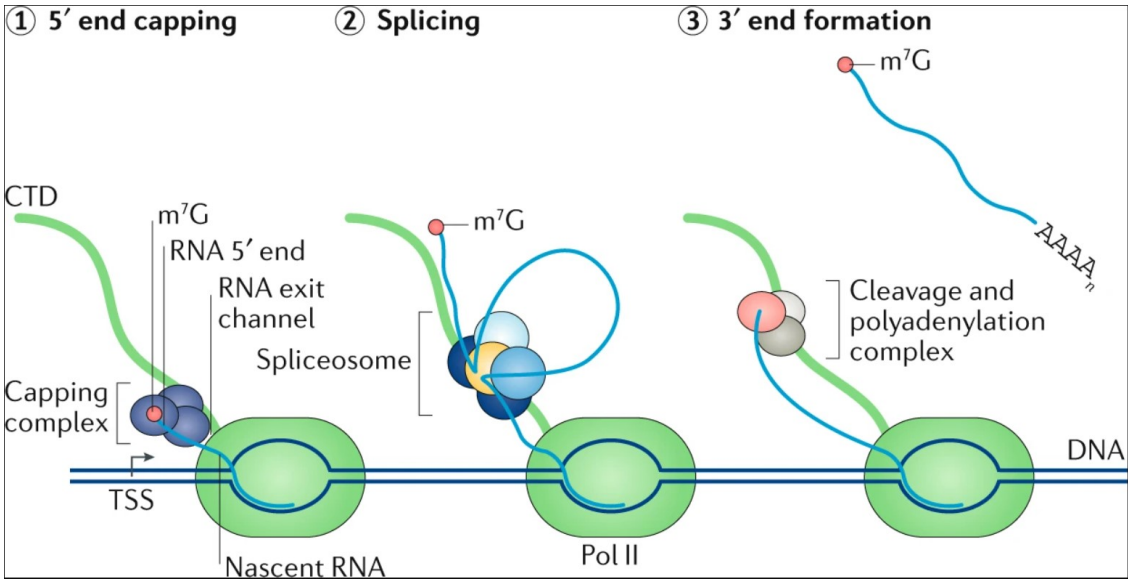


Figure 1.1: Co-transcriptional pre-mRNA processing. RNA polymerase begins transcription at the transcription start site (TSS). The 5'end of the nascent RNA is converted to an N7-methylated guanosine (m^7G) cap structure (1). During the process of transcription elongation, introns are excised by the spliceosome (2). Finally, as the end of the gene is transcribed by RNA polymerase, the emerging transcript is cleaved and polyadenylated by the cleavage and polyadenylation complex (3). Figure taken from Desterro, Bak-Gordon, and Carmo-Fonseca 2020.

1.2.1 5' Capping

This is the first of the pre-mRNA processing reactions and involves the acquisition of a cap structure to the 5' end of the new molecule (Figure 1.1). This occurs over the course of three individual reactions. To begin, the first triphosphate of the nascent transcript is hydrolysed to a diphosphate by an RNA 5'-triphosphatase. Next, a guanylyltransferase transfers a guanosine monophosphate from a donor GTP to the new diphosphate, producing a guanosine cap. Finally, a guanine-7-methyltransferase methylates this guanosine group at the N7 position (Ramanathan, Robb, and Chan 2016; Moore and Proudfoot 2009). This cap structure - at this point an N7-methylated guanosine - serves several critical functions.

The cap is required for mRNA stability - it physically shields the nascent transcript from degrading exoribonuclease activity (Liu and Kiledjian 2006) and is essential for cap-dependent protein synthesis as it directs the translational machinery to the 5' end of the RNA molecule (Gingras, Raught, and Sonenberg 1999). The binding of cap-binding complex (CBC) allows assembly of the spliceosome and therefore appropriate mRNA splicing to commence (Ohno, Sakamoto, and Shimura 1987; Fresco and Buratowski 1996; Rambout and Maquat 2020).

1.2.2 Splicing

The discovery of splicing in 1977 greatly expanded our understanding of the complexity of the genome (Berget, Moore, and Sharp 1977; Chow et al. 1977). Splicing is the process by which noncoding intron sequences are removed from a nascent transcript and its coding exons are fused (Figure 1.1). With approximately eight introns per human gene (Sakharkar, Chow, and Kangueane 2004), splicing is a well-orchestrated and efficient process that ensures that the new mRNA molecule contains a continuous open reading frame which can be translated into protein. Most splicing of nascent pre-mRNA occurs during transcriptional elongation, although some examples of post-transcriptional splicing have been recorded (Brugiolo, Herzel, and Neugebauer 2013; Jia, Long, et al. 2020).

Spliceosome

The large ribonucleoprotein spliceosome complex consists of 5 U-rich small nuclear RNAs (snRNAs) - U1, U2, U4, U5 and U6 - and is assembled from approximately 100 core proteins (Wilkinson, Charenton, and Nagai 2020). Conserved sequence

elements within the intron recruit the snRNPs (small nuclear ribonucleoproteins) until the greater macromolecular spliceosome complex is assembled (Frendewey and Keller 1985). The snRNPs coordinate base pairing between different snRNAs, and with the pre-mRNA, leading to the formation of secondary and tertiary structures which form the catalytic spliceosome core (Herzel et al. 2017). Splicing is a two-step transesterification reaction. First, the 5' exon-intron border splice site undergoes a nucleophilic attack by the 2' hydroxyl of an adenosine at the internal branch point sequence of the intron. This releases a free 5' exon, and a lariat-shaped molecule containing the intron and 3' exon. Next, the 3' hydroxyl of the now-freed 5' exon attacks the 3' intron-exon border, yielding one molecule containing the joined exons, and the remaining lariat-shaped intron (Proudfoot, Furger, and Dye 2002; Kastner et al. 2019).

Exon Junction Complex

Another multisubunit protein complex, the exon junction complex (EJC), binds to the transcript during pre-mRNA splicing near the exon-exon junction (Schlautmann and Gehring 2020). The EJC stays associated with the mRNA during its entire lifetime, and has important post-transcriptional implications for the metabolism of the transcript, including cytoplasmic localisation, and translational regulation (Woodward et al. 2017).

The EJC is also an instrumental component of the "canonical" nonsense-mediated degradation (NMD) pathway. After excision of introns by splicing, the EJC remains associated to the transcript close to the splice site (Supek, Lehner, and Lindeboom 2021). If a nonsense mutation has generated a premature termination codon on the transcript, the bound EJC will signal that the transcript is

defective during translation and trigger decay via the NMD pathway (Baudu et al. 2021). Since up to 10% of human mRNAs are regulated by NMD, this surveillance pathway plays a major role in the global regulation of mRNA expression (Metze et al. 2013).

Alternative Splicing

Most human genes have the potential to be spliced in different ways, leading to the generation of alternate isoforms and massively increasing genetic diversity by expanding the proteome without increasing the number of genes (Montes et al. 2019; Ule and Blencowe 2019). There are several different types of alternative splicing. These include cassette exon inclusion or skipping, usage of alternative 5' and 3' splice sites, intron retention (an intron that is normally excluded is kept in the final transcript), and mutually exclusive exons (exons that are never retained together in the same isoform) (Bhadra et al. 2020).

Splice site usage is substantially altered through the interaction of auxiliary splicing factors such as SR-family and HNRNP proteins (Cherry and Lynch 2020). These proteins bind short sequence motifs, and function in a combinatorial manner - even slight changes in abundance or activity of these RNA-binding proteins can result in major changes to the splicing fate of a transcript (Cherry and Lynch 2020). In this way, alternative splicing is thought to partly mediate an organism's response to stress, including that induced by heat, UV exposure or infections (Taylor and Sobczak 2020). SR-family proteins typically contribute a splice-enhancing effect to this complex interplay of RBPs, while HNRNPs commonly repress splicing, although these effects are often context-dependent (Stanley and Abdel-Wahab 2022).

1.2.3 Cleavage and Polyadenylation

Cleavage and polyadenylation (CPA) is the co-transcriptional pre-mRNA processing step that matures the 3' end of the emerging transcript (Figure 1.1) (Fischl et al. 2019). All protein-coding mRNAs, with the exception of replication-dependent histone transcripts, undergo endonucleolytic cleavage at the 3' end and the synthesis of a poly(A) tail, consisting of around 200 adenosine residues (Proudfoot, Furger, and Dye 2002; Colgan and Manley 1997; Eisen et al. 2020). The two CPA reactions, cleavage and subsequent polyadenylation, are tightly linked to transcription termination (Richard and Manley 2009). Functionally, CPA prepares the mRNA transcript for nuclear export and subsequent translation (Eckner, Ellmeier, and Birnstiel 1991). The addition of a poly(A) tail protects the newly transcribed RNA from 3' exonuclease degradation, thereby enhancing mRNA stability (Marzluff, Wagner, and Duronio 2008), while also promoting translation (Passmore and Coller 2021).

Cleavage and Polyadenylation Sequence Elements

The CPA site, also known as the polyadenylation site (PAS), is defined by cis-elements found both upstream and downstream. In metazoans, U-rich and UGUA elements are found 10-30nt upstream of the PAS. The poly(A) signal is also found here, which is comprised of a A[A/U]UAAA hexamer or a close variant thereof (Tian and Manley 2017; Tian, Hu, et al. 2005). U- and GU-rich elements are found 20-40nt downstream of the PAS (Tian, Hu, et al. 2005). While the upstream poly(A) signal can vary, it is usually ($\sim$75 %) the canonical hexamer AAUAAA or a similar variant (Beaudoing et al. 2000; Sun, Hamilton, and Tong 2020). No

clear consensus has emerged on what the exact nucleotide composition of the downstream elements (DSE) take, but it is commonly a GU-rich sequence followed by several uridines (Zarudnaya et al. 2003; Hu, Lutz, et al. 2005). The strength of a PAS is dependent on the combinatorial effect of its cis-elements. Therefore, PASs with weak or non-canonical sequences (for example A-rich sequences) often will have potent downstream U-rich elements (Jan et al. 2011; Nunes et al. 2010). On the other hand, strong upstream elements in PASs can compensate for weaker downstream ones (Cheng, Miura, and Tian 2006).

Cleavage Site

In mammals, the actual point of cleavage for the pre-mRNA lies directly after what is often a CA dinucleotide, normally found 15-30 nucleotides downstream of the poly(A) signal and up to 20 nucleotides upstream of the DSE (Yeh and Yong 2016; Derti et al. 2012). The importance of this dinucleotide for some pre-mRNA has been shown in clinical studies (Gruber and Zavolan 2019). The F2 gene which encodes for prothrombin, a clotting factor, normally features a CG dinucleotide at its 3' cleavage site. A gain-of-function mutation at this location, which changes the CG dinucleotide to the more widely found CA, was found to increase efficiency of CPA. This improved efficiency led to increased expression of *F2* mRNA and protein levels, and thus elevated prothrombin in plasma. This mutation (G20210A) is found in 1-2% of the general population and is a common risk factor for thromboembolism (Gehring et al. 2001; Ceelie et al. 2004). In this example, the length of the poly(A) tail remained the same between the wild type and mutated variants, highlighting the important role that nucleotide composition at the 3' cleavage site plays in CPA efficiency and consequently human health

(Gehring et al. 2001).

Core Cleavage and Polyadenylation Machinery

In metazoans, the polyadenylation machinery consists of more than 80 proteins, with over 20 of these considered as core proteins (Neve, Patel, et al. 2017). Essential components of this machinery include four multi-subunit factors: the cleavage and polydenylation factor (CPSF), the cleavage stimulation factor (CSTF), cleavage factor I (CFI) and cleavage factor II (CFII). Other critical factors include Symplekin, Poly(A) Polymerase (PAP), nuclear poly(A) binding protein 1 (PABPN1) and RNA Polymerase II (Pol II) (Figure 1.2). Table 1.1 summarises the major factors involved in the CPA process.

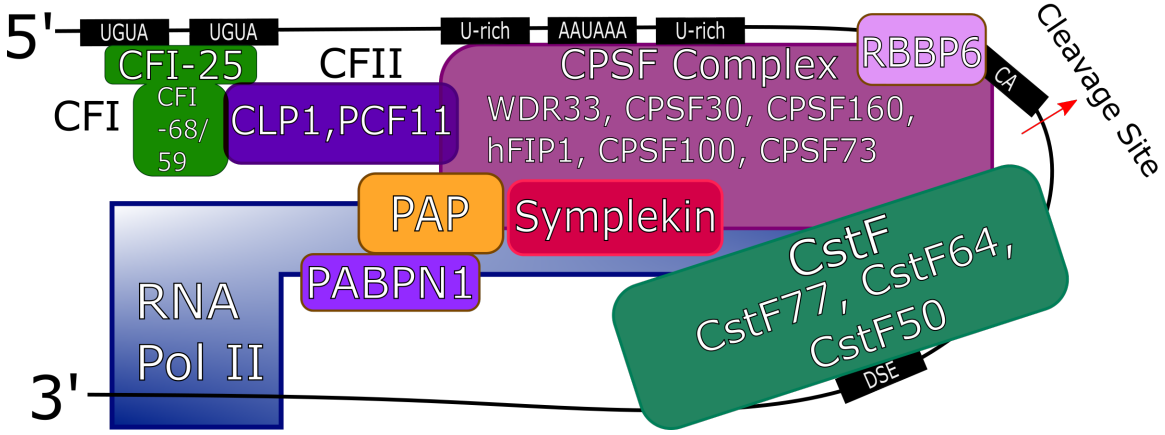


Figure 1.2: Core cleavage and polyadenylation machinery. The black line depicts a pre-RNA, from 5' to 3', and the black boxes represent cis-elements. AAUAAA represents the poly(A) signal (PAS) or its variants, while the UGUA and U-rich upstream elements and downstream elements (DSE) are also shown. The major components of the CPA machinery are made up of multi-protein complexes, including the cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulation factor (CSTF), cleavage factor I (CFI) and cleavage factor II (CFII). Different complexes are assigned unique colours in the diagram.

The **CPSF** complex consists of WDR33, CPSF30, CPSF160, FIP1, CPSF100 and CPSF73. Originally, it was believed that CPSF160 was the protein responsible for recognising the poly(A) signal, principally because it cross-linked to AAUAAA-containing RNA (Keller, Bienroth, et al. 1991). However, PAR-CLIP data showed that CPSF160 has a diffuse binding pattern along much of the 3' end of the pre-mRNA, with a shallow peak approximately 70 nucleotides upstream of the cleavage site, casting doubt over its role in the recognition of the polyadenylation signal (Martin, Gruber, et al. 2012). Recent structural data instead suggests that CPSF160 acts as a scaffold for the assembly of WDR33 and CPSF30, which directly contact the RNA and recognise the signal (Zhang, Sun, Shi, et al. 2020; Chan, Huppertz, et al. 2014). While the exact nature of this binding interaction is not fully understood, it was recently shown that both WDR33 and CPSF30 are required together to achieve high-affinity binding to the poly(A) signal, while either protein in complex alone with CPSF160 leads to much lower binding efficiency (Hamilton, Sun, and Tong 2019). The importance of CPSF30 in appropriate 3' end processing is highlighted by its interaction with the influenza non-structural protein 1 (NS1). This physical association inhibits the ability of CPSF30 to bind the poly(A) signal sequence and consequently causes aberrant gene expression (Nemeroff et al. 1998; Twu et al. 2006; Chan, Huppertz, et al. 2014; Sun, Zhang, et al. 2018). CPSF30 also interacts with FIP1, another member of the CPSF complex, through two zinc fingers (ZF4-ZF5) (Barabino, Ohnacker, and Keller 2000; Hamilton and Tong 2020). One FIP1 protein binds to each of the ZF4-ZF5 fingers, meaning that each CPSF30-FIP1 complex consists of one CPSF30 protein and two FIP1 molecules. FIP1 binds the U-rich sequences on either side of the poly(A) signal, and its most important known role is to recruit PAP (Neve, Patel,

et al. 2017), although it is also believed to be involved in PAS selection by interactions with the CPSF and CFI complexes (Zhu, Wang, et al. 2018). Kumar et al. showed in a recent structural study that the yeast FIP1 contains intrinsically disordered regions which remain highly dynamic even when FIP1 is in complex with the yeast CPSF orthologue, and this flexibility is important for several functions of the core polyadenylation machinery (Kumar et al. 2021). CPSF73 is the CPSF subunit that handles endonucleolytic cleavage of the emerging transcript at the PAS. CPSF73's cleavage ability is highly dependent on CPSF complex formation, and only has a weak cleaving potency on its own (Mandel et al. 2006). This is a result of CPSF73 residing in a closed formation when not in association with other CPSF subunits, blocking access to the RNA substrate (Mandel et al. 2006). Aberrant CPSF73 activity has been correlated with cancers, including non-small cell lung cancer (Ning et al. 2019), prostate cancer (Zhu, Yu, et al. 2009) and colorectal cancer (Luo, Geng, et al. 2019). Interestingly, CPSF73 has been suggested as a therapeutic target for protozoan parasites, including African trypanosomes and *Plasmodium falciparum* (drivers of African sleeping sickness and malaria respectively). Compounds have been identified which inhibit growth of these parasites by targeting the protozoan homologues of CPSF73, but not host CPSF73 (Wall et al. 2018; Sonoiki et al. 2017; Liu and Moore 2021). CPSF100 shares some weak sequence homology with CPSF73 but its role in CPA is much less clear, though because it cannot bind zinc it lacks any endonuclease activity (Neve, Patel, et al. 2017; Mandel et al. 2006). A recent study has identified CPSF100 residues that when mutated, abolish appropriate CPSF formation, suggesting that CPSF100 is crucial for CPSF complex assembly (Zhang, Sun, Shi, et al. 2020).

Recent work also demonstrates a key role for **RBBP6**, a 200kDa multidomain

protein, which is required for activation of the endonuclease activity of CPSF (Boreikaite et al. 2022). Importantly, RBBP6 does not co-purify with CPSF, indicating that it is not a member of the core CPSF complex. However, pull-down assays suggest that RBBP6 is recruited to the PAS and interacts with CPSF in an RNA-dependent manner (Boreikaite et al. 2022). Co-expression assays of RBBP6 and CPSF73 show that the ubiquitin-like domain of RBBP6 interacts with the CPSF73 nuclease domain. In this way, the authors propose a model wherein the assembly of CPSF, CSTF, CFII and RBBP6 on pre-mRNA leads to a conformational change in CPSF73 which opens its active site (Boreikaite et al. 2022).

The **CSTF** complex was the first of the core CPA machinery to be characterised structurally (Takagaki, Manley, et al. 1990). Today, it is known to be hexameric, containing two copies of each subunit, and considered essential for cleavage but not polyadenylation (Yang, Hsu, et al. 2018; Takagaki, Ryner, and Manley 1989). The complex binds the U- and GU-rich downstream elements via its CSTF64 subunit (MacDonald, Wilusz, and Shenk 1994). This binding site is normally within approximately 20 nucleotides of the cleavage site (Yao, Biesinger, et al. 2012). Interestingly CSTF64 has a second isoform in humans, found only in male germ line cells, which may be important in spermatogenesis and tissue-specific polyadenylation (Wallace et al. 1999; Monarez, MacDonald, and Dass 2007).

CSTF77 binds tightly to the hinge domain of CSTF64, and acts as a bridge between CSTF64 and CSTF50 (Legrand et al. 2007). CSTF77 possesses a nuclear localisation signal and is believed to be responsible for transporting the CSTF complex into the nucleus (Grozdanov et al. 2018). CSTF77 creates a link between the polyadenylation signal and DSE elements, through an association with

CPSF160 (Bai et al. 2007). A role in polyadenylation regulation is supported by recent evidence showing that modifying CSTF77 protein levels changes the RNA binding affinity of CSTF64 and its nuclear import (Grozdanov et al. 2018).

CSTF50 is the least studied CSTF member and has no yeast homologue, unlike most of the core CPA components, raising questions about its function in metazoans. CSTF50 contains 7-repeats of the WD40 domain which are involved with its interactions with CSTF77 (Takagaki and Manley 2000). Some evidence shows that CSTF50 might play a role in the efficient selection of downstream elements by the CSTF complex, but its exact function remains unknown (Yang, Hsu, et al. 2018).

CFI is a heterotetramer, consisting of a homodimer of two CFI25 subunits, which bind directly to UGUA upstream elements, and two larger CFI68 (also known as CPSF6) or CFI59 subunits (Yang, Coseno, et al. 2011). Structural data shows that the RNA-recognition motif of CFI68 is crucial for the association with CFI25 and the complex's RNA-binding ability (Li, Tong, et al. 2011; Yang, Coseno, et al. 2011). CFI is instrumental in the binding of the CPSF complex to the pre-mRNA (Rüegsegger, Beyer, and Keller 1996), and has a critical role in recognising the polyadenylation signal in instances when it is not the A[A/U]UAAA hexamer (Venkataraman, Brown, and Gilmartin 2005).

The **CFII** complex is the least characterised of the 4 core CPA complexes and consists of just two subunits, PCF11 and CLP1 (Schäfer et al. 2018). PCF11 is critical in the regulation of Pol II-directed transcription termination through its role in the degradation of the 3' product after cleavage (West and Proudfoot 2008). CLP1 on the other hand is suggested to bridge the CPSF and CFI during cleavage (Vries et al. 2000). CFII RNA-binding is mediated by PCF11 zinc fingers, and

G-rich sequences appear to be preferred (Schäfer et al. 2018).

Symplekin was first identified through its interaction with CSTF64's hinge region (Takagaki and Manley 2000). It is believed to act as a scaffold protein, bridging CSTF and CPSF (Takagaki and Manley 2000), and is required for 3'-end cleavage (Xiang et al. 2010). Symplekin is also required for cytoplasmic polyadenylation to occur in maturing oocytes (Hofmann et al. 2002).

PAP is responsible for generating the poly(A) tail of the emerging transcript (Raabe, Bollum, and Manley 1991), while PABPN1 regulates the length of the tail (Kuhn et al. 2009). PAP interacts with other core 3'-processing subunits, but it does not require any of these proteins to carry out the polyadenylation reaction *in vitro* (Edmonds 2002). It is recruited to the PAS via an interaction with CPSF (Kaufmann et al. 2004), where it adds adenosine residues in a template-independent manner. **PAPBN1** binds the developing poly(A) tail and enhances PAP's RNA-binding affinity, mediating an interaction with the CPSF complex that stimulates PAP's polyadenylation activity (Kuhn et al. 2009). PAP becomes more processive, and is able to generate a poly(A) tail of 200-250 nucleotides in length before dissociating (Kerwitz et al. 2003). The exact mechanism behind polyadenylation termination is unclear. It has been suggested that once the growing poly(A) tail reaches a size that can no longer be accommodated by a spherical particle structure 21nm in diameter as seen with electron microscopy, the interaction between the CPSF-PAP-PABN1 complex is weakened and the polyadenylation reaction is terminated (Keller, Kuhn, et al. 2000). Despite newly-generated poly(A) tails being approximately 200 nucleotides in length, the median tail length of cellular mRNAs is in fact close to 100 nucleotides (Subtelny et al. 2014; Eisen et al. 2020). This effect is most likely caused by the activity of cytoplasmic deadenylases

(Chang et al. 2014; Passmore and Coller 2022).

The carboxy-terminal domain (**CTD**) of RNA polymerase II is required for cleavage, and is the key component that couples 3'-end processing to transcription (McCracken et al. 1997). Phosphorylation of serine residues in the CTD recruits CPA factors (along with the 5'-capping complex and spliceosome) (Ahn, Kim, and Buratowski 2004).

Factor	**Function**
CPSF	*Recognises and binds PAS*
WDR33	Binds PAS
CPSF30	PAS recognition
CPSF160	Scaffold for protein binding
FIP1	5'-3' exoribonuclease mRNA decay
CPSF100	CPSF assembly
CPSF73	Endonucleolytic cleavage of emerging transcript
CSTF	*Essential for cleavage but not polyadenylation*
CSTF64	Binds U/GU-rich downstream elements
CSTF77	Acts as bridge between CSTF64 and CSTF50
CSTF50	Possible role in downstream element recognition
CFI	*Mediates CPSF-RNA binding*
CFI25	Binds UGUA upstream elements
CFI68	Enhances CFI25 RNA-binding
CFI59	Can replace CFI68 in complex
CFII	*Important for termination & protein-protein interactions*
PCF11	Critical for PolII-termination
CLP1	Possibly bridges CPSF and CFI
Other	
Symplekin	Scaffold that bridges CSTF and CPSF
PAP	Generates poly(A) tail
PABPN1	Limits length of poly(A) tail
PolII CTD	Platform for CPA & couples CPA to transcription
RBBP6	Activates endonuclease activity of CPSF

Table 1.1: Core cleavage and polyadenylation factors

1.3 Alternative Cleavage and Polyadenylation

Over 70% of human genes contain more than one polyadenylation site (PAS) (Derti et al. 2012). In a process termed alternative cleavage and polyadenylation (APA), differential selection of these PASs by the 3'-end processing machinery can lead to distinct RNA transcripts being generated. APA sites can be found both in the 3' untranslated region (termed UTR-APA) and the upstream coding region of the transcript (termed CR-APA). A schematic representation of both UTR-APA and CR-APA can be seen in Figure 1.3.

APA was first identified over three decades ago (Richard and Manley 2009; Barabino and Keller 1999) and is now understood to be a commonly used gene regulation mechanism in all eukaryotes. It is driven by the condition- and tissue-dependent expression of regulators that ultimately determine which APA site is used (Gruber and Zavolan 2019). Tissue-restricted genes are more commonly found to have a single 3'UTR isoform than those that are ubiquitously expressed, linking APA to cell identity (Lianoglou et al. 2013; Reyes and Huber 2018). When compared with each other, UTR-APA and CR-APA can lead to different ratios of shortening and lengthening in the same cells, suggesting differences in the mechanisms that regulate each category (Sandberg et al. 2008). 3'UTR lengths, along with the wider non-coding portion of the genome, have evolved in line with animal complexity. Median lengths have expanded from approximately 140 nucleotides in *C. elegans* to 2,000 nucleotides in humans, when looking only at those genes that have more than one polyadenylation site (Jan et al. 2011; Lianoglou et al. 2013). APA has been established to play a role in health and disease, including cancer (Xia et al. 2014; Fu et al. 2011; Zhang, Sun, Ren, et al. 2019; Bruhn et al. 2020;

Zhang, Liu, et al. 2021; Fischl et al. 2019), systemic lupus erythematosus (Graham et al. 2007), myotonic dystrophy (Batra et al. 2014), amyotrophic lateral sclerosis (Prudencio et al. 2015), and viral infection (Jia, Yuan, et al. 2017).

1.3.1 UTR-APA

Usage of an APA site in the 3'UTR region of the transcript does not affect its coding capacity, but because UTRs can contain cis-elements that contribute to RNA metabolism, UTR-APA has the potential to regulate genes post-transcriptionally. By modifying the length of the 3'-UTR, the ability of trans-elements such as protein factors or microRNAs (miRNAs) to bind the transcript is altered. In this way, the RNA molecule's stability, translational capability and cellular localisation can be regulated (Elkon, Ugalde, and Agami 2013). Furthermore, while UTR-APA does not change the coding potential of the new transcript, the distinct UTR lengths can affect the location of translation and in this way impact protein localisation (Berkovits and Mayr 2015; Arora et al. 2021).

When APA results in a short isoform, the absent UTR sequence that would be present in the longer isoform is referred to as the alternative UTR (aUTR), while the section of UTR found in both the short and long isoform is termed the constitutive UTR (cUTR). A diagram of UTR-APA can be seen in the top panel of Figure 1.3.

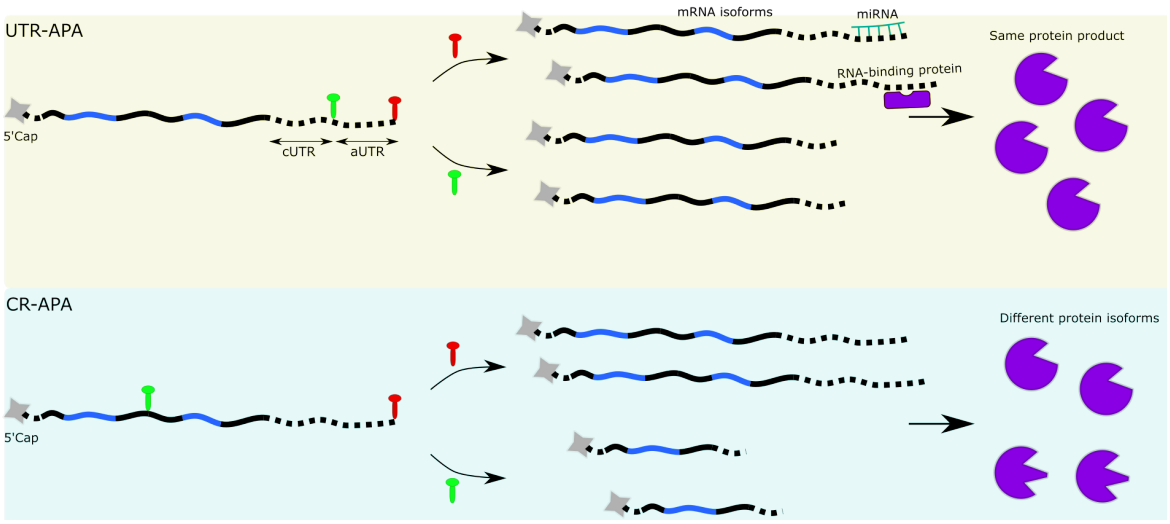


Figure 1.3: Diagrammatic overview of UTR-APA and CR-APA. There are two main types of APA, which are dependent on the placement of the polyadenylation sites (PASs). Differential usage of a PAS will lead to unique mRNA isoforms being generated. If the chosen PAS is within the coding region of the gene, this is termed CR-APA and translation of the resulting transcript will lead to a different protein isoform. In UTR-APA, the PAS is found within the 3' untranslated region of the transcript and will result in an mRNA molecule with a different length 3'UTR tail. The cUTR (constitutive UTR) is present in transcripts whether or not the proximal or distal PAS is used; the aUTR however is only present in the long isoform. The metabolism of the mRNA molecule may be modified in UTR-APA, but because the coding region is untouched, the coding potential of the gene is not changed. Solid black line, introns. Solid blue line, exons. Dashed line, untranslated region. Red pin, distal PAS. Green pin, proximal PAS.

Stability and translation of mRNA

Probably the best-studied impact of UTR-APA is its effect on retaining or deleting miRNA binding sites in nascent transcripts (Zhang, Liu, et al. 2021). miRNAs are small (∼22 nt) single-stranded non-coding RNA molecules that can silence gene expression by destabilising or cleaving mRNAs, or by hindering their translation (Bartel 2009; Fabian, Sonenberg, and Filipowicz 2010). Approximately 50% of the conserved microRNA binding sites in mammals are situated downstream of the proximal PAS (Sandberg et al. 2008; Ji, Lee, et al. 2009). The first study to observe differential UTR PAS selection reported a clear shift towards proximal PAS preference (shortening) in murine T-cell activation (Sandberg et al. 2008), and this trend has since been well-documented in proliferating cells also (Elkon, Ugalde, and Agami 2013). Targetting efficiency by miRNAs is modulated by UTR-APA - those miRNA binding sites near the end of a 3'UTR tend to be more active than those found more centrally (Hoffman et al. 2016). A global analysis study also found that UTR-APA influences about 10% of miRNA targetting between any two analysed cell types (Nam et al. 2014). One specific example of UTR-APA impacting miRNA binding capability and therefore stability of a gene is the myogensis regulator, PAX3. miRNA-206 is highly expressed in muscle stem cells, and a binding site for this miRNA is found in the PAX3 3'aUTR (Goljanek-Whysall et al. 2011; Morree et al. 2019). Varying ratios of long and short PAX3 isoforms are found in different muscle precursor populations. The long form, containing the miRNA-206 binding site, is dominant in limb satellite cells (SCs) whereas the short form, lacking the target site, is dominant in diaphragm SCs. Consequently PAX3 is able to maintain high levels in diaphragm SCs - which have high levels

of miRNA-206 - but not in limb SCs, where the PAX3 transcript can be targetted by miRNA-206 (Boutet et al. 2012).

A study utilising CRISPR technology demonstrated that the APA-generated long isoform of the homothorax (HTH) gene in *Drosophila* neurons exposed multiple miRNA binding sites that regulate its protein output and subsequently control normal adult behaviour (Garaulet et al. 2020). In an example of APA regulating miRNA binding potential in cancer, Li et al. demonstrated that FNDC3B 3'UTR shortening helps it escape miRNA-mediated gene repression in nasopharyngeal carcinoma, and that overexpression of FNDC3B promotes cancer progression (Li, Chen, et al. 2020).

The MCL1 gene, which plays an important role in the survival of T cells at multiple stages of their development and maintenance, is another example of a gene whose protein expression is regulated at the APA level by including or excluding 3'UTR cis-elements (Pereira-Castro, Garcia, et al. 2022). Two APA isoforms are produced of MCL1, independent of alternative splicing. Pereira-Castro et al. demonstrated in a recent study that expression of the shorter isoform is positively correlated with MCL1 protein levels, and the expression of this isoform is increased upon T cell activation. The longer MCL1 isoform, which contains an miRNA binding site, is less efficiently translated. Therefore, when T cells are activated, MCL1 protein expression is increased by higher production of the shorter MCL1 APA isoform (Pereira-Castro, Garcia, et al. 2022).

Cis-elements located in 3'UTRs can also provide platforms for destabilisation via RNA-binding proteins (RBP) (Figure 1.3). Just as with miRNAs, the deletion or retention of these elements can affect the transcript's stability. The most well-documented example is the AU-rich element (ARE), which is commonly found in

proteins that need to be tightly regulated: cytokines, transcription factors, proto-oncogenes and cell cycle regulators (Garneau, Wilusz, and Wilusz 2007; Brewer 1991). AREs have a strong degree of heterogeneity - it seems no two AREs are the same - but many decay-enhancing proteins are known to bind AREs. Even the presence of an ARE by itself is known to increase the exosome's affinity for the RNA molecule (Garneau, Wilusz, and Wilusz 2007). Because cis elements like miRNA sites and AREs have been linked to decay and negative gene regulation, it was commonly believed that a "long"-dominant APA profile - which would have a higher proportion of these cis elements - had a generally negative effect on gene expression, while an APA trend of shortening would lead to upregulation of genes. This view has been challenged by studies which show many UTR elements being positive for gene expression (Geisberg, Moqtaderi, Fan, et al. 2014; Oikonomou, Goodarzi, and Tavazoie 2014; Kristjánsdóttir, Fogarty, and Grimson 2015). One example showing how a long 3'UTR promotes mRNA stability is that of the prospero gene in *Drosophila*. Samuels et al. showed that an APA switch causes a 15kb extension of the 3'UTR of the prospero gene which stabilises the transcript by acting as a scaffold for the RNA-binding protein Syncrip (Samuels et al. 2020). Another example is the SERT (serotonin receptor) gene: its long isoform is translated at a higher rate (Yoon et al. 2013). On the other hand, Cheng et al. have recently demonstrated that widespread transcript shortening in differentiating syncytiotrophoblast cells generates mRNAs featuring higher stability (Cheng, Zheng, et al. 2020).

Regardless of UTR-APA trends being beneficial or detrimental to gene expression, it is clear from documented examples that mRNA stability and subsequent protein expression is impacted by UTR-APA, and that this has major biological

implications (Pereira-Castro and Moreira 2022).

mRNA localisation

Long APA isoforms tend to be more commonly found in the nucleus than the cytoplasm and even though this observation is conserved between cell types, the genes in question are cell-specific (Neve, Burger, et al. 2016). Indeed, a recent study has demonstrated that longer 3'UTR isoforms are specifically enriched in the nucleus, with this behaviour being attributed to possible nuclear retention elements in the aUTR (Tang, Yang, et al. 2022). At the individual gene level, one study showed that the inclusion of *Alu* repeats in the long APA isoform of the NICOLIN1 gene blocked its export from the nucleus (Chen, DeCerbo, and Carmichael 2008). Despite this observation, it is difficult to clearly distinguish this localisation effect on a global level from being the result of a difference in mRNA stability or nuclear export efficiency.

The impact of distinct aUTRs in subcellular localisation within the cytoplasm is better understood at the transcript-specific level. The biological relevance of this has been shown for several transcripts in neurons, with BDNF being the classic example. BDNF is a neurotrophic factor that is involved in synaptogenesis, and differential usage of its UTR polyadenylation sites direct transcripts to the soma or dendrites, depending on whether or not the transcript is the long or short APA isoform respectively (An, Gharami, et al. 2008). Another example of this effect can be seen with RANBP1, an effector of Ran GTPase, whose short APA isoforms are restricted to the cell body in contrast to the long form which is localised to the axon (Yudin et al. 2008). On a global scale, RNA sequencing of transcripts found in somata or the neuropil in neurons found a propensity for longer 3'UTRs in

either compartment (Tushev et al. 2018). These long isoforms contained regulatory elements that the short forms did not (Tushev et al. 2018). Conversely, another study found long isoforms only in the neurites as opposed to somata (Taliaferro et al. 2016). It is clear from studies such as these that UTR-APA creates distinct isoforms which contain or omit regulatory elements dependent on differential usage of UTR PASs. In a more recent study, Ciolli et al. identified multiple 3'UTR isoforms of the CDC42 gene using a combination of sequencing technologies (Ciolli Mattioli et al. 2019). The researchers then employed a mass spectrometry-based approach to demonstrate differential 3'UTR-bound proteomes for each isoform which influence the localisation and translation of the corresponding transcript (Ciolli Mattioli et al. 2019).

While there are clear examples of alternative PAS usage impacting transcript localisation and translation at the single gene level, further research will elucidate the detail in the underlying mechanisms and impacts of this phenomenon at a broader level.

Protein localisation and function

While UTR-APA was initially believed to only affect the properties of the nascent mRNA transcript, it was later discovered that UTR-APA can also impact the newly translated proteins. UTRs can act as scaffolds which enhance the binding of proteins to nascent proteins - regulating their function or transport without modifying the coding capacity of the mRNA transcript (Mayr 2019).

The example that typifies this phenomenon is the CD47 gene, which encodes a cell surface molecule that is a marker for "self" in the context of the immune system. The long CD47 mRNA isoform recruits the ELAVL1 RBP, and SET, to

the site of translation. Here, SET associates with the newly made cytoplasmic domain of CD47, leading to its translocation to the cellular membrane. In contrast, the short CD47 isoform which lacks the sequence required for ELAVL1-SET complex formation, generates a protein which is primarily localised to the endoplasmic reticulum (Berkovits and Mayr 2015). Thus, both the long and short CD47 isoforms generate the same CD47 protein, but the protein is localised to different cellular compartments and carries out different functions dependent on which isoform is translated. In addition to this effect, it was shown that both proteins even when localised to the same cellular compartment have different functional profiles (Berkovits and Mayr 2015).

Another study demonstrated a change in protein function caused by alternative 3'UTR length (Lee and Mayr 2019) by investigating the BIRC3 gene, which encodes an E3 ubiquitin ligase. Long BIRC3 transcripts are upregulated in cells derived from chronic lymphotic leukaemia, but Lee et al. noted that overall BIRC3 mRNA and protein levels remained similar between normal cells and the cancer phenotype, raising the question of purpose for the disparity in 3'UTR lengths. By employing CRISPR-Cas9 knockout cells and RNA interference, the authors observed that only the BIRC3 long isoform is necessary for B cell migration (Lee and Mayr 2019). Using quantitative mass spectrometry and other methods, it was shown that the BIRC3 protein encoded from the long mRNA isoform is a member of several protein complexes and whose interaction with these complexes is 3'UTR-dependent. In this way, the long 3'UTR instills new functions into the corresponding protein that it encodes when compared with the shorter 3'UTR, including the regulation of B cell migration through its interaction with IQGAP1 and RALA. Interestingly, the authors suggest that BIRC3 is one of many as yet

undiscovered examples that features a 3'UTR-dependent change in protein function. When comparing overall mRNA changes and 3'UTR ratios between normal and cancer cells, the vast majority of differential 3'UTR isoforms did not see a corresponding change in mRNA levels, thereby leading the authors to conclude that a portion of these changes impact protein function instead (Lee and Mayr 2019).

1.3.2 CR-APA

A large portion of PASs exist upstream of the UTR, most of which are found in introns. In mice, for example, approximately 40% of genes contain a PAS upstream of the 3'most exon (Hoque et al. 2013). CR-APA can result in alterations to the coding sequence and UTR of an mRNA, and consequently lead to the translation of a protein with a different C-terminus. CR-APA events can be further divided into two categories: those that generate a skipped terminal exon and those that generate a composite terminal exon (Figure 1.4). Skipped terminal exons are alternative terminal exons generated by splicing, and composite terminal exons are generated when an intronic PAS is used (Tian and Manley 2017). A PAS can also be selected in internal exons, but as these transcripts will not have in-frame stop codons, these are presumed to be rapidly degraded by the non-stop decay system (Vasudevan, Peltz, and Wilusz 2002). Most truncated proteins generated via CR-APA still conserve function, with the median truncated protein retaining two domains (Singh, Lee, et al. 2018). Genes with an intronic poly(A) site encode proteins that are larger and contain more domains than genes that do not have an intronic PAS (588 amino acids vs 432 amino acids respectively) (Singh, Lee, et

al. 2018). As CR-APA alters the coding potential of transcripts, its physiological relevance is more readily observed than that of UTR-APA.

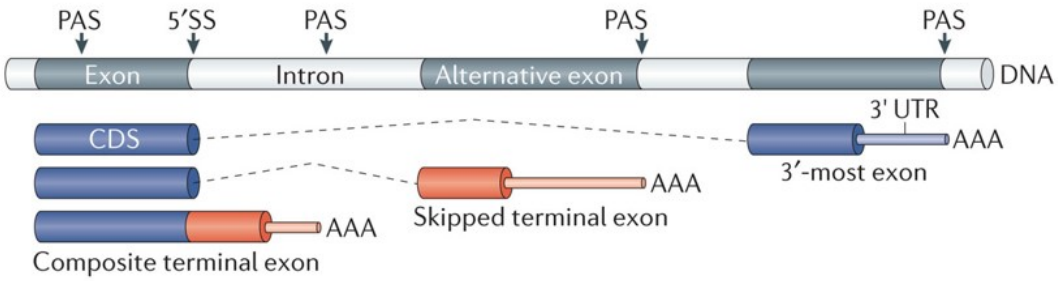


Figure 1.4: Coding region APA (CR-APA). CR-APA can be subdivided into two main types, each generating a different 3'-terminal exon. A canonical isoform is shown (top) which uses a polyadenylation site (PAS) located in the 3'-most exon. In a different scenario, a PAS in an alternative exon is used, which produces an isoform with a skipped terminal exon (middle). Finally, splicing inhibition at the 5' Splice Site (SS) leads to a PAS within an intron being selected. This leads to an isoform which includes a part of that intron and an exon that is described as a composite terminal exon. Figure adapted from Tian and Manley 2017.

The first discovered and probably most well-known example of CR-APA is that of the immunoglobulin M (IgM) heavy chain (Alt et al. 1980). When B cells transition to plasma cells, the IgM protein transitions from a membrane-bound state to a secreted form. This is induced by the CR-APA switch from the distal PAS to one found in a composite terminal exon. A later study observed that this CR-APA switch was associated with a specific increase in the levels of CSTF64, suggesting an interesting regulatory mechanism (Takagaki, Seipelt, et al. 1996). A more recent study notes that membrane proteins are in fact underrepresented in CR-APA genes, with a total of 499 membrane-protein encoding genes that undergo CR-APA while retaining at least one protein domain (Singh, Lee, et al. 2018). Corroborating the IgM story, many of these genes are cytokine receptors or growth factor receptors, and CR-APA generates secreted isoforms of the corresponding proteins.

While CR-APA can lead to proteins with distinct functions, the majority of all CR-APA events are thought to generate truncated proteins which retain at least one domain, with the potential to be functional (Singh, Lee, et al. 2018). Enzymes that undergo CR-APA truncation tend to keep their active sites, suggesting that while their function is retained, their binding affinity may be altered.

Just as CR-APA can qualitatively impact protein expression by altering the coding sequence of mRNA transcripts, it can quantitatively regulate the proteome by generating transcripts which likely lack any function (Tian and Manley 2017). This occurs when PASs located in promoter-proximal introns are used, resulting in a severely-truncated protein which probably lacks any function. This can be seen in the CSTF77 gene, where differential usage of an intronic PAS generates a shortened transcript which lacks sequences responsible for many of the CSTF77 protein's functions but contains stop codons (Pan et al. 2006). It was also observed that CSTF77 transcript isoform expression is regulated via a negative feedback loop, with the short and long isoforms having opposite expression patterns between tissues (Pan et al. 2006; Luo, Ji, et al. 2013).

1.3.3 APA regulation

Understanding how APA is regulated is an active area of research, given the observed and potential ability for APA to modulate gene expression. Much progress has been made in a relatively short space of time, due in part to recent technical advances which have enabled numerous datasets of PAS usage to be published and analysed.

Altering core CPA machinery levels

CPA is a comprehensive biological process, and because 3'-end processing is tightly regulated in order to achieve this, it is not surprising that one of the most important ways to regulate PAS selection is by altering the concentration levels of any of the numerous proteins (Curinha et al. 2014). One of the first observed examples of this was seen in B cells with the association of a change in CSTF64 levels with IgM switching from membrane-bound to secreted form (Takagaki, Seipelt, et al. 1996). Upon B cell activation, CSTF64 is upregulated, leading to the preferential usage of proximal PASs in some genes, including IgM (Chuvpilo et al. 1999). In line with these results, a later study showed that siRNA-mediated knockdown of CSTF64 and its paralogue, CSTF64τ, caused global lengthening (Yao, Choi, et al. 2013). This study also showed that there is at least some redundancy between these two proteins, as knockdown of each factor on its own was not sufficient to cause the global lengthening (Yao, Choi, et al. 2013). Another study demonstrated that some proximal PASs have a propensity for CSTF64/CSTF64τ binding via an uncommon sequence motif upstream - GU[G/U][G/U]U (Hwang, Park, et al. 2016), suggesting that some PASs are sub-optimal and only used when concentrations of 3'-processing factors are high.

The CFI complex is the polyadenylation factor currently considered the main regulator for UTR-APA (Ghosh et al. 2022). In contrast to CSTF64, it is understood to promote distal PAS usage and thus be responsible for a shift towards lengthening (Martin, Gruber, et al. 2012; Li, You, et al. 2015; Gruber, Martin, et al. 2012; Zhu, Wang, et al. 2018; Zhou, Qu, et al. 2020). A comprehensive study involving the knockout and overexpression of both CFI25 and CFI68 built

a reference set of 858 transcripts exhibiting APA, with 855 of these presenting as 3'UTR shortening when CFI is knocked down (Ghosh et al. 2022). The motif to which CFI25 binds - UGUA - is enriched in the upstream sequence of distal PASs (Martin, Gruber, et al. 2012; Li, You, et al. 2015; Gruber, Martin, et al. 2012; Ghosh et al. 2022), and knockdown via siRNA of either CFI25 or CFI68 results in increased usage of proximal PASs (Martin, Gruber, et al. 2012; Li, You, et al. 2015; Gruber, Martin, et al. 2012).

FIP1, a core member of the CPSF complex, mediates, at least in part, the interactions between CFI and CPSF, with CFI68 being more important to this interaction than CFI59 (Zhu, Wang, et al. 2018). A model proposed in order to explain this observation posits that CFI binds two UGUA elements - one upstream and one downstream of the PAS - and thereby causing a skipping of the PAS. In this model, CFI would have an inhibitory effect on 3'-end processing (Martin, Gruber, et al. 2012; Yang, Gilmartin, and Doublié 2011). Surprisingly, knockdown of CFI59 - one of the possible components of the CFI complex which is mutually exclusive with CFI68 - did not induce the same effect. Whether or not this model is reflected in biological reality and further identification of CFI59's role in PAS selection requires further work. Another more recently proposed model suggests that CFI sterically blocks cis-elements such as the poly(A) signal or cleavage site by binding UGUA elements adjacent to the proximal PAS (Schwich et al. 2021).

Altering CFII complex levels has also been shown to regulate PAS selection. Knockdown of the CFII complex member PCF11 enhances distal PAS usage in mouse myoblast cells (Li, You, et al. 2015) and in differentiating neuroblastomas (Ogorodnikov et al. 2018). This effect is consistent with PCF11's role in modulating PolII transcription termination (West and Proudfoot 2008). Supporting this,

Yang et al. found that PCF11 ubiquitination in cancer cells leads to loss of CFI25 from the CPA machinery, ultimately resulting in widespread 3'UTR shortening (Yang, Li, Connelly, et al. 2020). Another recent study also observed global UTR shortening upon PCF11 knockdown (Wang, Zheng, Wei, et al. 2019) but intriguingly also uncovered a new role for PCF11 in regulating gene expression via APA based on gene size. Large genes, which tend to have large introns, were found to be upregulated upon PCF11 knockdown. Large introns are more likely to harbour PASs (Tian, Pan, and Lee 2007), and it is thought that PCF11 suppresses selection of these intronic PASs (Wang, Zheng, Wei, et al. 2019). This gene regulation mechanism was seen to be active during cell differentiation and affected those genes with large introns the most (which tend to be involved with morphology, adhesion and migration). Surprisingly, PCF11 also was found to have an auto-regulation mechanism, much like CSTF77 that was discussed earlier. PCF11 was found to control its own gene expression levels via an intron PAS, which is highly conserved among vertebrates. Removing this intronic PAS resulted in a substantial upregulation of both the PCF11 mRNA and protein levels (Wang, Zheng, Wei, et al. 2019). CLP1, the other member of the CFII complex, has recently been identified as a promoter of the proximal PAS in human motor neuron cells, particularly favouring intronic PASs (LaForce et al. 2022). When CLP1 was knocked out in cells, the distal PAS is less favoured (LaForce et al. 2022), which mirrors the APA profile seen in PCF11 overexpressions (Wang, Zheng, Wei, et al. 2019).

PABPN1, which is critical in restricting the length of the polyadenylation tail (Kuhn et al. 2009), has also been found to play a role in regulating APA. PABPN1 suppresses proximal PASs and PABPN1 knockdown therefore induced 3'UTR shortening (Jenal et al. 2012). Another study found that knockdown of

PABPN1 encouraged use of PASs near gene promoters, which is probably related to its function in hyperpolyadenylation (Li, You, et al. 2015; Tian and Manley 2017). PABP1, another poly(A)-binding protein found predominantly in the cytoplasm, is also known to modulate 3'UTR length, although exactly how this occurs is poorly understood (Li, You, et al. 2015).

Intriguingly, APA seems to feature an in-built auto-regulation mechanism. When there is a demand for high levels of 3'-processing factors, such as during proliferation, the relevant factors are transcriptionally upregulated via E2F mediations (Elkon, Drost, et al. 2012). Consequently, when the 3'-processing factor levels are high, a highly-conserved alternate, intronic poly(A) site in the CSTF77 gene is favoured. Usage of this PAS generates a short, non-functional CSTF77 isoform, while at the same time reducing the levels of long, functional isoforms of CSTF77 (Luo, Ji, et al. 2013).

Altering FIP1 levels has also been linked to UTR shortening in embryonic stem cells (Lackford et al. 2014). In one study, of the more than 300 genes that underwent lengthening when FIP1 was knocked down, only 7 of these also experienced a change in overall transcript level, indicating that FIP1 regulates APA without impacting mRNA levels (Lackford et al. 2014). One study observed that FIP1 had the highest contribution of all CPSF subunits to APA regulation, concluding that FIP1 recruitment is a critical step for CPA (Li, You, et al. 2015). In an interesting example of a powerful link between APA and oxidative stress, one study found that FIP1 is upregulated during chronic kidney disease, and that FIP1 promotes 3'UTR shortening of NLRP3, a component of the inflammasome (Zheng, Tan, et al. 2021). The shortened isoform of NLRP3 in turn amplifies inflammasome activation and triggers reactive oxygen species (ROS) production.

In addition to these findings, the authors also found that silencing FIP1 reduced inflammation activation, cell damage and apoptosis, and that FIP1 upregulation leads to NLRP3 3'UTR shortening, which itself induces inflammasome activation, ROS production and apoptosis (Zheng, Tan, et al. 2021). The authors suggest that taken together, these data offer FIP1 as a potential therapeutic target for kidney injury (Zheng, Tan, et al. 2021). In another study, knocking down FIP1 was seen to have the greatest impact on APA when compared with the knockdown of other CPSF subunits, inducing 3'UTR lengthening (Li, You, et al. 2015).

Other factors regulating APA

Splicing factors are known to interact with APA (Niwa, Rose, and Berget 1990). This link is perhaps most clearly seen at intronic PASs, where splicing and APA most likely compete with one another. When splicing is inhibited, for example by knocking down a U2 snRNP component, intronic PAS selection is dampened (Li, You, et al. 2015; Tian and Manley 2017). Furthermore, it had long been wondered why the U1 snRNP (U1) is found at much higher levels than the other snRNPs even though they form spliceosomes in equal stoichiometry (Kambach, Walket, and Nagai 1999). This mystery was addressed when a study revealed that knockdown of the U1 snRNP caused CPA in many pre-mRNAs, commonly at the start of the transcript (Kaida et al. 2010), suggesting that U1 snRNPs protect nascent transcripts from being cleaved and polyadenylated prematurely. Reminiscent of PCF11 ablation, knockdown of U1 snRNPs disproportionately suppressed genes with long introns compared to short genes (Oh, Di, et al. 2017). A follow-up study showed that U1 homeostasis is important for balancing gene expression in cancer cells, and that shifts either way in its activity lead to the modulation of the

cancer cell phenotype (Oh, Venters, et al. 2020). Among other changes, widespread changes in UTR length were observed, with UTR shortening noted when U1 was inhibited and UTR lengthening when U1 was overexpressed (Oh, Venters, et al. 2020). RBPs interacting at or near PASs also have the potential to regulate APA by influencing PAS selection. Because of the already-discussed interconnectedness of splicing and APA, many RBPs that are already known to affect splicing have also been found to impact APA. The first such example that was discovered to regulate APA is NOVA2, an RBP specific to neurons (Licatalosi et al. 2008). Knocking out NOVA2 resulted in 3'UTR shortening, demonstrating that NOVA2 promotes UTR lengthening. NOVA2 was found to bind close to proximal PASs, suggesting that this mechanism is possibly induced by steric hindrance (Licatalosi et al. 2008). A later study identified NOVA2 and another protein, PTBP2, as two neuron-specific RNA-binding proteins that contribute to distinct APA profiles between excitatory and inhibitory neurons in microglia and astrocytes (Hwang, Saito, et al. 2017).

A *Drosophila* protein involved in alternative splicing, ELAV, was found to upregulate UTR lengthening when ectopically expressed in non-neuronal cells (Hilgers, Lemke, and Levine 2012). RNA immunoprecipitation in *Drosophila* neuronal cells pointed to ELAV binding proximal PASs to achieve this effect (Hilgers, Lemke, and Levine 2012). In the first study to demonstrate widespread loss of a tissue-specific APA profile, recent work deleting *Drosophila* ELAV family proteins show that these proteins are necessary and sufficient to broadly drive neural 3'UTR lengthening (Wei, Lee, et al. 2020). Other recent work further solidifies the role of the ELAV family in producing longer 3'UTR isoforms (Lee, Wei, et al. 2021; Wei and Lai 2022; Carrasco et al. 2020). The human ELAVL family of proteins (also called Hu), which are homologues of ELAV, block CPA at PASs containing U-rich

elements (Zhu, Zhou, et al. 2007). Three of these proteins - ELAVL3, ELAVL2 and ELAVL4 -, which are exclusively expressed in neurons, were also found to promote lengthening (Mansfield and Keene 2012; Zhu, Zhou, et al. 2007). ELAVL1, a family member which in contrast is ubiquitously expressed, was found to play a role along with its ELAV family members, in the regulation of its own APA (Dai, Zhang, and Makeyev 2012; Mansfield and Keene 2012). The nuclear ELAVL1 protein, and its ELAVL relatives are able to bind a GU-rich element near one of ELAVL1's major PASs, promoting the longer and less stable ELAVL1 isoform to be generated. Because of the neuron-specific expression of ELAVL3, ELAVL2 and ELAVL4, this mechanism balances the pro-differentiation activity of the neuron-specific members with the pro-proliferation activity of ELAVL1 (Mansfield and Keene 2012).

SRSF3 and SRSF7 are two other splicing factors that have also been found to play opposite roles in APA regulation (Müller-McNicoll et al. 2016). In mouse P19 cells, SRSF7 depletion leads to UTR lengthening, while SRSF3 knockdown leads to UTR shortening (Müller-McNicoll et al. 2016). It is thought that the shortening caused by SRSF3-knockdown can remove the binding site for NXF1, an important mRNA nuclear export factor, from transcripts. In this way, SRSF3 promotes distal PAS usage and upregulates mRNA nuclear export whereas SRSF7 promotes shortening and downregulates export (Müller-McNicoll et al. 2016). Schwich et al. also demonstrated that the APA regulation function of SRSF3 and SRSF7 can be independent of their splicing activities by binding RNA directly, upstream of proximal PASs (Schwich et al. 2021).

APA regulation by transcription

Since APA is most likely always co-transcriptional, transcription and the processes associated with it have also unsurprisingly been linked to APA regulation. This regulation can happen either directly via trans-factors such as transcription machinery directly modulating PAS choice, or by altering the rate of PAS elongation via mechanisms such as cis-elements which promote PolII pausing (Tian and Manley 2017).

Certain transcription factors can affect PAS selection directly or mediate recruitment to the PAS of factors that do. Knockdown of the PolII-associated factor complex (PAF) - which plays roles in transcriptional elongation and chromatin modification, among others - is associated with dramatic UTR shortening, via activation of intronic or proximal UTR PASs (Yang, Li, Hoque, et al. 2016). One of PAF's subunits, CDC73 is known to associate with the CPSF and the CSTF polyadenylation complexes (Rozenblatt-Rosen et al. 2009). A study using knockdowns and RNA-seq to provide further evidence that PAF is crucial to PolII-pausing (Yu et al. 2015) strengthens the link between PAF and PAS selection. Considering the strength of the link between PolII transcription and PAS selection, it is no surprise too that depleting or inactivating PolII alters PAS usage (Lima Neto et al. 2016).

It is well-established that cis-elements, particularly G-rich sequences, that can cause PolII pausing can also activate PAS usage (Yonaha and Proudfoot 1999). Alterations to the rate of PolII elongation can also impact PAS selection. A *Drosophila* study found that mutants with a reduced elongation rate also preferred proximal PASs (Pinto, Henriques, et al. 2011). It is posited that the mechanism

behind this is that a slower PolII exposes the proximal PAS of the nascent transcript to the polyadenylation machinery for a longer time, therefore increasing the opportunity that this PAS will be selected (Pinto, Henriques, et al. 2011). A recent study in yeast lends support to this model, where reduced elongation rates during a metabolic shift is accompanied by widespread transcript shortening (Geisberg, Moqtaderi, and Struhl 2020). The researchers also observed that a shift towards transcript lengthening positively correlated with increased elongation rates, although to a lesser extent (Geisberg, Moqtaderi, and Struhl 2020). These findings are consistent with an earlier study that demonstrated the same phenomenon - UTR shortening in elongation-compromised cells - in yeast (Cui and Denis 2003).

Epigenetic regulation of APA

Chromatin structure controls access to regulatory regions of the genome and as such, is tightly entwined with transcription, enabling chromatin and epigenetic modifications to regulate APA. Yeast studies show that the PAS region is devoid of nucleosomes (Spies et al. 2009; Tian and Manley 2017). A higher affinity for nucleosomes around commonly used PASs (Spies et al. 2009) and a correlation between nucleosome levels and PolII presence downstream of PASs suggests a role for chromatin organisation in APA regulation (Grosso et al. 2012). One study revealed such a link between heterochromatin and APA; DICER1, which as an endoribonuclease III mediates miRNA-degradation of APA isoforms, was found to promote heterochromatin formation and ultimately cleavage at proximal PASs (Neve, Burger, et al. 2016). Histone modification seems likely to affect APA, with several studies showing a correlation between it and PAS switching (Ji, Luo, et al. 2011; Li, Park, et al. 2016; Khaladkar, Smyda, and Hannenhalli 2011). Evidence

implicating two yeast histone methyltransferases, SET1 and SET2 (H3K4 and H3K6 respectively), in epigenetically regulating APA has recently been presented (Kaczmarek Michaels et al. 2020). Deletion of either methyltransferase modulates serine phosphorylation the PolII CTD and subsequent recruitment of the PAS machinery, both of which are suggested to account for the PAS switching (Kaczmarek Michaels et al. 2020). DNA methylation is an epigenetic modification that can impact PAS choice. The first examples of this were seen in mouse tissue (Wood et al. 2007; Cowley et al. 2012), where two retrogenes are situated inside introns of two host genes. The retrogene promoters are imprinted and silenced on the maternal allele but unmethylated and therefore active on the paternal allele. When a retrogene is transcribed, a PAS switch is induced in the gene on the other allele, suggesting a role - though inconclusive - for genetic imprinting in APA regulation (Wood et al. 2007; Cowley et al. 2012). A recent genome-wide study compared DNA methylation and PAS usage in cells that were DNA methylation-competent and cells that were DNA methylation-deficient and found a correlation between DNA methylation and APA (Nanavaty et al. 2020). Over 500 genes underwent a PAS switch in DNA methylation-deficient cells, with most of these experiencing isoform shortening. Binding sites for a multifunctional transcription factor, CTCF, were found to be highly enriched in the aUTRs of the shortened genes. As CTCF - which is highly conserved and has 11 zinc fingers (Filippova et al. 1996) - is well-documented in having DNA methylation-sensitive binding properties (Phillips and Corces 2009), this observation suggests an apparent link between DNA methylation and PAS usage. This study went on to conclude that the absence of DNA methylation enables CTCF binding, which enables recruitment of the cohesin complex and the formation of chromatin loops, ultimately promoting

proximal PAS usage (Nanavaty et al. 2020).

1.3.4 APA in Health and Disease

APA is commonly modulated in human health, including immunological (Bennett et al. 2001; Rogers et al. 1980; Nath et al. 2004), haemotological (Orkin et al. 1985; Rund et al. 1992; Higgs et al. 1983), neurological (Polymeropoulos et al. 1997; Tomé and Fardeau 1980) and endocrine (Duan et al. 2009; Lin, Gitelman, et al. 1991; Tee et al. 1995) diseases, in addition to cancers (Lin, Li, et al. 2012; Fu et al. 2011; Rehfeld et al. 2014; Fischl et al. 2019). In this section, examples of APA in cancer will be highlighted followed by an introduction into APA changes brought about during viral infection.

APA in Cancer

With the advent of widespread next-generation sequencing studies, much evidence has been presented demonstrating changes in PAS usage in cancer tissue. Some of these studies report a shift in the APA profile towards the generation of short isoforms in proliferating cells (Sandberg et al. 2008; Mayr and Bartel 2009; Fu et al. 2011), while others show a shift towards isoform lengthening (Lin, Li, et al. 2012; Singh, Alley, et al. 2009) demonstrating that APA profiles in cancer cells are complex. One large study that analysed 358 sets of paired tumour and non-tumour RNA-seq data deposited in the Cancer Genome Atlas across seven tumour types identified ∼1350 genes that undergo a PAS switch which have the potential to be used as biomarkers as the APA events were tumour-specific (Xia et al. 2014).

APA has been suggested to play a role in the activation of oncogenes (Yuan,

Hankey, et al. 2021; Masamha and Wagner 2018). This mechanism would naturally only involve those oncogenes that include more than one poly(A) site, and would mean that the oncogene is activated by removing regulatory (and possibly repressive) cis-elements from the mRNA transcript (Masamha and Wagner 2018). PAS switching in the cyclin D1 gene is one of the best physiological examples of oncogene activation by APA. In Non-Hodgkin's Mantle cell lymphoma, mutations within the UTR create a proximal PAS with the canonical hexamer, and tumours that used this PAS for cyclin D1 have higher levels of the transcript and higher proliferation rates (Wiestner et al. 2007). Reporter assays show that usage of the mutated proximal PAS cause several miRNA binding sites to be excluded from the transcript (Mayr and Bartel 2009; Chen, Bemis, et al. 2008; Deshpande et al. 2009). CDC6 is another oncogene that plays a critical role in the regulation of DNA replication (Yan et al. 1998). One study found that treating breast cancer cells with oestrogen led to shortening of CDC6's 3'UTR. As with cyclin D1, this PAS switch resulted in the shortened CDC6 isoform lacking miRNA binding sites which otherwise would repress its protein expression (Akman, Can, and Erson-Bensan 2012).

Widespread usage of intronic PASs in chronic lymphotic leukaemia has been reported, with most of the affected genes having tumour-suppressing functions (Lee, Singh, et al. 2018). DICER is one such gene, whose truncated form completely lacks the ability to cleave miRNA and consequently a decrease in miRNA expression is induced. Another tumour suppressor gene that is regulated by this mechanism is FOXN3, a transcriptional repressor. When an intronic PAS is used, FOXN de-represses its oncogenic targets MYC and PIM2 (Lee, Singh, et al. 2018).

The CFI complex has been implicated in oncogenesis. Expression of several

oncogenes increase when CFI25 is depleted, possibly because of its tendency to promote 3'UTR lengthening (Abadi et al. 2019; Masamha, Xia, et al. 2014). Indeed, CFI25 was shown to act as a tumour suppressor that inhibits growth of cervical cancer (Xing et al. 2021). CFI25 expression is downregulated in cervical cancer cells, and its tumour suppresive ability comes from direct involvement with PAS selection for genes involved in fatty acid metabolism and the Wnt and NF-κB signalling pathways (Xing et al. 2021). On the other hand, a cancer-promoting effect of CFI68 has been demonstrated in hepatocellular carcinoma cells by which it promotes 3'UTR shortening of the NQO1 metabolism regulator (Tan et al. 2021).

The CFII complex has also been found to play a role in altering PAS selection in cancer. PCF11 ubiquitination in cancer cells leads to loss of CFI25 from the CPA machinery, ultimately resulting in widespread 3'UTR shortening (Yang, Li, Connelly, et al. 2020). The authors of the study, Yang et al., defined a novel mechanism whereby MAGE11-A11, a member of the MAGE11 family which has been implicated in oncogenesis and is aberrantly expressed in cancer (Weon and Potts 2015), directly binds PCF11 and recruits it to the HUWE1 E3 ubiquitin ligase, promoting its ubiquitination (Yang, Li, Connelly, et al. 2020). Consequently, transcript shortening was seen to be promoted including in the case of oncogenes and tumour suppressors including the aforementioned cyclin D1.

Regulation of cancer-specific APA profiles has also been linked to HNRNPC, a protein associated with mRNA stabilisation, splicing and APA regulation (Fischl et al. 2019). Binding sites for HNRNPC were enriched in the aUTRs lengthening genes of metastasis-derived cells, while HNRNPC levels were elevated in metastatic cells. Knocking down HNRNPC using siRNA reverted a subset of APA profiles towards the non-malignant state (Fischl et al. 2019).

Host APA During Viral Infection

Viruses induce major changes in the host transcriptome, through host shutoff whereby the viral life cycle is supported by a corresponding downregulation of host gene expression (He et al. 2020) and via the host response, which plays a role in suppressing the impact of infection (Sun, Ye, et al. 2020).

Jia et al. observed major changes to the APA profile when human or mice cells were infected with vesicular stomatitis virus (VSV) (Jia, Yuan, et al. 2017). When gene ontology analyses were performed on those genes that switched PASs, immune-related categories such as inflammatory response and defence response to virus infection were significantly enriched. The authors observed that most of the genes involved in virus detection (Toll-like Receptor pathway, Jak-STAT pathway and RIG-I-like receptors) switched 3'UTR PASs when cells were infected with VSV, and also saw similar effects in cells infected with Sendai virus or type-1 herpes simplex virus (Jia, Yuan, et al. 2017). Using a luciferase activity assay, it was inferred that many of these genes experience differential protein expression as a result of these 3'UTR PAS switches, and these effects were similar between cell lines although different between genes. One such example is the short 3'UTR of DDX58 (also known as RIG-I) - a major detector of viral RNA in the cytoplasm - which resulted in increased protein expression of Renilla (Jia, Yuan, et al. 2017).

A study of APA in porcine epidemic diarrhea virus (PEDV) - a virus that causes high mortality rates in pigs, causing huge economic loss - identified over 800 genes that switched PAS throughout an infection timecourse (Wei, Li, et al. 2021). Just under half (360) of these genes underwent both shortening and lengthening which was timepoint-dependent, which the authors suggest indicates

that this dynamic cohort may play an active role in host-pathogen interactions (Wei, Li, et al. 2021). Genes involved in endoplasmic reticulum (ER) processing were particularly enriched in those genes that switched PAS, suggesting that virus infection impacts ER function.

Intriguingly, a recent study highlights how a virus-induced APA change can promote oncogenesis. Wu et al. found that human papillomavirus (HPV) causes global 3'UTR lengthening of transcripts in cancer samples and that the distinct 3'UTR APA profile could be used to differentiate clearly HPV-positive from HPV-negative cancer samples (Wu, Chen, et al. 2019). Lengthening of one of the transcripts, RBM25, was found to lead to higher recruitment of FUS and DGCR8 and consequently increase expression of PD1 (Wu, Chen, et al. 2019), a key inhibitory checkpoint for CD8 T cells which are known to be compromised in HPV infection (Eberhardt et al. 2021). Differential expression was also impacted between virus-infected and uninfected samples, with almost all of the core APA regulators upregulated in HPV-positive samples (Wu, Chen, et al. 2019).

Widespread APA changes have recently been reported in HeLa cells infected with herpes simplex virus 1 (HSV-1) (Wang, Liu, et al. 2021). Over a thousand genes were identified to PAS switch 12 hours post infection, with 71% of these lengthening and the remainder exhibiting transcript-shortening. Interestingly, by employing nascent RNA-sequencing (4sU-seq), the authors report that APA changes induced by HSV-1 are at least in part co-transcriptional (Wang, Liu, et al. 2021). The authors suggest that the HSV-1 viral immediate early protein ICP27 is responsible for many of the APA-induced changes seen in HSV-1 infection via an ICP27-CPSF interaction (Wang, Hennig, et al. 2020; Wang, Liu, et al. 2021). However, HSV-1 lacking ICP27 still induced APA changes in infected cells,

suggesting that other mechanisms are involved including the viral DNA, other viral proteins and/or the host stress response (Wang, Hennig, et al. 2020).

Most recently, changes in APA profiles have been detected in COVID patients, with a general trend towards 3'UTR shortening (An, Li, et al. 2021). Intriguingly, patients with shorter 3'UTRs had higher intensive care unit admission rates than patients with longer 3'UTRs, suggesting clinical implications for APA changes in viral infections. The same study reported an enrichment of innate immune response processes, such as neutrophil activation in those genes changing APA profiles, and proposed that APA was involved in regulating the innate immune response in COVID patients (An, Li, et al. 2021).

While it is clear that virus infection induces multi-faceted APA changes in the host transcriptome, and that these changes have physiological effects, the mechanisms behind these changes and the downstream effects of alternative PAS usage remain mostly unknown.

1.3.5 Techniques used to study APA

Since the discovery that APA is a widespread phenomenon important in establishing the transcriptome, numerous technologies and methods have been developed to study it. In the beginning, single-gene level APA research was employed mainly in the 1980s. The classic example of this is the previously-discussed seminal study describing the coding-region APA switch in IgM (Alt et al. 1980). With the development of technology, genome-wide APA study became possible. Here, we will discuss some of these major genome-wide methods.

cDNA and ESTs

The first genome-wide tools used to study APA employed cDNA sequences and expressed sequenced tags (ESTs) (Zheng and Tian 2014). This method was ideal for APA research, as a large proportion of cDNA/EST sequences are sourced from the 3' end of an mRNA (Tian, Hu, et al. 2005; Gautheret et al. 1998). This technique could be used to identify PASs by using cDNA/EST sequences deposited in Genbank or dbEST databases. The sequences were aligned to the genome and consecutive As at the 3' end were deduced to be at the poly(A) tail (Zheng and Tian 2014). This method was used to first show that genes having more than one PAS is a widespread phenomenon (Beaudoing et al. 2000; Tian, Hu, et al. 2005).

While cDNA and EST databases were invaluable to investigate APA at beyond the single-gene level in the early stages of the field, they were not without shortcomings. The most problematic was that a relatively small ($\sim$10 %) of ESTs contain poly(A) tail information (Gilat et al. 2006), meaning that APA research using ESTs was limited to only a low proportion of genes (Zheng and Tian 2014).

Microarrays

Microarray technology, used to quantify gene expression levels, is also appropriate for APA research. Research using array data generally focuses on tandem 3' UTR APA. The Affymetrix 3' microarray probes target UTR regions, and this property was exploited by Sandberg et al. in a method the authors called "PLATA" to compare transcripts in CD4-expressing resting mouse T cells with stimulated cells (Sandberg et al. 2008). A tandem UTR length index (TLI) was determined which represented the length of all tandem UTRs in a sample. The TLI estimates usage

of proximal or distal PASs based relative intensities of the probe of the tandem UTR, with a high TLI demonstrating increased usage of the distal PAS and a low TLI indicating increased usage of the proximal PAS. Similar methods were used to investigate 3'UTR length in during cell differentiation and development (Ji and Tian 2009) and between different tumour types (Singh, Alley, et al. 2009).

Despite the advancements in APA research that microarray technology allowed, it has limitations which restricts its usage. Most microarray APA studies are limited to analysing 2000 genes or less due to the low number of genes that have probes targetting both cUTR and aUTR. As well as this, microarrays cannot be used to identify PASs *de novo*, which is a problem when two or more PASs are close to each other (Zheng and Tian 2014).

RNA-seq technology

The advancement of next-generation sequencing technology led to the rapid buildup of sequencing data that is primed for APA analysis. Many RNA-seq protocols aim to sequence the whole transcript, therefore reducing the amount of reads found in the 3' region. To address this problem, RNA-seq reads can be enriched for the 3' end of transcripts, for example using oligo(dT) priming. The advent of single cell RNA-seq (scRNA-seq) also enables powerful analyses of APA at the single-cell or single-cluster level to be performed, but introduces analytical challenges. This section will discuss some examples of bioinformatic methods that have been developed to study APA on a comprehensive scale using RNA-seq data.

DaPars is a widely-used tool that identifies and quantifies dynamic APA events using standard RNA-seq data (Xia et al. 2014). It identifies a distal PAS, and employs a regression model to infer the position of the proximal PAS (Xia et al.

2014). DaPars exploits the principle that if RNA-seq read coverage is equal along the length of a transcript, a PAS would display as a distinct drop in sequenced reads. Consequently, one of the major innovations of DaPars was the ability to use existing RNA-seq data to determine APA events without needing any other experiments. However, reads are not uniform along gene loci and this increases false positives. Another key limitation of DaPars is that it cannot detect dynamic APA switching involving more than two poly(A) sites, as it only examines expression of a distal and proximal PAS.

APAtrap, like DaPars, uses RNA-seq data but attempts to predict all potential poly(A) sites in a gene (Ye, Long, et al. 2018). It employs a mean-squared error model for site prediction and has, unusually for tools in its class, been evaluated in *Arabidopsis* (as well as human), demonstrating flexibility between organisms. APAtrap is capable of identifying 3'UTR extension downstream of 3'ends, meaning that it can predict PASs in so-far unannotated regions. One limitation of APAtrap is that it does not take into account sample replicates, and instead averages the replicate values to form a contingency table.

QAPA is a tool that infers APA using a reference-based approach (Ha, Blencowe, and Morris 2018). It uses estimates of alternative 3'UTR expression combined with a database of annotated poly(A) sites to assess APA, and is fast and sensitive when compared to some other tools, including DaPars. A benchmarking study that compared several APA computational tools designed for standard RNA-seq found that QAPA was a standout in its ability to reliably quantify APA when paired with 3'-seq or Iso-seq derived PAS annotations (Shah et al. 2021).

DEXSeq (Anders, Reyes, and Huber 2012) is a statistical method originally designed for differential exon analyses based on the popular DESeq2 tool. As

DEXSeq analyses differential poly(A) site usage and not PAS discovery, an annotated PAS database is required as input to ensure appropriate selection of PASs by the tool. An example of this database is PolyA_DBv3, available for multiple organisms, which catalogues poly(A) sites aggregated by the 3' region extraction and deep sequencing (3'READS) method (Wang, Nambiar, et al. 2018). This method is notably not affected by internal A-rich sequences, addressing the internal priming issue that APA analysis often faces. One significant advantage of DEXSeq is that it offers reliable control of false discovery rates by accounting for biological variation between samples (Anders, Reyes, and Huber 2012).

APAlyzer is another tool that uses PASs annotated in the PolyA_DBv3 database (Wang and Tian 2020). It determines APA events (notably both 3'UTR and intronic APA) in RNA-seq data but does not identify *de novo* poly(A) sites. APAlyzer uses Fisher's exact test for differential PAS usage when no replicates are available, or a t-test during replicate design. Unlike some other tools, APAlyzer outputs several plots which aid in the visualisation of the APA profile.

LABRAT is a recently published quantification method for APA isoforms (Goering et al. 2021). LABRAT is capable of analysing both traditional RNA-seq data and also 3' end sequencing data, for which there is an inbuilt "librarytype" parameter. A distinguishing feature of LABRAT to other APA-quantifying tools is the use of Salmon (Patro et al. 2017), a novel transcript quantification method, which aids the tool's determination of relative PAS usage within genes. LABRAT classifies genes as having a tandem UTR structure or ALE (alternative last exon) structure, which allows comparisons between the two groups (Goering et al. 2021).

Several tools have also been developed recently to study alternative PAS usage in scRNA-seq data, although it is still an emerging branch of APA research. The

vast majority of scRNA-seq data is not generated with APA specifically in mind. However, because many scRNA-seq protocols are 3' end-linked, reads can contain information about poly(A) site positions making them suitable for APA analysis. The use of single-cell data with one novel tool, **scAPA**, demonstrated global 3' UTR shortening in T cells that were in the proliferative state (Shulman and Elkon 2019). scAPA carries out *de novo* PAS prediction within predicted peaks and performs differential PAS usage on single-cell clusters as input, using the chi-squared test. Another tool, **MAAPER** (Li, Zheng, et al. 2021), incorporates annotated PASs from the PolyA_DBv3 (Wang, Nambiar, et al. 2018) to achieve PAS identification. This gives MAAPER the capability of identifying PASs with single-nucleotide resolution, but limits its usage to the four organisms that PolyA_DBv3 is currently available for (human, mouse, rat and chicken). MAAPER, like scAPA, compares APA between cell clusters, but also offers a paired test to account for inter-sample variation. MAAPER uses a likelihood-ratio test to test if the overall PAS profile of a gene changes between samples and also a Fisher's exact test to establish if two specific PASs differ in isoform abundance between samples (Li, Zheng, et al. 2021).

1.4 Influenza

1.4.1 Introduction to Influenza A

The influenza A virus (IAV) is the causative agent of influenza, an infectious respiratory disease that affects 5-15% of the population of industrialised countries (Girard et al. 2005). In humans, symptoms can include fever, coughing, sore

throat, runny nose, headache, fatigue, and muscle aches. Infection can occur in any age group, though the risk of hospital admission and/or death is much higher in the elderly (Javanian et al. 2021). Pneumonia is the most common complication and is associated with high mortality rates, especially among the elderly and at-risk groups. IAV infection results in approximately 300,000-500,000 deaths annually (Girard et al. 2005), and brings with it a considerable global financial cost (Molinari et al. 2007). Influenza outbreaks have occurred at different times throughout history with varying global impacts; the most significant being the 1918-1919 pandemic which caused 50-100 million deaths (Taubenberger and Morens 2006). Waterfowl act as the main reservoir for influenza A, but the virus can also infect a wide variety of other animals, including poultry and terrestrial mammals (Edinger, Pohl, and Stertz 2014).

