

VIRAL AND HOST FACTORS REGULATING INFLUENZA VIRUS REPLICATION

BENJAMIN E. NILSSON



Sir William Dunn School of Pathology
and
Lincoln College

A thesis submitted for the degree of Doctor of Philosophy at the

University of Oxford, Trinity Term 2017

ABSTRACT

Viral and host factors regulating influenza virus replication

Benjamin E. Nilsson
Lincoln College

Doctor of Philosophy
Trinity Term, 2017

Avian influenza A viruses typically do not replicate very efficiently when exposed to a mammalian host species. One of the reasons for this is the low activity of the viral RNA-dependent RNA polymerase of avian influenza viruses in mammalian cells. This host restrictive effect can be overcome by adaptive mutations in the avian polymerase, many of which are found in the 627-domain of the PB2 polymerase subunit. Deletion of the 627-domain revealed that this domain is not required for enzymatic functions of the polymerase *in vitro*, but that it is essential for viral replication in a cellular environment in a nucleoprotein-independent manner. While the 627-domain is not necessary for viral RNA synthesis, it was demonstrated that it is involved in mediating encapsidation of nascent replication products.

Recently the host factor ANP32A was shown to be the main determinant of host range restriction of the viral polymerase. It was demonstrated that during viral infections ANP32A interacts with KPNA2, a host factor strongly linked to host range restriction of the viral polymerase. It was also revealed that avian polymerases specifically are restricted in vRNA synthesis, a defect that was reversed in the presence of avian ANP32A. ANP32A was shown to be an enhancer of vRNA synthesis *in vitro*.

Viral polymerase-polymerase interactions have been reported previously and presumably fulfil several essential functions during viral replication. Here the potential interaction interfaces of two different polymerase dimers were investigated and a role of polymerase dimers in replication and in *trans*-activation of cRNA-bound polymerases was found.

RNA-binding proteins are essential for RNA metabolism and therefore cell physiology. It has been reported that the RNA-binding proteome responds to biological stimuli. Here the response of the RNA-binding proteome to influenza virus infection was investigated using an *in vivo* UV crosslinking interactome capture technique. It was demonstrated that the RNA-binding proteome is significantly altered and that this effect is independent of protein abundance. Several host RNA-binding proteins were identified that change their RNA-binding behaviour and that could have pro- or antiviral functions during influenza virus infection.

DECLARATION

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

Benjamin E. Nilsson

London, October 2017

ACKNOWLEDGEMENTS

I would like to thank my supervisor Ervin Fodor for all his support, encouragement and inspiration during the course my studies. I did not realise molecular biology was that much fun before I joined his lab. I would also like to thank Aartjan te Velthuis for always being there for advice, discussions and late night chats. Another person who I am grateful to is Jane Sharps without whom the lab would not be able to function properly.

I would also like to thank my collaborators Alfredo Castello, Shabaz Mohammed, Svenja Hester, Ni Shuai, Bernd Fischer, Alastair Louey, Andre Furger, Michael Tellier and everyone in the Castello lab for all their work and support.

I am also grateful to the Wellcome Trust who funded my studies and to the Microbiology Society and Lincoln College for funds allowing me to visit several conferences.

I have to especially thank everyone in the Fodor lab and my friends who have put up with me and my commuting all these years as well as for fun memories of my time in Oxford. As always, I would not be where I am without my parents who have always supported me and made it possible for me to follow my dreams.

Finally, thank you Catherine, I could not have done this without you and all your support and love.

TABLE OF CONTENTS

ABSTRACT	2
DECLARATION	4
ACKNOWLEDGEMENTS	5
ABBREVIATIONS	10
Chapter 1: Introduction	14
1.1 General introduction.....	14
1.2 Influenza virus taxonomy and nomenclature.....	16
1.3 The influenza A virion.....	17
1.4 The influenza A virus life cycle	23
1.4.1 Attachment and entry	23
1.4.2 Fusion, uncoating and nuclear import	24
1.4.3 Viral gene expression	27
1.4.4 Nuclear export and trafficking of vRNP complexes.....	28
1.4.5 Genome packaging	30
1.4.6 Virion assembly and egress.....	31
1.5 Influenza virus RNA synthesis	32
1.5.1 Viral ribonucleoprotein complexes	32
1.5.2 The viral RNA-dependent RNA polymerase	35
1.5.2.1 The structure of the heterotrimeric viral RdRP.....	35
1.5.2.2 Conformational rearrangements of the viral RdRP	39
1.5.3 The viral nucleoprotein.....	41
1.5.4 Viral RNA promoters	41
1.5.5 Viral mRNA synthesis, processing and transport.....	43
1.5.5.1 Transcription	43
1.5.5.2 Polyadenylation.....	45
1.5.5.3 Splicing	46
1.5.5.4 Nuclear export of viral mRNA.....	47
1.5.6 Viral RNA replication.....	48
1.5.6.1 cRNA synthesis.....	48
1.5.6.2 vRNA synthesis	50
1.6 Host - virus interactions	51
1.6.1 Transcription and translation of eukaryotic mRNA.....	51
1.6.2 The RNA-binding proteome	54
1.6.3 Influenza virus mRNA processing	56
1.6.4 NS1-mediated host shut off.....	56
1.6.5 Pol II degradation by the viral RdRP.....	57
1.6.6 Host RNA degradation by PA-X.....	58
1.7 Influenza A virus host adaptation	59
1.7.1 Overview	59
1.7.2 Host adaptation and the PB2 polymerase subunit	60
1.7.2.1 The canonical host determinant PB2-627	60
1.7.2.2 Other PB2 host adaptive mutations.....	63
1.7.2.3 ANP32A is a determinant of host range restriction	67
1.7.3 Host adaptation and the PB1 polymerase subunit	69
1.7.4 Host adaptation and the PA polymerase subunit	72

1.7	Thesis objectives and outline	74
Chapter 2: Materials and Methods		76
2.1	Materials	76
2.1.1	General reagents	76
2.1.2	Enzymes	76
2.1.3	Radiochemicals	77
2.1.4	Synthetic DNA and RNA.....	77
2.1.5	Antibodies	77
2.1.6	Bacteria and bacterial culture media.....	78
2.1.7	Cell culture media.....	78
2.1.8	Cell lines	79
2.1.9	Viruses	79
2.1.10	Plasmids	79
2.2	Cloning	82
2.2.1	PCR mutagenesis	82
2.2.2	Plasmid construction.....	82
2.2.3	Transformation into competent E. coli.....	83
2.2.4	Plasmid preparation and sequencing	83
2.3	Eukaryotic cell culture	84
2.4	Virus culture	84
2.4.1	Reverse genetics.....	84
2.4.2	Infection of mammalian cells with influenza A virus.....	85
2.4.3	Plaque assay	85
2.4.4	Growth curve analysis	86
2.4.5	Purification of influenza A virus.....	86
2.5	Viral RNA analysis	87
2.5.1	RNP reconstitution assay.....	87
2.5.2	cRNA stabilisation assay	87
2.5.3	RNA extraction.....	88
2.5.4	Primer extension analysis.....	88
2.5.4.1	Primers	88
2.5.4.2	Radiolabelling of oligodeoxyribonucleotides	89
2.5.4.3	Primer extension.....	89
2.5.5	Polyacrylamide gel electrophoresis (PAGE).....	90
2.6	Protein purification and analysis	90
2.6.1	Recombinant influenza A virus RdRP purification.....	90
2.6.2	Recombinant ANP32A and ANP32B purification.....	91
2.6.3	Protein gel electrophoresis	92
2.6.3.1	SDS-PAGE	92
2.6.3.2	Western blotting.....	93
2.6.3.3	Staining of protein gels	93
2.7	In vitro activity assays	94
2.7.1	Capped-oligonucleotide cleavage assay.....	94
2.7.2	Capped-oligonucleotide extension assay	94
2.7.3	Globin mRNA transcription assay	96
2.7.4	Dinucleotide synthesis assay	97
2.7.5	Primer-independent replication assay	97
2.7.6	ApG extension assay	98
2.7.7	RNP activity assay	99
2.8	Viral RNA binding assays.....	99
2.8.1	Co-purification of viral RNA with the viral RdRP.....	99

2.8.2	smFRET promoter binding assay	100
2.9	RNA sequencing	101
2.9.1	3' UTR mRNA sequencing	101
2.9.1.1	Sample preparation and sequencing	101
2.9.1.2	RNA sequence analysis	101
2.9.2	Whole transcriptome sequencing	104
2.10	RBPome capture	105
2.10.1	Label-free RBPome capture	105
2.10.2	SILAC-labelled RBPome-capture	107
2.11	Mass spectrometry	108
2.11.1	Filter Aided Sample Preparation (FASP)	108
2.11.2	Mass spectrometry	109
2.11.3	Analysis of mass spectrometry data	110
2.12	Microscopy	111
2.12.1	Immunofluorescence confocal microscopy	111
2.12.2	Immunofluorescence microscopy	112
Chapter 3: The role of the PB2 627-domain in polymerase function		114
3.1	Introduction	114
3.2	Results	116
3.2.1	Assembly of the heterotrimeric viral RdRP	116
3.2.2	The PB2 627-domain does not affect binding of viral RNA	119
3.2.3	Requirement of PB2 domains for RdRP in vitro activity	122
3.2.4	The PB2 627-domain is required for cellular RdRP activity	128
3.2.5	The PB2 627-domain mediates RdRP recruitment	133
3.3	Discussion	138
3.4	Summary	142
Chapter 4: The role of ANP32A in viral polymerase function		143
4.1	Introduction	143
4.2	Results	145
4.2.1	The ANP32A interactome	145
4.2.2	The ANP32A interactome in influenza A virus infection	148
4.2.3	Avian ANP32A is required for avian IAV RdRP activity in mammalian host cells	154
4.2.4	Avian ANP32A is required for efficient vRNA synthesis by the avian RdRP in mammalian hosts	157
4.2.5	ANP32A enhances vRNA synthesis in vitro	159
4.3	Discussion	162
4.4	Summary	167
Chapter 5: Characterising the viral polymerase dimer interface		168
5.1	Introduction	168
5.2	Results	174
5.2.1	The N-terminus of PA is required for vRNA synthesis	174
5.2.2	RdRP dimer interfaces are important for viral replication	180
5.2.3	Avian ANP32A can restore replication by PA-C95A	183
5.3	Discussion	185
5.4	Summary	189
Chapter 6: The mRNA interactome in influenza virus infection		190
6.1	Introduction	190
6.2	Results	191

6.2.1	Capture of human RBPs in A549 cells	191
6.2.2	The dynamics of viral gene expression	195
6.2.3	The dynamics of viral RNA synthesis	198
6.2.4	The effect of influenza virus infection on the RBPome	204
6.2.5	Protein abundance does not drive changes in the RBPome	208
6.3	Discussion	215
6.4	Summary	222
REFERENCES		223
APPENDICES		259

ABBREVIATIONS

UNITS

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

bp	Base pair
Ci	Curie; $1 \text{ Ci} = 3.7 \times 10^{10}$ becquerels
Da	Dalton
g	Standard gravity: 9.80665 m/s^2
h	Hour(s)
kb	Kilobase
kDa	Kilodalton
min	Minute(s)
MOI	Multiplicity of infection
nt	Nucleotide(s)
PFU	Plaque-forming unit
pH	$-\log_{10} c$, where c is $[\text{H}^+]$ in mol/l
rpm	Revolutions per minute
s	Second(s)
U	Enzyme unit

OTHER ABBREVIATIONS

A	Adenine / Alanine
ACTB	β -actin
ACTD	Actinomycin D
ADAR1	Double-stranded RNA-specific adenosine deaminase
ANP32A	Acidic nuclear phosphoprotein 32 family member A
ANP32B	Acidic nuclear phosphoprotein 32 family member B
APA	Alternative polyadenylation
ATP	Adenosine-5'-triphosphate

C	Cysteine / Cytidine
CCR5	C-C chemokine receptor type 5
cDNA	Complementary DNA
chANP32A	Chicken acidic nuclear phosphoprotein 32 family member A
cRNA	Complementary RNA
cRNP	Complementary ribonucleoprotein
CTD	C-terminal domain
CTP	Cytidine-5'-triphosphate
D	Aspartic acid
DAPI	4',6-diamidino-2-phenylindole
ddH ₂ O	Double distilled water
DDX50	DEAD-Box Helicase 50
dH ₂ O	Distilled water
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide-5'-triphosphate
DTT	Dithiothreitol
E	Glutamic acid
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
F	Phenylalanine
FC	Fold change
FCS	Foetal calf serum
FDR	False discovery rate
G	Glycine / Guanine
GO	Gene ontology
GST	Glutathione S-transferase
GTP	Guanosine-5'-triphosphate
H	Histidine / Hydrogen
HA	Haemagglutinin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HNRNPR	Heterogeneous nuclear ribonucleoprotein R
h.p.i.	hours post-infection
HRP	Horseradish peroxidase

huANP32A	Human acidic nuclear phosphoprotein 32 family member A
huANP32B	Human acidic nuclear phosphoprotein 32 family member B
IPTG	Isopropyl β -D-1-thiogalactopyranoside
K	Lysine
KPNA1	Importin subunit α 5
KPNA2	Importin subunit α 1
KPNA6	Importin subunit α 7
M1	Matrix protein 1
M2	Matrix protein 2
MEM	Minimum Essential Medium
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MWCO	Molecular weight cut-off
N	Asparagine
NA	Neuraminidase
NEP	Nuclear export protein
NES	Nuclear export signal
NLS	Nuclear localisation signal
NP	Nucleoprotein
NS	Non-structural
NS1	Non-structural protein 1
NT60	Influenza A /Northern Territory /60/68 (H3N2) virus
ORF	Open reading frame
PABPC1	Poly(A) binding protein cytoplasmic 1
p.i.	post-infection
p.t.	post-transfection
PA	Polymerase acidic protein
PAGE	Polyacrylamide gel electrophoresis
PB1	Polymerase basic protein 1
PB2	Polymerase basic protein 2
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank

PMSF	Phenylmethylsulfonyl fluoride
PNK	T4 polynucleotide kinase
poly(A)	Polyadenylic acid
poly(U)	Polyuridylic acid
PTBP1	Polypyrimidine tract binding protein 1
RBD	RNA-binding domain
RBP	RNA-binding protein
RBPome	RNA-binding proteome
RdRP	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
rRNA	Ribosomal ribonucleic acid
S	Serine
SE	Standard error of the mean
SOC	Super Optimal broth with Catabolite repression
SD	Standard deviation
SDS	Sodium dodecyl sulfate
ss	single-stranded
T	Threonine
TAP	Tandem affinity purification
TBE	Tris-borate-EDTA
TEMED	<i>N,N,N',N'</i> -Tetramethylethylenediamine
TEV	Tobacco etch virus protease
Tris	Tris (hydroxymethyl) aminomethane
U	Uracil
UTP	Uridine-5'-triphosphate
UTR	Untranslated region
VEZT	Vezatin
vRNA	Viral ribonucleic acid
vRNP	Viral ribonucleoprotein
WHO	World Health Organisation
WSN	Influenza A /WSN/ 33 (H1N1) virus
Wt	Wild-type
Y	Tyrosine

Chapter 1: Introduction

1.1 General introduction

Influenza viruses cause a highly contagious disease of the respiratory tract that is characterised by common respiratory symptoms such as fever, cough, throat pain, headache, nasal congestion and sneezing, as well as muscle aching and malaise (Eccles, 2005). Symptoms resembling influenza were described by Hippocrates as early as in 412 BC in ancient Greece and it is entirely feasible that humans have been exposed to influenza since the onset of domestication and urbanisation when animal to human transmission would have been made much more likely through the concentration of human and animal populations in close vicinity. Influenza gained its current name in the 15th century from “influenza”, the Italian word for “influence”, due to the common belief that this illness was under the influence of the stars (Potter, 2001).

Influenza viruses are responsible for significant mortality and morbidity in the form of seasonal epidemics and occasional worldwide pandemics. In the northern and southern hemisphere, influenza epidemics occur every winter causing significant amounts of hospitalisation and death. The World Health Organisation (WHO) estimates that annual influenza epidemics result in up to five million cases of severe illness and between 250,000 to 500,000 deaths worldwide (WHO, 2016). Seasonal epidemics occur because of antigenic drift, a slow accumulation of mutations driven by the error-prone viral RNA polymerase allowing mutants to escape the host adaptive immune response. Occasional worldwide pandemics are the result of antigenic shift events, where transmission or reassortment of

influenza virus strains from non-human species to humans can lead to a global spread of influenza with significantly increased mortality rate due to a lack of cross protection from this zoonotic virus. Since 1900 the world has experienced five influenza pandemics. The 'Spanish influenza' pandemic of 1918 is the most severe pandemic to have been recorded in human history (Taubenberger & Kash, 2010; Taubenberger & Morens, 2006). The 1918 pandemic was caused by an H1N1 strain of influenza A virus of zoonotic origin and was estimated to have infected 500 million people (a third of the total world population at the time) and killing an estimated 50 to 100 million people (3-5% of the world's population at the time). The second pandemic was the 1957 'Asian influenza' caused by an H2N2 subtype, and the third pandemic was the 1969 'Hong Kong influenza' caused by an H3N2 subtype. These pandemics were much less severe than the 1918 pandemic, but still resulted in an estimated 1 to 4 million deaths worldwide. The fourth pandemic, the 1977 'Russian influenza', was again caused by an H1N1 virus which appears to be almost identical to the H1N1 strain circulating prior to 1950. Because of this, pre-existing immunity in the older generation lead to a very mild pandemic, mainly affecting children and young adults. In April 2009, a fifth influenza pandemic was described which was again caused by an H1N1 virus. This pandemic H1N1 (pH1N1) virus resulted from a triple reassortant of avian, swine and human influenza viruses reassorting with a Eurasian swine influenza virus, thus gaining the name 'Swine flu' (Neumann et al., 2009). It is estimated that in some regions, particularly Africa and Southeast Asia, up to 20-40% of the population were infected by pH1N1 and that 284,500 people died within the first twelve months of the pandemic (Dawood et al., 2012).

Since 1997, several highly pathogenic avian influenza (HPAI) viruses have emerged that are able to infect humans directly (Claas et al., 1998). Currently, HPAI A/H5N1 and low pathogenic avian influenza (LPAI) A/H7N9 strains are considered to represent the biggest threats to global health. However, other avian (e.g. H7N7 and H9N2) and swine viruses

(e.g. H1 and H3 subtypes) also sporadically infect humans. To date (July 2017), there have been 859 reported human cases of H5N1 with 453 deaths in 16 countries, resulting in a 53% mortality rate (WHO, 2017a). For H7N9 1557 human cases and at least 605 deaths have been reported as of July 2017 in China resulting in a mortality rate of 39% (WHO, 2017b). Although these viruses can cause severe disease in humans, infection almost entirely occurs through close contact with poultry. Despite a limited number of presumed human-to-human transmission incidents, transmission between humans with these viruses is highly inefficient. Nevertheless, recent studies have shown that these avian influenza viruses only require a few mutations to become transmissible in ferret and guinea pig models (Herfst et al., 2012; Imai et al., 2012; Zhang et al., 2013). Therefore, the concern is that viruses (like H5N1 and H7N9) that already have shown to be able to infect humans could acquire the necessary adaptive mutations to allow efficient human-to-human transmission.

In the present world with ever increasing globalisation, rapid population growth and cheap air travel, future influenza virus pandemics are inevitable. It is therefore crucial to understand the molecular biology of the virus in order to design and develop novel antiviral strategies and to heighten our pandemic preparedness.

1.2 Influenza virus taxonomy and nomenclature

Influenza viruses belong to the *Orthomyxoviridae* family which is divided into seven genera: *Influenzavirus A*, *Influenzavirus B*, *Influenzavirus C*, *Influenzavirus D*, *Thogotovirus*, *Isavirus* and *Quarjanavirus* (Palese & Shaw, 2013). *Orthomyxoviridae* are defined as single-stranded, negative-sense RNA viruses with a segmented RNA genome that replicate in the nucleus of infected host cells. Due to the segmented RNA genomes, these viruses can

undergo genetic reassortment and exchange genetic information with viruses of the same genus.

While influenza B viruses (humans, seals and ferrets), influenza C viruses (humans, dogs and swine) and influenza D viruses (cattle) only infect a relatively narrow range of hosts, influenza A viruses have a much broader host range. Aquatic birds are their primary reservoir, but influenza A viruses also infect humans, swine, horses, bats, seals, whales, poultry and non-aquatic birds (Neumann & Kawaoka, 2015).

Viruses within the influenza A virus genus can be classified further according to their serological subtype, determined by the viral haemagglutinin (HA) and neuraminidase (NA) glycoproteins. Currently, there are 18 HA (H1-H18) and 11 NA (N1-N11) subtypes described. However, it is likely that this is an underrepresentation of all subtypes present. Influenza virus isolates are named according to their genus, host species, geographical location of isolation, strain identification number, year of isolation and for influenza A viruses the HA and NA subtypes. Therefore, an influenza A virus isolated in 2011 from a mallard in Sweden with the strain number 133546 and a subtype of H10N4 would be named A/mallard/Sweden/133546/2011 (H10N4). If the influenza virus was isolated from a human, then the species name is omitted (Wright et al., 2013).

1.3 The influenza A virion

Influenza A virions are pleomorphic particles that take on spherical, filamentous or bacilliform morphologies (Wright et al., 2013). Spherical virions are approximately 100 nm in diameter, whereas filamentous particles can range from 300 nm to several micrometers in length (Calder et al., 2010; Harris et al., 2006). A schematic diagram and electron micrographs of influenza A virions are presented in Figure 1.1.

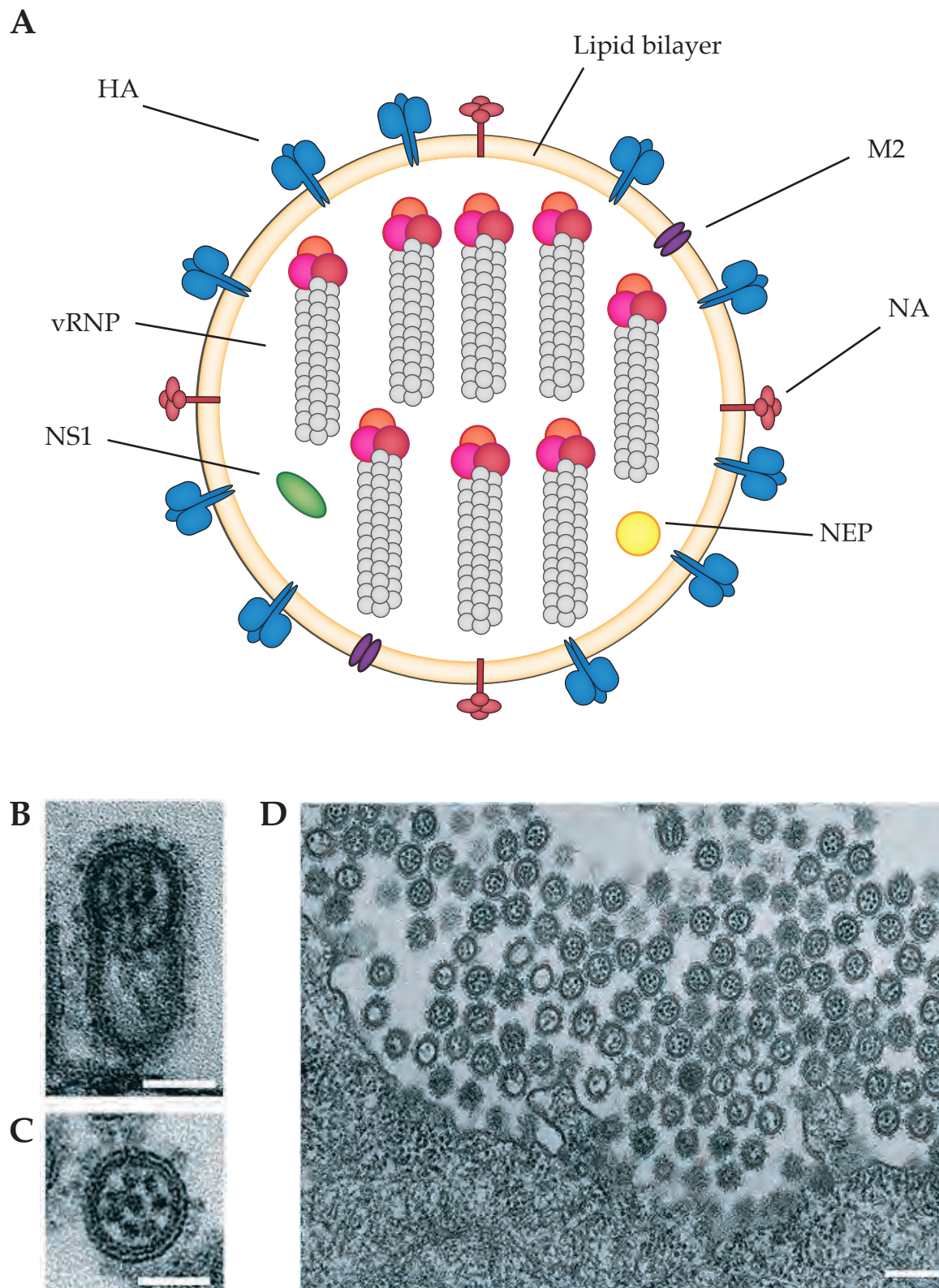


Fig. 1.1 The influenza A virion. (A) Schematic of the influenza A virion showing all structural proteins and vRNP complexes. (B to D) Negative stain electron micrograph of influenza A/WSN/33 (H1N1) virions in (B) vertical and (C) horizontal planes and (D) of budding virions from Madin-Darby Canine (MDCK) cells, adapted from Noda *et al.*, 2006. Scale bars are 50 nm (B and C) and 200 nm (D).

Table 1.1 Summary of influenza A virus genome segments and encoded proteins

Segment	Length (nucleotides)	Encoded protein(s)	Length (amino acids)	Reference
1	2341	PB2	759	Field & Winter, 1982
		PB2-S1	508	Yamayoshi <i>et al.</i> , 2016
2	2341	PB1	757	Winter & Fields, 1982
		PB1-N40	718	Wise <i>et al.</i> , 2009
		PB1-F2	87	Chen <i>et al.</i> , 2001
3	2233	PA	716	Field & Winter, 1982
		PA-X	252	Jagger <i>et al.</i> , 2012
		PA-N155	562	Muramoto <i>et al.</i> , 2013
		PA-N128	535	Muramoto <i>et al.</i> , 2013
4	1778	HA	566	Winter <i>et al.</i> , 1981a
5	1565	NP	498	Winter & Fields, 1981
6	1413	NA	454	Fields <i>et al.</i> , 1981
7	1027	M1	252	Winter & Fields, 1980
		M2	97	Winter & Fields, 1980
		M3	9	Lamb <i>et al.</i> , 1981
		M4	54	Shih <i>et al.</i> , 1998
		M42	99	Wise <i>et al.</i> , 2012
8	890	NS1	230	Winter <i>et al.</i> , 1981b
		NEP	121	Winter <i>et al.</i> , 1981b
		NS3	188	Selman <i>et al.</i> , 2012
		NEG8	216	Zhirnov <i>et al.</i> , 2007

Table 1.2 Summary of the known functions of influenza A virus encoded proteins

Encoded protein	Function
PB2	RdRP subunit responsible for cap-binding. Innate immunity antagonist. Host range determinant.
PB2-S1	PB2-related protein. Largely mitochondrial localisation, putative interferon antagonist.
PB1	RdRP subunit harbouring the catalytic domain.
PB1-N40	PB1-related protein. Unknown function.
PB1-F2	Apoptosis regulator and virulence factor.
PA	RdRP responsible for endonucleolytic cap cleavage.
PA-X	Modulates host immune response, contains endonuclease activity.
PA-N155	PA-related protein. Unknown function.
PA-N128	PA-related protein. Unknown function.
HA	Trimeric transmembrane surface glycoprotein. Binds sialic acid on host cell surface glycoprotein receptors. Facilitates virus receptor binding. Host range determinant. Virulence factor.
NP	Encapsidates vRNA and cRNA and the major constituent of viral ribonucleoprotein complexes. Necessary for transcription and replication.
NA	Tetrameric transmembrane surface glycoprotein. Cleaves sialic acid on the surface of host cells to facilitate virus release.
M1	Major structural protein of the virion, forming a scaffold matrix inside the virion. Essential for virus assembly and budding. Involved in nuclear export of vRNPs. Binds vRNP complexes as well as HA and NA.
M2	Tetrameric transmembrane protein. Acts as a proton ion channel that acidifies the virion and facilitates vRNP release. Important for viral morphogenesis, budding and membrane scission.
M3	Unknown function.
M4	Unknown function.
M42	M2-related protein. Largely retained in the Golgi compartment. Unknown function.
NS1	Interferon antagonist. Binds to dsRNA. Inhibits polyadenylation of pre-mRNAs. Binds to CPSF30 and PABP1I.
NEP	Mediates nuclear export of vRNPs. Regulates viral RNA transcription and replication. Minor virion component.
NS3	NS1-related protein. Potential role in host adaptation.
NEG8	Unknown function.

The virion is enveloped with a lipid bilayer which is acquired during budding from the host cell membrane. Embedded in the envelope are the trimeric haemagglutinin (HA) and tetrameric neuraminidase (NA) membrane glycoproteins as well as the tetrameric M2 ion channel. Approximately four HA trimers are embedded in the viral envelope for every NA tetramer, with M2 tetramers being far less abundant (Harris et al., 2006; Palese & Shaw, 2013). Inside the viral envelope the major structural component of the virion, the matrix protein M1, forms a matrix structure. The viral genome is packaged in eight individual viral ribonucleoprotein (vRNP) complexes, each of which consists of a single-stranded viral RNA (vRNA) segment that is encapsidated by viral nucleoprotein (NP) and bound at the 5' and 3' ends by the viral RNA-dependent RNA polymerase (RdRP), a heterotrimer formed by polymerase basic 1 (PB1), polymerase basic 2 (PB2) and polymerase acidic (PA) subunits (Arranz et al., 2012; Moeller et al., 2012). In addition to these structural proteins, non-structural protein 1 (NS1) and nuclear export protein (NEP) have been found inside influenza A virions, but it remains unclear whether their incorporation is required or an accidental by-product (Hutchinson et al., 2014; Richardson & Akkina, 1991). Influenza viruses also encode several auxiliary proteins, e.g. PA-X and PB1-F2, that fulfil a wide range of functions during infection and that are not known to be incorporated into influenza virions (Vasin et al., 2014). Furthermore, a number of host factors are also incorporated into influenza virions (Shaw et al., 2008).

The genome of influenza A viruses is comparably small and comprises a total of 13.6 kilobases (kb), with individual genome segments ranging from 0.9 to 2.3 kb in size.

A summary of the influenza A virus genome segments and the known encoded proteins is presented in Table 1.1. The known functions of each of the encoded proteins is described in Table 1.2.