IAV belongs to the *Orthomyxoviridae* family of viruses and has a 13.5kb genome that is made up of eight single-stranded, negative sense segments of RNA ranging in size from approximately 0.9kb to 2.3kb (Lee, Le Sage, et al. 2017). These are termed viral RNAs (vRNAs), and encode multiple proteins (Ruigrok et al. 2010). Two surface glycoproteins (HA and NA) control the ability to begin the infectious cycle and release progeny virus, respectively (Gamblin and Skehel 2010). Two matrix proteins (M1 and M2) are involved in supporting the viral envelope and acidifying the viral interior which promotes the release of the viral genome into the cytoplasm (Zhou, Zhang, et al. 2022; Eisfeld, Neumann, and Kawaoka 2015). Three proteins make up the IAV polymerase (FluPol); PB1, PB2 and PA. NP coats the vRNAs and is an essential protein for viral transcription and replication (Fodor 2013). Non-structural protein 1 (NS1) is multi-functional and is perhaps the main actor against host immunity (Hale, Randall, et al. 2008; Ji, Wang, and Liu 2021)

while nuclear export protein (NEP) is involved with vRNP export from the host cell nucleus (Bouvier and Palese 2008). As well as these ten classical proteins, IAV can encode for additional proteins that are the result of alternate reading frames. One example is PA-X, which is produced from a ribosomal frameshift during translation of the PA mRNA (Jagger et al. 2012). Table 1.2 summarises the primary IAV proteins in addition to PA-X and some of their key functions. Inside the host-derived envelope lie the eight NP-coated vRNAs, which are each in complex with a trimeric FluPol. Taken together, each vRNA-NP-FluPol subunit is known as a viral ribonucleoprotein (vRNP) (Eisfeld, Neumann, and Kawaoka 2015). A diagram of an IAV virion and a vRNP can be seen in Figure 1.5.

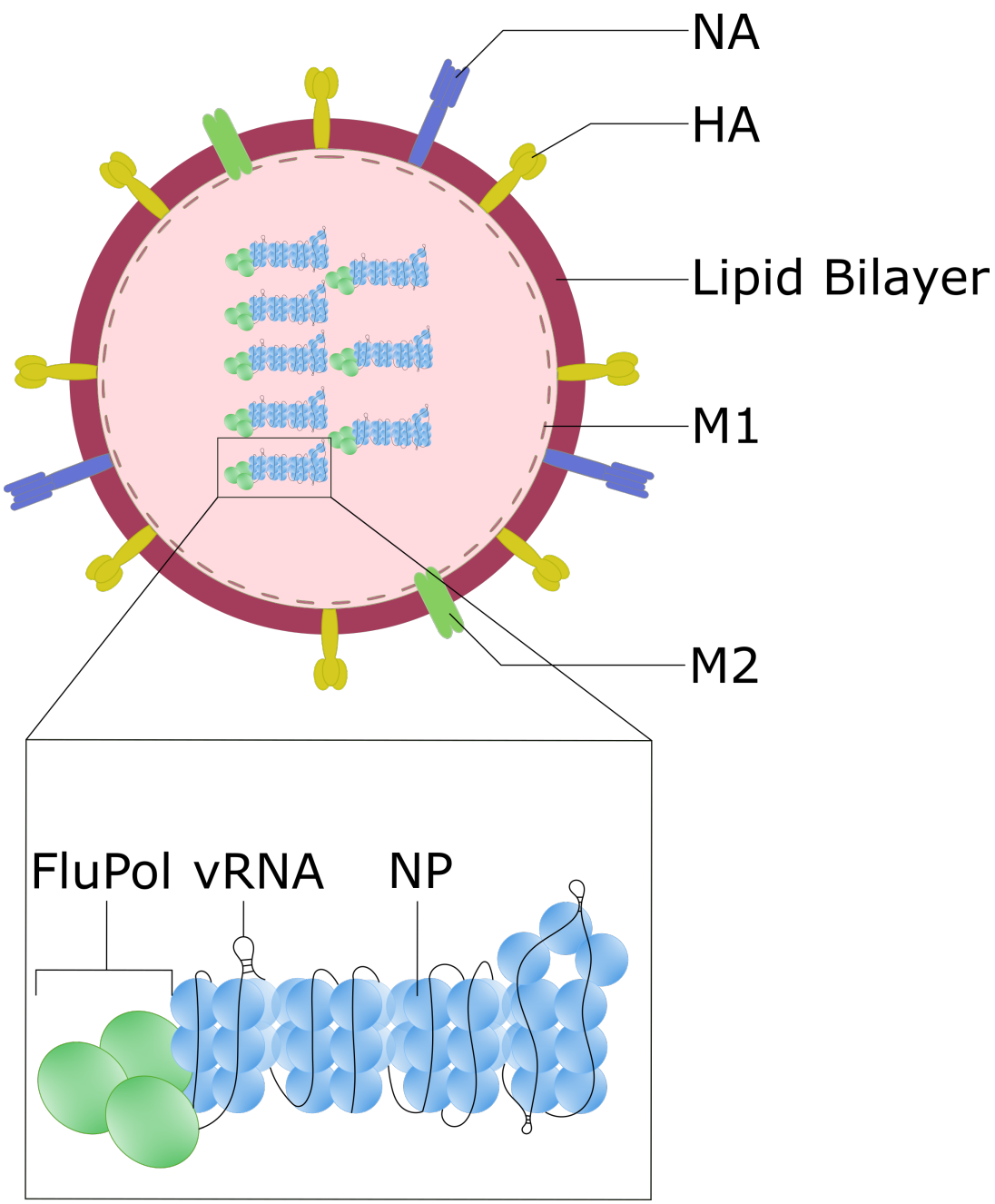


Figure 1.5: Schematic of an influenza virion. Surface glycoproteins (HA, NA and M2) are embedded in a host-derived lipid bilayer. Underneath, the M1 matrix protein provides structural support to the viral envelope. The influenza genome consists of eight negative-sense single stranded RNA segments, each of which are coated by the influenza NP protein to form a vRNP (expanded version shown). Both terminis of each vRNA segment are bound to a trimeric polymerase.

IAV Protein	**Function**
FluPol	*Transcription and replication of the viral genome*
PB1	Centre of FluPol complex, contains RNA synthesis catalytic residues
PB2	Multiple domains; Critical for transcription initiation
PA	Cap-snatching with PB2, Endonuclease activity
References	(Te Velthuis, Grimes, and Fodor 2021) (Fan et al. 2019; Fodor and Te Velthuis 2020)
Surface Glycoproteins	*IAV Classification Based on Subtypes*
HA	Homotrimer;Binds sialic acid on host receptor, critical for cell entry (Gamblin and Skehel 2010)
NA	Required for viral release from cell (Laborda, Wang, and Voglmeir 2016)
Matrix Proteins	*Mixed*
M1	Structural, promotes incorporation of IAV proteins into virions (Peukes et al. 2020)
M2	Ion channel; Activates endocytosed virus via acidification (Samji 2009; Liang et al. 2016)
Non-structural Proteins	*Modulate host response*
NS1	Inhibit host immune response, limit host polyadenylation (Hale, Randall, et al. 2008)
PA-X	Accessory protein; Host shut-off (Khaperskyy, Schmaling, et al. 2016)
Other	
NEP	Initially believed non-structural; Essential for replication and important for nuclear export (Iwatsuki-Horimoto et al. 2004)
NP	Critical component of vRNPs

Table 1.2: Classical Influenza A Proteins and Key Functions

1.4.2 Life Cycle of Influenza A

Entry into the host cell

IAV infection begins with attachment by the HA viral surface glycoprotein to cell surface proteins that contain sialic acid (Figure 1.6, 1) (Gamblin and Skehel 2010). This interaction triggers endocytosis of the virion. Once the virus is inside the endosome it is trafficked towards the nucleus with the low pH environment activating the viral M2 ion channel (Figure 1.6 (2)) (Rust et al. 2004). The activation and subsequent opening up of M2 causes further acidification of the viral interior, leading to vRNP release from the M1 matrix protein to which they are associated (Pinto and Lamb 2006). Concurrently, host cell proteases cleave the HA glycoprotein into two subunits, HA1 and HA2 (White and Whittaker 2016). Once released, the acidic environment causes the fusion peptide on the N-terminal of HA2 to become exposed (Bullough et al. 1994). Now activated, the fusion peptide of HA2 becomes embedded in the endosomal membrane while the C-terminus remains anchored in the viral membrane, resulting in fusion of the two membranes (Figure 1.6, 3). Upon fusion, the vRNPs are released into the cytoplasm. The newly-released vRNPs are thought to use the importin α-importin β pathway to enter the nucleus (Wang, Palese, and O'Neill 1997; Miyake et al. 2019). This is facillitated by the nuclear localisation signals present on the vRNP NP proteins that are exposed upon M1-removal (Hutchinson and Fodor 2013). The vRNP is delivered to the nuclear pore, where it is translocated into the nucleoplasm (Dou et al. 2018). Unusually for RNA viruses (Dou et al. 2018), the nucleus is the site of vRNA transcription and replication (Figure 1.6, 4), and both processes are carried out by FluPol (Fodor 2013; Fodor and Te Velthuis 2020).

Transcription (Including 5' Capping and 3' Polyadenylation)

FluPol transcribes IAV's negative-sense vRNA segments into positive-sense 5' capped and 3' polyadenylated mRNAs (Fodor 2013). Viral mRNAs are structurally identical to host mRNAs, so they are ready to be exported to the cytoplasm using endogenous host pathways and capable of translation by cellular machinery (Walker and Fodor 2019). **Transcription is initiated** with a process termed "cap-snatching" (Krug, Broni, and Bouloy 1979) and is dependent on RNA Pol II activity since a 5' capped host mRNA is used as a primer. FluPol first binds the CTD of RNA Pol II, bringing the nascent host 5' capped mRNA in close contact with the vRNP (Engelhardt, Smith, and Fodor 2005; Martin, Hengrung, et al. 2018). The endonuclease activity of the PA subunit of FluPol cleaves the host mRNA 10-14 bases downstream from the 5' terminal N7-methyl guanosine (m^7G)cap (Fodor and Te Velthuis 2020). Completing the cap-snatching procedure, the resulting short 5' capped RNA molecule - now a primer - base pairs with the vRNA template and is subsequently extended (Fodor and Te Velthuis 2020; Reich et al. 2014).

RNA synthesis occurs in the PB1 subunit of FluPol, which contains the catalytic core (Wandzik, Kouba, and Cusack 2021). Nucleotides for extensions are chosen based on their Watson-Crick complementarity to the template at a position in the active site called the +1 position. FluPol has multiple surface binding sites for the 3' vRNA promoter whose functions are mostly unknown. A recent study however suggests that one of these (mode B), is necessary for complete transcription (Walker, Sharps, and Fodor 2020). The suggested model proposes that the binding of the 3' vRNA promoter to FluPol once transcription has begun

enhances the structural integrity of the vRNP which may be required for FluPol to successfully move along the vRNA template (Walker, Sharps, and Fodor 2020). Splicing of IAV mRNA can occur co-transcriptionally via the same spliceosome-mediated mechanism that host pre-mRNA are subject to. For example, the M segment of IAV which gives rise to the M1 matrix protein can be spliced, ultimately leading to the generation of the M2 ion channel protein (Lamb and Choppin 1981).

Following extension, **termination of transcription** occurs with the "stuttering" of FluPol when it reaches a 5' proximal oligo(U) motif, leading to polyadenylation of the emerging mRNA (Robertson, Schubert, and Lazzarini 1981; Poon et al. 1999; Zheng, Lee, et al. 1999). The polyadenylation signal always begins at position 17 from the 5' end of the vRNA template, and consists of five or six consecutive uridines (Wandzik, Kouba, Karuppasamy, et al. 2020). While this process is not fully understood, a recently proposed model suggests that the uridine at position 17 slips in and out of the FluPol active site +1 position, and between every ATP incorporation the template oscillates between 9 and 10nt in length (Wandzik, Kouba, Karuppasamy, et al. 2020).

Once completed, the viral mRNAs are exported from the nucleus by the host NXF1 pathway and undergo translation in the cytoplasm or endoplasmic reticulum (Figure 1.6, 5) (York and Fodor 2013). Nuclear localisation signals on newly-synthesised viral proteins target them for nuclear import. Once in the nucleus, these proteins including NP, PB1, PB2, PA and NS1 contribute to viral transcription, replication and host shutoff (Dou et al. 2018).

Replication

To replicate the vRNA segments, two steps are required which are both carried out by FluPol. First, the negative sense vRNA must be converted to a full length, positive-sensed complementary copy which is termed cRNA ("complementary RNA"). Following this, the cRNA intermediate molecule is used as a template for vRNA synthesis (Fodor and Te Velthuis 2020). To begin the *de novo* synthesis of cRNA, the first two nucleotides at the 3' end of the vRNA template (UC) are used to transcribe a pppApG dinucleotide (Te Velthuis, Robb, et al. 2016; Fodor and Te Velthuis 2020). This process, termed "terminal initiation", is dependent on what is known as the priming loop, a loop structure that protrudes into the FluPol active site (Te Velthuis, Robb, et al. 2016). The pppApG dinucleotide is then extended to generate a full length cRNA. The nascent cRNA emerges from the viral polymerase, and assembles into a cRNP structure, which is a complex analogous to a vRNP (Fodor and Te Velthuis 2020). It has been commonly reported in the literature that cRNPs are not exported from the nucleus, enabling incorporation of solely vRNPs and not cRNPs into new virions (Tchatalbachev, Flick, and Hobom 2001; Veler et al. 2021).

Next, the new cRNA molecule acts as a template for vRNA synthesis, a process which is dependent on dimerisation of FluPol (Fan et al. 2019). FluPol produces a pppApG dinucleotide at the 4th and 5th cRNA positions, which is used as a primer. For vRNA synthesis to proceed, a second "regulatory" polymerase must dimerise with the "resident" polymerase. It is believed that this lock on vRNA synthesis is a mechanism for temporal control - vRNA production can only occur when there is enough newly-made viral polymerase accessible in the cell (Fan et al.

2019).

Interestingly, work in recent years has shown that two host factors, ANP32A and ANP32B, are essential for influenza FluPol activity (Zhang, Zhang, et al. 2019; Staller et al. 2019). It has also been demonstrated that species-specific differences in ANP32A have been shown to underpin the low activity of avian FluPol in mammalian cells (Long et al. 2016). Shedding light on this host-virus relationship, a cryo-electron microscopy study by Carrique et al. demonstrated a model whereby two FluPol molecules dimerise in an asymmetric fashion with the N-terminus of ANP32A acting as a bridge during both steps of replication (Carrique et al. 2020). This FluPol-ANP32A complex is believed to act as a replication platform for the IAV genome where one of the constituent FluPol molecules functions as a replicase, while the other assembles the viral genome (Carrique et al. 2020).

Export and trafficking of vRNPs

Progeny vRNPs are translocated to the cytoplasm via the cellular CRM1-dependent nuclear export pathway (Fodor 2013). The viral NEP protein interacts with vRNPs via the M1 protein, and directly associates with CRM1 (Paterson and Fodor 2012). CRM1 recognises NEP's nuclear export signal and facillitates export of the vRNP to the cytoplasm (Dou et al. 2018). Once in the cytoplasm, the vRNPs are trafficked to the plasma membrane for assembly in a Rab11-dependent manner (Castro Martin et al. 2017). Host factor Rab11a has been shown to be critical for vRNP trafficking and efficient IAV assembly, with loss of Rab11a shown to increase generation of non-infectious particles (Han et al. 2021). Veler et al. demonstrated that Rab11a specifically interacts with vRNPs and not cRNPs, and that FluPol associates directly with Rab11a via its PB2 C-terminal and nuclear localisation

signal domains (Veler et al. 2021).

IAV Assembly and Budding

IAV is not fully infectious unless all eight of its genome segments are present (Figure 1.6, 6). The mechanism that controls this assembly process are not fully understood, although it is not thought to be random as the majority of progeny virions contain all eight segments (Fournier et al. 2012; Lakdawala et al. 2014; Le Sage et al. 2020). One model for this process suggests that correct assembly is achieved by RNA-RNA and RNA-protein interactions between segments (Hutchinson, Kirchbach, et al. 2010; Dadonaite, Gilbertson, et al. 2019; Fournier et al. 2012; Li, Gu, et al. 2021).

The IAV membrane is host-derived, but when compared with the plasma membrane, it is enriched for cholesterol and sphingolipids indicating that virions bud from the distinct apical membrane regions known as rafts (Gerl et al. 2012). Here, the plasma membrane has become enriched with the viral transmembrane proteins HA, NA and M2 which have journeyed from their sites of translation and post-translational modification: the endoplasmic reticiulum membrane and Golgi apparatus, respectively. The exact budding process is not fully-understood, but the IAV virion induces membrane curvature and achieves scission, most likely with the activity of HA, NA, M1 and M2 (Figure 1.6, 7) (Dou et al. 2018). The final step that releases the virion is the cleavage of sialic acid residues from glycoproteins and glycolipids, breaking the binding of HA to the cell surface. This is carried out by NA, and enables more than one round of replication to occur (Bouvier and Palese 2008; Webster and Laver 1967).

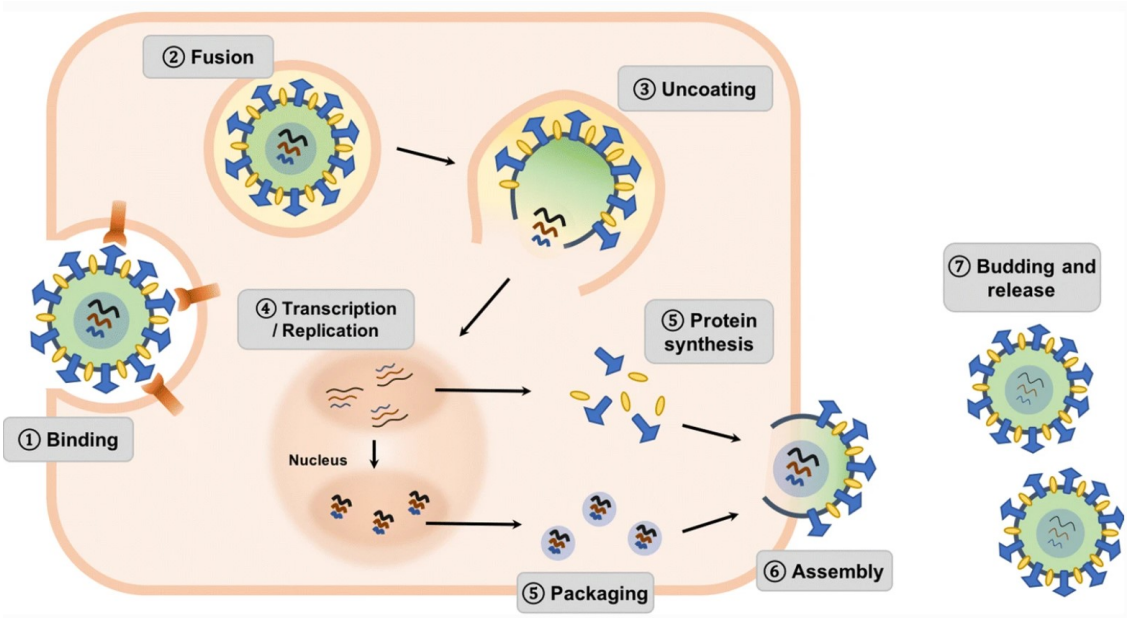


Figure 1.6: Influenza viral life cycle. The influenza virus particle binds sialic acid residues on the cell surface. The particle is then endocytosed, with the low pH environment triggering fusion of the viral and endosomal membranes. The newly-released vRNPs then uses the importin pathway to enter the nucleus, where transcription and replication of the viral genome occurs. Newly-made vRNPs are translocated to the cytoplasm where viral proteins are synthesised. Finally, vRNPs and viral proteins are packaged together at the cell membrane and viral progeny bud into the extracellular space. Figure taken from Ampomah and Lim 2020.

1.4.3 Interplay between IAV and host systems

Influenza A, like all viruses, employ multiple strategies to hijack host machinery and evade host immunity. One of these approaches aims to increase viral mRNA translation by reorganising host gene expression, in a process termed "host shut-off". Two of the major IAV proteins involved with host shutoff are NS1 and PA-X and both play major roles in the regulation of host gene expression (Chaimayo et al. 2018; Jagger et al. 2012; Hale, Randall, et al. 2008; Nogales, Martinez-Sobrido, Topham, et al. 2018), while FluPol has also been implicated in modulating host

RNA processing.

NS1

The NS1 protein of IAV is most commonly 230 amino acids in size, dependent on strain, and derives from the eighth genome segment (Hale, Randall, et al. 2008). NS1 is expressed at very high levels in infected cells (Shaw and Palese 2007), and its modulation of the immune system was first demonstrated in an early study which showed that the viral protein is required for viral replication only in interferon-deficient cells (García-Sastre et al. 1998). Since then, NS1 has become one of IAV's most characterised proteins and is now known to interact with over 50 host proteins (Nogales, Martinez-Sobrido, Topham, et al. 2018). One of these interactions, with CPSF30, allows IAV to directly interfere with host polyadenylation (Nemeroff et al. 1998). Binding of NS1 to CPSF30 hinders the ability of CPSF30 and the CPSF complex to recognise the poly(A) signal, thereby inhibiting or altering 3' maturation of host mRNAs (Nemeroff et al. 1998; Twu et al. 2006; Chan, Huppertz, et al. 2014; Ji, Wang, and Liu 2021). Notably, since polyadenylation of IAV viral RNAs is independent of the host cell 3'end machinery and is carried out by the viral polymerase (Poon et al. 1999), viral mRNAs are not impacted by this effect.

Das et al. reported the crystal structure of the C-terminal domain of NS1 (from the influenza A/Udorn/72 strain) associating with the second and third zinc fingers (ZF2-ZF3) of CPSF30 (Das et al. 2008). A later study showed that ZF2-ZF3 of CPSF30 are crucial for poly(A) signal recognition (Sun, Zhang, et al. 2018). Specifically, a tetramer is formed with each of the two zinc fingers interacting with a single NS1 protein. The residues involved in this interaction are

highly conserved amongst human IAVs, although there are exceptions (García-Sastre 2011). For example, the NS1 of the 2009 H1N1 pandemic swine flu lacked the ability to bind CPSF30 (Hale, Steel, et al. 2010), while mutations at three residues involved in the CPSF30-NS1 binding region restored this interaction. A recent study which compared groups of mice infected with IAV strains which did or did not have the capacity for NS1-CPSF30 interactions shows that the ability of NS1 to bind CPSF30 confers a replicative advantage to the virus (Mazel-Sanchez et al. 2021). This study also suggested a novel role for NS1 in the fine-tuning of the endoplasmic reticulum stress response to IAV's advantage by restricting host expression of ER stress response factors (Mazel-Sanchez et al. 2021).

A hallmark of IAV infection that has come to light in recent years is defective RNA Pol II transcription termination. Some studies claim that this effect is NS1-induced (Heinz et al. 2018; Zhao et al. 2018; Nacken et al. 2021), while another reports that it is independent of the NS1-CPSF30 interaction and instead largely due to cellular stress (Bauer et al. 2018). The difference in findings could be explained by the use of different IAV strains, suggesting that IAV subtypes have evolved subtle differences in the binding affinities of their NS1 proteins to host transcription factors (Nacken et al. 2021).

PA-X

PA-X is a viral protein involved in host shutoff that is translated as a +1 shift in the open-reading frame of the PA mRNA (Jagger et al. 2012). While translating, the ribosome shifts at a U-rich sequence near a rare codon (which promotes ribosomal frameshifting) (Firth et al. 2012). PA-X and PA share the same N-terminus, but PA-X contains a truncated C-terminus, with PA-X most commonly being 252

amino acids in size (Jagger et al. 2012).

PA-X protein levels are notably low in infected cells, especially when compared to NS1 (Jagger et al. 2012). This low expression level is likely partly explained by its ribosomal-shifting translation requirement. However, a recent study reports that the PA-X protein is rapidly turned over, and that this contributes to its regulation, with a region in the C-terminal being markedly unstable (Levene, Shrestha, and Gaglia 2021). The half-life of the PA-X protein differs between strains, with the H1N1 2009 pandemic strain PA-X having the longest half-life of those investigated (Levene, Shrestha, and Gaglia 2021). Interestingly, major host shutoff activity was seen in this strain (Nogales, Martinez-Sobrido, Chiem, et al. 2018; Desmet et al. 2013) despite the NS1 of this pandemic strain not inhibiting gene expression (Hale, Steel, et al. 2010).

Regardless, despite the low expression levels of PA-X, the viral protein is highly conserved and is present in all IAV strains (Shi et al. 2012). It has a wide-ranging ability to induce host shutoff, exceeding that of NS1 *in vitro* (Hayashi et al. 2016; Chaimayo et al. 2018). Both its N-terminal endonuclease domain (shared with PA) (Desmet et al. 2013; Khaperskyy, Emara, et al. 2014), and its unique C-terminal domain (Hayashi et al. 2016; Oishi, Yamayoshi, and Kawaoka 2015; Bavagnoli et al. 2015) contribute to this activity. IAVs modified to exclude PA-X generate more powerful inflammatory and immune responses in mice, chickens and pigs (Gao, Sun, et al. 2015; Gong et al. 2017; Hu, Mo, et al. 2015; Xu et al. 2017) - possibly because of decreased host shutoff. Indeed, a recent study demonstrated that PA-X interacts with host factor ANKRD17, a positive regulator of the RIG-I immune response (Li, Qi, et al. 2021). This interaction - dependent on the X domain and therefore specific to PA-X and not PA - was found to dampen the

host immune response to IAV RNA (Li, Qi, et al. 2021).

While PA-X is also found in the cytoplasm, it is enriched in the nucleus (Hayashi et al. 2016; Gaucherand et al. 2019; Li, Qi, et al. 2021). It selectively degrades host mRNA via a mechanism mediated by the host exonuclease XRN1, while ignoring viral mRNAs (Khaperskyy, Schmaling, et al. 2016). PA-X activity also seems to be linked to 3'-end RNA processing (Khaperskyy, Schmaling, et al. 2016) - suggesting that, as with NS1, IAV exploits the differing 3'-end processing pathways to protect its own mRNA. While NS1-induced shutoff more specifically targets the host immune response, PA-X is believed to contribute to a wider general shutoff (Dunagan, Hardy, and Takimoto 2021). Interestingly, a recent study also reports that the CFI complex - a critical component of the host polyadenylation machinery - regulates PA-X activity (Gaucherand et al. 2019).

FluPol

The viral polymerase itself has the potential to inhibit cellular RNA transcription. As mentioned previously, FluPol associates with host RNA Pol II when "cap-snatching" so it can remove the 5' cap from nascent cellular mRNAs and use it as a primer for the transcription of its own viral mRNAs. This interaction has implications for host transcription - it results in aberrant termination, thereby decreasing the amount of cellular mRNAs transcribed (Chan, Vreede, et al. 2006; Bauer et al. 2018). Bauer et al. recently reported a widespread loss of RNA Pol II gene occupancy in IAV-infected cells (Bauer et al. 2018). The authors of this study suggest that it is the FluPol cap snatching activity that is ultimately responsible for this phenomenon. To date only the interaction between FluPol and the C-terminal domain of RNA Pol II is known, but further interactions are likely

(Walker and Fodor 2019). Indeed, FluPol is known to associate with the RNA exosome complex which probably helps target FluPol to the gene promoter where it can cap snatch (Walker and Fodor 2019; Rialdi et al. 2017).

1.5 Thesis objective and outline

Is host alternative cleavage and polyadenylation altered by virus infection, and if so, how? Furthermore, to what extent do changes to the host APA landscape contribute physiologically to the virus infection and defence response of the host?. This thesis aims to contribute to our answering of these questions.

APA is a commonly-used gene regulation mechanism in all eukaryotes, and has been previously found to play a role in altering the expression patterns of proteins involved in virus detection during viral infection. Therefore, improving our understanding of how APA is utilised by cells to respond to infection with the epidemiologically-relevant influenza virus could uncover therapeutic targets for the development of an influenza antiviral drug.

To investigate host APA during influenza infection, human lung epithelial cells are infected with one of two different strains of influenza A, with the cellular RNA being extracted from the nuclei and cytoplasms separately at different points during infection. Deep sequencing is then utilised in combination with a bioinformatics approach to characterise changes in the host APA landscape. APA is recognised, as discussed in the above introduction, to be an extensive regulator of gene expression capable of inducing physiological changes, while the role of host APA during viral infection is poorly understood. The objective of this study is to quantify the extent of APA changes, identify the physiological implications of these poly(A)

site switches and explore the mechanisms behind switching poly(A) sites during viral infection.

Chapter 2

Materials and Methods

2.1 Cells

Adenocarcinomic human alveolar basal epithelial (A549) cells, Madin-Darby bovine kidney (MDBK) cells, and Madin-Darby canine kidney (MDCK) cells were sourced from the Cell Bank of the Sir William Dunn School of Pathology, University of Oxford.

2.2 Cell culture

A549 cells, routinely used to study influenza replication (Ujie et al. 2019), were used as a model to study host APA profiles during influenza infection, while influenza particles were propagated in commonly-used MDCK (Hamamoto et al. 2013) and MDBK (Dadonaite, Gilbertson, et al. 2019) cultures. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 10% foetal calf serum (FCS). Thrice a week, cells were passaged by 1:10 dilution from a confluent

density of 1x10^7 cells ml^{-1} using TrypLE express to detach the cell monolayer from the flask surface. All cells were grown and maintained in incubators at 37 degrees Celsius and 5% CO_2.

2.3 Influenza virus

Influenza A/Udorn/72 (H3N2) and influenza A/WSN/33 (H1N1) were generously supplied by the Fodor Lab.

2.3.1 Harvest of influenza A/Udorn/72 (Udorn) and influenza A/WSN/33 (WSN)

Mammalian cells (MDBK or MDCK) were washed with PBS and incubated with wildtype Udorn or WSN virus stocks (gifts from Ervin Fodor lab, Sir William Dunn School of Pathology, University of Oxford) at an MOI of 0.01 or 0.001 in the presence of MEM (0.5% FCS and 1% L-glutamine). Cells were incubated with the virus stocks for 1 hour at room temperature, with gentle rocking every 10 mins. For cells infected with Udorn, 0.5 µg of trypsin. Cells were then incubated at 37 degrees Celsius for 48-72 hours, at which point the virus was harvested by centrifugation at 4,000xg for 10 minutes at 4 degrees Celsius. Supernatants containing harvested virus were stored at -80 degrees Celsius or immediately subjected to purification.

2.3.2 Purification of influenza A virus

Harvested virus was transferred to thick-walled ultracentrifuge tubes. Virus was spun at 10,000 rpm for 15 mins at 4 degrees Celsius in a SW32 rotor and Optima

XPN-80 ultracentrifuge (Beckman Coulter). Supernatants were layered carefully onto a 30% sucrose cushion in virus resuspension buffer (5mM CaCl2, 100mM NaCl and 20mM Tris-HCl, pH 7.4) and pelleted by ultracentrifugation at 25,000 rpm for 90 mins at 4 degrees Celsius. Supernatant was then aspirated and the virus pellet was resuspended (using virus resuspension buffer) by vigorous pippetting and overnight incubation at 4 degrees Celsius. Collected virus was stored at -80 degrees Celsius.

2.3.3 Plaque assay

Viral titres were determined by plaque assay. Monolayers of MDBK or MDCK cells (MEM, 0.5% FCS and 1% L-glutamine) grown in 6-well plates were washed with PBS and infected with 200 µL of serially diluted virus for 1 hour at room temperature with gentle rocking every 5 mins to avoid drying out of cells. Virus was then aspirated from monolayers and 2mL of an agarose-MEM mixture (1% agarose) was layered atop each well and allowed to cool at room temperature. Once cooled, the plates were incubated at 37 degrees Celsius for 48-72 hours, at which point the agarose-MEM overlays were removed and cells were stained with Coomassie blue dye (0.1% Coomassie R250, 10% acetic acid, 40% methanol) for visualisation of plaques.

2.3.4 Infection of mammalian cells with influenza A

Purified influenza A/Udorn/72 or influenza A/WSN/33 virus was incubated with confluent monolayers of A549 cells in 15cm dishes at an MOI of 5 in DMEM (0.5% FCS) for 1 hour at room temperature. Subsequently, the media was aspirated

and replaced with DMEM and cells were incubated at 37 degrees Celsius. At various timepoints, described below, RNA was extracted from the infected cells. In parallel to the infections, a "mock" infection was set up to be used as a control representing uninfected cells. This control sample consisted of A549 cells and was not infected with influenza A virus, but was prepared in the same way as those to be infected. At 0hpi, RNA from the cells of this "mock" control sample was extracted.

For the Udorn infection, the full infection timecourse entailed a "mock" infection, as described above, and individual samples collected at three, six, twelve and eighteen hours post-infection.

For the WSN infection, two biological replicates of a full infection timecourse was carried out. Each biological replicate included a "mock" infection sample, as described above, and individual samples each collected at three, six, nine, twelve and eighteen hours post-infection.

2.3.5 Immunofluorescent microscopy of virus-infected cells

Sterilised 13mm coverslips were placed in each well of a 24-well plate, and A549 cells were grown atop them in the presence of DMEM (10% FCS) until confluent. Cells were washed with PBS and infected with virus at an MOI of 5. At various timepoints, cells were washed twice with PBS and fixed with 4% paraformaldehyde. Permeabilisation of cells was achieved by incubating with 0.25% Triton X-100 for 15 mins at room temperature. Cells were washed in blocking PBS (0.5% Triton X-100, 3% bovine serum albumin, 10% normal goat serum) overnight at 4 degrees Celsius. Cells were then stained using a rabbit polyclonal anti-NP antibody at a

1:500 dilution in blocking PBS for 1 hour at room temperature. After washing three times in PBS, the cells were then 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 subsequently washed twice with blocking PBS before mounting in Mowiol (Calbiochem) containing 1 µg 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.

2.4 Subcellular RNA fractionation

Protocol was followed as described by Neve et al. (Neve, Burger, et al. 2016). Briefly: approximately $5x10^7$ cells were washed twice with ice-cold PBS and detached from the cell culture flask using TrypLE. Cells were pelleted by centrifugation at 1000xg for 5 minutes, and the pellet was resuspended in 1mL lysis buffer (10mM Tris-HCl, 140mM NaCl, 1.5M $MgCl_2$, 0.5% NP40) with slow pipetting. The suspension was centrifuged at 1000xg for 3 minutes at 4 degrees Celsius. 500 µL of the supernatant was transferred to a tube and placed on ice, constituting the cytoplasmic fraction. The pellet (representing the nuclear fraction) was resuspended in 1mL of lysis buffer, at which point 100 µL of detergent stock (3.3% sodium deoxycholate, 6.6% Tween 40) was added dropwise under slow vortexing. The cytoplasmic fraction was spun at 12,000xg for 1 minute, and the supernatant was transferred to a new tube. The nuclear and cytoplasmic pellets were resuspended in 1mL and 500 µL Trizol respectively. Chloroform was added to each sample (20% volume) and the samples were vortexed vigorously before being

centrifuged at 13,000xg for 15 minutes at 4 degrees Celsius. The upper aqueous phase (containing RNA) was extracted and the RNA was precipitated overnight in 750 µL isopropanol. The RNA pellets were extracted and treated to DNase-I and RNaseOUT (Roche) as per manufacturer's instructions. RNA was purified by phenol-chloroform extraction using 2.5 times ethanol and 10% sodium acetate. Purified RNA was centrifuged at 13,000xg for 30 minutes at 4 degrees Celsius. The pellet was resuspended in RNA-free water and RNA concentration was quantified using a NanoDrop ND1000.

2.5 Denaturing agarose gel electrophoresis (RNA)

1% (w/v) denaturing formaldehyde-agarose gels were used to determine the quality of extracted RNA and purity of the cytoplasmic RNA obtained by subcellular fractionation. 2 µg of RNA was mixed with formaldehyde gel running buffer (40mM sodium acetate, 5mM EDTA, 100mM MOPs), formaldehyde and formamide. The RNA was incubated at 65 degrees Celsius for 15 minutes, and then mixed with gel loading buffer (50% glycerol, 1mM EDTA, 0.25% bromophenol blue, 0.25% xylene cyanol) and ethidium bromide. Gels were run for 2 hours at 65V and then visualised using a Kodak Gel Camera system.

2.6 RNA Sequencing

2.6.1 QuantSeq 3'mRNA-Seq Library Preparation

Libraries to be sequenced were prepared using the Lexogen QuantSeq 3'mRNA-Seq kit as per manufacturer's instructions. This method ensured that sequences closest

to the 3'-end of mRNAs (compared to whole-transcript sequencing), and that only one fragment per transcript would be captured, enabling an accurate assessment of gene expression. Input RNA for nuclear fractions was 500ng and 2-5 µg for cytoplasmic fractions (concentrations calculated using a NanoDrop ND1000). 14-16 PCR cycles were used for amplification. Resulting library fragment size was approximately 200bp.

2.6.2 High-sensitivity DNA (hsDNA) Bioanalyser Assay

As a quality control check, libraries generated with the Lexogen QuantSeq 3'mRNA-Seq protocol were assessed using a hsDNA chip on the Agilent 2100 Bioanalyzer as per manufacturer's instructions. Library concentrations were also quantified with this method.

2.6.3 RNA-seq on the Ion Proton Platform

3'QuantSeq libraries were diluted to a final concentration of 350pM. These were then loaded onto the Ion Chef platform for template preparation using the Ion PI Hi-Q Chef Kit and the resulting prepared chip was then sequenced using the Ion Proton Sequencing System as per manufacturer's instructions. Ion Proton sequencing works by measuring the pH when nucleotides are incorporated into a DNA strand via an ultrasensitive chip. The resulting read lengths are approximately 200bp, and single-end data is generated. Ion Sequences were then aligned to the human genome (hg19 build) by the Ion Torrent Server TMAP aligner using default settings. Each BAM file for further processing had a minimum of 1 million reads.

2.7 Bioinformatics

2.7.1 APA and gene expression pipeline

The bioinformatics pipeline used for APA and gene expression analyses was assembled in part together with Mark Hickling.

A schematic of the pipeline workflow can be seen in Figure 2.1. First, the FASTQ files - which were generated by the Ion Torrent sequencing platform - were aligned to the human genome by the Ion Torrent Server TMAP aligner. The resultant BAM files were then intersected with PolyA_DBv3, which is a database of annotated polyA sites that have been generated by the 3' region and extraction deep sequencing method (3'READS) (Wang, Nambiar, et al. 2018). Importantly, this step mitigates any potential internal priming issue, which is a problem often faced in APA analysis when oligodT primers are used during data collection. Windows were made by extending 20 nucleotides downstream and 200 nucleotides upstream of each annotated PAS. If windows overlapped on the same strand, they were merged into singular larger windows. Mapped reads were refined to their 3'-most nucleotide and those which fell within extended PAS windows were counted. Next, reads for each PAS were compiled into a counts table, which was then separately incorporated into (i) APA and (ii) differential gene expression analyses. For APA (i), reads for all PASs of each gene were compiled into counts tables and the location of reads within the gene body (3'UTR or coding-region) was also documented. Changes in relative PAS usage between replicates and conditions was then assessed using the DEXSeq method (Anders, Reyes, and Huber 2012), which outputs a p-adjusted value using the Benjamini-Hochberg method. A "deltaRA" value for each APA event was calculated to assess the change in relative abun-

dance between APA isoforms: values between -1 and 0 represents a shift towards lengthening while values between 0 and 1 represents movement towards shortening (with -1 and 1 meaning complete switches to the long or short APA isoforms respectively). DEXSeq was originally designed for differential exon usage analysis (Anders, Reyes, and Huber 2012), but has been repurposed in our pipeline to test for differential PAS usage. A major benefit of using DEXSeq is that it takes biological variation into account and controls for false discoveries by adjusting p-values for multiple testing using the Benjamini-Hochberg method.

APA events were then categorised into three groups so that appropriate downstream analyses could be carried out: All, UTR, and CR (see Figure 2.2). The "All" group contained reads present at PASs in both the 3'UTR and coding regions; "UTR" included reads only found in 3'UTRS, while the "CR" group consisted only of PAS reads present within introns or alternative exons. A visualisation of these groupings can be seen in Figure 2.2. For each category we then compared the two PASs with the most significant changes, while any gene with less than one read at either PAS in any sample is removed. We then calculated a measure of isoform relative abundance (deltaRA, see Equation 2.1).

$$\Delta RA = \left[\frac{\text{Short isoform}}{\text{Short isoform + Long isoform}}\right]_{Condition1} - \left[\frac{\text{Short isoform}}{\text{Short isoform + Long isoform}}\right]_{Condition2} \tag{2.1}$$

A deltaRA value between -1 and 0 represents relative isoform lengthening for a particular gene, whereas a value higher than 0 but less than 1 indicates shortening. The pipeline also extracts additional information for each APA event, including genomic location of PASs selected, PAS type, PAS signal and aUTR length (which,

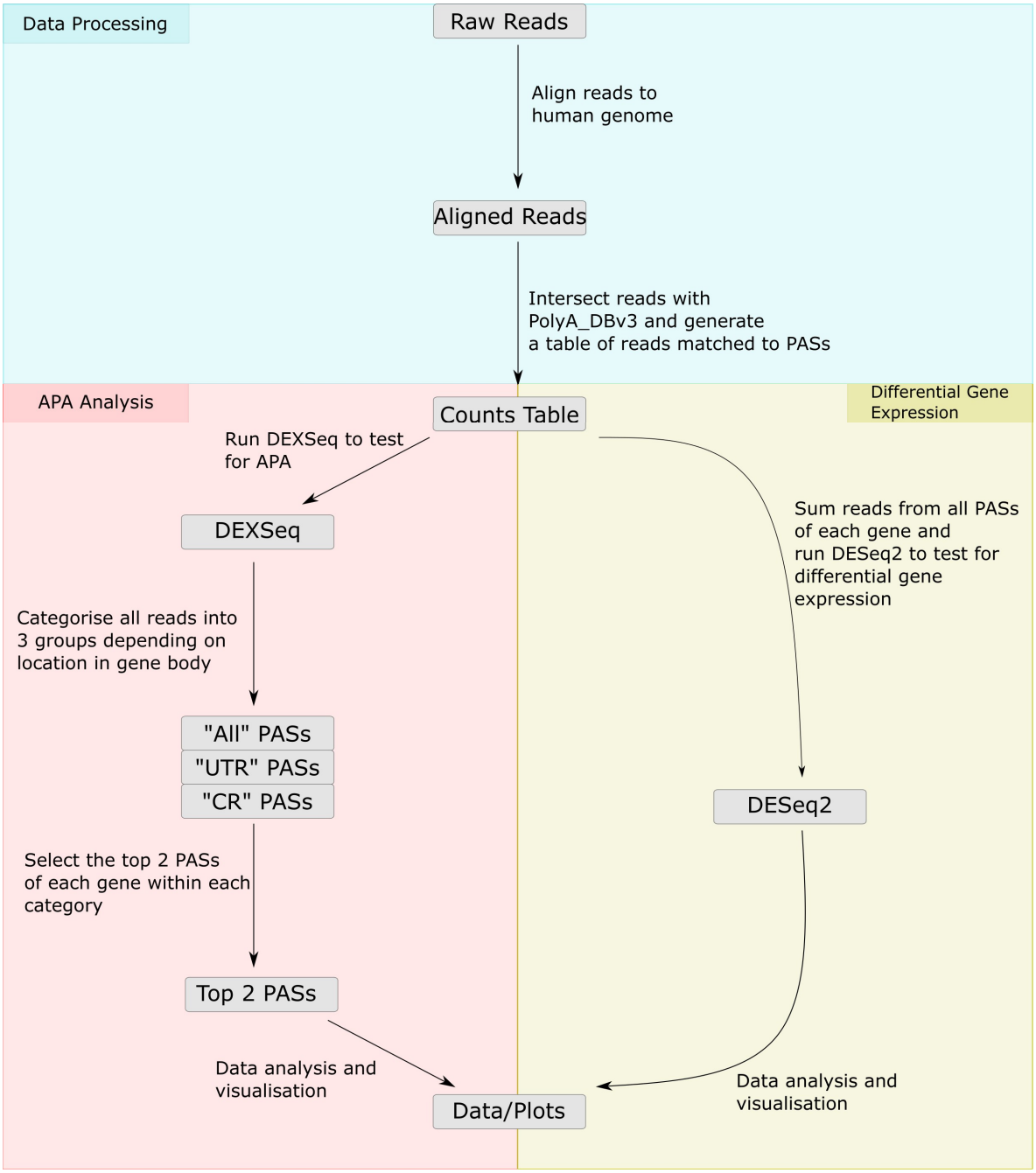


Figure 2.1: Bioinformatic pipeline used to detect APA and differential gene expression. Reads are aligned to the human genome, and intersected with PolyA_DBv3 to mitigate the internal priming issue. (i) For APA analysis, the counts table containing read numbers for every PAS was passed through DEXSeq to determine significant differences in PAS usage. Then, PASs were split into 3 groups: "All", "UTR" and "CR". Each of these groups were analysed separately. (ii) For differential gene expression analysis, all counts from each PAS were summed to provide the total number of reads for each gene. These totals were then provided to DESeq2 in order to analyse changes in gene expression between conditions.

for simplicity, is a term used in the pipeline to describe the number of bases between the two chosen PASs, regardless of PAS location). This additional information can then be utilised for further downstream analyses, which will be discussed in later sections.

For differential gene expression (ii), the counts table, generated after read alignment and PolyA_DBv3 intersection, was processed through DESeq2 (Love, Huber, and Anders 2014) - a commonly-used method to analyse differential count data. To do this, the reads of each gene were summed. Since a 3'end-biased method of library preparation was employed, each read represented a captured PASs. Genes not found in the RefSeq database were excluded. DESeq2 employs the Wald test for significance testing, and the p-values are adjusted for multiple testing by using the Benjamini-Hochberg method, as with DEXSeq. DESeq2 processing generated several values: base mean values for each gene (which represent the mean of normalised counts per sample once sequencing depth was normalised), log2-fold change values, p-values determined using a Wald test and p-adjusted values evaluated using the Benjamini-Hochberg method.

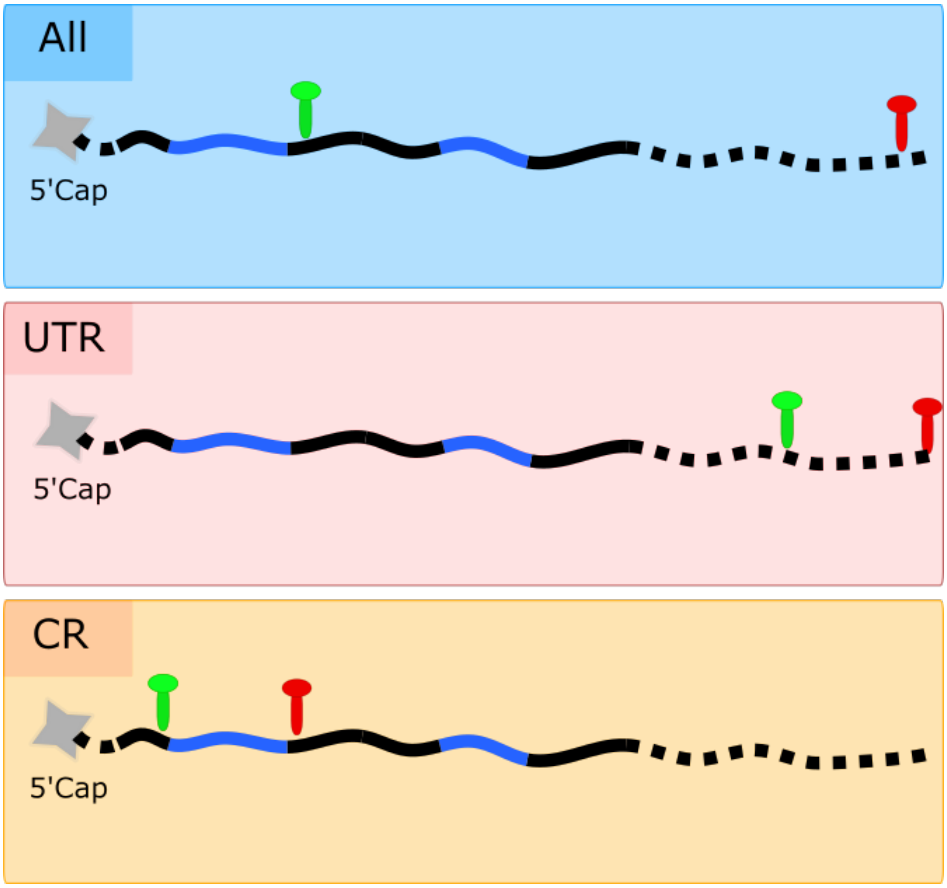


Figure 2.2: PAS groupings for APA analysis. To achieve a high-resolution look at changing APA profiles, poly(A) sites were separated into three groupings: "All" (containing PASs found anywhere throughout the pre-mRNA molecule), "UTR" (PASs only present in the 3'UTR), and "CR" (PASs found only within introns or alternative exons).

2.7.2 PAS signal and PAS type analyses

The poly(A) signal sequence and type for each annotated PAS were obtained from the PolyA_DBv3 (Wang, Nambiar, et al. 2018) database. The PAS signals and types of PASs with significant p-adjusted values as calculated by DEXSeq (Anders, Reyes, and Huber 2012) were then categorised into APA subtypes and APA direction, the distributions of which were then visualised using bar charts generated in R. These analyses were repeated on PASs with low p-adjusted values.

2.7.3 K-mer analyses

K-mer analyses were based on methods reported by Li. et al (Li, You, et al. 2015). Briefly, each PAS was broken down into 3 subregions referred to by their distances from the cleavage site: -100nt to -41nt (termed "Start"), -40nt to -1nt ("Mid") and +1nt to +100nt ("End") (depicted in Figure 2.3). RNA sequences were extracted for each subregion and scanned for 4- and 6-mers using a sliding window that shifts one nucleotide at a time. All K-mers were then compiled for each subregion of significant PASs, and their frequencies were tested against K-mer frequencies in similar subregions of other genes using the Fisher's exact test. The p-values of significant K-mers (less than 0.001) were designated to be enriched or depleted by being multiplied by 1 or -1 respectively.

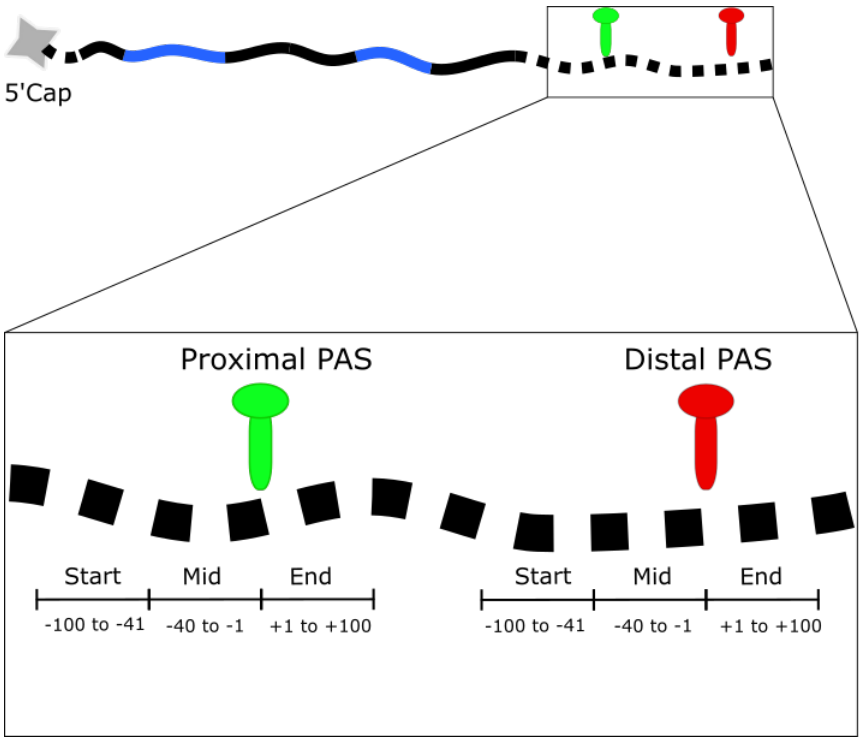


Figure 2.3: Schematic showing parameters of the K-mer analysis sequences surrounding PASs. Proximal and distal PASs were grouped separately, with each type being further broken down into three areas surrounding the PAS. "Start" consists of -100nt to -41nt; "Mid" includes -41nt to -1nt; "End" includes +1nt to +100nt.

2.7.4 aUTR motif discovery

RNA sequences of aUTRs from genes with significant DEXSeq p-adjusted values were extracted and scanned for motif enrichment using the XSTREME tool (Grant and Bailey 2021), part of the MEME SUITE (Bailey, Boden, et al. 2009). XSTREME first uses two different motif discovery algorithms, MEME (Bailey and Elkan 1995) and STREME (Bailey 2021), to discover novel sequence motifs. It then applies the SEA algorithm (Bailey and Grant 2021) to find enrichment of functional motifs, and then clusters the discovered motifs by similarity to each other.

aUTRs from genes which did not experience significant differential PAS usage were used as control sequences. Significant motifs were then scanned using the Tomtom tool (Gupta et al. 2007) to identify RNA-binding proteins which bind the generated motifs. Tomtom searches a database of motifs with given query motifs to find if any newly discovered motifs resemble previously-described putative motifs (Gupta et al. 2007).

2.7.5 Gene ontology analyses

Gene ontology analyses were carried out using the Enrichr platform (Chen, Tan, et al. 2013; Kuleshov et al. 2016; Xie et al. 2021). Enrichr is a commonly-used gene set search engine which utilises machine learning strategies to associate a collection of genes with a functional biological term. To do this, it draws upon hundreds of gene set libraries including differentially expressed genes after drug, gene, disease and pathogen perturbations derived from the national center for biotechnology information (NCBI) gene expression omnibus (GEO) (Kuleshov et al. 2016).

For biological process analyses, the "GO Biological Process 2021" gene library set was used, which draws from the Gene Ontology (Ashburner et al. 2000; "The Gene Ontology resource: enriching a GOld mine" 2021), a widely-used knowledge-base of gene functions.

For biological pathway analyses, the "BioPlanet 2019" gene library set was used (Huang, Grishagin, et al. 2019).

2.7.6 Venn diagrams

Venn diagrams showing overlapping genes between categories were generated using InteractiVenn (Heberle et al. 2015).

Chapter 3

Characterising the host alternative cleavage and polyadenylation landscape in influenza A H3N2 Udorn-infected cells

3.1 Introduction

Influenza A viruses (IAV) replicate within the nuclei of host cells, an unusual characteristic for RNA viruses (Dou et al. 2018). Due to this nuclear component of its life cycle, IAV can easily interfere with host cellular processes at the level of pre-mRNA processing. Indeed, a number of core IAV proteins have been identified which suppress host gene expression by directly compromising mRNA maturation.

This suppressive effect includes the IAV cap-snatching activity, which not only serves to equip the viral mRNAs with a functional cap, but destines the affected host mRNAs for decay (Fodor and Te Velthuis 2020). NS1, a flu host shut-off protein, is capable of compromising host mRNA 3'end formation by sequestering the essential 3'end processing factor CPSF30 (Das et al. 2008; Ji, Wang, and Liu 2021; Nemeroff et al. 1998; Twu et al. 2006; Chan, Huppertz, et al. 2014). PA-X, a viral endoribonuclease protein, is similarly enriched in the nucleus and selectively degrades host mRNA while ignoring viral mRNAs (Gaucherand et al. 2019; Li, Qi, et al. 2021; Khaperskyy, Schmaling, et al. 2016). PA-X is also thought to interact with another core CPA component, the CFI complex (Gaucherand et al. 2019).

It is well-established that alternative cleavage and polyadenylation (APA) can be induced by altering core CPA levels, including those of CPSF30 (Curinha et al. 2014; Ghosh et al. 2022). Additionally, changes in host APA profiles have been documented in viral infections, including vesicular stomatitis virus (Jia, Yuan, et al. 2017), porcine epidemic diarrhea virus (Wei, Li, et al. 2021), human papillomavirus (Wu, Chen, et al. 2019) and herpes simplex virus 1 infections (Wang, Liu, et al. 2021). These findings, taken together with the evidence that the IAV NS1 and PA-X proteins associate with host CPA factors, point to a strong case for IAV involvement with host cleavage and alternative polyadenylation. By producing mRNA isoforms with different 3'UTR lengths or coding-regions, APA has the potential to substantially alter host gene expression (see section 1.3 in the thesis introduction).

3.1.1 Aims

The work in this chapter set out to elucidate IAV's impact on the host APA profile, using the H3N2 influenza virus strain A/Udorn/72 (hereafter "Udorn"). Udorn is a well-characterised strain which is known to have a strong NS1-CPSF30 association (Twu et al. 2006; Bauer et al. 2018), and so was expected to induce a strong effect on host 3'-end processing. Importantly, since IAV has a strong nuclear component to its life cycle, a subcellular RNA fractionation approach was employed so that host mRNA could be analysed from both the nuclei and cytoplasms of cells. This enabled a high-resolution exploration of the APA profile of infected cells, and allowed us to distinguish co-transcriptional APA events from post-transcriptional APA events. In addition, the subcellular RNA fractionation technique provided powerful insights into gene expression change distinct to each subcellular compartment. Understanding more about how alternative polyadenylation is modulated during IAV infection will help us determine what biological functions and pathways are hijacked by IAV, and which specific genes (if any) may be targeted by therapeutics to hinder or block viral replication in cells.

3.2 Results

3.2.1 Design of the Udorn infection experiment

To analyse differential PAS selection and gene expression changes in Udorn-infected cells, a bulk 3'RNA-seq method was employed at various timepoints throughout an infection period. Initially, the experiment was intended to encompass a broad infection timecourse performed in duplicate, with RNA extracted from infected

cells at 3, 6, 12 and 18 hours post infection (hpi), and gene expression differences compared with uninfected cells at discrete timepoints. However, Udorn propagation to the scale necessary for this experiment proved more difficult than originally envisaged, and substantially more time was required for protocol optimisation than expected. Additionally, the coronavirus pandemic magnified this challenge by introducing departmental closures and self-isolation risks. With these difficulties in mind, it was instead decided to modify this experiment to compare gene changes in early infection (3hpi and 6hpi) with those changes observed later (12hpi and 18hpi) using the first full experimental replicate. This meant that analysis of the resulting data covered a window of time during an infection period rather than being restricted to single timepoints.

High infection rates needed to be achieved to ensure that sequencing reads would be sourced from infected cells - if too many uninfected cells were present, the signal from infected cells could be dampened and real gene expression changes during infection would be hidden. To achieve this necessary level of infection in a cell population, a high multiplicity of infection (MOI) was required - meaning that there would be sufficient infectious virus particles to infect the vast majority of cells. The standard Poisson distribution method (Noton et al. 2009; Martin, Harris, et al. 2020), described in Equation 3.1, was used to calculate an appropriate MOI of five meaning that approximately 95% of cells should be infected by at least one infectious particle.

$$P(k) = e^{-\mathrm{m}} m^{\mathrm{k}}/k! \tag{3.1}$$

As nuclear RNA was to be extracted and sequenced (as discussed in this chap-

ter's introduction), approximately 1.5x10^7 cells were required at each timepoint to produce enough nuclear RNA material for library preparation. Therefore, approximately 7.5x10^7 infectious Udorn particles were required for each timepoint. The experimental approach, from virus propagation to RNA-sequencing, is depicted in Figure 3.1.

Large-scale Udorn stocks were grown using Madin-Darby canine kidney (MDCK) cells and then purified by ultracentrifugation through a sucrose gradient to remove cell debris and contaminating host factors (Figure 3.1A). Next, the virus titre was measured by plaque assay which allowed measurement of plaque-forming units - a proxy for infectious Udorn particles. To do this, adhesive MDCK cells were infected with serial dilutions of purified virus, and the wells covered with a semi-solid agarose-media layer (Figure 3.1B). After a period of incubation and subsequent staining for live cells, small plaques were visible which represented singular infectious particles (see Figure 3.2A). It was important to know the virus titre at this stage so that Udorn infections of human cells could proceed at the known MOI quantity of five. Human lung adenocarcinoma (A549) cells - commonly used in influenza *in vitro* studies - were cultured and then infected, while a separate cell population was left uninfected (mock) for comparison (Figure 3.1C). To allow for efficient infection, cells were incubated at room temperature with the virus, which facillitated virion attachment to cell sialic acid residues by IAV's surface glycoprotein HA. Next, the cells were incubated at 37 degrees over a range of "hours post infection" (hpi): 3hpi, 6hpi, 12hpi and 18hpi, timepoints chosen to capture both early and late infection. Nuclear and cytoplasmic RNA was extracted from the cells at these timepoints (Figure 3.1D) which was then run on a denaturing gel as a quality control step, to ensure that the subcellular fractionation technique had

been successful. In humans, the 45S precursor ribosomal RNA (rRNA) is cleaved into the 28S and 18S rRNAs via the 32S pre-rRNA intermediate in a pathway that takes place in the nucleus, before the processed rRNAs are exported into the cytoplasm (Lafontaine 2015). Therefore the purity of the cytoplasmic fraction could be assessed by verifying the abscence of the pre-rRNAs. The lack of bands representing the 45S and 32S pre-rRNA subunits in the cytoplasmic RNA samples seen in RNA denaturing gels (an example of one is seen in Figure 3.2B) confirmed the purity of the cytoplasmic fractions. The RNA was then used to make 3'-end enriched libraries, which were sequenced on the Ion Torrent platform (Figure 3.1E). The reads were subsequently mapped and analysed using a custom-made APA pipeline (discussed below).

In parallel with the main infection experiment, a smaller-scale infection was carried out as a quality control step with the goal of ensuring that a high level of infection took place, as expected. To that end, cells were grown on coverslips and then infected at an MOI of five - the same level as in the main infection which was used for RNA sequencing. At the appropriate timepoints, these cells were fixed and incubated with antibodies targetting the NP IAV protein, while the nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI). As seen in Figure 3.3, weak NP signal could be seen mainly centered around cell nuclei during the early timepoints of 3hpi and 6hpi, while in 12hpi and 18hpi, much stronger NP signal can be seen both in the nuclei and cytoplasms of the vast majority of cells. Nuclear accumulation of NP is a hallmark of early infection, while it is known to spread into the cytoplasm later in infection (Kao et al. 2010). This confirmed that the virus stock was infectious, and could be used going forward to study host gene expression changes in infected cells.

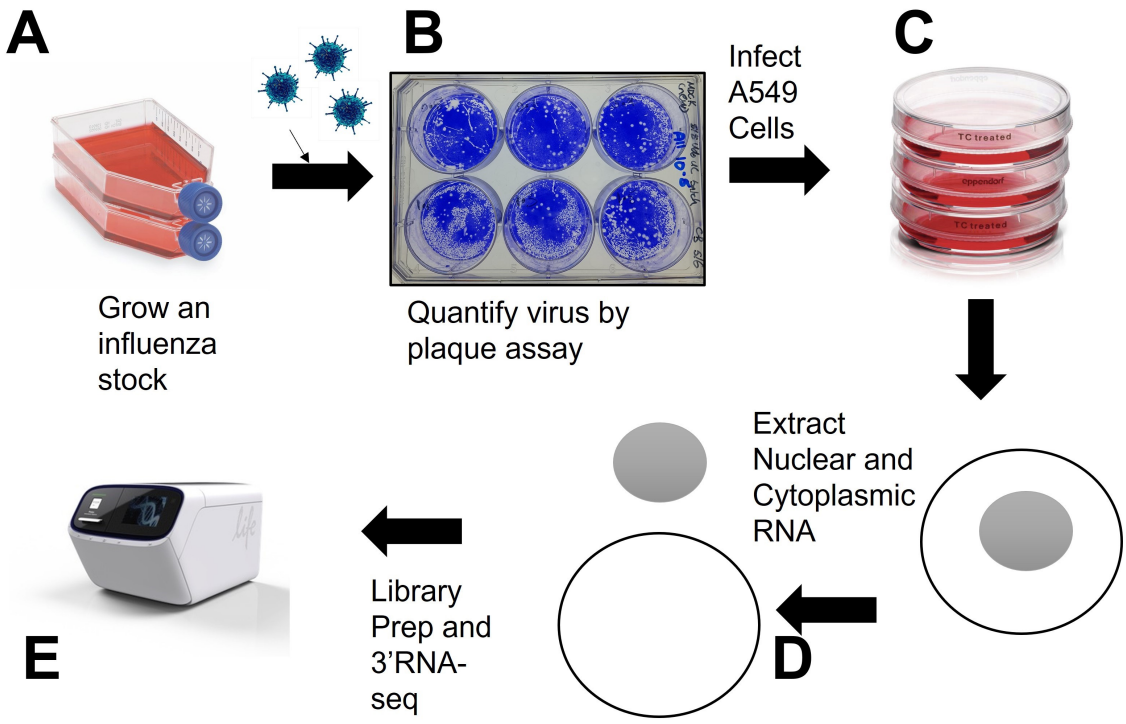


Figure 3.1: Experimental approach used for the IAV Udorn infection. First, large-scale Udorn stocks were propagated which were then quantified using the plaque assay technique. Using the resulting "plaque-forming unit" measure, A549 cells were infected at a ratio of five infectious units for every one cell. At different timepoints (3hpi, 6hpi, 12hpi and 18hpi), RNA from the nuclei and cytoplasms of infected cells was extracted. cDNA libraries were prepared using a 3'-end enriching method and RNA-sequencing was carried out.

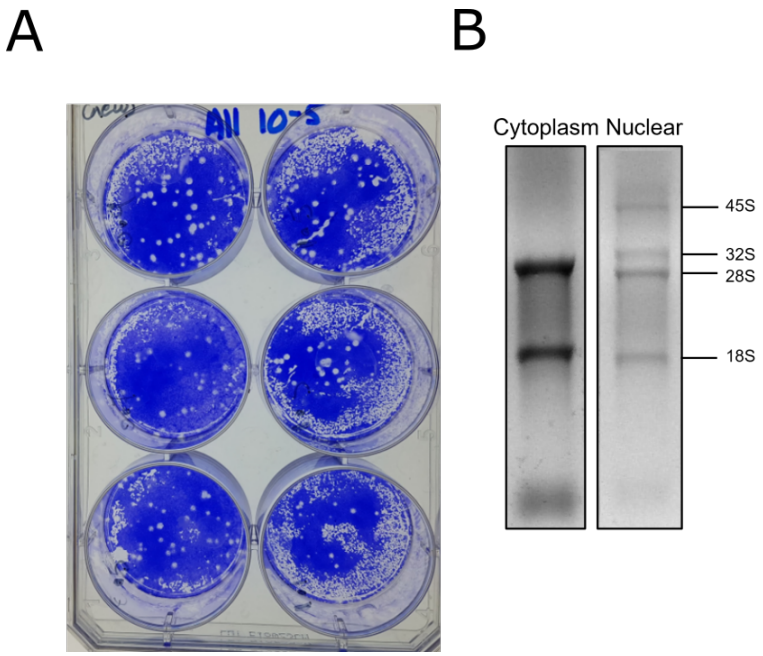


Figure 3.2: Udorn plaque assay and RNA gel. A) Plaque assays were used to quantify infectious particles. Each plaque represents one infectious particle. B) An RNA denaturing agarose gel showing the distribution of pre-rRNAs and processed rRNAs in the Udorn-infected cells, 12hours post-infection, in both the nuclear and cytoplasmic fractions.

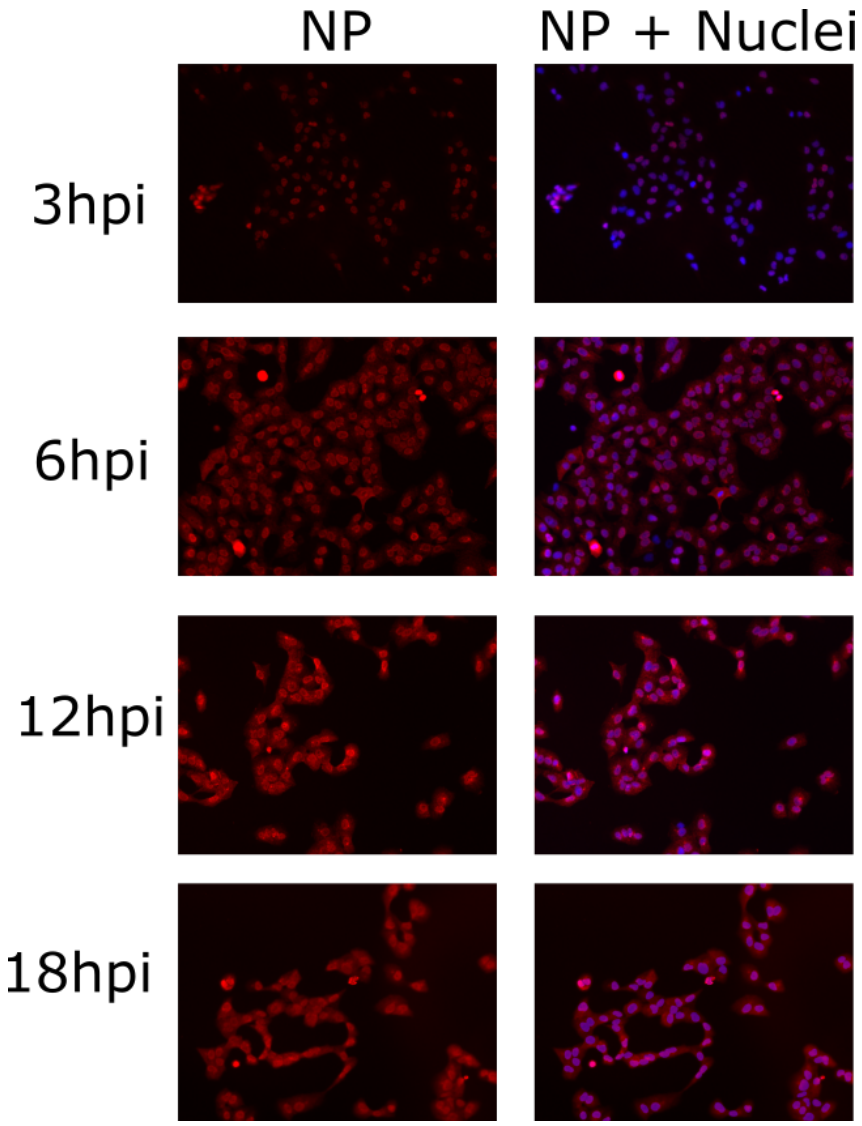


Figure 3.3: IAV (Udorn) infection efficiency of A549 cells. Human A549 cells were infected with Udorn at an MOI of five. At different timepoints post-infection (3hpi, 6hpi, 12hpi and 18hpi), the cells were fixed and viewed by immunofluorescence microscopy. Nuclei were stained blue with DAPI, while the IAV NP protein was visualised by staining with targetted antibodies.

3.2.2 Establishing a bioinformatic pipeline to analyse alternative cleavage and polyadenylation and differential gene expression

To carry out data analysis on the 3'RNA-seq data, we set up an in-house bioinformatic pipeline with two main aims in mind: 1) to robustly test for APA events and 2) to accurately measure differential gene expression between samples. As explored in the introduction of this thesis, there are several pipelines built to investigate APA that have been published in the literature and are available to use. However, one important criterion we had for the analysis of our 3'RNA-seq data was a pipeline that we could flexibly customise and use to perform further downstream analyses. With this in mind we assembled our own pipeline from the ground-up, which allowed us to understand and modify the analysis at any step from data collection to data visualisation. A detailed explanation of the bioinformatic process used to analyse alternative cleavage and polyadenylation and differential gene expression is found in chapter 2, section 2.7.1.

3.2.3 3'UTR APA changes are induced in Udorn-infected cells

To quantify changes in 3'UTR PAS usage of Udorn-infected cells, 3'RNA-seq data from early infection timepoints (3hpi and 6hpi) was compared with that from later in the infection (12hpi and 18hpi), and this was done separately with both nuclear and cytoplasmic RNA. To do this, the bioinformatic pipeline discussed in section 2.7.1 (chapter 2) was used. Any genes with a p-adjusted value of less than 0.1

determined by DEXSeq were considered significant APA events. If a gene had a deltaRA of less than 0, it was considered to be a lengthening event, while genes with deltaRA values of greater than 0 were considered to be shortening.

As seen in Figure 3.4A, with this approach we isolated 254 total 3'UTR APA events in the nuclear fraction in the later infection period, with 143 lengthening and 111 shortening compared to the early timepoints. The top 20 most significant APA events are shown in Table 3.1, with a complete list of all 254 events recorded in the Appendix, Table A.1. In the cytoplasmic samples, there were 271 3'UTR APA events in total, with 81 lengthening and 190 shortening. The top 20 most significant UTR APA cytoplasmic events are shown in Table 3.2, with a complete list of all 271 events recorded in the Appendix, Table A.2.

Noting that there were twice as many shortening events in the cytoplasms than lengthening events, it seemed plausible that the longer isoforms generated in the nucleus did not penetrate into the cytoplasm. To test whether this was the case, four categories were considered comprising all 3'UTR APA events for further comparisons between compartments: nuclear lengthening, nuclear shortening, cytoplasmic lengthening and cytoplasmic shortening. These categories were overlapped with one another to test how many APA events were persistent between the nucleus and cytoplasm (see Figure 3.4B). Interestingly, we found that only 6 of the 81 cytoplasmic lengthening events were shared with the nucleus, indicating that the vast majority of nuclear lengthened APA isoforms were not exported. Furthermore, our use of a subcellular fractionation approach determined that APA events in Udorn-infected cells broadly appeared to be localised to one compartment each, with few crossing between the nucleus and cytoplasm. Several different mechanisms could be at play here, such as the general property of

longer UTRs being negatively correlated with nuclear export (Chen and Steensel 2017) or, specifically in the context of IAV-induced host APA, the flu NS1 protein blocking nuclear export of host mRNAs (Satterly et al. 2007; Zhang, Xie, et al. 2019).

To summarise the quantification of 3'UTR APA events, Udorn infection induces a dramatic effect on the 3'UTR APA profile of host cells. We observed hundreds of 3'UTR APA events in both the nuclear and cytoplasmic fractions in the later infection phase, with many of these being unique to one compartment. The mechanisms behind these observations are difficult to delineate, but probably include the result of the NS1-induced blockage on host mRNA export (Satterly et al. 2007; Zhang, Xie, et al. 2019), and/or the general characteristic of mRNAs with longer 3'UTRs being retained at a higher rate than shorter counterparts (Chen and Steensel 2017). Compartmentalisation of APA transcripts could therefore be an effective method of gene regulation, especially if changes in differential gene expression levels aligned with differential PAS usage (discussed in later sections).

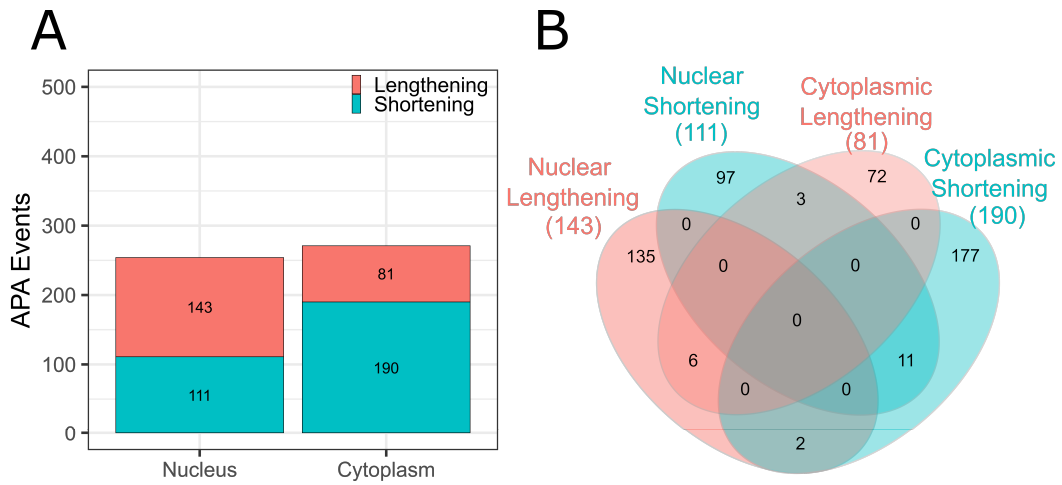


Figure 3.4: Significant 3'UTR APA events seen across the nuclei and cytoplasms of Udorn-infected cells. A) Bar chart showing the total number of significant 3'UTR APA events detected in both the nuclei and cytoplasms of Udorn-infected cells, split into lengthening and shortening categories. B) Venn diagram depicting overlaps of genes that undergo APA between the four categories: nuclear lengthening, nuclear shortening, cytoplasmic lengthening and cytoplasmic shortening.

Gene	P-adjusted	DeltaRA	Direction
AJUBA	4.54E-07	-0.441811384	Lengthening
SYNCRIP	1.29E-06	-0.461445756	Lengthening
HNRNPU	3.00E-06	-0.125205288	Lengthening
DCP1A	4.20E-06	0.492905109	Shortening
ATP6V1B2	5.64E-06	-0.576150393	Lengthening
DIABLO	6.04E-06	-0.55941173	Lengthening
UPF1	1.60E-05	0.701754386	Shortening
CKS1B	3.34E-05	-0.202687223	Lengthening
AKT2	8.68E-05	0.136051467	Shortening
GNPDA1	0.000116179	0.316079215	Shortening
NAV2	0.000148312	-0.624592075	Lengthening
COPS3	0.000154467	-0.409580007	Lengthening
ELAVL1	0.000160347	-0.530810587	Lengthening
TMEM135	0.000271801	-0.229437229	Lengthening
MYO18A	0.000326332	-0.720867869	Lengthening
TIAF1	0.000326332	-0.720867869	Lengthening
RAB3GAP1	0.000400197	0.089263041	Shortening
SYDE1	0.000442298	-0.605256926	Lengthening
PSMD11	0.000769277	0.042760727	Shortening
ASAP1	0.000775125	0.271000409	Shortening

Table 3.1: Table of the top 20 most significant genes undergoing 3'UTR APA switches (Nucleus, Udorn)

Gene	P-adjusted	DeltaRA	Direction
ARF3	3.30E-20	0.412364083	Shortening
MRPS16	2.19E-14	0.198032658	Shortening
VKORC1L1	2.36E-13	0.084989376	Shortening
JPT2	3.63E-12	0.494318477	Shortening
RPL37A	2.99E-10	-0.738090389	Lengthening
UBE2D3	3.09E-09	0.377963423	Shortening
PDXK	8.73E-09	0.388504898	Shortening
DESI1	1.86E-08	0.622911465	Shortening
PGAM5	6.62E-08	0.559710334	Shortening
F2R	8.92E-08	-0.078816621	Lengthening
CALM3	3.76E-07	0.58286639	Shortening
POLR2E	3.81E-07	0.094186184	Shortening
CTSB	4.52E-07	0.104904279	Shortening
EXOC7	7.56E-07	0.259066772	Shortening
ORAI3	1.41E-06	-0.628802982	Lengthening
INSIG1	1.57E-06	0.364571949	Shortening
FRMD8	2.34E-06	-0.567745872	Lengthening
NUAK2	2.98E-06	-0.709586283	Lengthening
PPIA	5.56E-06	0.059058037	Shortening
SLC35F6	5.56E-06	0.2233768	Shortening

Table 3.2: Table of the top 20 most significant genes undergoing 3'UTR APA switches (Cytoplasm, Udorn)

3.2.4 Genes regulated by 3' UTR APA are highly enriched for mRNA processing in Udorn-infected cells

In order to investigate what biological functions the genes regulated by Udorn-induced 3'UTR APA might be involved in, gene ontology (GO) analyses were performed on each of the four categories (Figure 3.5, with further detail in Tables A.3 - A.6 in the Appendix), using the Enrichr platform (Chen, Tan, et al. 2013; Kuleshov et al. 2016; Xie et al. 2021). Enrichr draws upon hundreds of gene set libraries to associate a collection of genes with functional biological terms. Interestingly, both nuclear GO "biological process" categories (lengthening, Table A.3, and shortening, Table A.4) were highly enriched for mRNA processing or mRNA stability activities - including mRNA processing, RNA splicing and RNA 3'-end processing - suggesting that APA may play a role itself in the regulation of factors involved in RNA metabolism during Udorn IAV infection, although it is not clear if these APA changes are induced by viral or host mechanisms. Additionally, some categories related to protein processing were also enriched in the nuclei, both lengthening and shortening, such as regulation of translation and protein modification and localisation. In the cytoplasmic categories, protein-associated processes such as protein localisation, protein transport and protein stabilisation were highly enriched in the genes that undergo transcript shortening (Table A.6), while those that lengthen (Table A.5) tended to be associated also with protein localisation, modification and pre-mRNA transcription. This is notable as IAV is known to use host protein transport systems during its replication cycle (Wang, Li, et al. 2012; Dou et al. 2018).