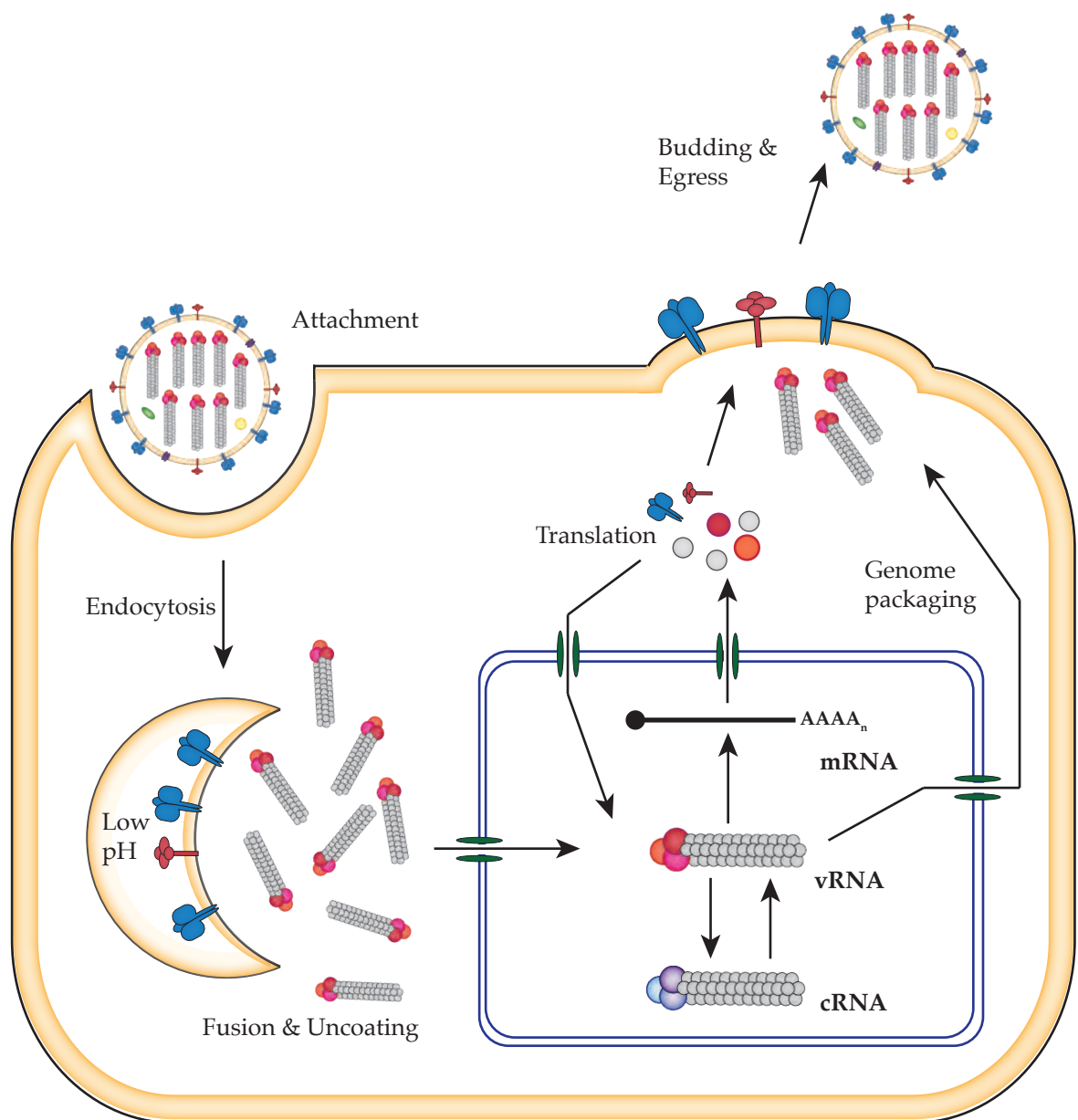


Fig. 1.2 The influenza A virus replication cycle. Influenza A virus infection begins with attachment via the viral HA glycoprotein to host cellular receptors that contain terminal α -2,3-linked or α -2,6-linked sialic acids. Viral entry occurs by receptor-mediated endocytosis. Acidification of the endosome results in a conformational change in the viral HA protein, exposing a fusion peptide that facilitates viral-endosome membrane fusion, releasing vRNP complexes into the cytoplasm. vRNP complexes are subsequently imported into the host cell nucleus, where the vRNA genome is transcribed and replicated by the viral RdRP. Genomic vRNA is replicated via a complementary positive-sense cRNA. Viral mRNAs are exported to the cytoplasm for translation of viral proteins. Progeny vRNP complexes are exported to the cytoplasm and are packaged into progeny virions at the plasma membrane along with M1, M2, NS1 and NEP. Progeny virions are released by budding.

1.4 The influenza A virus life cycle

Influenza A viruses commonly infect and replicate in epithelial cells of the respiratory and gastrointestinal tracts, but have also been shown to replicate in alveolar macrophages and dendritic cells (Hofmann et al., 1997; Seo et al., 2004; Smed-Sørensen et al., 2012). The life cycle of influenza viruses is relatively fast (8 to 10 hours in tissue culture) as is typical of viruses causing acute infections (Wright et al., 2013). Multiple interactions between the virus and the host cell take place during infection, which can regulate pathogenesis, immune response and host range restriction (Cauldwell et al., 2014; Matsuoka et al., 2013). A schematic diagram of the influenza A virus life cycle is presented in Figure 1.2.

1.4.1 Attachment and entry

Influenza A virus initiates infection and host cell entry by binding of the viral membrane glycoprotein HA to sialic acids present on cellular glycoproteins on the surface of the cell membrane. Sialic acids can be classified by their sialic acid linkage to the penultimate galactose in an N-linked glycan chain of the cellular receptor (Connor et al., 1994; Couceiro et al., 1993). Because influenza A viruses have evolved to utilise a single type of sialic acid specific to their host species, preference for a certain type of sialic acid is a strong determinant of host tropism. Sialic acids linked from C2 to C3 of galactose (α -2,3-linked sialic acid) are preferentially bound by avian influenza A viruses and are predominantly found in the gastrointestinal tract of birds and in the lower respiratory tract of mammals. In contrast, sialic acids linked from C2 to C6 of galactose (α -2,6-linked sialic acid) are preferentially bound by mammalian influenza A viruses and are predominantly found in the upper respiratory tract of mammals. It is possible for epithelia to have a mixed population of α -2,3-linked sialic acids and α -2,6-linked sialic acids, as is most notably the case for swine. Therefore, swine are considered a 'mixing

vessel' for genetic reassortment as their epithelia can be infected with avian and mammalian influenza viruses simultaneously (Ito et al., 1998). Once attachment of the virus has taken place through sialic acid binding, the virion is internalised into the host cell by receptor-mediated endocytosis. While clathrin-mediated endocytosis is the most widely accepted model for influenza virus entry, caveolae-dependent, macropinocytosis-dependent, caveolae-independent and clathrin-independent pathways have also been described (Edinger et al., 2014; Mercer et al., 2010). It remains debatable whether these different pathways represent flexible redundancy by the virus or whether different entry pathways are required in different cell types. In any case, endocytosis leads to the internalisation of the virion within an early endosome, from which the viral genome and the replication machinery needs to escape to establish a productive infection.

1.4.2 Fusion, uncoating and nuclear import

Once successfully internalised into the early endosome, the influenza A virion must fuse with the endosomal membrane to release its vRNP complexes into the host cell cytoplasm. This process relies on a structural rearrangement of the viral HA protein that is triggered by a decrease in pH within the maturing and acidifying endosome (Hamilton et al., 2012). For fusion to occur the viral HA protein must first have been cleaved from the HA0 precursor into HA1 and HA2 subunits by host-specific proteases either in the preceding round of viral infection (Boulay et al., 1987; Chen et al., 1998) or within the endosome (Zhirnov et al., 2002). The proteolytic cleavage sites in the HA protein have a relatively high sequence diversity across viral subtypes. However, multi basic cleavage sites, characteristic of highly pathogenic avian influenza viruses, are cleaved by ubiquitously expressed subtilisin-like proteases and therefore significantly enhance the virulence of

these viruses due to a higher likelihood of successful HA cleavage in a wide range of host species and tissues (Stieneke-Gröber et al., 1992; Taubenberger & Kash, 2010). The decrease in pH inside the acidifying endosome triggers a structural rearrangement of the cleaved HA protein resulting in the exposure of a hydrophobic fusion peptide in HA2. Insertion of this fusion peptide into the endosomal membrane causes fusion of the endosomal membrane with the viral membrane (Bullough et al., 1994). Simultaneously, the viral M2 ion channel pumps H⁺ ions from the endosome into the virion, acidifying the virion interior (Pinto et al., 1992), thus disrupting the weak interactions between M1 and vRNP complexes and releasing the vRNPs from the virion structure (Bui et al., 1996). Taken together, these processes result in the transfer of vRNP complexes from the virion interior into the host cell cytoplasm.

Once vRNP complexes have been released into the cytoplasm they are actively transported into the cell nucleus by the importin- α/β pathway (Hutchinson & Fodor, 2013). All protein constituents of a vRNP complex contain at least one nuclear localisation signal (NLS), however, nuclear import of vRNP complexes is mediated by the NLS of NP which is exposed after dissociation of the interaction of M1 with vRNPs (Cros et al., 2005; Martin & Helenius, 1991b; O'Neill et al., 1995). The NLS of NP is bound by importin- α (either importin- α -1 or importin- α -5) which subsequently itself is bound by importin- β , allowing the protein complex to dock at the nuclear pore complex (NPC) and to translocate into the nucleus.

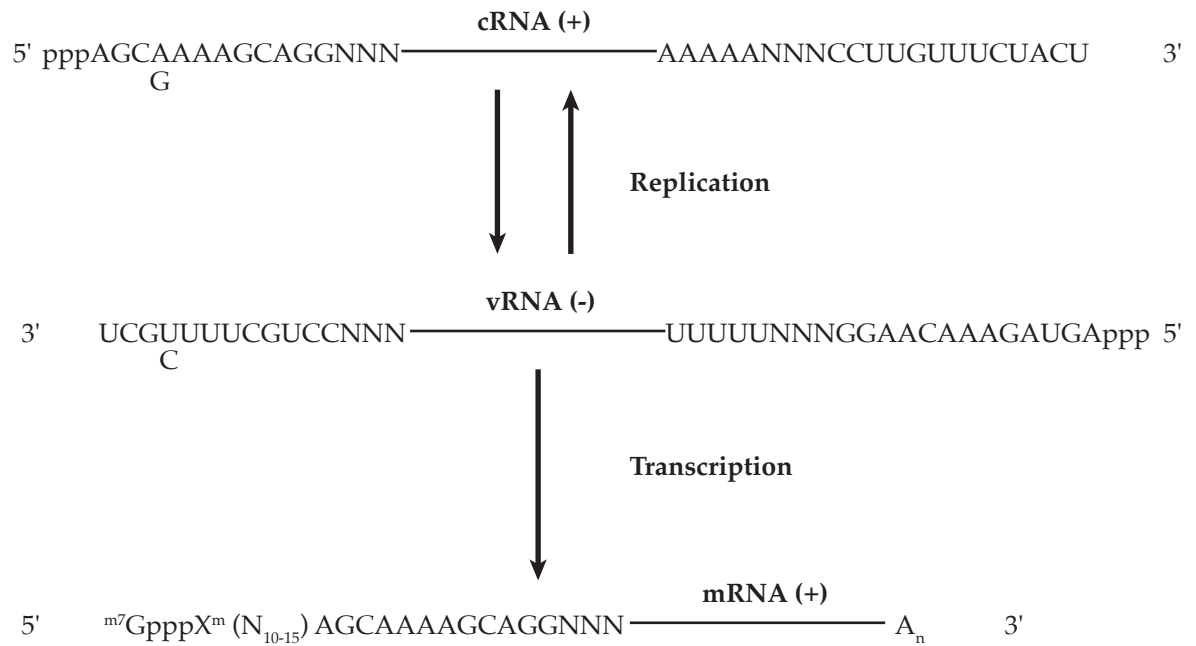


Fig. 1.3 Influenza A virus RNA synthesis. Schematic of the influenza virus RNA species synthesized by the viral RdRP. The parental negative-sense vRNA is shown in the center with the conserved non-coding sequences at each terminus, adjacent to segment-specific nucleotides (NNN). The polyadenylation signal consists of a poly(U) stretch at the 5' end of the vRNA. Viral mRNA derives its cap and 10–15 5' nucleotides from host pre-mRNAs and has a 3' poly(A) tail. The vRNA genome is replicated via cRNA which is an exact copy of the parental vRNA. The variation at position 4 in the 3' vRNA non-coding region is shown. Adapted from Palese & Shaw, 2013.

1.4.3 Viral gene expression

After vRNPs have been translocated into the host cell nucleus, the resident viral RdRP in vRNPs is responsible for transcribing viral messenger RNA (mRNA) and replicating the negative-sense viral RNA (vRNA) via a positive-sense complementary RNA (cRNA) intermediate to produce progeny vRNPs (Velthuis & Fodor, 2016). A schematic overview is presented in Figure 1.3 and a detailed account of viral RNA syntheses can be found in Chapter 1.5.

Viral gene expression appears to be regulated as different viral proteins are expressed in unequal amounts during infection (Enami et al., 1985; Hatada et al., 1989; Shapiro et al., 1987). Viral NS1 and NP are expressed in high amounts early on in infection, while HA, NA and M1 are all expressed at higher levels late in infection (Shapiro et al., 1987; Smith & Hay, 1982; Yángüez & Nieto, 2011). This differential temporal gene expression is beneficial to the virus. High levels of NP are required for efficient viral genome replication (Turrell et al., 2013) and NS1 plays an essential role in inhibiting the host innate immune response by antagonising interferon stimulation (Hale et al., 2008) and therefore NP and NS1 are necessary early on in infection. M1 has been shown to be inhibitory to viral genome replication (Perez & Donis, 1998; Watanabe et al., 1996) and is needed for nuclear export of vRNP complexes (Baudin et al., 2001; Bui et al., 2000; Martin & Helenius, 1991a), while HA and NA are required for virion assembly in the final stages of the viral life cycle, so late expression of these viral proteins is in the interest of the virus.

The mechanism of temporal regulation of viral gene expression is not understood yet, but several processes which might be involved have been proposed, such as differential RNA synthesis, RNA stability and regulation of translation (Shapiro et al., 1987; Smith & Hay, 1982; Yángüez & Nieto, 2011). Expression of the viral RdRP

subunits (PA, PB1, PB2) has been proposed to be regulated by the viral promoter sequence in the corresponding viral RNA segments (Lee & Seong, 1998; Lee et al., 2003). Viral gene segments encoding the RdRP subunits contain a C at position 4 from the 3' end of the vRNA, while all other segments contain a U residue at this position. This single nucleotide substitution seems to upregulate viral replication while downregulating viral mRNA synthesis, suggesting that the viral RdRP might interact differentially with the different promoters (Lee et al., 2003).

A recent study has suggested that suboptimal splicing of viral genome segments is involved in regulating the balance between viral transcription and replication (Chua et al., 2013). The usage of a suboptimal splice site in NS1 mRNA leads to a gradual accumulation of NEP mRNA and expression during an infection, which has been shown to be essential for properly timed nuclear export of vRNP complexes in infected cells.

1.4.4 Nuclear export and trafficking of vRNP complexes

After viral genome replication and vRNP assembly in the host cell nucleus, vRNP complexes must be transported through the nuclear membrane into the cytoplasm and to the cell membrane for packaging into progeny virions. The viral M1 and NEP proteins are essential for this process (Paterson & Fodor, 2012). M1 associates with vRNP complexes and is thought to be involved both in vRNP assembly and nuclear export of vRNP complexes (Bui et al., 2000; Huang et al., 2001; Martin & Helenius, 1991b). While M1 has been shown to be essential for nuclear export, on its own it is not sufficient. NEP has been

suggested to act as an adaptor between vRNP-M1 complexes and the importin- β protein CRM1 in a daisy chain model of nuclear vRNP export (Neumann et al., 2000; O'Neill et al., 1998). Here, the C-terminal domain of M1 forms a strong interaction with NP on vRNPs, while the N-terminal domain binds the C-terminus of NEP. This interaction is thought to block the NLS on the N-terminal domain of M1, thus preventing re-import of M1-vRNP complexes after nuclear export. CRM1, which is bound by its nuclear co-factor Ran-GTP, recognises the nuclear export signal (NES) on the N-terminus of NEP and mediates export of the whole vRNP-M1-NEP complex from the nucleus to the cytoplasm via the nuclear pore complex (NPC). In a slightly refined model, NEP interacts with CRM1 as before through its N-terminal domain, but its C-terminus interacts with both M1 and the viral RdRP via PB1 and PB2, facilitating M1 interaction with NP (Brunotte et al., 2014). Interaction between NEP and the RdRP could also explain why vRNPs but not complementary RNPs (cRNPs) are exported from the nucleus. The structure of the viral RdRP is affected by the binding of viral promoter RNA as suggested by recently published crystal structures (Hengrung et al., 2015; Pflug et al., 2014; Thierry et al., 2016), therefore allowing for a model where NEP can interact with vRNP-associated but not cRNP-associated RdRPs, thereby providing specificity for nuclear export.

It is likely however, that nuclear export is not solely achieved through this model and that there is some redundancy in the nuclear export mechanism. CRM1-dependent nuclear export mediated by both M1 and NP have also been proposed, as has a CRM1-independent M1-mediated pathway (Bui et al., 2000; Elton et al., 2001; Yu et al., 2012).

Several mechanisms have been proposed to promote the timing of nuclear export. Caspase 3 is activated late in infection and has been shown to increase diffusion limits of nuclear pores, thereby enhancing nuclear export (Wurzer et al., 2003).

Furthermore, accumulation of HA at the cell membrane activates the mitogen-activated protein kinase signalling cascade (Pleschka et al., 2001). Moreover, NEP is expressed from a spliced version of NS1 mRNA using a suboptimal 5' splice site, thus guaranteeing a delayed accumulation of NEP which acts as a molecular timer during infection (Chua et al., 2013).

After nuclear export, vRNPs accumulate at the perinuclear microtubule organising center (MTOC), where they interact with Rab11 through PB2. Rab11 subsequently directs transport of vRNP complexes to the cell membrane via vesicular anterograde microtubule transport (Amorim et al., 2011; Bruce et al., 2010; Momose et al., 2011). A recent study showed that different vRNPs form subcomplexes in the cytoplasm which then assemble into the complete eight segment vRNP complex during transport to the cell membrane (Lakdawala et al., 2014).

1.4.5 Genome packaging

For successful production of infectious viral particles, influenza A viruses need to make sure that exactly eight vRNPs are packaged into the progeny virions, one of each genome segment. Two general packaging mechanisms have been proposed: one that relies on random vRNP incorporation and which has been supported by the observation that most influenza virus particles are non-infectious because they lack one or more genome segments (Hutchinson et al., 2010; Brooke et al., 2013), and one that relies on selective packaging relying on RNA-RNA interactions to incorporate a single copy of each genome segment into progeny virions (Gerber et al., 2014). Indeed, microscopic and biochemical data support a selective packaging strategy by confirming that the majority to influenza

virions contain a single copy of each genomic segment (Bergmann & Muster, 1995; Chou et al., 2012; Gao et al., 2012; Smith & Hay, 1982). To achieve this, specific RNA-RNA interactions are thought to regulate assembly and incorporation of the full set of vRNP into virions (Hutchinson et al., 2010). Electron microscopic evidence and 3D tomographic analysis supports a 7 + 1 packaging arrangement, where a central vRNP is surrounded by the seven remaining vRNPs and that the relative position of vRNPs is conserved (Noda et al., 2006). The same study reports that the different vRNPs differ in length, concurring with the different lengths of genomic segments and that vRNPs form connections with each other, assembling into a supramacromolecular complex inside the virion (Fournier et al., 2012a; Fournier et al., 2012b; Sugita et al., 2013). Specific bipartite packaging signals are located in the 5' and 3' untranslated regions (UTRs) of vRNAs (Hutchinson et al., 2010). Specific regions of RNA that interact *in vitro* between all vRNA segments and RNA sequences that specifically interact between vRNA pairs have since been identified using reverse genetics, supporting a specific packaging mechanisms through intermolecular RNA-RNA interactions (Fournier et al., 2012b; 2012a; Gavazzi et al., 2013).

1.4.6 Virion assembly and egress

To complete the viral replication cycle, new virions must be assembled and exit the host cell. Influenza A virions assemble and bud from lipid raft domains on the apical surface of the cell membrane of polarised cells (Rodriguez Boulan & Sabatini, 1978; Scheiffele et al., 1999; Simons & Ikonen, 1997). The viral transmembrane proteins HA, NA and M2 are translocated from the sites of protein synthesis at the endoplasmic reticulum via the Golgi apparatus to the apical membrane. The HA and NA glycoproteins are actively targeted to the lipid raft domains, causing broadening and coalescence of these domains (Chen et al.,

2007a; Leser & Lamb, 2005; Takeda et al., 2003). Accumulation of HA and NA clusters in the lipid raft domains of the cell membrane has been suggested to cause membrane deformation and to initiate budding (Chen et al., 2007a). M1 is recruited to the budding virion by interacting with the cytoplasmic tails of HA and NA. M1 subsequently polymerises to form the interior matrix of the virion (Ali et al., 2000; Noton et al., 2007; Zhang et al., 2000). The M1 matrix then acts as a scaffold for the incorporation of the eight vRNP complexes (Noda et al., 2006). M2 incorporation is also mediated by M1 (Chen et al., 2008; McCown & Pekosz, 2006) and is thought to cause membrane curvature at the neck of the budding virus, resulting in scission of the membrane and release of the nascent virion into the extracellular space (Rossman et al., 2010; Rossman & Lamb, 2011). The NA glycoprotein on the surface of the nascent virion is responsible for hydrolysing sialic acids on the cell surface to prevent reattachment of progeny virions and reinfection of the same cell (Air & Laver, 1989; Huang et al., 2008).

1.5 Influenza virus RNA synthesis

1.5.1 Viral ribonucleoprotein complexes

Each viral genome segment inside an influenza virus virion exists in the form of a single ribonucleoprotein (RNP) complex. Here, the single-stranded viral RNA segment is wound around a double-helical scaffold of multiple viral nucleoproteins (NP) with a loop at one of the ends of the RNP structure and the viral RdRP bound to the 5' and 3' ends of the viral RNA at the other end (Figure 1.4A). When visualised by negative stain electron microscopy, vRNPs form ribbon-like superhelical filaments (Compans et al., 1972; Heggeness et al., 1982; Pons et al., 1969). Isolated cRNPs form a corresponding structure,

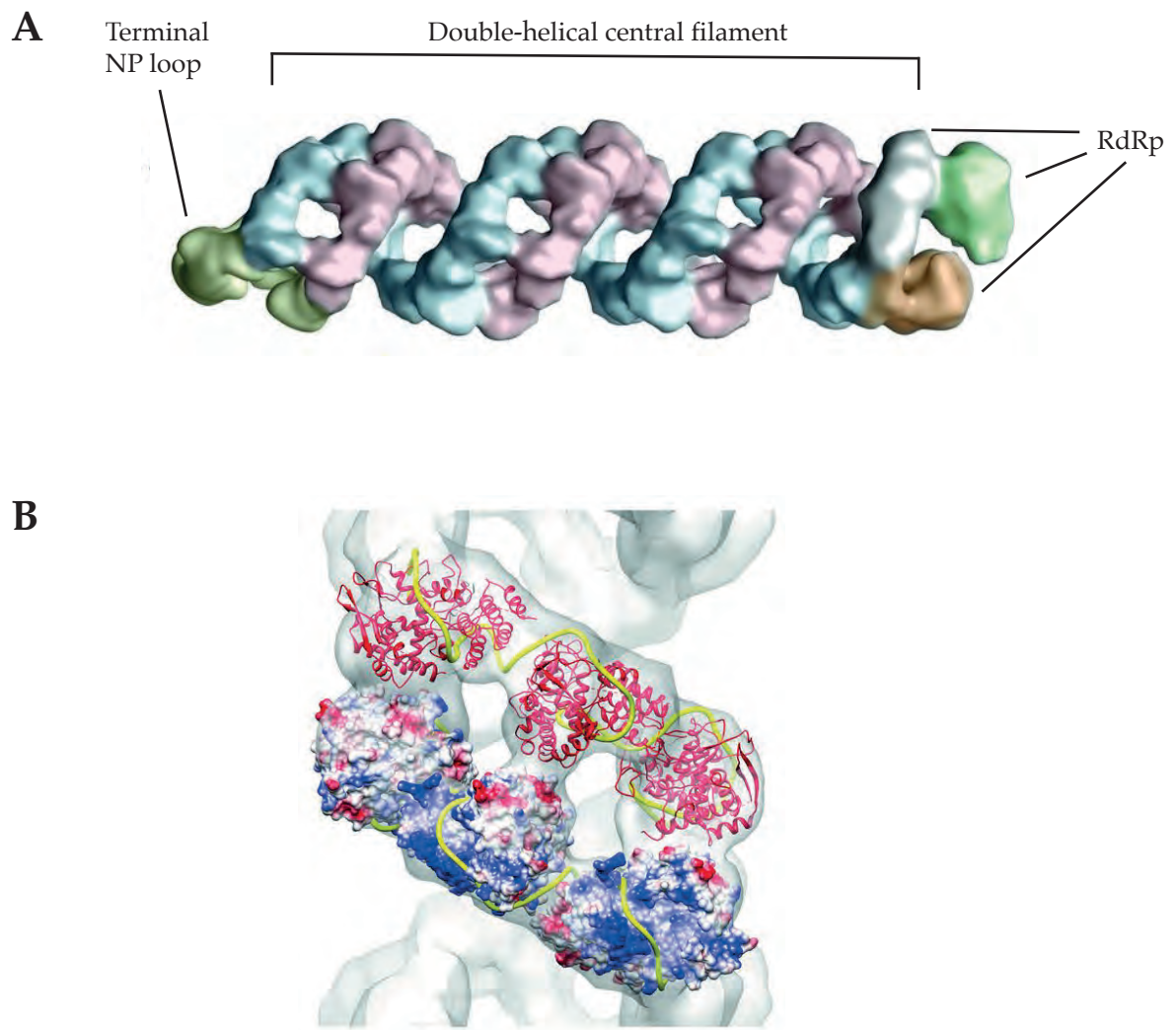


Fig. 1.4 The structure of the influenza A virus vRNP complex. (A) Cryo-electron microscopy model for the structure of the influenza A/WSN/33 virus vRNP. Double-helical central filament (blue and pink), terminal NP loop (yellow) and the viral RdRp (green, grey and brown) are shown. (B) Model for vRNA localisation (yellow thread) in the vRNP structure showing one NP strand with surface potential and the opposite strand as a ribbon representation. Adapted from Arranz et al., 2012.

containing NP and the viral RdRP, as observed in negative stain electron microscopy (York et al., 2013). The oligomerisation of NP is mainly responsible for the overall structural features of RNPs as NP in the absence and presence of viral RNA forms helical structures resembling RNPs *in vitro* (Ruigrok & Baudin, 1995; Yamanaka et al., 1990). The interaction of viral RNA with NP is sequence independent and instead relies on the negatively charged phosphate backbone of the RNA and positively charged basic patches on NP (Baudin et al., 1994; Elton et al., 1999; Ng et al., 2008; Ye et al., 2006). However, a recent study identified that the NP-binding region of vRNA is enriched for guanine but depleted of uracil, suggesting some sort of sequence bias towards NP binding (Lee et al., 2017). In the context of an RNP, each NP monomer binds approximately 24 nucleotides of RNA (Ortega et al., 2000). High resolution cryogenic electron microscopy (cryo-EM) of vRNP complexes has revealed more details about the vRNP structure as presented in Figure 1.4 (Arranz et al., 2012; Moeller et al., 2012). The double helical NP structure is formed by two NP strands with opposite polarity and contains both a major and minor groove between the strands. NP monomers interact with neighbouring NP monomers on the same strand (intra-strand interactions) as well as with NP monomers on the opposing strand (inter-strand interactions). The rise in the helix is approximately 30 Å with a twist of 60 - 70° per NP monomer, resulting in about five to six NP monomers per complete helical turn. Electrostatic modelling of the NP double helix shows a continuous positively charged pathway on the outer surface of the double helix that could bind the vRNA template (Figure 1.4B). This model is compatible with the fact that influenza virus vRNA is particularly sensitive to ribonuclease treatment compared to vRNA of other negative-strand RNA viruses (Duesberg, 1969; Ruigrok et al., 2010). Classically, the NP was thought to bind to vRNA in a uniform pattern, more recent evidence points to the fact that each vRNP segment has a unique NP-binding profile and that vRNA forms secondary structures protruding from the NP double helix (Lee et al., 2017).

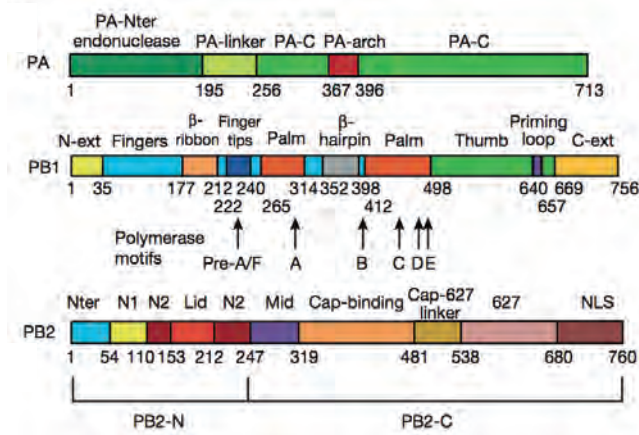
The loop end of the NP double helix is formed by a varying number of NP monomers, ranging from three to seven in the published vRNP structures (Arranz et al., 2012; Moeller et al., 2012). The RdRP-containing end of the double helix is found in two different conformations in one study, where a large domain of the RdRP is shifted compared to the other conformation (Arranz et al., 2012). In the second study the RdRP is depicted as a large domain with a smaller flexible domain (Moeller et al., 2012). This flexible domain is thought to be the PA C-terminal domain (He et al., 2008). However, the RdRP structure of a recombinant mini-RNP with nine NP monomers shows a more compact structure with no flexible appendices (Area et al., 2004; Coloma et al., 2009).