These findings demonstrate that IAV infection induces 3'UTR APA in genes

involved in host mRNA processing, including splicing, in the nucleus while protein-associated categories were over-represented in the cytoplasm. In the next sections, we explore wider gene expression level changes induced by IAV infection and investigate how 3'UTR APA might influence gene levels.

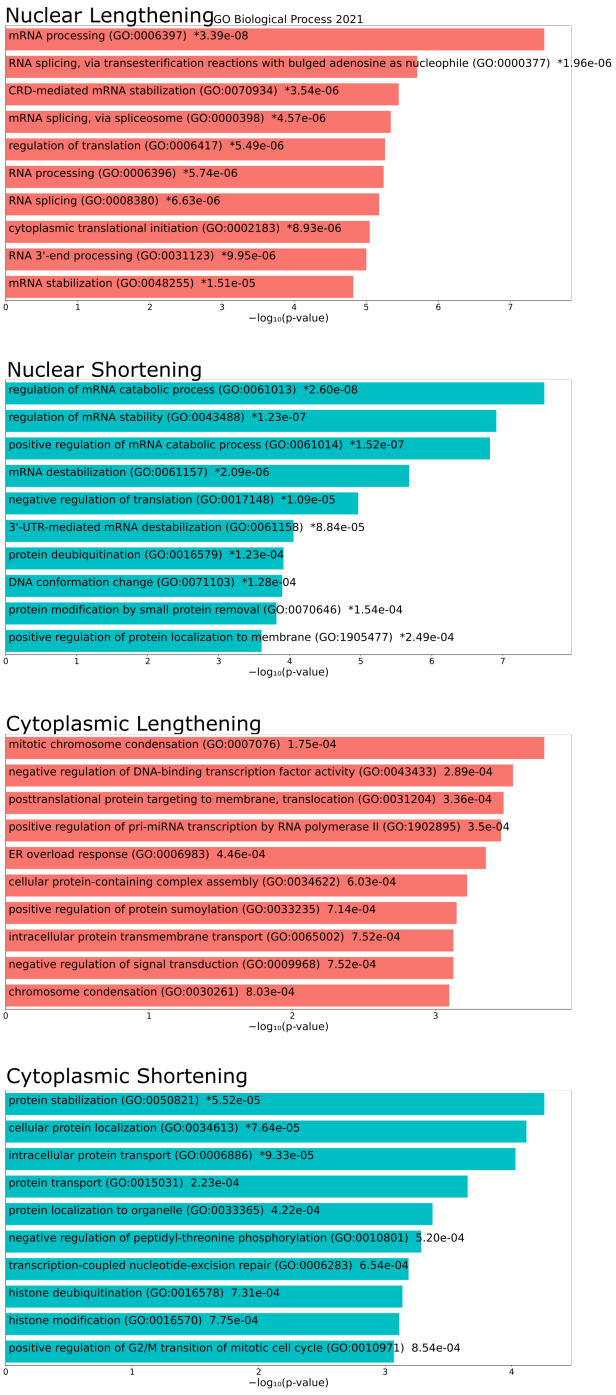


Figure 3.5: Gene ontology analyses of 3'UTR APA events detected in Udorn-infected cells in "biological process" category. Each category is shown separately, with -log10(p-values) adjusted for multiple testing by the Benjamini-Hochberg method represented on the x-axis.

3.2.5 Differential gene expression changes are induced in Udorn-infected cells

IAV infection is known to induce widespread changes to host gene expression, including modulating transcription termination (Bauer et al. 2018; Zhao et al. 2018; Heinz et al. 2018), altering genome structure (Heinz et al. 2018), degrading targeted transcripts (Gaucherand et al. 2019), inducing translational shutoff (Bercovich-Kinori et al. 2016) and alternative splicing (Thompson, Dittmar, et al. 2020). With this in mind, we next wanted to ask what genes were differentially expressed in Udorn-infected cells, and how expression patterns might differ between the nuclear and cytoplasmic fractions.

To test for gene expression changes, the reads for every PAS site of a gene were summed separately for each subcellular compartment and passed through DESeq2, which robustly quantifies differences in read counts per gene, with a p-adjusted vale of less than 0.1 considered a significant change. This method reported a total of 1096 gene expression changes in the nuclear fraction, with over twice as many upregulated genes (751) than downregulated genes (345). Almost three times the number of total changes seen in the nucleus were observed in the cytoplasm (2961), although the difference between upregulation (1630) and downregulation (1331) was less pronounced. A bar chart depicting these changes can be seen in Figure 3.6A, with the top 100 most significantly differentially regulated genes for the nuclear and cytoplasmic samples shown in Tables A.7 and A.8 respectively. To determine if genes experienced increased regulation in more than one of the four categories, a Venn diagram was employed which overlapped upregulated or downregulated genes in both compartments (Figure 3.6B). Approximately half (171) of

the genes that were downregulated in the nucleus were also downregulated in the cytoplasm, while 216 of the 751 genes that were upregulated in the nucleus were similarly upregulated in the cytoplasm also. As expected, this analysis confirmed that extensive changes in gene expression levels were induced during Udorn infection, and we next moved to explore what biological functions may be affected by these changes and what relationship these changes have with 3'UTR APA changes.

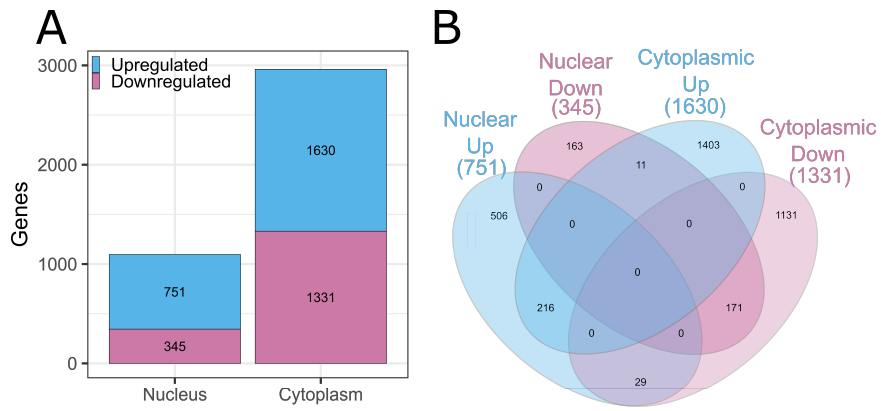


Figure 3.6: Differential gene expression changes observed in Udorn-infected cells. A) Bar chart displaying the total number of changes in gene expression observed in both the nuclear and cytoplasmic fractions, including a breakdown between upregulated and downregulated genes. B) Venn diagram showing overlaps of the four categories of gene expression: upregulation and downregulation in the nuclear and cytoplasmic fractions.

Genes enriched for translation and protein processing are differentially expressed in Udorn-infected cells

To assess what biological processes may be affected during Udorn infection by changes in gene expression levels, gene ontology analyses were performed (summarised in Figure 3.7, with a more detailed breakdown found in Tables A.9 - A.12 in the Appendix). Interestingly, protein-processing terms including translation,

protein targeting to endoplasmic reticulum (ER) and SRP (signal recognition particle) dependent targeting to membrane were highly significantly upregulated in both the nucleus (Table A.9) and cytoplasm (Table A.11). Upregulation of these processes is most likely beneficial to the virus as IAV requires plentiful supply of translation factors for production of its own proteins. Additionally, influenza membrane proteins are synthesised by ER-associated ribosomes and IAV proteins including HA are dependent on SRP (Dou et al. 2018). Oxidative phosphorylation-related terms were enriched in upregulated genes in the cytoplasm (mitochondrial ATP synthesis coupled electron transport, aerobic electron transport chain). This was inclusive of complex I (NDUF family), complex III (UQCR family) and complex IV (COX family) genes in particular. Continuous activity of the oxidative phosphorylation pathway is critical for IAV replication, and it has been reported that genes associated with this process are relatively untouched by host shutoff during infection (Bercovich-Kinori et al. 2016). Categories associated with viral processes were also highly enriched (not shown) in the cytoplasmic upregulated gene group, including positive regulation of viral transcription, positive regulation of viral processes, negative regulation of viral life cycle and regulation of viral genome replication (all terms having adjusted p-values less than 0.05). While cytoplasmic upregulation of these categories seemed to have contradictory effects on the viral life cycle, it is possible that these gene expression changes are reflective of both viral and host immune mechanisms at play. For example, the flu PA-X protein selectively degrades host mRNA while ignoring viral mRNA, which could be one way in which IAV manipulates host gene expression to produce a more favourable environment for viral replication (Gaucherand et al. 2019; Li, Qi, et al. 2021; Khaperskyy, Schmaling, et al. 2016). Gene downregulation did not seem to

be shared to the same degree between nucleus and cytoplasm, with regulation of tRNA processes, ncRNA nuclear export and metabolic processes downregulated in the nucleus (Table A.10) while transcription featured heavily in the "cytoplasmic down" category (Table A.12). Additionally, while not in the top 10 most significant GO terms and thus not shown in Figure 3.7, genes associated with "regulation of defense response to virus by host" were also highly significantly enriched (p-adj of 0.019) in the "cytoplasmic down" group.

In summary, genes that were upregulated during Udorn-infection were highly enriched for protein generation and processing, processes which are important for the generation of IAV proteins (Dou et al. 2018). In the cytoplasm, many genes involved in oxidative phosphorylation were upregulated, a process known to be important for IAV propagation (Bercovich-Kinori et al. 2016), including those encoding subunits of complex I, complex III and complex IV of the electron transport chain. Additionally, genes which have been shown to have pro- and antiviral effects were shown to be upregulated in the cytoplasm, most likely resulting from a mixture of viral and host mechanisms.

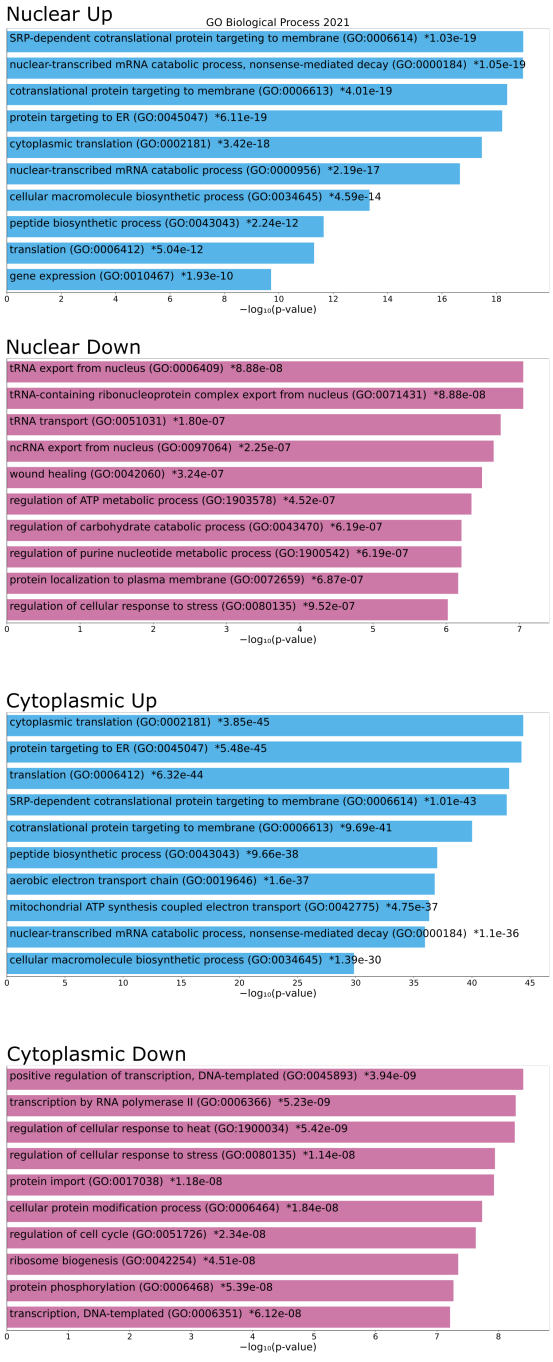


Figure 3.7: Gene ontology analyses of genes differentially expressed in Udorn-infected cells ("biological process" category). Gene expression changes are broken down into those that are upregulated or downregulated, in both the nuclear and cytoplasmic fractions, with -log10(pvalues) adjusted for multiple testing by the Benjamini-Hochberg method represented on the x-axis.

Genes associated with immunity and viral genome replication switch 3'UTR PASs and change expression in the cytoplasm

As discussed previously, 3'UTR APA does not lead to coding region changes for transcripts. However, one way it can affect protein output is by affecting stability of the RNA molecule and therefore regulation of the gene's overall expression. Therefore, we next wanted to explore what impact 3'UTR APA may have on gene expression levels in Udorn-infected cells. While deciphering the exact regulatory mechanisms contributing to the gene expression profile of cells is a major task, comparing UTR APA changes with differential gene expression data could indicate to us whether or not 3'UTR PAS switching contributes to this multi-faceted process. With this in mind, we aimed to delineate the impact of 3'UTR APA on differential gene expression in Udorn-infected cells by comparing those genes that switch PASs in the 3'UTR and are also differentially expressed during Udorn infection.

Noting that the cytoplasm is the site of translation, it was anticipated that gene expression differences there would be more physiologically relevant than those in the nucleus. Thus, we extracted those genes that underwent 3'UTR APA in the cytoplasm while also experiencing an overall change in gene expression in the cytoplasm. A total of 87 genes were identified in this category, and to explore their functions a gene ontology analysis was performed. Categories related to immunity (cytokine production, immunoglobulin production), protein transport (nuclear import) and particularly strikingly, regulation of viral genome replication were enriched (Table 3.3). As these 87 genes were those that experienced a PAS switch in the cytoplasm while also changing gene expression, it can be speculated that

APA regulation of these 87 genes, enriched for immunity and viral genome replication, contributes to their expression levels and potentially their protein output. This is notable as the most significantly enriched categories were those related to immunity and viral replication, suggesting that virus infection induces broad regulation of APA in the host - either via IAV or host mechanisms - thereby altering expression of genes which ultimately impact its ability to replicate.

GO Biological Process	**Genes**	**P-adjusted**
Positive regulation of cytokine production	SLC7A5, MAVS, TMED10, PDE4D, F2R, CEBPG, POLR2E, PHB	0.045115052
Regulation of viral genome replication	MAVS, RAD23A, ZC3HAV1, PPIA	0.045115052
Positive regulation of protein localization to nucleus	WWTR1, MAVS, SMAD3, RAN	0.045115052
Regulation of immunoglobulin production	GPI, PHB, SUPT6H	0.045115052
Positive regulation of protein import into nucleus	MAVS, SMAD3, RAN	0.045115052

Table 3.3: Table of GO biological processes for upregulated genes in cytoplasm which were also undergoing APA in Udorn-infected cells. The top 5 most significant terms are shown.

3.2.6 Coding-region APA changes are induced in Udorn-infected cells

Quantification of coding-region APA events

Coding-region APA is the most obvious way that APA can induce physiological changes, as it results in a change in the coding sequence of a transcript and thus alters the final protein product. Therefore, we next asked what coding-region APA changes were induced as Udorn infection proceeds. To do this, we collated all PAS switches that occurred between the top two most significant poly(A) sites as determined by DEXSeq into two categories: 1) those involving two poly(A) sites that are both within the coding region, a situation which we term CR-CR APA and 2) those involving a PAS in an intron/internal exon and also a PAS within the 3'UTR, which we term CR-UTR APA. Both of these classifications of coding-region APA result in an alteration of the final protein product, with the only and major difference between them being the capacity that CR-UTR APA has to make the full-length protein product with the involvement of the 3'UTR. As described in chapter 2, section 2.7.1, the flexibility of our pipeline allowed the analysis of both of these categories, which enabled us to have a more complete look at the overall coding-region APA profile of the infected cells. The analysis of the CR-CR category was carried out directly as an integral part of our pipeline while the CR-UTR analysis was extracted from the "All" group referred to in chapter 2, section 2.7.1, and a schematic representation of both of these groups can be reviewed in Figure 2.2.

Using this approach, we measured a total of 138 CR-CR APA events occurring in Udorn-infected cells with an apparent trend towards shortening, particularly in

the cytoplasm where there was approximately a 3:1 ratio of shortening to lengthening. For APA involving at least one PAS in the coding region and also one PAS in the UTR (CR-UTR APA), we detected a total of 171 changes in the nucleus and 240 in the cytoplasm. The direction of APA appeared to largely be the same in CR-UTR, with more shortening events overall, although this was more pronounced in the nucleus where almost 90% (153) of the total events were shortening. Bar charts displaying these coding-region APA events can be seen in Figure 3.8. Interestingly, the APA trends for coding-region APA and 3'UTR APA (discussed in section 3.2.3) were substantially different. Whereas a bias towards lengthening was reported in the nucleus for 3'UTR, shortening dominated coding-region APA, particularly in CR-UTR APA. Differences in how coding-region APA and 3'UTR APA are regulated have been reported before (Di Giammartino, Nishida, and Manley 2011), so it seems likely that different mechanisms are at least partly involved for this disparity during Udorn infection.

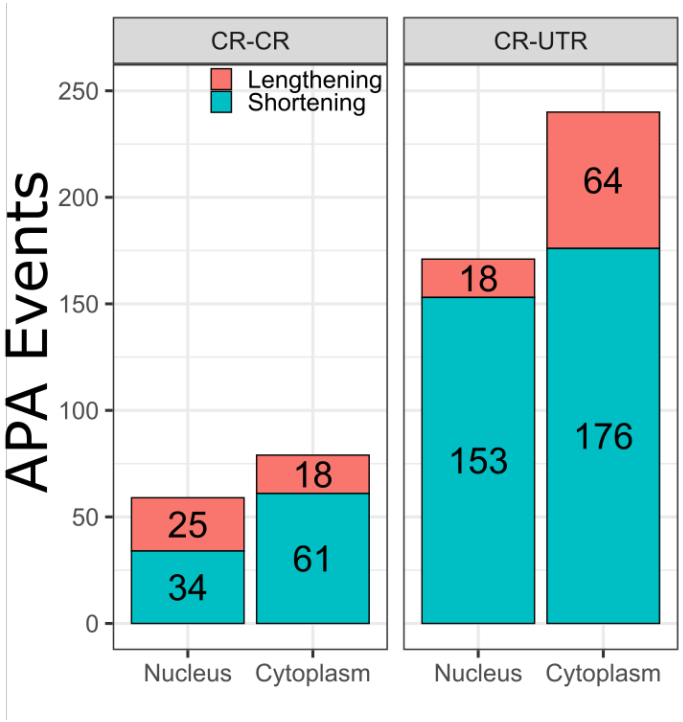


Figure 3.8: Coding-region APA in Udorn-infected cells. Lengthening and shortening events are highlighted separately. The left panel depicts significant changes in APA usage when both PASs (poly(A) sites) are located in the coding region (CR) of the pre-mRNA. The right panel shows the numbers of CR-UTR APA events detected, in both the nuclear and cytoplasmic fractions.

Lengthening CR-UTR APA genes are associated with IAV infection and viral RNA transcription and replication

To investigate the functions of genes that undergo an APA-induced coding-region change, we used the Enrichr tool to search for biological processes and pathways that the genes are associated with. We focused on the cytoplasmic fraction for this analysis, as we anticipated that coding-region APA changes would contribute the maximum physiological impact in the cytoplasm as the site of translation. In CR-UTR APA, lengthening events mean that the pool of transcripts transition from a state of incomplete coding capacity towards a state of full coding potential, thereby increasing the protein output of genes. Shortening, in contrast, compromises the coding capacity of full-length transcripts and therefore either decreases the protein output or generates truncated proteins with reduced or modified functions. With this in mind, we processed lengthening and shortening events in separate GO analyses.

Surprisingly, despite consisting of a much smaller share of APA events, the lengthening category of CR-UTR APA genes yielded many times more significant results for both the "biological processes" category and the "pathways" category than those genes that experienced shortening, suggesting a higher degree of shared function. Genes in the lengthening category for biological processes were associated mainly with protein processing, including translation and protein targeting (Table 3.4). For enriched pathways (Table 3.5), two influenza-related pathways were the most significant ("Influenza infection" and "Influenza viral RNA transcription and replication"), with others being related to disease, gene expression and translation. The genes associated with the two influenza pathways were: HSP90AA1,

RPL21, RPS8, RPS6, RPL10A and RAN. RPL21, RPS8, RPS6 and RPL10A are ribosomal proteins which have critical roles in translation. Lengthening CR-UTR events for these genes would most likely be beneficial to influenza, whether or not they are induced by the virus or the host, as more available translation machinery would mean more IAV proteins being produced. RAN is believed to play a role in influenza nuclear import (Eisfeld, Neumann, and Kawaoka 2015; Moreira, Yamauchi, and Matthias 2021), while an interaction between HSP90AA1 and flu PB2, enhancing viral RNA synthesis, has also been shown (Wang, Zheng, Lin, et al. 2020). Interestingly a paralog of HSP90AA1, HSP90AB1, which shares over 85% amino acid similarity, also experiences CR-UTR lengthening, although it is not highlighted by the GO analysis..

The GO analysis of the shortening category (Table 3.6) yielded fewer significant results despite consisting of almost three times as many genes, suggesting less homogeneity and possibly a more generalised, less specific response to Udorn infection. This could represent part of the IAV-induced host shutoff which limits host gene expression to create a more favourable environment for flu replication. There were two categories only in the "biological processes" analysis with a p-adjusted value of less than 0.1 and these were related to glucuronidation. In the pathway analysis (Table 3.7), there was low significance again, with an enrichment for a variety of terms including gene expression, apoptosis and glucuronidation. The genes highlighted associated with glucuronidation in the shortening cohort are UGT1A10, UGT1A9 and UGT1A6.

In this section we have demonstrated how lengthening CR-UTR APA changes could promote a positive environment for viral replication, by generating transcripts which encode for proteins known to be beneficial to IAV propagation.

Biological Process	P-adjusted	Odds Ratio	Genes
Cellular macromolecule biosynthetic process (GO:0034645)	0.017367819	7.852048688	RPL21, RPS8, RPS6, RPL10A, PWP1, MRPL43, PTMS
Positive regulation of tau-protein kinase activity (GO:1902949)	0.017367819	160.7419355	HSP90AA1, HSP90AB1
Neuron projection extension (GO:1990138)	0.017367819	32.63278689	HSP90AA1, HSP90AB1, CPNE1
Peptide biosynthetic process (GO:0043043)	0.017367819	10.67634676	RPL21, RPS8, RPS6, RPL10A, MRPL43
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	0.017367819	15.3875969	RPL21, RPS8, RPS6, RPL10A
Cellular response to heat (GO:0034605)	0.017367819	29.66169896	HSP90AA1, HSP90AB1, HSF1
Cytoplasmic translation (GO:0002181)	0.017367819	14.86666667	RPL21, RPS8, RPS6, RPL10A
Cotranslational protein targeting to membrane (GO:0006613)	0.017367819	14.70074074	RPL21, RPS8, RPS6, RPL10A
Positive regulation of cellular protein metabolic process (GO:0032270)	0.019715872	13.4952381	HSP90AA1, ASPH, CIRBP, VIM
Protein targeting to ER (GO:0045047)	0.019715872	13.35824916	RPL21, RPS8, RPS6, RPL10A

Table 3.4: Gene ontology analysis of lengthening CR-UTR APA changes in Udorn-infected cells (Cytoplasmic).

BioPlanet Pathway	P-adjusted	Odds Ratio	Genes
Influenza infection	0.001255915	15.06085193	HSP90AA1, RPL21, RPS8, RPS6, RPL10A, RAN
Influenza viral RNA transcription and replication	0.005095167	13.99435028	HSP90AA1, RPL21, RPS8, RPS6, RPL10A
T cell receptor regulation of apoptosis	0.008276191	5.328374656	HSP90AA1, ILF3, HSP90AB1, PSPC1, HSF1, PSME1, BIRC5, VIM, RPL10A
Disease	0.014240018	4.742009569	HSP90AA1, PSMA4, TSG101, RPL21, RPS8, RPS6, PSME1, RPL10A, RAN
Cytoplasmic ribosomal proteins	0.016247252	12.71282051	RPL21, RPS8, RPS6, RPL10A
Gene expression	0.029966244	3.668522385	SF3B2, PSMA4, THRA, RPL21, RPS8, RPS6, MBD2, PSME1, RPL10A, RAN
Sema3A/PAK-dependent axon repulsion	0.029966244	49.43672457	HSP90AA1, HSP90AB1
Hypoxia-inducible factor in the cardivascular system	0.029966244	45.90322581	HSP90AA1, ASPH
Translation	0.031637106	8.974603175	RPL21, RPS8, RPS6, RPL10A
Antigen processing and presentation	0.044918358	12.52080706	HSP90AA1, HSP90AB1, PSME1

Table 3.5: BioPlanet pathway analysis of lengthening CR-UTR APA changes in Udorn-infected cells (Cytoplasmic).

Biological Process	P-adjusted	Odds Ratio	Genes
Flavonoid glucuronidation (GO:0052696)	0.069000438	57.27745665	UGT1A10, UGT1A9, UGT1A6
Xenobiotic glucuronidation (GO:0052697)	0.069000438	42.95375723	UGT1A10, UGT1A9, UGT1A6
Glucuronate metabolic process (GO:0019585)	0.123602168	28.6300578	UGT1A10, UGT1A9, UGT1A6
Cellular glucuronidation (GO:0052695)	0.192377784	21.46820809	UGT1A10, UGT1A9, UGT1A6
Regulation of telomere maintenance via telomerase (GO:0032210)	0.296857751	9.581395349	TEN1, SRC, SMG6, CCT4
Regulation of cellular ketone metabolic process (GO:0010565)	0.340370733	7.660465116	PSMA3, PSMC2, PSMF1, PDK2
Positive regulation of programmed cell death (GO:0043068)	0.340370733	3.348064406	ABR, AKAP13, FLCN, DAPK1, PDCD2, CLU, TGM2, HTATIP2
Regulation of cytoplasmic transport (GO:1903649)	0.340370733	25.30651341	SRC, ATP9A
Regulation of hematopoietic progenitor cell differentiation (GO:1901532)	0.340370733	6.470029479	PSMA3, PSMC2, PDCD2, PSMF1
Regulation of mRNA catabolic process (GO:0061013)	0.340370733	4.925026241	PSMA3, PSMC2, IGF2BP1, PSMF1, EXOSC3

Table 3.6: Gene ontology analysis of shortening CR-UTR APA changes in Udorn-infected cells (Cytoplasmic).

BioPlanet Pathway	P-adjusted	Odds Ratio	Genes
Irinotecan pathway	0.094088471	34.35953757	UGT1A10, UGT1A9, UGT1A6
Nuclear beta-catenin signaling and target gene transcription regulation	0.126793617	7.699415205	YWHAE, MDFIC, RUVBL2, IGF2BP1, ADCY7
Apoptosis	0.126793617	3.986568987	PSMA3, FNTA, DAPK1, PSMC2, PSMF1, SPTAN1, HIST1H1C, TNFRSF1A
Ascorbate and aldarate metabolism	0.126793617	14.92912792	UGT1A10, UGT1A9, UGT1A6
Glucuronidation	0.126793617	14.30635838	UGT1A10, UGT1A9, UGT1A6
Autodegradation of Cdh1 by Cdh1-APC/C	0.126793617	8.514211886	PSMA3, PSMC2, ANAPC5, PSMF1
Pentose and glucuronate interconversions	0.126793617	13.20453535	UGT1A10, UGT1A9, UGT1A6
Gene expression	0.126793617	2.263371086	EIF2B5, ZNF264, NUP210, RNMT, EHMT2, SMG6, AIMP2, PSMA3, RPL18A, HNRNPUL1, MED31, SNAPC4, PSMC2, ZNF567, PSMF1, FARSA, EXOSC3, ZNF354A
FSH regulation of apoptosis	0.126793617	3.654341737	AKAP12, ALDH1A3, DAPK1, CREM, CLU, ADCY7, TP53I11, ATP9A
Disease	0.126793617	2.509315376	NUP210, SRC, RNMT, ASNS, ADCY7, PSMA3, RPL18A, PSMC2, IGF2BP1, RCC1, PSMF1, EXOSC3, EXTL3, MBTPS2

Table 3.7: BioPlanet pathway analysis of shortening CR-UTR APA changes in Udorn-infected cells (Cytoplasmic).

3.3 Discussion

The aim of this chapter was to characterise the modulation of APA in influenza A H3N2 infected cells. An experimental approach utilising a powerful RNA subcellular fractionation technique with 3'RNA sequencing was employed (Figure 3.1), which enabled us to distinguish co-transcriptional and post-transcriptional APA events, while also elucidating differences in gene expression between nucleus and cytoplasm.

First, IAV Udorn propagation was carried out at a large scale, in order to ensure that enough infectious particles would be produced which could subsequently be used to extract nuclear RNA from sufficient numbers of infected mammalian cells. The propagated virus was immediately purified by ultracentrifugation through a sucrose gradient, to remove any contaminating host factors and cell debris. As lab-grown virus cultures often contain quantities of non-infectious particles (Tapia et al. 2019), the propagated virus had to be assayed to quantify the amount of infectious particles within the culture. This was done by plaque assay, which involves infecting a monolayer of cells with serially diluted concentrations of virus stock and then overlaying a semi-solid medium (culture media mixed with agarose). The cells were then incubated at 37 degrees Celsius for 2-3 days, at which point the semi-solid layer was removed and a stain used to identify dead cells in the monolayer. An example of a Udorn plaque assay can be seen in Figure 3.2A. Once the required amount of virus was prepared, human lung epithelial cells (A549) were infected over a timecourse (3hpi, 6hpi, 12hpi and 18hpi) chosen to provide a broad view of the progression of the Udorn infection. To ensure that an appropriate level of Udorn infection was achieved, we ran a small-scale infection which we

then visualised using immunofluorescence microscopy. We observed little sign of infection at 3hpi, early signs at 6hpi and strong indications at 12hpi and 18hpi. RNA was extracted from the human cells at these timepoints, and separated into fractions containing RNA from the nuclei and cytoplasms. As a quality control check, the RNA was run on a denaturing agarose gel to check the purity of the fractionation. The cytoplasmic RNA was indeed pure, as seen in Figure 3.2B (the nuclear pre-rRNAs are seen only in the nucleus band). 3'-enriched cDNA libraries were then made from the RNA samples which were sequenced on an Ion Torrent sequencing platform.

Next, to analyse the data, we assembled our own in-house bioinformatic pipeline, an outline of which can be viewed in Figure 2.1, chapter 2. This enabled us to use robust statistical methodologies (DEXSeq for APA analysis and DESeq2 for differential gene expression), while also empowering us with the ability to flexibly ask custom questions from the RNA-seq data which would not have been possible (or more difficult) using pre-built pipelines. This flexibility, for example, allowed us to analyse cis-element sequence differences in PASs regulated by APA (analyses which are discussed later in this section). Importantly, compared to similar pipelines which simply average counts between samples within a condition, DEXSeq takes biological variability into account and controls for false discovery with the Benjamini-Hochberg method. Our pipeline also accepts the PolyA_DBv3 (Wang, Nambiar, et al. 2018) as input, which is a database of annotated PASs found in the human genome generated by 3' region and extraction deep sequencing method (3'READS). This is an important step, as intersecting our reads with the PolyA_DBv3 nullifies the internal priming issue, which is a common problem faced in many APA analyses. DEXSeq calculates difference in PAS usage between

conditions, and calculates a p-adjusted value for each PAS. In our pipeline, the two most significant PASs are compared for relative usage - which constitutes an "APA event" if significant - and we calculate a value, called deltaRA, based on the relative isoform abundance resulting from each of these PASs (Equation 2.1, chapter 2). This value ultimately describes to what extent a gene is "lengthening" (overall relative isoform abundance shifts to the distal PAS) or "shortening" (relative abundance shifts to the isoform generated by usage of the proximal PAS). To achieve a high-resolution characterisation of the global APA profile, we split APA events into those occurring solely within the 3'UTR (UTR-APA), those solely occurring within the coding region (CR-APA), and APA events happening anywhere within the pre-mRNA ("All")(Figure 2.2). Crucially, the "All" group can be further broken down into those APA events involving one coding-region PAS and one 3'UTR PAS. These CR-UTR APA events directly impact the coding capacity of transcripts, so are particularly physiologically relevant. To investigate differential gene expression, we sum all reads for each transcript and pass a table containing these values through DESeq2.

We first looked at 3'UTR APA, and identified a total of 254 significant APA events (DEXSeq p-adj less than 0.1) in the nuclear sample and 271 APA events in the cytoplasmic sample (Figure 3.4A). There was a preference towards lengthening in the nucleus (143 vs 111 shortening events), whereas the opposite was observed in the cytoplasm (81 lengthening vs 190 shortening). These findings matched observations in the literature showing that longer 3'UTRs can host cis-elements which are detrimental to the stability of the transcript, including miRNA binding sites (Zhang, Liu, et al. 2021; Sandberg et al. 2008; Ji, Lee, et al. 2009) or decay-enhancing AU-rich elements (Garneau, Wilusz, and Wilusz 2007; Brewer

1991). Additionally, lengthened 3'UTRs have been reported to block nuclear export (Chen, DeCerbo, and Carmichael 2008). A recently published study details how long 3'UTR APA isoforms are specifically enriched in the nucleus, with this behaviour possibly being a result of nuclear retention elements present in the aUTRs of these isoforms (Tang, Yang, et al. 2022). On the other hand, shorter APA isoforms have been found to be abundant in the cytoplasm, possibly as a result of incomplete splicing (Neve, Burger, et al. 2016). The enrichment of longer 3'UTR APA isoforms in the nucleus and shorter 3'UTR APA isoforms in the cytoplasm during Udorn infection therefore likely results from a combinatorial effect of the aforementioned mechanisms. This APA isoform bias between compartments is possibly a gene expression regulatory tool used by the cell or induced by Udorn: genes involved in critical biological functions such as mRNA metabolism and splicing were enriched in the nucleus while important processes related to protein metabolism were enriched in the cytoplasm (Figure 3.5). While it is important to note that 3'UTR APA does not regulate coding capacity of transcripts, it does have the potential to modulate protein output.

Next, we identified widespread gene expression changes induced by Udorn infection. We observed 1096 changes in the nucleus and 2961 in the cytoplasm (Figure 3.6A). Approximately half of differentially expressed genes in the nucleus were not observed in the cytoplasm (Figure 3.6B), which is an important finding since the cytoplasm is the site of translation. Gene ontology analyses revealed highly significant enrichment of upregulated genes in the cytoplasm associated with many biological processes key to IAV replication within the cell, including protein-processing terms like "translation", "protein targeting to ER" and several oxidative phosphorylation-associated terms. Translation of viral proteins via host

machinery is important for the replicating virus, and continuous activity of the oxidative phosphorylation pathway is known to be crucial for the virus's replication (Bercovich-Kinori et al. 2016). Indeed, it has been reported previously that genes associated with oxidative phosphorylation are relatively untouched by IAV-induced host shutoff (Bercovich-Kinori et al. 2016), so it is particularly interesting to see that not only is this reflected in our data, but that these genes are in fact upregulated in the cytoplasm. Genes enriched for positive regulation of viral transcription and viral processes were also upregulated, suggesting a possible mechanism by which IAV orchestrates the regulation of host genes beneficial to its replication. Downregulation of genes included a wider assortment of biological processes, including host transcription related categories in the cytoplasm and "regulation of defense response to virus by host".

Detecting widespread gene expression changes induced by Udorn infection is an unsurprising finding, as IAV is known to induce substantial gene expression alterations (Gaucherand et al. 2019; Bercovich-Kinori et al. 2016; Thompson, Dittmar, et al. 2020). However our use of an RNA subcellular fractionation technique allowed us to go further and test for differential gene expression changes solely in the cytoplasm and, combined with our APA analysis, enabled us to derive potential regulatory links between Udorn-induced APA and host gene expression change. We overlapped genes regulated by 3'UTR APA with those genes that are differentially expressed, and found that 87 genes fit this criteria in the cytoplasm. These genes were highly enriched for virus- or immune-related biological processes (Table 3.3), suggesting a clear APA regulatory pathway for these genes which could impact their overall gene expression and alter either the ability of the host to generate an adequate immune response or the ability of Udorn to replicate within the cell, as

it is likely that APA changes are a product of both viral and host mechanisms. MAVS and ZC3HAV1 are two genes in this category that are associated with regulation of the viral genome. MAVS is a critical anti-viral immunity protein (Ren et al. 2020), and is a well-documented interactor of influenza A (Qian et al. 2017; Wang, Zhu, et al. 2019; Ren et al. 2020). ZC3HAV1 is a host antiviral factor known to regulate IAV replication through its interaction with the host immune response (Zhang, Goraya, et al. 2020). Further work investigating the functional impact of host 3'UTR APA on protein output of these two genes would affirm our findings.

It is easier to infer physiological consequences from coding-region APA than 3'UTR APA, even though there is a bias in the literature for 3'UTR APA research (Neve, Patel, et al. 2017). We observed substantial numbers of coding-region APA events in Udorn infected cells (Figure 3.8), with many more CR-UTR APA events than CR-CR and a heavy bias towards shortening. This is particularly interesting, as CR-UTR APA events involve genes switching from transcripts that contain the full coding capacity to truncated transcripts, or vice versa. CR-UTR APA is therefore a key regulatory mechanism that enables swift activation or inhibition of genes (Tian and Manley 2017). As mentioned, the heavy bias towards shortening coding-region APA events indicates inhibition of gene expression during IAV infection by coding-region APA. Indeed, gene ontology analyses of this large cohort revealed surprisingly few significantly enriched categories, suggesting that this mechanism could reflect a general host shutoff effect (Figure 3.8, bottom two charts). Conversely, the lengthening genes were highly enriched for protein-synthesis terms and influenza infection. RAN and HSP90AA1, are two genes in particular that switch to a distal PAS and therefore gain full coding potential, and

both have clear interactions with IAV. RAN is believed to play a role in influenza nuclear import (Eisfeld, Neumann, and Kawaoka 2015; Moreira, Yamauchi, and Matthias 2021) while HSP90AA1 enhances viral RNA synthesis via an interaction with IAV PB2 (Wang, Zheng, Lin, et al. 2020). Our findings here suggest that IAV-induced host shutoff inhibits host gene expression via coding-region APA in a broad, non-specific manner, while genes beneficial to the viral life cycle are either untouched or upregulated. IAV PA-X, an auxiliary flu protein, could be implicated in this process, which selectively degrades host mRNAs (Khaperskyy, Schmaling, et al. 2016) and is thought to induce a more generalised form of host shutoff than NS1 (Dunagan, Hardy, and Takimoto 2021).

3.4 Summary

In this chapter, we set up a robust bioinformatic pipeline, and used powerful computational tools to eliminate potential experimental artifacts and account for biological variation between samples. We identified widespread APA and differential gene expression changes in Udorn-infected cells, and because of our subcellular RNA fractionation approach, we were able to pinpoint these to the appropriate cellular compartment. We observed APA and gene expression changes in biological process categories functionally relevant to viral replication and host immune response. We identified several coding-region APA changes which could modulate viral replication in cells, although additional work would confirm their physiological relevance. Taken together, our findings report broad changes in the APA profiles of cells infected with the H3N2 Udorn IAV strain, many of which are likely to be physiologically relevant.

Chapter 4

Characterising the host alternative cleavage and polyadenylation landscape in influenza A H1N1 WSN-infected cells

4.1 Introduction

In the previous chapter, we investigated gene expression and alternative cleavage and polyadenylation (APA) profile changes induced by H3N2 influenza A/Udorn/72 strain ("Udorn") infection. We demonstrated the presence of widespread APA and gene expression changes in Udorn-infected cells by developing and employing a custom-made pipeline which accounted for biological variation between samples.

However, as Udorn proved difficult to propagate in large quantities, we were not able to perform true biological replicates of the experiment. Our original vision of using the newly-established pipeline to compare infected cells with uninfected cells through a timecourse with replicates had to be modified and we instead compared early Udorn infection (by pooling the 3hpi and 6hpi timepoints) to late infection (by pooling the 12hpi and 18hpi timepoints).

To support the Udorn findings, we next used the pipeline to investigate IAV-induced APA changes on a more discrete dataset that included repeats at multiple timepoints. To that end, we utilised a 3'RNA-seq dataset generated in our lab using the same methodology as previously described for the Udorn infected cells, with some subtle but important differences (depicted in Figure 3.1). Briefly, this means that an influenza stock (using a new H1N1 influenza A/WSN/33 strain) was propagated and quantified using plaque assays, human lung adenocarcinoma (A549) cells were subsequently infected at a ratio of five infectious particles to every one mammalian cell, nuclear and cytoplasmic RNA was extracted at various timepoints (0hpi, 3hpi, 6hpi, 9hpi, 12hpi and 18hpi) and 3'-end enriched libraries were sequenced using an Ion Torrent platform. To sum up the key differences with this dataset: 1) instead of Udorn, the H1N1 influenza A/WSN/33 strain (hereafter referred to as "WSN") was used, which was much easier to propagate in bulk; 2) the timecourse included the timepoints 0hpi, 3hpi, 6hpi, 9hpi, 12hpi and 18hpi; 3) importantly, two biological replicates of the infection were generated in full. The experiments done to generate this dataset were carried out together with Alastair Louey.

Methodology used in this chapter

Most analyses in this chapter used the purpose-built pipeline described in detail in section 2.7.1, chapter 2. Briefly, differential usage of PASs was determined by the statistical method DEXSeq (Anders, Reyes, and Huber 2012) with relative switching between the top two most significant PASs (with a p-adjusted value of less than 0.1) considered "lengthening" if the distal PAS was preferred in WSN infection, or "shortening" if the proximal was favoured. Significant alteration of PAS usage was considered an "APA event", and APA events were split into three groupings (see Figure 2.2, chapter 2) which separate APA events into those occurring in the UTR, coding region, or anywhere within the gene. For differential gene expression analysis, DESeq2 was employed, and genes were considered to be significantly downregulated or upregulated if they were assigned a p-adjusted value of less than 0.1. Both DEXSeq and DESeq2 take biological variation into account (rather than simply averaging values between replicates/conditions) which is an important factor to consider for the WSN infection data which was done in duplicate. In practical terms, the replicates for each timepoint were processed through the pipeline and compared for significant changes with the uninfected samples, also in replicate. This was done separately for RNA from both the nuclear and the cytoplasmic fractions.

Background on the H1N1 influenza A/WSN/33 strain

WSN is a strain of the influenza A virus which, unlike Udorn (H3N2), expresses the H1 and N1 antigens on its surface. It has been commonly used in lab settings for many decades to investigate influenza properties (Choppin et al. 1969). WSN

mainly produces spherical particles whereas Udorn populations consist of a mixture of spherical and filamentous particles (Roberts et al. 2013), a distinct property believed to be partly responsible for the difference in ease of lab propagation between the two strains (Dadonaite, Vijayakrishnan, et al. 2016).

Aims

This chapter aimed to generate a clear profile of host APA and differential gene expression changes during an infection timecourse with IAV. The WSN dataset - with replicates generated for each timepoint - enabled us to carry out granular analyses of profile changes to better understand how the presence of IAV induces host APA in the host cell environment and how these changes have physiological implications by impacting not only the ability of the virus to replicate but also by impacting host defence. Additionally, the utilisation of distinct timepoints enabled a more detailed study of cis-element mechanisms behind the regulation of APA in IAV-infected cells than was possible using the Udorn data from the previous chapter. Furthermore, observations from the analyses performed in this chapter were compared with the Udorn findings to explore whether changes in host APA profiles were strain-specific or shared between IAV strains.

4.2 Results

4.2.1 WSN infection induces widespread 3'UTR APA changes

As discussed in the previous chapter, genes with multiple PASs in the 3'UTR (untranslated region) have the potential to be regulated by APA. Differential selection

of PASs (poly(A) sites) within the 3'UTR (untranslated region) can ultimately lead to modified protein output, by retaining or removing cis-elements affecting the stability, translational efficiency or nuclear export capacity of the nascent transcript (Tian and Manley 2017). When we compared early infection against late infection in the Udorn dataset, we saw widespread induction of 3'UTR APA profile changes (documented in the previous chapter). Here, we utilised the WSN dataset to track host 3'UTR APA changes in a stepped way at discrete timepoints (0hpi, 3hpi, 6hpi, 9hpi, 12hpi and 18hpi) throughout an infection period. To this end, we analysed the 3'RNA-seq data from WSN-infected cells using our purpose-built pipeline to test for significant changes in host 3'UTR PAS usage (described in section 2.7.1).

At the early infection phase (3hpi), though 3'UTR APA events were approaching significance, none were below the p-adjusted threshold of 0.1 in either the nucleus or cytoplasm at 3hpi (hours post infection) (Figure 4.1). This was unsurprising, as it would be expected that some time would be needed for viral proteins to be translated and imported back into the nucleus, and for the host immune response to react. Towards the mid-infection phase (6hpi), a major jump was observed in the nucleus to 534 significant 3'UTR APA events, suggesting that as the viral infection progressed, a stronger effect was induced in the host transcriptome. 3'UTR PAS switching by genes in the nucleus steadily increased between 6hpi and 9hpi, and peaked at 12hpi with 1047 before declining somewhat to 687 at 18hpi. In the cytoplasm, we observed a low number of significant 3'UTR APA events in early-mid infection (3hpi, 6hpi and 9hpi), until a rapid linear rise was seen from 76 events at 9hpi to 720 at 18hpi. Tables detailing the genes experiencing the most significant 3'UTR PAS switches can be seen in the Appendix (Nuclear data shown in Tables A.13 - A.16 and Cytoplasmic data shown in Tables A.17 - A.20).

A striking difference when comparing host 3'UTR APA profiles in the nucleus and cytoplasm was the timing offset between substantial numbers of APA events occurring. Hundreds of genes switched 3'UTR PASs in the nucleus at 6hpi onwards, whereas only at 12hpi in the cytoplasm is the one-hundred threshold reached. This observation suggests that much of the 3'UTR APA induced in WSN infection was occurring co-transcriptionally in the nucleus, and that many of these transcripts were not exported to the cytoplasm (or if they were, they were degraded rapidly). As mentioned in the previous chapter, this effect could in part be attributed to the IAV NS1 protein, which interacts with host cleavage and polyadenylation (CPA) by binding CPSF30 (discussed in the introduction), and is mainly localised in the nucleus early in infection (Ayllon and García-Sastre 2014). Furthermore, there were substantially more shortening events in the cytoplasm than the nucleus, as a proportion of total significant APA events (Figure 4.1): approximately half of all APA events in the cytoplasm and a fifth of those in the nucleus featured upregulated proximal PASs. This suggests that the nuclear export rate and/or stability of nuclear distal APA isoforms was modulated in such a way as to impede their ability to penetrate into the cytoplasm. For example, lengthened 3'UTRs have been reported to block nuclear export (Chen, DeCerbo, and Carmichael 2008; Tang, Yang, et al. 2022) and contain stability-compromising cis-elements (Zhang, Liu, et al. 2021; Sandberg et al. 2008; Ji, Lee, et al. 2009). Finally, when compared with data from the Udorn infection (discussed in the previous chapter), approximately half of the genes identified to be regulated by 3'UTR APA in that dataset were also seen in the WSN infection (see Appendix Figure A.1), indicating that different IAV strains induce similar host 3'UTR APA profiles.

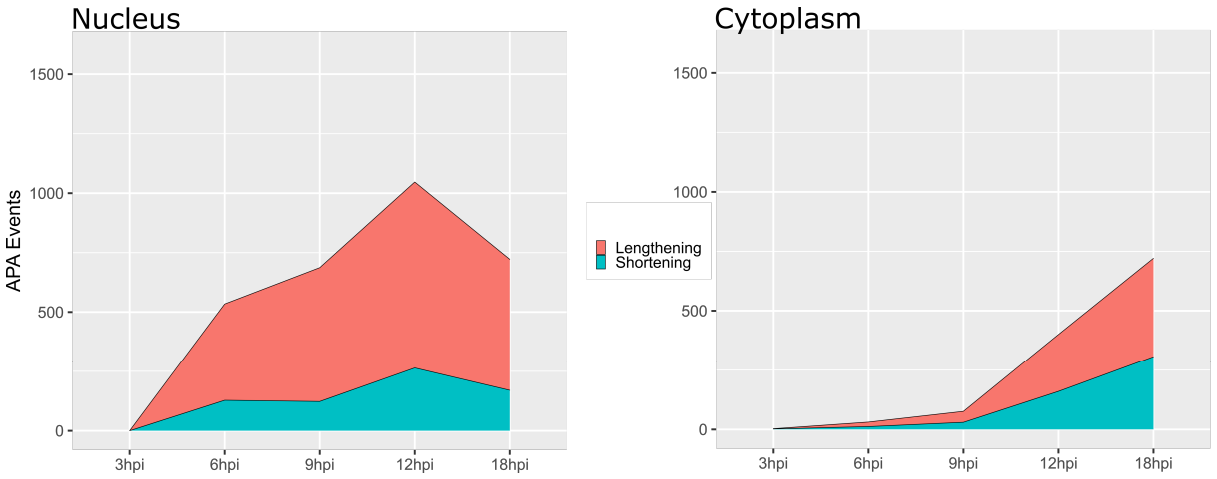


Figure 4.1: Overview of 3'UTR APA changes within nuclear and cytoplasmic fractions during the WSN infection. The numbers of 3'UTR APA events were tracked across the WSN infection in both the nucleus (left) and cytoplasm (right).

The 3'UTR APA landscape is highly dynamic in WSN-infected cells

IAV infection is a multi-phase process, consisting of various stages such as the attachment of the virion to cell surface, endocytosis of the virion, replication of viral RNA and release of progeny virions (reviewed in section 1.4) (Dou et al. 2018). As such, we expected the 3'UTR APA landscape to be highly dynamic between various points in the timecourse. Indeed, IAV NS1, the major host shutoff protein believed to interact with CPSF30 and therefore thought to be largely responsible for dysregulating host polyadenylation, is believed to accumulate in the nucleus early in infection, and then later transported to the cytoplasm (Ylösmäki et al. 2016).

Do the same genes switch 3'UTR PASs right throughout the infection, or are most APA events seen only in the early or late infection? Are these genes consistently lengthening or shortening, or do they change direction between timepoint

or subcellular compartment? We next utilised the WSN dataset to address these questions. One of the major advantages of the WSN infection data (over the Udorn-infection data used in chapter two) is the ability to detect APA changes at discrete timepoints during the infection period, enabling the tracking of APA events at the gene-level throughout IAV infection. To examine the dynamic nature of 3'UTR APA changes, we collated all genes regulated by 3'UTR APA in both cellular compartments and extracted their deltaRA values, which measures the overall APA direction for each gene (lengthening or shortening). The deltaRA measure is explained in detail in section 2.7.1, chapter 2, but in brief, a deltaRA value between 0 and 1, represents a shift towards shortening, with 1 representing a total shift to the shortened APA isoform. Likewise, a value between 0 and -1 represents lengthening, with -1 indicating complete shift of a gene's APA isoforms towards the distal APA isoform.

A heatmap of this data revealed the APA trends of those genes which switch PASs throughout infection, in both the nuclear and cytoplasmic samples. In Figure 4.2, each row represents a single gene that is significantly regulated by APA in at least one timepoint, with colour representing the gene's deltaRA value at each timepoint in both cellular compartments. Interestingly a substantial number of genes demonstrated a sustained impact from 3'UTR APA changes - being consistently lengthened or shortened in multiple timepoints - in either the nucleus or cytoplasm, whereas surprisingly few APA events appeared to penetrate between compartments. This was particularly evident with those lengthening genes in the nucleus, most of which did not appear to also lengthen in the cytoplasm. These findings further demonstrate that many nuclear APA events do not penetrate into the cytoplasm.



Figure 4.2: All significant 3'UTR APA events across the WSN infection, in both cytoplasmic and nuclear samples. Each row in the heatmap represents a gene that is regulated by APA in at least one timepoint during WSN infection, with the colour representing that gene's deltaRA value at each timepoint, in both the nucleus and the cytoplasm. n=2079.

Further analysis revealed that 193 genes PAS switched in every timepoint from 6hpi onwards in the nucleus, with hundreds more switching in at least two timepoints (Figure 4.3A). This was a starkly different situation to the cytoplasmic APA events, where only 5 genes were seen to be regulated by 3'UTR APA in every timepoint (excluding 3hpi), with almost 300 additional genes switching 3'UTR PASs in at least two timepoints (Figure 4.3B). Furthermore, roughly 1,189 genes uniquely switched 3'UTR PASs in the nucleus, with an additional 333 genes regulated by APA in both the nucleus and cytoplasm and 556 genes exclusively switching their 3'UTR PASs in the cytoplasm (Figure 4.3C).

Overall, this analysis showed that while many of the 3'UTR APA events that we detected in WSN-infected cells were sustained between multiple infection timepoints, they were often exclusive to either of the subcellular compartments. As mentioned earlier, this effect could be a result of mechanisms mentioned previously, such as longer 3'UTRs blocking nuclear export or the selective degradation of transcripts by PA-X.

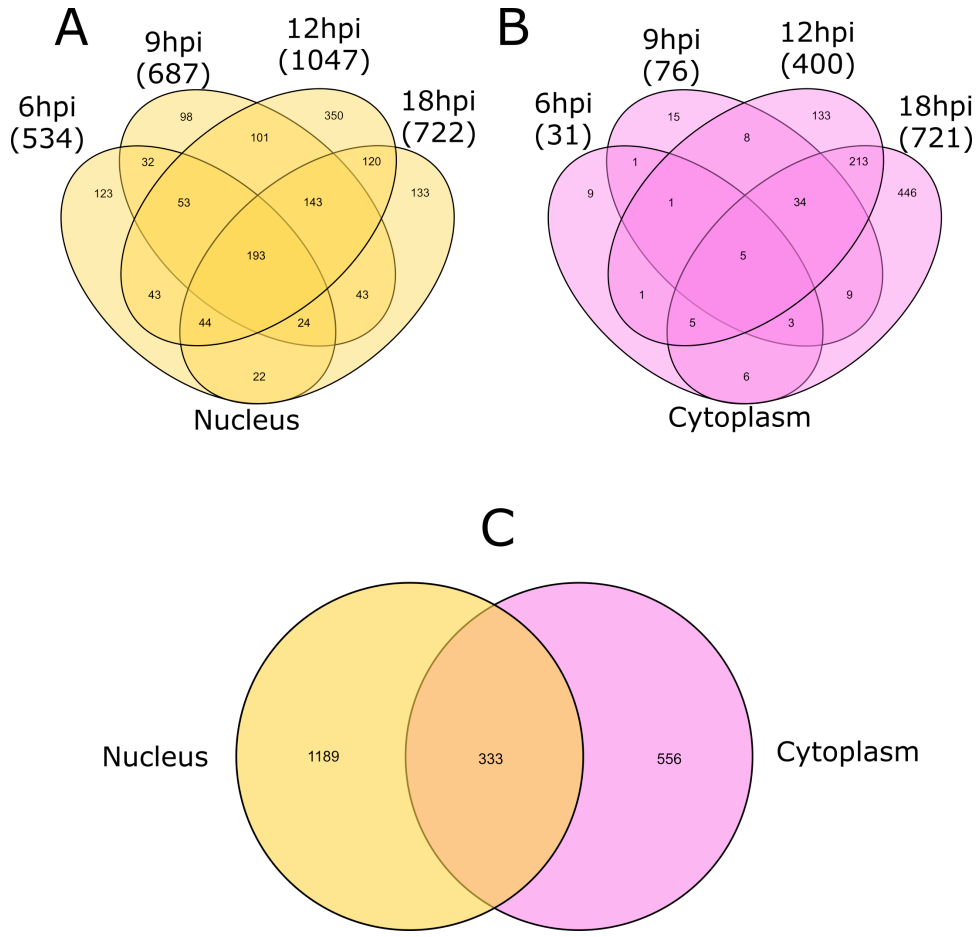


Figure 4.3: Overlaps of genes experiencing a 3'UTR APA change in WSN-infected cells. Venn diagrams show the total number of 3'UTR APA events are shown in each timepoint, both in the nucleus (A) and in the cytoplasm (B). The first timepoint (3hpi) has been excluded due to an absence of significant APA events. C) The total number of unique genes regulated by 3'UTR APA at any timepoint in the nucleus and cytoplasm are shown overlapped.

4.2.2 Genes regulated by 3'UTR APA are highly enriched for mRNA processing in WSN-infected cells

As we had observed substantial changes in the host 3'UTR APA landscape during WSN infection, we next wanted to ask what physiological consequences these changes may have. To that end, we explored which biological functions might be enriched in genes that switch 3'UTR PASs. We separated 3'UTR APA events into lengthening and shortening groups in either the nuclear or cytoplasmic samples. To focus on those APA events most persistent throughout the infection, we extracted genes that either lengthened or shortened in multiple timepoints (two or more). In total, there were 637 genes that lengthened by switching to distal PASs in the nucleus in multiple timepoints, and 147 that shortened by switching to proximal PASs. In the cytoplasm, 171 genes lengthened and 103 genes shortened at multiple points during the infection. These lower numbers in the cytoplasmic fraction reflect the lower numbers of total 3'UTR APA events than when compared with the nucleus (as seen earlier in Figure 4.1).

We performed gene ontology analyses on the genes we extracted using the Enrichr tool (Chen, Tan, et al. 2013; Kuleshov et al. 2016; Xie et al. 2021) to identify functions that may be enriched in each category. The top ten most significant biological processes associated with each category of genes are shown in Figure 4.4, and listed in further detail in Tables A.21 - A.24 in the Appenidx. Interestingly, RNA-processing related functions were massively enriched in the nuclear lengthening category (Table A.21), with processes such as RNA splicing, mRNA processing, rRNA processing and gene expression also highly enriched. Some of the genes shared between the mRNA-processing and mRNA splicing categories

included several SR-family proteins (SRSF1, SRSF3, SRSF6, SRSF7, SRSF9 and SRSF10 and a HNRNP family-member, HNRNPK). Notably, IAV hijacks host splicing machinery to process some of their own RNA segments, and the SR-family is believed to be involved in this (Artarini et al. 2019; Nilsson, Abdurahman, and Schwartz 2020). And, while HNRNPK has been implicated in the usage of host splicing by IAV for its own purposes (Thompson, Muñoz-Moreno, et al. 2018; Tsai et al. 2013), it has also been recently reported that HNRNPK is involved in many host alternative splicing events that are induced upon IAV-infection (Thompson, Dittmar, et al. 2020). Therefore, regulation of these genes at the 3'UTR APA level could ultimately influence directly the processing of host and viral mRNA.

There were fewer significantly enriched biological processes in the nuclear shortening category (Table A.22), with transcription regulation featuring heavily. In the cytoplasmic lengthening category, terms related to protein localisation and transport were enriched, in addition to processes associated with cell cycle regulation and ribosome biogenesis. Intriguingly, lengthening genes in the cytoplasm (Table A.23) were also highly enriched for mRNA processing (a function also enriched in nuclear lengthening genes). Two of these - CPSF6 and WDR33 - are important for assembly of the CPSF complex (a critical cleavage and polyadenylation multi-subunit factor), while another two, CELF1 and HNRNPA0 have been reported as potential host factors important for IAV replication (Ackerman et al. 2019). Most (6 of 9) of the genes linked to mRNA processing in the "cytoplasmic lengthening" category did not also lengthen in the nucleus, indicating that lengthening of these genes could be a result of post-transcriptional APA events. There were no biological process categories in the cytoplasmic shortening group (Table A.24) with p-adjusted values of less than 0.1, although several were approaching

that significance threshold, including protein oligomerisation and protein localisation/transport.

3'UTR APA has the potential to modulate the expression of genes by altering transcript properties such as translational efficiency, transcript stability or transcript localisation and therefore regulate gene expression. We identified many genes involved in critical biological processes that switch 3'UTR PASs during WSN infection. The SR-family was particularly enriched in the nuclear lengthening category and, along with HNRNPK (which also lengthens in the nucleus), has already been implicated in the splicing of IAV RNA segments (Thompson, Muñoz-Moreno, et al. 2018) and in the alternative splicing of host transcripts during IAV infection (Thompson, Dittmar, et al. 2020). Furthermore, HNRNPK has been seen to play a role in the selection of alternative PAS sites by arresting binding of the CFI complex (Naganuma et al. 2012; Mikula et al. 2013). As well as this, CELF1 and HNRNPA0 were found to lengthen in the cytoplasm, with both known to be host factors linked to IAV replication. These findings suggest that 3'UTR APA plays a major regulatory role in the expression of host splicing genes which could have physiological implications for viral infection. 3'UTR APA was also seen to regulate the expression of host factors important for IAV replication. Whether or not these 3'UTR APA changes are induced by viral or host mechanisms, and are overall beneficial or suppressive to viral replication, is not clear from these findings.

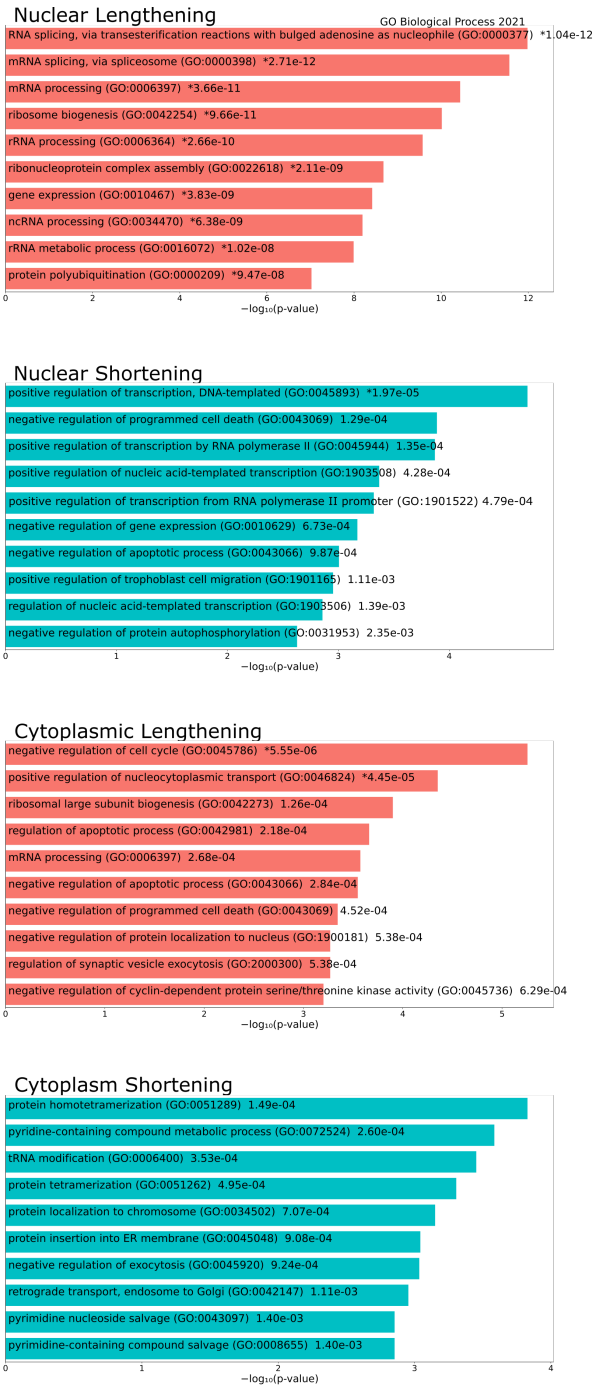


Figure 4.4: Gene ontology analyses of 3'UTR APA events detected in WSN-infected cells in "biological process" category. Each category is shown separately, with -log10(pvalues) adjusted for multiple testing by the Benjamini-Hochberg method represented on the x-axis.

4.2.3 aUTR motif analysis of lengthening nuclear transcripts

In the previous sections, we have quantitatively described host 3'UTR APA events during a WSN infection period, and have identified a bias towards nuclear lengthening when compared with cytoplasmic 3'UTR APA events. We also reported physiologically relevant functions which are enriched in those genes regulated by 3'UTR APA. We next wanted to explore mechanisms that may give rise to this bias. It is well-established that aUTRs (alternative 3'UTRs; i.e. UTRs present in the distal but not the proximal APA isoforms) can possess cis-elements that affect properties of the transcript. The mechanism behind this model involves certain cis-elements present in the aUTR acting as binding platforms for RNA-binding proteins (RBPs) and miRNAs, which together can affect the stability or nuclear export efficiency of the nascent transcript (Zhang, Liu, et al. 2021; Samuels et al. 2020; Tushev et al. 2018). Therefore, we were interested in what motifs may be present in aUTRs of genes regulated by APA during WSN infection.

To investigate whether or not aUTR cis-elements may play a role in the abundance of distal APA isoforms in the nucleus relative to the cytoplasm, we extracted the aUTR sequences of those lengthening genes present in the nucleus but not in the cytoplasm at 12 hours post infection, when this disparity was most pronounced. This cohort consisted of 694 aUTR sequences in total, once a small number of outliers with aUTR lengths of over 20,000 nucleotides were removed. Using the computational MEME SUITE tool (Bailey, Boden, et al. 2009) we compared the aUTR sequences of these 694 genes to aUTRs from genes which did not PAS switch. This was a necessary control, as aUTRs are relatively abundant in

certain motifs (such as miRNA and RBP binding sites) (Tushev et al. 2018).

Using this approach, over 20 significant motifs were identified and processed with the Tomtom tool (Gupta et al. 2007) to identify RNA-binding proteins which bind these motifs, indicating that aUTR motifs play a critical role in the compartmentalisation of long nuclear transcripts during flu infection. The 5 most significant motifs and RBPs are shown in Figure 4.5, with significance given as E-values, which is the p-value calculated by Fisher's exact test, multiplied by the number of motifs reported. Strikingly, almost every aUTR (693 out of 694) was significantly enriched with U-rich HuR/ELAVL1 binding motifs. ELAVL1 is a multi-functional protein that has previously been reported as a positive regulator of 3'UTR lengthening (Elkon, Ugalde, and Agami 2013; Turner and Diaz-Munoz 2018), and the high significance of a binding motif for ELAVL1 in the aUTRs of lengthened transcripts in the nucleus suggests that it plays a role in positively regulating the selection of distal PAS choice in the nucleus during WSN infection. Additionally, significantly enriched motifs associated with other RBPs such as SAMD4A, SRSF10, PCBP2 and HNRNPK could implicate these proteins in the positive regulation of lengthening by APA. As these motifs were identified from APA transcripts in the nucleus, it is possible that regulation by these RBPs occurs co-transcriptionally or shortly after transcription. For example, it is believed that regulation of APA by ELAVL1 is brought about by the binding of ELAVL1 at U-rich sites in the 3'UTR of pre-mRNAs, thereby blocking the binding of core cleavage and polyadenylation factors at particular PASs (Zhu, Zhou, et al. 2007). Additionally, both SRSF10 (Jobbins et al. 2022) and HNRNPK (Naganuma et al. 2012; Mikula et al. 2013) have been been reported to alter polyadenylation site choice. Another possibility is that these RBPs bind APA

isoforms post-transcriptionally in the nucleus and consequently affect transcript properties like stability or nuclear export. For instance, the Drosophila variant of SAMD4A has roles in the translational suppression and destabilisation of mRNA (Chen, Dumelie, et al. 2014; Nelson, Leidal, and Smibert 2004) while PCBP family members (including PCBP1/2 and HNRNPK) also have similar gene regulation functions (Yuan, Chen, and Cai 2021).

The highly enriched motifs of RBPs occurring in aUTRs of genes in the nucleus noted in this section represent a mechanism for gene regulation by 3'UTR APA during WSN infection. Further work is required to address more fully the roles that these RBPs may play in PAS selection. This could involve depleting protein levels of these factors and probing if overall PAS usage shifts in the absence of the RBP.

	E-value	RBP	aUTRs
	6.4e-035	SAMD4A	261
	2.7e-030	PCBP2	336
	3.8e-029	HuR, ZC3H14	693
	6.2e-024	SRSF10	207
	1.0e-022	PCBP1, HNRNPK	298

Figure 4.5: Enriched aUTR motifs in distal APA isoforms in WSN-infected cells, nucleus, 12hpi. The top 5 most significant motifs detected in the aUTRs of lengthening genes in the nucleus are shown, including the RBPs that may bind them. Total number of aUTRs searched is 694.

4.2.4 Differential gene expression changes are induced in WSN-infected cells

Just as the spread of infection timepoints in the WSN dataset provided a granular view of 3'UTR APA changes at discrete stages during infection, we next wanted to explore how overall gene expression levels are altered in the context of IAV infection at distinct timepoints. These findings could then be compared with those of our Udorn dataset, but could also be importantly utilised in conjunction with our APA analyses to explore the link between APA and gene expression.

As in chapter 3, to carry out our differential gene expression analysis we used the DESeq2 computational method as a core component of our pipeline, and we considered a p-adjusted value of less than 0.1 to represent a significantly regulated gene. The first step in our analysis to investigate interplay between 3'UTR APA and differential gene expression was to quantify all significantly regulated genes using the method described in both the nucleus (Figure 4.6, left) and the cytoplasm (Figure 4.6, right). Somewhat mirroring the host APA profiles, more substantial differential gene expression changes were detected in the nucleus earlier in infection compared to the cytoplasm; there were 2287 significantly regulated genes in the nucleus at 9hpi and just 702 in the cytoplasm at the same timepoint. At every nuclear timepoint, there was a higher proportion of upregulated to downregulated genes (1505 to 1111 at 12hpi respectively). In contrast to this, the breakdown between upregulation and downregulation was much more even in the cytoplasm, with closer to equal ratios at each timepoint. While there were substantially more regulated genes in early timepoints in the nucleus, at 18hpi the inverse was true with 3339 differentially expressed genes in the cytoplasm compared to 2558 in the

nucleus. Additionally, gene expression changes appeared to persist throughout the infection: 1424 and 1005 genes were upregulated or downregulated respectively in the nucleus in multiple timepoints, whereas approximately 800 genes were upregulated and a further 800 downregulated in the cytoplasm in more than one timepoint. Overall, the trends of upregulation and downregulation in both subcellular compartments matched those seen in our Udorn dataset. Venn diagrams showing overlapping gene expression changes between timepoints of the WSN infection can be seen in Appendix A.3.

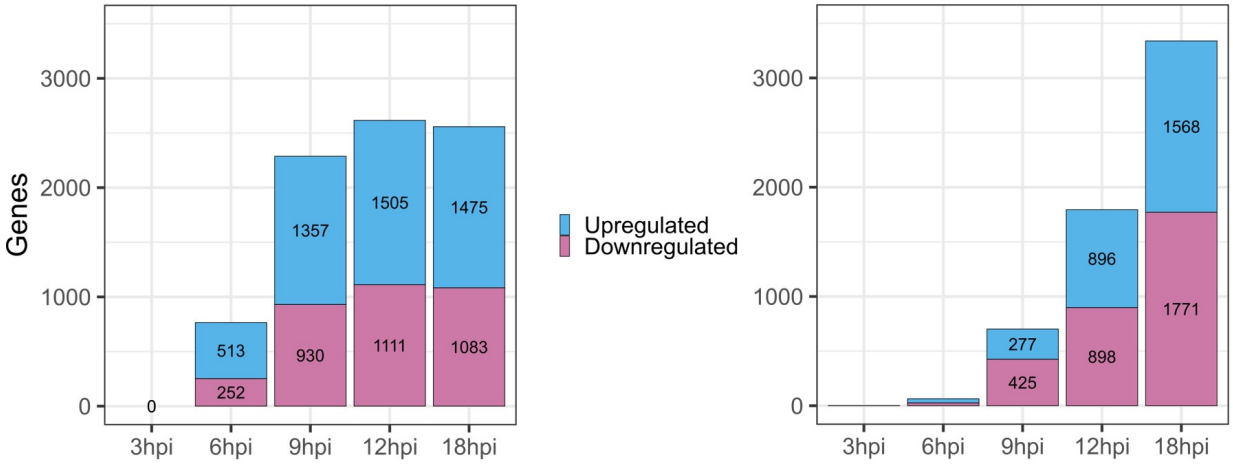


Figure 4.6: Barcharts of differential gene expression changes in WSN-infected cells. Numbers of significantly regulated genes in the nucleus are shown in the right chart, while those detected in the cytoplasm are shown in the left.

When analysing the differentially regulated genes in each category for potentially enriched biological processes, we found a bias for genes associated with mRNA catabolic processes, SRP-dependent protein targeting, protein targeting to endoplasmic reticulum, translation and regulation of transcription in both the nuclear and cytoplasmic samples, while terms associated with DNA replication and repair were downregulated in both compartments (Figure 4.7, Tables A.25

- A.28 in Appendix). These findings informed us that the expected robust host gene expression response was induced during WSN infection, and largely matched what has been reported in the literature (Ashraf et al. 2020; Thompson, Dittmar, et al. 2020). Importantly however, the GO analyses showed functionally distinct enrichment patterns from those genes regulated by 3'UTR APA (compare Figure 4.7 with Figure 4.4), suggesting that 3'UTR APA changes are not simply resulting from wider transcriptional changes and that different mechanisms could be at play for induction of differential gene expression and APA. Interestingly, genes associated with mitotic spindle organisation and microtubule cytoskeleton organisation involved in mitosis and mitotic sister chromatid segregation were highly significantly downregulated in the nucleus. A link between influenza infection, the interferon response and mitotic spindle disorganisation has recently been drawn, and it is suggested that interferon induces a systemic effect on cell division (McDougall et al. 2019).

Altogether, this gene expression analysis confirmed that widespread changes in differential gene regulation were induced during WSN infection, and we could proceed to investigate interplay of gene expression change with host APA. This functional analysis, when compared with that discussed in section 4.2.2, also strongly suggests that viral-induced host APA changes are not a side effect of wider transcriptional dysregulation. Additionally, gene expression in the nucleus was overall consistently enhanced at every timepoint, while near-equal numbers of genes in the cytoplasm were upregulated and downregulated. Furthermore, many gene expression changes were persistent throughout the infection (detected in at least two timepoints), suggesting sustained transcriptomic responses to IAV infection.

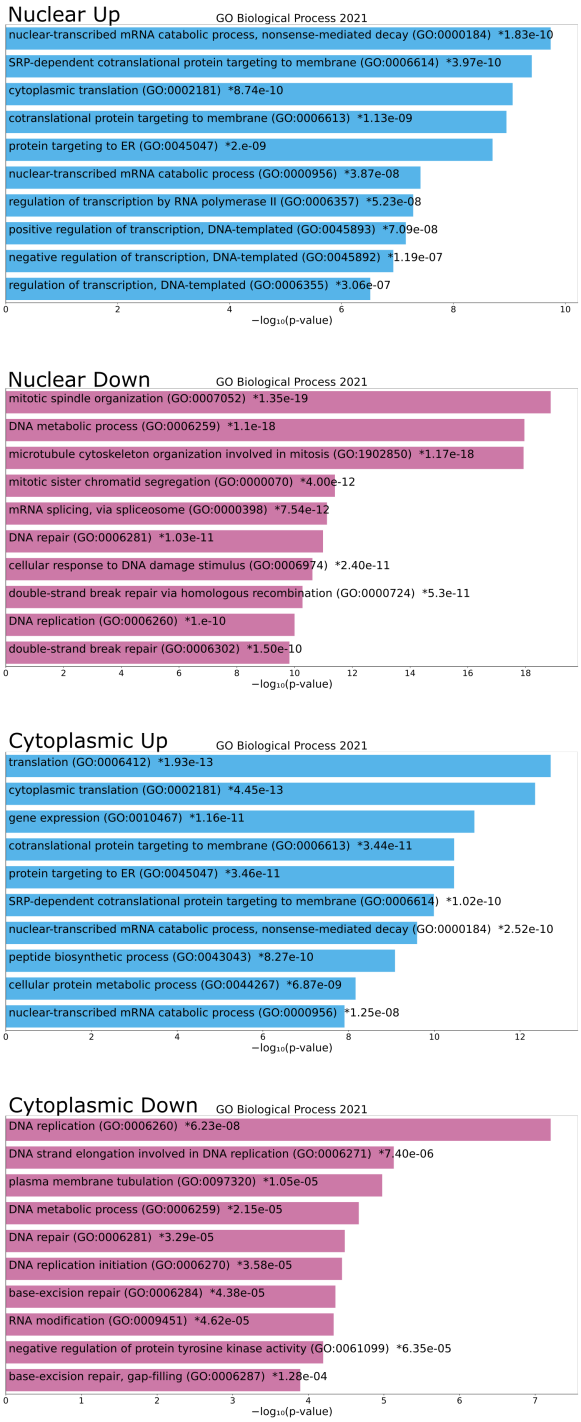


Figure 4.7: Gene ontology analyses of genes differentially expressed in WSN-infected cells ("biological process" category). Gene expression changes are broken down into those that are upregulated or downregulated, in both the nuclear and cytoplasmic fractions, with -log10(pvalues) adjusted for multiple testing by the Benjamini-Hochberg method represented on the x-axis.

4.2.5 Interplay between host 3'UTR APA and differential gene expression in WSN-infected cells

In the previous section we confirmed a strong induction of differential gene expression in WSN-infected cells, which corroborated what has previously been reported in the literature, although our data adds another dimension by separately analysing the nuclear and cytoplasmic samples. However, little is understood about what interplay there might be between these viral-induced alterations of gene expression levels and host 3'UTR APA, a regulatory mechanism which has the potential to affect important properties of transcripts. Earlier (in section 4.2.1), we highlighted substantial changes in the host 3'UTR APA landscape of cells during WSN infection. Next, we wanted to ask if there were any genes that switched PASs in their 3'UTRs which also experienced differential gene regulation in WSN-infected cells compared to non-infected cells. We specifically isolated those genes that either lengthen or shorten in the nucleus, and experienced a significant change in their cytoplasmic gene expression levels. This group therefore represents those genes where PAS switching in the nucleus to another 3'UTR PAS is most likely to contribute to a change in protein levels.

We noted in section 4.2.2 that several SR-family genes and HNRNPK switched to distal 3'UTR PASs in the nucleus, so we next asked if their gene expression levels in the cytoplasm were altered. We found that the expression of many SR genes and HNRNPK was also upregulated in the cytoplasms of WSN-infected cells. The SR-family of proteins bind pre-mRNA in the nucleus and have known involvement in alternative splicing (Shepard and Hertel 2009) and links to alternative polyadenylation (Shen et al. 2019). HNRNPK also binds pre-mRNA and

is known to regulate transcript stability by binding 3'UTRs (Yin et al. 2020). Furthermore, HNRNPK is also known to regulate polyadenylation site choice - in one study, HNRNPK was found to disrupt binding of the important CFI complex near a proximal alternative polyadenylation site, thereby suppressing a poly(A) site and leading to the generation of an alternate, longer isoform (Naganuma et al. 2012). This study's findings could in part explain the abundance of nuclear 3'UTR lengthening that we have documented in flu infection (discussed earlier in section 4.2.1). Interestingly, the SR-family proteins and HNRNPK have also been reported to be splicing factors hijacked by IAV to aid in the splicing of its own RNA segments (Artarini et al. 2019; Nilsson, Abdurahman, and Schwartz 2020; Thompson, Muñoz-Moreno, et al. 2018; Tsai et al. 2013) while HNRNPK also is thought to be responsible for many host alternative splicing events induced by WSN infection (Thompson, Dittmar, et al. 2020).

We also found that PSMB1, a gene that encodes a subunit of the 20S proteasome complex, switches to a longer 3'UTR APA isoform while also experiencing a gene expression increase in the cytoplasm. The PSMB family plays a role in the immune system (Yasutomo 2019), and PSMB1 has recently been reported to negatively regulate virus- and IAV-induced immune responses (Wu, Niu, et al. 2019) (which, if overexcited, could be damaging). Overexpression of PSMB1 substantially inhibited IFN-beta promoter activation while silencing increased IFN-beta levels (Wu, Niu, et al. 2019). VAPA, another gene that lengthens in the nucleus while having increased gene expression levels in the cytoplasm, has also been implicated in the host immune response during IAV infection. Interactions of VAPA with the antiviral effector protein IFITM3 restricts IAV infection by mediating the accumulation of cholesterol in endosomes, thereby blocking viral entry

(Amini-Bavil-Olyaee et al. 2013). These findings suggest that 3'UTR lengthening of PSMB1 and VAPA induced by APA in the nucleus could positively regulate their gene expression levels in the cytoplasm during IAV infection, thereby modulating the host immune response.