1.5.2 The viral RNA-dependent RNA polymerase

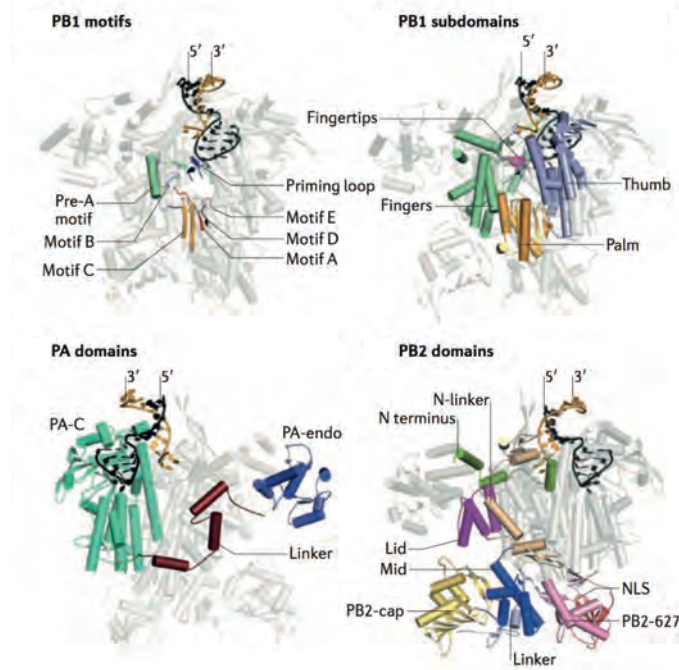
1.5.2.1 The structure of the heterotrimeric viral RdRP

In recent years, several structures of the influenza A, influenza B and influenza C virus RdRP have been solved through X-ray crystallography and cryo-EM (Chang et al., 2015; Hengrung et al., 2015; Pflug et al., 2014; Reich et al., 2014; Thierry et al., 2016). The influenza virus RdRP consists of the three subunits PA, PB1 and PB2 (Fig. 1.5A). The PB1 subunit forms the core of the RdRP and displays the typical motifs of RNA-dependent RNA polymerases (Fig. 1.5B), which are pre-A motif (or motif F) and motifs A-E (Velthuis, 2014). It also has the canonical right-hand-like fold with fingers, palm and thumb regions. The N-terminal endonuclease domain and the C-terminal domain of the PA subunit (or P3 in influenza C viruses) sit on opposite ends of PB1 and a long linker region wrapping

A



B



C

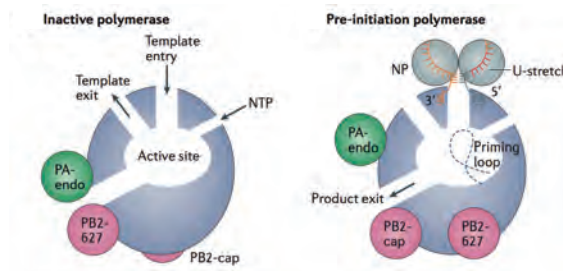


Fig. 1.5 The structure of the influenza A virus RdRP. (A) Schematic of the domains of the RdRP subunits (adapted from Pflug *et al.*, 2014). (B) Models of the influenza A virus RdRP bound to vRNA (PDB entry 4WSB) with the different domains and motifs of PA, PB1 and PB2 highlighted. The 5' and 3' termini of the vRNA are shown in dark grey and yellow, respectively. (C) Schematics of the RdRP in the inactive state, showing the positions of the active site and proposed channels. In the absence of vRNA, the PA endonuclease and PB2 cap-binding domains are arranged in a conformation incompatible with cap-snatching. In the transcription pre-initiation state, a conformational rearrangement of the PA endonuclease, the PB2 cap-binding and PB2 627-domains results in an structure compatible with cap-snatching (adapted from te Velthuis & Fodor, 2016).

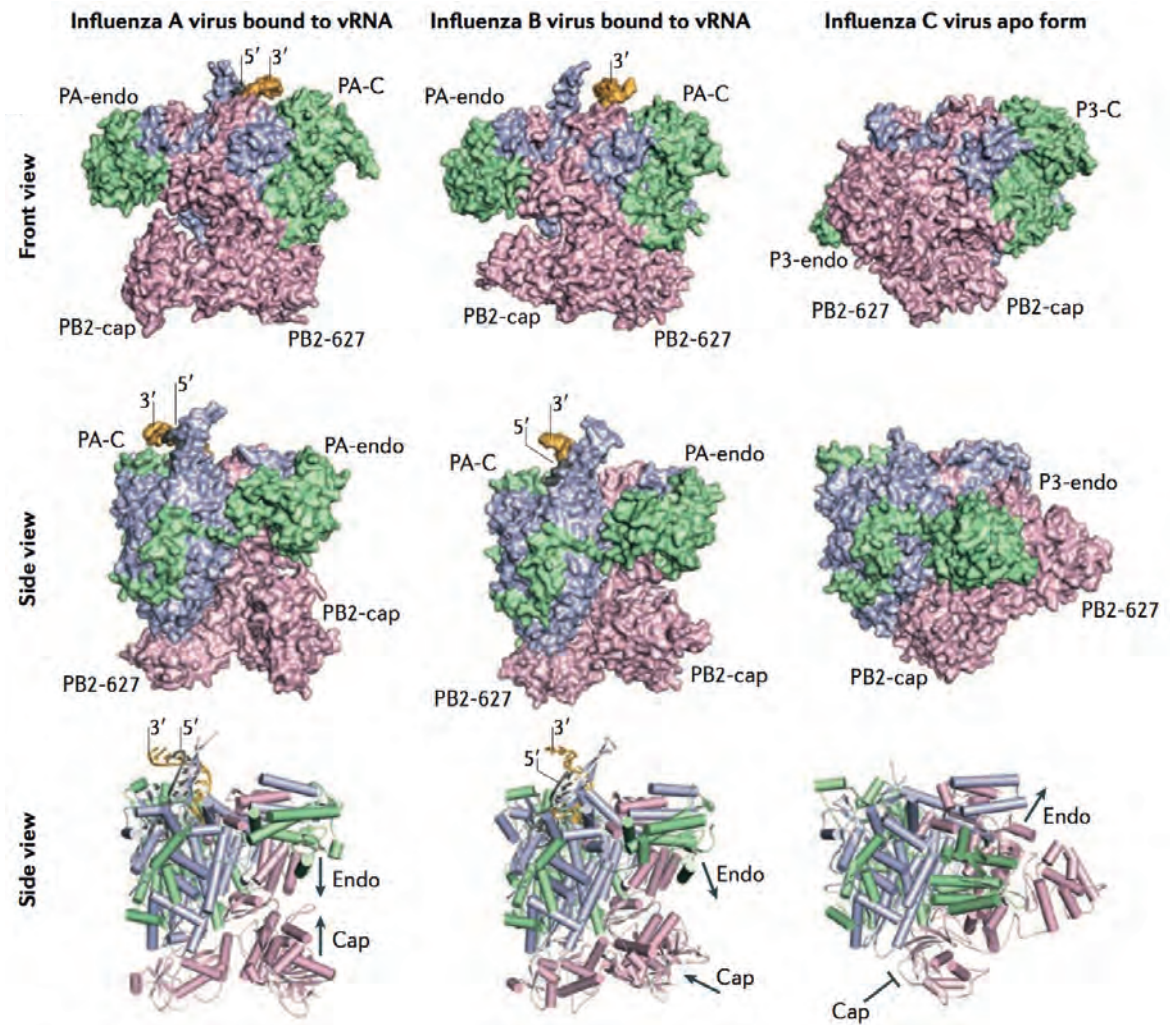


Fig. 1.6 The structure of the RdRP of influenza A, influenza B and influenza C viruses. (A) Front and side views of the influenza A virus (Protein Data Bank (PDB) entry 4WSB), influenza B virus (PDB entry 4WSA) and influenza C virus (PDB entry 5D98) RdRP, shown in surface (top and middle rows) and cartoon model (bottom row) representations. The three subunits, PB1, PB2 and PA (or P3 for influenza C virus) are coloured in purple, pink and green, respectively. The PB2 cap-binding domain (PB2-cap), PB2 627-domain (PB2-627), PA and P3 endonuclease domain (PA-endo and P3-endo), and PA and P3 C-terminal domain (PA-C and P3-C) are indicated. In the influenza A and influenza B virus RdRP structures, the 5' and 3' termini of a vRNA are shown in dark grey and yellow, respectively; by contrast, the influenza C virus RdRP is in the apo form (unbound to RNA). In the influenza A virus RdRP structure, the cap-binding and endonuclease domains face each other, configured for cap-snatching (left, bottom row). In the influenza B virus RdRP structure, the cap-binding domain is rotated to face the centre of the RdRP, which is consistent with the insertion of the capped primer into the active site through the product exit channel (middle, bottom row). In the influenza C virus RdRP structure, the cap-binding pocket of PB2 is blocked, and the P3 endonuclease domain is oriented away from the PB2 cap-binding domain, incompatible with cap-snatching (right, bottom row). Adapted from te Velthuis & Fodor (2016).

around the PB1 fingers and palm connects the two PA domains. PA mediates the cleavage of cellular 7-methylguanosine capped RNAs to produce 5' capped RNA primers for viral transcription through its N-terminal endonuclease domain. The PA endonuclease domain has a similar fold to that of type II restriction endonucleases (Dias et al., 2009; Yuan et al., 2009). Its activity depends on bivalent cations and both RNA and DNA can act as substrates. PA was also shown to be responsible for binding to the C-terminal domain (CTD) of the cellular RNA polymerase II (Pol II) and thus for allowing efficient viral transcription (Lukarska et al., 2017). The C-terminal domain of PA interacts with the N-terminal domain of PB1 (He et al., 2008; Obayashi et al., 2008; Pflug et al., 2014).

The N-terminal regions of PB2 form an interface with the PB1 thumb and are critical for viral RNA synthesis (Gastaminza et al., 2003; Hara et al., 2017; Pflug et al., 2014), while the remaining part of PB2 displays a flexible structure which is discussed in more detail in Chapter 1.5.2.2. The C-terminal region of PB2 contains the cap-binding domain which is responsible for binding the 5' cap of cellular RNAs prior to endonucleolytic cleavage by the PA endonuclease (Blaas et al., 1982; Guilligay et al., 2008). This interaction is mediated by the residues H357 and F404 and occurs through an aromatic sandwich mechanism (Guilligay et al., 2008). The 627-domain of PB2 has been heavily implicated with host adaptation of the avian viral RdRP to mammalian hosts (see Chapter 1.6). The nuclear localisation domain is located at the far C-terminus and consists of a classical bipartite NLS which interacts with importin- α 5 during nuclear import of PB2 (Tarendeau et al., 2007).

A large central cavity is formed between PB1 and the N-terminus of PB2 (Fig. 1.5C), which contains the active site (Hengrung et al., 2015; Pflug et al., 2014; Reich et al., 2014). The priming loop, a β -hairpin structure protruding from the PB1 thumb towards the active site, is also found in this central cavity. The priming loop supports the sugar-base of the

first NTP during *de novo* initiation and is essential for primer-independent replication of the vRNA template (Velthuis et al., 2016).

Several channels lead to and from the active site: the template entry channel, the template exit channel, the NTP entry channel and the product exit channel. The template entry channel is formed by all three RdRP subunits and leads from the site of 3' vRNA binding into the active site. The putative template exit channel is near the template entry channel, which would allow replication of the viral genome with minimal disruption of RNP complexes through dissociation and re-association of viral RNA from the same NP. The NTP entry channel is formed by highly conserved basic residues of PB1. The product exit channel is characterised by the path that the 3' end of the capped primer takes into the active site (Velthuis & Fodor, 2016).

1.5.2.2 Conformational rearrangements of the viral RdRP

The viral RdRP can be seen in different conformations in the various published crystal structures. Whereas the vRNA-bound influenza A and influenza B virus RdRP (Pflug et al., 2014; Reich et al., 2014) are seen in the same conformational arrangement, the RNA-free (apo) influenza C virus RdRP (Hengrung et al., 2015) and the cRNA-bound influenza B virus RdRP (Thierry et al., 2016) are seen in different conformations (Fig. 1.6). The core of the RdRP, consisting of PB1, the N-terminus of PB2 (Nter, N1, N2, lid domains) and the C-terminus of PA/P3 (linker, arch, Cter domains), remains immobile during these structural rearrangements, while the C-terminal domains of PB2 (cap-binding, mid-link, 627- and NLS domains) and the PA/P3 endonuclease domain undergo significant conformational rearrangements.

In the RNA-free apo structure of the influenza C virus RdRP, the PB2 cap-binding domain is immobilised through interactions with several RdRP domains and the cap-binding site is sterically hidden. The P3 endonuclease domain is located on the far side of the RdRP

and is interacting with the PB2 NLS domain. This 'closed' conformation is incompatible with cap-snatching and transcription and could represent the structure of the transcriptionally inactive newly assembled RdRP (Hengrung et al., 2015; Thierry et al., 2016).

The 5' cRNA-bound influenza B virus RdRP structure is similar to the 'closed' apo structure, however the PB2 627- and NLS domains and the PA endonuclease domain are slightly rearranged. This conformation might represent the RdRP during cRNA to vRNA replication (Thierry et al., 2016).

In the vRNA-bound influenza A and influenza B virus RdRP the PB2 cap-binding and the PA endonuclease domains are located close to and facing each other. The cap-binding domain is not immobilised by other domains and is flexible to rotate freely. The endonuclease domain is stabilised by the PB1 C-terminus and the PB2 N-terminus and is not blocked by the PB2 NLS domain anymore. This 'open' conformation is compatible with cap-snatching and could represent the structure in a transcription pre-initiation state (Pflug et al., 2014; Reich et al., 2014).

The evidence suggests that multiple conformations of the viral RdRP exist in equilibrium and that the RdRP therefore is inherently flexible (Hengrung et al., 2015; Thierry et al., 2016). Interaction of the RdRP with viral RNA influences this equilibrium and certain conformations are likely stabilised or induced by host or viral factors. It is very likely that there are several other relevant conformations of the RdRP that so far have not been described.

1.5.3 The viral nucleoprotein

The viral nucleoprotein NP accounts for most of the mass in viral RNP complexes and forms a structural scaffold around which the viral RNA is wound. The structural function of NP is essential for replication and transcription of the viral genome (Huang et al., 1990). However, NP is not required for the catalytic activity of the RdRP itself and rather acts as an elongation factor, which is shown by the fact that short vRNA-like segments can be replicated and transcribed in the absence of NP (Lee et al., 2002; Resa-Infante et al., 2010; Turrell et al., 2013).

Structural analysis of NP reveals three major structural domains: a head, a body and a tail-loop domain (Ng et al., 2008; Ye et al., 2006). A positively charged basic groove between the head and body domains is proposed to bind viral RNA (Ng et al., 2008; Turrell et al., 2013). Oligomerisation of NP, which is required for forming the structural scaffold of RNPs, relies on the tail-loop domain of NP docking into the basic groove next to the RNA binding region of another NP molecule. In the context of an RNP, NP also needs to be bound to the viral RNA-bound RdRP which is mediated through all three RdRP subunits (Biswas et al., 1998; Poole et al., 2004; Vidic et al., 2016).

1.5.4 Viral RNA promoters

Influenza virus vRNA segments have non-coding regions (NCRs) at the 5' and 3' termini flanking the coding region consisting of at least one open reading frame (Robertson, 1979; Skehel & Hay, 1978). Some parts of each NCR sequence are segment-specific, the terminal sequences however are conserved in all influenza vRNA segments. These conserved terminal sequences comprise 13 nucleotides at the 5' end and 12 nucleotides at the 3' end of the vRNA segments forming the viral RNA promoter region and have been shown to

be necessary for RdRP activity (Fodor et al., 1994; Tiley et al., 1994). The 5' and 3' ends of the viral promoter have partial and inverted complementarity (Desselberger et al., 1980). Initial models tried to explain this property by proposing a “panhandle” structure due to Watson-Crick base pairing between the termini (Hsu et al., 1987) or an “RNA-fork” model where the extreme termini remain single-stranded (Fodor et al., 1994; Kim et al., 1997). A revised model proposed a “corkscrew”-like structure where parts of the single-stranded 5' and 3' termini form secondary structures (Flick et al., 1996; Flick & Hobom, 1999). At each end two short stem loops are formed containing a stem of two base pairs and a four-nucleotide loop. This model was supported by the fact that both stem loops appear to be important for endonuclease activity of the RdRP (Leahy et al., 2001a; 2001b) and that the 5' stem loop is necessary for polyadenylation of viral mRNA (Pritlove et al., 1999). It is likely that the vRNA promoter forms a “panhandle” structure in an unbound state, whereas it adopts the “corkscrew” conformation when bound by the viral RdRP (Tomescu et al., 2014). Recently published crystal structure of the viral RdRP containing bound promoter vRNA confirm the existence of a 5' stem loop where nucleotides 1-10 form a hook-like structure that is tightly bound in a deep binding site on the surface close to the template entry channel (Pflug et al., 2014; Reich et al., 2014). The 3' promoter strand however does not appear as a stem loop in these published structures. Nucleotides 1 - 9 of the 3' end of the vRNA promoter bind instead in a binding site on the surface of the RdRP and remain in a dynamic single-stranded form. The distal regions of the 5' (nucleotides 11 - 16) and 3' vRNA promoter (nucleotides 10 - 15) form double-stranded regions as predicted by the “corkscrew” model.

The cRNA promoter is complementary to the vRNA promoter and is thought to adopt a structure analogous to that of the vRNA promoter. Supporting this is the fact that the 5' and 3' end stem loops are required for vRNA synthesis in cell culture (Azzeh et al., 2001; Crow et al., 2004). The recently published RdRP structures confirm that the 5' end of the

cRNA promoter is bound in the same location by the influenza B virus RdRP as the 5' vRNA end and that it also forms a hook-like structure (Thierry et al., 2016).