Next, we looked at those genes which lengthened in the nucleus but which also experienced downregulated cytoplasmic gene expression, as this group represents those genes for which nuclear 3'UTR lengthening is most likely to contribute to overall downregulation. Interestingly, upon running a gene ontology analysis for this group (not shown), we found that TLR6 and IL17RA, two immune receptors which are positive regulators of cytokine production were included in this category. TLR6 stimulation, along with other TLRs, was found to confer protection against influenza pneumonia in mice and this effect was inducible before and after infection (Tuvim et al. 2012). IL17RA, a receptor for the IL17A and IL17F cytokines, has been reported to play a critical role in the host immune response to influenza infection by mediating the recruitment of neutrophils, with IL17RA knockout mice experiencing less lung injury (Tuvim et al. 2012). The PAS-switching of these two cytokine regulators could play a role in their negative changes in gene expression. By promoting their lengthening and nuclear retention, 3'UTR APA could represent an important layer of regulation of the host immune response during IAV infection.

Shortening of 3'UTRs, too, has the potential to alter transcript properties and subsequently affect overall gene expression levels, so we asked what genes that switched to proximal 3'UTR PASs were also upregulated in the cytoplasm. Interestingly JUN, which encodes a protein that forms a major transcription factor (AP-1), was seen to have increased gene expression in the cytoplasm while also shortening in the nucleus. AP-1 induces expression of antiviral cytokines (as well

as other genes), and thus forms part of the innate immune response (Ludwig et al. 2001). IFIT3 is another gene which regulates the immune response (Johnson et al. 2018) and was seen to switch to a proximal 3'UTR PAS in the nucleus while also increasing gene expression in the cytoplasm. IFIT3 has recently been reported to be a pro-viral regulator of IAV infection, with loss of IFIT3 resulting in reduced IAV replication (Tran et al. 2020). 3'UTR shortening of JUN and IFIT3 could contribute to their increased gene expression levels in infected cells and thereby ultimately modulate the host immune response (in the case of JUN) and promote IAV replication (in the case of IFIT3).

In this section, we've outlined physiologically relevant genes which switch PASs in their 3'UTRs, while also experiencing significant differences in their overall gene expression levels in the cytoplasm. We propose that 3'UTR APA of these genes could therefore impact their overall gene expression levels in the cytoplasm (and consequently their translational rate/stability) by modulating transcript properties. Based on these findings, 3'UTR APA of these genes could have a mix of viral and host benefits, and it is likely that both viral and host mechanisms are at play. A table summarising the genes, their APA status and cytoplasmic gene expression levels can be seen in Table 4.1.

Genes	**Nuclear 3'UTR APA**	**DEXSeq P-adjusted**	**Log2Fold change in cyto-plasm**	**DESeq P-adjusted**
SRSF1	Lengthens	0.0064	Not signifi-cant	0.71
SRSF7	Lengthens	0.0096	1.6	0.057
SRSF9	Lengthens	0.17	Not signifi-cant	0.69
SRSF3	Lengthens	0.000003	1.4	0.049
SRSF10	Lengthens	0.028	2	0.00036
SRSF6	Lengthens	0.000000009	Not signifi-cant	0.85
HNRNPK	Lengthens	0.057	1.5	0.013
PSMB1	Lengthens	0.046	1.8	0.036
VAPA	Lengthens	0.004	1.8	0.0077
TLR6	Lengthens	0.05	-3.1	0.0088
IL17RA	Lengthens	0.023	-3.0	0.00013
JUN	Shortens	0.02	2.0	0.0016
IFIT3	Shortens	0.001	3.4	0.0013

Table 4.1: Table exemplifying relationship between 3'UTR APA and differential transcript levels (WSN 18hpi)

4.2.6 Coding-region APA changes in WSN-infected cells

WSN induces widespread coding-region APA profile changes

As outlined in the previous chapter, coding-region APA changes are more easily linked with potential physiological changes than 3'UTR APA as the coding capacity of transcripts is altered, thereby resulting in proteins with changed amino acid sequence. Both coding-region APA subtypes, CR-CR APA and CR-UTR APA, involve alterations of the coding sequence, but only CR-UTR APA has the potential to yield transcripts with complete coding sequences (refer to section 3.2.6 in the previous chapter for a deeper explanation of coding-region APA). Coding-region APA is therefore a powerful gene expression regulatory mechanism which can increase or decrease the available pool of complete transcripts to be translated into full-length or truncated proteins. With this in mind, we investigated the coding-region APA events that occur in WSN-infected cells, and detailed some of their potential physiological implications for the host cell and virus.

To investigate CR-CR APA in WSN-infected cells, we extracted all significant APA events from the "CR" category of our pipeline, which determines APA events involving two PASs which are both in the coding-region of a gene (see Figure 2.2, chapter 2). For CR-UTR APA, we isolated those APA events from the "All" output of our pipeline (which calculates APA in genes between any two PASs anywhere in the gene body, see section 2.7.1 for a deeper explanation of the APA categories used in our pipeline) and isolated those where relative usage between a coding-region PAS and a 3'UTR PAS is significantly altered (Figure 2.2, chapter 2). We performed these analyses separately for the nuclear and cytoplasmic datasets.

By and large, the quantity of coding-region APA events seemed to mirror the

numbers of 3'UTR APA events at each timepoint throughout the infection: nuclear coding-region APA changes peaked at 9hpi (640 in total), whereas cytoplasmic events peaked at 18hpi. Interestingly however, shortening events outnumbered lengthening events at every single timepoint in both compartments. At 12 hours post infection, 508 (∼79 % of CR-UTR APA events) coding-region APA changes in the nucleus involved a relative shift in usage from a distal 3'UTR PAS to a proximal coding-region PAS (Figure 4.8). Similarly, in the cytoplasm, 175 (∼68 %) CR-UTR APA events were considered to be shortening (Figure 4.8). This is a departure from the trend seen in 3'UTR APA, where lengthening events accounted for a substantially higher share of 3'UTR APA in the nucleus and closer to an equal share in the cytoplasm (refer to Figure 4.1). Additionally, there appeared to be consistently higher numbers of CR-UTR events than CR-CR events, probably reflective of the fact that there are more 3'UTR PASs than coding-region PASs and consequently more opportunities for PAS switching involving 3'UTR PASs (Tian and Manley 2017). Lastly, we overlapped those genes which experienced CR-UTR cytoplasmic PAS switches between the WSN and Udorn infections (from the previous chapter), and found that over a third of the genes identified to be regulated by CR-UTR APA in the Udorn infection were common with those in the WSN infection (see Appendix Figure A.2). Acknowledging differences in experimental designs, this suggested that there was a strong shared impact on host CR-UTR APA profiles between different IAV strains.

To identify the potential physiological processes and pathways that coding-region APA may in part regulate in WSN-infected cells, we performed gene ontology analyses on genes experiencing CR-UTR APA in the cytoplasm, the site of translation (summarised in Figure 4.9, with more detail in Tables A.29 - A.32).

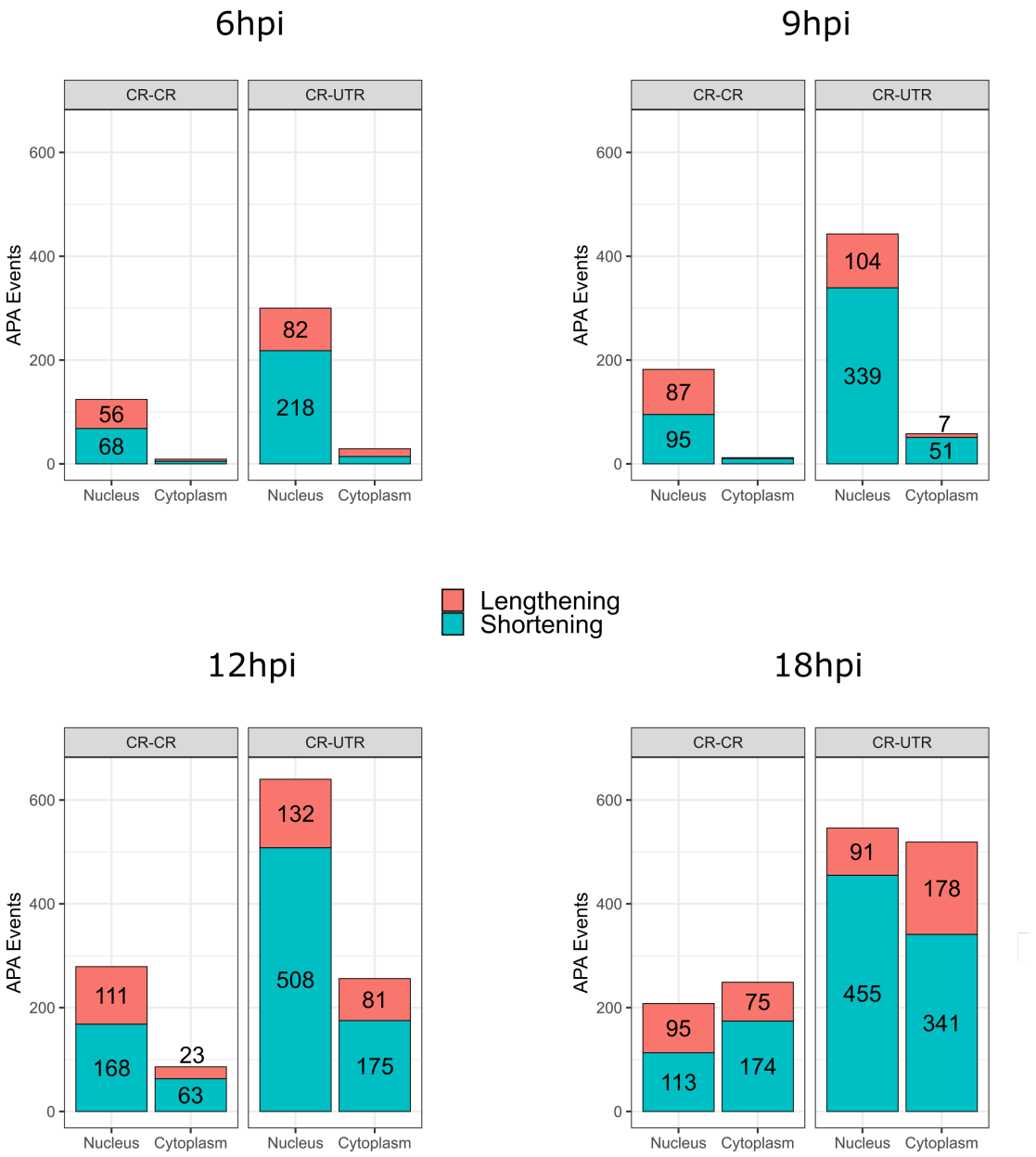


Figure 4.8: Coding-region APA events in WSN-infected cells. CR-CR refers to PAS switching between two poly(A) sites within the coding region. CR-UTR refers to alternative PAS usage between one coding region PAS and one 3'UTR PAS. The 3hpi timepoint has been excluded due to minimal events.

We focused on this group of genes in particular as CR-UTR APA represents a clear switch from full-length to truncated protein depending on the APA direction: shortening represents a relative shift in the pool of gene transcripts towards proximal isoforms with truncated coding capacity whereas lengthening indicates a switch from truncated transcripts to isoforms with full coding capacity. Genes that switched from a proximal coding region PAS to a distal 3'UTR PAS ("cytoplasmic lengthening") were enriched for the regulation of mRNA splicing, flavonoid glucuronidation and RNA-dependent DNA biosynthesis (Table A.29). Enriched pathways for cytoplasmic lengthening genes included gene expression, mRNA processing, TNF-alpha signaling and mRNA splicing (Table A.30).

One of the genes highlighted in the "regulation of mRNA splicing" category for cytoplasmic lengthening genes was RBM39, a transcriptional co-activator of an important protein complex involved in the transcriptional response to many conditions including viral antigens and cytokines, including NF-κB (Song et al. 2021). Interestingly, RBM39 has recently been reported to bind viral RNA during porcine reproductive and respiratory syndrome virus (PRRSV), thereby enhancing PRRSV replication (Song et al. 2021). Therefore, lengthening of the RBM39 transcripts from coding-region PASs to 3'UTR PASs, thereby translating relatively more full-length protein could be ultimately beneficial to IAV replication in a similar manner. Additionally, RBM39 has been reported to interact with the IAV PA-X protein, although any effect it might have on PA-X activity was difficult to assess due to the effect of RBM39 silencing on cell viability (Gaucherand et al. 2019). HNRNPL was another protein in this category (regulation of mRNA splicing), and is also known to regulate APA by binding 3'UTRs (Huang, Li, et al. 2012).

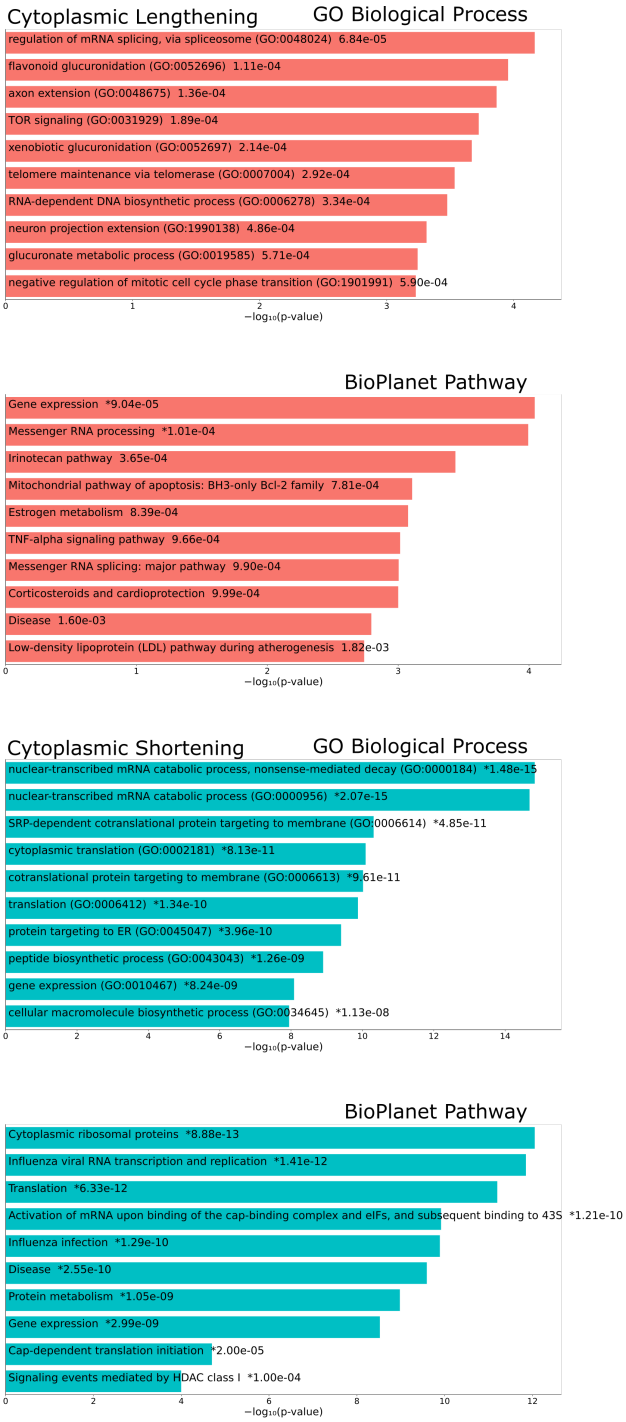


Figure 4.9: Gene ontology analyses of genes that switch PASs between coding region and the 3'UTR in WSN-infected cells ("biological process" and "pathways" category). Genes regulated by CR-UTR APA in the cytoplasm at any infection timepoint are shown, with the enriched biological processes and pathways shown in separate bar charts. -log10(pvalues) adjusted for multiple testing by the Benjamini-Hochberg method represented on the x-axis.

Two genes involved in host cleavage and polyadenylation, WDR33 and CSTF77, also experienced lengthening in the cytoplasm from proximal coding-region PASs to distal 3'UTR PASs. WDR33 binds the poly(A) signal sequence, along with CPSF30, while CSTF77 is a member of the CSTF complex and is believed to be responsible for the transport of the complex into the nucleus (reviewed in section 1.2.3). In addition, modulation of protein levels of these two key cleavage and polyadenylation factors could regulate the global APA profile, an effect seen with other core CPA factors (Curinha et al. 2014).

In the cytoplasmic shortening category, highly enriched biological processes included mRNA catabolism, SRP-dependent cotranslational protein targeting, translation, protein targeting to ER and gene expression (Table A.31). These categories are critically important functions for protein translation and protein processing, so the shortening of genes involved in these processes via coding-region APA may be a way of regulating gene expression at this level. However, it is difficult to disentangle whether or not the overall effect of CR-UTR shortening would be beneficial or detrimental to viral replication, as at least some of the coding-region APA changes could be induced as a host immune response rather than viral-induced, as part of the host shutoff effect. Future work exploring this question would be intriguing, for example by using RNAi techniques to knock down expression of short APA isoforms and subsequently testing viral replication. Regardless, enriched biological pathways included several protein translation-related groups like cytoplasmic ribosomal proteins, translation, eIFs, protein metabolism and cap-dependent translation initiation - again suggesting widespread modulations to protein translation induced by coding-region APA (Table A.32). Furthermore, pathways involved in influenza viral RNA transcription/replication and influenza

infection were also enriched. These are listed in Table 4.2, and feature many ribosomal proteins or subunits. Influenza viruses depend on cellular pathways for translation of their own proteins, but it is not clear whether coding-region APA of cellular ribosomal genes would be beneficial or detrimental to IAV, or whether it is induced by a specific virus-host interaction or as part of a more general host immune response. Future investigation could further elucidate the regulatory role of CR-UTR APA on the protein output of these genes and what impact it may have on the IAV replication cycle.

Enriched Pathways	**Genes**
Influenza viral RNA transcription & replication	RPS7, RPLP1, RPL12, RPL35A, RPSA, RPL8, RPS4Y1, SMG6, RPS4X, PPP2CA, RPS15A, PPP2R1A, RPS20, RPL37, RPS11, RPL29, POLR2I, RPS21, RPS13, RPS12
Influenza infection	SEH1L, NUP210, RPS7, RPLP1, RPL12, RPL35A, RPSA, RPL8, RPS4Y1, RPS4X, RPS15A, RPS20, RPL37, RPS11, RPL29, POLR2I, RPS21, RPS13, RPS12

Table 4.2: Table listing genes included in the influenza-related enriched pathways for coding-region APA in WSN-infected cells (cytoplasmic shortening).

Host factors important for IAV infection are regulated by coding-region APA

In the previous section we discussed an overview of the CR-UTR APA changes detected in WSN-infected cells and highlighted those that may play a role in influenza-related biological pathways. In this section, we focus on certain genes regulated by CR-UTR APA in the cytoplasms of WSN-infected cells which are likely to have implications for the host immune response based on published findings.

One of the genes significantly regulated by CR-UTR APA is ZC3HAV1, a gene that positively upregulates the interferon response and has been shown to restrict IAV replication (Zhang, Goraya, et al. 2020). ZC3HAV1 encodes the ZAP protein (Zinc Anti-viral Protein) and has been reported to increase in gene expression upon IAV infection (Zhang, Goraya, et al. 2020). Indeed, our gene expression data corroborates this finding (3.5-fold at 12hpi in the cytoplasm and 4.5-fold at 18hpi in the cytoplasm in WSN-infected cells). Importantly however, the ZAP protein has a long and short isoform (Kerns, Emerman, and Malik 2008). A recent study demonstrated that the short isoform was generated by APA, mediated by CSTF64 and stimulated by type I interferon signalling (Schwerk et al. 2019). Of note for our study, the long isoform of the ZAP protein (ZAPL) has also been found to associate with NS1-CPSF30 complexes (Liu, Zhou, et al. 2015; Chen, Liu, et al. 2014). We observed that APA regulation of ZC3HAV1 led to an overall decrease in the short isoform of ZAP (ZAPS) relative to the long isoform (ZAPL) throughout infection in the cytoplasm (6hpi, 9hpi, 12hpi and 18hpi). Figure 4.10 (top left panel) shows the usage of each annotated ZC3HAV1 PAS at 12hpi and 18hpi in

the cytoplasm. A clear drop in usage of the coding-region PAS that leads to the ZAPS transcript was seen in infected cells relative to the uninfected mock cells (red line). Corresponding to this drop is a comparative rise in selection of PASs in the UTR of the ZC3HAV1 gene, the usage of which generates ZAPL transcripts. The differential usage of these PASs leads to the inclusion/exclusion of approximately 14,000 bases in the nascent transcript, altering its coding capacity and determining whether the ZAPL or ZAPS protein is translated. Interestingly, both forms of ZAP have been found to inhibit IAV replication via different mechanisms. ZAPL promotes proteasomal degradation of two IAV proteins, PB2 and PA, in a manner dependent on its C-terminal PARP domain (Liu, Zhou, et al. 2015). Despite lacking this domain, ZAPS was also found to inhibit expression of the IAV proteins PA, PB1 and NA, suggesting that balancing expression of these protein isoforms is important for the regulation of IAV replication (Tang, Wang, and Gao 2017). This balancing act probably explains the CR-UTR APA regulation of ZC3HAV1 that we observed throughout the WSN infection at almost every timepoint, the deltaRA values of which can be viewed in Table 4.3.

Intriguingly, another gene that was found to be regulated by CR-UTR in WSN-infected cells - DHX30 - has previously been reported to enhance the antiviral activity of ZAP (Ye, Liu, et al. 2010). Furthermore, DHX30 has been demonstrated to associate with the NS1-CPSF30 complex by mass spectrometry (Chen, Liu, et al. 2014; Liu, Zhou, et al. 2015), suggesting that it could play a role in altering APA during IAV infection via this interaction. Moreover, DHX30 has recently been found to inhibit IAV replication via an interaction with NS1 - IAV replication was enhanced 5-8 fold when DHX30 was knocked down in cells (Chen, Ma, et al. 2020). In contrast to ZC3HAV1 which lengthened to distal 3'UTR PASs relative to

a proximal coding-region PAS, DHX30 shortens in our data during WSN infection and we observed increased usage of a coding-region proximal PAS approximately 33,000 bases upstream of the 3'-most exon, excluding 20 exons in total (Figure 4.10, top right). This likely results in fewer full-length DHX30 protein isoforms being generated, a situation which may enhance viral replication based on findings from the literature, cited above. This could therefore make DHX30 an interesting example of a gene that is regulated by NS1-induced CR-UTR APA to the benefit of IAV replication, although further work would be needed to confirm NS1's role. DHX30's deltaRA values throughout the WSN infection timecourse can be seen in Table 4.3.

Gene	**6hpi deltaRA**	**9hpi deltaRA**	**12hpi deltaRA**	**18hpi deltaRA**
ZC3HAV1	-0.43	-0.75	-0.79	-0.76
DHX30	NS	NS	0.17	0.24
AIMP2	0.04	0.20	0.44	0.84
IFNAR2	0.28	0.51	NS	0.68

Table 4.3: Table showing deltaRA values for ZC3HAV1, DHX30, AIMP2 and IFNAR2. DeltaRA measures the relative abundance change between the short and long APA isoforms. 0 to 1 represents a shift in abundance towards the short isoform, with 0 to -1 representing lengthening. -1 and 1 represent complete shifts in the transcript pool to the long or short APA isoforms respectively. Refer to section 2.7.1 for further explanation of deltaRA.

AIMP2 is another host factor regulated by CR-UTR APA during WSN infection that is known to impact IAV's life cycle, and has been reported to interact with the IAV NEP protein, a protein involved in the nuclear export of viral ribonucleoproteins (vRNPs) (Dou et al. 2018). It was found that the presence of NEP protects AIMP2 from ubiquitin-mediated degradation, and that AIMP2 enhances

IAV replication (Gao, Wu, et al. 2015). The mechanism behind this positive regulation of viral replication is believed to be an AIMP2-mediated switch of the IAV M1 protein from a ubiqutinated to a SUMOylated state, thereby enhancing M1-led nuclear export of vRNPs. In our WSN infection data, we observed CR-UTR shortening of AIMP2 at multiple timepoints throughout the timecourse (6hpi, 9hpi, 12hpi and 18hpi). Usage of the proximal coding region PAS (upregulated in WSN infection) leads to over 12,000 bases and four exons being excluded from the transcript (Figure 4.10 bottom left) - suggesting that CR-UTR APA during WSN infection leads to the protein pool of AIMP2 shifting towards substantially truncated isoforms which have reduced or altered function. Based on the reported positive effect of AIMP2 on IAV replication (Gao, Wu, et al. 2015), this CR-UTR APA switch could be part of the wider host stress response and potentially detrimental to IAV replication.

The IFNAR2 gene, responsible for encoding one of the two subunits of the type I interferon receptor (IFNAR), was also found to be regulated by CR-UTR APA. IAV induces a strong type I interferon response (Killip, Fodor, and Randall 2015), so any potential regulation of the receptor responsible for recognising IAV pathogen-associated molecular patterns via APA could be highly physiologically relevant. Importantly, the IFN response is enhanced by a positive feedback loop: IFN produced in response to virus infection binds the interferon receptor which ultimately leads to induction of genes including IFN. This can thus augment autocrine effects in the host cell, or have paracrine effects on nearby cells (Jia, Rahbar, et al. 2010). During WSN infection, we observed IFNAR2 shortening at multiple points throughout the infection (6hpi, 9hpi and 18hpi) (see Table 4.3). This manifested in a switch from a 3'UTR distal PAS to a proximal coding-region

PAS with the exclusion of an exon (approximately 1000 bases) (see Figure 4.10, bottom right). The IFNAR2 receptor subunit has been shown to be required for effective anti-IAV immune responses (Shepardson et al. 2018), suggesting that any modulation of its activity could be beneficial to IAV replication.

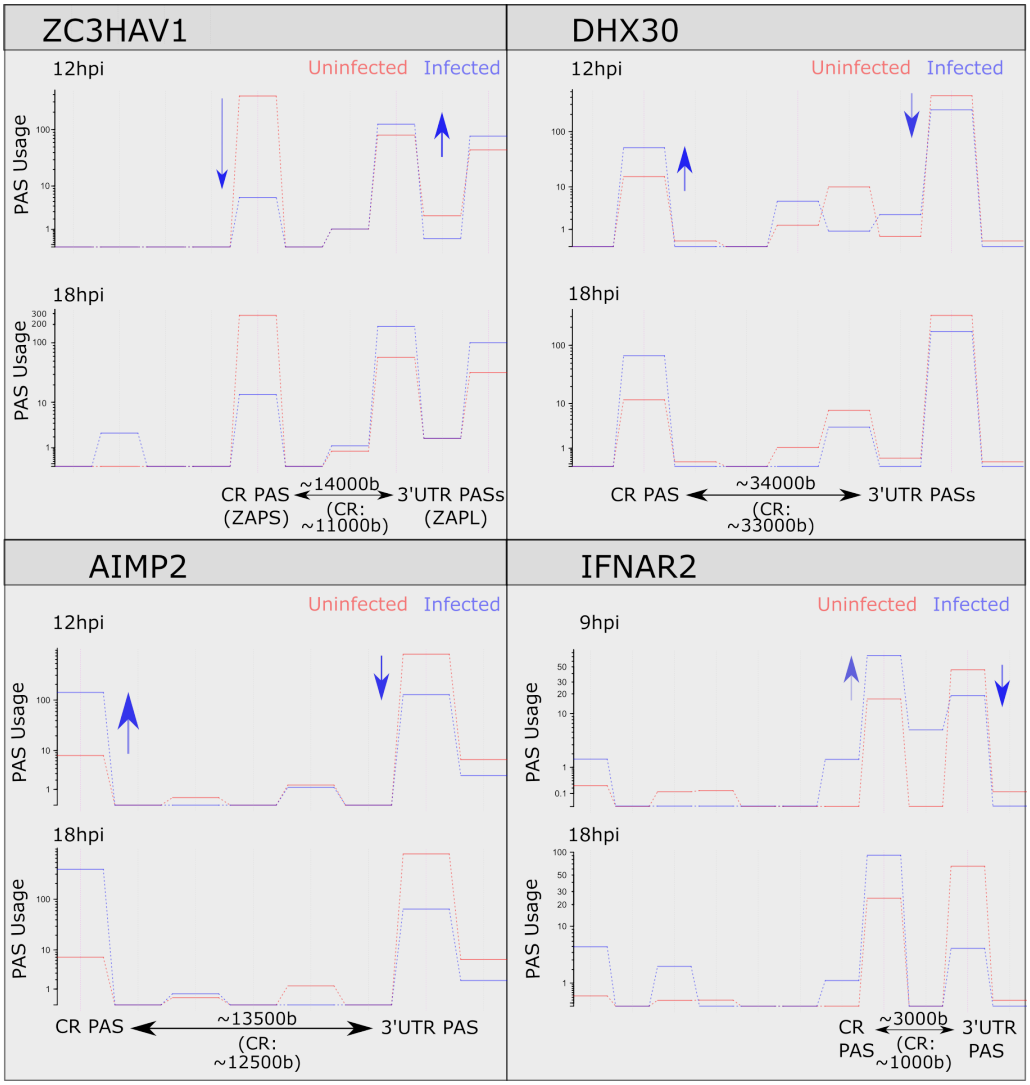


Figure 4.10: Plots showing differential CR-UTR PAS usage in WSN-infected cells for ZC3HAV1, DHX30, AIMP2 and IFNAR2. ZC3HAV1 is relatively lengthened by CR-UTR APA while DHX30, AIMP2 and IFNAR2 are shortened. Purple and red lines represent PAS usage in infected and uninfected cells respectively, and the differences in distance between switching PASs are depicted in bases (b).

In this section we have highlighted four host genes, the regulation of which by CR-UTR APA could have major physiological implications for infected cells. Namely, ZC3HAV1 lengthens, ultimately producing an APA isoform which encodes ZAPL an interactor of the NS1-CPSF30 complex. Modulating levels of ZAPS and ZAPL (both inhibiting IAV infection by different mechanisms) could have implications for viral replication, and form part of the host stress response. DHX30 shortens by increasing relative usage of a coding-region PAS, which could enhance viral replication based on published findings which showed an increase in IAV replication when DHX30 was knocked down. AIMP2 is believed to enhance viral replication through an interaction with IAV NEP and shortens via a coding-region PAS in WSN-infected cells which could have inhibitory implications for IAV replication. IFNAR2 - a critical subunit of the type I interferon receptor - was also seen to shorten in our data, suggesting a virus-induced dampening effect on the host immune response regulated by CR-UTR APA. We therefore propose a model by which APA is a target for IAV to manipulate in such a way that contributes to host shutoff and the suppression of the immune response, thereby changing a hostile replicative environment to a more beneifical one. Future work investigating the impacts on IAV infection of the CR-UTR APA isoforms of these genes could therefore lead to important pharmacological insights.

4.3 Discussion

In this chapter, we aimed to build on the findings from our Udorn study by investigating changes induced throughout an infection timecourse using a different IAV strain and the same powerful subcellular RNA fractionation approach. We

utilised a purpose-built bioinformatic pipeline (the design of which is discussed in chapter 3, section 2.7.1) to robustly characterise changes in host differential gene expression and APA at distinct infection timepoints, and to investigate mechanistic causes behind PAS switching by analysing cis-element frequency patterns in those PASs that are differentially regulated. Overall, our WSN findings fully corroborated our findings in Udorn, suggesting that flu-induced APA changes are not strain-specific.

We identified widespread changes in the 3'UTR APA landscape of WSN-infected cells. As expected due to the multi-stage life cycle of IAV (Dou et al. 2018), the number of changes was highly dynamic throughout the timecourse (Figure 4.1). 3'UTR APA events steadily increased from 6hpi onwards, peaking at over 1,000 significant events at 12 hours post infection in the nuclear samples with a bias towards lengthening. In the cytoplasm, 3'UTR APA events were slower to manifest but rapidly increased, with 76 events at 9hpi and 720 by 18hpi. A benefit of our subcellular RNA fractionation approach meant that we were able to further explore whether APA isoforms were unique to each compartment - an important point to consider since the cytoplasm is the site of translation. To investigate this, we visualised all 3'UTR APA isoforms in both fractions over the timecourse in a heatmap and Venn diagrams (Figures 4.2 and 4.3 respectively). We found that while the majority of APA events found at one timepoint were seen in at least one other timepoint in the same compartment, relatively few events persisted between compartments. This suggested that while 3'UTR APA was highly induced at the point of transcription in the nucleus, post-transcriptional regulation of APA isoforms was also widespread in WSN-infected cells, ultimately resulting in biased distribution of isoforms in both compartments. This phenomenon is well-reported

in the literature and can result from aUTRs containing cis-elements which affect transcript stability, decay rate or nuclear export (Neve, Patel, et al. 2017). Additionally, it is thought that the PA-X IAV protein selectively degrades host mRNA with length being a determinant (Gaucherand et al. 2019), which could also play a role in the post-transcriptional fate of APA isoforms in IAV-infected cells once they reach the cytoplasm.

What physiological processes might be impacted by infection-induced host 3'UTR APA? Gene ontology analyses revealed (Figure 4.4) that mRNA processing-related categories including splicing were highly enriched in those genes which switched to distal APA isoforms in the nucleus. IAV is known to hijack host splicing machinery for the splicing of its own segments, suggesting a mechanism by which 3'UTR APA in WSN-infected cells regulates the expression of core splicing factors, which themselves are important for the IAV life cycle. However, it is not clear from the findings in this study whether or not these 3'UTR APA changes are induced by viral or host mechanisms, or if they have a cumulative beneficial or suppressive effect on viral replication. Further work addressing these questions could yield intriguing results. Additionally, we found that two core host cleavage and polyadenylation factors (CPSF6 and WDR33) lengthened in the cytoplasm via 3'UTR APA. The next step for this element of the study would be to examine whether the translation rates of these two important host factors are affected adversely due to this 3'UTR lengthening. These findings could then reveal if 3'UTR APA of these two genes contributes to regulation of the wider host APA profile in IAV-infected cells.

We next examined what binding motifs could be present in aUTR motifs that had the potential to alter transcript properties, such as stability or nuclear export

efficiency. We focused on those genes that lengthened in the nucleus but not in the cytoplasm at 12hpi, which totalled 694 distal isoforms. We identified over 20 significant motifs that were enriched in this group than those least likely to PAS switch in WSN-infected cells (the top 5 are shown in Figure 4.5). In particular, the binding motif for HuR/ELAVL1 was highly enriched in the aUTRs of genes lengthened by 3'UTR APA in the nucleus, suggesting that ELAVL1 contributes positively to the over-representation of lengthened APA isoforms in the nucleus. Other RBP binding motifs that were enriched included those for SAMD4A, SRSF10, PCBP2 and HNRNPK. These proteins have been reported to influence transcript stability and decay rate, thereby opening up the possibility that these RBPs regulate gene expression via 3'UTR APA. As binding motifs for these RBPs were highly over-represented in those genes which lengthened via 3'UTR APA in the nucleus but not in the cytoplasm, it seems likely that these RBPs could be involved in shaping host APA profiles during virus infection. This concept has already been proven for HNRNPK, which was shown to induce lengthening of a long noncoding RNA by arresting the binding of the CFI complex, mediated by an RNA HNRNPK-binding sequence near the PAS (Naganuma et al. 2012). Additional work investigating the effects on APA isoform abundance in the presence or absence of these proteins would clarify the impact of aUTR motifs in WSN-infected cells.

Our analyses of gene expression patterns of virus-infected cells confirmed that viral-induced APA changes to host profiles were functionally distinct to the wider transcriptional response induced by influenza infection. We identified thousands of differentially expressed genes during the infection (Figure 4.6), with these changes peaking towards the mid- and late-infection phases (9hpi onwards). The changes we saw largely matched what has been reported in the literature (Ashraf et al. 2020;

Thompson, Dittmar, et al. 2020), and we contributed an additional dimension to these reports by analysing nuclear and cytoplasm gene expression separately and over a timecourse (Appendix A.3). Based on this, we were assured that a robust immune response had been induced by our experimental approach. By utilising both our gene expression and 3'UTR APA analyses, we revealed interplay between the two and asked if PAS switching might influence the overall gene expression of specific genes. We identified those 3'UTR APA events which are most likely to cause changes in overall gene expression levels and thus be physiologically relevant to IAV infection. We found that several SR-family members and HNRNPK, proteins with established roles in binding mRNA and with links to splicing (Shepard and Hertel 2009; Thompson, Dittmar, et al. 2020) and alternative polyadenylation (Shen et al. 2019; Hu, Lutz, et al. 2005; Tian and Graber 2012), lengthened in the nucleus while also experiencing upregulated gene expression in the cytoplasm. Noting that we previously demonstrated that HNRNPK binding sites were enriched in those genes which were nuclear-retained, we suggest an intriguing role for HNRNPK in virus-induced APA: 3'UTR lengthening of HNRNPK contributes to an upregulation of cytoplasmic HNRNPK transcript levels, thus leading to increased HNRNPK proteins, which then subsequently contributes to a wider lengthening shift in host nuclear APA profiles by binding mRNA cis-sequences. Further work is needed to test this model, perhaps by investigating if HNRNPK knockdown suppresses the lengthening effect on APA profiles.

Additionally, interplay between 3'UTR APA and gene expression patterns were identified for host factors important for viral and IAV replication. PSMB1 and VAPA, were relatively lengthened in the nucleus while their expression levels were upregulated in the cytoplasm. We also found that two cytokine receptors, TLR6

and IL17RA, which have been shown to play roles in the host's response to IAV infection, lengthened in the nucleus but had downregulated gene expression levels in the cytoplasm - suggesting that 3'UTR APA of these two important immune genes could play a role in their overall gene expression regulation. Finally, we identified JUN and IFIT3, both regulators of the innate immune response, as genes which shorten their 3'UTRs in the nucleus while also having upregulated gene expression levels in the cytoplasm. These immunologically-relevant genes (Table 4.1) are examples of 3'UTR APA potentially impacting gene expression levels in the cytoplasm and consequently modulating IAV infection and/or the host stress response, although further work confirming this link would clarify the physiological impact of 3'UTR APA on these genes.

As discussed previously, 3'UTR APA has the potential to impact gene expression by altering the properties of transcripts such as stability or export (Zhang, Liu, et al. 2021; Sandberg et al. 2008; Ji, Lee, et al. 2009). However, APA that involves at least one coding-region poly(A) site (coding-region APA) can directly alter the resulting protein product. Coding-region APA has the clear potential to enact physiological change by altering protein identity, and subsequently protein function. Therefore, we next investigated to what extent there may be coding-region APA changes induced during WSN-infection. We identified hundreds of CR-CR APA (PAS switching between two coding region PASs) and CR-UTR APA (PAS switching between one coding region PAS and one 3'UTR PAS) significant events, peaking at 12hpi and 18hpi (Figure 4.8). Consistently, more CR-UTR APA events than CR-CR APA events were observed at every timepoint. As CR-UTR APA events represent shifts to truncated proteins (shortening) or full proteins (lengthening), we performed gene ontology analyses on those genes that were subjected

to CR-UTR APA in the cytoplasm to investigate what physiological functions could be impacted. For those genes experiencing lengthening via CR-UTR APA (and thus restoring the full coding capacity to the transcript), enriched processes included regulation of mRNA splicing and DNA biosynthesis, while several biological pathways were enriched including gene expression, TNF-alpha signaling and mRNA processing/splicing (Figure 4.9). One example gene in this category, RBM39, has already been reported to enhance viral replication of another RNA virus, porcine reproductive and respiratory syndrome virus, so lengthening of this gene via CR-UTR APA could also be promote IAV replication. For those genes that switch to truncated APA isoforms by shortening, we saw an abundance of translation-related genes such as cytoplasmic ribosomal proteins/subunits (outlined in Table 4.2), suggesting that CR-UTR APA plays a major role in the regulation of host and viral gene expression via translation during IAV infection. We also documented CR-UTR APA changes in several key host factors which have known roles in viral or IAV infection. These are highlighted in Table 4.3, with their CR-UTR APA changes depicted in schematics in Figure 4.10. One of these, ZC3HAV1, lengthens and switches to a long isoform (which encodes the ZAPL protein) with a corresponding collapse in reads encoding the "short" ZAPS protein product. Interestingly, both protein isoforms have been reported to inhibit IAV replication, but via different mechanisms (Liu, Zhou, et al. 2015; Tang, Wang, and Gao 2017). Another gene, DHX30, which shortens and switches to a substantially truncated APA isoform, has been reported to bind the NS1-CPSF30 complex (Chen, Liu, et al. 2014; Liu, Zhou, et al. 2015) and also enhances the antiviral activity of ZAP (Ye, Liu, et al. 2010). One study found that DHX30 knockdown resulted in a 5-8 fold increase in IAV replication (Chen, Ma, et al. 2020), suggesting that

truncation of DHX30 mRNA could be beneficial to viral infection. AIMP2 and IFNAR2 also both shorten via proximal coding-region PASs, with AIMP2 believed to positively impact viral replication (Gao, Wu, et al. 2015) whereas IFNAR2 has an inhibitory effect on IAV infection (Shepardson et al. 2018). Modulation of the activity of these CR-UTR APA example genes could have major implications for IAV replication and/or for the host immune response to viral stress.

Interestingly, we also found that two core cleavage and polyadenylation factors, CSTF77 and WDR33 also experienced CR-UTR lengthening in the cytoplasm, indicating that the protein levels of these two regulators of APA are themselves regulated by APA during IAV infection. Modulation of CSTF77 protein levels has previously been shown to induce widespread APA changes (Luo, Ji, et al. 2013), suggesting a clear mechanism by which flu-induced coding-region APA of CSTF77 has the potential to regulate host APA. The link between WDR33 protein levels and mechanistic regulation of APA is less clear, with one study reporting only a mild effect upon WDR33 knockdown (Li, You, et al. 2015). How WDR33's activity is impacted by the NS1-CPSF30 association is unknown, although an NS1-WDR33 interaction has been reported (Wang, Fu, et al. 2017). One can speculate that a shift in usage from a coding region PAS to a distal 3'UTR PAS, thereby increasing levels of WDR33 protein, could represent a mechanism used by the host to circumvent WDR33 deregulation caused by interaction with NS1.

Summary

In this chapter, we used a purpose-built bioinformatic pipeline to characterise host gene expression changes, focusing on the host APA landscape, in influenza A/WSN/33-infected cells through a timecourse. We corroborated our findings

from the Udorn analysis, revealing that flu-induced APA changes are largely strain-independent. We identified widespread 3'UTR and coding-region APA changes in both nuclear and cytoplasmic samples. In particular, we noted how the aUTRs of genes regulated by 3'UTR APA are highly enriched in binding sites for several RNA-binding proteins which have the capacity to regulate gene expression by altering transcript properties. We also highlighted several physiologically relevant genes in the context of IAV infection which experience coding-region changes in their transcripts due to APA. Overall, we documented widespread APA changes in host cells infected by influenza A/WSN/33 and explored their potential physiological ramifications.

Chapter 5

Investigating the characteristics of differentially regulated poly(A) sites in influenza-infected cells

5.1 Introduction

In chapters 3 and 4 of this thesis, changes in the host APA landscape are explored during infection with the H3N2 A/Udorn/72 and H1N1 A/WSN/33 strains of influenza respectively. But what decides whether or not a poly(A) site is used or not used during virus infection? And is there any pattern to this process which may partly explain why some PASs are up- or down-regulated?

The combinatorial strength of a poly(A) site is well-understood to be defined by cis-elements, including the core poly(A) signal and surrounding auxiliary elements (Zheng and Tian 2014; Tian and Manley 2017). Additionally, where a PAS is located within the gene body affects its strength - PASs which are furthest from

the coding region are generally considered to be stronger than those closest to the gene body (Legendre and Gautheret 2003; Tang, Yang, et al. 2022). Are these poly(A) characteristics determining parameters for the regulation of cleavage and polyadenylation in the context of influenza infection?

5.1.1 Aims

In this chapter, we use bioinformatic methods in combination with the sequencing data discussed in chapters 3 and 4, to investigate characteristics of poly(A) sites which are differentially regulated in virus infection. We uncover highly significant frequency biases for sequences in and around poly(A) sites which are differentially used during influenza infection. These insights add to our understanding of the complex host-virus relationship, and of alternative cleavage and polyadenylation, and may prove useful in the identification of a target for antiviral drug development.

5.2 Results

5.2.1 A-rich poly(A) sites are enriched in upregulated PASs during flu infection

To investigate the relationship between the core poly(A) signal and APA events, we extracted all significant genes (as determined by DEXSeq), and separated the APA events therein into lengthening and shortening categories. The benefit of our subcellular RNA fractionation approach (discussed in the previous two chapters) meant that we could focus only on the nuclear fraction, which was important

since we were most interested in the mechanisms behind PAS site choice during transcription and we could thus minimise post-transcriptional effects. We then calculated the usage frequencies of each PAS signal for every APA type (All, UTR and CR). We also performed the same analysis on the least significant APA events, for comparison purposes, with equivalent numbers of genes. This method was performed on each dataset from the Udorn and WSN infections separately. With this approach, we observed relative increases in A-rich PASs (AAAAAA), particularly in the upregulated proximal PASs for shortening genes.

For Udorn, these increases were: 11.9% to 31.3% for "All" and 7.2% to 15.3% for "UTR", (Figure 5.1, top) when compared with genes not experiencing significant differences in PAS usage (Figure 5.1, bottom). We also observed increases in A-rich prevalence for upregulated distal PASs in lengthening genes: ∼3 % in "All" to 11.6%, and ∼2 % in "UTR" to 11.2%.

We also found an overall increase in the usage of A-rich poly(A) sites (AAAAAA) in PASs regulated by APA during WSN infection, in all types of APA. This increase was most dramatic in the "All" category of APA, where 32.5% of upregulated proximal PASs were A-rich compared (Figure 5.2, top left) compared to 13.5% in proximal PASs not differentially regulated (Figure 5.2, bottom left). Increases in A-rich PAS signals were also seen in those PASs regulated only by 3'UTR-APA, although less pronounced: 5.6% to 13.5% (Figure 5.2 middle). Upregulated distal PASs were also enriched for A-rich poly(A) signals in those genes regulated by APA from less than ∼3 % in "All" to 8.1% (Figure 5.2, left) and ∼1 % to 7.1% in 3'UTR APA (Figure 5.2, middle). A-rich PAS signals were also used more frequently in those PASs regulated by CR-CR APA (16.7% to 24% for distal lengthening, 18.3% to 29.2% for proximal shortening) as seen in Figure 5.2 (right). While only the

data from eighteen hours post infection are depicted in Figure 5.2, these trends were also seen at other times during the infection, and figures representing the PAS signal data at these timepoints can be seen in the Appendix (section A.15). The A-rich usage fluctuations seen in our analysis were accompanied by changes in usage of the canonical hexamer, particularly that of AAUAAA (28.1% to 21.6% in shortening events, 54.8% to 46.9% in lengthening events). Interestingly, the NS1 protein of the Udorn IAV strain is known to have a strong interaction with CPSF30 (Das et al. 2008), the core cleavage and polyadenylation factor which plays a major role in recognising the AAUAAA signal, suggesting a possible mechanism by which PASs featuring the canonical hexamer are less favourably selected during Udorn infection.

Based on this data, there is a clear bias for A-rich PASs in genes regulated by APA during influenza infection which is not strain-specific. Importantly, A-rich PASs have been reported to be sufficient to direct cleavage and polyadenylation if there is a strong downstream element present (Nunes et al. 2010; Zheng, Liu, and Tian 2016; Wang, Zheng, Yehia, et al. 2018). Additionally, because our pipeline uses the PolyA_DBv3 database (Wang, Nambiar, et al. 2018) we can be confident that the PASs we detect are not experimental artifacts resulting from internal priming.

5.2.2 Distal-most PASs are less frequently used in genes regulated by APA during flu infection

Next, we wanted to investigate what role PAS location within the gene body might play in flu-induced APA. Therefore, we performed an additional analysis focused

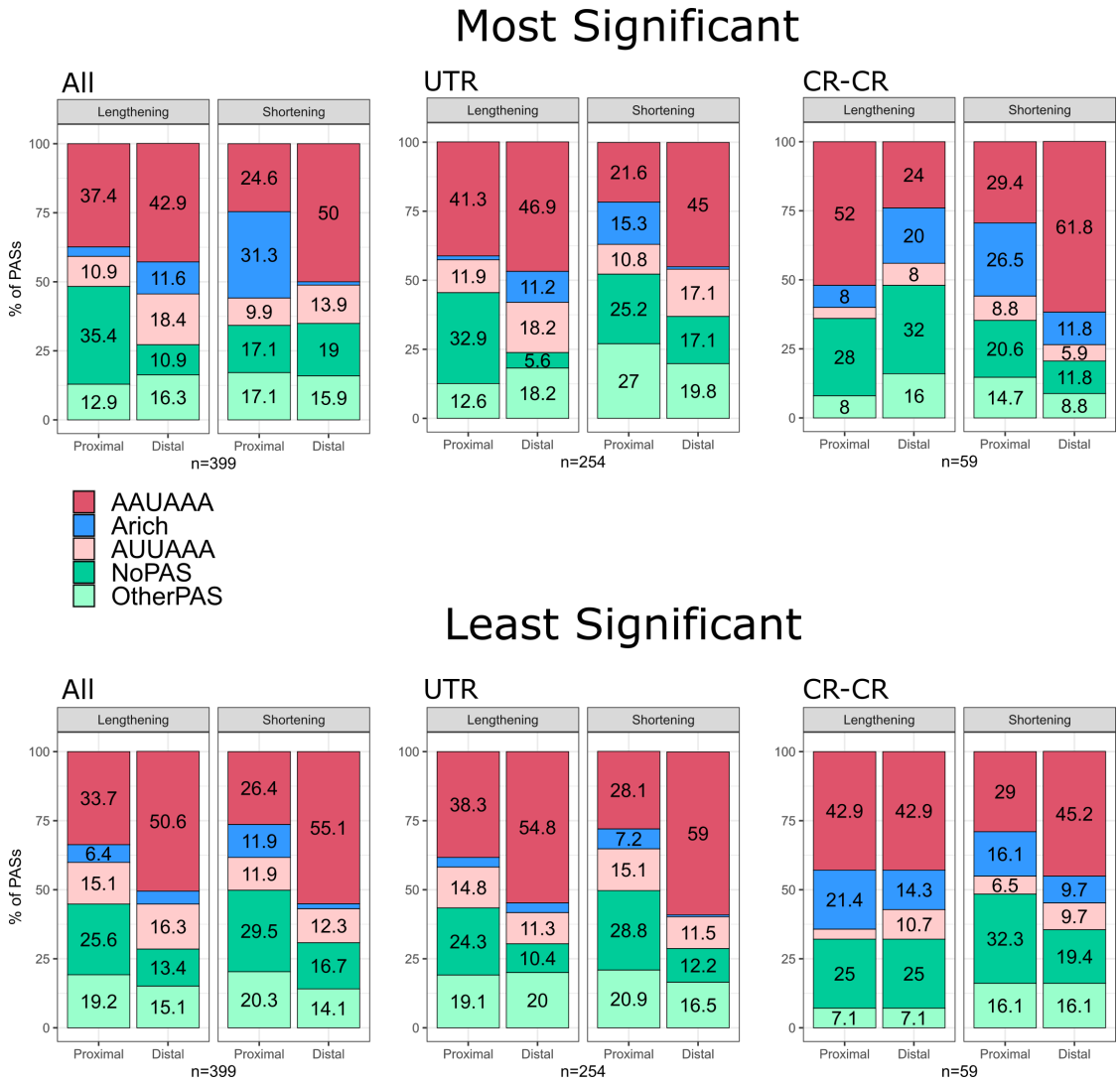


Figure 5.1: Relative amounts in the nucleus of each signal sequence in PASs regulated by APA (Top) and those not (Bottom), Udorn infection. The distribution of each annotated PAS signal can be seen in those PASs differentially regulated during Udorn infection (Top), or those not undergoing APA (Bottom). "All" refers to APA changes involving PASs found anywhere within the pre-mRNA. "UTR" includes APA changes involving PASs found only within the 3'UTR, and "CR-CR" APA comprises those PASs found only within the coding-region. The total number of genes used in each APA category is shown under each chart.

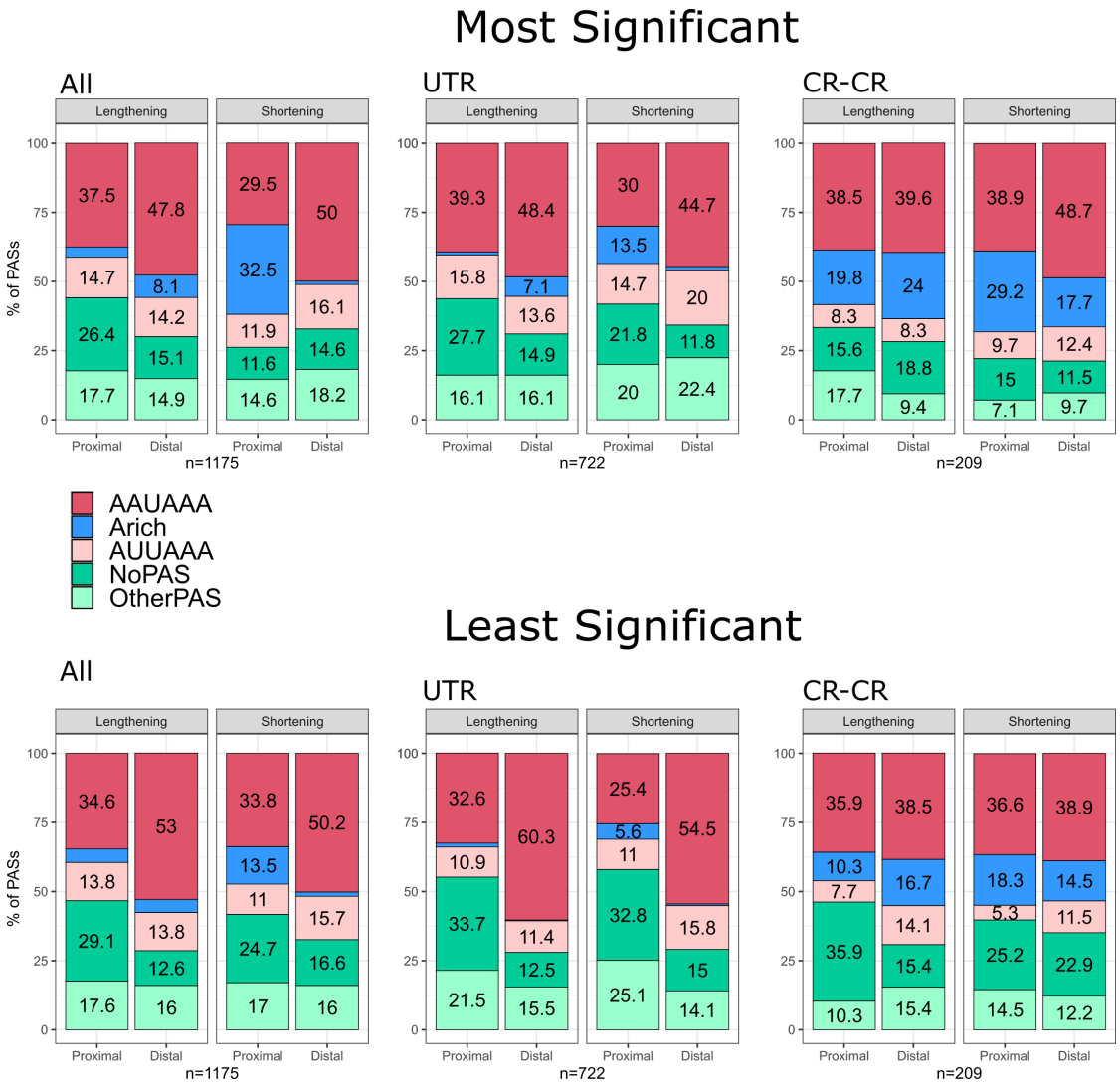


Figure 5.2: Relative amounts in the nucleus of each signal sequence in PASs regulated by APA (Top) and those not (Bottom), 18hpi (WSN infection). The distribution of each annotated PAS signal can be seen in those PASs differentially regulated during WSN infection (Top), or those not undergoing APA (Bottom). "All" refers to APA changes involving PASs found anywhere within the pre-mRNA. "UTR" includes APA changes involving PASs found only within the 3'UTR, and "CR-CR" APA comprises those PASs found only within the coding-region. The total number of genes used in each APA category is shown under each chart.

on "PAS types". "PAS type" here refers to the PAS location within the pre-mRNA (5'UTR, 3'UTR, CDS, Intron or Pseudogene) and are those annotated in the PolyA_DBv3 database (Wang, Nambiar, et al. 2018), including 5'UTR, CDS, intergenic and 3'UTR, where 3'UTR PASs are further split into first (F), middle (M), last (L), or single (S). To compare usage of PAS types, we collated the PAS types of both the proximal and distal PASs for lengthening and shortening genes, in each APA type, for those genes regulated by APA during influenza infection (with data from each strain analysed separately).

In Udorn (Figure 5.3), for 3'UTR APA, the distal-most PAS (the last annotated PAS) in each gene accounts for 37% of those PASs upregulated in lengthening genes (distal), and 42% of those downregulated in shortening genes (distal). This is a notable decrease when compared to other genes less likely to be regulated by APA in Udorn infection, with the last 3'UTR PAS site accounting for over half of distal PASs (62% in lengthening and 52.9% in shortening). Conversely, there is a higher usage of middle-UTR PASs in genes regulated by APA during Udorn infection. This general trend, while less pronounced, is also seen in the "All" APA category (which includes CR-CR, CR-UTR APA and UTR-APA).

In WSN, we again found a decrease in usage of distal-most PASs by UTR APA (3'UTR L) in upregulated distal PASs (60.7% to 46.2%), which can be seen in Figure 5.4 (middle). The same effect was also observed in the "All" category with usage of the last 3'UTR PAS decreasing from 50.8% to 41.5% (Figure 5.4, left). While Figure 5.4 displays only the findings from the 18hpi timepoint, the same trend was also observed at other timepoints during WSN infection which can be seen in the Appendix (section A.16).

Interestingly, distal-most PASs are stronger than proximal PASs (Legendre

and Gautheret 2003; Tang, Yang, et al. 2022), a phenomenon believed to help distal PASs compete with proximal PASs, which are transcribed first. Therefore, the stark increased usage of middle 3'UTR PASs relative to distal-most PASs for lengthening 3'UTR APA in both strains could suggest weaker recognition of the stronger distal-most PASs, or stronger recognition of PASs closer to the gene body, possibly due to a fluctuation in availability of cleavage and polyadenylation factors (which have varying affinities for distinct cis-element sequences (Zheng and Tian 2014)). For example, CPSF5 is known to bind the UGUA cis-element, which is more distributed around distal PASs than proximal PASs (Zhu, Wang, et al. 2018). Furthermore, the PA-X influenza protein is enriched in the nucleus and believed to interact with CPSF5 (Gaucherand et al. 2019), suggesting one possible mechanism by which usage of distal-most PASs is dysregulated in flu-infected cells. Additionally, the decreased usage of AAUAAA PASs in the Udorn dataset (seen in Figure 5.1) aligns with the observations of less frequently used distal-most sites, as 3'-most PASs mainly feature the canonical hexamer whereas PASs proximal to the stop codon more commonly use one-base variants (Beaudoing et al. 2000).

As these proximal PASs tend to have less conserved auxiliary sequences than their distal-most counterparts (Tang, Yang, et al. 2022; Shi 2012), the next section details analyses of cis-element sequences surrounding PASs regulated by APA during virus infection.

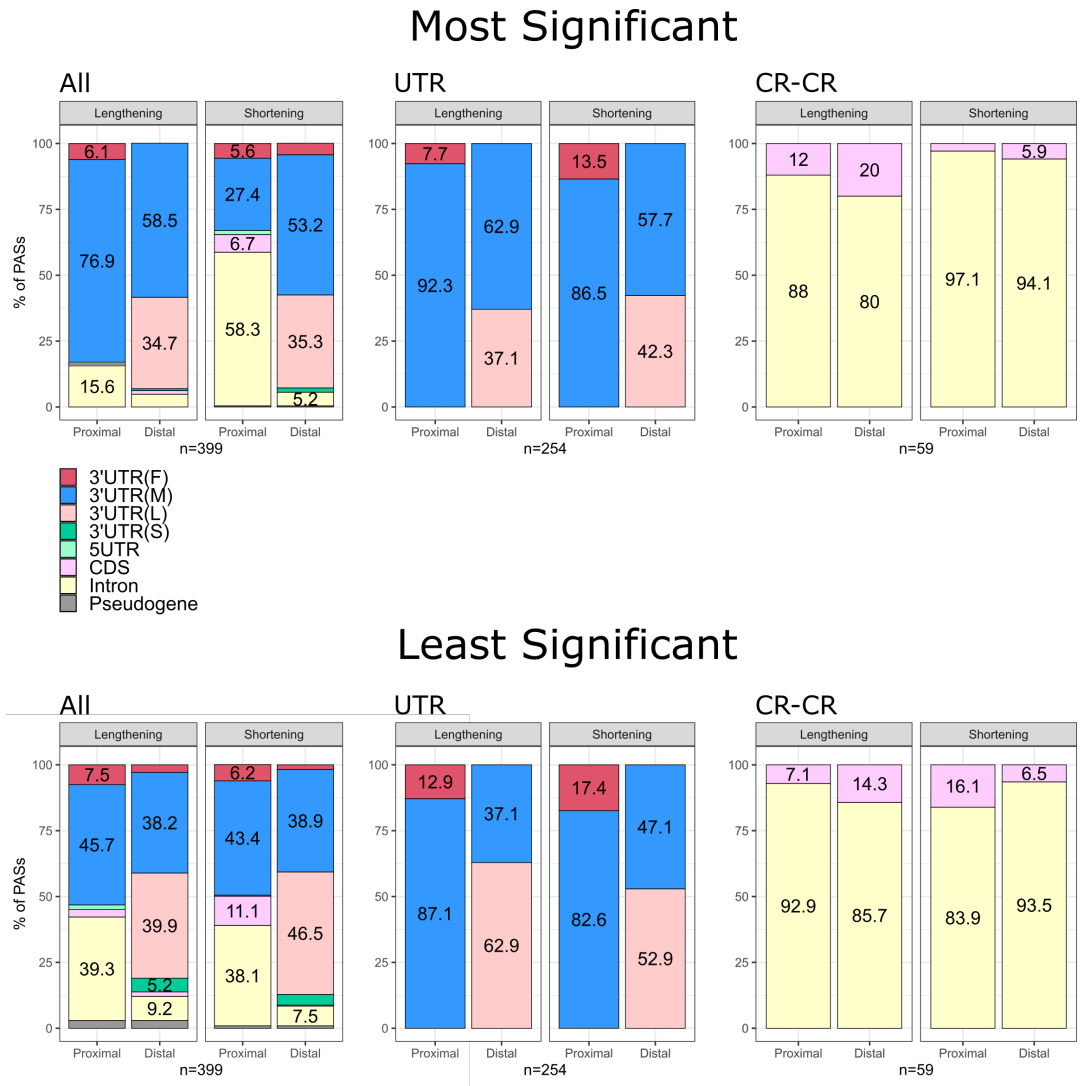


Figure 5.3: Location of PAS within the gene body of those regulated by APA (Top) and those not (Bottom) in the nucleus, Udorn infection. PASs involved in APA in Udorn-infected cells were classified depending on their location in the gene body. Those situated in the 3'UTR were further broken down into first (F), middle (M) and last (L) PAS. If there was just one annotated PAS in the 3'UTR, it is referred to as single (S). "All" refers to APA changes involving PASs found anywhere within the pre-mRNA. "UTR" includes APA changes involving PASs found only within the 3'UTR, and "CR-CR" APA comprises those PASs found only within the coding-region. The total number of genes used in each APA category is shown under each chart.

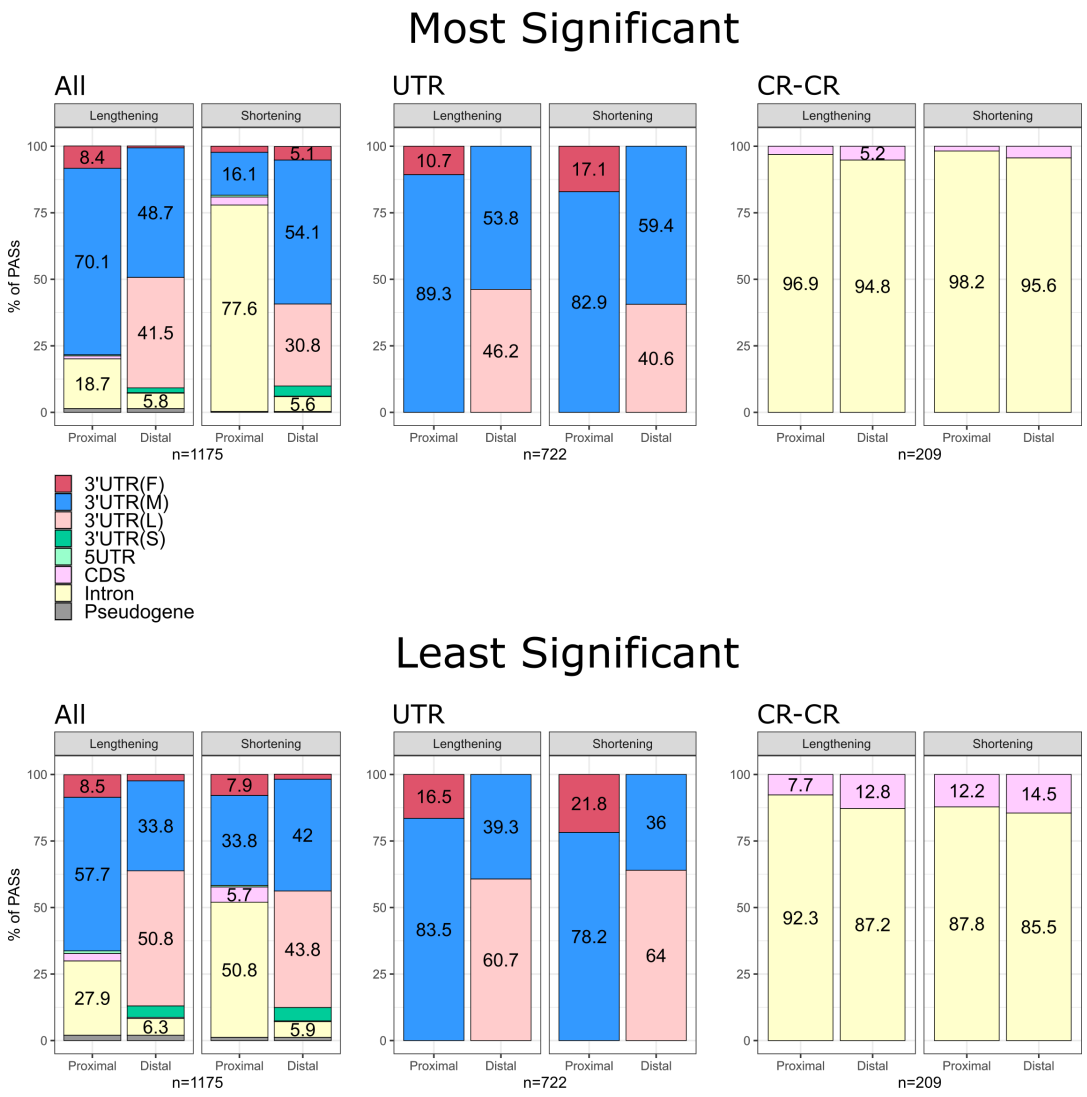


Figure 5.4: Location of PAS within the gene body of those regulated by APA (Top) and those not (Bottom) in the nucleus, 18hpi (WSN infection). PASs involved in APA in WSN-infected cells were classified depending on their location in the gene body. Those situated in the 3'UTR were further broken down into first (F), middle (M) and last (L) PAS. If there was just one annotated PAS in the 3'UTR, it is referred to as single (S). "All" refers to APA changes involving PASs found anywhere within the pre-mRNA. "UTR" includes APA changes involving PASs found only within the 3'UTR, and "CR-CR" APA comprises those PASs found only within the coding-region. The total number of genes used in each APA category is shown under each chart.

5.2.3 Differences in cis-elements underlie differential PAS usage during flu infection

In the previous section, we identified differences in usage of PASs containing certain poly(A) signal sequences. The core poly(A) signal sequences (e.g. the canonical hexamer) found at PASs are upstream elements, usually found within 40 nucleotides upstream of the cleavage site. However, this core signal is one component of a multi-partite sequence element, normally also consisting of less-defined regions upstream and downstream of the PAS, and it is the combined effect of these RNA sequence elements which give PASs their overall strength (Tian and Manley 2017). Noting our previous observations that usage of A-rich PASs increases while usage of distal-most PASs decreases in genes regulated by APA during flu infection, we next asked if there were any biases for cis-elements surrounding PASs which were upregulated or downregulated in WSN infection. As proximal and distal PASs are known to have differences in the cis-elements surrounding them (Hu, Lutz, et al. 2005; Tang, Yang, et al. 2022), we separated the proximal and distal PASs of genes significantly regulated by APA (P-adj less than 0.1 by DEXSeq). Additionally, we broke down each PAS into 3 parts referred to by their distance from the cleavage site: -100nt to -41nt (termed "Start"), -40nt to -1nt ("Mid") and +1nt to +100nt ("End"). The bioinformatic method applied here is outlined in chapter 2 of this thesis (section 2.7.3), with a visual guide found in Figure 2.3. We focused only on the nuclear fraction for this analysis, as we expected cis-elements to have the most impact co-transcriptionally.

Figure 5.5 displays how many 4-mers were enriched or depleted in each category of APA during Udorn infection in PASs regulated by APA compared to those that

are not. We observed generally higher 4-mer biases in the start and end subregions of each PAS category. This could suggest that, while there is variation in the PAS signal sequence between PASs regulated by APA and those that are not in Udorn infection, there is greater variation in the cis-elements downstream and upstream of the PAS. CR-CR APA seemed to have the lowest number of 4-mers, which is somewhat unsurprising, as it was the smallest APA category (with a total of 59 significant APA events).

In the WSN dataset, we similarly saw substantial numbers of both types of K-mers (4-mers and 6-mers) at every timepoint (except 3hpi) across all types of APA, strongly indicating that cis-elements contribute to PAS switching during WSN infection. Total numbers of significantly biased 4-mers and 6-mers at each timepoint throughout the WSN infection can be seen plotted in Figure 5.6.

To further explore the potential impacts of cis-element differences on PAS selection during WSN infection, we divided PASs in the same manner as was done for the Udorn analysis. First, PASs were split into those that are upregulated and downregulated in either proximal or distal PASs, since different cis-elements are found between these PAS types (Hu, Lutz, et al. 2005; Tang, Yang, et al. 2022), and then each PAS was further broken down into subregions encompassing the PAS and auxiliary regions. We saw the largest frequency biases for 4-mers in the "start" and "end" regions of PASs, with the highest number seen in downregulated distal PASs in the "end" subregion (Figure 5.7). Interestingly, this showed that while there was sequence variation in all subregions, the "mid" subregion (where the poly(A) signal sequence would be expected) had relatively fewer differences in cis-element sequences in PASs regulated by APA during WSN infection compared to both upstream ("start") and downstream ("end") of the cleavage site.

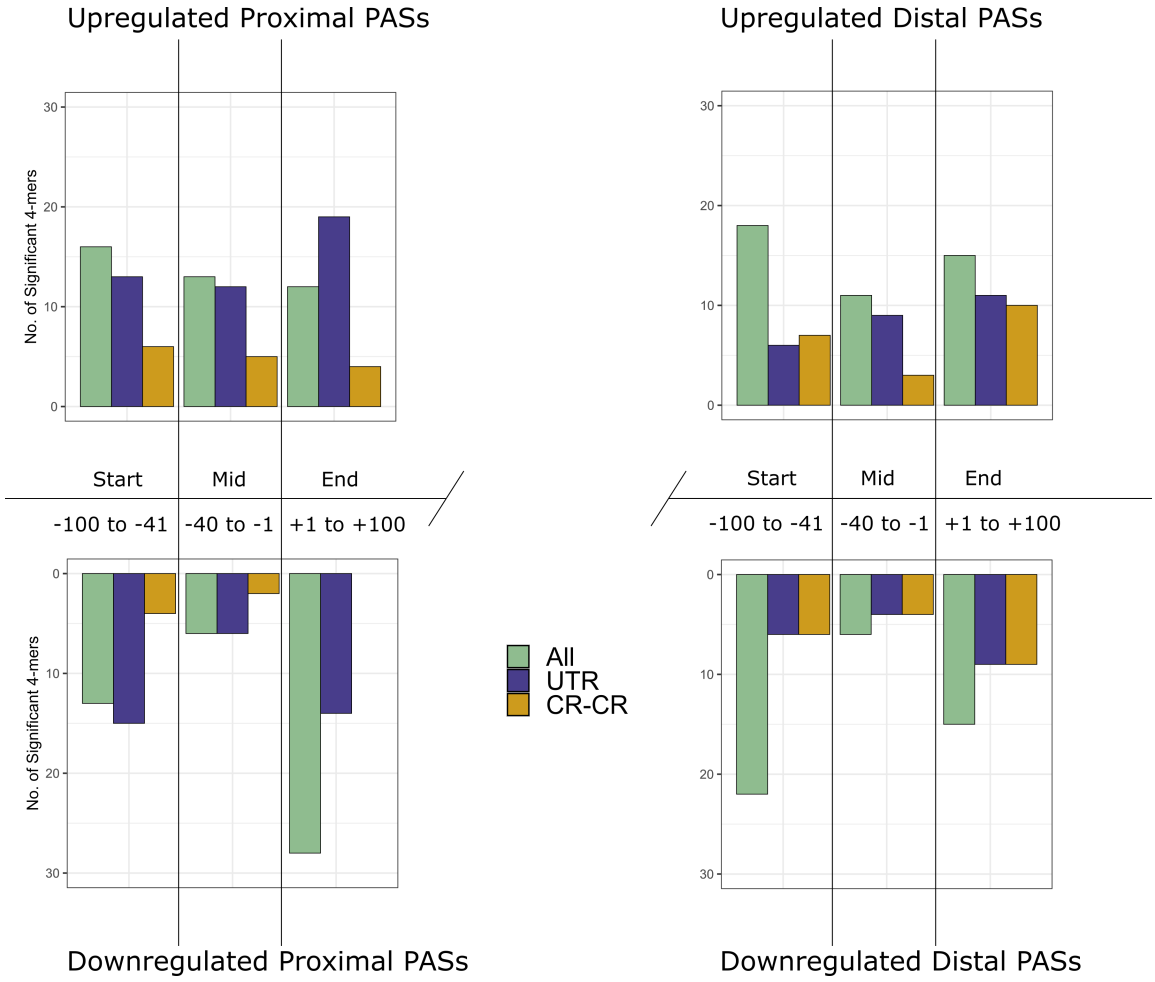


Figure 5.5: Numbers of significantly enriched or depleted 4-mers in PASs regulated by APA during Udorn infection (Nucleus). Proximal and distal PASs were separated, which were then further broken down into upregulated/downregulated PASs at proximal and distal sites. Three regions surrounding the PAS ("Start", "Mid" and "End") were scanned for significantly enriched or depleted 4-mers in PASs regulated by APA compared to those not, as calculated by Fisher's exact test (P-val less than 0.001).

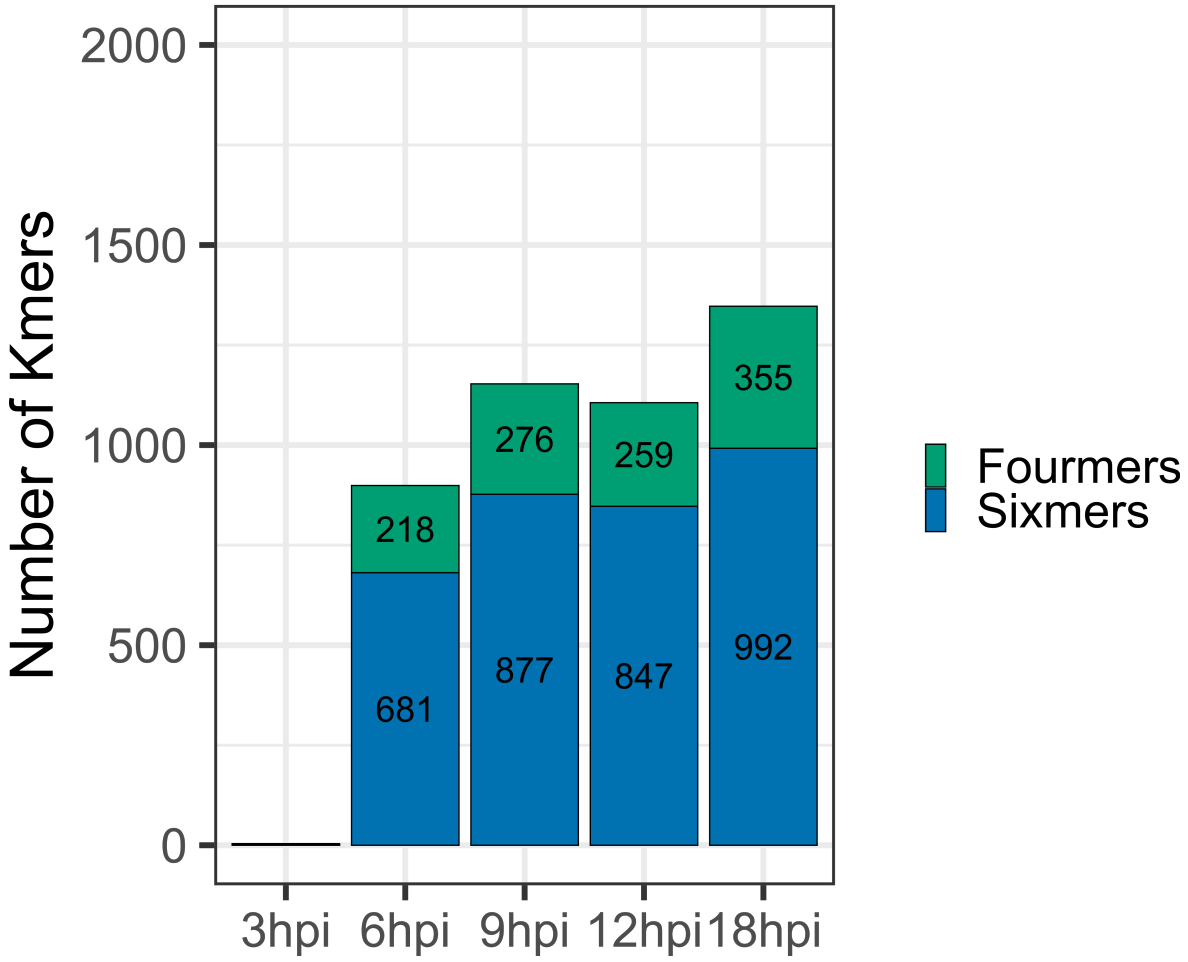


Figure 5.6: Unique 4-mers or 6-mers with significantly biased frequency at each timepoint throughout the WSN-infection (Nucleus), "All" APA.

These two subregions should correspond to where upstream and downstream elements reside, respectively (Elkon, Ugalde, and Agami 2013). Thus, relatively high amounts of K-mer differences in PASs regulated by APA suggest that cis-elements play a substantial role in differential PAS selection during WSN infection, with the upstream and downstream elements playing particularly important roles.

Having observed changes in K-mer frequencies at poly(A) sites regulated by APA during flu infection, and with numbers of K-mer biases pinpointed to subregions around PASs (Udorn: Figure 5.5, WSN: Figure 5.7), we next asked which specific 4-mers and 6-mers might be enriched or depleted in each subregion around upregulated or downregulated PASs during flu infection, with each strain. Since the combination of cis-sequences found at a PAS substantially influences its strength (Di Giammartino, Nishida, and Manley 2011), identifying the cis-regulatory elements which are significantly enriched or depleted in PASs undergoing APA during flu infection could share important insights as to how host APA is modulated in the presence of IAV specifically, and viruses more generally. To identify 4-mers and 6-mers which may contribute most to the alteration of host APA during IAV infection, we sorted all significantly-changing K-mers and sorted them based on occurrences.

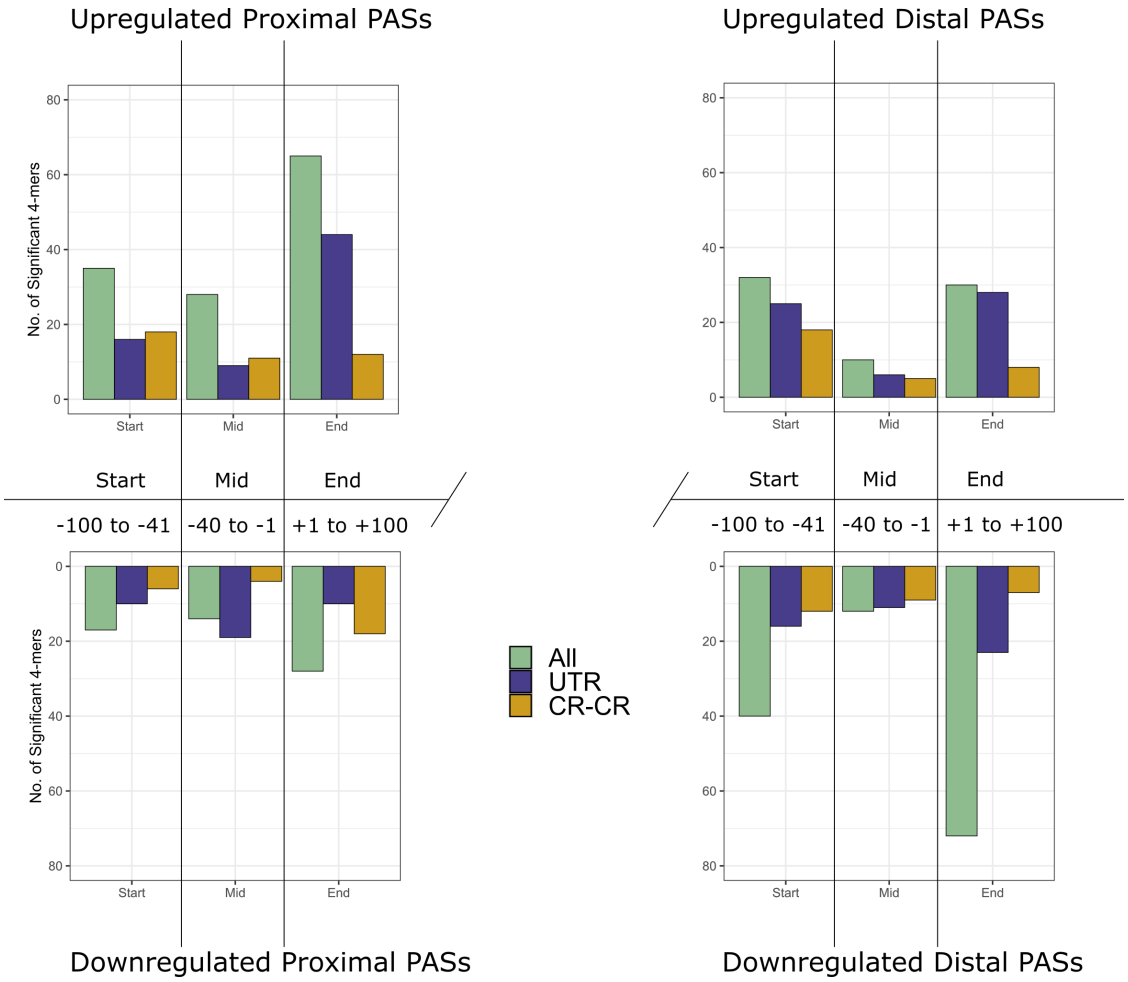


Figure 5.7: Numbers of significantly enriched or depleted 4-mers in PASs regulated by APA during WSN infection, 18hpi (Nucleus). Proximal and distal PASs were separated, which were then further broken down into upregulated proximal/distal PASs, or downregulated proximal/distal PASs. Three regions surrounding the PAS ("start", "mid" and "end") were scanned for significantly enriched or depleted 4-mers in PASs regulated by APA compared to those not, as calculated by Fisher's exact test.

Summary heatmaps for these analyses can be seen in Figures 5.8 and 5.9 for Udorn, and Figures 5.10 and 5.11 for WSN (18 hours post-infection). Corresponding heatmaps for the other WSN timepoints can be found in Appendix A.17.

We found that the A-rich 4-mer or 6-mer AAAA(AA) was significantly enriched in almost all upregulated proximal and distal PASs, and depleted in downregulated mid PAS sites and proximal PASs in the "All" APA category. This most likely reflected the shift to A-rich PASs that we reported earlier. Interestingly, this supports the findings we noted earlier (in section 5.2.1), which showed relative increases in usage of A-rich PAS signals in upregulated PASs, and decreases in downregulated PASs, although this does not explain the AAAAAA bias in the "start" and "end" PAS subregions.

Interestingly, the UUUU(UU) K-mer, a well-established binding site for the FIP1 cleavage and polyadenylation factor, was seen to be consistently depleted in the "start" subregion of downregulated PASs and enriched in upregulated PASs during infection with both influenza strains, including at all timepoints of WSN infection. FIP1 is believed to bind U-rich sequences upstream of the PAS (Lackford et al. 2014; Li, You, et al. 2015; Yuan, Hankey, et al. 2021). Interestingly, Li et al. reported enrichment of UUUU in the same "start" PAS region of downregulated poly(A) sites and depletion in upregulated distal "start" PASs when FIP1 was knocked down (Li, You, et al. 2015), suggesting that differences in FIP1 function during flu infection may account for this APA shift. Indeed, FIP1 has been reported to strongly interact with WSN NS1 (Wang, Fu, et al. 2017), an interaction which could explain our observations. This could indicate that activity of FIP1 (Lackford et al. 2014; Li, You, et al. 2015; Yuan, Hankey, et al. 2021) is modulated during WSN infection which thereby influences PAS selection. Addi-

tionally, ELAVL1 also binds U-rich motifs, and has been shown to influence PAS selection by blocking the binding of core CPA factors (Elkon, Ugalde, and Agami 2013; Turner and Diaz-Munoz 2018; Zhu, Zhou, et al. 2007).

G/U-rich motifs (CUCU, GUUU, UCUU, GGGG, GUGU and related 6-mers) also had biased frequency in several different categories. G-rich hexamers were found to change in frequency, particularly downstream of the PAS in the "end" region. UGUGUG/GUGUGU, known binding sites for CSTF64 (Hu, Lutz, et al. 2005; Yuan, Hankey, et al. 2021), were highly significantly biased particularly in the early- and mid-infection phases (see Appendix A.11 and A.13 to refer to 6hpi and 9hpi respectively). Importantly, depletion of either CSTF64 or its paralogue CSTF64τ has been reported to induce changes in APA profiles (Yao, Biesinger, et al. 2012). Additionally, Wang et al. observed significant alterations of the frequency biases of both UGUGUG and GUGUGU upon PCF11 knockdown (Wang, Zheng, Wei, et al. 2019). Downstream auxiliary sequences are known to be G- and C-rich (Hu, Lutz, et al. 2005; Tian and Graber 2012), potentially suggesting that those PASs with non-traditional downstream elements are more likely to be regulated by APA during flu infection.

Several U-rich hexamers such as UUGUUU, UUAGUU, CUUUUU, UUUUUC, UAGUUU and UUGCUU were enriched or depleted particularly in the "start" (upstream) and "end" (downstream) regions, which is notable as these sequences are binding sites for HNRNPC, a known regulator of APA (Fischl et al. 2019; Yang, Wang, et al. 2020). Indeed, HNRNPC modulates PAS usage by binding poly(U) tracts near the PAS, with the number and length of uridine stretches contributing to its regulatory impact (Gruber, Schmidt, et al. 2016). Additionally, C-rich and G-rich motifs such as CCCC(CC), GGGG(GG), GGGC, CGGG, GCGG and

AGGG were depleted, particularly in the "end" subregion at all timepoints of WSN infection. Interestingly, poly(C) regions act as binding motifs for HNRNPK, which can influence PAS site selection (Naganuma et al. 2012; Mikula et al. 2013), and other PCBP proteins.

The 6-mers UAUAUA and AUAUAU - and their 4-mer derivatives UAUA and AUAU - were also highly biased in frequency, mostly being enriched in upregulated PASs and depleted in downregulated PASs, particularly in the "end" subregion. UAUAUA and derivatives have been reported to be sequence elements usually upstream of the cleavage site, and are particularly common in yeast (Hu, Lutz, et al. 2005; Tian and Graber 2012; Wang, Zheng, Yehia, et al. 2018). This observed effect could suggest that the presence of these sequences downstream of the cleavage site could positively contribute to selection of PASs during IAV infection.

Overall, the findings in this section uncovered some of the mechanisms behind APA regulation during flu infection. The evidence presented here indicates that cis-element sequences surrounding PASs strongly influence PAS selection during flu infection. We found substantial increases in usage of PASs featuring a core A-rich poly(A) signal in both shortening and lengthening events. Furthermore, PASs enriched with AAAA(AA) sequences were more likely to be differentially used, while the converse was also found to be true: PASs lacking these A-rich sequences were least likely to be selected by APA. These alterations result from direct flu-host interactions such as that between the flu NS1 protein and the host core CPA factor CPSF30, or the proposed association between another flu, PA-X, and host CPSF5. Alternatively, secondary infection effects could explain changes in activity of regulators of APA.

Together, these findings strongly indicate that biases in PAS cis-element se-

quences regulate APA during IAV infection, particularly in regions upstream and downstream of the cleavage site. The findings here, derived from two different influenza A strains, suggest that many of these biases are not strain-specific. The AAAA(AA) K-mer was consistently enriched in upregulated PASs and depleted in downregulated PASs, suggesting a key role for it in the usage of PASs during IAV infection. Interestingly, PASs featuring A-rich poly(A) signal sequences rely on auxiliary elements surrounding the cleavage site (Nunes et al. 2010; Zheng, Liu, and Tian 2016; Wang, Zheng, Yehia, et al. 2018), suggesting that the PASs most likely to be regulated by APA during flu infection are reliant on auxiliary elements. With this in mind, we identified auxiliary sequences whose frequencies were biased in PASs regulated by APA included binding sites for FIP1, CSTF64 and HNRNPC. As mentioned earlier, these proteins each have the potential to regulate APA. Considering this, we suggest that mechanistically, differences in PAS sequence biases could partly explain APA shifts induced by influenza infection. These APA shifts could be brought about by altered activity patterns of RNA-binding proteins that bind these cis-elements such as FIP1, CSTF64 and HNRNPC during the context of virus infection. In fact, as mentioned earlier, a reported strong interaction between FIP1 and flu NS1 (Wang, Fu, et al. 2017), could account for these biased frequencies of FIP1 binding sites in PASs regulated by APA.

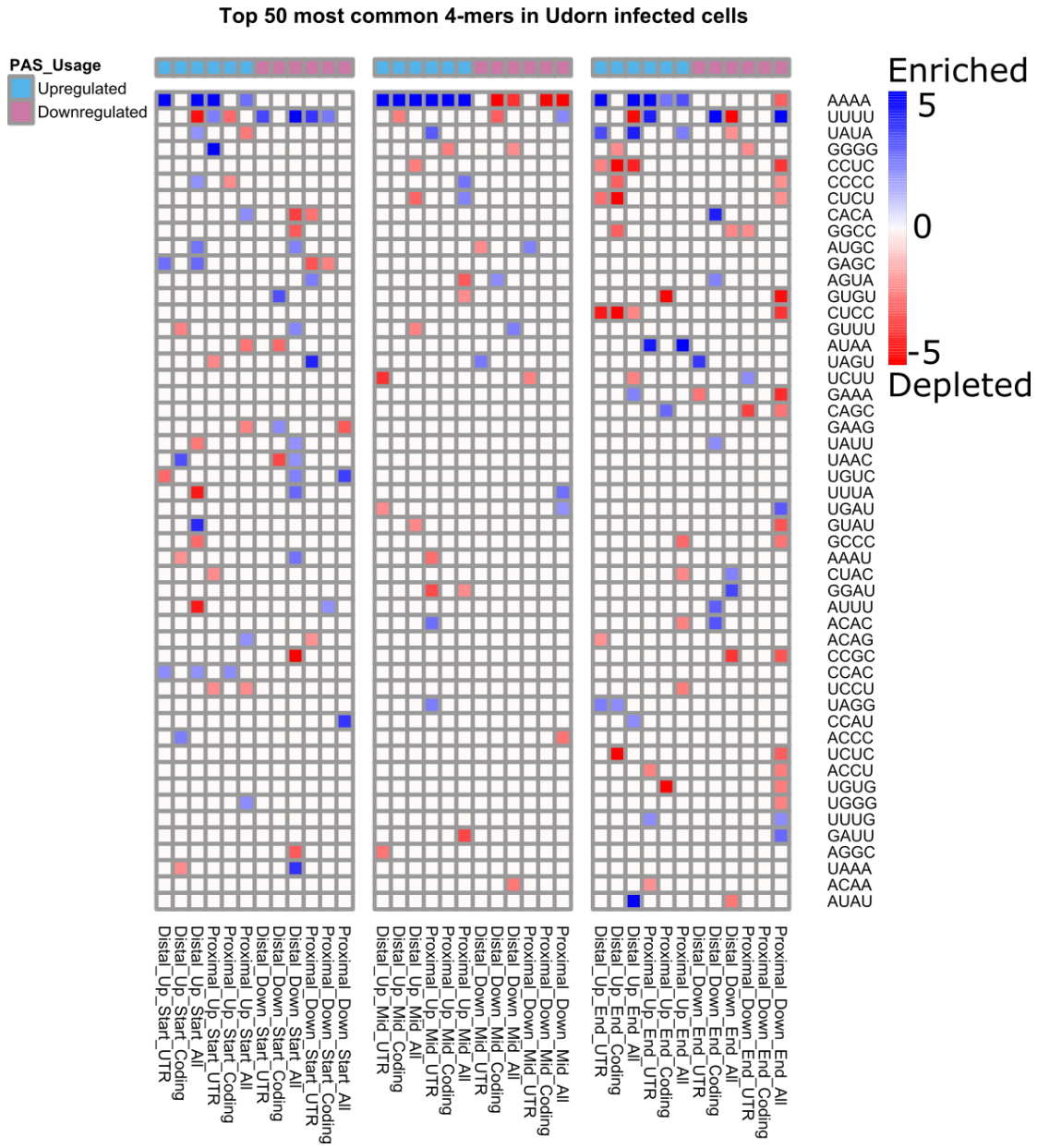


Figure 5.8: Top 50 most common enriched and depleted 4-mers (Top 50) in PASs regulated by APA during Udorn infection (Nucleus). PASs have been split into proximal and distal categories, while further being broken into 3 regions surrounding the PAS (Start, Mid and End). The frequency of the top 50 4-mers for each APA type - All, Coding (CR-CR) and UTR is shown (P-val less than 0.001). Significantly regulated PASs were compared with equivalent PASs of other genes, with the Fisher's exact test used to determine significance values for each 4-mer. The -log10(P-value) was then multiplied by 1 if the 4-mer was enriched, or by -1 if the 4-mer was depleted, and these values were then capped at either 5 or -5 for easier visualisation.

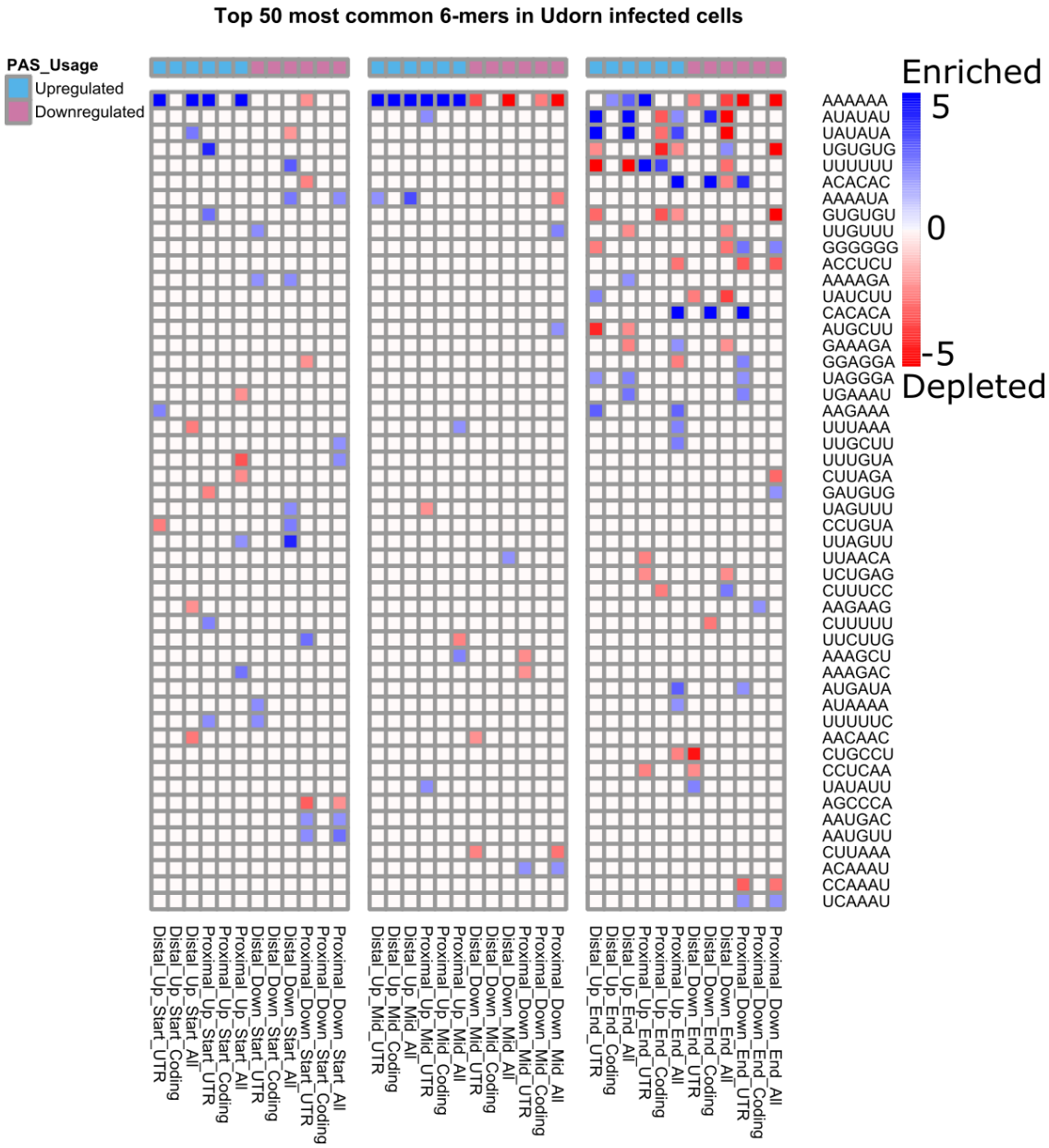


Figure 5.9: Top 50 most common enriched and depleted 6-mers in PASs regulated by APA during Udorn infection (Nucleus). PASs have been split into proximal and distal categories, while further being broken into 3 regions surrounding the PAS (Start, Mid and End). The frequency of the top 50 6-mers for each APA type - All, Coding (CR-CR) and UTR is shown (P-val less than 0.001). Significantly regulated PASs were compared with equivalent PASs of other genes, with the Fisher's exact test used to determine significance values for each 6-mer. The -log10(P-value) was then multiplied by 1 if the 6-mer was enriched, or by -1 if the 6-mer was depleted, and these values were then capped at either 5 or -5 for easier visualisation.

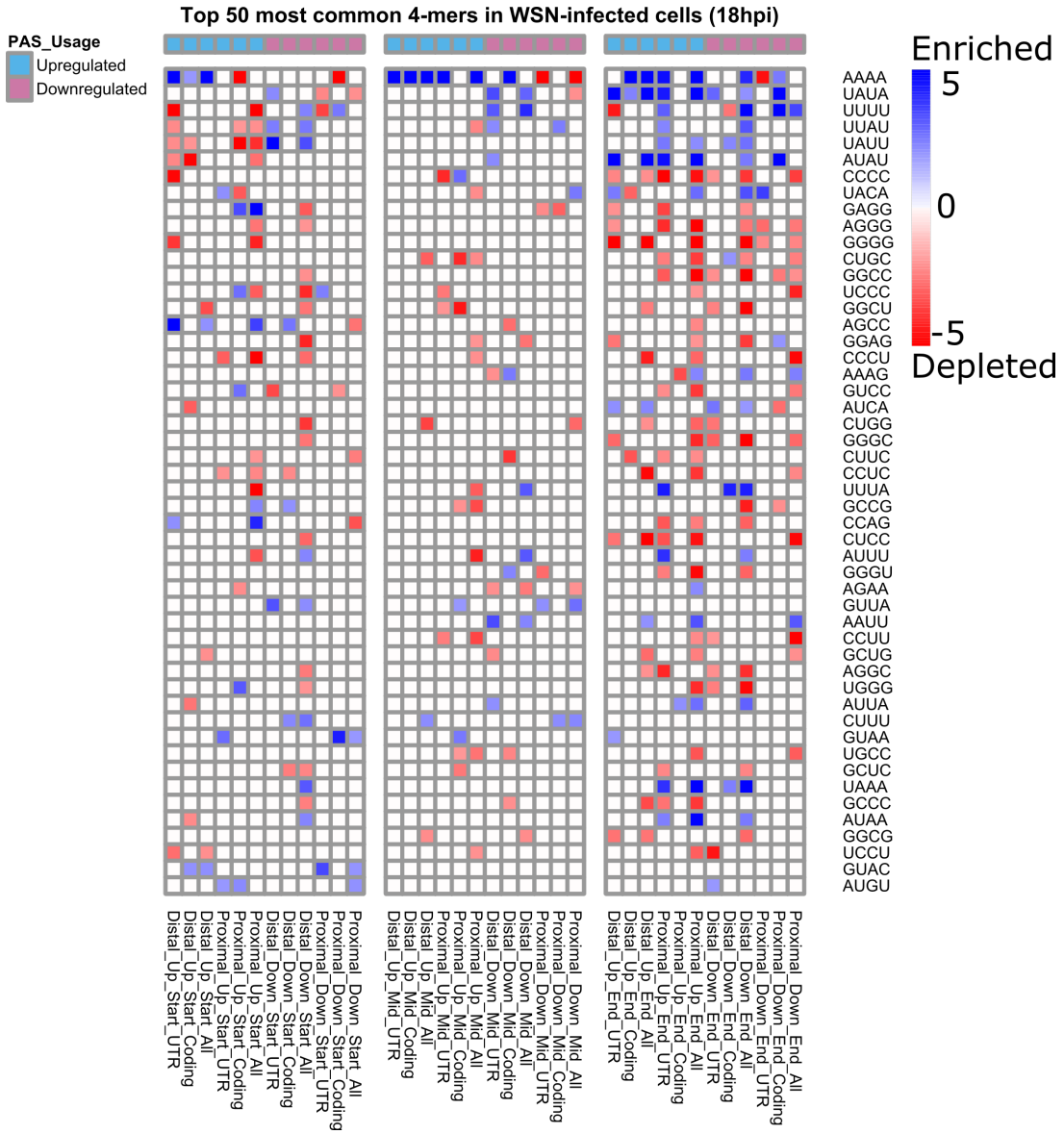


Figure 5.10: Top 50 most common enriched and depleted 4-mers (Top 50) in PASs regulated by APA during WSN infection, 18 hours post-infection (Nucleus). PASs have been split into proximal and distal categories, while further being broken into 3 regions surrounding the PAS (Start, Mid and End). The frequency of the top 50 4-mers for each APA type - All, Coding (CR-CR) and UTR is shown (P-val less than 0.001). Significantly regulated PASs were compared with equivalent PASs of other genes, with the Fisher's exact test used to determine significance values for each 4-mer. The -log10(P-value) was then multiplied by 1 if the 4-mer was enriched, or by -1 if the 4-mer was depleted, and these values were then capped at either 5 or -5 for easier visualisation.

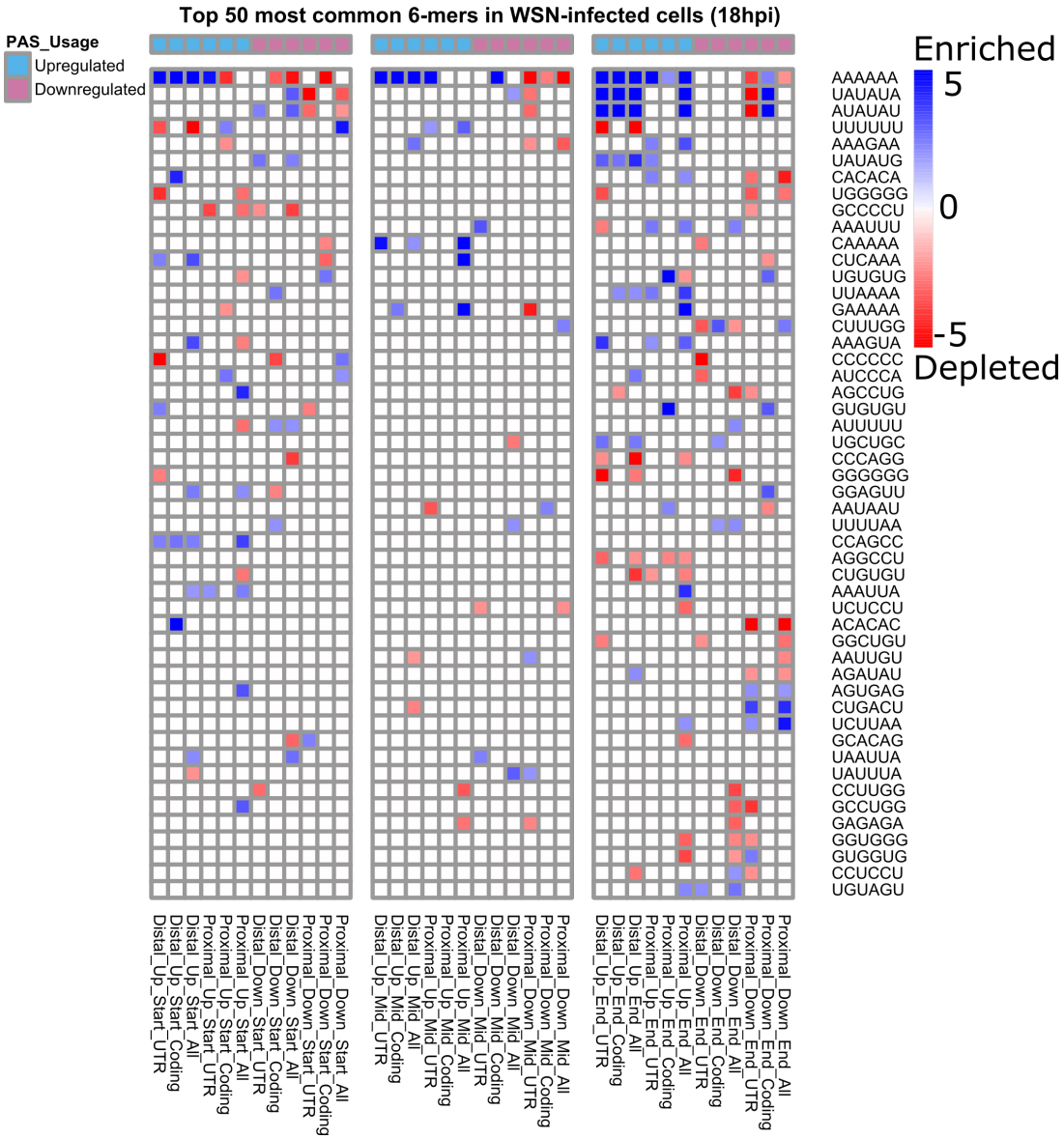


Figure 5.11: Top 50 most common enriched and depleted 6-mers (Top 50) in PASs regulated by APA during WSN infection, 18 hours post-infection (Nucleus). PASs have been split into proximal and distal categories, while further being broken into 3 regions surrounding the PAS (Start, Mid and End). The frequency of the top 50 6-mers for each APA type - All, Coding (CR-CR) and UTR is shown (P-val less than 0.001). Significantly regulated PASs were compared with equivalent PASs of other genes, with the Fisher's exact test used to determine significance values for each 6-mer. The -log10(P-value) was then multiplied by 1 if the 6-mer was enriched, or by -1 if the 6-mer was depleted, and these values were then capped at either 5 or -5 for easier visualisation.

5.3 Discussion

Cis-elements in and around PASs ultimately control cleavage and polyadenylation (Tian and Manley 2017), and by comparing genes regulated by APA with those that are not, we uncovered which precise cis-element sequences may make genes more likely to PAS switch during influenza infection. We therefore wanted to explore their distributions in those PASs differentially used during flu infection with the aim of uncovering mechanistic reasons underpinning APA in host cells under viral stress. In this chapter, we performed bioinformatic analyses on the data described in chapters 3 and 4, with the aim of uncovering the mechanisms underpinning APA in host cells under viral stress.

Focusing only on nuclear RNA, we first investigated usage of poly(A) signals - the core cis-element - in those PASs regulated by APA, and observed increases in A-rich PAS signals, in every subtype of APA at every timepoint, suggesting a bias for PASs with this signal sequence during flu infection. We also found overall decreases in the usage of distal-most PASs (3'UTR L) relative to PASs further upstream for lengthening genes at every timepoint. This observation was particularly interesting as distal-most PASs are considered to be stronger (Legendre and Gautheret 2003; Tang, Yang, et al. 2022), suggesting that the availability or efficiency of certain cleavage and polyadenylation factors is modulated during WSN infection and thereby influencing PAS selection, and inducing a shift in usage to weaker PASs.

We also examined what downstream or upstream elements may be biased in frequency in PASs regulated by APA by performing an extensive K-mer analysis on three subregions of each PAS. We identified substantial numbers of 4- and 6-

mers enriched or depleted in both infections, with particularly high concentrations around the upstream and downstream subregions. These two subregions correspond to reported locations of upstream and downstream auxiliary elements, suggesting that PAS selection during flu infection relies heavily on these cis-elements. We reported the top 50 K-mers for each timepoint. We saw a strong enrichment of A-rich motifs in upregulated PASs and a depletion in downregulated PASs, suggesting a strong preference for PASs with A-rich sequences downstream and upstream of the poly(A) signal. This A-rich motif was also depleted in downregulated PASs in the "mid" subregion (-40 to -1 nucleotides from the cleavage site). This affirmed our earlier findings that there was a bias in PAS selection towards A-rich PASs during influenza infection, although it was unexpected that this was also observed immediately upstream and downstream of the PAS, and not entirely centered around the PAS signal.

The UAUAUA sequence and its derivatives, a reported upstream element (Hu, Lutz, et al. 2005; Tian and Graber 2012; Wang, Zheng, Yehia, et al. 2018), were enriched in upregulated PASs and depleted in downregulated PASs. C-rich and G-rich motifs, known to feature in downstream elements (Hu, Lutz, et al. 2005; Tian and Graber 2012), were depleted in "end" subregions.

Furthermore, we identified biases for several G/U-rich 4-mers and 6-mers which are often binding sequences for several CPA regulators. One important CPA factor, CSTF64, is known to bind UGUGUG and GUGUGU (Hu, Lutz, et al. 2005), and we observed significant depletion of both hexamers in the "end" regions (+1 to +100) of PASs, particularly in the early- and mid-infection timepoints for the WSN infection, suggesting an important role for CSTF64 in APA regulation at these phases.

Several U-rich sequences which are binding sites for HNRNPC were also determined to change frequency in PASs regulated by APA, including UUGUUU, UUAGUU, CUUUUU, UUUUUC, UAGUUU and UUGCUU. HNRNPC is a known regulator of APA (Fischl et al. 2019; Yang, Wang, et al. 2020), so this could suggest a role for HNRNPC in PAS selection during flu infection.

We also reported a strong depletion of UUUU(UU) sequences in upregulated PASs, which is noteworthy as FIP1 binds this sequence (Lackford et al. 2014; Li, You, et al. 2015; Yuan, Hankey, et al. 2021). Interestingly, FIP1 has been found to strongly interact with WSN NS1 in a study exploring influenza protein interactomes (Wang, Fu, et al. 2017). We therefore propose that dysregulation of FIP1 activity by association with NS1 contributes to flu-induced host APA. Knockdown of FIP1 has already been shown to induce 3'UTR lengthening (Li, You, et al. 2015) which, combined with our findings, suggests that FIP1 interaction with NS1 could contribute to the strong lengthening effect we observe in IAV infection, particularly within nuclear RNA. Interestingly, we observed enrichment and depletion of the UUUU tetramer in the same PAS subregions reported by Li et. al when FIP1 was knocked down in their study (Li, You, et al. 2015). Investigating the correlation of 3'UTR APA patterns for genes in IAV-infected cells and in cells with depleted FIP1 levels would be the next step in examining the contribution of FIP1 to APA regulation during flu infection.

Altogether, these findings constitute strong evidence that cis-elements surrounding PASs play a major role in the regulation of APA during IAV infection. An increase in usage of A-rich PASs and a decrease in usage of strong distal-most PASs suggest that flu infection induces usage of weaker PASs which are reliant on auxiliary cis-elements. Moreover, we identified important cis-element sequences

which are significantly enriched or depleted in PASs regulated by IAV infection. A clear example of this is a binding site for FIP1, which is depleted in PASs that are preferentially used, indicating that the activity of this core 3'-end factor is modulated in the context of flu infection. Indeed, FIP1 has been reported to strongly interact with NS1, suggesting that its activity is directly modulated by virus infection which consequently has implications on host APA profiles.

5.4 Summary

In this chapter, we utilised bioinformatic techniques to uncover part of the underlying mechanism that influences which PASs are upregulated or downregulated during flu infection. We found that there is a preference for A-rich PASs to be regulated by APA during flu infection. We also identified auxiliary sequences downstream and upstream of the core poly(A) signal which are enriched or depleted in differentially regulated PASs during IAV infection. Intriguingly, we uncovered a shift in usage towards A-rich poly(A) sites accompanied by depletions of U-rich auxiliary elements. As FIP1 binds U-rich sequences on either side of the poly(A) signal, and NS1 has been reported to strongly interact with FIP1, our findings suggest that FIP1 dysregulation during influenza infection contributes to wider host APA profile changes.

These are important insights which add to our understanding of the transcriptomic changes induced by virus infection, and also to the wider APA field. Further investigation by examining the activity of the proteins highlighted in this work will broaden our understanding of precisely how these sequence biases influence APA during flu infection.

Chapter 6

Discussion and Future Directions

6.1 Final discussions, future directions and conclusions

6.1.1 Thesis summary

This thesis explored host alternative cleavage and polyadenylation profile changes induced by two different influenza A strains. A powerful experimental approach was followed by which host nuclear and cytoplasmic mRNAs were independently sequenced, addressing the common issue faced in the field of being unable to distinguish co-transcriptional and post-transcriptional APA events. To analyse the resulting data, a novel bioinformatic pipeline was used to identify differential PAS usage. Deeper downstream analyses enabled by the pipeline, including K-mer and cis-element studies, partly uncovered mechanisms of APA regulation during flu infection.

In the first results chapter of the thesis, assembly and use of a new custom-built

computational pipeline is outlined, which is utilised to identify host APA changes induced by a H3N2 influenza A strain, Udorn, by comparing PAS usage in early infection (3hpi and 6hpi) with that in late infection (12hpi and 18hpi). The second results chapter of the thesis builds on these findings in a more granular way, by investigating host gene expression changes at discrete timepoints with replicates during an infection period using the H1N1 influenza A/WSN/33 strain. Lastly, the final results chapter uncovers cis-element sequences which are enriched or depleted in PASs regulated by APA during flu infection, which provides some insight into the mechanisms underlying PAS selection during viral stress.

6.1.2 Widespread host gene expression changes are induced during IAV infection

Infection of lung epithelial cells by either of the two different IAV strains used in this study (Udorn or WSN) brought about dramatic changes in the host APA landscape. Hundreds of PAS-switching events were induced during both infections, with the more granular WSN experiment revealing their chronological spread. With this dataset, it was shown that APA (both 3'UTR and coding-region) events steadily increased from 6hpi onwards. Importantly, the RNA subcellular fractionation approach used in this study enabled us to pinpoint APA changes to either the nucleus or cytoplasm and thereby distinguish co-transcriptional and post-transcriptional APA. Gene ontology analyses of the genes regulated by APA revealed a functionally distinct enrichment pattern when compared with those genes which are differentially expressed as part of the wider transcriptomic response to IAV infection. We identified a bias for longer APA isoforms in the nu-

cleus, suggesting that these genes are regulated by mechanisms affecting nuclear retention or stability of transcripts during IAV infection. Indeed, when we isolated those genes which lengthened only in the nucleus (and not in the cytoplasm), we found substantial enrichment of 3'UTR motifs recognised by several RNA-binding proteins which could alter transcript properties, such as ELAVL1 and HNRNPK). We also established a link between 3'UTR APA and gene expression levels by taking full advantage of our separate datasets for the nucleus and cytoplasm and identifying several RNA-binding proteins (including members of the SR-family as well as HNRNPK) whose overall gene expression levels in the cytoplasm are positively related with their regulation by 3'UTR APA in the nucleus. This could therefore represent a model by which expression of important mRNA processing genes in the cytoplasm is regulated by switching 3'UTR PAS sites in the nucleus during flu infection. Confirming whether or not the alteration of 3'UTR length for these genes affects their protein output, and consequently the ability of IAV to replicate, could be the next step for further research.

For coding-region APA, we detected the reverse trend, with a general bias towards shortening. As these APA events involved a switch from a 3'UTR PAS to an internal proximal PAS, these changes most likely represent a form of gene suppression. We highlighted some of these changes which we speculate could have important physiological implications for the infected cell by impacting the ability of the virus to replicate either as host factors which interact directly with viral replication or as host stress response factors. This category of genes included ZC3HAV1, which is known to inhibit IAV replication by promoting degradation of IAV proteins as well as by directly inhibiting expression of flu proteins, and DHX30 which is reported to inhibit IAV replication by associating with the NS1-

CPSF30 complex. Further work identifying the most impactful genes regulated by coding-region APA in the context of IAV replication could yield pharmacological targets for future drug discovery.

Taking our findings on host APA changes together, it seems likely that IAV targets and modulates host APA to convert a hostile cellular environment to a more beneficial one for its replication. However, based on our data alone, it is not clear what proportion of these APA changes are induced by either the virus alone or by host mechanisms, and disentangling these two elements should be a priority for future work.

6.1.3 Mechanisms underpinning APA during IAV infection

What are the mechanisms behind differential PAS selection during flu infection? Our experimental approach allowed us to address this question by focusing on nuclear APA changes. By first comparing the aUTRs of longer APA isoforms with other aUTRs, we found a strong enrichment for motifs associated with RNA-binding proteins such as ELAVL1 and HNRNPK. Both of these proteins have previously been reported to alter PAS choice by blocking the binding of core cleavage and polyadenylation factors, leading us to hypothesise that these proteins could play a major role in the generation of long APA isoforms during IAV infection. Furthermore, other RBPs which bind motifs that were enriched in "long" 3'UTRs (such as SAMD4A, PCBP1 and PCBP2) have also previously been shown to impact gene expression, which may also partly explain the observed nuclear lengthening bias.

To further investigate the mechanistic causes which influence PAS choice in

the context of IAV infection, we studied cis-element sequences found at differentially regulated PASs. The overall strength of a poly(A) site is determined by the combined effect of cis-elements surrounding the cleavage site, so identifying trends in those PASs which are regulated during flu infection is a necessary step in uncovering the mechanisms behind IAV-induced APA. We analysed multiple PAS subregions and looked for sequences biases, including differences in usage of the core poly(A) signal sequence itself. Our extensive analyses revealed the same general trends between both IAV strains. Interestingly, we observed an increased usage of PASs featuring A-rich poly(A) signals. It has been previously reported in the literature that A-rich poly(A) signals, in conjunction with strong downstream sequence elements, are sufficient to direct 3'-end processing, although the exact mechanisms behind this are not fully understood. Downstream and upstream elements are well known components of the multi-partite poly(A) signal, and core factors bind these sequences during the course of 3'-end formation. We identified many K-mers which were enriched or depleted upstream and downstream of differentially regulated PASs during IAV infection. We observed significant frequency biases for the GUGU(GU) K-mer, which is a CSTF64 binding site. Notably, the WSN analysis revealed that frequency biases for this K-mer were mainly found in the early- and mid-infection phases, suggesting that this is when APA regulation mediated by CSTF64 is most prominent. Interestingly, the UUUU(UU) K-mer, a well-established binding site for the FIP1 cleavage and polyadenylation factor, was found to be consistently depleted in downregulated PASs and enriched in upregulated PASs. This finding raises the intriguing prospect that FIP1 activity is modulated during flu infection, reinforced by findings in the literature which report that knockdown of FIP1 leads to similar patterns of UUUU(UU) frequency

bias in PASs regulated by APA. Moreover, FIP1 has previously been reported to strongly interact with NS1. Future work is needed to ask if altered FIP1 activity is responsible for this effect in the context of flu infection.

Additionally, investigation into how exactly these cis-element sequence biases regulate APA during virus infection would yield valuable insights into how host cleavage and polyadenylation is modulated by viral infection and would also be of interest to the wider APA field. This could be examined by altering protein levels of 3-end processing factors to identify changes in usage of particular sequences or by performing similar analyses in cells infected with other viruses.

6.1.4 Concluding remarks

Viral infection results in massive changes to the host gene expression landscape which directly influence the infection outcome. This thesis investigated the impact of influenza A infection on an important element of host gene expression, cleavage and polyadenylation, using H3N2 and H1N1 IAV strains (Udorn and WSN, respectively). Overall, transcriptome changes induced by flu infection were largely strain-independent. Widespread changes in the host APA landscape were identified. These changes, or APA events, were further categorised into those occurring entirely within the 3'UTR and those that alter the coding capacity of mRNA transcripts. APA events involving physiologically relevant genes in the context of IAV infection were highlighted.

Mechanisms mediating these poly(A) site switches were uncovered by analysing nucleotide sequences surrounding PASs. The findings in this study suggest that aUTR-binding proteins (such as ELAVL1 and HNRNPK) influence PAS choice

in the nucleus and contribute to nuclear retention during flu infection. Detailed analyses of PAS cis-elements revealed shifts in usage to A-rich poly(A) sites and biases in sequences surrounding the cleavage site. One of these sequences is a binding site for FIP1, and is depleted in PASs used during flu infection. We therefore propose that deregulation of FIP1 activity, potentially by an established strong NS1 interaction, contributes to widespread flu-induced APA changes. Other proteins which may regulate APA during flu-infection based on these findings include CSTF64, HNRNPC and HNRNPK. The evidence presented in this work therefore demonstrates that cis-element sequences play a major role in the regulation of APA during IAV infection.

There are many exciting questions to ask as part of further research. Are the APA changes unique to influenza A infection, or are they shared with other viruses? How responsible is IAV itself for these changes, or are they largely attributable to the host immune response? What is the exact physiological impact of viral-induced APA on viral replication? How exactly do these cis-element sequence biases influence PAS choice in flu-infected cells?

Appendix A

Appendix

A.1 Full lists of genes experiencing a significant PAS switch during Udorn infection (Nucleus)

Gene	P-adjusted	DeltaRA	Direction
AJUBA	4.54E-07	-0.441811384	Lengthening
SYNCRIP	1.29E-06	-0.461445756	Lengthening
HNRNPU	3.00E-06	-0.125205288	Lengthening
DCP1A	4.20E-06	0.492905109	Shortening
ATP6V1B2	5.64E-06	-0.576150393	Lengthening
DIABLO	6.04E-06	-0.55941173	Lengthening
UPF1	1.60E-05	0.701754386	Shortening
CKS1B	3.34E-05	-0.202687223	Lengthening
AKT2	8.68E-05	0.136051467	Shortening
GNPDA1	0.000116179	0.316079215	Shortening
NAV2	0.000148312	-0.624592075	Lengthening
COPS3	0.000154467	-0.409580007	Lengthening
ELAVL1	0.000160347	-0.530810587	Lengthening
TMEM135	0.000271801	-0.229437229	Lengthening
MYO18A	0.000326332	-0.720867869	Lengthening
TIAF1	0.000326332	-0.720867869	Lengthening
RAB3GAP1	0.000400197	0.089263041	Shortening
SYDE1	0.000442298	-0.605256926	Lengthening
PSMD11	0.000769277	0.042760727	Shortening
ASAP1	0.000775125	0.271000409	Shortening
EIF4EBP1	0.000775125	0.567481884	Shortening

IAH1	0.000787851	-0.612580231	Lengthening
TRIM44	0.000787851	-0.50480226	Lengthening
GTPBP4	0.000969392	-0.549521002	Lengthening
KLF5	0.00118675	0.311789564	Shortening
ZFP36L2	0.001335184	-0.376666667	Lengthening
RAB14	0.001455865	0.247636214	Shortening
ISG20L2	0.001467194	0.36419164	Shortening
DDX3Y	0.00154545	0.482370408	Shortening
NBPF9	0.00154545	-0.270509978	Lengthening
IWS1	0.001553595	-0.313512361	Lengthening
UBA1	0.001567851	-0.578616352	Lengthening
INTS2	0.001575442	0.254032258	Shortening
PLEKHG2	0.001575442	-0.637837838	Lengthening
PTP4A2	0.001617757	-0.274299313	Lengthening
HSPA4L	0.001645214	-0.46278826	Lengthening
AK2	0.001880671	0.323892974	Shortening
NAGLU	0.002309107	-0.432075171	Lengthening
NGRN	0.002368487	-0.201864553	Lengthening
CYLD	0.002744626	-0.557166529	Lengthening
LARP4	0.002816637	-0.055519897	Lengthening
SLC25A37	0.002926097	0.07509005	Shortening
NEDD9	0.003608756	0.320467836	Shortening
NUCKS1	0.003608756	0.056143518	Shortening
PHLDA3	0.003608756	-0.370776256	Lengthening
BCL9L	0.004262003	0.35530063	Shortening
MYADM	0.004262003	0.395053763	Shortening
YLPM1	0.004704041	-0.451550495	Lengthening
DSTYK	0.004749086	0.590022506	Shortening
SET	0.005466356	0.047027349	Shortening
TOMM20	0.005587625	0.360468027	Shortening
EIF3M	0.006024925	-0.447045913	Lengthening
NUCB1	0.006148301	-0.370965374	Lengthening
ABL2	0.006301886	-0.459028703	Lengthening
DHX9	0.006423142	-0.136064584	Lengthening
AEBP2	0.006438612	-0.349509804	Lengthening
KIAA1549	0.006613936	0.339922952	Shortening
MALSU1	0.006777264	-0.665916592	Lengthening
MORF4L1	0.007220816	-0.016798854	Lengthening
RYBP	0.007456057	0.366077505	Shortening
NCBP2	0.007792768	-0.673740053	Lengthening
YTHDF2	0.009038382	0.33894276	Shortening
MAML1	0.009188476	-0.398408488	Lengthening
SRSF6	0.010596147	-0.431338028	Lengthening
CHCHD3	0.010614163	0.373257679	Shortening
PUM1	0.012173201	0.384313725	Shortening
FNIP1	0.012258859	0.348615269	Shortening
DCAF8	0.012359371	0.015815085	Shortening
F2R	0.013146401	-0.529330573	Lengthening

SLC38A2	0.01379099	0.027116602	Shortening
RABGAP1	0.013902415	0.321288087	Shortening
DLGAP5	0.014027784	-0.558441558	Lengthening
MRPL27	0.014255275	-0.401244677	Lengthening
RNF20	0.014255275	-0.258716994	Lengthening
RSPRY1	0.01435683	0.25572796	Shortening
IRS2	0.014801543	0.178035982	Shortening
KLHDC2	0.01498616	-0.693163752	Lengthening
OS9	0.01498616	-0.486534553	Lengthening
KDM5B	0.015056755	-0.001373626	Lengthening
FRMD8	0.015133069	-0.626984127	Lengthening
WDR1	0.015379354	0.217618947	Shortening
HMGCS1	0.015477991	-0.27248104	Lengthening
TNPO2	0.015588354	0.263285701	Shortening
NR2F2	0.015597844	-0.600272294	Lengthening
TCERG1	0.015687506	-0.334786118	Lengthening
MRPL45	0.015976151	-0.205271565	Lengthening
GTF3C4	0.015998739	0.279188504	Shortening
EIF3H	0.016417431	-0.410780889	Lengthening
RACGAP1	0.016987597	-0.306878307	Lengthening
CAP1	0.017480529	0.237131715	Shortening
COPS2	0.017511477	-0.322097021	Lengthening
EIF5B	0.017550322	-0.113305322	Lengthening
METAP2	0.017964461	0.040699223	Shortening
SON	0.018037077	-0.482703546	Lengthening
GTF2A2	0.01824312	-0.493157263	Lengthening
NSD3	0.01893511	0.356145967	Shortening
PHF8	0.019053217	-0.458221024	Lengthening
DARS	0.019489707	0.064267069	Shortening
ATXN7L3	0.019690028	-0.402362869	Lengthening
ASNSD1	0.019717102	-0.373313783	Lengthening
SPAG9	0.021155324	0.189840723	Shortening
MDM2	0.021303957	0.033683835	Shortening
RAB5A	0.02189017	-0.456676657	Lengthening
PLEC	0.022556948	0.074789916	Shortening
LRP6	0.02302671	0.5	Shortening
TMEM87A	0.023243235	0.3761223	Shortening
MKKS	0.024043586	-0.407518797	Lengthening
TOP1	0.02455258	0.327490592	Shortening
CDC27	0.024896009	0.383965015	Shortening
FAM126A	0.025445027	-0.568326948	Lengthening
GSK3B	0.025768524	-0.251559774	Lengthening
ZNF227	0.026137343	-0.5	Lengthening
SLC38A1	0.026700627	0.488506833	Shortening
CD63	0.027650317	-0.061407495	Lengthening
MTCH2	0.027863101	-0.437397067	Lengthening
CLASP1	0.028338598	-0.063470067	Lengthening
PGD	0.029999807	0.057052213	Shortening

ZNF687	0.030735999	-0.52688172	Lengthening
ABHD2	0.031100534	0.392642643	Shortening
ARL4C	0.031100534	0.08958728	Shortening
FAT1	0.032069422	0.219016414	Shortening
ALKBH3	0.032566398	-0.317073171	Lengthening
CSNK1A1	0.032821128	-0.253803832	Lengthening
RAD23A	0.032821128	0.098917952	Shortening
ATP6V1D	0.034426658	-0.232489451	Lengthening
ARL8B	0.034595596	0.32983804	Shortening
DDX21	0.035137583	0.191804341	Shortening
DENR	0.035285622	-0.448818898	Lengthening
POLR1C	0.035285622	0.105649303	Shortening
TSC22D1	0.036371465	-0.313679745	Lengthening
PAK2	0.036671897	-0.535846438	Lengthening
HIPK3	0.037871354	0.484313725	Shortening
DNAJA1	0.038069399	-0.578935949	Lengthening
HSPA5	0.03922148	-0.145960754	Lengthening
DYNC1I2	0.040947534	-0.177178125	Lengthening
H2AFJ	0.041772751	0.079217148	Shortening
SPEN	0.042693672	-0.245675377	Lengthening
HMGB1	0.044609714	0.18959538	Shortening
SIAH2	0.044717348	0.290167866	Shortening
ARID2	0.045543091	0.485271318	Shortening
FBXO30	0.045619294	-0.580503145	Lengthening
SRRM1	0.045857031	-0.510992672	Lengthening
B4GALT5	0.046318726	0.152614568	Shortening
UBA2	0.046426404	0.597435897	Shortening
CACUL1	0.046798906	0.18791964	Shortening
SLC12A4	0.047084984	-0.459401709	Lengthening
FOXN2	0.048015184	-0.375879165	Lengthening
CLTC	0.048191636	-0.016687809	Lengthening
ARL5B	0.050024157	0.309195924	Shortening
NAMPT	0.050024157	0.334309529	Shortening
MLEC	0.051620826	0.316821192	Shortening
ZC3H11A	0.052278208	-0.369269761	Lengthening
RBM17	0.052621466	0.296899841	Shortening
PCGF3	0.052999745	-0.732251949	Lengthening
CSNK2A1	0.053665146	0.222433059	Shortening
MRPS7	0.053988721	-0.249295207	Lengthening
SLC35B1	0.055296386	-0.305662806	Lengthening
CHORDC1	0.055658049	0.426942568	Shortening
GFM1	0.056681951	-0.305288462	Lengthening
DNAJC2	0.057810251	-0.54157375	Lengthening
HMGB2	0.05803266	-0.255714286	Lengthening
EPC1	0.058625962	0.287558988	Shortening
B4GALT1	0.059137633	-0.071642984	Lengthening
TRMT10C	0.059252485	-0.680250784	Lengthening
TOB2	0.059323114	0.4489093	Shortening

YIPF6	0.060697818	0.148063006	Shortening
THUMPD1	0.061218565	0.299534884	Shortening
CRK	0.061449373	0.185774799	Shortening
LMAN1	0.062012395	0.322368421	Shortening
RBM15	0.062012395	-0.294460641	Lengthening
FOSL2	0.062504991	-0.063101088	Lengthening
ARID4B	0.062566632	0.048780488	Shortening
E2F1	0.062677006	0.40286186	Shortening
ETAA1	0.062762247	-0.095497574	Lengthening
INO80D	0.062762247	0.502380952	Shortening
KHSRP	0.062762247	0.010565062	Shortening
NMT2	0.062774304	-0.366220562	Lengthening
EI24	0.063185263	0.21911793	Shortening
ETS1	0.063559468	0.366970532	Shortening
EIF2S3	0.063659887	-0.384931717	Lengthening
PPP1R15A	0.063659887	-0.087962253	Lengthening
ZNF131	0.064796524	0.098554845	Shortening
ZRANB2	0.065005017	0.519119769	Shortening
ARMCX3	0.065341615	0.200005553	Shortening
UAP1	0.067547312	-0.532894737	Lengthening
RBM14	0.068297748	-0.427070707	Lengthening
RPL23A	0.068297748	0.009635347	Shortening
MAVS	0.068509867	-0.319049437	Lengthening
AGO2	0.068829103	0.174000309	Shortening
MRPL47	0.068898329	-0.590345727	Lengthening
POGZ	0.069270278	0.26126591	Shortening
PSMC2	0.070766613	-0.505335827	Lengthening
MAPK1IP1L	0.071369417	-0.36407767	Lengthening
H3F3A	0.072728368	0.192599912	Shortening
IGF2BP1	0.072728368	0.709119497	Shortening
NFE2L1	0.072728368	-0.356057866	Lengthening
COL5A1	0.073497537	0.512922932	Shortening
MRPS18C	0.074013605	-0.647450111	Lengthening
MFN1	0.074196949	-0.075757576	Lengthening
MIEF1	0.07458577	0.227992958	Shortening
JMJD1C	0.075040279	-0.075362319	Lengthening
ZNF557	0.075284987	0.328798186	Shortening
RCAN1	0.075550544	-0.500827939	Lengthening
PDPK1	0.076819884	0.402258292	Shortening
DUSP1	0.077089847	-0.628818826	Lengthening
SRPRA	0.077499233	-0.551164432	Lengthening
USP9X	0.077513788	0.222375667	Shortening
STX3	0.077948068	-0.248177722	Lengthening
NUAK2	0.078416484	-0.203790577	Lengthening
EIF5	0.078797982	-0.57170641	Lengthening
ANXA3	0.07907388	-0.530645161	Lengthening
FBXO46	0.07907388	0.45959596	Shortening
FLCN	0.079238895	-0.43020595	Lengthening

MFSD14A	0.079533399	-0.299662484	Lengthening
ESF1	0.079556708	0.671122995	Shortening
ARHGEF18	0.079657666	-0.562026413	Lengthening
MFSD11	0.079657666	-0.336509146	Lengthening
WDR73	0.079724798	-0.353577626	Lengthening
ZNF160	0.079724798	-0.266666667	Lengthening
SLC15A4	0.079891447	0.374626308	Shortening
RSL1D1	0.079960154	-0.352104635	Lengthening
SRSF2	0.081379376	-0.479453974	Lengthening
AHCYL1	0.082854382	0.144924257	Shortening
TARDBP	0.083188672	-0.282694414	Lengthening
DPYSL3	0.083612039	0.288331308	Shortening
PDPR	0.085631492	0.24885906	Shortening
IVNS1ABP	0.086077227	-0.473126125	Lengthening
ADIPOR2	0.086434121	0.22421574	Shortening
YWHAG	0.08774103	0.039725122	Shortening
P2RY11	0.088439182	-0.340026189	Lengthening
CAMSAP2	0.088819517	-0.401238461	Lengthening
GPATCH8	0.090817591	-0.054659498	Lengthening
FAM129B	0.091982746	-0.045236045	Lengthening
TYW5	0.092325958	-0.396959459	Lengthening
NRDC	0.095038294	-0.200595644	Lengthening
ZFYVE27	0.095878902	-0.355311355	Lengthening
ASDURF	0.097004356	-0.193173821	Lengthening
DDX52	0.097060204	-0.013708514	Lengthening
DHX36	0.09718996	0.269585253	Shortening
DLD	0.097191919	-0.292951103	Lengthening
RAB21	0.097215988	0.101633987	Shortening
NFKBIA	0.097414125	0.443129869	Shortening
SEC23IP	0.097414125	-0.370454545	Lengthening
GET4	0.097484053	0.350717703	Shortening
PNPT1	0.097760029	-0.486833856	Lengthening
IFRD1	0.097788977	-0.391284319	Lengthening
USP7	0.098416082	-0.072298027	Lengthening
FST	0.098755687	0.217142857	Shortening
PI4K2A	0.098847289	0.407062147	Shortening
CYCS	0.098877451	-0.161567831	Lengthening
SLC39A10	0.099055671	0.505457191	Shortening
GYG1	0.099163828	-0.273927393	Lengthening
KAT6A	0.099163828	0.235390947	Shortening
GNAI3	0.099167531	-0.368724133	Lengthening

Table A.1: Table of all genes undergoing significant 3'UTR APA switches (Nuclear, Udorn)

A.2 Full lists of genes experiencing a significant PAS switch during Udorn infection (Cytoplasm)

Gene	P-adjusted	DeltaRA	Direction
ARF3	3.30E-20	0.412364083	Shortening
MRPS16	2.19E-14	0.198032658	Shortening
VKORC1L1	2.36E-13	0.084989376	Shortening
JPT2	3.63E-12	0.494318477	Shortening
RPL37A	2.99E-10	-0.738090389	Lengthening
UBE2D3	3.09E-09	0.377963423	Shortening
PDXK	8.73E-09	0.388504898	Shortening
DESI1	1.86E-08	0.622911465	Shortening
PGAM5	6.62E-08	0.559710334	Shortening
F2R	8.92E-08	-0.078816621	Lengthening
CALM3	3.76E-07	0.58286639	Shortening
POLR2E	3.81E-07	0.094186184	Shortening
CTSB	4.52E-07	0.104904279	Shortening
EXOC7	7.56E-07	0.259066772	Shortening
ORAI3	1.41E-06	-0.628802982	Lengthening
INSIG1	1.57E-06	0.364571949	Shortening
FRMD8	2.34E-06	-0.567745872	Lengthening
NUAK2	2.98E-06	-0.709586283	Lengthening
PPIA	5.56E-06	0.059058037	Shortening
SLC35F6	5.56E-06	0.2233768	Shortening
UQCR11	6.66E-06	0.054103675	Shortening
CHMP1A	9.23E-06	-0.068031832	Lengthening
SESN2	9.25E-06	0.333710338	Shortening
PSME4	2.01E-05	-0.604779412	Lengthening
MED29	4.70E-05	-0.162257496	Lengthening
ANKRD9	4.72E-05	-0.760009736	Lengthening
KIF1C	5.14E-05	0.056267683	Shortening
SERP1	7.17E-05	0.172735418	Shortening
C12orf49	7.37E-05	0.640324519	Shortening
CALM1	8.13E-05	0.162833457	Shortening
TUSC2	8.41E-05	0.334063515	Shortening
RXRA	0.000133072	0.408937145	Shortening
CDC25B	0.000156748	0.060892816	Shortening
NCLN	0.000157287	0.123107307	Shortening
KLHL9	0.000203023	-0.698597268	Lengthening
TOB2	0.000278776	0.443331077	Shortening
CTBP2	0.000325922	-0.018195056	Lengthening
LRRC14	0.000554895	-0.422920892	Lengthening

UCK2	0.000571978	0.528912543	Shortening
WWTR1	0.000649309	0.212724675	Shortening
MAVS	0.00066855	0.532217492	Shortening
ARL4C	0.00067484	0.125454581	Shortening
CDK5R1	0.000680862	0.51287794	Shortening
PSMB2	0.000725578	0.269876976	Shortening
YTHDC1	0.000797131	-0.227996782	Lengthening
EIF4B	0.000941475	-0.232949455	Lengthening
BCAR3	0.00101161	0.418598234	Shortening
EIF3A	0.00101161	-0.424419423	Lengthening
GSK3B	0.001051735	-0.401561543	Lengthening
RFFL	0.001092755	0.475693766	Shortening
CENPA	0.001103991	-0.360003779	Lengthening
YRDC	0.001240793	0.328151476	Shortening
SLC25A37	0.001502556	0.044743507	Shortening
CPSF4	0.001605152	0.183729646	Shortening
SOX4	0.002077487	0.282739871	Shortening
NARS	0.002147799	-0.210045582	Lengthening
MTRNR2L1	0.002681201	0.151692823	Shortening
M6PR	0.002752461	0.257029319	Shortening
SLC7A5	0.002849357	0.035698782	Shortening
PDE4D	0.002951114	-0.528057486	Lengthening
GPI	0.003038459	0.185449941	Shortening
CARHSP1	0.003061089	0.281782822	Shortening
SPAG9	0.003203106	0.498240322	Shortening
RNF10	0.003222138	0.197701914	Shortening
SYNJ2	0.0032244	0.030732861	Shortening
ADAM10	0.003267297	-0.29270298	Lengthening
PSMC4	0.003267297	0.024261259	Shortening
PGRMC1	0.003391531	0.231328851	Shortening
CEBPG	0.003605128	-0.48136695	Lengthening
SEC63	0.003696672	-0.213780074	Lengthening
CHMP3	0.003725907	0.250598281	Shortening
GNL3	0.004097375	-0.6	Lengthening
AP3S2	0.00414093	0.455357143	Shortening
GRAMD1A	0.00471682	-0.000322025	Lengthening
METTL9	0.004744776	0.181691853	Shortening
ARL8B	0.005002884	0.367437614	Shortening
SEL1L3	0.005049329	0.075364922	Shortening
CRTAP	0.005388271	0.017154431	Shortening
CLPB	0.005784143	0.394462719	Shortening
GATC	0.00586262	0.582167832	Shortening
ARL1	0.005937691	0.273525482	Shortening
FAM210B	0.006035783	0.455128205	Shortening
ZDHHC9	0.006036747	-0.167507431	Lengthening
PSENEN	0.006053621	0.091015673	Shortening
ENY2	0.006275581	0.090180103	Shortening
VPS26B	0.006617967	0.242734361	Shortening

DNM2	0.006836708	0.170551132	Shortening
TSKU	0.007127164	-0.062788238	Lengthening
LY6E	0.007182105	0.087463586	Shortening
POLR2J3	0.00724851	0.495543672	Shortening
RRP7A	0.00724851	0.019451389	Shortening
ACOT9	0.007253446	0.2888811	Shortening
PIAS4	0.007398191	-0.398496241	Lengthening
STAT6	0.007825706	0.343650794	Shortening
ACTR1A	0.00785466	0.236478342	Shortening
TMEM120B	0.00815537	0.245908347	Shortening
PSMD11	0.008225025	-0.080920018	Lengthening
PJA2	0.009235909	0.375688705	Shortening
EHD2	0.009491705	-0.230603249	Lengthening
UBE2Q2	0.009630537	0.437859609	Shortening
COL5A1	0.010011335	0.470093235	Shortening
SARM1	0.010011335	-0.457692308	Lengthening
MMP24OS	0.010041577	0.122280688	Shortening
HMGN4	0.010766599	-0.255813953	Lengthening
UBALD2	0.012256592	0.271858946	Shortening
GNPNAT1	0.012700778	0.324565884	Shortening
MXD4	0.012948728	0.682299632	Shortening
UBFD1	0.013092387	0.328018707	Shortening
RAN	0.013432814	0.034109877	Shortening
TMED10	0.014299672	0.184969136	Shortening
MRGBP	0.014301906	0.233110068	Shortening
GALNT1	0.014456621	-0.197734936	Lengthening
GPS1	0.014456621	0.028658986	Shortening
YIF1B	0.01505509	0.31510099	Shortening
RAB10	0.015924286	0.194066042	Shortening
JPT1	0.016257006	0.07758729	Shortening
TMEM199	0.016257006	0.172622002	Shortening
SNX8	0.016381144	0.285085295	Shortening
FBXO9	0.016434299	0.290052356	Shortening
EFHD2	0.016536895	0.35984199	Shortening
STX12	0.016950903	0.37400319	Shortening
EZR	0.018004037	-0.01845238	Lengthening
ISY1	0.018094253	0.362030905	Shortening
FOXC1	0.018962255	0.398838846	Shortening
RFC5	0.019977358	0.243169399	Shortening
SET	0.020051205	0.046741103	Shortening
RNF121	0.020066889	0.262130295	Shortening
TMEM106C	0.02095649	0.249547646	Shortening
PAQR5	0.021079959	0.113632257	Shortening
BLCAP	0.021299803	0.299681575	Shortening
EHD1	0.021472113	0.120186824	Shortening
WIPI2	0.022626794	0.036871218	Shortening
TMEM56	0.022642156	-0.460618847	Lengthening
GCLC	0.023523256	0.258193052	Shortening

PAIP1	0.023983448	0.202240896	Shortening
ZSWIM7	0.024255601	0.076373432	Shortening
CHCHD4	0.024256507	-0.105603391	Lengthening
CTC1	0.02453306	-0.48	Lengthening
NCOA5	0.025752327	0.073707371	Shortening
ERCC2	0.026391925	0.228389155	Shortening
AKR1C2	0.026419387	0.001491289	Shortening
TRAPPC2L	0.026743875	0.043153215	Shortening
NIPSNAP1	0.027096364	-0.144418627	Lengthening
CDT1	0.027432566	0.106153576	Shortening
ATG4B	0.02800126	0.208461654	Shortening
ADD2	0.028228283	0.412123386	Shortening
PLEKHB2	0.030306044	0.220600775	Shortening
BHLHE40	0.030726826	-0.150833215	Lengthening
NEK9	0.030983323	0.318283166	Shortening
WSB2	0.030989713	0.010784983	Shortening
FAM168B	0.031131341	0.261516698	Shortening
MRPS17	0.031239184	0.433207572	Shortening
RNF103-CHMP3	0.031633603	0.250598281	Shortening
NFKBIA	0.032017095	0.60393978	Shortening
RAD23A	0.032519613	-0.090547318	Lengthening
ASRGL1	0.0340292	-0.458275058	Lengthening
IGFLR1	0.034218947	0.228373574	Shortening
RNF149	0.034292027	-0.258196721	Lengthening
MEF2D	0.034807847	0.612104508	Shortening
TXNRD1	0.035425212	0.003286666	Shortening
ACP1	0.035590082	-0.057411156	Lengthening
HNRNPK	0.035590082	0.055064971	Shortening
TMEM38B	0.035686486	-0.394871795	Lengthening
NTMT1	0.035885965	0.050171144	Shortening
USP22	0.035885965	0.02246986	Shortening
TSPAN9	0.036028757	0.193133386	Shortening
CMIP	0.037303726	-0.056649637	Lengthening
RBPMS	0.037325667	0.241687764	Shortening
NDUFA3	0.037927329	0.192010412	Shortening
DNAJB1	0.038514008	0.085766936	Shortening
TMCO1	0.038542163	-0.119814611	Lengthening
KEAP1	0.038763052	-0.127864782	Lengthening
DVL3	0.03991018	0.384626785	Shortening
SMAD3	0.040310905	-0.396743597	Lengthening
DDX46	0.040422401	0.04923231	Shortening
PPP2R2C	0.040959201	0.302380952	Shortening
KCTD1	0.042527058	-0.21242236	Lengthening
BCAT1	0.042574836	0.41212406	Shortening
SURF4	0.043806157	0.150713693	Shortening
ZFP91	0.043806157	-0.217937307	Lengthening
FOXRED2	0.044871858	0.376759248	Shortening
NAA10	0.044999624	0.064590051	Shortening

GMFB	0.045292494	0.515544957	Shortening
UNKL	0.045358464	0.346127103	Shortening
CDC25C	0.045879293	0.491812535	Shortening
PAIP2	0.046028896	0.241504329	Shortening
HSBP1	0.047188041	0.148955864	Shortening
ARPC5L	0.047679769	-0.142737278	Lengthening
HOXB7	0.048507929	-0.217201892	Lengthening
ARSE	0.04860789	0.083615247	Shortening
RBMS1	0.04860789	0.343283582	Shortening
GLS	0.048661243	0.099802372	Shortening
GET4	0.048848243	0.225605875	Shortening
LDAH	0.048848243	0.225641026	Shortening
RPL23	0.04886968	0.044444444	Shortening
ABCE1	0.048884129	0.422991485	Shortening
DUSP8	0.049995347	-0.483739837	Lengthening
C19orf25	0.050072747	0.12693205	Shortening
VMA21	0.050702625	0.315401386	Shortening
LARP4	0.05154466	-0.258054674	Lengthening
ERO1A	0.052406083	0.443644887	Shortening
C6orf120	0.053670256	0.283333333	Shortening
DIAPH1	0.054663604	-0.185870656	Lengthening
CYHR1	0.054760233	-0.155400853	Lengthening
AIMP1	0.05487757	-0.362637363	Lengthening
DCAF8	0.054962143	0.087732919	Shortening
FECH	0.055689482	0.272182254	Shortening
CKAP2L	0.056741927	-0.349655172	Lengthening
PAPPA	0.056741927	-0.368903803	Lengthening
DNPEP	0.056938945	0.039675694	Shortening
ZNF24	0.057153299	-0.185524508	Lengthening
FGF2	0.057404363	-0.068627451	Lengthening
INPP5K	0.057404363	0.30286819	Shortening
PHB	0.058059928	0.156083422	Shortening
FAM120A	0.058539597	-0.12435587	Lengthening
EIF5	0.058707847	0.220795107	Shortening
SMIM20	0.05888849	0.31475469	Shortening
HES4	0.059190066	-0.245756801	Lengthening
COMMD5	0.059604497	-0.302083333	Lengthening
TIMM10B	0.059785466	0.333503242	Shortening
HSPA5	0.061636546	-0.01483228	Lengthening
TRIM8	0.061911518	-0.305107527	Lengthening
ABHD17A	0.062103962	0.280256289	Shortening
DSTN	0.062240258	0.079222362	Shortening
RAB35	0.062713718	0.225143851	Shortening
IER3IP1	0.06372186	0.221761944	Shortening
TSN	0.065093196	-0.169043785	Lengthening
SUPT6H	0.066049513	-0.168000102	Lengthening
UBR7	0.066341126	0.386708861	Shortening
TP53I11	0.066403259	0.238356365	Shortening

BPTF	0.066803245	0.511688312	Shortening
SEC24A	0.068349784	-0.294546607	Lengthening
NQO1	0.068424057	0.084260502	Shortening
ZC3HAV1	0.069940146	0.234065041	Shortening
ZNF609	0.069940146	0.387362637	Shortening
APH1B	0.070485807	0.208333333	Shortening
POLR3H	0.070843547	0.307075732	Shortening
CXXC5	0.071182728	-0.238665759	Lengthening
SLC19A1	0.071485382	0.161875183	Shortening
USP3	0.073505483	0.115278973	Shortening
POLR3E	0.073585944	-0.317829457	Lengthening
DMWD	0.074150375	-0.195896627	Lengthening
PRMT3	0.074187896	0.068762816	Shortening
TCTN3	0.074187896	0.28231003	Shortening
ANXA4	0.074787173	0.146502836	Shortening
WIZ	0.075177333	0.095711297	Shortening
RB1	0.075971324	-0.192749779	Lengthening
PHACTR4	0.076680103	0.45	Shortening
MXD1	0.07895134	-0.332918223	Lengthening
MEGF8	0.079912808	-0.268115942	Lengthening
NSD2	0.079912808	0.322419355	Shortening
ESYT2	0.07994457	0.356459756	Shortening
RAB11A	0.08106249	0.0610339	Shortening
LMAN2	0.082460148	-0.15044016	Lengthening
CLINT1	0.08266203	-0.263780518	Lengthening
COPZ1	0.083684897	0.177956031	Shortening
ALKBH5	0.086219989	-0.138750465	Lengthening
C1RL	0.086282731	0.101214575	Shortening
DPYSL3	0.088342728	0.265013744	Shortening
MRPL19	0.088922143	0.289417386	Shortening
CSNK2A1	0.089908317	-0.392977799	Lengthening
PIGK	0.092310102	0.342476805	Shortening
PTMA	0.093080507	0.094918828	Shortening
PYGB	0.093823139	-0.201705023	Lengthening
SAP18	0.095204717	0.015905434	Shortening
TNPO1	0.095204717	-0.266268281	Lengthening
SLC2A3	0.096683636	0.376901798	Shortening
GLUD1	0.098812105	-0.274488304	Lengthening
ALDH3A2	0.098839451	0.249257816	Shortening
ASAP2	0.098968179	0.215646259	Shortening
HMGB3	0.099479804	0.330759331	Shortening

Table A.2: Table of all genes undergoing significant 3'UTR APA switches (Cytoplasm, Udorn)

A.3 Gene ontology analyses of genes experiencing a 3'UTR PAS switch in in Udorn-infected cells in the "biological process" category (Nuclear)

Term	P-adjusted	Odds Ratio	Genes
mRNA processing (GO:0006397)	4.48E-05	7.426519217	SPEN, TCERG1, RBM14, RBM15, DHX9, NCBP2, HNRNPU, ELAVL1, SRRM1, SYNCRIP, SON, ZC3H11A, SRSF2, SRSF6
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	0.001252756	6.811458333	SPEN, SYNCRIP, RBM14, RBM15, DHX9, NCBP2, SRSF2, HNRNPU, SRSF6, ELAVL1, SRRM1
CRD-mediated mRNA stabilization (GO:0070934)	0.001252756	212.7321429	SYNCRIP, DHX9, HNRNPU
mRNA splicing, via spliceosome (GO:0000398)	0.001252756	6.208491762	SPEN, SYNCRIP, RBM14, RBM15, DHX9, NCBP2, SRSF2, HNRNPU, SRSF6, ELAVL1, SRRM1
Regulation of translation (GO:0006417)	0.001252756	7.824428155	SYNCRIP, EIF5B, EIF5, DHX9, NCBP2, EIF3H, MALSU1, LARP4, ELAVL1
RNA processing (GO:0006396)	0.001252756	7.778007024	IVNS1ABP, TCERG1, SYNCRIP, SON, NCBP2, SRSF2, HNRNPU, TARDBP, SRRM1
RNA splicing (GO:0008380)	0.001252756	11.17986425	IVNS1ABP, TCERG1, SYNCRIP, NCBP2, SRSF2, TARDBP, SRRM1
Cytoplasmic translational initiation (GO:0002183)	0.001463134	38.06618705	EIF3M, EIF2S3, EIF5, DENR
RNA 3'-end processing (GO:0031123)	0.001463134	13.76016684	PNPT1, NCBP2, ZC3H11A, SRSF2, SRSF6, SRRM1
mRNA stabilization (GO:0048255)	0.001999664	18.41137124	SYNCRIP, DHX9, HNRNPU, TARDBP, ELAVL1

Table A.3: Gene ontology analysis of lengthening 3'UTR APA events detected in Udorn-infected cells in the "biological process" category (Nuclear).

Term	P-adjusted	Odds Ratio	Genes
Regulation of mRNA catabolic process (GO:0061013)	3.19E-05	15.44195731	UPF1, PSMD11, SET, YTHDF2, KHSRP, DHX36, IGF2BP1, PUM1, DCP1A
regulation of mRNA stability (GO:0043488)	6.20E-05	12.72133963	PSMD11, SET, YTHDF2, KHSRP, DHX36, E2F1, IGF2BP1, PUM1, DCP1A
Positive regulation of mRNA catabolic process (GO:0061014)	6.20E-05	29.85112782	UPF1, YTHDF2, KHSRP, DHX36, AGO2, PUM1
mRNA destabilization (GO:0061157)	6.40E-04	28.38193253	UPF1, YTHDF2, KHSRP, DHX36, PUM1
Negative regulation of translation (GO:0017148)	0.00266838	13.47278912	YTHDF2, DHX36, AGO2, EIF4EBP1, IGF2BP1, PUM1
3'-UTR-mediated mRNA destabilization (GO:0061158)	0.018075767	42.47008547	UPF1, KHSRP, DHX36
Protein deubiquitination (GO:0016579)	0.019595085	5.886718896	NFKBIA, PSMD11, INO80D, USP9X, SIAH2, MDM2, RAD23A, TOMM20
DNA conformation change (GO:0071103)	0.019595085	36.8037037	KAT6A, HMGB1, TOP1
Protein modification by small protein removal (GO:0070646)	0.020975858	5.686422258	NFKBIA, PSMD11, INO80D, USP9X, SIAH2, MDM2, RAD23A, TOMM20
Positive regulation of protein localization to membrane (GO:1905477)	0.030605573	9.725334119	PDPK1, AKT2, E2F1, MIEF1, YWHAG

Table A.4: Gene ontology analysis of shortening 3'UTR APA events detected in Udorn-infected cells in the "biological process" category (Nuclear).

A.4 Gene ontology analyses of genes experiencing a 3'UTR PAS switch in in Udorn-infected cells in the "biological process" category (Cytoplasm)

Term	P-adjusted	Odds Ratio	Genes
Mitotic chromosome condensation (GO:0007076)	0.074999804	31.88301282	RB1, CSNK2A1, CHMP1A
Negative regulation of DNA-binding transcription factor activity (GO:0043433)	0.074999804	9.430301297	RB1, LRRC14, CEBPG, BHLHE40, KEAP1
Posttranslational protein targeting to membrane, translocation (GO:0031204)	0.074999804	100.8303797	HSPA5, SEC63
Positive regulation of pri-miRNA transcription by RNA polymerase II (GO:1902895)	0.074999804	24.67493797	SMAD3, FGF2, GNL3
ER overload response (GO:0006983)	0.074999804	84.02109705	GSK3B, TMCO1
Cellular protein-containing complex assembly (GO:0034622)	0.074999804	7.973845657	CHCHD4, PSMD11, SMAD3, CENPA, EIF4B
Positive regulation of protein sumoylation (GO:0033235)	0.074999804	63.00949367	PIAS4, GNL3
Intracellular protein transmembrane transport (GO:0065002)	0.074999804	18.64727955	CHCHD4, HSPA5, SEC63
Negative regulation of signal transduction (GO:0009968)	0.074999804	6.025440613	RB1, LRRC14, CSNK2A1, MEGF8, FGF2, TSKU
Chromosome condensation (GO:0030261)	0.074999804	18.20238095	RB1, CSNK2A1, CHMP1A

Table A.5: Gene ontology analysis of lengthening 3'UTR APA events detected in Udorn-infected cells in the "biological process" category (Cytoplasmic).