1.5.5 Viral mRNA synthesis, processing and transport

1.5.5.1 Transcription

A model of influenza virus transcription can be seen in Figure 1.7A. Influenza viruses do not encode an mRNA capping enzyme and therefore rely on the host transcription machinery to provide a 5' 7-methylguanosine (m7G) cap structure which is obtained from host RNA polymerase II complex derived products in a process called cap-snatching (Engelhardt & Fodor, 2006). Therefore, viral transcription is sensitive to inhibitors of cellular DNA-dependent RNA polymerase II such as actinomycin D and α -amanitin. Host small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and promoter-associated capped small RNAs (generated by RNA polymerase II pausing during transcription initiation) are preferred over mRNAs or pre-mRNAs in cap-snatching (Gu et al., 2015; Koppstein et al., 2015; Sikora et al., 2014). Cap-snatching can only occur when the viral RdRP is in its pre-initiation state, where the PB2 cap-binding and the PA endonuclease domains are aligned to allow binding and cleavage of capped RNA. The PB2 cap-binding domain is responsible for binding the 5' cap of host capped RNA (Cianci et al., 1995; Li et al., 1998). The PA endonuclease domain subsequently endonucleolytically cleaves the host RNA 10 - 15 nucleotide downstream of the 5' cap, usually after a purine residue, creating a short capped RNA primer (Beaton & Krug, 1981; Dias et al., 2009; Plotch et al., 1981; Yuan et al., 2009). The 3' end of the capped primer is then transferred from the endonuclease active site into the RdRP active site through the product exit

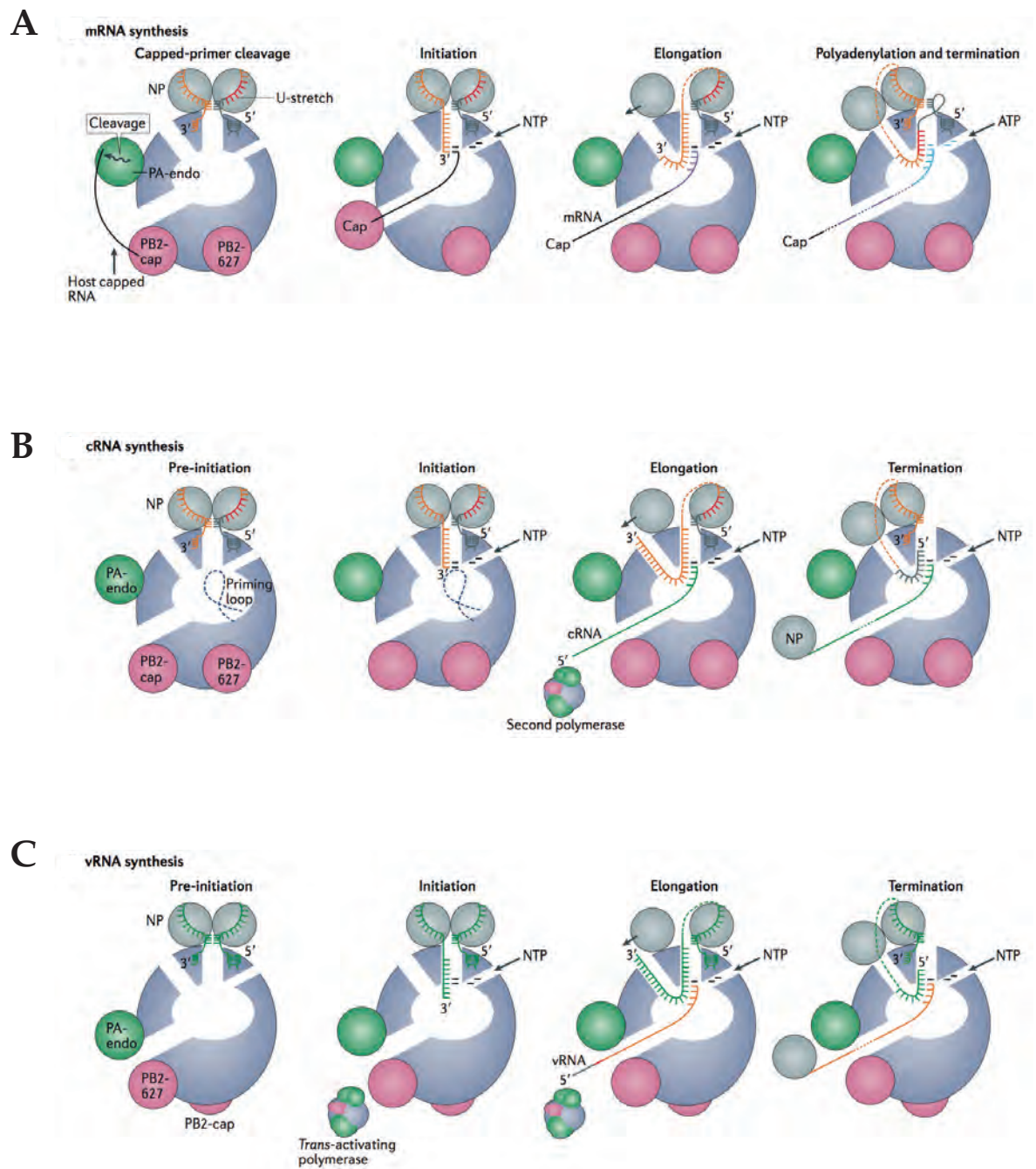


Fig. 1.7 Models of viral transcription and replication. Schematics of (A) mRNA transcription, (B) cRNA synthesis and (C) vRNA synthesis. Adapted from te Velthuis & Fodor (2016).

channel. This process is likely facilitated by a rotation of the PB2 cap-binding domain (Reich et al., 2014). At the same time the 3' end of the vRNA template needs to move from its binding site on the surface of the RdRP through the template entry channel into the RdRP active site. The 3' vRNA end is stabilised in the active site either by interaction with the capped primer or by base pairing of the initiating nucleotide. Transcription is initiated by addition of a G (at position C2 of the 3' vRNA template) or C (at position G3 of the 3' vRNA template) residue to the 3' of the capped primer (Beaton & Krug, 1981; Fodor et al., 1995). Transcription initiation is favoured by primers with a 3'-CA (Beaton & Krug, 1981; Shaw & Lamb, 1984) and the initiation process is independent of the priming loop in the RdRP active site (Velthuis et al., 2016).

Transcription elongation is then carried out in a vRNA template-dependent manner. The structural implications of transcription are not known yet, but the current model, which is compatible with the published crystal structures, proposes a local dissociation of the vRNA template from the NP double-helical structure at the template entry channel as the vRNA template is fed through the template entry channel, the RdRP active site and the template exit channel. A limited progressive local restructuring of the NP scaffold upon vRNA dissociation could allow reassociation of the vRNA template with NP at the exit channel (Velthuis & Fodor, 2016).

1.5.5.2 Polyadenylation

During host transcription, nascent mRNA transcripts produced by the RNA polymerase II have a signal sequence (AAUAAA) which facilitates cleavage and subsequent polyadenylation of the transcript through several cleavage and polyadenylation factors and the polyadenylate polymerase (Proudfoot, 2011; Proudfoot & Brownlee, 1976). Influenza virus mRNA transcripts do not contain this polyadenylation signal and therefore mediate polyadenylation in a host-independent manner through the viral RdRP.

Elongation is carried out in a vRNA template-dependent manner until a 5-7 nucleotide poly(U) stretch reaches the active site of the RdRP. This poly(U) stretch is located 16 nucleotides from the 5' terminus of the vRNA template (Robertson et al., 1981). Stuttering of the RdRP results in incorporation of multiple AMPs and in the formation of a poly(A) tail (Fodor et al., 1994). Stuttering is most likely caused by steric hindrance by the 5' end of the vRNA template which continues to be anchored by its hook structure in its binding pocket and possibly is stabilised by the return of the already transcribed 3' end to its binding site and base pairing with the 5' end (Fodor et al., 1994; Pritlove et al., 1999; Tiley et al., 1994).

1.5.5.3 Splicing

Splicing is a co-transcriptional event performed by the cellular spliceosome. Here a complex formed by small nuclear RNAs (snRNAs) and proteins remove intron sequences between a splice donor and a splice acceptor site in the mRNA transcript (Matera & Wang, 2014). Only three viral segments produce transcripts through alternative splicing. Segment 1 (PB2), segment 7 (M) and segment 8 (NS) produce unspliced and spliced mRNA transcripts.

Segment 1 produces the unspliced PB2 mRNA transcript which encodes the PB2 RdRP subunit, but which can also be spliced into PB2-S1 mRNA encoding the auxiliary PB2-S1 protein (Yamayoshi et al., 2016).

Segment 7 produces the unspliced M1 mRNA transcript which encodes the matrix protein (M1). This transcript, however, can be alternatively spliced to produce different alternative mRNA transcripts. The M2 mRNA transcript encodes the M2 ion channel (Lamb & Choppin, 1981; Lamb et al., 1981). The M3 mRNA produces a potential 9 amino acid peptide with an unknown biological function (Lamb et al., 1981). The M4 mRNA transcript encodes a putative polypeptide similar to M1 with an internal deletion (Shih et

al., 1998). The M42 mRNA transcript encodes the M42 ion channel, an alternative to the M2 ion channel with an antigenically distinct ectodomain (Wise et al., 2012b).

Segment 8 produces the unspliced NS1 mRNA encoding the multifunctional NS1 protein (Winter et al., 1981b), which can be spliced into the NS2 mRNA encoding the nuclear export protein NEP (Lamb & Lai, 1980) as well as the NS3 mRNA encoding NS3 with an unknown function (Selman et al., 2012).

Splicing of viral mRNA is carried out by the cellular spliceosome using the canonical splice donor and acceptor sequences, but its mechanisms has not been fully explained yet (Lamb & Horvath, 1991). In contrast to cellular mRNA, which is efficiently spliced, only approximately 10% of viral transcripts are spliced (Lamb et al., 1980; 1981). Splicing of viral transcripts has been shown to be regulated by both host and viral factors. The viral NS1 protein has been shown to be involved in regulating splicing of the M1 mRNA (Robb & Fodor, 2012). Its role in regulating NS1 mRNA splicing is controversial and has been both confirmed (Garaigorta & Ortín, 2007; Smith & Inglis, 1985) and refuted (Robb et al., 2010). Numerous host factors have been described to be involved in regulating viral mRNA splicing, most notably NS1-binding protein (NS1-BP) and heterogeneous nuclear ribonucleoprotein K (hnRNPK) (Tsai et al., 2013) as well as the viral RdRP and the cellular splicing factor SF2/ASF (Shih & Krug, 1996; Shih et al., 1995), which all have been implicated in regulating M1 mRNA splicing. However, recent large-scale genome wide screens have identified numerous other host factors thought to be involved in regulating viral transcript splicing (Dubois et al., 2014; Fournier et al., 2014).

1.5.5.4 Nuclear export of viral mRNA

When the RdRP transitions from transcription initiation to elongation the 5' cap is released from the PB2 cap-binding domain and bound by the host nuclear cap-binding complex. This results in a viral mRNA associated ribonucleoprotein complex (mRNP).

Assembly and nuclear export of these viral mRNPs however is heavily reliant on numerous cellular factors and ongoing host RNA polymerase II transcription (Amorim et al., 2007; Read & Digard, 2010; Tokunaga et al., 2006; York & Fodor, 2013). It has been shown that cellular RNA-binding proteins and the nuclear export machinery of mRNA are hijacked in influenza A virus infection (Bier et al., 2011). Cellular mRNAs are being exported from the nucleus through the transcription export (TREX) complex. REF/Aly, which is a component of the TREX, recruits nuclear RNA export factor 1 (NXF1) and its co-factor p15 to mediate nuclear export of mRNPs to the cytoplasm via the nuclear pore complex (Katahira, 2012). Viral mRNPs are also thought to be exported through the NXF1/p15 pathway (Hao et al., 2008; Read & Digard, 2010; Wang et al., 2008). Alternative mRNA export pathways exist in the cell and it is likely that the virus has developed a redundancy of mechanisms for nuclear export of mRNPs.

The viral NS1 protein has also been heavily implicated in regulating viral mRNP export. It has been shown that NS1 inhibits nuclear export of cellular mRNPs through binding to the cleavage and polyadenylation specificity factor 30 (CPSF30) and poly(A)-binding protein II (PABPII) and preventing 3' processing of pre-mRNA (Chen & Krug, 2000). It also forms an inhibitory complex with NXF1/p15, thus blocking export of mRNPs (Satterly et al., 2007) and co-purifies with viral mRNA and NXF1 (Wang et al., 2008).

1.5.6 Viral RNA replication

1.5.6.1 cRNA synthesis

Replication of the viral RNA genome is a two-stage process, which requires synthesis of a positive-sense complementary RNA (cRNA) intermediary to synthesise copies of the

negative-sense single-stranded RNA genome. A model of cRNA synthesis is presented in Figure 1.7B. In contrast to viral mRNA, both cRNA and vRNA do not possess a 5' cap, but instead contain a 5' triphosphate. Thus, replication is initiated *de novo* in a primer-independent manner. Biochemical *in vitro* data suggests that the resident RdRP of imported vRNPs is capable of cRNA synthesis (Vreede & Brownlee, 2007; York et al., 2013). During replication initiation of cRNA, the 3' terminus of vRNA needs to be translocated into the active site of the RdRP. The first nucleotide to be incorporated by the RdRP is GTP through base-pairing to the penultimate C2 residue in the 3' terminus of the vRNA. This allows formation of a phosphodiester bond with an ATP nucleotide which base-pairs with the terminal U1 residue of the 3' vRNA, resulting in a pppApG dinucleotide. This process is called terminal initiation (Deng et al., 2006b; Vreede et al., 2008). This *de novo* formation of a dinucleotide requires support from a priming loop, as is seen in other *de novo*-initiating viral RdRPs (Velthuis et al., 2016). After initiation, the vRNA template is thread through the active site and the nascent cRNA strand is elongated similarly to the elongation process in mRNA transcription with little disruption of the vRNP structure. However, unlike in transcription, the 5' end of vRNA needs to be released from its binding pocket to be copied in full and it remains unclear how this is regulated. The nascent newly synthesised cRNA associates with a newly synthesised viral RdRP and the first NP monomer through an interaction between the RdRP and NP as soon as 5' end emerges from the product exit channel. Subsequent NP monomers are recruited through the groove and tail-loop interactions between cRNA-bound NP and free NP (Chan et al., 2010; Ng et al., 2008; Turrell et al., 2013; Ye et al., 2006).

1.5.6.2 vRNA synthesis

In the second stage of viral replication, cRNA acts as a template for vRNA synthesis (Figure 1.7C). Contrary to terminal replication initiation on a vRNA template, initiation occurs internally on a cRNA template on residues U4 and C5 of the 3' cRNA, resulting in a pppApG dinucleotide. The pppApG dinucleotide is then realigned to residues U1 and C2 at the 3' cRNA terminus, allowing subsequent elongation of the full-length nascent vRNA strand (Deng et al., 2006b). The difference in terminal and internal replication initiation has been proposed to stem from slight differences in the viral promoter structure and its affinity for binding to the RdRP (Azzeh et al., 2001; Crow et al., 2004; Gonzalez & Ortín, 1999). The priming loop, which is critical for terminal initiation, is not essential for internal initiation (te Velhuis et al., 2016). While cRNA can be synthesised entirely by the resident RdRP of the replicating vRNP, vRNA synthesis requires a second RdRP. Purified cRNPs have been shown to be inactive *in vitro* in the absence of additional purified RdRP, which interestingly does not need to be catalytically active (York et al., 2013). This suggests that the second RdRP does not fulfil a catalytic function and therefore has a *trans*-activating, rather than *trans*-acting role.