Term	P-adjusted	Odds Ratio	Genes
Protein stabilization (GO:0050821)	0.045973456	56.32079499	STX12, CRTAP, NCLN, RPL23, DVL3, WIZ, PHB, ZSWIM7, SOX4
Cellular protein localization (GO:0034613)	0.045973456	39.29522008	EHD1, ARF3, ARL4C, STX12, TMED10, ABHD17A, ARL1, PIGK, VPS26B, WIPI2, SNX8, COPZ1
Intracellular protein transport (GO:0006886)	0.045973456	37.62397063	EHD1, ARF3, ARL4C, STX12, TMED10, DESI1, SURF4, ARL1, VPS26B, SNX8, COPZ1, RAN
Protein transport (GO:0015031)	0.082398012	30.88802372	EHD1, ARF3, ARL4C, RAB10, STX12, TMED10, ARL1, CHMP3, VPS26B, SNX8, COPZ1, DNM2
Protein localization to organelle (GO:0033365)	0.118375917	43.31252417	ALDH3A2, EHD1, SERP1, RAB35, ARL1, RAB11A, GET4
Negative regulation of peptidyl-threonine phosphorylation (GO:0010801)	0.118375917	171.5240652	INPP5K, CALM3, CALM1
Transcription-coupled nucleotide-excision repair (GO:0006283)	0.118375917	57.53886095	ISY1, RFC5, GPS1, ERCC2, POLR2E
Histone deubiquitination (GO:0016578)	0.118375917	143.3167277	ENY2, USP22, USP3
Histone modification (GO:0016570)	0.118375917	42.60916835	ENY2, MRGBP, USP3, NTMT1, USP22, PHB
Positive regulation of G2/M transition of mitotic cell cycle (GO:0010971)	0.118375917	131.9744275	CDC25C, RAB11A, CDC25B

Table A.6: Gene ontology analysis of shortening 3'UTR APA events detected in Udorn-infected cells in the "biological process" category (Cytoplasmic).

A.5 Top 100 differentially expressed genes in Udorn-infected cells

Gene	Base Mean	Log2Fold Change	P-adjusted
EGR1	771.9850332	4.08943888	3.26E-13
ATP6V0A4	81.24147355	5.375693846	6.30E-12
MED12L	95.56456443	4.266016419	8.01E-12
SRPK2	395.0071037	2.904927897	4.57E-11
ZNF462	232.4423156	3.532310738	5.31E-11
TEX14	130.1633517	4.445252912	9.97E-11
MROH8	88.31064739	5.014897507	1.05E-10
NCALD	92.94205482	4.297410694	2.65E-10
RYR2	110.017905	4.361441459	2.65E-10
ZFP82	63.71423302	4.348263451	2.65E-10
INTS6	398.6860343	2.87465009	4.41E-10
NUCB1	866.510662	3.585127948	4.51E-10
NUAK2	2245.824187	3.284112193	1.16E-09
CENPC	179.4528207	3.802627365	1.89E-09
FLNB	993.9517238	-3.095909883	1.89E-09
NSUN6	324.8589869	3.228668774	2.05E-09
HDAC9	53.35035929	5.971601922	2.44E-09
LARP4	1168.080496	2.945579593	8.44E-09
KIAA1549	193.1429471	3.505340726	1.84E-08
FYN	116.6258191	3.814003674	3.00E-08
BAG3	375.8193383	2.178440341	1.83E-07
CDADC1	90.69353452	3.624004836	2.08E-07
KMT2C	242.4167899	2.520328995	2.11E-07
ZNF654	115.462645	2.733267569	4.28E-07
ABCC2	155.7131198	-3.066741014	6.55E-07
WAC	1029.971066	2.239083426	1.42E-06
STX3	344.8637602	2.700643177	1.42E-06
DYSF	64.25565544	3.890125755	1.53E-06
KCNK4	81.54188372	3.26685094	1.78E-06
ZNF529	58.53216853	3.896616576	1.90E-06
HIST1H2BF	54.04015287	3.702764299	2.12E-06
CREB5	58.94909476	3.651472034	2.14E-06
MYH9	621.2880505	-2.936831254	2.41E-06
MAP4K4	1175.721986	2.454586849	2.48E-06
HSPB8	129.1536592	2.72542067	2.57E-06
ZFP69B	46.28338074	4.02617665	3.43E-06
C12orf43	139.6265897	3.095351642	3.62E-06
B4GALT5	714.9440344	2.287560112	4.58E-06
DKK1	99.00733907	-2.951770297	4.58E-06
MAP2K3	993.228784	1.814834312	4.58E-06
POU2F2	80.43263059	3.928534466	4.58E-06
SLC33A1	140.6082598	2.415403058	4.58E-06

TRAF1	140.5897983	2.868046416	5.08E-06
FBN2	191.4025571	2.220347586	5.77E-06
PPP2R3C	225.9081433	2.119226584	5.77E-06
NEDD1	431.3038256	2.300360764	6.03E-06
WDR73	306.9062094	2.308118119	6.81E-06
TMEM223	106.1650644	2.990865074	8.41E-06
XPA	261.1598726	2.534931319	8.47E-06
GPX1	110.0442029	2.596337896	9.32E-06
BRPF1	230.1832448	2.139872595	1.06E-05
MED26	93.97241652	2.532382444	1.15E-05
HSPA4L	88.21436291	2.560680539	1.37E-05
INTS2	146.401568	2.328481166	1.37E-05
SLC6A6	101.0875848	-2.736545619	1.38E-05
TSPYL4	145.3631453	3.131433493	1.40E-05
AHNAK	629.1816873	-2.146729661	1.43E-05
NPDC1	109.9473891	2.430090723	1.43E-05
SNX31	50.1033874	4.234275564	1.43E-05
PPM1N	25.97191848	4.793210516	1.51E-05
ACTN4	1254.842952	-2.657498003	1.53E-05
SBF2	236.4571462	2.204391733	1.58E-05
CDKN1C	46.5797513	4.064544852	1.69E-05
GCAT	33.2279057	4.254548361	1.91E-05
ARL2	152.9174823	2.154794407	2.03E-05
POLG2	509.1379965	2.604178992	2.38E-05
FKBP2	101.0913102	2.563539746	2.41E-05
ZNF542P	63.31692411	3.382160333	2.41E-05
ZEB2	147.5573076	2.653717248	2.43E-05
SETD5	2172.772024	1.713850575	2.93E-05
SPTBN1	1081.67467	-2.069279912	4.31E-05
IQGAP1	790.927011	-2.179879872	4.80E-05
RAD23A	205.609626	2.189657521	4.80E-05
ASDURF	58.68159593	2.940834322	5.34E-05
KHSRP	1514.268422	1.829344271	5.34E-05
PKLR	24.21777811	6.375009843	5.42E-05
USP9X	741.8911439	1.934981628	6.12E-05
PAK1	156.9222637	2.132783766	7.32E-05
RPL7A	1280.458125	1.569756623	7.90E-05
UBN2	288.3402985	2.233784551	9.10E-05
JAZF1	56.10035082	2.692248445	0.000115222
DYNC1H1	187.7106777	-2.038328201	0.000123024
CHAC1	50.81951763	-3.849311197	0.000131828
MCRIP1	174.7353302	2.498937615	0.00013651
SPOP	771.5940091	1.63449978	0.000138199
HMGA1	858.19489	1.685283974	0.000157701
HAP1	75.85211577	2.456487462	0.000158845
MB21D2	191.779259	2.19321138	0.000158845
MTF2	136.8070632	2.11646799	0.000172705
TXLNB	13.83179781	8.13098006	0.000183301

ACTB	1366.91415	-1.489469302	0.000184443
CYP24A1	132.8064235	-2.370605975	0.000184443
VEZF1	397.9774403	2.157177473	0.000184443
TNKS1BP1	41.62128	-4.275535164	0.000185357
MARVELD2	114.5684616	2.44571141	0.000206942
SESN2	454.9557	1.882836466	0.000212566
BLNK	19.40594483	6.043067554	0.000214071
CDKL1	67.85275051	2.834139558	0.000214071
CXXC4	29.89296842	3.377207616	0.000232155
MTERF2	180.7012517	2.75688367	0.000234946

Table A.7: Top 100 most significant differentially regulated genes in Udorn-infected cells (Nuclear). "Base Mean" refers to the mean of normalised counts, normalised for sequencing depth.

Gene	Base Mean	Log2Fold Change	P-adjusted
INTS6	1292.049074	6.726262253	2.54E-51
VKORC1L1	1751.468602	4.278537569	1.86E-34
MROH8	4192.454436	4.556758671	3.55E-33
LMO7	1623.334227	4.924263059	5.87E-27
PUS10	111.5698422	6.165642224	2.36E-19
CHAC1	473.9092956	-5.063270268	4.08E-18
KLF4	381.4417827	4.523180508	1.03E-17
ISG15	189.8462835	4.878021207	1.63E-17
PHYH	331.8049998	3.467944412	4.77E-16
CENPC	3399.014992	3.521587551	1.17E-15
HILPDA	391.3986145	-4.260954472	5.00E-15
POLE2	635.723869	4.103491941	5.00E-15
OAZ1	4876.280102	2.678153857	6.54E-15
IL6	94.35995228	4.952437186	9.44E-15
HIST1H2BI	181.5684371	4.809528517	1.25E-14
FXYD2	775.9061426	2.832384118	1.49E-14
FXYD6-FXYD2	777.9006186	2.809065733	1.81E-14
CYB5D2	732.7302964	3.783139237	3.81E-14
STAG3L4	199.4045118	3.926300509	1.00E-13
CAPNS1	9237.249351	2.653885704	1.33E-13
EFNA1	375.7205932	-4.246793666	1.63E-13
GPX4	797.0585122	3.474473122	1.88E-13
RCAN3	290.4306225	3.960762464	3.17E-13
DGCR6L	472.1501183	2.843492603	3.77E-13
ADM5	89.98158147	4.267129418	8.37E-13
TEX14	240.1321393	4.171222125	9.08E-13
MIF	10452.08525	2.648094672	1.00E-12
SERF2	11268.66164	2.468469717	1.00E-12
COX5B	2603.510608	2.616051594	1.45E-12
NR4A1	338.242463	-2.903992899	3.19E-12
POLR2E	3380.462947	2.372679985	3.52E-12

INO80E	477.0684745	2.600076221	6.75E-12
MFSD2B	1143.724025	2.585683908	1.14E-11
POLR2J	597.6980593	2.528704526	1.18E-11
TEN1	223.4177792	3.122265873	1.18E-11
SH3BGRL3	930.6650933	2.84700025	1.21E-11
ADAMTS6	630.5208578	2.987356643	1.98E-11
GRIA4	58.31823095	5.523134961	2.97E-11
HMGA1	14740.75363	2.580819919	3.07E-11
DKK1	233.3421244	-3.296766008	3.08E-11
GCN1	196.865238	-3.21953097	4.25E-11
ELF3	414.1201009	-3.070022391	4.33E-11
MAPK3	516.6089338	2.531452417	4.33E-11
ECSIT	1096.132655	2.324653184	4.56E-11
EMD	6852.159081	2.426495041	6.05E-11
NDUFA13	947.2751447	2.392753951	6.39E-11
RPS14P3	10092.24897	2.446537708	6.62E-11
SNORC	255.7102998	2.704350006	9.49E-11
DNLZ	326.8227346	2.804273393	1.08E-10
USF2	1409.011132	2.343585217	1.08E-10
ARL6IP4	540.1929579	2.393960447	1.09E-10
CA5BP1	120.9356066	3.446351288	1.26E-10
DUSP5	148.7485473	-3.435113612	1.54E-10
HM13	727.2353634	2.318023547	1.58E-10
MALSU1	181.7617157	-3.373087412	1.74E-10
PLK2	135.5131341	-3.641690504	2.78E-10
SNRPB	21805.58034	2.284309661	3.21E-10
CRTC2	213.8841582	3.197414473	3.27E-10
OST4	1236.113338	2.223765137	3.31E-10
RPL18	11020.62139	2.162596627	3.31E-10
F2R	1044.018467	3.005619607	3.71E-10
LARP6	1098.858642	2.320157169	4.41E-10
UNC5C	51.35916807	6.078170722	4.41E-10
HPS6	131.5541063	-4.044916187	4.78E-10
PAM16	300.7544691	2.844029773	4.78E-10
ATP6V0B	3325.195624	2.218553149	5.79E-10
AKT1	712.9843235	2.390669038	6.50E-10
RPS28	1794.780673	2.266594102	7.33E-10
BANF1	405.9375316	2.682400593	7.40E-10
RABAC1	386.263274	2.587431313	7.57E-10
LAMTOR4	1097.320169	2.277180513	8.07E-10
HEXA	949.7334921	2.204854721	8.42E-10
UQCR11	1404.772708	2.285680572	9.30E-10
RBPMS	719.1373296	2.339311529	9.83E-10
C19orf33	2341.169131	2.278496648	9.93E-10
HSF1	1260.858376	2.477284779	9.98E-10
TRAPPC3L	259.8420017	3.14279473	1.08E-09
ROMO1	4327.662342	2.128486182	1.21E-09
SMOX	1055.634518	2.505142124	1.24E-09

KRT10	785.492124	2.556427085	1.41E-09
TAGLN2	12219.11091	2.511940298	1.51E-09
RPS2	16076.07382	2.223386793	1.56E-09
PPDPF	5295.739783	2.385078781	1.57E-09
CYSTM1	2125.431363	2.403521038	1.85E-09
COX8A	29964.3577	2.007300331	2.02E-09
TCEA2	220.6429354	2.619769235	2.02E-09
EEF1A2	478.2425841	2.309610554	2.30E-09
MZT2B	1148.34172	2.156264762	2.32E-09
C17orf49	783.1053915	2.22701748	2.73E-09
KRCC1	153.2536653	3.554453349	2.81E-09
ZFAND5	461.4743515	-2.448882247	3.10E-09
LAGE3	1604.535429	2.304354605	4.26E-09
PFDN2	573.4475118	2.22826828	6.82E-09
TNFRSF1A	469.1511029	-3.206983339	8.07E-09
AMOTL2	207.327406	-2.956647761	8.53E-09
TSKU	536.7283688	-2.608231849	8.53E-09
GAPDH	20415.74487	2.21077195	9.86E-09
NT5C3B	388.6408745	2.377269297	9.86E-09
CRIP1	2963.683981	2.118217539	1.00E-08
APRT	1522.403304	2.07828426	1.08E-08

Table A.8: Top 100 most significant differentially regulated genes in Udorn-infected cells (Cytoplasmic). "Base Mean" refers to the mean of normalised counts, normalised for sequencing depth.

A.6 Gene ontology analyses of differentially expressed genes in Udorn-infected cells

Term	P-adjusted	Odds Ratio	Genes
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	1.65E-16	12.63457609	SRP19, RPL10, RPLP0, RPL10A, RPL7A, RPS14, RPS16, RPS19, RPL36, RPLP2, RPL15, RPS11, RPS27A, RPL39, RPS9, RPL41, RPS8, RPS5, RPL22, RPL13A, RPL23A, RPS28, RPL27A, RPL37A, RPS20, RPL29, RPL28, UBA52, RPS21
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	1.65E-16	10.53204897	SMG1, RPL10, RPLP0, RPL10A, RPS14, CTIF, RPL7A, RPS16, RPS19, RPL36, RPLP2, RPL15, RPS11, RPS27A, RPL39, RPS9, RPL41, RPS8, RPS5, RPL22, RPL13A, RPL23A, RPS28, RPL27A, RPL37A, RPS20, EIF3E, RPL29, DCP1A, RPL28, UBA52, RPS21
Cotranslational protein targeting to membrane (GO:0006613)	4.20E-16	11.85459195	SRP19, RPL10, RPLP0, RPL10A, RPL7A, RPS14, RPS16, RPS19, RPL36, RPLP2, RPL15, RPS11, RPS27A, RPL39, RPS9, RPL41, RPS8, RPS5, RPL22, RPL13A, RPL23A, RPS28, RPL27A, RPL37A, RPS20, RPL29, RPL28, UBA52, RPS21
Protein targeting to ER (GO:0045047)	4.80E-16	10.93002489	SRP19, RPL10, ZFAND2A, RPLP0, RPL10A, RPS14, RPL7A, RPS16, RPS19, RPL36, RPLP2, RPL15, RPS11, RPS27A, RPL39, RPS9, RPL41, RPS8, RPS5, RPL22, RPL13A, RPL23A, RPS28, RPL27A, RPL37A, RPS20, RPL29, RPL28, UBA52, RPS21
Cytoplasmic translation (GO:0002181)	2.15E-15	11.42998191	RPL10, RPLP0, RPL10A, RPL7A, RPS14, RPS16, RPS19, RPL36, RPL15, RPS11, RPS27A, RPL39, RPS9, RPL41, RPS8, RPS5, RPL22, RPL13A, RPL23A, RPS28, EIF3M, RPL27A, RPL37A, RPS20, RPL29, RPL28, UBA52, RPS21
Nuclear-transcribed mRNA catabolic process (GO:0000956)	1.15E-14	7.128764569	SMG1, RPL10, RPLP0, RPL10A, RPS14, CTIF, RPL7A, RPS16, RPS19, POLR2D, RPL36, RPLP2, RPL15, RPS11, RPS27A, RPL39, RPS9, RPL41, RPS8, RPS5, RPL22, RPL13A, RPL23A, RPS28, CNOT7, RPL27A, CSDE1, RPL37A, RPS20, EIF3E, RPL29, DCP1A, RPL28, UBA52, RPS21, DCP1B
Cellular macromolecule biosynthetic process (GO:0034645)	2.06E-11	4.374645084	SET, RNASEH2A, RPL10, RPLP0, MRPL37, RPL10A, EEF1B2, RPS14, RPL7A, RPS16, RPS19, POLD2, POLR2D, RPL36, E2F1, EPC1, RPLP2, RPL15, RPS11, RPS27A, RPL39, POLR2J, ST3GAL2, POLR2L, RPL41, RPS9, RPS8, RPS5, RPL22, RPL13A, RPL23A, EEF1G, RPS28, EEF1D, RPL27A, POLE2, RPL37A, RPS6KB2, RPS20, TOP1, RPL29, RPL28, RPS21, B4GALT5
Peptide biosynthetic process (GO:0043043)	8.79E-10	5.773061463	RPL10, RPLP0, MRPL37, RPL10A, RPS14, RPL7A, RPS16, RPS19, RPL36, RPLP2, RPL15, RPS11, RPS27A, RPL39, RPS9, RPL41, RPS8, RPS5, RPL22, RPL13A, RPL23A, RPS28, RPL27A, RPL37A, RPS6KB2, RPS20, RPL29, RPL28, RPS21
Translation (GO:0006412)	1.76E-09	4.841902769	RPL10, RPLP0, MRPL37, RPL10A, EEF1B2, RPS14, RPL7A, RPS16, RPS19, RPL36, RPLP2, RPL15, RPS11, RPS27A, RPL39, RPL41, RPS9, RPS8, RPS5, RPL22, RPL13A, RPL23A, EEF1G, RPS28, EEF1D, RPL27A, RPL37A, RPS6KB2, RPS20, RPL29, RPL28, UBA52, RPS21
Gene expression (GO:0010467)	6.05E-08	3.471021686	SMG1, RPL10, YTHDC1, RPLP0, U2AF1, MRPL37, RPL10A, RPS14, RPL7A, RPS16, RPS19, PABPN1, POLR2D, RPL36, E2F1, EPC1, RPLP2, RPL15, RPS11, RPS27A, RPL39, POLR2J, SRSF11, POLR2L, RPL41, RPS9, RPS8, RPS5, RPL22, RPL13A, RPL23A, RPS28, RPL27A, RPL37A, RPS6KB2, SRSF3, RPS20, HNRNPH3, RPL29, RPL28, RPS21

Table A.9: Gene ontology analysis of significantly upregulated genes in Udorn-infected cells in the "biological process" category (Nuclear).

Term	P-adjusted	Odds Ratio	Genes
tRNA export from nucleus (GO:0006409)	1.13E-04	17.92193563	NUP93, SEC13, NUP107, NUP188, NUP43, RAE1, NOL6, NUP37
tRNA-containing ribonucleoprotein complex export from nucleus (GO:0071431)	1.13E-04	17.92193563	NUP93, SEC13, NUP107, NUP188, NUP43, RAE1, NOL6, NUP37
tRNA transport (GO:0051031)	1.43E-04	16.06548654	NUP93, SEC13, NUP107, NUP188, NUP43, RAE1, NOL6, NUP37
ncRNA export from nucleus (GO:0097064)	1.43E-04	15.52917903	NUP93, SEC13, NUP107, NUP188, NUP43, RAE1, NOL6, NUP37
wound healing (GO:0042060)	1.65E-04	9.748756219	PPP3CA, MACF1, SMAD3, DST, AHNAK2, MYH9, HMOX1, NRG1, TSKU, PLEC
Regulation of ATP metabolic process (GO:1903578)	1.92E-04	11.17477204	NUP93, SEC13, NUP107, NUP188, SHMT2, NUP43, RAE1, OGT, NUP37
Regulation of carbohydrate catabolic process (GO:0043470)	1.94E-04	13.30733362	NUP93, SEC13, NUP107, NUP188, NUP43, RAE1, OGT, NUP37
Regulation of purine nucleotide metabolic process (GO:1900542)	1.94E-04	13.30733362	NUP93, SEC13, NUP107, NUP188, NUP43, RAE1, OGT, NUP37
Protein localization to plasma membrane (GO:0072659)	1.94E-04	6.217944951	C2CD5, JUP, SEC16A, DENND4C, RDX, TTC7A, TNFRSF1A, IKBKB, TSPAN14, MAP7, FLNA, NECTIN3, SPTBN1
Regulation of cellular response to stress (GO:0080135)	2.22E-04	6.645928948	NUP93, SEC13, CEBPB, NUP107, NUP188, RPA2, NUP43, SLC38A2, RAE1, MTOR, HSPA1A, NUP37

Table A.10: Gene ontology analysis of significantly downregulated genes in Udorn-infected cells in the "biological process" category (Nuclear).

Term	P-adjusted	Odds Ratio	Genes
Cytoplasmic translation (GO:0002181)	1.13E-41	23.39158163	RPL30, RPL32, RPL34, RPLP1, RPLP0, RPL10A, RPL8, RPL9, RPS15, RPS14, RPS16, RPL18A, RPS19, RPS18, RPL36, RPL35, RPL38, RPL37, RPS11, RPL39, RPS10, RPS13, RPS12, RPS9, RPL21, RPS8, RPS5, RPL22, RPL13A, RPSA, FTSJ1, RPL37A, RPL27, RPL29, RPL28, UBA52, RPL10, RPL12, RPL11, RPS15A, RPS3, RPL13, RPL15, RPS2, RPL18, RPL19, RWDD1, EIF2B3, RPL41, TMA7, RPL35A, RPL23A, RPS28, RPS27, RPS29, RPL27A, DRG1, DRG2, RPS20, RPS21, RPL26L1, RPS23
Protein targeting to ER (GO:0045047)	1.13E-41	20.03665714	RPL30, RPL32, RPL34, ZFAND2A, RPLP1, ZFAND2B, RPLP0, SRP14, RPL8, RPL10A, RPL9, RPS15, RPS14, RPS16, RPS19, RPL18A, RPS18, SEC61G, RPL36, RPL35, RPLP2, RPL38, RPL37, RPS11, RPL39, RPS10, RPS13, RPS12, RPS9, RPL21, RPS8, RPS5, RPL22, RPL13A, RPSA, RPL37A, CHMP4B, RPL27, CHMP4A, RPL29, RPL28, UBA52, RPL10, RPL12, RPL11, RPS4Y1, RPS15A, RPS3, RPL13, RPL15, RPS2, RPL18, RPL19, RPL41, SGTA, RPL35A, RPL23A, RPS28, RPS27, SPCS1, RPS29, RPL27A, RPS20, RPS21, RPS23

Translation (GO:0006412)	8.71E-41	8.947195634	RPL30, MRPS16, RPL32, RPL34, MRPS11, RPLP1, RPLP0, MRPL37, RPL8, RPL10A, RPL9, RPS15, MRPL41, RPS14, TRMT112, MRPL2, RPS16, RPS19, RPL18A, RPS18, RPL36, RPL35, RPLP2, RPL38, RPL37, RPS11, RPL39, RPS10, RPS13, RPS12, RPS9, RPL21, MRPS24, RPS8, RPS5, RPL22, DHPS, MRPS2, MRPS21, RPL13A, RPSA, MRPS6, FTSJ1, MRPS5, MRPL43, TUFM, EEF1G, MRPL51, PPA1, EEF1D, EEF1A2, RPL37A, RPL27, RPL29, RPL28, UBA52, RPL10, RPL12, MRPS33, RPL11, MRPS34, RPS4Y1, MRPL57, MRPL11, RPS15A, RPS3, RPL13, RPL15, RPS2, RPL18, RPL19, EEFSEC, RWDD1, RPL41, NDUFA7, TMA7, MRPL28, RPL35A, RPL23A, MRPL23, EEF2, MRPL24, RPS28, RPS27, RPS29, RPL27A, DRG1, DRG2, RPS20, RPS21, RPL26L1, RPS23
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	1.05E-40	23.36305732	RPL30, RPL32, RPL34, RPLP1, ZFAND2B, RPLP0, SRP14, RPL10A, RPL8, RPL9, RPS15, RPS14, RPS16, RPL18A, RPS19, RPS18, RPL36, RPL35, RPLP2, RPL38, SEC61B, RPL37, RPS11, RPL39, RPS10, RPS13, RPS12, RPS9, RPL21, RPS8, RPS5, RPL22, RPL13A, RPSA, RPL37A, RPL27, RPL29, RPL28, UBA52, RPL10, RPL12, RPL11, RPS4Y1, RPS15A, RPS3, RPL13, RPL15, RPS2, RPL18, RPL19, RPL41, RPL35A, RPL23A, RPS28, RPS27, RPS29, RPL27A, RPS20, RPS21, RPS23
cotranslational protein targeting to membrane (GO:0006613)	8.02E-38	19.67382013	TRAM1, RPL30, RPL32, RPL34, RPLP1, RPLP0, SRP14, RPL10A, RPL8, RPL9, RPS15, RPS14, RPS16, RPL18A, RPS19, RPS18, RPL36, RPL35, RPLP2, RPL38, RPL37, RPS11, RPL39, RPS10, RPS13, RPS12, RPS9, RPL21, RPS8, RPS5, RPL22, RPL13A, RPSA, RPL37A, RPL27, RPL29, RPL28, UBA52, RPL10, RPL12, RPL11, RPS4Y1, RPS15A, RPS3, RPL13, RPL15, RPS2, RPL18, RPL19, RPL41, RPL35A, RPL23A, RPS28, RPS27, RPS29, RPL27A, RPS20, RPS21, RPS23
Peptide biosynthetic process (GO:0043043)	6.67E-35	9.880141388	RPL30, MRPS16, RPL32, RPL34, MRPS11, RPLP1, RPLP0, MRPL37, RPL8, RPL10A, RPL9, RPS15, MRPL41, GGTLC2, RPS14, RPS16, RPS19, RPL18A, RPS18, RPL36, RPL35, RPLP2, RPL38, RPL37, RPS11, GGT1, RPL39, RPS10, RPS13, RPS12, RPS9, RPL21, RPS8, RPS5, RPL22, DHPS, RPL13A, MRPS21, RPSA, MRPS6, MRPS5, MRPL43, MRPL51, EEF1A2, RPL37A, RPL27, RPL29, RPL28, RPL10, RPL12, MRPS33, RPL11, RPS4Y1, MRPL11, RPS15A, RPS3, RPL13, RPL15, RPS2, RPL18, RPL19, RPL41, MRPL28, RPL35A, RPL23A, MRPL23, MRPL24, RPS28, RPS27, RPS29, RPL27A, RPS20, RPS21, RPS23
Aerobic electron transport chain (GO:0019646)	9.43E-35	27.08053974	COX7B, NDUFA13, NDUFA11, NDUFA12, COX4I1, NDUFA10, COX6A1, COX7C, COX8A, NDUFC2, NDUFC1, SDHC, COX6B1, COX7A2L, NDUFS8, NDUFS7, NDUFS6, NDUFS5, UQCRC1, NDUFB9, NDUFB8, NDUFB7, NDUFB6, NDUFB10, UQCRB, NDUFB11, NDUFB5, NDUFB4, NDUFB3, NDUFB2, NDUFB1, UQCR11, UQCR10, COX5B, COX5A, UQCRH, NDUFV1, NDUFA9, NDUFA8, NDUFA7, NDUFA6, UQCC3, NDUFA3, NDUFA2, NDUFA1, COX6C, UQCRHL, UQCRQ, CYCS
Mitochondrial ATP synthesis coupled electron transport (GO:0042775)	2.46E-34	25.84819734	COX7B, NDUFA13, NDUFA11, NDUFA12, COX4I1, NDUFA10, COX6A1, COX7C, COX8A, NDUFC2, NDUFC1, SDHC, COX6B1, COX7A2L, NDUFS8, NDUFS7, NDUFS6, NDUFS5, UQCRC1, NDUFB9, NDUFB8, NDUFB7, NDUFB6, NDUFB10, UQCRB, NDUFB11, NDUFB5, NDUFB4, NDUFB3, NDUFB2, NDUFB1, UQCR11, UQCR10, COX5B, COX5A, UQCRH, NDUFV1, NDUFA9, NDUFA8, NDUFA7, NDUFA6, UQCC3, NDUFA3, NDUFA2, NDUFA1, COX6C, UQCRHL, UQCRQ, CYCS
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	5.05E-34	13.695617	RPL30, RPL32, RPL34, RPLP1, RPLP0, SMG9, RPL8, RPL10A, RPL9, RPS15, RPS14, RPS16, RPS19, RPL18A, RPS18, RPL36, RPL35, RPLP2, RPL38, RPL37, RPS11, RPL39, RPS10, RPS13, RPS12, RPS9, RPL21, RPS8, RPS5, RPL22, RPL13A, RPSA, RPL37A, RPL27, PABPC1, RPL29, RPL28, UBA52, RPL10, RPL12, RPL11, RPS4Y1, HELZ2, RPS15A, RPS3, RPL13, RPL15, RPS2, RPL18, RPL19, RPL41, RPL35A, UPF3A, RPL23A, RPS28, RPS27, RPS29, RPL27A, RPS20, RPS21, RPS23

Cellular macromolecule biosynthetic process (GO:0034645)	5.76E-28	5.129460235	RPL30, MRPS16, RPL32, RPL34, MRPS11, RPLP1, RPLP0, MRPL37, RPL8, RPL10A, RPL9, PTMS, RPS15, MRPL41, RPS14, RPS16, RPS19, RPL18A, RPS18, PCLAF, RPL36, RPL35, RPLP2, RPL38, RPL37, RPS11, RPL39, RPS10, RPS13, RPS12, RPS9, RPL21, RPS8, RPS5, RPL22, DHPS, MRPS21, RPL13A, RPSA, MRPS6, MRPS5, MRPL43, TUFM, EEF1G, MRPL51, EEF1D, EEF1A2, RPL37A, RPL27, RPL29, RPL28, RPL10, RPL12, MRPS33, RPL11, RPS4Y1, MRPL11, RPS15A, POLR2C, ST3GAL4, RPS3, E2F1, POLR2E, RPL13, POLR2F, POLR2G, APOE, RPL15, RPS2, POLR2H, RPL18, POLR2I, POLR2J, POLR2K, RPL19, POLR2L, EEFSEC, RPL41, DUT, MRPL28, RPL35A, RPL23A, MRPL23, EEF2, MRPL24, RPS28, RPS27, RPS29, RPL27A, POLE2, DRG1, RPS20, RPS21, RPS23, NFKBIB

Table A.11: Gene ontology analysis of significantly upregulated genes in Udorn-infected cells in the "biological process" category (Cytoplasmic).

Term	**P-adjusted**	**Odds Ratio**	**Genes**
Positive regulation of transcription, DNA-templated (GO:0045893)	7.34E-06	1.828126584	CCNT1, PRDM4, ICE1, MAML1, FHOD1, ETS2, HOXA10, ZMIZ1, KAT8, HOXA1, KAT7, PIM2, IER5, RNF111, MEN1, ACVR1, ZNF281, DYRK1A, WDR75, HNF1B, MED4, HCFC1, AR, TBL1XR1, KAT6A, KMT2D, DLX1, ZNF395, CDCA4, SHC1, PRKDC, FOXO1, DHX33, DEAF1, NUP85, HIVEP1, ABL1, FADD, TFAP2A, ZNF263, TFAP2C, JUND, SMAD3, FZD5, NFATC3, NR2F2, INO80, PBX1, FOSL2, SMAD7, CDK9, NR4A2, CDK8, NR4A1, MAVS, MMS19, CSRNP1, ELL, HMGB2, PKD2, WDR43, IRF2BPL, PAGR1, PPP3CA, HHEX, MED13, LBH, SIN3A, UBTF, TEAD2, NKX3-1, UTP15, RBM14, EDN1, NCOA3, ARNT, RGMB, PPRC1, ARID1A, TGFBR1, TNFRSF1A, METTL23, DNAJC2, RBL1, ELF3, ELF4, TRAF6, ARHGEF2, CEBPB, KDM3A, MYRF, HEATR1, RNF4, EGFR, LRP6, CAMKK2, SCAF8, NFIL3, SERTAD1, SERTAD2, HSF2, ZNF746, E2F2, HES1, VHL, BRD8, TRIM44, MAP3K2, MLXIP, ARNT2, NFYA, USP22, FOXJ3, DAB2IP, TBX3, MTOR, HIPK2, PER1, KLF7, TFAP4, KLF5, NFIA, NFIC, KANSL3, F2RL1, CEBPZ, NAA15, ERCC6, LPIN2, LPIN3, FOXA2
Transcription by RNA polymerase II (GO:0006366)	7.34E-06	2.725371747	GMEB2, INTS14, ELL, CCNT1, ICE1, MAML1, PHAX, TRIM24, TEAD2, ABT1, KLF13, TSC22D1, NCBP1, MED4, MED26, MED7, NCOR2, AR, TBL1XR1, SLBP, ZNF473, DDX5, CTDP1, RNMT, CSTF2T, RTF1, RPAP1, GTF2E1, TAF5L, IWS1, HDAC7, SCAF8, MNT, POLR2A, DEAF1, TAF1C, POLR2D, LEO1, E2F2, CPSF4, RNGTT, NFYA, MBD1, NR2F6, NR4A2, CDK9, NR4A1, NFIC, INTS5, CDK12, ERCC6
Regulation of cellular response to heat (GO:1900034)	7.34E-06	5.48783724	NUP214, NDC1, HSPA8, HSP90AA1, HSP90AB1, NUP210, NUP133, NUP155, RPA1, RPA2, MTOR, NUP93, DNAJC2, DNAJB1, BAG2, NUP85, TPR, NUP35, NUP43, NUP88, IER5, HSPA1A
Regulation of cellular response to stress (GO:0080135)	9.59E-06	4.227111845	DYRK3, CEBPB, HSP90AB1, DNAJB1, NUAK1, BAG2, NUP85, TPR, ABL1, NUP88, NUP43, IER3, MCL1, NUP214, NDC1, HSPA8, HSP90AA1, NUP210, NUP155, NUP133, RPA1, RPA2, MTOR, NUP93, DNAJC2, NUP35, HSPA1A
Protein import (GO:0017038)	9.59E-06	4.956317765	NUP214, ATAD1, HSP90AA1, ABCD3, NUP133, NUP155, PPP1R10, HSPA4, SNX33, IPO7, IPO4, PEX26, NUP93, IPO8, IPO9, NUP85, PEX5, TPR, FAM53B, HEATR3, NUP88, TNPO1, KPNB1

Cellular protein modification process (GO:0006464)	1.25E-05	1.845616144	GMFB, CCNT1, CSF1, PRAG1, MAML1, FBXO21, LOXL2, GOLGA2, PSMD7, PPP2R1B, PSMD5, SIK1B, PIM1, PDK4, PIM3, MAP3K9, PIM2, MEN1, ACVR1, WSB1, DYRK1A, ALG1, HASPIN, CDC25B, CCNE1, TSSK6, ULK1, SIK1, SIK2, TRIB1, ALG10B, MTMR2, RPN2, PRKDC, KLHL11, MTMR4, FBXO41, TGOLN2, UGCG, NUAK1, IRAK1, HLCS, STK36, STC2, ABL1, MAP3K20, NAE1, SOCS5, TGFB2, ICMT, KLHL25, TIPARP, WFS1, PLK2, PMM2, SSH1, CLK2, CDK9, CDK8, CDK6, GGCX, AAK1, TAB1, CDK12, FGFR1, PTPRU, DYRK3, DYRK2, PIK3CD, CHM, LAMC1, NAGPA, PTPRF, DUSP11, PPP3CA, DUSP12, DUSP4, DUSP5, BRD2, DUSP1, PDPK1, DUSP6, CKAP4, TGFBR1, DUSP7, CTTNBP2NL, PTP4A2, PEAK1, FAM20B, B3GNT3, FAM20C, B3GNT2, BIRC6, PRKAA1, CTDP1, TULP4, ACVR1B, THBS1, PPM1F, AURKB, CAMKK2, CAND1, PDF, GCNT2, CCL2, FEM1C, VHL, MAPK6, PTPN1, FEM1B, FBXL18, LAMB1, MTOR, HIPK2, KBTBD6
Regulation of cell cycle (GO:0051726)	1.36E-05	2.707840709	ANKRD17, PPP1R15A, HSP90AB1, XIAP, TXLNG, PKD2, THBS1, RRP8, LRP6, TBRG4, CCNB2, CCNB1, CCND1, MAP3K20, ABL1, MFN2, PIM3, PIM2, WHAMM, CDK5RAP3, MEN1, NKX3-1, TGFB2, JUND, SMAD3, DUSP1, DAB2IP, DHCR24, CDC25A, TGFBR1, TBX3, HCFC1, RHOB, HIPK2, CDK9, NR4A1, CENPF, TFAP4, CDK6, MYBBP1A, MDM2, SIK1, NF2, TARDBP, PRR11, TRIM32, BIRC3
Ribosome biogenesis (GO:0042254)	2.29E-05	3.184320201	DDX27, DDX47, POP4, NIP7, HEATR1, NOL9, MALSU1, PWP2, WDR43, MRM2, NOL6, EXOSC6, EXOSC9, BMS1, RRS1, SURF6, HEATR3, RRP12, EXOSC3, UTP15, NOP56, NOM1, NOP14, WDR36, NOP16, IMP3, UTP3, WDR74, WDR75, URB2, DDX52, WDR55, MRTO4, NAF1, NOL11
Protein phosphorylation (GO:0006468)	2.43E-05	2.213023256	DYRK3, GMFB, DYRK2, CCNT1, PRAG1, MAML1, PIK3CD, IGF1R, EEF2K, SIK1B, PIM1, PDK4, PIM3, MAP3K9, PIM2, EPHB4, EPHB3, ACVR1, BRD2, PDPK1, CAD, DYRK1A, HASPIN, TGFBR1, CIT, CDC25B, INCENP, CCNE1, PEAK1, TYRO3, ULK3, FAM20B, TSSK6, FAM20C, BIRC6, ULK1, SIK1, SIK2, TRIB1, PRKAA1, PRKDC, PRKX, ACVR1B, EGFR, AURKB, CAMKK2, NUAK1, IRAK1, STK36, ABL1, MAP3K20, CCL2, MAPK6, TGFB2, PLK2, DAB2IP, MTOR, HIPK2, SMAD7, CLK2, CDK9, CDK8, CDK6, AAK1, CDK12, FGFR1
Transcription, DNA-templated (GO:0006351)	2.49E-05	2.968772584	DDX5, GMEB2, CTDP1, ELL, CCNT1, RNMT, RTF1, RPAP1, GTF2E1, TAF5L, IWS1, MNT, TAF1B, POLR2A, DEAF1, TAF1C, POLR2D, TRIM24, LEO1, UBTF, ABT1, IVNS1ABP, BRF2, KLF13, RNGTT, TSC22D1, NFYA, NCBP1, MBD1, TWNK, MED4, NR4A2, CDK9, POLR3A, NFIC, POLR1C, MMS19, ERCC6

Table A.12: Gene ontology analysis of significantly downregulated genes in Udorn-infected cells in the "biological process" category (Cytoplasmic).

A.7 Top 50 most significant genes undergoing 3'UTR APA switches (Nucleus, WSN)

Gene	P-adjusted	DeltaRA	Direction

MYO1E	2.01E-17	-0.092501756	Lengthening
SYNCRIP	5.93E-16	-0.507887254	Lengthening
LAMA5	5.57E-15	-0.219562247	Lengthening
RPL27A	3.11E-14	-0.01579519	Lengthening
SLC52A2	5.61E-14	-0.272022948	Lengthening
PRAG1	1.56E-13	-0.195164228	Lengthening
FOSL2	9.60E-12	0.388860717	Shortening
HUWE1	7.90E-11	-0.163097222	Lengthening
F2R	8.08E-11	-0.715479494	Lengthening
ALYREF	2.49E-10	-0.231591387	Lengthening
ZNF207	8.14E-10	-0.082173175	Lengthening
DCTN4	4.56E-09	-0.527348658	Lengthening
ETAA1	4.58E-09	-0.379746835	Lengthening
TRIM11	5.05E-09	-0.132894113	Lengthening
PSPC1	9.23E-09	-0.352095808	Lengthening
FDFT1	1.50E-08	-0.212523182	Lengthening
TMEM120A	1.54E-08	-0.258168668	Lengthening
IVNS1ABP	2.05E-08	-0.307562045	Lengthening
ZCCHC8	2.12E-08	-0.294988979	Lengthening
PAAF1	3.06E-08	-0.502923977	Lengthening
RRP36	7.58E-08	-0.345350692	Lengthening
IER5	1.00E-07	-0.234248993	Lengthening
ACACA	1.53E-07	0.11622276	Shortening
GPN2	1.82E-07	-0.502513058	Lengthening
SMC1A	1.82E-07	-0.163187727	Lengthening
PPIA	2.25E-07	-0.238209081	Lengthening
CREBZF	3.94E-07	-0.483562108	Lengthening
CD63	4.41E-07	-0.017729771	Lengthening
CYP4F11	5.89E-07	-0.49866007	Lengthening
MPST	5.89E-07	-0.191903225	Lengthening
PCYT2	6.45E-07	-0.699874214	Lengthening
KNTC1	7.77E-07	-0.469021694	Lengthening
MLH3	1.12E-06	-0.680451312	Lengthening
ZWINT	1.24E-06	-0.063500326	Lengthening
ITGA2	1.26E-06	-0.361917968	Lengthening
DGKE	1.39E-06	0.362300485	Shortening
TRPM7	3.94E-06	-0.131617647	Lengthening
ADAM10	5.15E-06	-0.659975093	Lengthening
SMG9	5.64E-06	-0.558997253	Lengthening
CCND3	6.63E-06	-0.066486576	Lengthening
S100A16	7.83E-06	-0.089498716	Lengthening
TFAP2A	8.36E-06	0.37274347	Shortening
KNOP1	9.69E-06	-0.196680915	Lengthening
DTNB	1.03E-05	-0.500505107	Lengthening
LUC7L3	1.16E-05	-0.528107621	Lengthening
EIF4A3	1.16E-05	-0.040018015	Lengthening
EIF3M	1.24E-05	-0.312877264	Lengthening
ITGA3	1.31E-05	-0.299228667	Lengthening

ELAVL1	1.35E-05	-0.390103567	Lengthening
MVP	1.36E-05	-0.054133513	Lengthening

Table A.13: Table of the top 50 most significant genes undergoing 3'UTR APA switches (Nuclear, WSN, 6hpi)

Gene	P-adjusted	DeltaRA	Direction
LAMA5	2.27E-27	-0.473747201	Lengthening
FOSL2	4.11E-27	0.429971989	Shortening
FDFT1	8.71E-24	-0.362333978	Lengthening
EIF3M	1.83E-16	-0.597294532	Lengthening
HILPDA	6.11E-16	-0.221029495	Lengthening
F2R	8.90E-16	-0.779284484	Lengthening
AJUBA	2.37E-15	-0.317688137	Lengthening
DCTN4	6.86E-15	-0.657436578	Lengthening
DDX5	1.71E-14	-0.333803413	Lengthening
CD63	1.72E-14	-0.041442621	Lengthening
DYNC1H1	2.93E-14	-0.601222651	Lengthening
HUWE1	1.96E-12	-0.323406501	Lengthening
PRAG1	2.63E-12	-0.156101606	Lengthening
NMD3	3.37E-12	0.36440678	Shortening
SRSF6	3.76E-12	-0.870292386	Lengthening
ZCCHC8	4.17E-12	-0.388527419	Lengthening
KNTC1	1.03E-11	-0.728342652	Lengthening
ITGA3	1.05E-11	-0.297276998	Lengthening
NCOA3	1.29E-11	-0.7821884	Lengthening
SLC52A2	1.42E-11	-0.224268689	Lengthening
MYO1E	4.33E-11	-0.174217419	Lengthening
CREBZF	7.33E-11	-0.528535653	Lengthening
IER5	1.04E-10	-0.239763526	Lengthening
TRPM7	1.15E-10	-0.364379748	Lengthening
ZNF207	1.89E-10	-0.397087549	Lengthening
PTCD3	2.61E-10	-0.438449322	Lengthening
ITGA2	3.46E-10	-0.708898944	Lengthening
DNAJC2	6.95E-10	-0.645989975	Lengthening
IVNS1ABP	9.03E-10	-0.305006554	Lengthening
SYNCRIP	3.28E-09	-0.054545455	Lengthening
RAB11A	1.20E-08	0.281196581	Shortening
SMG9	1.28E-08	-0.855809833	Lengthening
PPIA	1.42E-08	-0.372699872	Lengthening
FARSA	1.50E-08	-0.227594665	Lengthening
CCDC127	2.24E-08	-0.713733075	Lengthening
AIMP2	2.45E-08	-0.011698593	Lengthening
MFSD11	6.60E-08	-0.640839387	Lengthening
ETAA1	7.05E-08	-0.384126984	Lengthening
RBM14	7.29E-08	-0.525787327	Lengthening
DRG1	7.45E-08	-0.349517585	Lengthening

ATP6V1B2	7.75E-08	-0.286383399	Lengthening
CDC42SE2	8.46E-08	-0.509132948	Lengthening
RRP36	9.01E-08	-0.375066454	Lengthening
NFIB	1.31E-07	-0.488490414	Lengthening
ZNF281	1.46E-07	-0.008725248	Lengthening
TRMU	1.64E-07	-0.804773562	Lengthening
BMS1	1.91E-07	-0.065537893	Lengthening
ZNF398	2.34E-07	-0.173736206	Lengthening
SLC25A13	2.39E-07	0.031603774	Shortening
POLR1A	2.55E-07	-0.577978884	Lengthening

Table A.14: Table of the top 50 most significant genes undergoing 3'UTR APA switches (Nuclear, WSN, 9hpi)

Gene	P-adjusted	DeltaRA	Direction
DYNC1H1	2.84E-74	-0.693726551	Lengthening
LAMA5	5.23E-51	-0.480542066	Lengthening
NUAK2	2.97E-34	-0.538849952	Lengthening
LUC7L3	6.14E-31	0.108910205	Shortening
EIF3M	3.83E-30	-0.575992679	Lengthening
AJUBA	4.70E-30	-0.334343409	Lengthening
HILPDA	1.44E-28	-0.103836187	Lengthening
F2R	3.53E-28	-0.814799105	Lengthening
HSPA5	6.37E-27	-0.034328814	Lengthening
NMD3	2.90E-26	0.413978495	Shortening
SMG9	6.92E-25	-0.661579064	Lengthening
RBM14	3.00E-24	-0.614935855	Lengthening
SUPT6H	1.16E-22	-0.854005784	Lengthening
DCTN4	2.33E-22	-0.664807585	Lengthening
ITPRIP	3.33E-21	-0.577944862	Lengthening
MDM2	3.49E-21	-0.689378588	Lengthening
SF3B3	3.66E-21	-0.282938076	Lengthening
XPO5	5.71E-21	-0.098617512	Lengthening
TP53BP1	8.21E-21	-0.573994734	Lengthening
FDFT1	2.82E-20	-0.221445347	Lengthening
STX3	3.29E-20	-0.74266055	Lengthening
DDX5	2.81E-19	-0.475502254	Lengthening
PDCD11	2.83E-19	0.046699438	Shortening
IVNS1ABP	9.76E-19	-0.345740111	Lengthening
PRAG1	2.44E-18	-0.165541626	Lengthening
ZNF281	1.05E-17	0.417968446	Shortening
RNH1	1.69E-17	-0.126127206	Lengthening
DNAJC2	2.12E-17	-0.694779983	Lengthening
PTCD3	4.51E-17	-0.443278455	Lengthening
ZNF207	4.77E-17	-0.454530947	Lengthening
PPAN	1.57E-16	-0.887192287	Lengthening
CD63	2.40E-16	-0.020943406	Lengthening

ESF1	3.30E-16	-0.509653371	Lengthening
KNTC1	1.01E-15	-0.720766894	Lengthening
LARP4	1.96E-15	-0.289960751	Lengthening
SERPINB9	2.51E-15	-0.16025641	Lengthening
SINHCAF	2.90E-15	0.498251748	Shortening
PSPC1	4.31E-15	-0.387654321	Lengthening
EIF4A3	7.57E-15	-0.219243795	Lengthening
RFFL	2.14E-14	0.547825737	Shortening
KRIT1	2.16E-14	0.623210634	Shortening
WDR43	3.06E-14	-0.45876312	Lengthening
TRPM7	4.62E-14	-0.163083164	Lengthening
BNIP2	7.76E-14	-0.370112233	Lengthening
PPDPF	9.22E-14	-0.129357493	Lengthening
NOM1	1.09E-13	-0.282286599	Lengthening
DHX37	1.10E-13	-0.744443408	Lengthening
YLPM1	3.23E-13	-0.513931644	Lengthening
MYO1E	4.77E-13	-0.142020836	Lengthening
MFSD11	6.70E-13	-0.628571429	Lengthening

Table A.15: Table of the top 50 most significant genes undergoing 3'UTR APA switches (Nuclear, WSN, 12hpi)

Gene	P-adjusted	DeltaRA	Direction
NUAK2	2.61E-38	-0.562403671	Lengthening
DYNC1H1	5.91E-31	-0.488524239	Lengthening
LAMA5	1.69E-29	-0.523853515	Lengthening
FOSL2	1.92E-24	0.42227451	Shortening
VCP	5.28E-24	-0.513997207	Lengthening
SERPINB9	2.24E-23	-0.12254902	Lengthening
SUPT6H	3.96E-23	-0.8972203	Lengthening
SMG9	2.74E-20	-0.817270595	Lengthening
HSPA5	3.78E-20	-0.075399451	Lengthening
HILPDA	4.58E-19	-0.266992054	Lengthening
F2R	7.63E-19	-0.849505929	Lengthening
ZNF207	1.23E-18	-0.593424606	Lengthening
ATP6V1B2	4.24E-17	-0.585177855	Lengthening
RBM14	7.51E-17	-0.781708328	Lengthening
MYO1E	1.22E-15	-0.277377076	Lengthening
MDM2	2.30E-15	-0.590639093	Lengthening
TOMM70	2.10E-14	-0.528690041	Lengthening
CD63	5.27E-14	-0.035370606	Lengthening
PSMA3	4.27E-13	-0.290879813	Lengthening
DHX37	4.93E-13	-0.772185797	Lengthening
MVP	5.31E-13	-0.164031621	Lengthening
RNF34	7.30E-13	-0.601906151	Lengthening
DCTN4	1.02E-12	-0.716190984	Lengthening
EIF4A3	1.16E-12	-0.33990847	Lengthening

STX3	1.41E-12	-0.743722944	Lengthening
NCOA3	1.71E-12	-0.8077988	Lengthening
EIF3M	2.12E-12	-0.62906774	Lengthening
MON2	4.42E-12	-0.679792321	Lengthening
SLC25A13	4.42E-12	0.077656405	Shortening
PLEC	1.88E-11	-0.184220876	Lengthening
TRPM7	1.90E-11	-0.484363439	Lengthening
ESF1	3.69E-11	-0.528857409	Lengthening
FRMD8	6.06E-11	-0.377665838	Lengthening
NOP58	1.88E-10	-0.484601774	Lengthening
CSNK2A1	3.25E-10	-0.166579178	Lengthening
DRG1	7.38E-10	-0.424501425	Lengthening
C14orf119	9.06E-10	-0.352325795	Lengthening
KRIT1	9.47E-10	0.683816695	Shortening
MFSD11	1.12E-09	-0.65	Lengthening
PUM2	1.13E-09	-0.008221994	Lengthening
ZCCHC8	2.02E-09	-0.329443066	Lengthening
YLPM1	2.16E-09	-0.768756329	Lengthening
DKK1	2.34E-09	-0.305385794	Lengthening
DERL2	3.00E-09	-0.449970868	Lengthening
FDFT1	3.92E-09	-0.162242719	Lengthening
GNAI3	4.12E-09	-0.746019865	Lengthening
HUWE1	6.65E-09	-0.316779995	Lengthening
NOM1	7.78E-09	-0.283577871	Lengthening
SRSF6	8.97E-09	-0.855309415	Lengthening
LARP4	9.62E-09	-0.276198246	Lengthening

Table A.16: Table of the top 50 most significant genes undergoing 3'UTR APA switches (Nuclear, WSN, 18hpi)

A.8 Top 50 most significant genes undergoing 3'UTR APA switches (Cytoplasm, WSN)

Gene	P-adjusted	DeltaRA	Direction
RNF14	4.32E-25	0.617579921	Shortening
SMAD3	5.16E-24	-0.307311359	Lengthening
SOX12	1.35E-12	-0.073219577	Lengthening
DHX37	4.06E-08	-0.203667064	Lengthening
CBX2	0.000617432	-0.139491841	Lengthening
ZC3HAV1	0.002760362	-0.065270936	Lengthening
CEBPG	0.003458881	-0.503462584	Lengthening
ACACA	0.003546127	0.089811366	Shortening

DYNC1H1	0.005830303	-0.083494686	Lengthening
PPAN	0.010653125	-0.161575013	Lengthening
GTF3C4	0.011499741	-0.394797253	Lengthening
PLIN2	0.01239492	-0.038794053	Lengthening
LRPAP1	0.012827054	0.049470363	Shortening
MASTL	0.013311951	0.215686275	Shortening
AP5B1	0.016554979	-0.200716497	Lengthening
TNFRSF10A	0.01766156	0.555175714	Shortening
CORO1C	0.019507944	0.192618492	Shortening
DDX51	0.021718452	0.236297102	Shortening
CGGBP1	0.026765752	-0.349685176	Lengthening
TPRKB	0.02933688	0.28609826	Shortening
KCTD1	0.039646399	-0.188078725	Lengthening
JMJD4	0.047662834	0.32548068	Shortening
TK1	0.049091741	0.014192308	Shortening
YWHAZ	0.051219885	-0.222210133	Lengthening
ATMIN	0.051428076	-0.052709455	Lengthening
MRPS25	0.070437456	0.072968338	Shortening
TRIM13	0.072497277	-0.293576882	Lengthening
COMMD5	0.086051413	-0.102286606	Lengthening
C8orf88	0.086208729	0.510513036	Shortening
TOMM70	0.09198181	-0.153034357	Lengthening
CRTC1	0.09853269	-0.542857143	Lengthening

Table A.17: Table of the most significant genes undergoing 3'UTR APA switches (Cytoplasmic, WSN, 6hpi)

Gene	P-adjusted	DeltaRA	Direction
DHX37	9.91E-23	-0.630408427	Lengthening
CBX2	3.70E-14	-0.348089445	Lengthening
PPAN	2.03E-09	-0.753228935	Lengthening
MCL1	3.73E-08	-0.543901275	Lengthening
JMJD4	6.53E-08	0.548733618	Shortening
SUPT6H	9.49E-08	-0.143296272	Lengthening
ARMC7	1.17E-07	-0.532608696	Lengthening
ATP9A	1.89E-07	0.090064713	Shortening
HSPA5	3.44E-07	-0.009255097	Lengthening
SMAD3	1.94E-06	-0.21373726	Lengthening
RNF14	2.13E-06	0.616155419	Shortening
PCNA	7.87E-06	-0.179597976	Lengthening
C2CD2L	3.87E-05	-0.489053254	Lengthening
FAM102A	5.04E-05	-0.100280298	Lengthening
H3F3B	5.04E-05	-0.125228336	Lengthening
GPAT4	0.000114352	-0.235704561	Lengthening
RRP7A	0.000176489	0.060131433	Shortening
YKT6	0.000236416	0.153330431	Shortening
MXD4	0.000504575	0.453439153	Shortening

PTP4A1	0.000915194	-0.066334305	Lengthening
CTC1	0.001152677	-0.545957619	Lengthening
TOR1A	0.001357314	-0.248663102	Lengthening
PSKH1	0.00136525	-0.29905014	Lengthening
CHMP1A	0.001735495	-0.49811463	Lengthening
TRMT61A	0.002199719	0.472277424	Shortening
PLEC	0.002430207	-0.50770472	Lengthening
DESI1	0.002650959	0.283916746	Shortening
ZC3HAV1	0.003471353	-0.035551207	Lengthening
HEXIM1	0.00370239	-0.373091179	Lengthening
SLC35B4	0.003910004	-0.504951003	Lengthening
UBE2D3	0.004410716	0.304029304	Shortening
CSNK2A1	0.004553073	-0.482629139	Lengthening
PCLAF	0.005596152	0.075633171	Shortening
ARF3	0.005829115	-0.037971185	Lengthening
F2R	0.00756962	-0.022432862	Lengthening
COLGALT1	0.010030302	-0.013419006	Lengthening
ZNF24	0.010202918	-0.689427972	Lengthening
CCNYL1	0.010615194	0.670311342	Shortening
SERPINB9	0.01099251	0.096555218	Shortening
MAK16	0.012289874	-0.720833333	Lengthening
SPOP	0.012463229	-0.239594502	Lengthening
APPBP2	0.01389995	0.204545455	Shortening
ARMC6	0.014378227	-0.052902448	Lengthening
ANKRD9	0.014904885	-0.29530009	Lengthening
ELL	0.015050315	0.248035915	Shortening
LAMA5	0.018410565	-0.321334923	Lengthening
ALG2	0.018590393	-0.548051948	Lengthening
YTHDF3	0.018838184	-0.018856963	Lengthening
LGR4	0.023264756	-0.380459057	Lengthening
KCTD1	0.033085783	-0.258468336	Lengthening

Table A.18: Table of the top 50 most significant genes undergoing 3'UTR APA switches (Cytoplasmic, WSN, 9hpi)

Gene	P-adjusted	DeltaRA	Direction
DHX37	9.79E-51	-0.743406965	Lengthening
HSPA5	3.61E-35	-0.022461226	Lengthening
SOX12	1.33E-29	-0.551725765	Lengthening
B4GALT5	3.33E-18	-0.177771245	Lengthening
SUPT6H	1.53E-16	-0.140720801	Lengthening
GRB2	7.03E-16	0.500698992	Shortening
ARMC7	4.00E-15	-0.283333333	Lengthening
PLEC	6.71E-15	-0.637752414	Lengthening
TMEM167A	2.64E-12	0.678061395	Shortening
CRLF1	2.18E-11	-0.684308049	Lengthening
AP2B1	1.63E-10	0.560796169	Shortening

SRXN1	3.33E-10	-0.040090593	Lengthening
CTC1	7.33E-10	-0.805935749	Lengthening
ANKRD9	9.48E-10	-0.609792844	Lengthening
PCNA	2.67E-09	-0.313037192	Lengthening
CCT5	4.03E-09	-0.052900271	Lengthening
TXNDC15	4.63E-09	0.491181658	Shortening
FRMD8	1.29E-08	-0.383064345	Lengthening
DCAF8	1.47E-08	0.2741	Shortening
PGAM5	2.02E-08	0.509039909	Shortening
GPRC5B	2.60E-08	-0.227661314	Lengthening
PIM1	4.64E-08	-0.623863636	Lengthening
LARP4	6.85E-08	-0.343604108	Lengthening
AP1S1	1.27E-07	-0.149934874	Lengthening
H3F3B	1.50E-07	-0.153717292	Lengthening
SERPINB9	1.50E-07	-0.032882012	Lengthening
ZC3HAV1	1.80E-07	-0.0263091	Lengthening
WDR33	2.49E-07	-0.597020195	Lengthening
PLIN2	5.67E-07	-0.068480136	Lengthening
NDUFAF4	6.95E-07	-0.577123502	Lengthening
CDK9	9.04E-07	-0.516409266	Lengthening
TRMT61A	1.10E-06	0.519777424	Shortening
DKK1	1.80E-06	-0.340210225	Lengthening
TSKU	2.01E-06	-0.128082649	Lengthening
MAK16	2.20E-06	-0.731536697	Lengthening
CHMP1A	2.61E-06	-0.552114394	Lengthening
PLXND1	4.49E-06	0.745310245	Shortening
CEBPG	4.63E-06	-0.572900851	Lengthening
C2CD2L	4.75E-06	-0.263648724	Lengthening
RPL7L1	5.04E-06	-0.162362758	Lengthening
OSER1	6.55E-06	-0.621106337	Lengthening
DENND1A	7.22E-06	0.049451626	Shortening
CCNYL1	7.39E-06	0.742578548	Shortening
MCL1	1.08E-05	-0.666301532	Lengthening
SLC35B4	1.23E-05	-0.389203862	Lengthening
TP53BP1	1.26E-05	-0.351027397	Lengthening
EMC8	1.58E-05	0.19691112	Shortening
MAPK9	1.70E-05	0.370043614	Shortening
DYNC1H1	2.05E-05	-0.434265239	Lengthening
CSDE1	2.56E-05	-0.189925148	Lengthening

Table A.19: Table of the top 50 most significant genes undergoing 3'UTR APA switches (Cytoplasmic, WSN, 12hpi)

Gene	P-adjusted	DeltaRA	Direction
DHX37	5.92E-70	-0.788108185	Lengthening
SOX12	1.42E-42	-0.618361366	Lengthening
PLEC	1.47E-40	-0.759134982	Lengthening

B4GALT5	1.63E-29	-0.279338814	Lengthening
SUPT6H	2.18E-28	-0.147291538	Lengthening
ATP9A	1.02E-27	0.041886099	Shortening
ZSWIM7	8.06E-27	0.474243703	Shortening
TSKU	1.04E-26	-0.116180369	Lengthening
SERPINB9	1.74E-26	-0.223791103	Lengthening
ANKRD9	3.73E-26	-0.623518334	Lengthening
HSPA5	2.49E-25	-0.023806349	Lengthening
PPAN	2.79E-25	-0.840185457	Lengthening
COX7C	5.51E-24	0.254123966	Shortening
GPAT4	7.14E-23	-0.468463182	Lengthening
CTC1	4.40E-21	-0.892133677	Lengthening
TRAF4	1.44E-19	0.550059654	Shortening
NLN	1.48E-19	-0.065864662	Lengthening
AGR2	7.77E-19	0.35898764	Shortening
MEF2D	1.63E-17	-0.65819398	Lengthening
DKK1	5.46E-17	-0.346048424	Lengthening
FRMD8	1.29E-16	-0.597610629	Lengthening
ACACA	1.76E-16	0.060928183	Shortening
MCL1	1.24E-15	-0.678566523	Lengthening
NAP1L1	2.67E-15	-0.13433939	Lengthening
RPL13	1.26E-14	0.014907481	Shortening
PLIN2	2.30E-14	-0.086655532	Lengthening
MXD4	4.98E-14	0.539495798	Shortening
CBX2	5.02E-14	-0.138839047	Lengthening
INSIG1	3.77E-13	-0.55365223	Lengthening
CSDE1	3.97E-13	-0.271706776	Lengthening
HEXIM1	5.49E-13	-0.441341904	Lengthening
DCAF8	1.11E-12	0.145810526	Shortening
JMJD4	1.17E-12	0.550094162	Shortening
HNRNPA0	1.71E-12	-0.202777907	Lengthening
HCCS	3.84E-12	-0.401660312	Lengthening
SLC7A5	3.84E-12	0.009415575	Shortening
CRLF1	4.52E-12	-0.835698448	Lengthening
MDM2	5.34E-12	-0.804713805	Lengthening
ATMIN	1.86E-11	-0.132592447	Lengthening
DYNC1H1	2.09E-11	-0.444338854	Lengthening
EEF1A1	4.73E-11	0.021740726	Shortening
GRB2	1.53E-10	0.414839791	Shortening
CHMP1A	2.78E-10	-0.635540423	Lengthening
C2CD2L	1.25E-09	-0.350164366	Lengthening
TUBB6	1.37E-09	0.195945946	Shortening
DYNLL2	1.51E-09	-0.247520048	Lengthening
PCLAF	3.01E-09	0.10706854	Shortening
SSR1	3.72E-09	0.20028828	Shortening
MAP1LC3B	5.78E-09	0.072352422	Shortening
SRSF3	9.17E-09	0.204581693	Shortening

Table A.20: Table of the top 50 most significant genes undergoing 3'UTR APA switches (Cytoplasmic, WSN, 18hpi)

A.9 Overlaps of genes regulated by APA in the Udorn and WSN datasets

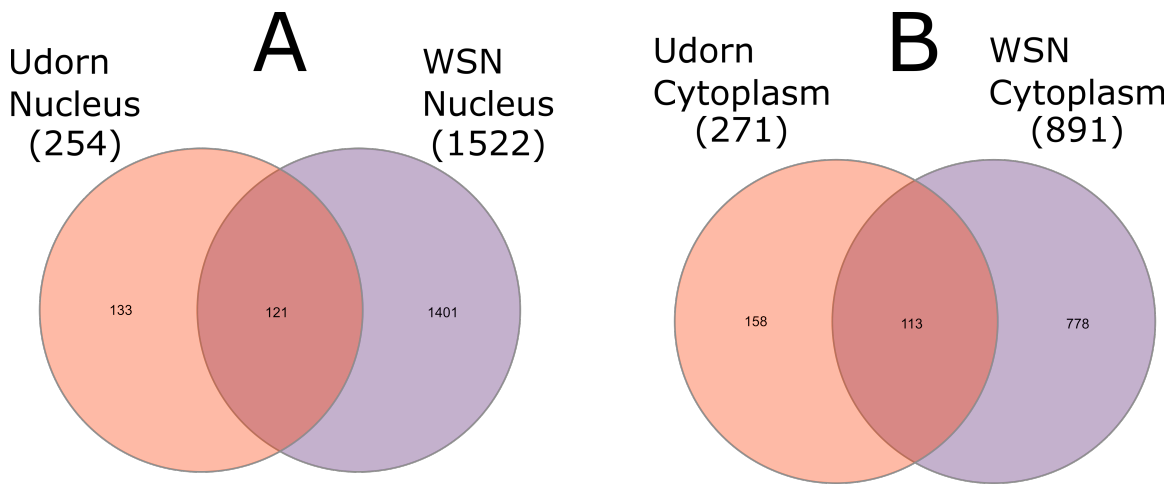


Figure A.1: Overlaps of genes experiencing a 3'UTR PAS switch in the Udorn and WSN datasets. Venn diagrams depicting the numbers of genes regulated by 3'UTR APA in A) the nuclei samples and B) the cytoplasmic samples of the Udorn and WSN datasets. For WSN, genes that switch 3'UTR PASs at any timepoint are included.

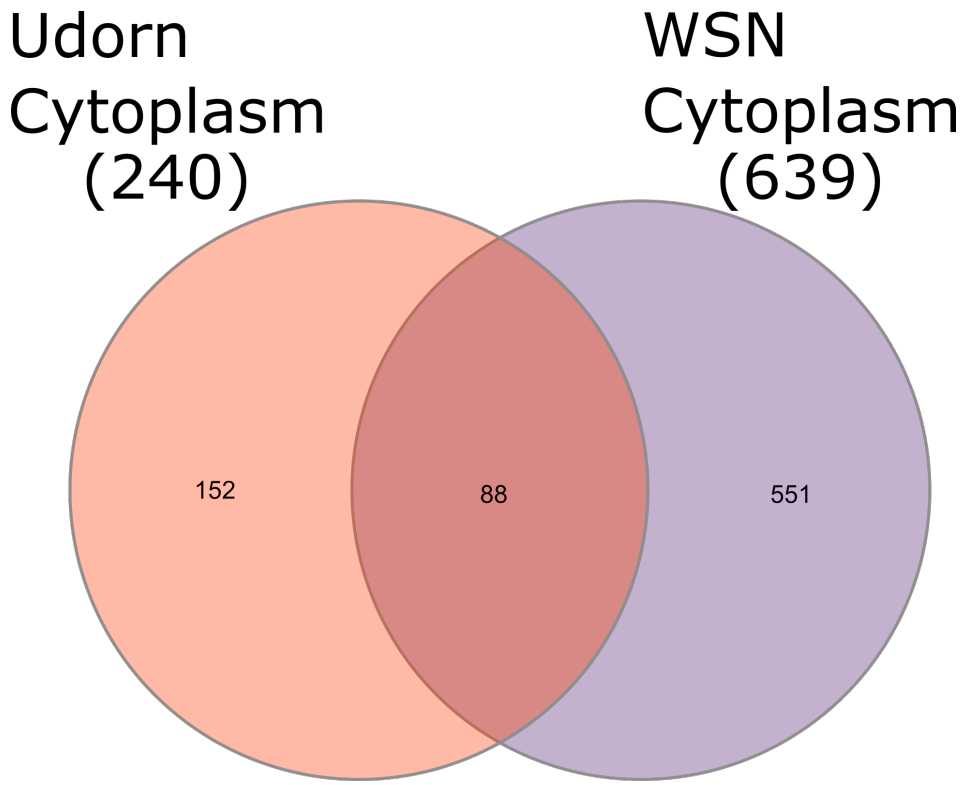


Figure A.2: Overlaps of genes regulated by CR-UTR APA in the Udorn and WSN datasets. Venn diagrams depicting the numbers of genes regulated by CR-UTR APA in the cytoplasmic samples of the Udorn and WSN datasets. For WSN, genes switching between coding-region PASs and 3'UTR PASs at any timepoint are included.

A.10 Gene ontology analyses of genes experiencing a 3'UTR PAS switch in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Nuclear)

Term	P-adjusted	Odds Ratio	Genes
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	3.08E-09	4.974849256	ZCCHC8, DDX5, SF3B3, RBM8A, DDX46, EIF4A3, SRSF1, ELAVL1, SYNCRIP, SART1, SART3, POLR2D, POLR2E, SF3B1, SRSF10, POLR2L, SF3A1, RBM14, ALYREF, RBM19, PRPF40B, CDC5L, TFIP11, PLRG1, CRNKL1, U2SURP, WDR33, PRPF4, HNRNPK, PSPC1, SRSF3, SRSF6, SRSF7, SRSF9
mRNA splicing, via spliceosome (GO:0000398)	4.01E-09	4.652135837	ZCCHC8, DDX5, SF3B3, RBM8A, DDX46, EIF4A3, SRSF1, ELAVL1, SYNCRIP, SART1, SART3, POLR2D, POLR2E, SF3B1, SRSF10, POLR2L, SF3A1, RBM14, ALYREF, RBM19, PRPF40B, CDC5L, TFIP11, PLRG1, CRNKL1, U2SURP, WDR33, PRPF4, HNRNPK, PSPC1, GEMIN4, SRSF3, SRSF6, SRSF7, SRSF9
mRNA processing (GO:0006397)	3.60E-08	4.189995612	ZCCHC8, TCERG1, DDX5, SF3B3, RBM8A, DDX46, EIF4A3, SRSF1, ELAVL1, SYNCRIP, SART1, SART3, POLR2D, POLR2E, SF3B1, SRSF10, POLR2L, SF3A1, RBM14, ALYREF, RBM19, PRPF40B, CDC5L, TFIP11, PLRG1, CRNKL1, U2SURP, WDR33, PRPF4, HNRNPK, PSPC1, SRSF3, SRSF6, SRSF7, SRSF9
Ribosome biogenesis (GO:0042254)	7.14E-08	5.149985097	DDX47, POP4, WDR3, MALSU1, RPF2, WDR43, BMS1, DHX37, RRS1, RPL14, RRP36, RPL37, EXOSC1, NOM1, NOP58, UTP6, KRR1, WDR18, DDX52, BYSL, TBL3, RPS28, RPL27A, ESF1, NAF1, UBA52, MPHOSPH6
rRNA processing (GO:0006364)	1.57E-07	5.303557234	DDX47, POP4, WDR3, WDR43, SART1, BMS1, DHX37, RPL14, RRP36, RPL37, EXOSC1, NOP58, UTP6, KRR1, WDR18, DDX52, BYSL, RSL1D1, TBL3, RPS28, NSA2, RPL27A, ESF1, UBA52, MPHOSPH6
Ribonucleoprotein complex assembly (GO:0022618)	1.04E-06	5.705928854	MCTS1, MRPS11, SRSF1, PPAN, RPF2, MRPS7, CRNKL1, SART1, RPS28, EIF5, SART3, POLR2D, GEMIN4, RRS1, LUC7L3, DENR, NAF1, SRSF6, SF3B1, SRSF10, SRSF9
Gene expression (GO:0010467)	1.62E-06	3.448887923	SMG1, NUP107, MRPS16, RBM8A, MRPS11, EIF4A3, SRSF1, GSPT1, HSPD1, DUSP11, SYNCRIP, POLR2D, RPL14, POLR2E, EIF4EBP2, RPL37, POLR2L, MEF2A, RANBP2, ALYREF, MRPL27, TFIP11, MRPS7, WDR33, PRPF4, RPS28, HNRNPK, RPL27A, DRG1, SRSF3, RPS20, SRSF6, SRSF7, FARSB, SRSF9
ncRNA processing (GO:0034470)	2.36E-06	4.45331068	ZCCHC8, POP1, DDX47, POP4, WDR3, WDR43, BMS1, DHX37, RPL14, RRP36, RPL37, EXOSC1, UTP6, NOP58, KRR1, WDR18, DDX52, BYSL, TBL3, RPS28, RPL27A, INTS4, ESF1, UBA52, MPHOSPH6
rRNA metabolic process (GO:0016072)	3.36E-06	4.911800232	NOP58, UTP6, DDX47, KRR1, POP4, WDR3, WDR18, DDX52, WDR43, BYSL, TBL3, RPS28, RPL27A, BMS1, DHX37, RPL14, RRP36, ESF1, RPL37, UBA52, EXOSC1, MPHOSPH6
Protein polyubiquitination (GO:0000209)	2.80E-05	3.320242708	PSMD12, RNF34, NEDD4L, FBXO22, DDA1, FBXL20, NPEPPS, RNF216, UBB, UBE2Q1, TNKS2, PSMB1, ARIH2, FBXW8, HUWE1, UBE2G1, TRAF2, FBXO11, KLHL42, PSMA5, PSMA3, PSMC4, UBE2S, PSMC2, RLIM, UBE2V2, MDM2, TRIP12, UBA52, THOP1

Table A.21: Gene ontology analysis of genes lengthened by 3'UTR APA in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Nuclear). 271

Term	P-adjusted	Odds Ratio	Genes
Positive regulation of transcription, DNA-templated (GO:0045893)	0.02646269	2.964937931	TFAP2A, JUN, FOXC1, CCNT2, FZD2, PRDM4, RNF14, DDX58, ZNF281, TCF12, SMAD5, FOSL2, CAMKK2, SETX, NFKBIA, ING5, AGO2, ASXL2, EPC1, EP300, ZNF268, TEAD2, BCL9L
Negative regulation of programmed cell death (GO:0043069)	0.060314181	4.227697771	TFAP2A, SETX, NFKBIA, AVEN, DSTYK, AMIGO2, UBE2O, CFLAR, TM7SF3, ZNF268, IFIT3
Positive regulation of transcription by RNA polymerase II (GO:0045944)	0.060314181	2.950008643	TFAP2A, JUN, FOXC1, CCNT2, PRDM4, DDX58, TCF12, SMAD5, FOSL2, SETX, NFKBIA, AGO2, ASXL2, EPC1, EP300, CDK12, ZNF268, BCL9L
Positive regulation of nucleic acid-templated transcription (GO:1903508)	0.128629926	3.422079453	TFAP2A, ING5, JUN, FOXC1, FZD2, RNF14, ZNF281, EPC1, EP300, SMAD5, TEAD2, CAMKK2
Positive regulation of transcription from RNA polymerase II promoter involved in cellular response to chemical stimulus (GO:1901522)	0.128629926	22.79770115	SESN2, EP300, SMAD5
Negative regulation of gene expression (GO:0010629)	0.150732042	4.041901303	SLC7A5, FBXW7, ZNF281, AGO2, IGF2BP1, EPC1, PPM1D, LSM14A, LGR4
Negative regulation of apoptotic process (GO:0043066)	0.187134472	3.282484832	TFAP2A, SETX, NFKBIA, AVEN, DSTYK, CAV1, UBE2O, CFLAR, FNIP1, ZNF268, IFIT3
Positive regulation of trophoblast cell migration (GO:1901165)	0.187134472	54.37534247	AGO2, ACVR1B
Regulation of nucleic acid-templated transcription (GO:1903506)	0.207354752	3.352657005	FOXC1, BTG1, RNF14, ZNF281, JMJD1C, EP300, TERF2IP, ACVR1B, ZNF268, TEAD2
Negative regulation of protein autophosphorylation (GO:0031953)	0.249305756	33.97945205	ERRFI1, CAV1

Table A.22: Gene ontology analysis of genes shortened by 3'UTR APA in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Nuclear).

A.11 Gene ontology analyses of genes experiencing a 3'UTR PAS switch in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Cytoplasm)

Term	P-adjusted	Odds Ratio	Genes
Negative regulation of cell cycle (GO:0045786)	0.007366376	11.55128633	HEXIM1, LATS2, THAP5, SMAD3, HPGD, CHMP1A, IPO5
Positive regulation of nucleocytoplasmic transport (GO:0046824)	0.029510983	14.53717308	GSK3B, SMAD3, CPSF6, ANP32B, IPO5
Ribosomal large subunit biogenesis (GO:0042273)	0.05571031	11.45563021	NIP7, MAK16, PPAN, RRS1, RRP15
Regulation of apoptotic process (GO:0042981)	0.062791463	2.908804299	GSK3B, ARF4, HSPA5, GADD45B, ZFAND6, SERPINB9, DKK1, MLLT11, UFM1, SLK, LATS2, TCIM, PIM1, CYP1B1, BNIP2, ANP32B, MCL1
mRNA processing (GO:0006397)	0.062791463	4.184836153	SMU1, RBM14, PRPF38A, CPSF6, SART3, CELF1, ELAVL1, SRSF8, WDR33, HNRNPA0
Negative regulation of apoptotic process (GO:0043066)	0.062791463	3.374289316	GSK3B, ARF4, HSPA5, ZFAND6, F2R, SERPINB9, DKK1, UFM1, TCIM, PIM1, BNIP2, CRLF1, MCL1
Negative regulation of programmed cell death (GO:0043069)	0.079332723	3.615692568	GSK3B, ARF4, UFM1, HSPA5, TCIM, ZFAND6, PIM1, BNIP2, SERPINB9, DKK1, MCL1
Negative regulation of protein localization to nucleus (GO:1900181)	0.079332723	22.11272321	GSK3B, UFM1, LATS2
Regulation of synaptic vesicle exocytosis (GO:2000300)	0.079332723	22.11272321	GSK3B, PREPL, RAP1A
Negative regulation of cyclin-dependent protein serine/threonine kinase activity (GO:0045736)	0.082888536	20.81092437	HEXIM1, LATS2, IPO5

Table A.23: Gene ontology analysis of genes lengthened by 3'UTR APA in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Cytoplasm).

Term	P-adjusted	Odds Ratio	Genes
Protein homotetramerization (GO:0051289)	0.11233548	16.70791246	GOLGA2, TK1, ACACA, THG1L
Pyridine-containing compound metabolic process (GO:0072524)	0.11233548	131.3135314	PDXK, PNPO
tRNA modification (GO:0006400)	0.11233548	13.13859911	TYW3, TPRKB, TRMT61A, THG1L
Protein tetramerization (GO:0051262)	0.118089677	11.95838987	GOLGA2, TK1, ACACA, THG1L
Protein localization to chromosome (GO:0034502)	0.125995581	19.22516129	XRCC5, CTCF, WAPL
Protein insertion into ER membrane (GO:0045048)	0.125995581	17.52617647	ALDH3A2, CYB5A, EMC8
Negative regulation of exocytosis (GO:0045920)	0.125995581	56.26591231	ATP9A, RAB7A
Retrograde transport, endosome to Golgi (GO:0042147)	0.132548905	9.53006253	ARFRP1, YKT6, SNX5, RAB7A
Pyrimidine nucleoside salvage (GO:0043097)	0.133393455	43.7579758	UCK1, TK1
Pyrimidine-containing compound salvage (GO:0008655)	0.133393455	43.7579758	UCK1, TK1

Table A.24: Gene ontology analysis of genes shortened by 3'UTR APA in 2 or more timepoints detected in WSN-infected cells in the "biological process" category (Cytoplasm).

A.12 Venn diagrams depicting differential gene expression in WSN-infected cells.

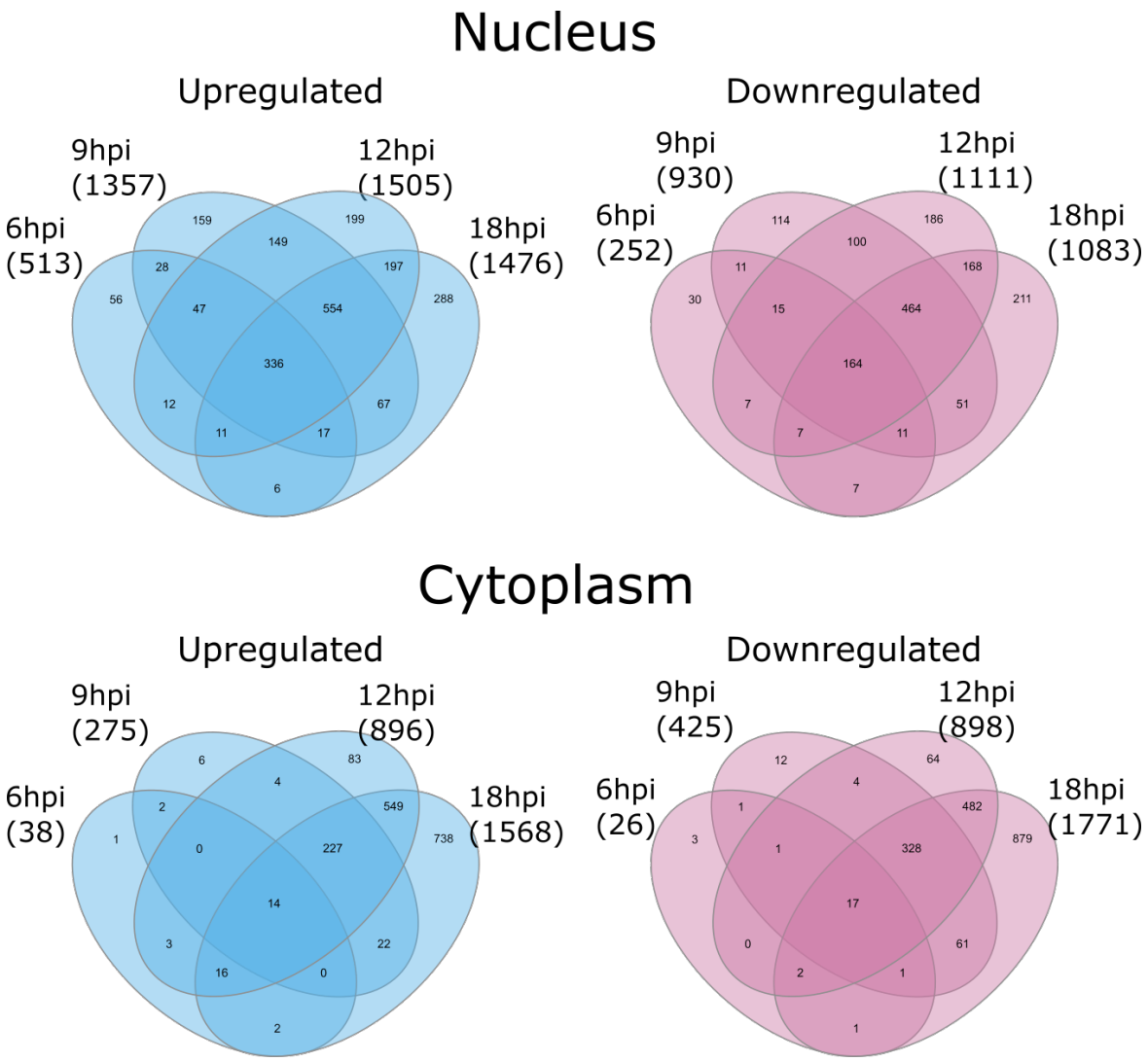


Figure A.3: Overlaps of differentially expressed genes in WSN-infected cells. Venn diagrams depicting all significantly regulated genes in both the nucleus (top) and cytoplasm (bottom) are shown with overlaps. The number of differentially expressed genes in each timepoint is shown in brackets.

A.13 Gene ontology analyses of differentially expressed genes in WSN-infected cells

Term	P-adjusted	Odds Ratio	Number of Genes
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	7.37E-07	4.794990579	30
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	8.01E-07	5.379470672	26
Cytoplasmic translation (GO:0002181)	1.14E-06	5.137766105	26
Cotranslational protein targeting to membrane (GO:0006613)	1.14E-06	5.061937221	26
Protein targeting to ER (GO:0045047)	1.61E-06	4.704630223	27
Nuclear-transcribed mRNA catabolic process (GO:0000956)	2.60E-05	3.292159849	34
Negulation of transcription by RNA polymerase II (GO:0006357)	3.01E-05	1.535460677	221
Positive regulation of transcription, DNA-templated (GO:0045893)	3.57E-05	1.719566228	133
Negative regulation of transcription, DNA-templated (GO:0045892)	5.31E-05	1.791686117	111
Regulation of transcription, DNA-templated (GO:0006355)	1.18E-04	1.494294381	220

Table A.25: Gene ontology analysis of significantly upregulated genes in WSN-infected cells in the "biological process" category (Nuclear).

Term	P-adjusted	Odds Ratio	Number of Genes
Mitotic spindle organization (GO:0007052)	5.02E-16	7.160232968	42
DNA metabolic process (GO:0006259)	1.45E-15	4.895756283	55
Microtubule cytoskeleton organization involved in mitosis (GO:1902850)	1.45E-15	7.940332395	37
Mitotic sister chromatid segregation (GO:0000070)	3.72E-09	6.611123058	26
mRNA splicing, via spliceosome (GO:0000398)	5.61E-09	3.735511017	44
DNA repair (GO:0006281)	6.40E-09	3.56761342	46
Cellular response to DNA damage stimulus (GO:0006974)	1.28E-08	3.262652705	50
Double-strand break repair via homologous recombination (GO:0000724)	2.46E-08	6.341418458	24
DNA replication (GO:0006260)	4.13E-08	5.812638308	25
Double-strand break repair (GO:0006302)	5.59E-08	4.513763876	31

Table A.26: Gene ontology analysis of significantly downregulated genes in WSN-infected cells in the "biological process" category (Nuclear).

Term	P-adjusted	Odds Ratio	Number of Genes
Translation (GO:0006412)	6.62E-10	4.948138129	37
Cytoplasmic translation (GO:0002181)	7.64E-10	8.148092825	24
Gene expression (GO:0010467)	1.33E-08	3.527421389	46
Cotranslational protein targeting to membrane (GO:0006613)	2.38E-08	7.139228886	22
Protein targeting to ER (GO:0045047)	2.38E-08	6.722778799	23
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	5.86E-08	7.103433613	21
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	1.24E-07	5.972671569	23
Peptide biosynthetic process (GO:0043043)	3.56E-07	4.686206897	27
Cellular protein metabolic process (GO:0044267)	2.63E-06	2.862547737	45
Nuclear-transcribed mRNA catabolic process (GO:0000956)	3.98E-06	4.194053527	26

Table A.27: Gene ontology analysis of significantly upregulated genes in WSN-infected cells in the "biological process" category (Cytoplasmic).

Term	P-adjusted	Odds Ratio	Number of Genes
DNA replication (GO:0006260)	1.95E-04	4.921826465	20
DNA strand elongation involved in DNA replication (GO:0006271)	0.010988144	13.63503724	7
Plasma membrane tubulation (GO:0097320)	0.010988144	18.34881486	6
DNA metabolic process (GO:0006259)	0.016791648	2.537055941	29
DNA repair (GO:0006281)	0.018054879	2.428898136	30
DNA replication initiation (GO:0006270)	0.018054879	6.658275474	9
Base-excision repair (GO:0006284)	0.018054879	5.649596965	10
RNA modification (GO:0009451)	0.018054879	5.027825086	11
Negative regulation of protein tyrosine kinase activity (GO:0061099)	0.02208211	8.81989833	7
Base-excision repair, gap-filling (GO:0006287)	0.039964811	6.350395339	8

Table A.28: Gene ontology analysis of significantly downregulated genes in WSN-infected cells in the "biological process" category (Cytoplasmic).

A.14 Gene ontology and BioPlanet pathway analyses of genes experiencing a CR-UTR PAS switch in WSN-infected cells (Cytoplasmic, any timepoint)

Biological Process	P-adjusted	Odds Ratio	Genes
Regulation of mRNA splicing, via spliceosome (GO:0048024)	0.066265187	7.653711621	HNRNPL, RBM39, RBM25, WTAP, NSRP1, SRSF10, JMJD6
Flavonoid glucuronidation (GO:0052696)	0.066265187	44.72850679	UGT1A10, UGT1A9, UGT1A6
Axon extension (GO:0048675)	0.066265187	17.96	CYFIP1, NRP2, HSP90AA1, MAP1B
TOR signaling (GO:0031929)	0.066265187	16.32561983	FLCN, LARP1, RPS6, AKT1
Xenobiotic glucuronidation (GO:0052697)	0.066265187	33.54298643	UGT1A10, UGT1A9, UGT1A6
Telomere maintenance via telomerase (GO:0007004)	0.073856727	14.36436364	ANKRD28, HSP90AA1, PPP6R3, TERF2
RNA-dependent DNA biosynthetic process (GO:0006278)	0.073856727	13.81118881	ANKRD28, HSP90AA1, PPP6R3, TERF2
Neuron projection extension (GO:1990138)	0.083350667	12.38056426	CYFIP1, NRP2, HSP90AA1, MAP1B
Glucuronate metabolic process (GO:0019585)	0.083350667	22.35746606	UGT1A10, UGT1A9, UGT1A6
Negative regulation of mitotic cell cycle phase transition (GO:1901991)	0.083350667	6.301472157	PSMA5, PSMC6, PRMT2, PSMD2, JADE1, BRD7

Table A.29: Gene ontology analysis of lengthening CR-UTR APA changes in WSN-infected cells (Cytoplasmic).

BioPlanet Pathway	P-adjusted	Odds Ratio	Genes
Gene expression	0.031774815	2.508965826	CARS, RPL30, TRAM1, RPL32, CSTF3, SRSF1, NR3C1, CCAR1, NUP85, PSMD2, AKT1, ZNF544, SUPT16H, HNRNPA3, ZNF480, CBX3, RPS6, UPF3A, MED6, HNRNPL, PSMA5, PSMC6, CLNS1A, EIF3J, EIF3H
Messenger RNA processing	0.031774815	4.741416881	HNRNPL, RBM39, HNRNPA3, PRMT2, NUP85, CSTF3, SRSF1, SRSF10, CCAR1, PPM1G
Irinotecan pathway	0.076263595	26.83167421	UGT1A10, UGT1A9, UGT1A6
Mitochondrial pathway of apoptosis: BH3-only Bcl-2 family	0.078291735	5.953725174	ADORA2B, CSNK2A2, CAPN2, AKT1, VDAC1, BID
Estrogen metabolism	0.078291735	19.1616031	COMT, UGT1A9, UGT1A6
TNF-alpha signaling pathway	0.078291735	5.701883148	MAP2K3, HSP90AA1, PSMD2, AKT1, BID, SQSTM1
Messenger RNA splicing: major pathway	0.078291735	7.143944336	HNRNPL, HNRNPA3, CSTF3, SRSF1, CCAR1
Corticosteroids and cardioprotection	0.078291735	17.88325792	HSP90AA1, AKT1, NR3C1
Disease	0.094758181	2.389892573	RPL30, NPM1, SUPT16H, HSP90AA1, CDKN1B, RPL32, SH3KBP1, GOSR2, RPS6, PDIA6, PSMA5, ERN1, PSMC6, NUP85, PSMD2, AKT1, LRRFIP1
Low-density lipoprotein (LDL) pathway during atherogenesis	0.094758181	44.53153153	CSF1, ACAT1

Table A.30: BioPlanet pathway analysis of lengthening CR-UTR APA changes in WSN-infected cells (Cytoplasmic).

Biological Process	P-adjusted	Odds Ratio	Genes
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	2.54E-12	11.8992674	RPS7, RPLP1, RPL12, PYM1, RPL35A, RPSA, RPL8, RPS4Y1, GSPT1, SMG6, RPS4X, PPP2CA, RPS15A, PPP2R1A, RPS20, EIF3E, RPL37, RPS11, RPL29, RPS21, RPS13, RPS12
Nuclear-transcribed mRNA catabolic process (GO:0000956)	2.54E-12	8.890957072	RPLP1, RPL12, RPL8, RPS4Y1, GSPT1, SMG6, RPS4X, PPP2CA, EXOSC4, RPS15A, PPP2R1A, RPL37, RPS11, RPS13, RPS12, RPS7, PYM1, SAMD4B, RPL35A, RPSA, CNOT10, RPS20, EIF3E, RPL29, CNOT9, RPS21
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	3.80E-08	10.49240285	RPS7, RPLP1, RPL12, RPL35A, RPSA, RPL8, RPS4Y1, RPS4X, RPS15A, RPS20, RPL37, RPS11, RPL29, RPS21, RPS13, RPS12
Cytoplasmic translation (GO:0002181)	4.52E-08	10.08205725	RPS7, RPLP1, RPL12, RPL35A, RPSA, RPL8, GSPT1, RPS4X, RPS15A, RPS20, RPL37, RPS11, RPL29, RPS21, RPS13, RPS12
Cotranslational protein targeting to membrane (GO:0006613)	4.52E-08	9.952289833	RPS7, RPLP1, RPL12, RPL35A, RPSA, RPL8, RPS4Y1, RPS4X, RPS15A, RPS20, RPL37, RPS11, RPL29, RPS21, RPS13, RPS12
Translation (GO:0006412)	5.37E-08	5.911498443	NDUFA7, RPS7, MRPS11, RPLP1, RPL12, MRPS12, RPL35A, RPSA, RPL8, RPS4Y1, GSPT1, RPS4X, AIMP2, RPS15A, EEF1D, RPS20, RPL37, RPS11, RPL29, FARSA, RPS21, RPS13, RPS12
Protein targeting to ER (GO:0045047)	1.33E-07	8.918625264	RPS7, RPLP1, RPL12, RPL35A, RPSA, RPL8, RPS4Y1, RPS4X, RPS15A, RPS20, RPL37, RPS11, RPL29, RPS21, RPS13, RPS12
Peptide biosynthetic process (GO:0043043)	3.75E-07	6.473193473	RPS7, MRPS11, RPLP1, RPL12, MRPS12, RPL35A, RPSA, RPL8, RPS4Y1, GSPT1, RPS4X, RPS15A, RPS20, RPL37, RPS11, RPL29, RPS21, RPS13, RPS12
Gene expression (GO:0010467)	2.22E-06	4.041006227	SEH1L, MRPS11, RPLP1, RPL12, MRPS12, RPL8, RPS4Y1, GSPT1, SMG6, RPS4X, RPS15A, RPL37, RPS11, POLR2I, RPS13, RPS12, NUP210, RPS7, RPL35A, RPSA, TFIP11, TSPAN14, HNRNPUL1, RPS20, SRSF5, RPL29, RPS21
Cellular macromolecule biosynthetic process (GO:0034645)	2.74E-06	4.246678465	MRPS11, RPLP1, RPL12, MRPS12, RPL8, RPS4Y1, GSPT1, RPS4X, RPS15A, ORC3, NT5M, RPL37, RPS11, POLR2I, RPS13, RPS12, RPS7, RPL35A, RPSA, POLA2, EEF1D, POLE2, RPS20, RPL29, RPS21

Table A.31: Gene ontology analysis of shortening CR-UTR APA changes in WSN-infected cells (Cytoplasmic).

BioPlanet Pathway	P-adjusted	Odds Ratio	Genes
Cytoplasmic ribosomal proteins	5.99E-10	10.42964152	RPS7, RPLP1, RPL12, RPL35A, RPSA, RPL8, RPS4Y1, RPS4X, RPS15A, EIF3K, RPS20, EIF3E, RPL37, RPS11, RPL29, RPS21, EIF3D, RPS13, RPS12
Influenza viral RNA transcription and replication	5.99E-10	9.321368749	RPS7, RPLP1, RPL12, RPL35A, RPSA, RPL8, RPS4Y1, SMG6, RPS4X, PPP2CA, RPS15A, PPP2R1A, RPS20, RPL37, RPS11, RPL29, POLR2I, RPS21, RPS13, RPS12
Translation	1.80E-09	7.914977717	EIF2B5, RPS7, RPLP1, RPL12, RPL35A, RPSA, RPL8, RPS4Y1, RPS4X, RPS15A, EIF3K, EEF1D, RPS20, EIF3E, RPL37, RPS11, RPL29, RPS21, EIF3D, RPS13, RPS12
Activation of mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	2.17E-08	14.25331089	RPS7, RPSA, RPS4Y1, RPS4X, RPS15A, EIF3K, RPS20, EIF3E, RPS11, RPS21, EIF3D, RPS13, RPS12
Influenza infection	2.17E-08	7.533488192	SEH1L, RPS7, NUP210, RPLP1, RPL12, RPL35A, RPSA, RPL8, RPS4Y1, RPS4X, RPS15A, RPS20, RPL37, RPS11, RPL29, POLR2I, RPS21, RPS13, RPS12
Disease	3.83E-08	3.349296391	SEH1L, TSG101, SRC, MAPKAP1, RPLP1, RPL12, RPL8, RPS4Y1, PSMD8, PPP2CA, RPS4X, PSMB7, RPS15A, EXOSC4, PPP2R1A, MKNK1, LMNA, RCC1, NELFB, RPL37, NELFA, POLR2I, RPS11, EXTL3, RPS13, RPS12, SEC31A, SUPT16H, NUP210, RPS7, DDX11, SLC11A1, RPL35A, RPSA, PRKAR1B, DDIT3, PSMC2, NMT1, EXOC4, RPS20, RPL29, RPS21
Protein metabolism	1.31E-07	3.876557564	RPLP1, RPL12, RPL8, RPS4Y1, GALNT10, PIGX, RPS4X, RPS15A, SAMM50, ARSK, KIF13A, MGAT1, RPL37, RPS11, ARSD, RPS13, RPS12, SEC31A, EIF2B5, TIMM8A, RPS7, RPL35A, RPSA, EIF3K, EEF1D, RFT1, EIF3E, RPS20, PFDN2, RPL29, EIF3D, RPS21
Gene expression	3.42E-07	2.762385147	EHMT2, RPLP1, UBE2D3, RPL8, SMG6, PSMD8, RPS4X, PPP2R1A, XPO5, SNAPC4, NELFB, RPL37, NELFA, TNPO1, RPS11, RPS13, RPS12, SUPT16H, NUP210, RPS7, RPSA, HNRNPUL1, EEF1D, SRSF5, RPL29, SEH1L, RPL12, RPS4Y1, PPP2CA, PSMB7, RPS15A, EXOSC4, POLR2I, EIF2B5, DARS, ZNF263, SMAD3, RPL35A, CNOT10, NR2F6, AIMP2, EIF3K, NFIB, PSMC2, EIF3E, RPS20, FARSA, EIF3D, RPS21, TNRC6A
Cap-dependent translation initiation	0.001795948	6.71080379	EIF3K, RPLP1, RPL12, RPL35A, EIF3E, RPL37, RPL8, RPL29, EIF3D
Signaling events mediated by HDAC class I	0.008056916	6.143194335	MBD3, HDAC3, SIN3B, NUP210, SIN3A, HDAC11, HDAC6, TNFRSF1A

Table A.32: BioPlanet pathway analysis of shortening CR-UTR APA changes in WSN-infected cells (Cytoplasmic).

A.15 Usage of each PAS signal for PASs regulated by APA in the nucleus during WSN infection: 6hpi, 9hpi and 12hpi

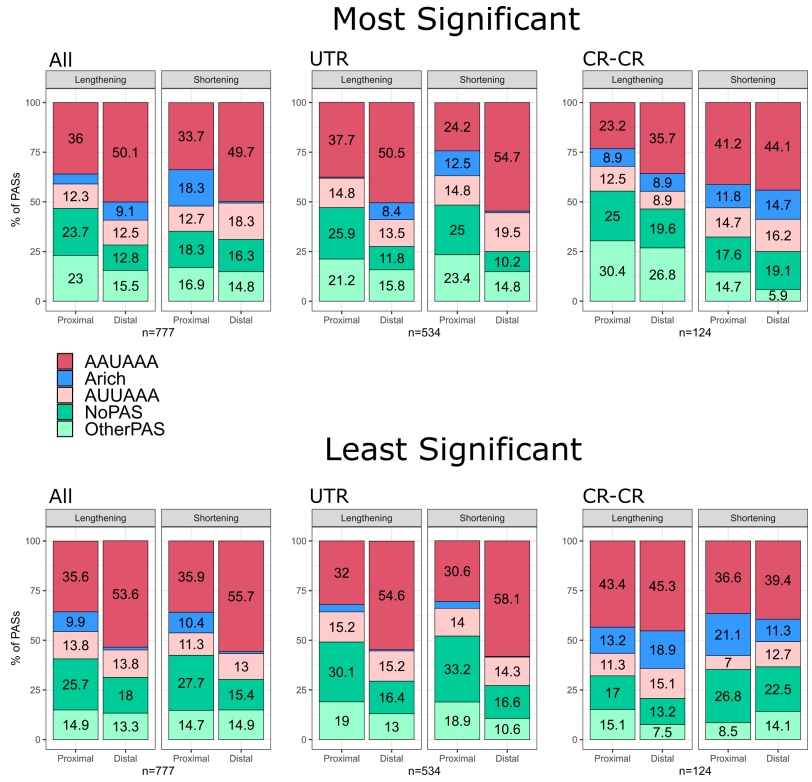


Figure A.4: Relative amounts in the nucleus of each signal sequence in PASs regulated by APA (Top) and those not (Bottom), 6hpi (WSN infection). The distribution of each annotated PAS signal can be seen in those PASs differentially regulated during WSN infection (Top), or those not undergoing APA (Bottom). "All" refers to APA changes involving PASs found anywhere within the pre-mRNA. "UTR" includes APA changes involving PASs found only within the 3'UTR, and "CR-CR" APA comprises those PASs found only within the coding-region. The total number of genes used in each APA category is shown under each chart.

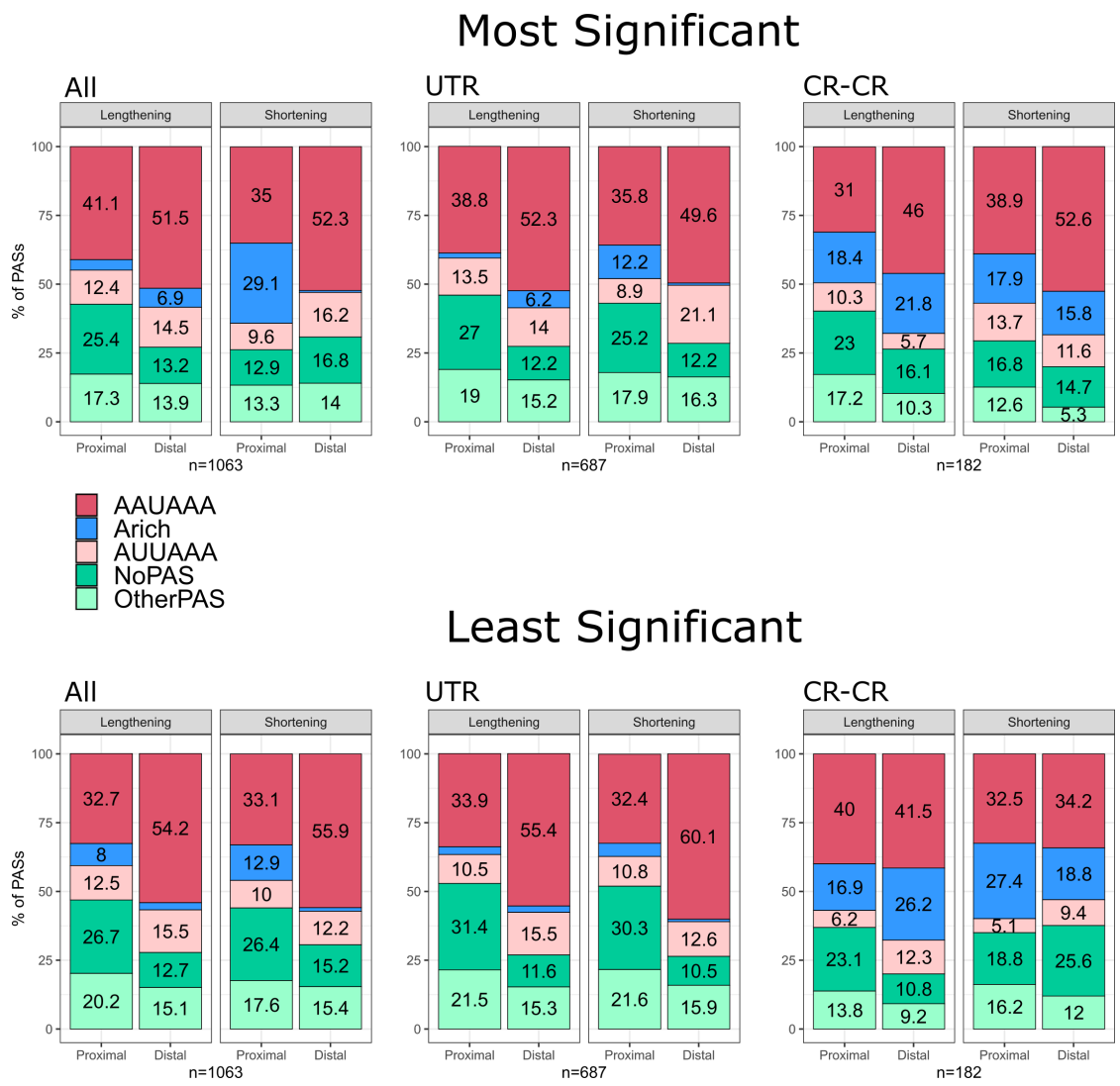


Figure A.5: Relative amounts in the nucleus of each signal sequence in PASs regulated by APA (Top) and those not (Bottom), 9hpi (WSN infection). The distribution of each annotated PAS signal can be seen in those PASs differentially regulated during WSN infection (Top), or those not undergoing APA (Bottom). "All" refers to APA changes involving PASs found anywhere within the pre-mRNA. "UTR" includes APA changes involving PASs found only within the 3'UTR, and "CR-CR" APA comprises those PASs found only within the coding-region. The total number of genes used in each APA category is shown under each chart.

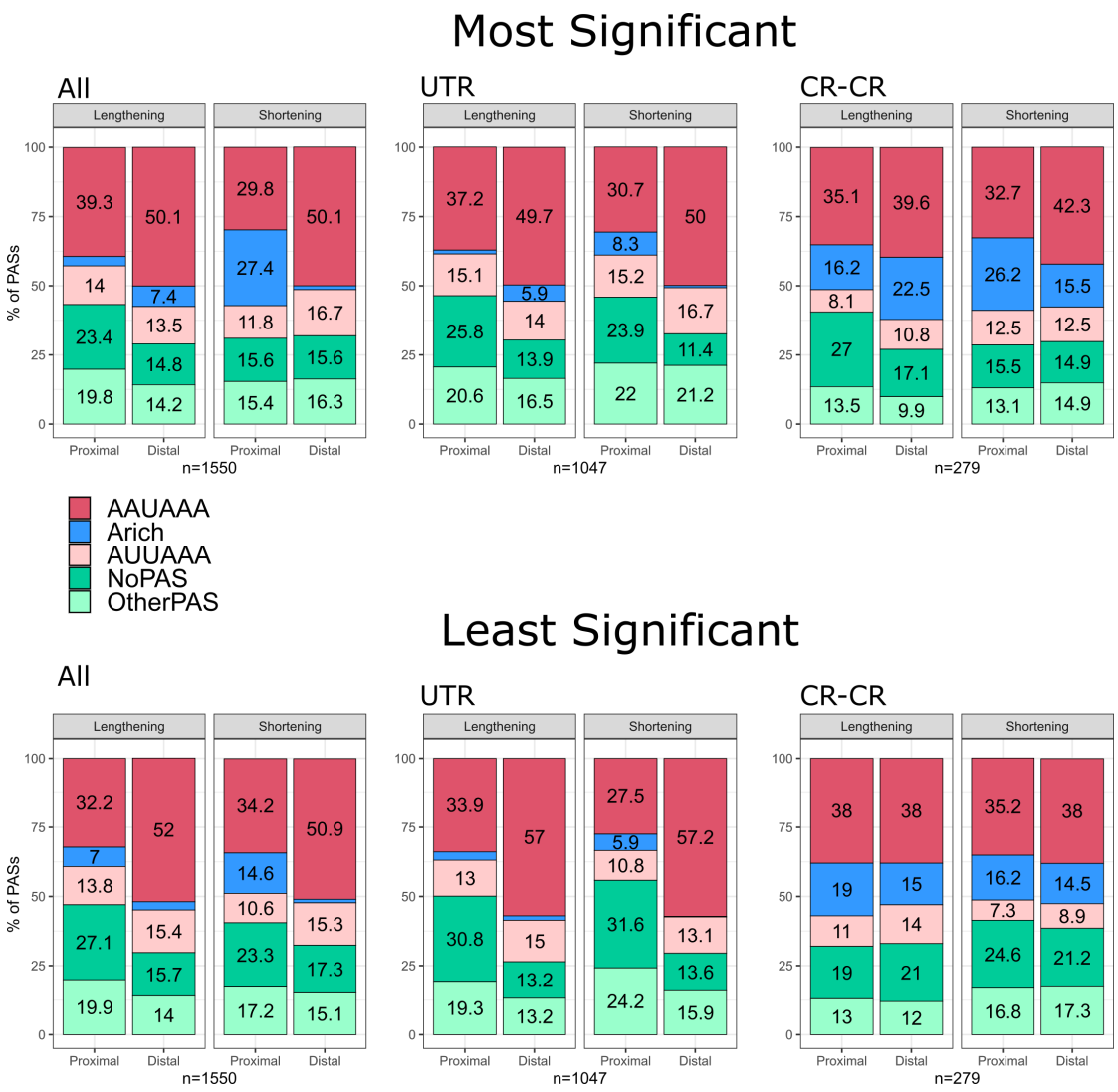


Figure A.6: Relative amounts in the nucleus of each signal sequence in PASs regulated by APA (Top) and those not (Bottom), 12hpi (WSN infection). The distribution of each annotated PAS signal can be seen in those PASs differentially regulated during WSN infection (Top), or those not undergoing APA (Bottom). "All" refers to APA changes involving PASs found anywhere within the pre-mRNA. "UTR" includes APA changes involving PASs found only within the 3'UTR, and "CR-CR" APA comprises those PASs found only within the coding-region. The total number of genes used in each APA category is shown under each chart.

A.16 Location of PAS within gene body for PASs regulated by APA in the nucleus during WSN infection: 6hpi, 9hpi and 12hpi

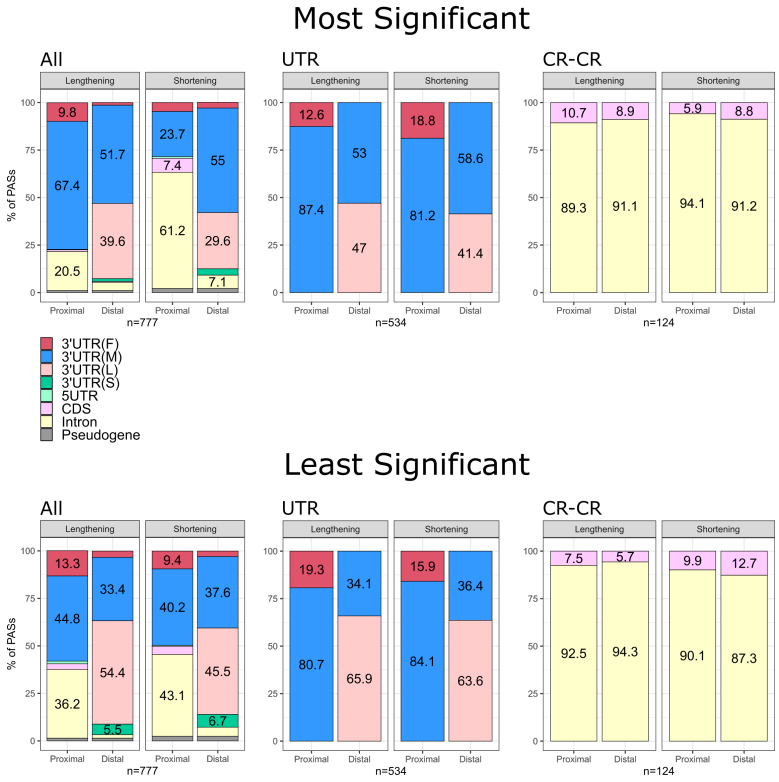


Figure A.7: Location of PAS within the gene body of those regulated by APA (Top) and those not (Bottom) in the nucleus, 6hpi (WSN infection). PASs involved in APA in WSN-infected cells were classified depending on their location in the gene body. Those situated in the 3'UTR were further broken down into first (F), middle (M) and last (L) PAS. If there was just one annotated PAS in the 3'UTR, it is referred to as single (S). "All" refers to APA changes involving PASs found anywhere within the pre-mRNA. "UTR" includes APA changes involving PASs found only within the 3'UTR, and "CR-CR" APA comprises those PASs found only within the coding-region. The total number of genes used in each APA category is shown under each chart.

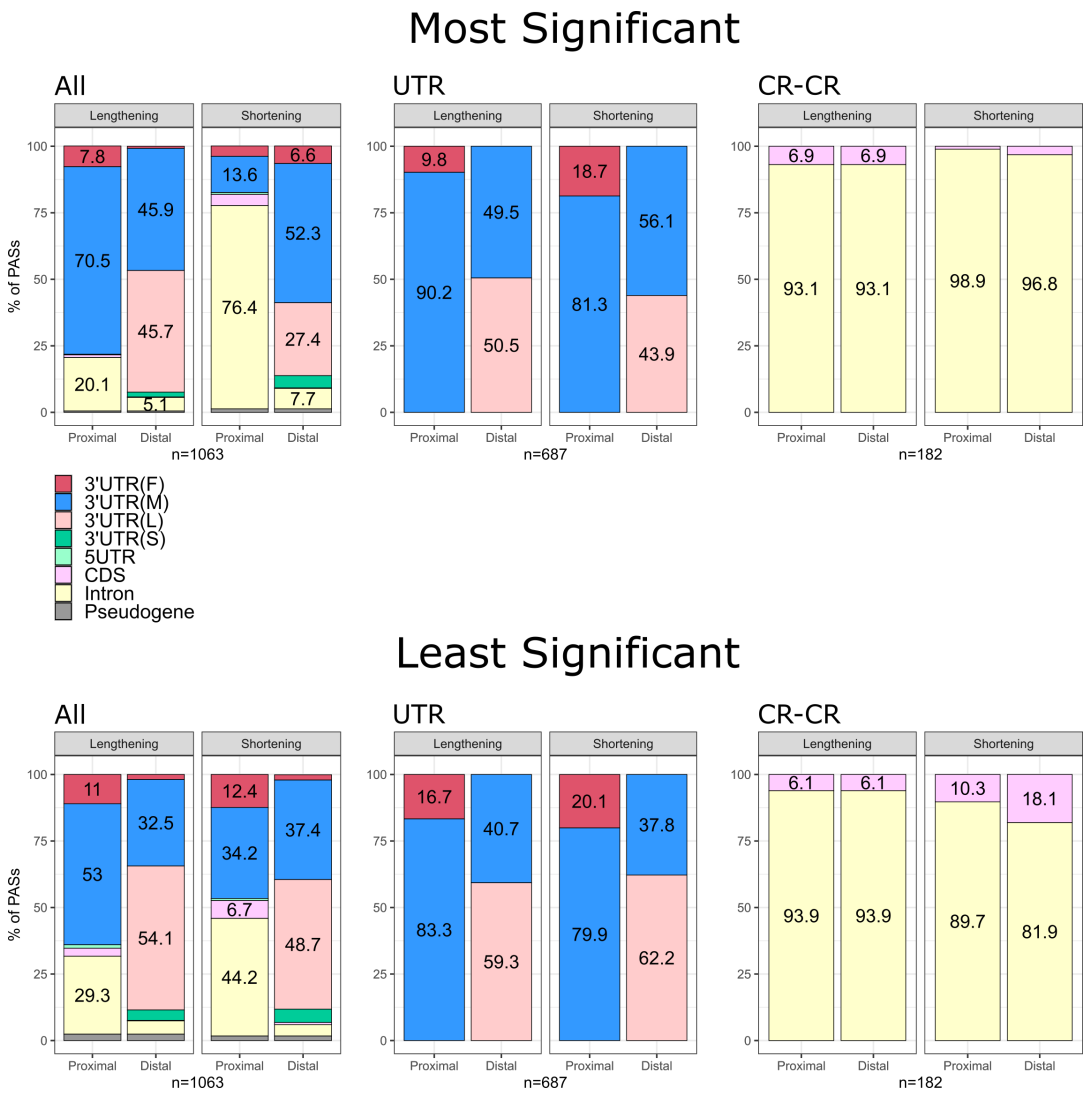


Figure A.8: Location of PAS within the gene body of those regulated by APA (Top) and those not (Bottom) in the nucleus, 9hpi (WSN infection). PASs involved in APA in WSN-infected cells were classified depending on their location in the gene body. Those situated in the 3'UTR were further broken down into first (F), middle (M) and last (L) PAS. If there was just one annotated PAS in the 3'UTR, it is referred to as single (S). "All" refers to APA changes involving PASs found anywhere within the pre-mRNA. "UTR" includes APA changes involving PASs found only within the 3'UTR, and "CR-CR" APA comprises those PASs found only within the coding-region. The total number of genes used in each APA category is shown under each chart.

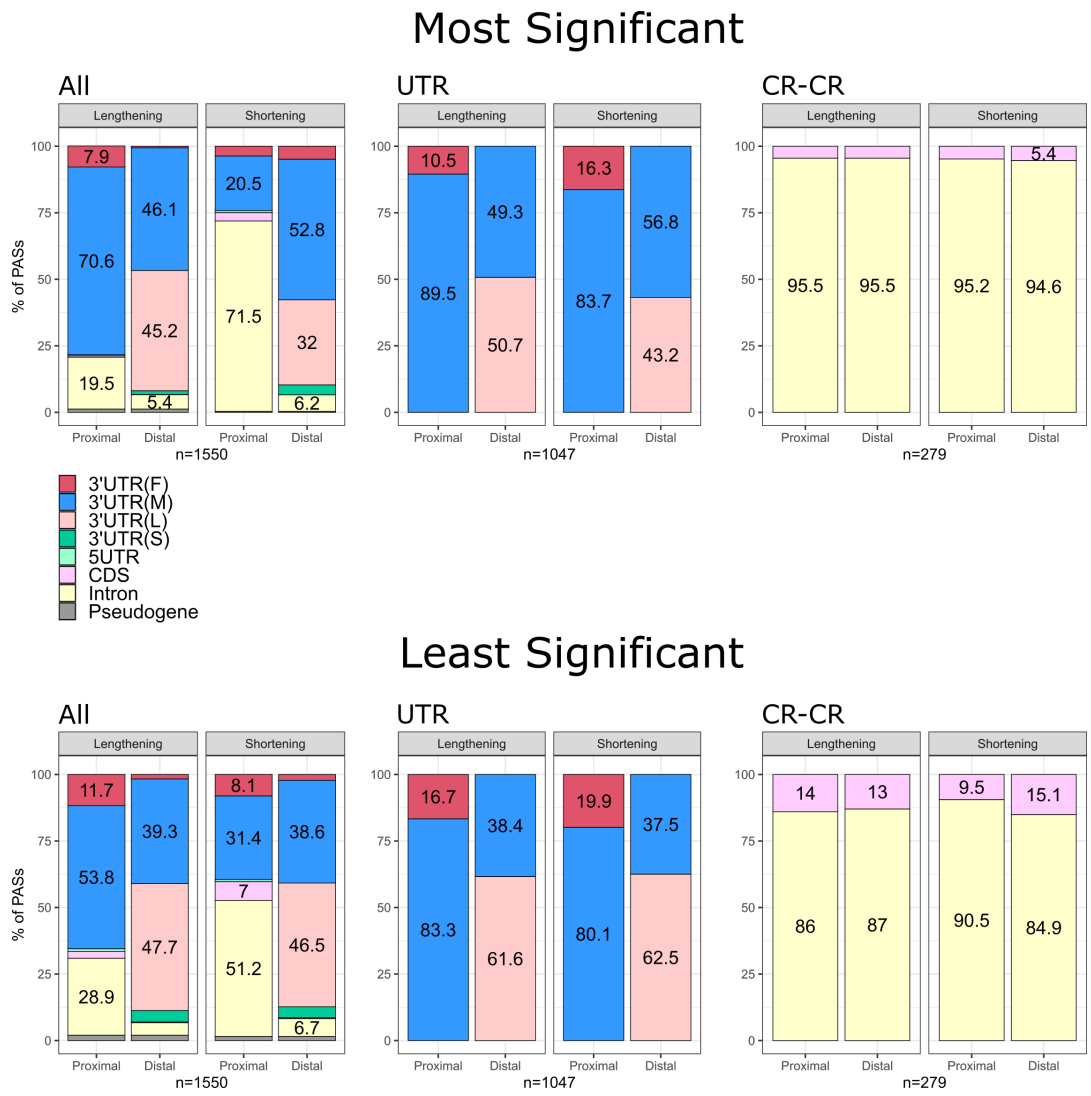


Figure A.9: Location of PAS within the gene body of those regulated by APA (Top) and those not (Bottom) in the nucleus, 12hpi (WSN infection). PASs involved in APA in WSN-infected cells were classified depending on their location in the gene body. Those situated in the 3'UTR were further broken down into first (F), middle (M) and last (L) PAS. If there was just one annotated PAS in the 3'UTR, it is referred to as single (S). "All" refers to APA changes involving PASs found anywhere within the pre-mRNA. "UTR" includes APA changes involving PASs found only within the 3'UTR, and "CR-CR" APA comprises those PASs found only within the coding-region. The total number of genes used in each APA category is shown under each chart.

A.17 K-mers (4- and 6-mers) enriched or depleted in PASs upregulated and downregulated during WSN infection (Nucleus): 6hpi, 9hpi and 12hpi

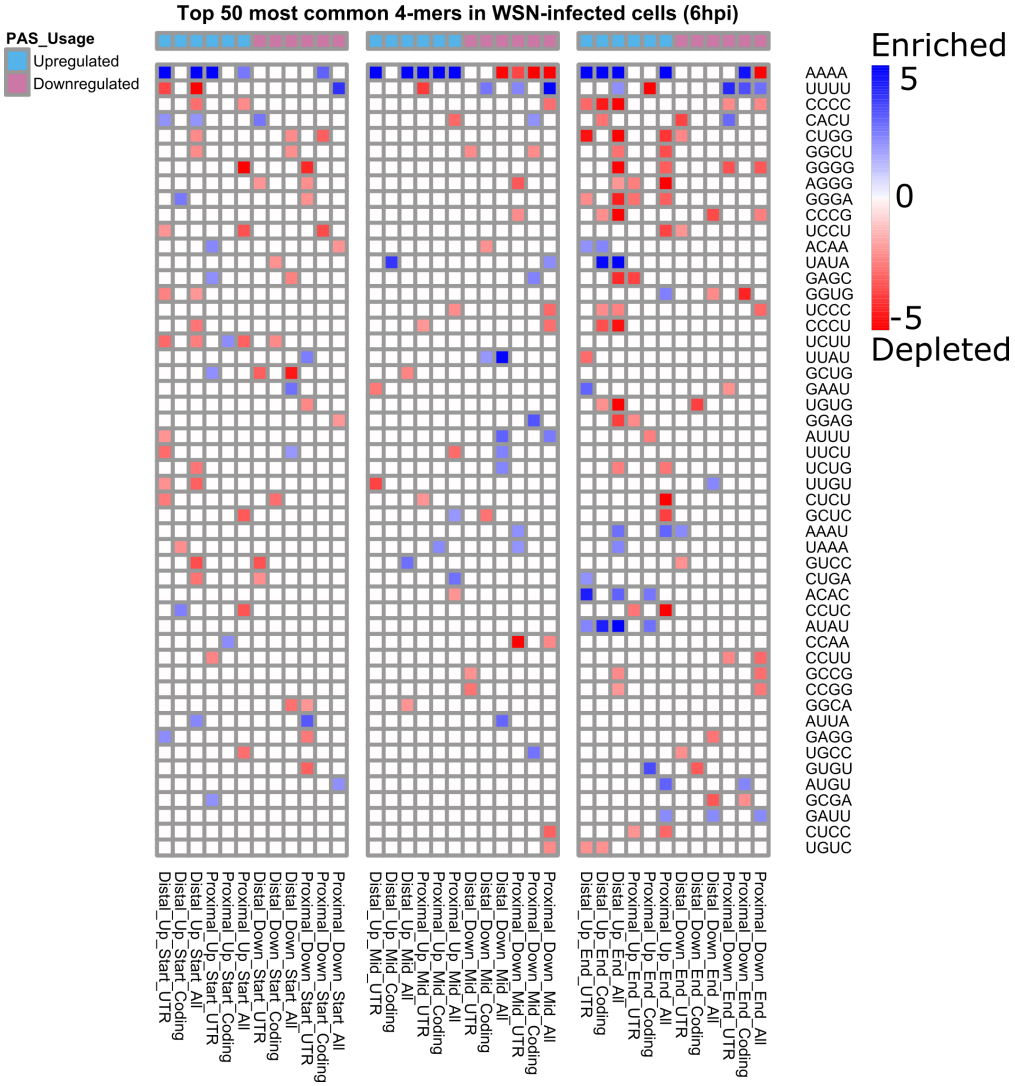


Figure A.10: Top 50 most common enriched and depleted 4-mers (Top 50) in PASs regulated by APA during WSN infection, 6 hours post-infection (Nucleus). PASs have been split into proximal and distal categories, while further being broken into 3 regions surrounding the PAS (Start, Mid and End). The frequency of the top 50 4-mers for each APA type - All, Coding (CR-CR) and UTR is shown (P-val less than 0.001). Significantly regulated PASs were compared with equivalent PASs of other genes, with the Fisher's exact test used to determine significance values for each 4-mer. The -log10(P-value) was then multiplied by 1 if the 4-mer was enriched, or by -1 if the 4-mer was depleted, and these values were then capped at either 5 or -5 for easier visualisation.

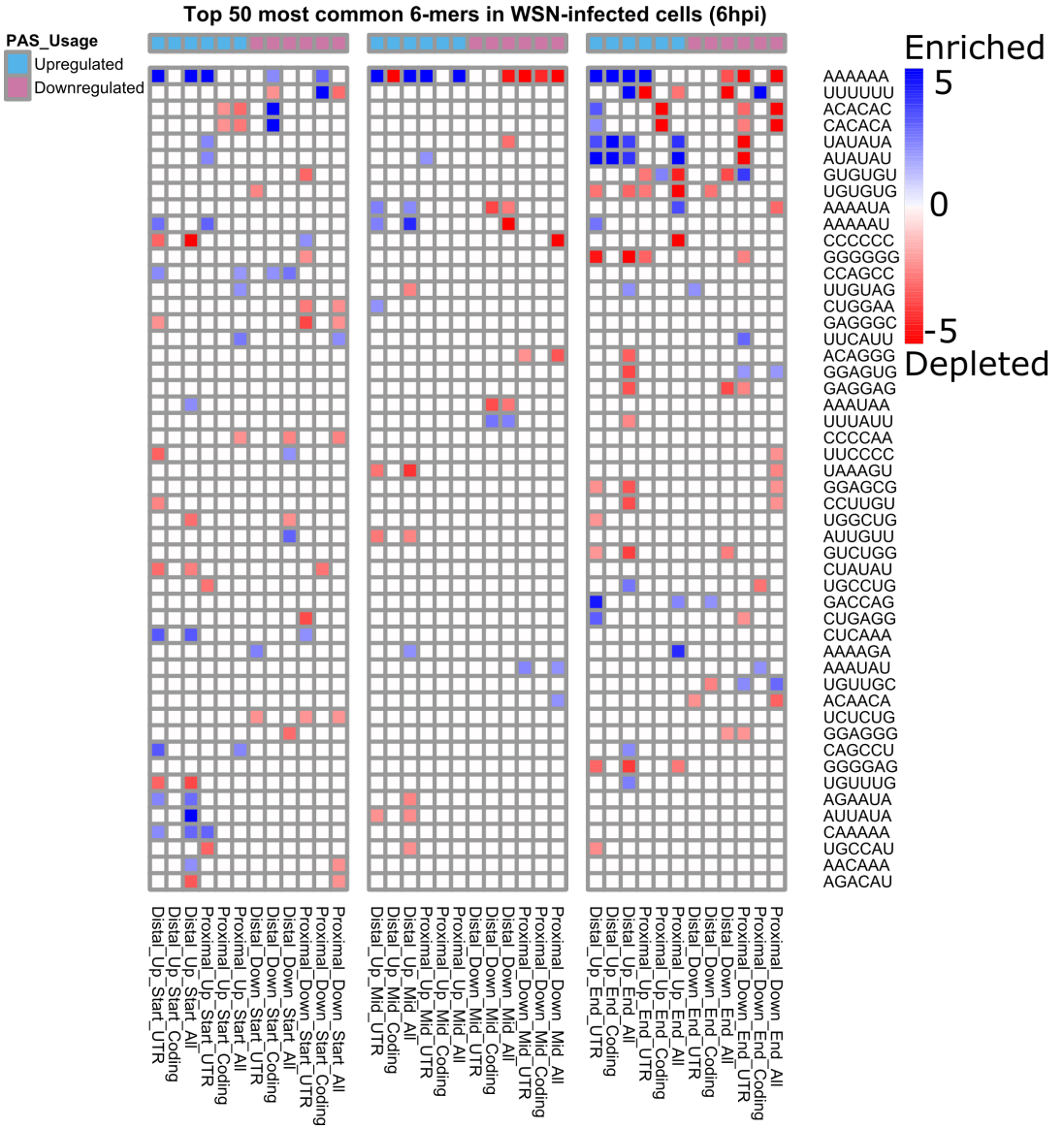


Figure A.11: Top 50 most common enriched and depleted 6-mers (Top 50) in PASs regulated by APA during WSN infection, 6 hours post-infection (Nucleus). PASs have been split into proximal and distal categories, while further being broken into 3 regions surrounding the PAS (Start, Mid and End). The frequency of the top 50 6-mers for each APA type - All, Coding (CR-CR) and UTR is shown (P-val less than 0.001). Significantly regulated PASs were compared with equivalent PASs of other genes, with the Fisher's exact test used to determine significance values for each 6-mer. The -log10(P-value) was then multiplied by 1 if the 6-mer was enriched, or by -1 if the 6-mer was depleted, and these values were then capped at either 5 or -5 for easier visualisation.

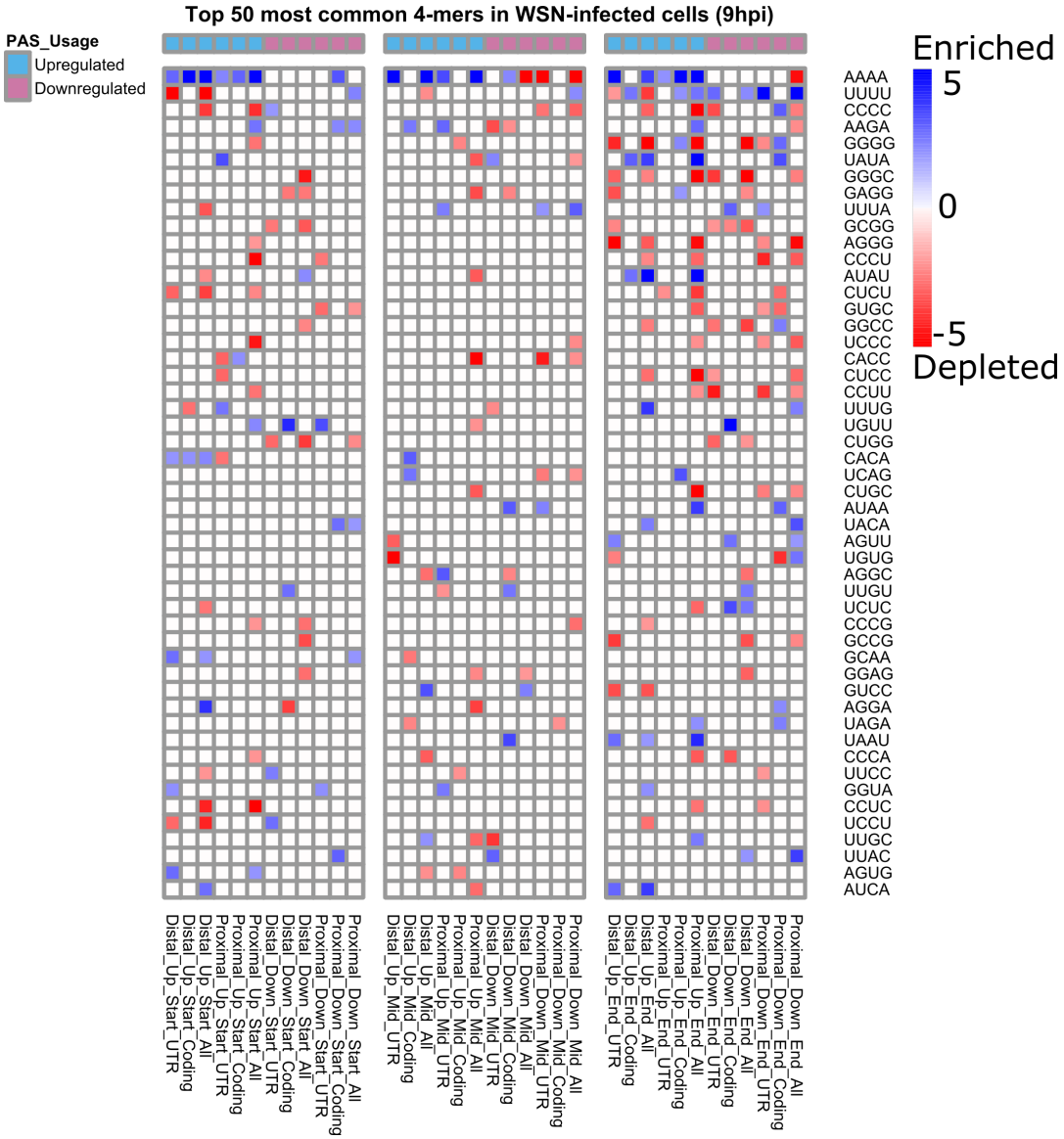


Figure A.12: Top 50 most common enriched and depleted 4-mers (Top 50) in PASs regulated by APA during WSN infection, 9 hours post-infection (Nucleus). PASs have been split into proximal and distal categories, while further being broken into 3 regions surrounding the PAS (Start, Mid and End). The frequency of the top 50 4-mers for each APA type - All, Coding (CR-CR) and UTR is shown (P-val less than 0.001). Significantly regulated PASs were compared with equivalent PASs of other genes, with the Fisher's exact test used to determine significance values for each 4-mer. The -log10(P-value) was then multiplied by 1 if the 4-mer was enriched, or by -1 if the 4-mer was depleted, and these values were then capped at either 5 or -5 for easier visualisation.

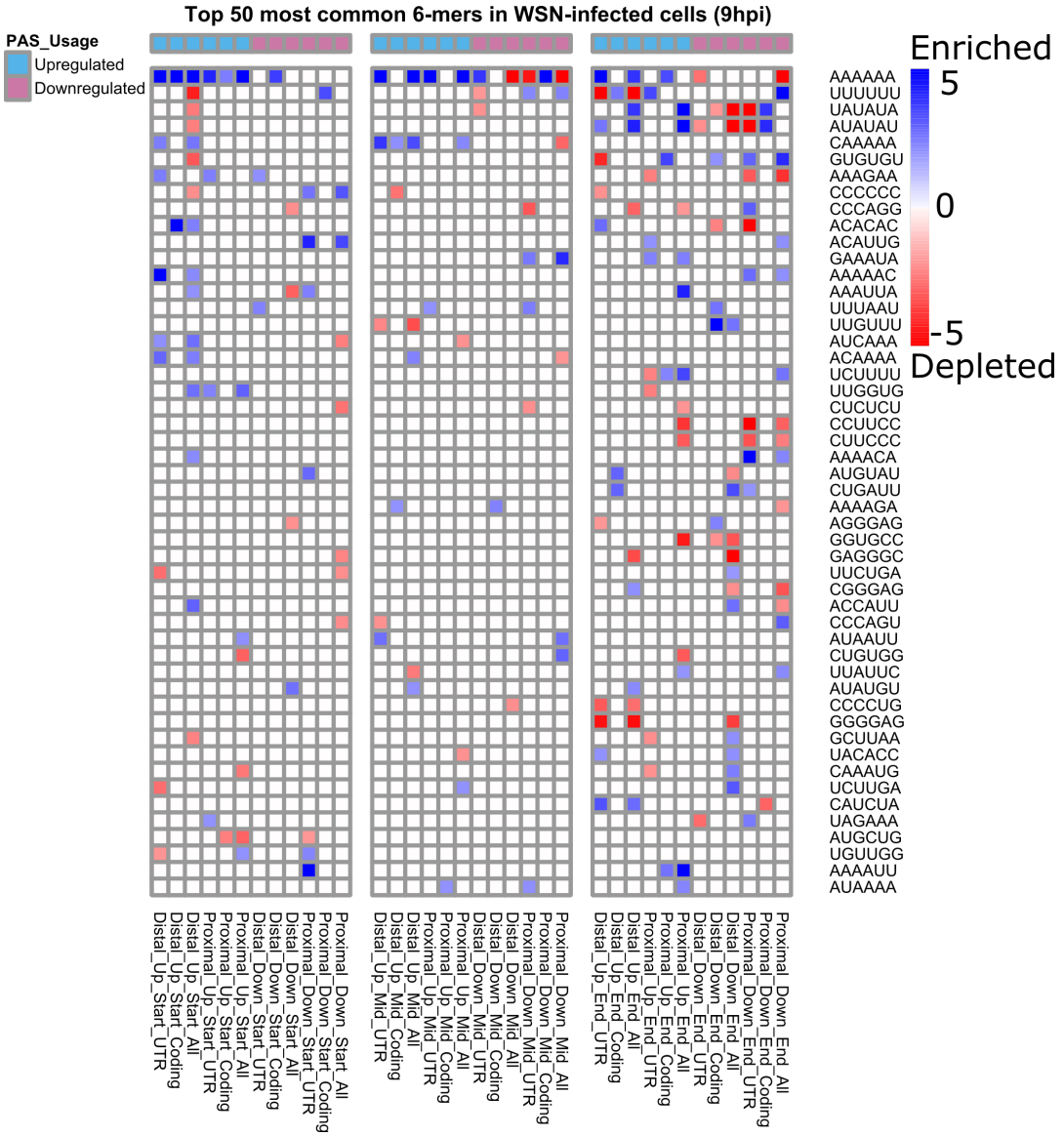


Figure A.13: Top 50 most common enriched and depleted 6-mers (Top 50) in PASs regulated by APA during WSN infection, 9 hours post-infection (Nucleus). PASs have been split into proximal and distal categories, while further being broken into 3 regions surrounding the PAS (Start, Mid and End). The frequency of the top 50 6-mers for each APA type - All, Coding (CR-CR) and UTR is shown (P-val less than 0.001). Significantly regulated PASs were compared with equivalent PASs of other genes, with the Fisher's exact test used to determine significance values for each 6-mer. The -log10(P-value) was then multiplied by 1 if the 6-mer was enriched, or by -1 if the 6-mer was depleted, and these values were then capped at either 5 or -5 for easier visualisation.

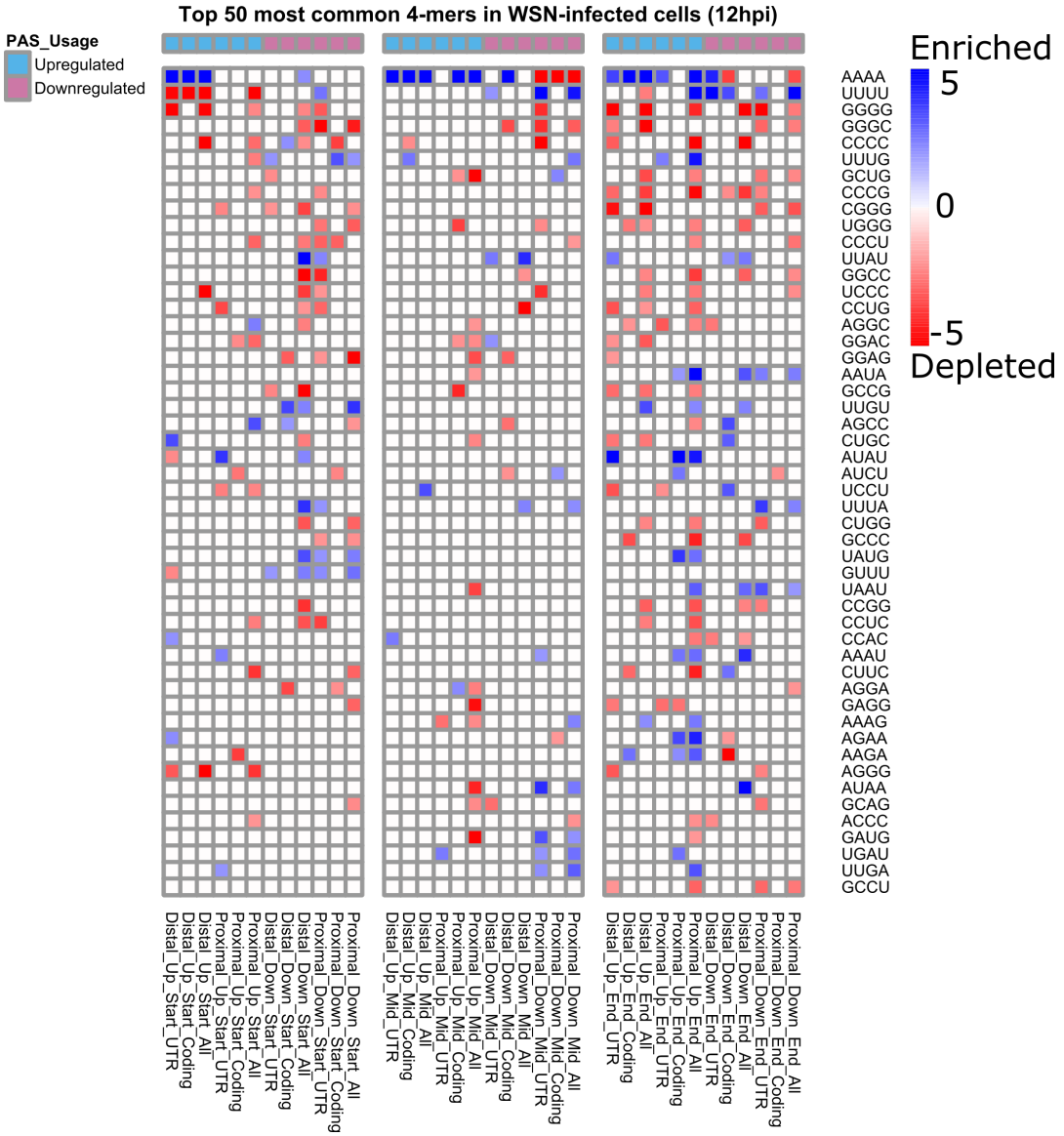


Figure A.14: Top 50 most common enriched and depleted 4-mers (Top 50) in PASs regulated by APA during WSN infection, 12 hours post-infection (Nucleus). PASs have been split into proximal and distal categories, while further being broken into 3 regions surrounding the PAS (Start, Mid and End). The frequency of the top 50 4-mers for each APA type - All, Coding (CR-CR) and UTR is shown (P-val less than 0.001). Significantly regulated PASs were compared with equivalent PASs of other genes, with the Fisher's exact test used to determine significance values for each 4-mer. The -log10(P-value) was then multiplied by 1 if the 4-mer was enriched, or by -1 if the 4-mer was depleted, and these values were then capped at either 5 or -5 for easier visualisation.

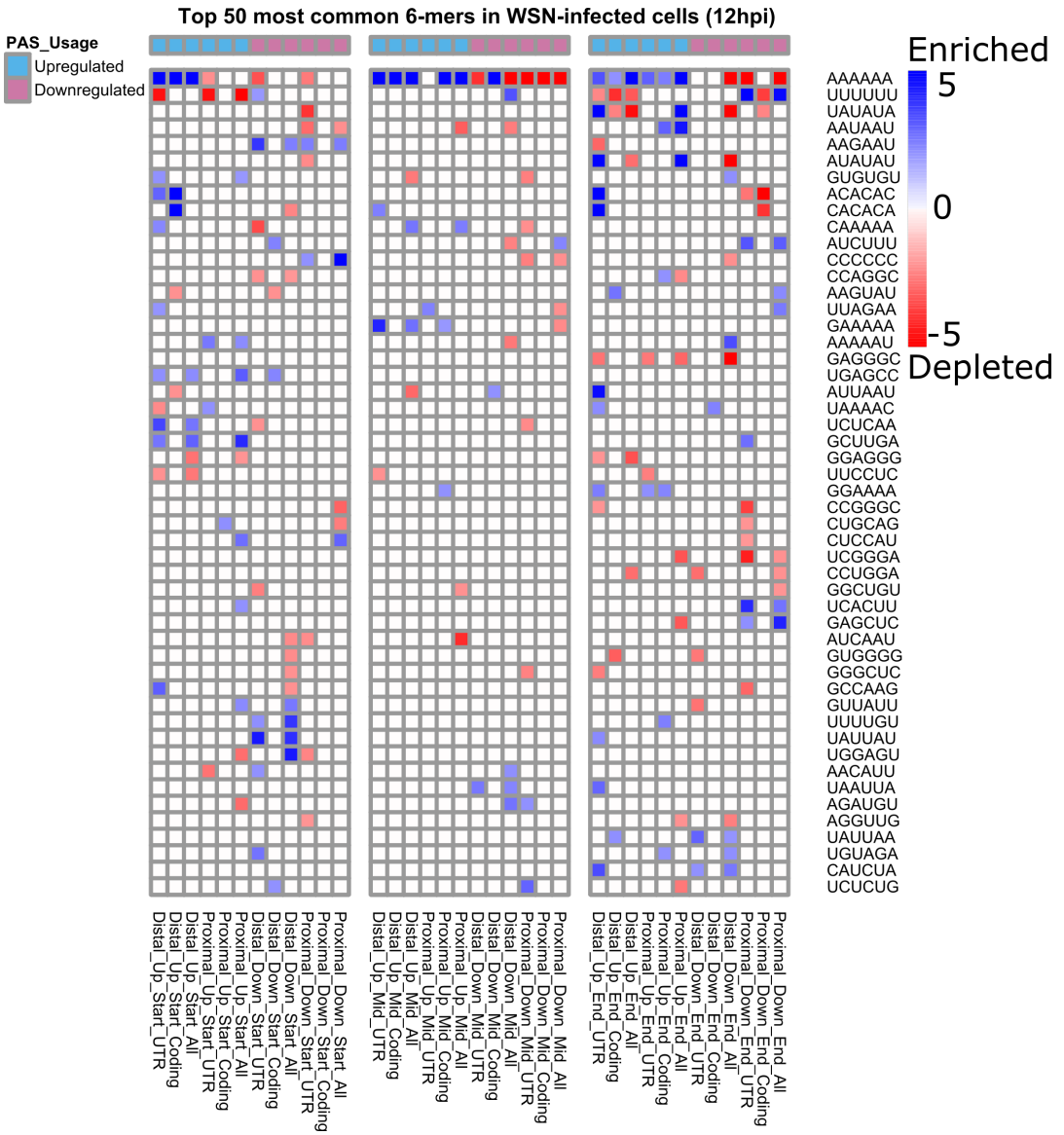


Figure A.15: Top 50 most common enriched and depleted 6-mers (Top 50) in PASs regulated by APA during WSN infection, 12 hours post-infection (Nucleus). PASs have been split into proximal and distal categories, while further being broken into 3 regions surrounding the PAS (Start, Mid and End). The frequency of the top 50 6-mers for each APA type - All, Coding (CR-CR) and UTR is shown (P-val less than 0.001). Significantly regulated PASs were compared with equivalent PASs of other genes, with the Fisher's exact test used to determine significance values for each 6-mer. The -log10(P-value) was then multiplied by 1 if the 6-mer was enriched, or by -1 if the 6-mer was depleted, and these values were then capped at either 5 or -5 for easier visualisation.

Bibliography

Abadi, Mohammad Hassan Jafari Najaf et al. (2019). "CFIm25 and alternative polyadenylation: Conflicting roles in cancer". In: *Cancer letters* 459, pp. 112–121.

Ackerman, Emily E, John F Alcorn, Takeshi Hase, and Jason E Shoemaker (2019). "A dual controllability analysis of influenza virus-host protein-protein interaction networks for antiviral drug target discovery". In: *BMC bioinformatics* 20.1, pp. 1–13.

Ahn, Seong Hoon, Minkyu Kim, and Stephen Buratowski (2004). "Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3 end processing". In: *Molecular cell* 13.1, pp. 67–76.

Akman, Begum H, Tolga Can, and A Elif Erson-Bensan (2012). "Estrogen-induced upregulation and 3-UTR shortening of CDC6". In: *Nucleic acids research* 40.21, pp. 10679–10688.

Alt, Frederick W et al. (1980). "Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3 ends". In: *Cell* 20.2, pp. 293–301.

Amini-Bavil-Olyaee, Samad et al. (2013). "The antiviral effector IFITM3 disrupts intracellular cholesterol homeostasis to block viral entry". In: *Cell host & microbe* 13.4, pp. 452–464.

Ampomah, Patrick B and Lina HK Lim (2020). "Influenza A virus-induced apoptosis and virus propagation". In: *Apoptosis* 25.1, pp. 1–11.

An, Juan Ji, Kusumika Gharami, et al. (2008). "Distinct role of long 3 UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons". In: *Cell* 134.1, pp. 175–187.

An, Sanqi, Yueqi Li, et al. (2021). "Genome-wide profiling reveals alternative polyadenylation of innate immune-related mRNA in patients with COVID-19". In: *Frontiers in immunology* 12.

Anders, Simon, Alejandro Reyes, and Wolfgang Huber (2012). "Detecting differential usage of exons from RNA-seq data". In: *Nature Precedings*, pp. 1–1.

Arora, Ankita, Raeann Goering, Hei Yong G Lo, Joelle Lo, Charlie Moffatt, and J Matthew Taliaferro (2021). "The Role of Alternative Polyadenylation in the Regulation of Subcellular RNA Localization." In: *Frontiers in Genetics* 12, pp. 818668–818668.

Artarini, Anita et al. (2019). "Regulation of influenza A virus mRNA splicing by CLK1". In: *Antiviral research* 168, pp. 187–196.

Ashburner, Michael et al. (2000). "Gene ontology: tool for the unification of biology". In: *Nature genetics* 25.1, pp. 25–29.

Ashraf, Usama et al. (2020). "Influenza virus infection induces widespread alterations of host cell splicing". In: *NAR genomics and bioinformatics* 2.4, lqaa095.

Ayllon, Juan and Adolfo García-Sastre (2014). "The NS1 protein: a multitasking virulence factor". In: *Influenza Pathogenesis and Control-Volume II*, pp. 73–107.

Bai, Yun, Thierry C Auperin, Chi-Yuan Chou, Gu-Gang Chang, James L Manley, and Liang Tong (2007). "Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors". In: *Molecular cell* 25.6, pp. 863–875.

Bailey, Timothy L (2021). "STREME: accurate and versatile sequence motif discovery". In: *Bioinformatics* 37.18, pp. 2834–2840.

Bailey, Timothy L, Mikael Boden, et al. (2009). "MEME SUITE: tools for motif discovery and searching". In: *Nucleic acids research* 37.suppl_2, W202–W208.

Bailey, Timothy L and Charles Elkan (1995). "The value of prior knowledge in discovering motifs with MEME." In: *Ismb*. Vol. 3, pp. 21–29.

Bailey, Timothy L and Charles E Grant (2021). "SEA: Simple Enrichment Analysis of motifs". In: *bioRxiv*.

Barabino, Silvia ML and Walter Keller (1999). "Last but not least: regulated poly (A) tail formation". In: *Cell* 99.1, pp. 9–11.

Barabino, Silvia ML, Martin Ohnacker, and Walter Keller (2000). "Distinct roles of two Yth1p domains in 3-end cleavage and polyadenylation of yeast pre-mRNAs". In: *The EMBO journal* 19.14, pp. 3778–3787.

Bartel, David P (2009). "MicroRNAs: target recognition and regulatory functions". In: *cell* 136.2, pp. 215–233.

Batra, Ranjan et al. (2014). "Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease". In: *Molecular cell* 56.2, pp. 311–322.

Baudu, Timothée et al. (2021). "The NMD Pathway Regulates GABARAPL1 mRNA during the EMT". In: *Biomedicines* 9.10, p. 1302.

Bauer, David LV et al. (2018). "Influenza virus mounts a two-pronged attack on host RNA polymerase II transcription". In: *Cell reports* 23.7, pp. 2119–2129.

Bavagnoli, Laura et al. (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". In: *Nucleic acids research* 43.19, pp. 9405–9417.

Beaudoing, Emmanuel, Susan Freier, Jacqueline R Wyatt, Jean-Michel Claverie, and Daniel Gautheret (2000). "Patterns of variant polyadenylation signal usage in human genes". In: *Genome research* 10.7, pp. 1001–1010.

Bennett, Craig L et al. (2001). "A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA→ AAUGAA) leads to the IPEX syndrome". In: *Immunogenetics* 53.6, pp. 435–439.

Bentley, David L (2014). "Coupling mRNA processing with transcription in time and space". In: *Nature Reviews Genetics* 15.3, pp. 163–175.

Bercovich-Kinori, Adi et al. (2016). "A systematic view on influenza induced host shutoff". In: *Elife* 5, e18311.

Berget, Susan M, Claire Moore, and Phillip A Sharp (1977). "Spliced segments at the 5 terminus of adenovirus 2 late mRNA". In: *Proceedings of the National Academy of Sciences* 74.8, pp. 3171–3175.

Berkovits, Binyamin D and Christine Mayr (2015). "Alternative 3 UTRs act as scaffolds to regulate membrane protein localization". In: *Nature* 522.7556, pp. 363–367.

Bhadra, Malini, Porsha Howell, Sneha Dutta, Caroline Heintz, and William B Mair (2020). "Alternative splicing in aging and longevity". In: *Human genetics* 139.3, pp. 357–369.

Boreikaite, Vytaute, Thomas S Elliott, Jason W Chin, and Lori A Passmore (2022). "RBBP6 activates the pre-mRNA 3 end processing machinery in humans". In: *Genes & development* 36.3-4, pp. 210–224.

Boutet, Stéphane C et al. (2012). "Alternative polyadenylation mediates microRNA regulation of muscle stem cell function". In: *Cell stem cell* 10.3, pp. 327–336.

Bouvier, Nicole M and Peter Palese (2008). "The biology of influenza viruses". In: *Vaccine* 26, pp. D49–D53.

Brewer, GARY (1991). "An A+ U-rich element RNA-binding factor regulates c-myc mRNA stability in vitro". In: *Molecular and cellular biology* 11.5, pp. 2460–2466.

Brugiolo, Mattia, Lydia Herzel, and Karla M Neugebauer (2013). "Counting on co-transcriptional splicing". In: *F1000prime reports* 5.

Bruhn, Oliver et al. (2020). "Alternative polyadenylation of ABC transporters of the C-family (ABCC1, ABCC2, ABCC3) and implications on posttranscriptional micro-RNA regulation". In: *Molecular Pharmacology* 97.2, pp. 112–122.

Bullough, Per A, Frederick M Hughson, John J Skehel, and Don C Wiley (1994). "Structure of influenza haemagglutinin at the pH of membrane fusion". In: *Nature* 371.6492, pp. 37–43.

Carrasco, Judit et al. (2020). "ELAV and FNE determine neuronal transcript signatures through EXon-activated rescue". In: *Molecular Cell* 80.1, pp. 156–163.

Carrique, Loïc et al. (2020). "Host ANP32A mediates the assembly of the influenza virus replicase". In: *Nature* 587.7835, pp. 638–643.

Castro Martin, Isabel Fernández de, Guillaume Fournier, Martin Sachse, Javier Pizarro-Cerda, Cristina Risco, and Nadia Naffakh (2017). "Influenza virus genome reaches the plasma membrane via a modified endoplasmic reticulum and Rab11-dependent vesicles". In: *Nature communications* 8.1, pp. 1–12.

Ceelie, H, CC Spaargaren-van Riel, RM Bertina, and HL Vos (2004). "G20210A is a functional mutation in the prothrombin gene; effect on protein levels and 3-end formation". In: *Journal of Thrombosis and Haemostasis* 2.1, pp. 119–127.

Chaimayo, Chutikarn, Megan Dunagan, Tsuyoshi Hayashi, Netty Santoso, and Toru Takimoto (2018). "Specificity and functional interplay between influenza virus PA-X and NS1 shutoff activity". In: *PLoS Pathogens* 14.11, e1007465.

Chan, Annie Y, Frank T Vreede, Matt Smith, Othmar G Engelhardt, and Ervin Fodor (2006). "Influenza virus inhibits RNA polymerase II elongation". In: *Virology* 351.1, pp. 210–217.

Chan, Serena L, Ina Huppertz, et al. (2014). "CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3 processing". In: *Genes & development* 28.21, pp. 2370–2380.

Chang, Hyeshik, Jaechul Lim, Minju Ha, and V Narry Kim (2014). "TAIL-seq: genome-wide determination of poly (A) tail length and 3' end modifications". In: *Molecular cell* 53.6, pp. 1044–1052.

Chen, Edward Y, Christopher M Tan, et al. (2013). "Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool". In: *BMC bioinformatics* 14.1, pp. 1–14.

Chen, Guifang, Chien-Hung Liu, Ligang Zhou, and Robert M Krug (2014). "Cellular DDX21 RNA helicase inhibits influenza A virus replication but is counteracted by the viral NS1 protein". In: *Cell host & microbe* 15.4, pp. 484–493.

Chen, Guifang, Li-Chung Ma, et al. (2020). "A double-stranded RNA platform is required for the interaction between a host restriction factor and the NS1 protein of influenza A virus". In: *Nucleic acids research* 48.1, pp. 304–315.

Chen, Linan, Jason G Dumelie, et al. (2014). "Global regulation of mRNA translation and stability in the early Drosophila embryo by the Smaug RNA-binding protein". In: *Genome biology* 15.1, pp. 1–21.

Chen, Ling-Ling, Joshua N DeCerbo, and Gordon G Carmichael (2008). "Alu element-mediated gene silencing". In: *The EMBO journal* 27.12, pp. 1694–1705.

Chen, Robert W, Lynne T Bemis, et al. (2008). "Truncation in CCND1 mRNA alters miR-16-1 regulation in mantle cell lymphoma". In: *Blood, The Journal of the American Society of Hematology* 112.3, pp. 822–829.

Chen, Tao and Bas van Steensel (2017). "Comprehensive analysis of nucleocytoplasmic dynamics of mRNA in Drosophila cells". In: *PLoS Genetics* 13.8, e1006929.

Cheng, Larry C, Dinghai Zheng, et al. (2020). "Widespread transcript shortening through alternative polyadenylation in secretory cell differentiation". In: *Nature communications* 11.1, pp. 1–14.

Cheng, Yiming, Robert M Miura, and Bin Tian (2006). "Prediction of mRNA polyadenylation sites by support vector machine". In: *Bioinformatics* 22.19, pp. 2320–2325.