In the *trans*-activating RdRP model, the additional RdRP induces a conformational rearrangement of the resident RdRP in the cRNP that allows translocation of the 3' cRNA terminus into the active site and internal initiation. The *trans*-activating RdRP could also bind to the 5' end of the emerging nascent vRNA strand, recruit the first NP and ultimately assemble into a vRNP.

In the *trans*-acting RdRP model (Jorba et al., 2009), a catalytically active RdRP, possibly associated with small viral RNA is recruited to the replicating cRNP and receives the 3' terminus of the cRNA template. The *trans*-acting RdRP then carries out the vRNA synthesis as described above. A third RdRP is then recruited to the *trans*-acting RdRP and binds to the 5' end of the emerging nascent vRNA and assembles the vRNP. This model is

supported by genetic complementation experiments and observation of branched RNPs in cryo-EM (Jorba et al., 2009; Moeller et al., 2012). Small viral RNAs (svRNAs) are 22-27-nucleotide-long RNAs that correspond to the 5' termini of vRNA that are highly expressed in infected cells and that have been implicated in viral replication (Perez et al., 2010; Umbach et al., 2010).

1.6 Host - virus interactions

1.6.1 Transcription and translation of eukaryotic mRNA

In eukaryotes, DNA is wound around histones to form densely packaged chromatin structures. Host pre-mRNA is synthesised by the DNA-dependent RNA polymerase II (Pol II) holoenzyme. Pol II does not track along the DNA template, but instead is fixed in transcription factories in the nucleus to which chromatin loops are tethered, containing high concentrations of Pol II and transcription factors (Cook, 2010). Host pre-mRNAs are subsequently synthesised by Pol II utilising numerous RNA processing host factors. Phosphorylation of a series of tandem repeats (serine-2 or serine-5) on the carboxyl terminal domain (CTD) of the RBP1 subunit of Pol II regulates capping of nascent pre-mRNAs. A 5' 7-methylguanosine cap is added to pre-mRNAs by the capping enzyme complex (CEC) containing a phosphatase, a guanylyl transferase and a methyltransferase. The cap is subsequently bound by the cap binding complex (CBC), consisting of CBC20 and CBC80 (Shandilya & Roberts, 2012; Topisirovic et al., 2011).

Splicing is performed co-transcriptionally by the spliceosome (Reed, 2003). The spliceosome is an RNP complex formed by snRNA and proteins that are responsible for

the removal of introns between splice donor and splice acceptor sites in pre-mRNA transcripts. The exon junction complex (EJC), composed of the four major components eIF4A-III, Magoh, Y14 and MLN51, is located onto the junction of two newly joined exons of pre-mRNA and together they form a stable mRNP complex which is exported from the nucleus. EJCs can regulate the localisation, stability and translation of spliced mRNA transcripts.

Two *cis* RNA sequences in host pre-mRNA transcripts are required for polyadenylation, cleavage and maturation of pre-mRNAs into mRNA transcripts. The poly(A) signal is a 6-nucleotide-long motif and is found 10-30 nucleotides upstream of a cleavage site, which is bound by cleavage and polyadenylation specificity factor (CPSF). Downstream sequence elements (DSEs) or upstream elements (USEs) enhance cleavage efficiency by binding the cleavage stimulation factor (CSTF). Together with cleavage factors CFI and CFII these two protein complexes mediate cleavage of the mRNA transcript. Polyadenylation is subsequently catalysed by the poly(A) polymerase (PAP) at a slow rate, however, association with poly(A) binding protein II (PABPII) results in a rapid increase in the rate of polyadenylation (Proudfoot, 2011).

The fate of Pol II transcripts is initially determined by the hnRNP C1/C2 complex, separating long (> 200 nucleotides) transcripts, which are exported via the mRNP export pathway, and short snRNA transcripts, which are exported by CRM1 (McCloskey et al., 2012). During Pol II transcription mRNA transcripts acquire numerous RNA-binding proteins, such as the CBC, heterogeneous nuclear ribonucleoproteins (hnRNPs), poly(A)-binding proteins, the EJC, serine/arginine-rich (SR) proteins and the transcription export (TREX) complex. The main nuclear export factor for host mRNA is the nuclear export factor 1 (NXF1), which forms a heterodimer with p15. NXF1-p15 is recruited to cellular mRNAs through SR proteins associated with mRNP complexes and directs the export of

mRNP complexes through the nuclear pore complex (Carmody & Wentz, 2009; Kelly & Corbett, 2009; Koehler & Hurt, 2007; Proudfoot, 2011).

Translation of mRNA into proteins takes place in ribosomes in the cytoplasm (Hershey et al., 2012; Roux & Topisirovic, 2012; Sonenberg & Hinnebusch, 2009). Eukaryotic ribosomes consist of a small subunit (40S) composed of 18S rRNA and approximately 30 proteins and a large subunit (60S) composed of 28S, 5.8S and 5S rRNA and approximately 45 proteins. Translation initiation requires several eukaryotic initiation factors. Factors eIF-1, eIF-1A and eIF-3 bind to the 40S ribosomal subunit, eIF-2 interacts with the initiating methionyl tRNA, eIF-4E recognises the 5' 7-methylguanosine cap of mRNA and eIF-4G interacts both with eIF-4E and PABPII associated with the 3' poly(A) tail of mRNA. Factors eIF-4E and eIF-4G then guide the mRNA with the help of eIF-4A and eIF-4B to the 40S ribosomal subunit where eIF-4G interacts with eIF-3. The 40S ribosomal subunit then scans the mRNA transcript for the AUG start codon. Following recognition of the start codon, eIF-5 triggers the release of initiation factors and the association of the 40S subunit with the 60S subunit to form the fully assembled ribosome.

Following translation initiation, the polypeptide chain is elongated. During elongation aminoacyl tRNAs are paired with the corresponding mRNA codon with the help of elongation factor eEF-1 α . Successful pairing releases eEF-1 α and peptide bonds are formed between the current and previous aminoacyl tRNA catalysed by the 60S ribosomal subunit. Elongation factor eEF-2 then stimulates the ribosome to move one codon further along the mRNA transcript and the elongation process repeats itself.

Elongation of the polypeptide chain continues until a stop codon is reached. No tRNA complementary to stop codons exist, instead the release factor eRF-1 recognises the termination sequence and terminates protein synthesis by interacting with eRF-3 to hydrolyse the bond between polypeptide and the final tRNA, resulting in release of the

polypeptide chain and dissociation of the ribosome (Hershey et al., 2012; Roux & Topisirovic, 2012; Sonenberg & Hinnebusch, 2009).

1.6.2 The RNA-binding proteome

RNA-binding proteins (RBPs) play a key role in regulating all aspects of RNA metabolism (Glisovic et al., 2008). Both the catalogue of cellular RBPs and the activity of these RBPs respond to different biological and environmental stimuli. In fact, numerous diseases have been linked to defects in RBP expression and function, particularly neurological disorders, muscular atrophies, cardiovascular diseases and cancer (Cooper et al., 2009; Darnell, 2010; Liao et al., 2016; Lukong et al., 2008). Because of the importance of RNA metabolism in cell physiology, RBPs are typically highly conserved, ubiquitously expressed and abundant (Beckmann et al., 2015; Gerstberger et al., 2014a; 2014b).

RBPs assemble with RNA into ribonucleoprotein (RNP) complexes and collectively all the RBPs are known as the RNA-binding proteome (RBPome). Until recently, our understanding of how RBPs interact with RNA was based on a limited number of known modular RNA-binding domains (RBDs), including the RNA recognition motif (RRM), the K-homology domain (KH), the double-stranded RBD (dsRBD), zinc fingers (Znf) and the DEAD box helicase domain (Lunde et al., 2007). However, computational predictions suggested that 3 - 11 % of a given proteome is involved in RNA-binding, a significantly larger proportion of the proteome than expected based on the presence of classical RBDs (Anantharaman et al., 2002; Gerstberger et al., 2014b). Numerous noncanonical RBDs have been reported since (Lee & Hong, 2004; Niessing et al., 2004; Rammelt et al., 2011; Zalfa et al., 2005). These novel RBPs exhibit protein domains with dual functions (Walden et al., 2006), previously unknown protein domains (Jia et al., 2008) or disordered domains

lacking any stable tertiary structure (Lobley et al., 2007; Oberstrass et al., 2005; Phan et al., 2011; Xie et al., 2007).

Recently, systematic large-scale proteome-wide approaches have significantly expanded the known RBPome by both classical and noncanonical RBPs. Dozens of unorthodox novel RBPs were identified in microarray studies in yeast (Scherrer et al., 2010; Tsvetanova et al., 2010). More significantly, recently several studies based on *in vivo* crosslinking, poly(A)-RNA capture and MS have identified more than a thousand proteins interacting with poly(A)-RNA, including hundreds of novel RBPs (Baltz et al., 2012; Castello et al., 2012). Interestingly, disordered regions were significantly enhanced in both known and novel RBPs, suggesting a significant role of disordered protein regions in RNA-binding (Castello et al., 2012; Kwon et al., 2013).

Using this RBPome capture technique, the RBPomes of HeLa (Castello et al., 2012), HEK-293 (Baltz et al., 2012), human macrophages (Liepelt et al., 2016), mouse embryonic stem cells (Kwon et al., 2013), *Saccharomyces cerevisiae* (Beckmann et al., 2015; Matia-Gonzalez et al., 2015; Mitchell et al., 2013), *Caenorhabditis elegans* (Matia-Gonzalez et al., 2015), *Drosophila melanogaster* embryos (Sysoev et al., 2016; Wessels et al., 2016), *Leishmania donovani* (Nandan et al., 2017) and *Arabidopsis thaliana* (Marondedze et al., 2016; Reichel et al., 2016; Zhang et al., 2016) have been described. Comparison of the RBPomes of these different cell lines and species revealed a conserved core of approximately 250 RBPs that are conserved in eukaryotes and most of which have known functions in RNA metabolism (Beckmann et al., 2015).

1.6.3 Influenza virus mRNA processing

As described in Chapter 1.5.5, the influenza A virus relies on the host machinery for viral transcription and nuclear export of mRNPs. Early in the transcription cycle the viral RdRP interacts with the CTD of Pol II that is phosphorylated at serine-5 (Martínez-Alonso et al., 2016). This close association with Pol II and the host capping machinery allows the PB2 subunit of the viral RdRP to bind to the 5' cap of nascent host RNAs and the PA subunit of the viral RdRP cleaves the cap 10-15 nucleotides downstream of the cap, thus initiating viral transcription. Splicing of viral transcripts is also performed by the host splicing machinery, but it has been shown that the viral NS1 protein can regulate the viral splicing process. In contrast to cellular polyadenylation, polyadenylation of viral transcripts is performed by the viral RdRP. For nuclear export of viral mRNPs, influenza viruses rely and interact with the NXF1/p15 pathway, as well as the CRM1 pathway. Viral mRNAs are translated using the host cellular machinery using the mechanism described in Chapter 1.6.1.

1.6.4 NS1-mediated host shut off

Viral NS1 is a multifunctional protein involved in inhibiting the type I IFN response through many different processes (Krug, 2015). Its structure is formed by an RNA-binding N-terminal domain and an effector C-terminal domain (Hale et al., 2008). One way that NS1 blocks the type I IFN pathway is by inhibiting processing of 3' ends of Pol II transcribed pre-mRNAs. NS1 was shown to interact with CPSF30, a component of the CPSF complex responsible for cleavage of pre-mRNAs and recruitment of PAP for polyadenylation (Nemeroff et al., 1998). Because polyadenylation is coupled to nuclear export of mRNA, NS1 interaction with CPSF30 blocks mRNA export to translating

ribosomes. In addition to this mechanism, the C-terminal effector domain of NS1 also binds to PABPII and prevents it from stimulating polyadenylation by PAP. This results in mRNAs with short poly(A) tails that are exported very inefficiently from the nucleus (Chen et al., 1999).

NS1 also interferes directly with mRNA export by binding to NXF1, p15, Rae1 and E1B-AP5. These proteins interact with mRNA or nucleoporins and mediate nuclear export of mRNAs and binding to NS1 blocks their cellular functions (Kuss et al., 2013). Additionally, NS1 interacts with and promotes degradation of the nucleoporin Nup98, leading to export defects (Satterly et al., 2007).

Because viral mRNA synthesis does not rely on host mediated cleavage and polyadenylation and at least some viral mRNAs are exported in an NXF1-independent manner, viral mRNA is resistant to NS1-mediated host shut off.

1.6.5 Pol II degradation by the viral RdRP

Host shut off is also mediated by the viral RdRP (Vreede & Fodor, 2010). Active Pol II transcription and physical contact between the Pol II CTD and the viral RdRP are required for viral mRNA synthesis (Chan et al., 2006; Engelhardt et al., 2005; Spooner & Barry, 1977). Phosphorylation of serine-2 and serine-5 dictates the stage of transcription of Pol II and the viral RdRP interacts with the initiating serine-5 phosphorylated form of Pol II which is compatible with capping of cellular mRNAs (Engelhardt et al., 2005; Martínez-Alonso et al., 2016). Except for cap-snatching, this interaction also leads to ubiquitination and subsequent degradation of the large subunit of Pol II by the proteasome (Rodriguez et al., 2007; Vreede et al.,

2010). Degradation of Pol II results in inhibition of the host antiviral response through blocking of IFN-stimulated genes (Vreede et al., 2010). Degradation of Pol II also might play a role as a switch from viral mRNA synthesis to viral genome replication, which does not require Pol II association (Vreede & Fodor, 2010).

1.6.6 Host RNA degradation by PA-X

PA-X is produced by a ribosomal frameshift in translating the PA mRNA transcript (Firth et al., 2012; Jagger et al., 2012) and is responsible for degradation of host RNA. PA-X has the same N-terminal endonuclease as PA, but terminates in a short 41 or 61 amino acid-long C-terminal domain (Jagger et al., 2012; Khapersky et al., 2016; Oishi et al., 2015). PA-X can cleave both ssRNA and dsRNA, whereas PA preferentially cleaves ssRNA, fitting their roles in host shut off and cleaving pre-mRNAs for cap-snatching respectively (Bavagnoli et al., 2015). Like PA, PA-X preferentially degrades Pol II transcripts, as opposed to Pol I or Pol III transcripts, and does not affect viral transcripts. PA-X degradation has been linked to cellular 3' end processing in the nucleus. PA-X cleaves RNAs internally and cleaved RNA fragments are degraded by cellular exonucleases, e.g. Xrn1 (Khapersky et al., 2016). Viruses lacking PA-X exhibit lower viral protein synthesis, which could suggest that host RNA degradation allows preferential access of viral transcripts to the cellular translational machinery (la Luna et al., 1995).

Expression of PA-X mediates nuclear accumulation of cytoplasmic poly(A) binding protein (PABPC1), leading to nuclear retention and hyperadenylation of host mRNAs (Khapersky et al., 2014). Therefore, nuclear PABPC1 could also reduce the amount of translatable host mRNAs and induce host shut off.

Because of its ability to degrade host mRNAs, PA-X is associated with reduced gene expression and inhibition of the host immune response.

1.7 Influenza A virus host adaptation

1.7.1 Overview

The natural hosts of influenza viruses are aquatic birds, mainly wild waterfowl, which act as a natural reservoir for the virus. Avian influenza viruses can infect some mammalian species, albeit not very efficiently, but usually this leads to sporadic dead-end infections. In rare cases, sustained human epidemics can occur after one of these zoonotic infections. For this to happen, the avian virus needs to adapt to the mammalian host species to be able to interact with the host machinery which the virus relies on for efficient replication. Recent studies have shown that avian influenza viruses could require as little as five adaptive mutations to enable efficient transmission between mammalian hosts, four of them in HA and one of them in PB2 (Herfst et al., 2012; Imai et al., 2012). Studies revealed that these mutations could change the avian influenza virus receptor binding specificity from α -2,3-linked sialic acid to α -2,6-linked sialic acid, which is preferred by mammalian-adapted influenza viruses (Chutinimitkul et al., 2010; Herfst et al., 2012), and enable efficient replication of an avian RdRP in mammalian host cells at lower temperatures typical for mammalian hosts (Subbarao et al., 1993).