Cherry, Sara and Kristen W Lynch (2020). "Alternative splicing and cancer: insights, opportunities, and challenges from an expanding view of the transcriptome". In: *Genes & Development* 34.15-16, pp. 1005–1016.

Choppin, Purnell W et al. (1969). "Replication of influenza virus in a continuous cell line: high yield of infective virus from cells inoculated at high multiplicity." In: *Virology* 39.1, pp. 130–34.

Chow, Louise T, Richard E Gelinas, Thomas R Broker, and Richard J Roberts (1977). "An amazing sequence arrangement at the 5 ends of adenovirus 2 messenger RNA". In: *Cell* 12.1, pp. 1–8.

Chuvpilo, Sergei et al. (1999). "Alternative polyadenylation events contribute to the induction of NF-ATc in effector T cells". In: *Immunity* 10.2, pp. 261–269.

Ciolli Mattioli, Camilla et al. (2019). "Alternative 3 UTRs direct localization of functionally diverse protein isoforms in neuronal compartments". In: *Nucleic acids research* 47.5, pp. 2560–2573.

Cohen, Sarah, Shelly Au, and Nelly Panté (2011). "How viruses access the nucleus". In: *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1813.9, pp. 1634–1645.

Colgan, Diana F and James L Manley (1997). "Mechanism and regulation of mRNA polyadenylation". In: *Genes & development* 11.21, pp. 2755–2766.

Cowley, Michael, Andrew J Wood, Sabrina Böhm, Reiner Schulz, and Rebecca J Oakey (2012). "Epigenetic control of alternative mRNA processing at the imprinted Herc3/Nap1l5 locus". In: *Nucleic acids research* 40.18, pp. 8917–8926.

Cui, Yajun and Clyde L Denis (2003). "In vivo evidence that defects in the transcriptional elongation factors RPB2, TFIIS, and SPT5 enhance upstream poly (A) site utilization". In: *Molecular and cellular biology* 23.21, pp. 7887–7901.

Curinha, Ana, Sandra Oliveira Braz, Isabel Pereira-Castro, Andrea Cruz, and Alexandra Moreira (2014). "Implications of polyadenylation in health and disease". In: *Nucleus* 5.6, pp. 508–519.

Dadonaite, Bernadeta, Brad Gilbertson, et al. (2019). "The structure of the influenza A virus genome". In: *Nature microbiology* 4.11, pp. 1781–1789.

Dadonaite, Bernadeta, Swetha Vijayakrishnan, Ervin Fodor, David Bhella, and Edward C Hutchinson (2016). "Filamentous influenza viruses". In: *The Journal of general virology* 97.8, p. 1755.

Dai, Weijun, Gen Zhang, and Eugene V Makeyev (2012). "RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage". In: *Nucleic acids research* 40.2, pp. 787–800.

Das, Kalyan et al. (2008). "Structural basis for suppression of a host antiviral response by influenza A virus". In: *Proceedings of the National Academy of Sciences* 105.35, pp. 13093–13098.

Derti, Adnan et al. (2012). "A quantitative atlas of polyadenylation in five mammals". In: *Genome research* 22.6, pp. 1173–1183.

Deshpande, Anagha et al. (2009). "3'UTR mediated regulation of the cyclin D1 proto-oncogene". In: *Cell Cycle* 8.21, pp. 3592–3600.

Desmet, Emily A, Kendra A Bussey, Raychel Stone, and Toru Takimoto (2013). "Identification of the N-terminal domain of the influenza virus PA responsible for the suppression of host protein synthesis". In: *Journal of virology* 87.6, pp. 3108–3118.

Desterro, Joana, Pedro Bak-Gordon, and Maria Carmo-Fonseca (2020). "Targeting mRNA processing as an anticancer strategy". In: *Nature Reviews Drug Discovery* 19.2, pp. 112–129.

Di Giammartino, Dafne Campigli, Kensei Nishida, and James L Manley (2011). "Mechanisms and consequences of alternative polyadenylation". In: *Molecular cell* 43.6, pp. 853–866.

Dou, Dan, Rebecca Revol, Henrik östbye, Hao Wang, and Robert Daniels (2018). "Influenza A virus cell entry, replication, virion assembly and movement". In: *Frontiers in immunology* 9, p. 1581.

Duan, Haichuan, Nadia Cherradi, Jean-Jacques Feige, and Colin Jefcoate (2009). "cAMP-dependent posttranscriptional regulation of steroidogenic acute regulatory (STAR) protein by the zinc finger protein ZFP36L1/TIS11b". In: *Molecular Endocrinology* 23.4, pp. 497–509.

Dunagan, Megan M, Kala Hardy, and Toru Takimoto (2021). "Impact of influenza A virus shutoff proteins on host immune responses". In: *Vaccines* 9.6, p. 629.

Eberhardt, Christiane S et al. (2021). "Functional HPV-specific PD-1+ stem-like CD8 T cells in head and neck cancer". In: *Nature* 597.7875, pp. 279–284.

Eckner, Richard, W Ellmeier, and ML Birnstiel (1991). "Mature mRNA 3 end formation stimulates RNA export from the nucleus." In: *The EMBO journal* 10.11, pp. 3513–3522.

Edinger, Thomas O, Marie O Pohl, and Silke Stertz (2014). "Entry of influenza A virus: host factors and antiviral targets". In: *Journal of General Virology* 95.2, pp. 263–277.

Edmonds, Mary (2002). "A history of poly A sequences: from formation to factors to function". In.

Eisen, Timothy J et al. (2020). "The dynamics of cytoplasmic mRNA metabolism". In: *Molecular Cell* 77.4, pp. 786–799.

Eisfeld, Amie J, Gabriele Neumann, and Yoshihiro Kawaoka (2015). "At the centre: influenza A virus ribonucleoproteins". In: *Nature Reviews Microbiology* 13.1, pp. 28–41.

Elkon, Ran, Jarno Drost, et al. (2012). "E2F mediates enhanced alternative polyadenylation in proliferation". In: *Genome biology* 13.7, pp. 1–15.

Elkon, Ran, Alejandro P Ugalde, and Reuven Agami (2013). "Alternative cleavage and polyadenylation: extent, regulation and function". In: *Nature Reviews Genetics* 14.7, pp. 496–506.

Engelhardt, Othmar G, Matt Smith, and Ervin Fodor (2005). "Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II". In: *Journal of virology* 79.9, pp. 5812–5818.

Fabian, Marc Robert, Nahum Sonenberg, and Witold Filipowicz (2010). "Regulation of mRNA translation and stability by microRNAs". In: *Annual review of biochemistry* 79, pp. 351–379.

Fan, Haitian et al. (2019). "Influenza A virus RNA polymerase structures provide insights into viral genome replication". In: *Nature* 573.7773, p. 287.

Filippova, Galina N et al. (1996). "An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes". In: *Molecular and cellular biology* 16.6, pp. 2802–2813.

Firth, AE et al. (2012). "Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC_UUU_CGU and is in the+ 1 direction". In: *Open biology* 2.10, p. 120109.

Fischl, Harry et al. (2019). "hnRNPC regulates cancer-specific alternative cleavage and polyadenylation profiles". In: *Nucleic acids research* 47.14, pp. 7580–7591.

Fodor, E (2013). "The RNA polymerase of influenza a virus: mechanisms of viral transcription and replication." In: *Acta virologica* 57.2, pp. 113–122.

Fodor, Ervin and Aartjan JW Te Velthuis (2020). "Structure and function of the influenza virus transcription and replication machinery". In: *Cold Spring Harbor perspectives in medicine* 10.9, a038398.

Fournier, Emilie et al. (2012). "A supramolecular assembly formed by influenza A virus genomic RNA segments". In: *Nucleic acids research* 40.5, pp. 2197–2209.

Frendewey, David and Walter Keller (1985). "Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences". In: *Cell* 42.1, pp. 355–367.

Fresco, LUCILLE D and STEPHEN Buratowski (1996). "Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing." In: *Rna* 2.6, pp. 584–596.

Fu, Yonggui et al. (2011). "Differential genome-wide profiling of tandem 3 UTRs among human breast cancer and normal cells by high-throughput sequencing". In: *Genome research* 21.5, pp. 741–747.

Gamblin, Steven J and John J Skehel (2010). "Influenza hemagglutinin and neuraminidase membrane glycoproteins". In: *Journal of biological chemistry* 285.37, pp. 28403–28409.

Gao, Huijie, Yipeng Sun, et al. (2015). "The contribution of PA-X to the virulence of pandemic 2009 H1N1 and highly pathogenic H5N1 avian influenza viruses". In: *Scientific reports* 5.1, pp. 1–11.

Gao, Shijuan, Jiaoxiang Wu, et al. (2015). "Interaction of NS2 with AIMP2 facilitates the switch from ubiquitination to SUMOylation of M1 in influenza A virus-infected cells". In: *Journal of virology* 89.1, pp. 300–311.

Garaulet, Daniel L, Binglong Zhang, Lu Wei, Elena Li, and Eric C Lai (2020). "miRNAs and neural alternative polyadenylation specify the virgin behavioral state". In: *Developmental cell* 54.3, pp. 410–423.

García-Sastre, Adolfo (2011). "Induction and evasion of type I interferon responses by influenza viruses". In: *Virus research* 162.1-2, pp. 12–18.

García-Sastre, Adolfo et al. (1998). "Influenza A virus lacking the NS1 gene replicates in interferon-deficient systems". In: *Virology* 252.2, pp. 324–330.

Garneau, Nicole L, Jeffrey Wilusz, and Carol J Wilusz (2007). "The highways and byways of mRNA decay". In: *Nature reviews Molecular cell biology* 8.2, pp. 113–126.

Gaucherand, Lea et al. (2019). "The influenza A virus endoribonuclease PA-X usurps host mRNA processing machinery to limit host gene expression". In: *Cell reports* 27.3, pp. 776–792.

Gautheret, Daniel, Olivier Poirot, Fabrice Lopez, Stéphane Audic, and Jean-Michel Claverie (1998). "Alternate polyadenylation in human mRNAs: a large-scale analysis by EST clustering". In: *Genome research* 8.5, pp. 524–530.

Gehring, Niels H et al. (2001). "Increased efficiency of mRNA 3 end formation: a new genetic mechanism contributing to hereditary thrombophilia". In: *Nature genetics* 28.4, pp. 389–392.

Geisberg, Joseph V, Zarmik Moqtaderi, Xiaochun Fan, Fatih Ozsolak, and Kevin Struhl (2014). "Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast". In: *Cell* 156.4, pp. 812–824.

Geisberg, Joseph V, Zarmik Moqtaderi, and Kevin Struhl (2020). "The transcriptional elongation rate regulates alternative polyadenylation in yeast". In: *Elife* 9, e59810.

Gerl, Mathias J et al. (2012). "Quantitative analysis of the lipidomes of the influenza virus envelope and MDCK cell apical membrane". In: *Journal of Cell Biology* 196.2, pp. 213–221.

Ghosh, Souvik et al. (2022). "CFIm-mediated alternative polyadenylation remodels cellular signaling and miRNA biogenesis". In: *Nucleic Acids Research* 50.6, pp. 3096–3114.

Gilat, Roi, Sergey Goncharov, Nir Esterman, and Dorit Shweiki (2006). "Underrepresentation of PolyA/PolyT tailed ESTs in Human ESTdb: an obstacle to alternative polyadenylation inference". In: *Bioinformation* 1.6, p. 220.

Gingras, Anne-Claude, Brian Raught, and Nahum Sonenberg (1999). "eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation". In: *Annual review of biochemistry* 68.1, pp. 913–963.

Girard, Marc P, Thomas Cherian, Yuri Pervikov, and Marie Paule Kieny (2005). "A review of vaccine research and development: human acute respiratory infections". In: *Vaccine* 23.50, pp. 5708–5724.

Glaunsinger, Britt and Don Ganem (2004). "Lytic KSHV infection inhibits host gene expression by accelerating global mRNA turnover". In: *Molecular cell* 13.5, pp. 713–723.

Goering, Raeann, Krysta L Engel, Austin E Gillen, Nova Fong, David L Bentley, and J Matthew Taliaferro (2021). "LABRAT reveals association of alternative polyadenylation with transcript localization, RNA binding protein expression, transcription speed, and cancer survival". In: *BMC genomics* 22.1, pp. 1–27.

Goljanek-Whysall, Katarzyna et al. (2011). "MicroRNA regulation of the paired-box transcription factor Pax3 confers robustness to developmental timing of myogenesis". In: *Proceedings of the National Academy of Sciences* 108.29, pp. 11936–11941.

Gong, Xiao-Qian et al. (2017). "PA-X protein decreases replication and pathogenicity of swine influenza virus in cultured cells and mouse models". In: *Veterinary microbiology* 205, pp. 66–70.

Graham, Robert R et al. (2007). "Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus". In: *Proceedings of the National Academy of Sciences* 104.16, pp. 6758–6763.

Grant, Charles E and Timothy L Bailey (2021). "XSTREME: Comprehensive motif analysis of biological sequence datasets". In: *BioRxiv*.

Grosso, Ana Rita, Sérgio Fernandes de Almeida, José Braga, and Maria Carmo-Fonseca (2012). "Dynamic transitions in RNA polymerase II density profiles during transcription termination". In: *Genome Research* 22.8, pp. 1447–1456.

Grozdanov, Petar N, Elahe Masoumzadeh, Michael P Latham, and Clinton C MacDonald (2018). "The structural basis of CstF-77 modulation of cleavage and polyadenylation through stimulation of CstF-64 activity". In: *Nucleic acids research* 46.22, pp. 12022–12039.

Gruber, Andreas J, Ralf Schmidt, et al. (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". In: *Genome research* 26.8, pp. 1145–1159.

Gruber, Andreas J and Mihaela Zavolan (2019). "Alternative cleavage and polyadenylation in health and disease". In: *Nature Reviews Genetics* 20.10, pp. 599–614.

Gruber, Andreas R, Georges Martin, Walter Keller, and Mihaiela Zavolan (2012). "Cleavage factor Im is a key regulator of 3 UTR length". In: *RNA biology* 9.12, pp. 1405–1412.

Gupta, Shobhit, John A Stamatoyannopoulos, Timothy L Bailey, and William Stafford Noble (2007). "Quantifying similarity between motifs". In: *Genome biology* 8.2, pp. 1–9.

Ha, Kevin CH, Benjamin J Blencowe, and Quaid Morris (2018). "QAPA: a new method for the systematic analysis of alternative polyadenylation from RNA-seq data". In: *Genome biology* 19.1, pp. 1–18.

Hale, Benjamin G, Richard E Randall, Juan Ortín, and David Jackson (2008). "The multifunctional NS1 protein of influenza A viruses". In: *Journal of general virology* 89.10, pp. 2359–2376.

Hale, Benjamin G, John Steel, et al. (2010). "Inefficient control of host gene expression by the 2009 pandemic H1N1 influenza A virus NS1 protein". In: *Journal of virology* 84.14, pp. 6909–6922.

Hamamoto, Itsuki, Hiroshi Takaku, Masato Tashiro, and Norio Yamamoto (2013). "High yield production of influenza virus in Madin Darby canine kidney (MDCK) cells with stable knockdown of IRF7". In: *PloS one* 8.3, e59892.

Hamilton, Keith, Yadong Sun, and Liang Tong (2019). "Biophysical characterizations of the recognition of the AAUAAA polyadenylation signal". In: *Rna* 25.12, pp. 1673–1680.

Hamilton, Keith and Liang Tong (2020). "Molecular mechanism for the interaction between human CPSF30 and hFip1". In: *Genes & development* 34.23-24, pp. 1753–1761.

Han, Julianna et al. (2021). "Host factor Rab11a is critical for efficient assembly of influenza A virus genomic segments". In: *PLoS pathogens* 17.5, e1009517.

Hayashi, Tsuyoshi, Chutikarn Chaimayo, James McGuinness, and Toru Takimoto (2016). "Critical role of the PA-X C-terminal domain of influenza A virus in its subcellular localization and shutoff activity". In: *Journal of virology* 90.16, pp. 7131–7141.

He, Tianqiong et al. (2020). "Host shutoff activity of VHS and SOX-like proteins: role in viral survival and immune evasion". In: *Virology journal* 17.1, pp. 1–11.

Heberle, Henry, Gabriela Vaz Meirelles, Felipe R da Silva, Guilherme P Telles, and Rosane Minghim (2015). "InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams". In: *BMC bioinformatics* 16.1, pp. 1–7.

Heinz, Sven et al. (2018). "Transcription elongation can affect genome 3D structure". In: *Cell* 174.6, pp. 1522–1536.

Herzel, Lydia, Diana SM Ottoz, Tara Alpert, and Karla M Neugebauer (2017). "Splicing and transcription touch base: co-transcriptional spliceosome assembly and function". In: *Nature reviews Molecular cell biology* 18.10, pp. 637–650.

Higgs, DR, SEY Goodbourn, J Lamb, JB Clegg, DJ Weatherall, and NJ Proudfoot (1983). "α-Thalassaemia caused by a polyadenylation signal mutation". In: *Nature* 306.5941, pp. 398–400.

Hilgers, Valérie, Sandra B Lemke, and Michael Levine (2012). "ELAV mediates 3 UTR extension in the Drosophila nervous system". In: *Genes & development* 26.20, pp. 2259–2264.

Hoffman, Yonit et al. (2016). "3'UTR shortening potentiates microRNA-based repression of pro-differentiation genes in proliferating human cells". In: *PLoS genetics* 12.2, e1005879.

Hofmann, Ilse, Martina Schnolzer, Isabelle Kaufmann, and Werner W Franke (2002). "Symplekin, a constitutive protein of karyo-and cytoplasmic particles involved in mRNA biogenesis in Xenopus laevis oocytes". In: *Molecular biology of the cell* 13.5, pp. 1665–1676.

Hoque, Mainul et al. (2013). "Analysis of alternative cleavage and polyadenylation by 3 region extraction and deep sequencing". In: *Nature methods* 10.2, pp. 133–139.

Hu, Jiao, Yiqun Mo, et al. (2015). "PA-X decreases the pathogenicity of highly pathogenic H5N1 influenza A virus in avian species by inhibiting virus replication and host response". In: *Journal of virology* 89.8, pp. 4126–4142.

Hu, JUN, Carol S Lutz, Jeffrey Wilusz, and BIN Tian (2005). "Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation". In: *Rna* 11.10, pp. 1485–1493.

Huang, Ruili, Ivan Grishagin, et al. (2019). "The NCATS BioPlanet–an integrated platform for exploring the universe of cellular signaling pathways for toxicology, systems biology, and chemical genomics". In: *Frontiers in pharmacology*, p. 445.

Huang, Yan, Wencheng Li, et al. (2012). "Mediator complex regulates alternative mRNA processing via the MED23 subunit". In: *Molecular cell* 45.4, pp. 459–469.

Hutchinson, Edward C and Ervin Fodor (2013). "Transport of the influenza virus genome from nucleus to nucleus". In: *Viruses* 5.10, pp. 2424–2446.

Hutchinson, Edward C, Johann C von Kirchbach, Julia R Gog, and Paul Digard (2010). "Genome packaging in influenza A virus". In: *Journal of general virology* 91.2, pp. 313–328.

Hwang, Hun-Way, Christopher Y Park, et al. (2016). "PAPERCLIP identifies microRNA targets and a role of CstF64/64tau in promoting non-canonical poly (A) site usage". In: *Cell reports* 15.2, pp. 423–435.

Hwang, Hun-Way, Yuhki Saito, et al. (2017). "cTag-PAPERCLIP reveals alternative polyadenylation promotes cell-type specific protein diversity and shifts Araf isoforms with microglia activation". In: *Neuron* 95.6, pp. 1334–1349.

Iwatsuki-Horimoto, Kiyoko, Taisuke Horimoto, Yutaka Fujii, and Yoshihiro Kawaoka (2004). "Generation of influenza A virus NS2 (NEP) mutants with an altered nuclear export signal sequence". In: *Journal of virology* 78.18, pp. 10149–10155.

Jagger, BW et al. (2012). "An overlapping protein-coding region in influenza A virus segment 3 modulates the host response". In: *Science* 337.6091, pp. 199–204.

Jan, Calvin H, Robin C Friedman, J Graham Ruby, and David P Bartel (2011). "Formation, regulation and evolution of Caenorhabditis elegans 3 UTRs". In: *Nature* 469.7328, pp. 97–101.

Javanian, Mostafa, Mohammad Barary, Sam Ghebrehewet, Veerendra Koppolu, VeneelaKrishnaRekha Vasigala, and Soheil Ebrahimpour (2021). "A brief review of influenza virus infection". In: *Journal of Medical Virology* 93.8, pp. 4638–4646.

Jenal, Mathias et al. (2012). "The poly (A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites". In: *Cell* 149.3, pp. 538–553.

Ji, Zhe, Ju Youn Lee, Zhenhua Pan, Bingjun Jiang, and Bin Tian (2009). "Progressive lengthening of 3 untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development". In: *Proceedings of the National Academy of Sciences* 106.17, pp. 7028–7033.

Ji, Zhe, Wenting Luo, et al. (2011). "Transcriptional activity regulates alternative cleavage and polyadenylation". In: *Molecular systems biology* 7.1, p. 534.

Ji, Zhe and Bin Tian (2009). "Reprogramming of 3 untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types". In: *PloS one* 4.12, e8419.

Ji, Zhu-xing, Xiao-quan Wang, and Xiu-fan Liu (2021). "NS1: A Key Protein in the "Game" Between Influenza A Virus and Host in Innate Immunity". In: *Frontiers in Cellular and Infection Microbiology* 11, p. 638.

Jia, Danlin, Ramtin Rahbar, et al. (2010). "Influenza virus non-structural protein 1 (NS1) disrupts interferon signaling". In: *PloS one* 5.11, e13927.

Jia, Jinbu, Yanping Long, et al. (2020). "Post-transcriptional splicing of nascent RNA contributes to widespread intron retention in plants". In: *Nature Plants* 6.7, pp. 780–788.

Jia, Xin, Shaochun Yuan, et al. (2017). "The role of alternative polyadenylation in the antiviral innate immune response". In: *Nature communications* 8.1, pp. 1–12.

Jobbins, Andrew M et al. (2022). "Dysregulated RNA polyadenylation contributes to metabolic impairment in non-alcoholic fatty liver disease". In: *Nucleic acids research* 50.6, pp. 3379–3393.

Johnson, Britney et al. (2018). "Human IFIT3 modulates IFIT1 RNA binding specificity and protein stability". In: *Immunity* 48.3, pp. 487–499.

Kaczmarek Michaels, Katarzyna, Salwa Mohd Mostafa, Julia Ruiz Capella, and Claire L Moore (2020). "Regulation of alternative polyadenylation in the yeast Saccharomyces cerevisiae by histone H3K4 and H3K36 methyltransferases". In: *Nucleic acids research* 48.10, pp. 5407–5425.

Kaida, Daisuke et al. (2010). "U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation". In: *nature* 468.7324, pp. 664–668.

Kambach, Christian, Stefan Walket, and Kiyoshi Nagai (1999). "Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles". In: *Current opinion in structural biology* 9.2, pp. 222–230.

Kao, Richard Y et al. (2010). "Identification of influenza A nucleoprotein as an antiviral target". In: *Nature biotechnology* 28.6, pp. 600–605.

Kastner, Berthold, Cindy L Will, Holger Stark, and Reinhard Lührmann (2019). "Structural insights into nuclear pre-mRNA splicing in higher eukaryotes". In: *Cold Spring Harbor perspectives in biology* 11.11, a032417.

Kaufmann, Isabelle, Georges Martin, Arno Friedlein, Hanno Langen, and Walter Keller (2004). "Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly (A) polymerase". In: *The EMBO journal* 23.3, pp. 616–626.

Keller, Rebecca W, Uwe Kuhn, Mateo Aragon, Larissa Bornikova, Elmar Wahle, and David G Bear (2000). "The nuclear poly (A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly (A) tail". In: *Journal of molecular biology* 297.3, pp. 569–583.

Keller, Walter, Silke Bienroth, Katharine M Lang, and Gerhard Christofori (1991). "Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3 processing signal AAUAAA." In: *The EMBO journal* 10.13, pp. 4241–4249.

Kerns, Julie A, Michael Emerman, and Harmit S Malik (2008). "Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein". In: *PLoS genetics* 4.1, e21.

Kerwitz, Yvonne et al. (2003). "Stimulation of poly (A) polymerase through a direct interaction with the nuclear poly (A) binding protein allosterically regulated by RNA". In: *The EMBO journal* 22.14, pp. 3705–3714.

Khaladkar, Mugdha, Mark Smyda, and Sridhar Hannenhalli (2011). "Epigenomic and RNA structural correlates of polyadenylation". In: *RNA biology* 8.3, pp. 529–537.

Khaperskyy, Denys A, Mohamed M Emara, Benjamin P Johnston, Paul Anderson, Todd F Hatchette, and Craig McCormick (2014). "Influenza a virus host shutoff disables antiviral stress-induced translation arrest". In: *PLoS pathogens* 10.7, e1004217.

Khaperskyy, Denys A, Summer Schmaling, Jonah Larkins-Ford, Craig McCormick, and Marta M Gaglia (2016). "Selective degradation of host RNA polymerase II transcripts by influenza A virus PA-X host shutoff protein". In: *PLoS pathogens* 12.2, e1005427.

Killip, Marian J, Ervin Fodor, and Richard E Randall (2015). "Influenza virus activation of the interferon system". In: *Virus research* 209, pp. 11–22.

Kristjánsdóttir, Katla, Elizabeth A Fogarty, and Andrew Grimson (2015). "Systematic analysis of the Hmga2 3 UTR identifies many independent regulatory sequences and a novel interaction between distal sites". In: *Rna* 21.7, pp. 1346–1360.

Krug, Robert M, Barbara A Broni, and Michele Bouloy (1979). "Are the 5 ends of influenza viral mRNAs synthesized in vivo donated by host mRNAs?" In: *Cell* 18.2, pp. 329–334.

Kuhn, Uwe, Miriam Gündel, Anne Knoth, Yvonne Kerwitz, Sabine Rüdel, and Elmar Wahle (2009). "Poly (A) tail length is controlled by the nuclear poly (A)-

binding protein regulating the interaction between poly (A) polymerase and the cleavage and polyadenylation specificity factor". In: *Journal of Biological Chemistry* 284.34, pp. 22803–22814.

Kuleshov, Maxim V et al. (2016). "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update". In: *Nucleic acids research* 44.W1, W90–W97.

Kumar, Ananthanarayanan, WH Conny, Juan B Rodríguez-Molina, Xiao-Han Li, Stefan MV Freund, and Lori A Passmore (2021). "Dynamics in Fip1 regulate eukaryotic mRNA 3 end processing". In: *Genes & development* 35.21-22, pp. 1510–1526.

Laborda, Pedro, Su-Yan Wang, and Josef Voglmeir (2016). "Influenza neuraminidase inhibitors: synthetic approaches, derivatives and biological activity". In: *Molecules* 21.11, p. 1513.

Lackford, Brad et al. (2014). "Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal". In: *The EMBO journal* 33.8, pp. 878–889.

Lafontaine, Denis LJ (2015). "Noncoding RNAs in eukaryotic ribosome biogenesis and function". In: *Nature structural & molecular biology* 22.1, pp. 11–19.

LaForce, Geneva R et al. (2022). "Suppression of premature transcription termination leads to reduced mRNA isoform diversity and neurodegeneration". In: *Neuron*.

Lakdawala, Seema S et al. (2014). "Influenza a virus assembly intermediates fuse in the cytoplasm". In: *PLoS pathogens* 10.3, e1003971.

Lamb, Robert A and Purnell W Choppin (1981). "Identification of a second protein (M2) encoded by RNA segment 7 of influenza virus". In: *Virology* 112.2, pp. 729–737.

Le Sage, Valerie, Jack P Kanarek, Dan J Snyder, Vaughn S Cooper, Seema S Lakdawala, and Nara Lee (2020). "Mapping of influenza virus RNA-RNA interactions reveals a flexible network". In: *Cell reports* 31.13, p. 107823.

Lee, Nara, Valerie Le Sage, Adalena V Nanni, Dan J Snyder, Vaughn S Cooper, and Seema S Lakdawala (2017). "Genome-wide analysis of influenza viral RNA and nucleoprotein association". In: *Nucleic acids research* 45.15, pp. 8968–8977.

Lee, Seungjae, Lu Wei, et al. (2021). "ELAV/Hu RNA binding proteins determine multiple programs of neural alternative splicing". In: *PLoS genetics* 17.4, e1009439.

Lee, Shih-Han and Christine Mayr (2019). "Gain of additional BIRC3 protein functions through 3-UTR-mediated protein complex formation". In: *Molecular cell* 74.4, pp. 701–712.

Lee, Shih-Han, Irtisha Singh, Sarah Tisdale, Omar Abdel-Wahab, Christina S Leslie, and Christine Mayr (2018). "Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia". In: *Nature* 561.7721, pp. 127–131.

Legendre, Matthieu and Daniel Gautheret (2003). "Sequence determinants in human polyadenylation site selection". In: *BMC genomics* 4.1, pp. 1–9.

Legrand, Pierre, Noel Pinaud, Lionel Minvielle-Sebastia, and Sebastien Fribourg (2007). "The structure of the CstF-77 homodimer provides insights into CstF assembly". In: *Nucleic acids research* 35.13, pp. 4515–4522.

Levene, Rachel Emily, Shailab D Shrestha, and Marta Maria Gaglia (2021). "The influenza A virus host shutoff factor PA-X is rapidly turned over in a strain-specific manner". In: *Journal of Virology* 95.8, e02312–20.

Li, Heng, Shuilong Tong, et al. (2011). "Structural basis of pre-mRNA recognition by the human cleavage factor Im complex". In: *Cell research* 21.7, pp. 1039–1051.

Li, Mai, Wenbao Qi, et al. (2021). "Influenza A virus protein PA-X suppresses host Ankrd17-mediated immune responses". In: *Microbiology and Immunology* 65.1, pp. 48–59.

Li, Wei Vivian, Dinghai Zheng, Ruijia Wang, and Bin Tian (2021). "MAAPER: model-based analysis of alternative polyadenylation using 3 end-linked reads". In: *Genome biology* 22.1, pp. 1–21.

Li, Wencheng, Ji Yeon Park, Dinghai Zheng, Mainul Hoque, Ghassan Yehia, and Bin Tian (2016). "Alternative cleavage and polyadenylation in spermatogenesis connects chromatin regulation with post-transcriptional control". In: *BMC biology* 14.1, pp. 1–17.

Li, Wencheng, Bei You, et al. (2015). "Systematic profiling of poly (A)+ transcripts modulated by core 3'end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation". In: *PLoS genetics* 11.4, e1005166.

Li, Xiuli, Min Gu, Qinmei Zheng, Ruyi Gao, and Xiufan Liu (2021). "Packaging signal of influenza A virus". In: *Virology Journal* 18.1, pp. 1–10.

Li, Ying-Qing, Yang Chen, et al. (2020). "FNDC3B 3-UTR shortening escapes from microRNA-mediated gene repression and promotes nasopharyngeal carcinoma progression". In: *Cancer Science* 111.6, pp. 1991–2003.

Liang, Ruibin, Jessica MJ Swanson, Jesper J Madsen, Mei Hong, William F DeGrado, and Gregory A Voth (2016). "Acid activation mechanism of the influenza A M2 proton channel". In: *Proceedings of the National Academy of Sciences* 113.45, E6955–E6964.

Lianoglou, Steve, Vidur Garg, Julie L Yang, Christina S Leslie, and Christine Mayr (2013). "Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression". In: *Genes & development* 27.21, pp. 2380–2396.

Licatalosi, Donny D et al. (2008). "HITS-CLIP yields genome-wide insights into brain alternative RNA processing". In: *Nature* 456.7221, pp. 464–469.

Lima Neto, Quirino Alves de et al. (2016). "Structural and functional analysis of four non-coding Y RNAs from Chinese hamster cells: identification, molecular dynamics simulations and DNA replication initiation assays". In: *BMC molecular biology* 17.1, pp. 1–9.

Lin, DONG, SE Gitelman, P Saenger, WL Miller, et al. (1991). "Normal genes for the cholesterol side chain cleavage enzyme, P450scc, in congenital lipoid adrenal hyperplasia." In: *The Journal of clinical investigation* 88.6, pp. 1955–1962.

Lin, Yuefeng, Zhihua Li, et al. (2012). "An in-depth map of polyadenylation sites in cancer". In: *Nucleic acids research* 40.17, pp. 8460–8471.

Liu, Chien-Hung, Ligang Zhou, Guifang Chen, and Robert M Krug (2015). "Battle between influenza A virus and a newly identified antiviral activity of the PARP-containing ZAPL protein". In: *Proceedings of the National Academy of Sciences* 112.45, pp. 14048–14053.

Liu, H and Megerditch Kiledjian (2006). "Decapping the message: a beginning or an end". In: *Biochemical Society Transactions* 34.1, pp. 35–38.

Liu, Huiyun and Claire L Moore (2021). "On the cutting edge: regulation and therapeutic potential of the mRNA 3 end nuclease". In: *Trends in Biochemical Sciences* 46.9, pp. 772–784.

Long, Jason S et al. (2016). "Species difference in ANP32A underlies influenza A virus polymerase host restriction". In: *Nature* 529.7584, pp. 101–104.

López-Denman, Adam J, Daniel E Tuipulotu, Jessica B Ross, Alice M Trenerry, Peter A White, and Jason M Mackenzie (2021). "Nuclear localisation of West Nile virus NS5 protein modulates host gene expression". In: *Virology* 559, pp. 131–144.

Love, Michael I, Wolfgang Huber, and Simon Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome biology* 15.12, pp. 1–21.

Ludwig, Stephan, Christina Ehrhardt, Elisabeth R Neumeier, Michael Kracht, Ulf R Rapp, and Stephan Pleschka (2001). "Influenza virus-induced AP-1-dependent gene expression requires activation of the JNK signaling pathway". In: *Journal of Biological Chemistry* 276.14, pp. 10990–10998.

Luo, Kaili, Jingwen Geng, et al. (2019). "LncRNA CASC9 interacts with CPSF3 to regulate TGF-β signaling in colorectal cancer". In: *Journal of Experimental & Clinical Cancer Research* 38.1, pp. 1–16.

Luo, Wenting, Zhe Ji, et al. (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". In: *PLoS genetics* 9.7, e1003613.

Lyles, Douglas S (2000). "Cytopathogenesis and inhibition of host gene expression by RNA viruses". In: *Microbiology and molecular biology reviews* 64.4, pp. 709–724.

MacDonald, Clinton C, Jeffrey Wilusz, and Thomas Shenk (1994). "The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream

of the cleavage site and influences cleavage site location". In: *Molecular and cellular biology* 14.10, pp. 6647–6654.

Mandel, Corey R et al. (2006). "Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease". In: *Nature* 444.7121, pp. 953–956.

Mansfield, Kyle D and Jack D Keene (2012). "Neuron-specific ELAV/Hu proteins suppress HuR mRNA during neuronal differentiation by alternative polyadenylation". In: *Nucleic acids research* 40.6, pp. 2734–2746.

Martin, Brigitte E, Jeremy D Harris, Jiayi Sun, Katia Koelle, and Christopher B Brooke (2020). "Cellular co-infection can modulate the efficiency of influenza A virus production and shape the interferon response". In: *PLoS pathogens* 16.10, e1008974.

Martin, Georges, Andreas R Gruber, Walter Keller, and Mihaela Zavolan (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". In: *Cell reports* 1.6, pp. 753–763.

Martin, Itziar Serna, Narin Hengrung, et al. (2018). "A mechanism for the activation of the influenza virus transcriptase". In: *Molecular cell* 70.6, pp. 1101–1110.

Marzluff, William F, Eric J Wagner, and Robert J Duronio (2008). "Metabolism and regulation of canonical histone mRNAs: life without a poly (A) tail". In: *Nature Reviews Genetics* 9.11, pp. 843–854.

Masamha, Chioniso P and Eric J Wagner (2018). "The contribution of alternative polyadenylation to the cancer phenotype". In: *Carcinogenesis* 39.1, pp. 2–10.

Masamha, Chioniso P, Zheng Xia, et al. (2014). "CFIm25 links alternative polyadenylation to glioblastoma tumour suppression". In: *Nature* 510.7505, pp. 412–416.

Mayr, Christine (2019). "What are 3 UTRs doing?" In: *Cold Spring Harbor Perspectives in Biology* 11.10, a034728.

Mayr, Christine and David P Bartel (2009). "Widespread shortening of 3 UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells". In: *Cell* 138.4, pp. 673–684.

Mazel-Sanchez, Beryl et al. (2021). "Influenza A viruses balance ER stress with host protein synthesis shutoff". In: *Proceedings of the National Academy of Sciences* 118.36.

McCracken, Susan et al. (1997). "The C-terminal domain of RNA polymerase II couples mRNA processing to transcription". In: *Nature* 385.6614, pp. 357–361.

McDougall, William M et al. (2019). "Viral infection or IFN-α alters mitotic spindle orientation by modulating Pericentrin levels". In: *Iscience* 12, pp. 270–279.

Metze, Stefanie, Veronika A Herzog, Marc-David Ruepp, and Oliver Mühlemann (2013). "Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways". In: *Rna* 19.10, pp. 1432–1448.

Mikula, Michal, Karol Bomsztyk, Krzysztof Goryca, Krzysztof Chojnowski, and Jerzy Ostrowski (2013). "Heterogeneous nuclear ribonucleoprotein (HnRNP) K genome-wide binding survey reveals its role in regulating 3-end RNA processing and transcription termination at the early growth response 1 (EGR1) gene through XRN2 exonuclease". In: *Journal of Biological Chemistry* 288.34, pp. 24788–24798.

Miyake, Yasuyuki et al. (2019). "Influenza virus uses transportin 1 for vRNP debundling during cell entry". In: *Nature microbiology* 4.4, pp. 578–586.

Molinari, Noelle-Angelique M et al. (2007). "The annual impact of seasonal influenza in the US: measuring disease burden and costs". In: *Vaccine* 25.27, pp. 5086–5096.

Monarez, Roberto R, Clinton C MacDonald, and Brinda Dass (2007). "Polyadenylation proteins CstF-64 and τCstF-64 exhibit differential binding affinities for RNA polymers". In: *Biochemical Journal* 401.3, pp. 651–658.

Montes, Matias, Brianne L Sanford, Daniel F Comiskey, and Dawn S Chandler (2019). "RNA splicing and disease: animal models to therapies". In: *Trends in Genetics* 35.1, pp. 68–87.

Moore, Melissa J and Nick J Proudfoot (2009). "Pre-mRNA processing reaches back totranscription and ahead to translation". In: *Cell* 136.4, pp. 688–700.

Moreira, Etori Aguiar, Yohei Yamauchi, and Patrick Matthias (2021). "How Influenza Virus Uses Host Cell Pathways during Uncoating". In: *Cells* 10.7, p. 1722.

Morree, Antoine de et al. (2019). "Alternative polyadenylation of Pax3 controls muscle stem cell fate and muscle function". In: *Science* 366.6466, pp. 734–738.

Müller-McNicoll, Michaela et al. (2016). "SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export". In: *Genes & development* 30.5, pp. 553–566.

Nacken, Wolfgang, André Schreiber, Dörthe Masemann, and Stephan Ludwig (2021). "The Effector Domain of the Influenza A Virus Nonstructural Protein NS1 Triggers Host Shutoff by Mediating Inhibition and Global Deregulation of Host Transcription When Associated with Specific Structures in the Nucleus". In: *Mbio* 12.5, e02196–21.

Naganuma, Takao, Shinichi Nakagawa, Akie Tanigawa, Yasnory F Sasaki, Naoki Goshima, and Tetsuro Hirose (2012). "Alternative 3-end processing of long noncoding RNA initiates construction of nuclear paraspeckles". In: *The EMBO journal* 31.20, pp. 4020–4034.

Nam, Jin-Wu et al. (2014). "Global analyses of the effect of different cellular contexts on microRNA targeting". In: *Molecular cell* 53.6, pp. 1031–1043.

Nanavaty, Vishal et al. (2020). "DNA methylation regulates alternative polyadenylation via CTCF and the cohesin complex". In: *Molecular cell* 78.4, pp. 752–764.

Nath, Swapan K, Jennifer A Kelly, John B Harley, and R Hal Scofield (2004). "Mapping the Systemic Lupus Erythematosus Susceptibility Genes". In: *Autoimmunity*, pp. 11–29.

Nelson, Meryl R, Andrew M Leidal, and Craig A Smibert (2004). "Drosophila Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression". In: *The EMBO journal* 23.1, pp. 150–159.

Nemeroff, Martin E, Silvia ML Barabino, Yongzhong Li, Walter Keller, and Robert M Krug (1998). "Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3 end formation of cellular pre-mRNAs". In: *Molecular cell* 1.7, pp. 991–1000.

Neve, Jonathan, Kaspar Burger, et al. (2016). "Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation". In: *Genome research* 26.1, pp. 24–35.

Neve, Jonathan, Radhika Patel, Zhiqiao Wang, Alastair Louey, and André Martin Furger (2017). "Cleavage and polyadenylation: Ending the message expands gene regulation". In: *RNA biology* 14.7, pp. 865–890.

Nilsson, Kersti, Samir Abdurahman, and Stefan Schwartz (2020). "Influenza virus natural sequence heterogeneity in segment 8 affects interactions with cellular RNA-binding proteins and splicing efficiency". In: *Virology* 549, pp. 39–50.

Ning, Yue, Wanxia Liu, Xiaoying Guan, Xiaobin Xie, and Yajie Zhang (2019). "CPSF3 is a promising prognostic biomarker and predicts recurrence of non-small cell lung cancer". In: *Oncology Letters* 18.3, pp. 2835–2844.

Niwa, Maho, Scott D Rose, and Susan M Berget (1990). "In vitro polyadenylation is stimulated by the presence of an upstream intron." In: *Genes & development* 4.9, pp. 1552–1559.

Nogales, Aitor, Luis Martinez-Sobrido, Kevin Chiem, David J Topham, and Marta L DeDiego (2018). "Functional evolution of the 2009 pandemic H1N1 influenza virus NS1 and PA in humans". In: *Journal of Virology* 92.19, e01206–18.

Nogales, Aitor, Luis Martinez-Sobrido, David J Topham, and Marta L DeDiego (2018). "Modulation of innate immune responses by the influenza A NS1 and PA-X proteins". In: *Viruses* 10.12, p. 708.

Noton, Sarah L et al. (2009). "Studies of an influenza A virus temperature-sensitive mutant identify a late role for NP in the formation of infectious virions". In: *Journal of virology* 83.2, pp. 562–571.

Nunes, Nuno Miguel, Wencheng Li, Bin Tian, and Andre Furger (2010). "A functional human Poly (A) site requires only a potent DSE and an A-rich upstream sequence". In: *The EMBO journal* 29.9, pp. 1523–1536.

Ogorodnikov, Anton et al. (2018). "Transcriptome 3 end organization by PCF11 links alternative polyadenylation to formation and neuronal differentiation of neuroblastoma". In: *Nature communications* 9.1, pp. 1–16.

Oh, Jung-Min, Chao Di, et al. (2017). "U1 snRNP telescripting regulates a size–function-stratified human genome". In: *Nature structural & molecular biology* 24.11, pp. 993–999.

Oh, Jung-Min, Christopher C Venters, et al. (2020). "U1 snRNP regulates cancer cell migration and invasion in vitro". In: *Nature communications* 11.1, pp. 1–8.

Ohno, Mutsuhito, Hiroshi Sakamoto, and Yoshiro Shimura (1987). "Preferential excision of the 5'proximal intron from mRNA precursors with two introns as mediated by the cap structure". In: *Proceedings of the National Academy of Sciences* 84.15, pp. 5187–5191.

Oikonomou, Panos, Hani Goodarzi, and Saeed Tavazoie (2014). "Systematic identification of regulatory elements in conserved 3 UTRs of human transcripts". In: *Cell reports* 7.1, pp. 281–292.

Oishi, Kohei, Seiya Yamayoshi, and Yoshihiro Kawaoka (2015). "Mapping of a region of the PA-X protein of influenza A virus that is important for its shutoff activity". In: *Journal of virology* 89.16, pp. 8661–8665.

Orkin, Stuart H, TC Cheng, SE Antonarakis, and HH Kazazian Jr (1985). "Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene." In: *The EMBO journal* 4.2, pp. 453–456.

Pan, Zhenhua, Haibo Zhang, Lisa K Hague, Ju Youn Lee, Carol S Lutz, and Bin Tian (2006). "An intronic polyadenylation site in human and mouse CstF-77 genes suggests an evolutionarily conserved regulatory mechanism". In: *Gene* 366.2, pp. 325–334.

Passmore, Lori A and Jeff Coller (2021). "Roles of mRNA poly (A) tails in regulation of eukaryotic gene expression". In: *Nature Reviews Molecular Cell Biology*, pp. 1–14.

Passmore, Lori A and Jeff Coller (2022). "Roles of mRNA poly (A) tails in regulation of eukaryotic gene expression". In: *Nature Reviews Molecular Cell Biology* 23.2, pp. 93–106.

Paterson, Duncan and Ervin Fodor (2012). "Emerging roles for the influenza A virus nuclear export protein (NEP)". In: *PLoS pathogens* 8.12, e1003019.

Patro, Rob, Geet Duggal, Michael I Love, Rafael A Irizarry, and Carl Kingsford (2017). "Salmon provides fast and bias-aware quantification of transcript expression". In: *Nature methods* 14.4, pp. 417–419.

Pereira-Castro, I and A Moreira (2022). "On the function and relevance of alternative 3'-UTRs in gene expression regulation (vol 125, e1653, 2021)". In: *WILEY INTERDISCIPLINARY REVIEWS-RNA* 13.1.

Pereira-Castro, Isabel, Beatriz C Garcia, et al. (2022). "MCL1 alternative polyadenylation is essential for cell survival and mitochondria morphology". In: *Cellular and Molecular Life Sciences* 79.3, pp. 1–25.

Peukes, Julia et al. (2020). "The native structure of the assembled matrix protein 1 of influenza A virus". In: *Nature* 587.7834, pp. 495–498.

Phillips, Jennifer E and Victor G Corces (2009). "CTCF: master weaver of the genome". In: *Cell* 137.7, pp. 1194–1211.

Pinto, Lawrence H and Robert A Lamb (2006). "The M2 proton channels of influenza A and B viruses". In: *Journal of Biological Chemistry* 281.14, pp. 8997–9000.

Pinto, Pedro AB, Telmo Henriques, et al. (2011). "RNA polymerase II kinetics in polo polyadenylation signal selection". In: *The EMBO journal* 30.12, pp. 2431–2444.

Polymeropoulos, Mihael H et al. (1997). "Mutation in the α-synuclein gene identified in families with Parkinson's disease". In: *science* 276.5321, pp. 2045–2047.

Poon, Leo LM, David C Pritlove, Ervin Fodor, and George G Brownlee (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". In: *Journal of virology* 73.4, pp. 3473–3476.

Proudfoot, Nick J, Andre Furger, and Michael J Dye (2002). "Integrating mRNA processing with transcription". In: *Cell* 108.4, pp. 501–512.

Prudencio, Mercedes et al. (2015). "Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS". In: *Nature neuroscience* 18.8, pp. 1175–1182.

Qian, Wei et al. (2017). "The C-terminal effector domain of non-structural protein 1 of influenza A virus blocks IFN-β production by targeting TNF receptor-associated factor 3". In: *Frontiers in immunology* 8, p. 779.

Raabe, Tobias, FJ Bollum, and James L Manley (1991). "Primary structure and expression of bovine poly (A) polymerase". In: *Nature* 353.6341, pp. 229–234.

Ramanathan, Anand, G Brett Robb, and Siu-Hong Chan (2016). "mRNA capping: biological functions and applications". In: *Nucleic acids research* 44.16, pp. 7511–7526.

Rambout, Xavier and Lynne E Maquat (2020). "The nuclear cap-binding complex as choreographer of gene transcription and pre-mRNA processing". In: *Genes & development* 34.17-18, pp. 1113–1127.

Rehfeld, Anders et al. (2014). "Alternative polyadenylation of tumor suppressor genes in small intestinal neuroendocrine tumors". In: *Frontiers in endocrinology* 5, p. 46.

Reich, Stefan et al. (2014). "Structural insight into cap-snatching and RNA synthesis by influenza polymerase". In: *Nature* 516.7531, pp. 361–366.

Ren, Zhihua, Ting Ding, Zhicai Zuo, Zhiwen Xu, Junliang Deng, and Zhanyong Wei (2020). "Regulation of MAVS expression and signaling function in the antiviral innate immune response". In: *Frontiers in immunology* 11, p. 1030.

Reyes, Alejandro and Wolfgang Huber (2018). "Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues". In: *Nucleic acids research* 46.2, pp. 582–592.

Rialdi, Alexander et al. (2017). "The RNA exosome syncs IAV-RNAPII transcription to promote viral ribogenesis and infectivity". In: *Cell* 169.4, pp. 679–692.

Richard, Patricia and James L Manley (2009). "Transcription termination by nuclear RNA polymerases". In: *Genes & development* 23.11, pp. 1247–1269.

Roberts, Kari L, George P Leser, Chunlong Ma, and Robert A Lamb (2013). "The amphipathic helix of influenza A virus M2 protein is required for filamentous bud formation and scission of filamentous and spherical particles". In: *Journal of virology* 87.18, pp. 9973–9982.

Robertson, JAMES S, MANFRED Schubert, and ROBERT A Lazzarini (1981). "Polyadenylation sites for influenza virus mRNA". In: *Journal of virology* 38.1, pp. 157–163.

Rogers, J et al. (1980). "Two mRNAs with different 3 ends encode membrane-bound and secreted forms of immunoglobulin μ chain". In: *Cell* 20.2, pp. 303–312.

Rozenblatt-Rosen, Orit et al. (2009). "The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3 mRNA processing factors". In: *Proceedings of the National Academy of Sciences* 106.3, pp. 755–760.

Rüegsegger, Ursula, Katrin Beyer, and Walter Keller (1996). "Purification and Characterization of Human Cleavage Factor Im Involved in the 3 End Processing of Messenger RNA Precursors ()". In: *Journal of Biological Chemistry* 271.11, pp. 6107–6113.

Ruigrok, Rob WH, Thibaut Crépin, Darren J Hart, and Stephen Cusack (2010). "Towards an atomic resolution understanding of the influenza virus replication machinery". In: *Current opinion in structural biology* 20.1, pp. 104–113.

Rund, Deborah, Carol Dowling, Khamis Najjar, Eliezer A Rachmilewitz, Haig H Kazazian, and Ariella Oppenheim (1992). "Two mutations in the beta-globin polyadenylylation signal reveal extended transcripts and new RNA polyadenylylation sites." In: *Proceedings of the National Academy of Sciences* 89.10, pp. 4324–4328.

Rust, Michael J, Melike Lakadamyali, Feng Zhang, and Xiaowei Zhuang (2004). "Assembly of endocytic machinery around individual influenza viruses during viral entry". In: *Nature structural & molecular biology* 11.6, pp. 567–573.

Sakharkar, Meena Kishore, Vincent TK Chow, and Pandjassarame Kangueane (2004). "Distributions of exons and introns in the human genome". In: *In silico biology* 4.4, pp. 387–393.

Salzberg, Steven L (2018). "Open questions: How many genes do we have?" In: *BMC biology* 16.1, pp. 1–3.

Samji, Tasleem (2009). "Influenza A: understanding the viral life cycle". In: *The Yale journal of biology and medicine* 82.4, p. 153.

Samuels, Tamsin J et al. (2020). "Neuronal upregulation of Prospero protein is driven by alternative mRNA polyadenylation and Syncrip-mediated mRNA stabilisation". In: *Biology open* 9.5, bio049684.

Sandberg, Rickard, Joel R Neilson, Arup Sarma, Phillip A Sharp, and Christopher B Burge (2008). "Proliferating cells express mRNAs with shortened 3'untranslated regions and fewer microRNA target sites". In: *Science* 320.5883, pp. 1643–1647.

Satterly, Neal et al. (2007). "Influenza virus targets the mRNA export machinery and the nuclear pore complex". In: *Proceedings of the National Academy of Sciences* 104.6, pp. 1853–1858.

Schäfer, Peter et al. (2018). "Reconstitution of mammalian cleavage factor II involved in 3 processing of mRNA precursors". In: *Rna* 24.12, pp. 1721–1737.

Schlautmann, Lena P and Niels H Gehring (2020). "A day in the life of the exon junction complex". In: *Biomolecules* 10.6, p. 866.

Schwerk, Johannes et al. (2019). "RNA-binding protein isoforms ZAP-S and ZAP-L have distinct antiviral and immune resolution functions". In: *Nature immunology* 20.12, pp. 1610–1620.

Schwich, Oliver Daniel et al. (2021). "SRSF3 and SRSF7 modulate 3 UTR length through suppression or activation of proximal polyadenylation sites and regulation of CFIm levels". In: *Genome biology* 22.1, pp. 1–34.

Shah, Ankeeta, Briana E Mittleman, Yoav Gilad, and Yang I Li (2021). "Benchmarking sequencing methods and tools that facilitate the study of alternative polyadenylation". In: *Genome biology* 22.1, pp. 1–21.

Shaw, Megan and Peter Palese (Jan. 2007). "Orthomyxoviridae: The viruses and their replication". In: *Fields Virology*, pp. 1647–1689.

Shen, Ting et al. (2019). "Alternative polyadenylation dependent function of splicing factor SRSF3 contributes to cellular senescence". In: *Aging (Albany NY)* 11.5, p. 1356.

Shepard, Peter J and Klemens J Hertel (2009). "The SR protein family". In: *Genome biology* 10.10, pp. 1–9.

Shepardson, Kelly M et al. (2018). "IFNAR2 is required for anti-influenza immunity and alters susceptibility to post-influenza bacterial superinfections". In: *Frontiers in immunology*, p. 2589.

Shi, Mang, Brett W Jagger, Helen M Wise, Paul Digard, Edward C Holmes, and Jeffery K Taubenberger (2012). "Evolutionary conservation of the PA-X open reading frame in segment 3 of influenza A virus". In: *Journal of virology* 86.22, pp. 12411–12413.

Shi, Yongsheng (2012). "Alternative polyadenylation: new insights from global analyses". In: *Rna* 18.12, pp. 2105–2117.

Shulman, Eldad David and Ran Elkon (2019). "Cell-type-specific analysis of alternative polyadenylation using single-cell transcriptomics data". In: *Nucleic acids research* 47.19, pp. 10027–10039.

Singh, Irtisha, Shih-Han Lee, et al. (2018). "Widespread intronic polyadenylation diversifies immune cell transcriptomes". In: *Nature communications* 9.1, pp. 1–16.

Singh, Priyam, Travis L Alley, et al. (2009). "Global changes in processing of mRNA 3 untranslated regions characterize clinically distinct cancer subtypes". In: *Cancer research* 69.24, pp. 9422–9430.

Song, Yinna et al. (2021). "RBM39 alters phosphorylation of c-Jun and binds to viral RNA to promote PRRSV proliferation". In: *Frontiers in immunology* 12.

Sonoiki, Ebere et al. (2017). "A potent antimalarial benzoxaborole targets a Plasmodium falciparum cleavage and polyadenylation specificity factor homologue". In: *Nature communications* 8.1, pp. 1–11.

Spies, Noah, Cydney B Nielsen, Richard A Padgett, and Christopher B Burge (2009). "Biased chromatin signatures around polyadenylation sites and exons". In: *Molecular cell* 36.2, pp. 245–254.

Staller, Ecco et al. (2019). "ANP32 proteins are essential for influenza virus replication in human cells". In: *Journal of virology* 93.17, e00217–19.

Stanley, Robert F and Omar Abdel-Wahab (2022). "Dysregulation and therapeutic targeting of RNA splicing in cancer". In: *Nature Cancer* 3.5, pp. 536–546.

Subtelny, Alexander O, Stephen W Eichhorn, Grace R Chen, Hazel Sive, and David P Bartel (2014). "Poly (A)-tail profiling reveals an embryonic switch in translational control". In: *Nature* 508.7494, pp. 66–71.

Sun, Jiya, Fei Ye, et al. (2020). "Comparative transcriptome analysis reveals the intensive early stage responses of host cells to SARS-CoV-2 infection". In: *Frontiers in microbiology*, p. 2881.

Sun, Yadong, Keith Hamilton, and Liang Tong (2020). "Recent molecular insights into canonical pre-mRNA 3'-end processing". In: *Transcription* 11.2, pp. 83–96.

Sun, Yadong, Yixiao Zhang, et al. (2018). "Molecular basis for the recognition of the human AAUAAA polyadenylation signal". In: *Proceedings of the National Academy of Sciences* 115.7, E1419–E1428.

Supek, Fran, Ben Lehner, and Rik GH Lindeboom (2021). "To NMD or Not To NMD: nonsense-mediated mRNA decay in cancer and other genetic diseases". In: *Trends in Genetics* 37.7, pp. 657–668.

Takagaki, Yoshio and James L Manley (2000). "Complex protein interactions within the human polyadenylation machinery identify a novel component". In: *Molecular and cellular biology* 20.5, pp. 1515–1525.

Takagaki, Yoshio, James L Manley, Clinton C MacDonald, Jeffrey Wilusz, and Thomas Shenk (1990). "A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs." In: *Genes & Development* 4.12a, pp. 2112–2120.

Takagaki, Yoshio, Lisa C Ryner, and James L Manley (1989). "Four factors are required for 3'-end cleavage of pre-mRNAs." In: *Genes & development* 3.11, pp. 1711–1724.

Takagaki, Yoshio, Rebecca L Seipelt, Martha L Peterson, and James L Manley (1996). "The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation". In: *Cell* 87.5, pp. 941–952.

Taliaferro, J Matthew et al. (2016). "Distal alternative last exons localize mRNAs to neural projections". In: *Molecular cell* 61.6, pp. 821–833.

Tan, Sheng et al. (2021). "CPSF6 links alternative polyadenylation to metabolism adaption in hepatocellular carcinoma progression". In: *Journal of Experimental & Clinical Cancer Research* 40.1, pp. 1–16.

Tang, Peng, Yang Yang, et al. (2022). "Alternative polyadenylation by sequential activation of distal and proximal PolyA sites". In: *Nature Structural & Molecular Biology*, pp. 1–11.

Tang, Qiannan, Xinlu Wang, and Guangxia Gao (2017). "The short form of the zinc finger antiviral protein inhibits influenza A virus protein expression and is antagonized by the virus-encoded NS1". In: *Journal of virology* 91.2, e01909–16.

Tapia, Felipe, Tanja Laske, Milena A Wasik, Markus Rammhold, Yvonne Genzel, and Udo Reichl (2019). "Production of defective interfering particles of in-

fluenza A virus in parallel continuous cultures at two residence times—insights from qPCR measurements and viral dynamics modeling". In: *Frontiers in bioengineering and biotechnology*, p. 275.

Taubenberger, Jeffery K and David M Morens (2006). "1918 Influenza: the mother of all pandemics". In: *Revista Biomedica* 17.1, pp. 69–79.

Taylor, Katarzyna and Krzysztof Sobczak (2020). "Intrinsic regulatory role of RNA structural arrangement in alternative splicing control". In: *International journal of molecular sciences* 21.14, p. 5161.

Tchatalbachev, Svetlin, Ramon Flick, and Gerd Hobom (2001). "The packaging signal of influenza viral RNA molecules". In: *Rna* 7.7, pp. 979–989.

Te Velthuis, Aartjan JW, Jonathan M Grimes, and Ervin Fodor (2021). "Structural insights into RNA polymerases of negative-sense RNA viruses". In: *Nature Reviews Microbiology* 19.5, pp. 303–318.

Te Velthuis, Aartjan JW, Nicole C Robb, Achillefs N Kapanidis, and Ervin Fodor (2016). "The role of the priming loop in influenza A virus RNA synthesis". In: *Nature microbiology* 1.5, pp. 1–7.

Tee, Meng-Kian et al. (1995). "T→ A transversion 11 bp from a splice acceptor site in the human gene for steroidogenic acute regulatory protein causes congenital lipoid adrenal hyperplasia". In: *Human molecular genetics* 4.12, pp. 2299–2305.

"The Gene Ontology resource: enriching a GOld mine" (2021). In: *Nucleic acids research* 49.D1, pp. D325–D334.

Thompson, Matthew G, Mark Dittmar, et al. (2020). "Viral-induced alternative splicing of host genes promotes influenza replication". In: *Elife* 9, e55500.

Thompson, Matthew G, Raquel Muñoz-Moreno, et al. (2018). "Co-regulatory activity of hnRNP K and NS1-BP in influenza and human mRNA splicing". In: *Nature communications* 9.1, pp. 1–12.

Tian, Bin and Joel H Graber (2012). "Signals for pre-mRNA cleavage and polyadenylation". In: *Wiley Interdisciplinary Reviews: RNA* 3.3, pp. 385–396.

Tian, Bin, Jun Hu, Haibo Zhang, and Carol S Lutz (2005). "A large-scale analysis of mRNA polyadenylation of human and mouse genes". In: *Nucleic acids research* 33.1, pp. 201–212.

Tian, Bin and James L Manley (2017). "Alternative polyadenylation of mRNA precursors". In: *Nature reviews Molecular cell biology* 18.1, pp. 18–30.

Tian, Bin, Zhenhua Pan, and Ju Youn Lee (2007). "Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing". In: *Genome research* 17.2, pp. 156–165.

Tomé, FMS and M Fardeau (1980). "Nuclear inclusion in oculopharyngeal dystrophy". In: *Acta neuropathologica* 49.1, pp. 85–87.

Tran, Vy et al. (2020). "Influenza virus repurposes the antiviral protein IFIT2 to promote translation of viral mRNAs". In: *Nature microbiology* 5.12, pp. 1490–1503.

Tsai, Pei-Ling, Ni-Ting Chiou, Sharon Kuss, Adolfo García-Sastre, Kristen W Lynch, and Beatriz MA Fontoura (2013). "Cellular RNA binding proteins NS1-BP and hnRNP K regulate influenza A virus RNA splicing". In: *PLoS pathogens* 9.6, e1003460.

Turner, Martin and Manuel D Diaz-Munoz (2018). "RNA-binding proteins control gene expression and cell fate in the immune system". In: *Nature immunology* 19.2, pp. 120–129.

Tushev, Georgi, Caspar Glock, Maximilian Heumüller, Anne Biever, Marko Jovanovic, and Erin M Schuman (2018). "Alternative 3 UTRs modify the localization, regulatory potential, stability, and plasticity of mRNAs in neuronal compartments". In: *Neuron* 98.3, pp. 495–511.

Tuvim, Michael J, Brian E Gilbert, Burton F Dickey, and Scott E Evans (2012). "Synergistic TLR2/6 and TLR9 activation protects mice against lethal influenza pneumonia". In: *PloS one* 7.1, e30596.

Twu, Karen Y, Diana L Noah, Ping Rao, Rei-Lin Kuo, and Robert M Krug (2006). "The CPSF30 binding site on the NS1A protein of influenza A virus is a potential antiviral target". In: *Journal of virology* 80.8, pp. 3957–3965.

Ujie, Michiko et al. (2019). "Long-term culture of human lung adenocarcinoma A549 cells enhances the replication of human influenza A viruses". In: *The Journal of General Virology* 100.10, p. 1345.

Ule, Jernej and Benjamin J Blencowe (2019). "Alternative splicing regulatory networks: functions, mechanisms, and evolution". In: *Molecular cell* 76.2, pp. 329–345.

Vasudevan, Shobha, Stuart W Peltz, and Carol J Wilusz (2002). "Non-stop decay—a new mRNA surveillance pathway". In: *Bioessays* 24.9, pp. 785–788.

Veler, Hana et al. (2021). "The C-terminal domains of the PB2 subunit of the influenza A virus RNA polymerase directly interact with cellular GTPase Rab11a". In: *Journal of Virology*, jvi–01979.

Venkataraman, Krishnan, Kirk M Brown, and Gregory M Gilmartin (2005). "Analysis of a noncanonical poly (A) site reveals a tripartite mechanism for vertebrate poly (A) site recognition". In: *Genes & development* 19.11, pp. 1315–1327.

Vries, Henk de, Ursula Rüegsegger, Wolfgang Hübner, Arno Friedlein, Hanno Langen, and Walter Keller (2000). "Human pre-mRNA cleavage factor IIm contains homologs of yeast proteins and bridges two other cleavage factors". In: *The EMBO journal* 19.21, pp. 5895–5904.

Walker, Alexander P and Ervin Fodor (2019). "Interplay between influenza virus and the host RNA polymerase II transcriptional machinery". In: *Trends in microbiology* 27.5, pp. 398–407.

Walker, Alexander P, Jane Sharps, and Ervin Fodor (2020). "Mutation of an influenza virus polymerase 3 RNA promoter binding site inhibits transcription elongation". In: *Journal of Virology* 94.13, e00498–20.

Wall, Richard J et al. (2018). "Clinical and veterinary trypanocidal benzoxaboroles target CPSF3". In: *Proceedings of the National Academy of Sciences* 115.38, pp. 9616–9621.

Wallace, A Michelle et al. (1999). "Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells". In: *Proceedings of the National Academy of Sciences* 96.12, pp. 6763–6768.

Wandzik, Joanna M, Tomas Kouba, and Stephen Cusack (2021). "Structure and function of influenza polymerase". In: *Cold Spring Harbor Perspectives in Medicine* 11.9, a038372.

Wandzik, Joanna M, Tomas Kouba, Manikandan Karuppasamy, et al. (2020). "A structure-based model for the complete transcription cycle of influenza polymerase". In: *Cell* 181.4, pp. 877–893.

Wang, Lingyan, Bishi Fu, et al. (2017). "Comparative influenza protein interactomes identify the role of plakophilin 2 in virus restriction". In: *Nature communications* 8.1, pp. 1–12.

Wang, Ping, Peter Palese, and Robert E O'Neill (1997). "The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza a virus nucleoprotein NP is a nonconventional nuclear localization signal". In: *Journal of virology* 71.3, pp. 1850–1856.

Wang, Ruifang, Yinxing Zhu, et al. (2019). "Influenza M2 protein regulates MAVS-mediated signaling pathway through interacting with MAVS and increasing ROS production". In: *Autophagy* 15.7, pp. 1163–1181.

Wang, Ruijia, Ram Nambiar, Dinghai Zheng, and Bin Tian (2018). "PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes". In: *Nucleic acids research* 46.D1, pp. D315–D319.

Wang, Ruijia and Bin Tian (2020). "APAlyzer: a bioinformatics package for analysis of alternative polyadenylation isoforms". In: *Bioinformatics* 36.12, pp. 3907–3909.

Wang, Ruijia, Dinghai Zheng, Lu Wei, Qingbao Ding, and Bin Tian (2019). "Regulation of intronic polyadenylation by PCF11 impacts mRNA expression of long genes". In: *Cell reports* 26.10, pp. 2766–2778.

Wang, Ruijia, Dinghai Zheng, Ghassan Yehia, and Bin Tian (2018). "A compendium of conserved cleavage and polyadenylation events in mammalian genes". In: *Genome research* 28.10, pp. 1427–1441.

Wang, Song, Hua Li, et al. (2012). "Transport of influenza virus neuraminidase (NA) to host cell surface is regulated by ARHGAP21 and Cdc42 proteins". In: *Journal of Biological Chemistry* 287.13, pp. 9804–9816.

Wang, Xingbo, Tuyuan Zheng, Lulu Lin, et al. (2020). "Influenza a virus induces autophagy by its hemagglutinin binding to cell surface heat shock protein 90AA1". In: *Frontiers in Microbiology* 11, p. 2399.

Wang, Xiuye, Thomas Hennig, et al. (2020). "Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27". In: *Nature communications* 11.1, pp. 1–13.

Wang, Xiuye, Liang Liu, et al. (2021). "Mechanism and consequences of herpes simplex virus 1-mediated regulation of host mRNA alternative polyadenylation". In: *PLoS genetics* 17.3, e1009263.

Webster, RG and WG Laver (1967). "Preparation and properties of antibody directed specifically against the neuraminidase of influenza virus". In: *The Journal of Immunology* 99.1, pp. 49–55.

Wei, Lu and Eric C Lai (2022). "Regulation of the Alternative Neural Transcriptome by ELAV/Hu RNA Binding Proteins." In: *Frontiers in Genetics* 13, pp. 848626–848626.

Wei, Lu, Seungjae Lee, et al. (2020). "Overlapping activities of ELAV/Hu family RNA binding proteins specify the extended neuronal 3 UTR landscape in drosophila". In: *Molecular cell* 80.1, pp. 140–155.

Wei, Xiaona, Jie Li, Yun Zhang, Lang Gong, Chunyi Xue, and Yongchang Cao (2021). "Profiling of alternative polyadenylation and gene expression in PEDV-infected IPEC-J2 cells". In: *Virus Genes* 57.2, pp. 181–193.

Weon, Jenny L and Patrick Ryan Potts (2015). "The MAGE protein family and cancer". In: *Current opinion in cell biology* 37, pp. 1–8.

West, Steven and Nicholas J Proudfoot (2008). "Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination". In: *Nucleic acids research* 36.3, pp. 905–914.

White, Judith M and Gary R Whittaker (2016). "Fusion of enveloped viruses in endosomes". In: *Traffic* 17.6, pp. 593–614.

Wiestner, Adrian et al. (2007). "Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival". In: *Blood, The Journal of the American Society of Hematology* 109.11, pp. 4599–4606.

Wilkinson, Max E, Clément Charenton, and Kiyoshi Nagai (2020). "RNA splicing by the spliceosome". In: *Annual review of biochemistry* 89, pp. 359–388.

Wood, Andrew J, Roland G Roberts, David Monk, Gudrun E Moore, Reiner Schulz, and Rebecca J Oakey (2007). "A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation". In: *PLoS Genetics* 3.2, e20.

Woodward, Lauren A, Justin W Mabin, Pooja Gangras, and Guramrit Singh (2017). "The exon junction complex: a lifelong guardian of mRNA fate". In: *Wiley Interdisciplinary Reviews: RNA* 8.3, e1411.

Wu, Fangyi, Zhenmin Niu, Bin Zhou, Pengcheng Li, and Feng Qian (2019). "PSMB1 negatively regulates the innate antiviral immunity by facilitating degradation of IKK-ε". In: *Viruses* 11.2, p. 99.

Wu, Yingcheng, Hao Chen, et al. (2019). "HPV shapes tumor transcriptome by globally modifying the pool of RNA binding protein-binding motif". In: *Aging (Albany NY)* 11.8, p. 2430.

Xia, Zheng et al. (2014). "Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3-UTR landscape across seven tumour types". In: *Nature communications* 5.1, pp. 1–13.

Xiang, Kehui et al. (2010). "Crystal structure of the human symplekin–Ssu72–CTD phosphopeptide complex". In: *Nature* 467.7316, pp. 729–733.

Xie, Zhuorui et al. (2021). "Gene set knowledge discovery with enrichr". In: *Current protocols* 1.3, e90.

Xing, Yifei et al. (2021). "Downregulation of NUDT21 contributes to cervical cancer progression through alternative polyadenylation". In: *Oncogene* 40.11, pp. 2051–2064.

Xu, Guanlong et al. (2017). "PA-X protein contributes to virulence of triple-reassortant H1N2 influenza virus by suppressing early immune responses in swine". In: *Virology* 508, pp. 45–53.

Yan, Zhen et al. (1998). "Cdc6 is regulated by E2F and is essential for DNA replication in mammalian cells". In: *Proceedings of the National Academy of Sciences* 95.7, pp. 3603–3608.

Yang, Fei, Wei Wang, Murat Cetinbas, Ruslan I Sadreyev, and Michael D Blower (2020). "Genome-wide analysis identifies cis-acting elements regulating mRNA polyadenylation and translation during vertebrate oocyte maturation". In: *Rna* 26.3, pp. 324–344.

Yang, Qin, Molly Coseno, Gregory M Gilmartin, and Sylvie Doublié (2011). "Crystal structure of a human cleavage factor CFIm25/CFIm68/RNA complex provides an insight into poly (A) site recognition and RNA looping". In: *Structure* 19.3, pp. 368–377.

Yang, Qin, Gregory M Gilmartin, and Sylvie Doublié (2011). "The structure of human cleavage factor Im hints at functions beyond UGUA-specific RNA binding: a role in alternative polyadenylation and a potential link to 5'capping and splicing". In: *RNA biology* 8.5, pp. 748–753.

Yang, Seung Wook, Lei Li, Jon P Connelly, et al. (2020). "A cancer-specific ubiquitin ligase drives mRNA alternative polyadenylation by ubiquitinating the mRNA 3 end processing complex". In: *Molecular cell* 77.6, pp. 1206–1221.

Yang, Wen, Peter L Hsu, Fan Yang, Jae-Eun Song, and Gabriele Varani (2018). "Reconstitution of the CstF complex unveils a regulatory role for CstF-50 in recognition of 3-end processing signals". In: *Nucleic acids research* 46.2, pp. 493–503.

Yang, Yan, Wencheng Li, Mainul Hoque, et al. (2016). "PAF complex plays novel subunit-specific roles in alternative cleavage and polyadenylation". In: *PLoS genetics* 12.1, e1005794.

Yao, Chengguo, Jacob Biesinger, et al. (2012). "Transcriptome-wide analyses of CstF64–RNA interactions in global regulation of mRNA alternative polyadenylation". In: *Proceedings of the National Academy of Sciences* 109.46, pp. 18773–18778.

Yao, Chengguo, Eun-A Choi, et al. (2013). "Overlapping and distinct functions of CstF64 and CstF64τ in mammalian mRNA 3 processing". In: *Rna* 19.12, pp. 1781–1790.

Yasutomo, Koji (2019). "Dysregulation of immunoproteasomes in autoinflammatory syndromes". In: *International Immunology* 31.10, pp. 631–637.

Ye, Congting, Yuqi Long, Guoli Ji, Qingshun Quinn Li, and Xiaohui Wu (2018). "APAtrap: identification and quantification of alternative polyadenylation sites from RNA-seq data". In: *Bioinformatics* 34.11, pp. 1841–1849.

Ye, Peiying, Shufeng Liu, Yiping Zhu, Guifang Chen, and Guangxia Gao (2010). "DEXH-Box protein DHX30 is required for optimal function of the zinc-finger antiviral protein". In: *Protein & cell* 1.10, pp. 956–964.

Yeh, Hsin-Sung and Jeongsik Yong (2016). "Alternative polyadenylation of mRNAs: 3-untranslated region matters in gene expression". In: *Molecules and cells* 39.4, p. 281.

Yin, Ziwei et al. (2020). "RNA-binding motifs of hnRNP K are critical for induction of antibody diversification by activation-induced cytidine deaminase". In: *Proceedings of the National Academy of Sciences* 117.21, pp. 11624–11635.

Ylösmäki, Leena, Riku Fagerlund, Inka Kuisma, Ilkka Julkunen, and Kalle Saksela (2016). "Nuclear translocation of crk adaptor proteins by the influenza a virus ns1 protein". In: *Viruses* 8.4, p. 101.

Yonaha, Masatomo and Nick J Proudfoot (1999). "Specific transcriptional pausing activates polyadenylation in a coupled in vitro system". In: *Molecular cell* 3.5, pp. 593–600.

Yoon, YoneJung et al. (2013). "Anxiety-associated alternative polyadenylation of the serotonin transporter mRNA confers translational regulation by hnRNPK". In: *Proceedings of the National Academy of Sciences* 110.28, pp. 11624–11629.

York, Ashley and Ervin Fodor (2013). "Biogenesis, assembly, and export of viral messenger ribonucleoproteins in the influenza A virus infected cell". In: *RNA biology* 10.8, pp. 1274–1282.

Yu, Ming et al. (2015). "RNA polymerase II–associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II". In: *Science* 350.6266, pp. 1383–1386.

Yuan, Chendong, Mingxiang Chen, and Xiaolu Cai (2021). "Advances in poly (rC)-binding protein 2: Structure, molecular function, and roles in cancer". In: *Biomedicine & Pharmacotherapy* 139, p. 111719.

Yuan, Fuwen, William Hankey, Eric J Wagner, Wei Li, and Qianben Wang (2021). "Alternative polyadenylation of mRNA and its role in cancer". In: *Genes & Diseases* 8.1, pp. 61–72.

Yudin, Dmitry et al. (2008). "Localized regulation of axonal RanGTPase controls retrograde injury signaling in peripheral nerve". In: *Neuron* 59.2, pp. 241–252.

Zarudnaya, Margarita I, Iryna M Kolomiets, Andriy L Potyahaylo, and Dmytro M Hovorun (2003). "Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures". In: *Nucleic acids research* 31.5, pp. 1375–1386.

Zhang, Baoge, Mohsan Ullah Goraya, et al. (2020). "Zinc finger CCCH-type antiviral protein 1 restricts the viral replication by positively regulating type I interferon response". In: *Frontiers in Microbiology*, p. 1912.

Zhang, Haili, Zhenyu Zhang, et al. (2019). "Fundamental contribution and host range determination of ANP32A and ANP32B in influenza A virus polymerase activity". In: *Journal of virology* 93.13, e00174–19.

Zhang, Jin, Wenqiang Sun, Cong Ren, Xiangmudong Kong, Wensheng Yan, and Xinbin Chen (2019). "A PolH transcript with a short 3 UTR enhances PolH expression and mediates cisplatin resistance". In: *Cancer research* 79.14, pp. 3714–3724.

Zhang, Ke, Yihu Xie, et al. (2019). "Structural basis for influenza virus NS1 protein block of mRNA nuclear export". In: *Nature microbiology* 4.10, pp. 1671–1679.

Zhang, Yi, Lian Liu, et al. (2021). "Alternative polyadenylation: methods, mechanism, function, and role in cancer". In: *Journal of Experimental & Clinical Cancer Research* 40.1, pp. 1–19.

Zhang, Yixiao, Yadong Sun, Yongsheng Shi, Thomas Walz, and Liang Tong (2020). "Structural insights into the human pre-mRNA 3-end processing machinery". In: *Molecular cell* 77.4, pp. 800–809.

Zhao, Nan et al. (2018). "Influenza virus infection causes global RNAPII termination defects". In: *Nature structural & molecular biology* 25.9, pp. 885–893.

Zheng, Dinghai, Xiaochuan Liu, and Bin Tian (2016). "3 READS+, a sensitive and accurate method for 3 end sequencing of polyadenylated RNA". In: *Rna* 22.10, pp. 1631–1639.

Zheng, Dinghai and Bin Tian (2014). "RNA-binding proteins in regulation of alternative cleavage and polyadenylation". In: *Systems Biology of RNA Binding Proteins*, pp. 97–127.

Zheng, Hongyong, Hye Annie Lee, Peter Palese, and Adolfo García-Sastre (1999). "Influenza A virus RNA polymerase has the ability to stutter at the polyadenylation site of a viral RNA template during RNA replication". In: *Journal of virology* 73.6, pp. 5240–5243.

Zheng, Tong, Yuqin Tan, et al. (2021). "Alternative polyadenylation trans-factor FIP1 exacerbates UUO/IRI-induced kidney injury and contributes to AKI-CKD transition via ROS-NLRP3 axis". In: *Cell Death & Disease* 12.6, pp. 1–16.

Zhou, Ao, Wenhua Zhang, Xia Dong, Mengyun Liu, Hongbo Chen, and Bin Tang (2022). "The battle for autophagy between host and influenza A virus". In: *Virulence* 13.1, pp. 46–59.

Zhou, Zijing, Jing Qu, et al. (2020). "Stiff matrix instigates type I collagen biogenesis by mammalian cleavage factor I complex-mediated alternative polyadenylation". In: *JCI insight* 5.3.

Zhu, Hui, Hua-Lin Zhou, Robert A Hasman, and Hua Lou (2007). "Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences". In: *Journal of Biological Chemistry* 282.4, pp. 2203–2210.

Zhu, Yong, Xiuye Wang, et al. (2018). "Molecular mechanisms for CFIm-mediated regulation of mRNA alternative polyadenylation". In: *Molecular cell* 69.1, pp. 62–74.

Zhu, ZH, YP Yu, YK Shi, JB Nelson, and JH Luo (2009). "CSR1 induces cell death through inactivation of CPSF3". In: *Oncogene* 28.1, pp. 41–51.