While these mutations appear to be the minimal amount that are required for efficient airborne transmission of avian influenza A viruses between mammalian hosts, there are numerous other host adaptive mutations that can be found in many of the viral proteins and that can work in tandem with or compensate for the lack of other mutations. Apart

from the HA and RdRP viral proteins, adaptive mutations have also been found in NS1, NEP, NP and NA (Cauldwell et al., 2014; Mänz et al., 2012; Schrauwen & Fouchier, 2014).

1.7.2 Host adaptation and the PB2 polymerase subunit

1.7.2.1 The canonical host determinant PB2-627

The PB2 subunit of the RdRP is the main determinant of influenza A virus RdRP host range and its role in overcoming host range restriction of influenza viruses has been known for a long time (Almond, 1977). The most well studied and most critical host adaptation is a glutamic acid to lysine substitution in residue 627 of the PB2 subunit (Subbarao et al., 1993). This single mutation has been shown to effectively overcome the host restriction of an avian RdRP in a mammalian host, increasing viral replication, pathogenicity and transmissibility (Chen et al., 2007b; Hatta et al., 2001; Shinya et al., 2004; Steel et al., 2009). PB2-627K is present in most circulating human influenza A virus strains and was found in the 1918, 1957 and 1968 pandemic strains (Scholtissek et al., 1978; Taubenberger et al., 2005), as well as in 32% of all human H5N1 isolates since 1997 and most of the human H7N9 isolates from the recent outbreak in China (Liu et al., 2013; Long et al., 2013).

PB2-627 is located on the surface of the 627-domain, which derives its name from PB2-627 (Figure 1.8). The surface contains a basic patch which is disrupted by the avian signature glutamic acid and which is fully restored by the substitution for lysine (Kirui et al., 2014; Tarendeau et al., 2008). PB2-627 sits on top of a loop adjacent to the basic patch (Kuzuhara et al., 2009).

Although PB2-E627K or a similar host adaptive mutation is necessary for efficient adaptation by avian influenza A viruses to mammalian hosts, the exact molecular mechanism of how this is achieved remains unknown. One of the first factors to be proposed to drive adaptation is temperature. Avian influenza viruses replicate in the gastrointestinal tracts of birds where the mean temperature is 41°C, whereas temperatures in the upper respiratory tract of mammals is only about 33°C. Studies have shown that avian influenza virus replication is attenuated in mammalian cells at 33°C and that the E627K substitution restored some of the viral replication (Ayora-Talavera et al., 2009; Hatta et al., 2007; Massin et al., 2001; Scull et al., 2009). However, when avian and mammalian influenza virus RdRPs were studied outside of a cell-based context *in vitro*, it was found that avian-adapted RdRPs were not restricted at lower temperatures and more active at higher temperature than mammalian-adapted RdRPs (Aggarwal et al., 2011). It was also shown that an avian RdRP was restricted in human cells even at high temperatures, suggesting that temperature cannot be the sole factor dictating host restriction (Bussey et al., 2010; Massin et al., 2001). These findings suggested that the avian RdRP is not inherently attenuated at lower temperatures, but that it possibly fails to effectively assemble vRNPs. Indeed, the interaction between the viral RdRP and NP has previously been implicated to be a factor determining host range restriction. PB2-627K has been shown to stabilise the interaction between PB2 and NP (Labadie et al., 2007; Mehle & Doudna, 2008; Ng et al., 2012a; Rameix-Welti et al., 2009; Ng et al., 2012b). It has been proposed that there therefore is a requirement for co-selection of PB2 and NP during viral evolution as the effect of PB2-E627K depended on the origin of the NP segment (Bogs et al., 2011; Ng et al., 2012b). However, recent studies have conclusively shown that host adaptive mutations including PB2-E627K do not enhance PB2-NP interaction (Cauldwell et al., 2013) and that host restriction occurs independently of the presence of NP (Paterson et al., 2014).

Conflicting evidence suggests that the nature of PB2-627 affects viral RNA promoter binding. The presence of lysine in PB2-627 was on the one hand shown to enhance binding of PB2 to viral RNA promoters (Kuzuhara et al., 2009), but on the other hand reduce vRNA and cRNA promoter binding as well as replication initiation *in vitro* (Nakazono et al., 2012). These findings are difficult to interpret as they are observed *in vitro* in the absence of all cellular factors which contribute to viral RdRP function and because the 627-domain is located far away from the active site and the promoter binding region of the viral RdRP (Pflug et al., 2014).

The viral RNA promoters do nevertheless have a significant role in determining host restriction. Avian RdRP restriction in mammalian host cells can be overcome if mutations (especially G3A, U5C and C8U) are introduced in the 3' vRNA promoter (Crescenzo-Chaigne et al., 2002; Neumann & Hobom, 1995), resulting in perfect base pairing of the 5' and 3' terminal ends of the promoter sequence. This means, that the 5' terminus of the resulting cRNA promoter is identical to the 5' terminus of the vRNA promoter. RdRP enhancement is only observed if PB2-627E, but not PB2-627K is present (Crescenzo-Chaigne et al., 2002). This effect was shown to be due to the 3' promoter sequence and not the complementarity of the promoter ends (Paterson et al., 2014).

Since, evidence has pointed to interaction with host factors as the most significant factor determining host restriction. When an avian influenza virus infects a mammalian host cell, it encounters a whole new cellular environment which might affect any step of viral replication. Two studies using RdRP activity assays in heterokaryons of human and avian cells identified that one or more host factors are crucial for host restriction. While one study concluded that mammalian cells harbour a dominant inhibitory factor that restricts avian RdRP activity (Mehle & Doudna, 2008), the other study concluded that mammalian cells lack an activating factor required for avian RdRP activity (Moncorge et al., 2010). As it is known that the viral RdRP interacts with numerous host factors, it is not surprising

that there would be negative and positive regulators of avian and mammalian-adapted RdRPs (Watanabe et al., 2014; York et al., 2014).

In a study trying to identify influenza RdRP-binding host proteins whose binding profiles depend on the nature of PB2-627, the DEAD box RNA helicase 17 protein (DDX17) was shown to play a small but significant role in regulating avian and mammalian RdRP activity (Bortz et al., 2011). Silencing of DDX17 was shown to restrict mammalian, but enhance avian RdRP activity in mammalian host cells, while silencing the chicken homologue inhibited both avian and mammalian-adapted RdRPs in avian cells. While DDX17 is unlikely to be solely responsible for avian virus host range restriction, more work is required to elucidate the exact role it plays in regulating avian and mammalian RdRP activity.

1.7.2.2 Other PB2 host adaptive mutations

Apart from PB2-E627K several other host adaptive mutations also cluster around the 627-domain of PB2. The 2009 pandemic pH1N1 virus retained the avian signature glutamic acid of PB2-627. In fact, it was shown that introducing a lysine at this position did not enhance virulence or replication of the virus, suggesting that it is possible for other mutations to fully compensate for the lack of PB2-E627K (Herfst et al., 2010; Jagger et al., 2010; Zhu et al., 2010). Instead of E627K, pH1N1 contained mutations G590S and Q591R, located adjacent to PB2-627 on the same surface region. They are able to mask the negative charge of the glutamic acid and establish a positively charged basic region on the surface of the PB2 C-terminus (Mehle & Doudna, 2009; Tarendeau et al., 2008; Yamada et al., 2010).

Bioinformatics analysis of human avian influenza virus isolates identified PB2-598, which is located close to the 590/591/627 cluster, as a potential host range determinant.

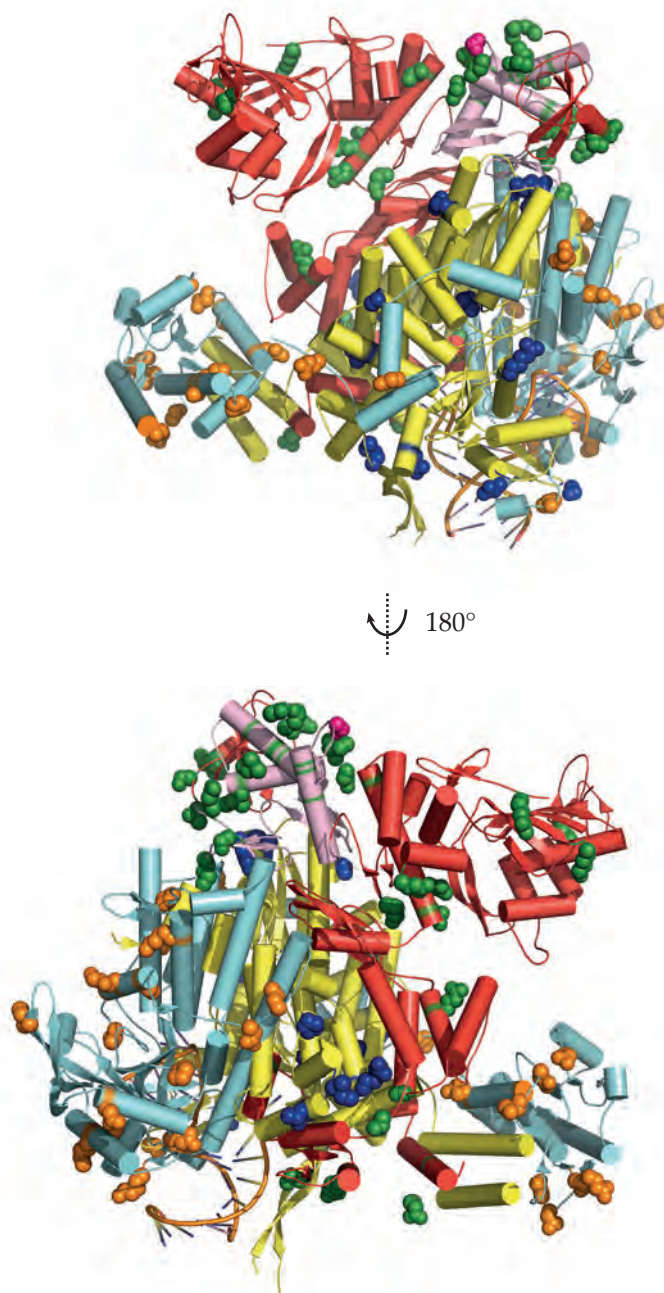


Fig. 1.8 Host adaptive mutations in the viral RdRP structure. Crystal structure of the vRNA-bound influenza A/Little yellow-shouldered bat/Guatemala/060/2010 (H17N10) virus RdRP (PDB 4WSB). PA, PB1 and PB2 are shown in cyan, yellow and red, respectively. The PB2 627-domain is highlighted in pink. Host adaptive residues in PA (orange), PB1 (blue) and PB2 (red) are shown as spheres. PB2-627 is highlighted in magenta on top of the PB2 627-domain.

Table 1.3 List of all described host adaptive mutations in the PB2 subunit

Subunit	Residue	Adaptive mutation	Reference
PB2	9	D9N	(Naffakh et al., 2000)
	44	A44S	(Chen et al., 2006)
	64	M64T	(Miotto et al., 2008)
	81	T81M	(Naffakh et al., 2000)
	105	T105V	(Miotto et al., 2008)
	158	E158G	(Zhou et al., 2011)
	199	A199S	(Taubenberger et al., 2005)
	253	D253N	(Mok et al., 2011)
	256	D256G	(Manzoor et al., 2009)
	271	T271A	(Bussey et al., 2011)
	292	I292T	(Miotto et al., 2008)
	368	R368K	(Miotto et al., 2008)
	475	L475M	(Naffakh et al., 2000)
	493	R493K	(Tamuri et al., 2009)
	526	K526R	(Song et al., 2014)
	559	T559I	(Hu et al., 2014)
	567	D567N	(Miotto et al., 2008)
	569	T569A	(Tamuri et al., 2009)
	588	A588I	(Naffakh et al., 2000)
	590	G590S	(Mehle & Doudna, 2009)
	591	Q591R	(Mehle & Doudna, 2009)
	598	V598T	(Hu et al., 2017)
	613	V613T	(Chen et al., 2006)
	627	E627K	(Subbarao et al., 1993)
	661	A661T	(Miotto et al., 2008)
	674	A674T	(Chen et al., 2006)
	676	T676I	(Neumann et al., 2014)
	682	G682S	(Tamuri et al., 2009)
	684	A684S	(Hayashi et al., 2015)
	701	D701N	(Li et al., 2005)
	702	K702R	(Naffakh et al., 2000)
	714	S714I	(Gabriel et al., 2005)

Functional characterisation of PB2-V598T/I showed that PB2-598 plays a similar role to PB2-590/591/627 in determining viral host range (Hu et al., 2017).

A threonine-to-alanine substitution in PB2-271 is found in pH1N1 and most mammalian-adapted influenza A virus strains. This mutation has been found to enhance RdRP activity in human cells (Bussey et al., 2010; Cauldwell et al., 2013; Foeglein et al., 2011), increase viral replication in mice (Bussey et al., 2010) and re-introducing the avian signature alanine into human influenza A virus strains resulted in loss of transmissibility in guinea pigs and ferrets (Zhang et al., 2012). The mechanism of its host adaptive effect is still unknown.

PB2-D701N is another mutation strongly linked to host adaptation. It is not present in currently circulating human influenza A virus strains, but was found in some human H5N1 isolates and the dominant Eurasian swine H1N1 strains (Qi et al., 2012). It has been shown to increase virus replication in mice (Gabriel et al., 2005; 2008), enhance avian H5N1 replication in mice and humans (de Jong et al., 2006; Li et al., 2005), compensate for the lack of K627E and enable transmissibility in guinea pigs (Gao et al., 2009; Steel et al., 2009). Introduction of PB2-D701N into pH1N1 led to increased pathogenicity and viral replication in mice and transmissibility in ferrets (Zhou et al., 2013), suggesting that E627K, G590S and Q591R on the one hand, and D701N on the other hand, have different mechanisms of how they assert their host adaptive effects. PB2-701 has been implicated to be involved in nuclear import of PB2 through the interaction of the PB2-NLS domain with importin- α 5 (Tarendeau et al., 2007). Nuclear import of viral protein complexes appears to play a significant role in host range restriction. The avian RdRP needs to adapt to binding to human importin- α isoforms for efficient viral replication in mammalian cells (Gabriel et al., 2005; 2008; 2011; Hudjetz & Gabriel, 2012; Resa-Infante et al., 2008; 2014). Interaction with importin- α could either increase nuclear import of vRNP components or facilitate the use of importin- α as a co-factor for viral RdRP activity. It is important to note

however, that these studies were based on silencing of importin- α , therefore there could be the alternative conclusion that importin- α is necessary to import another host factor which acts as a required co-factor for viral RdRP activity into the nucleus. Avian influenza A viruses undergo a switch in importin- α dependency when subject to host adaptation. In mammalian cells, avian influenza A viruses rely on the presence of importin- $\alpha 3$, whereas mammalian-adapted influenza A viruses depend on importin- $\alpha 7$ (Hudjetz & Gabriel, 2012).

Large-scale screenings and bioinformatics analyses have identified numerous other host adaptive mutations in the PB2 subunit which are presented in Table 1.3 and Figure 1.8.

1.7.2.3 ANP32A is a determinant of host range restriction

A recent study identified the host factor acidic (leucine-rich) nuclear phosphoprotein 32 family member A (ANP32A), also known as pp32 or PHAPI, to be at the centre of avian to mammalian host adaptation of the viral RdRP (Long et al., 2016). ANP32A is a multifunctional host protein, which consists of an N-terminal leucine-rich repeat (LRR) domain with four LRR motifs and a C-terminal low-complexity acidic region (LCAR) (Chen et al., 1996; Matilla & Radrizzani, 2005). The structure of the LRR domain has been solved and characterised, showing the LRR motifs forming canonical parallel beta-sheet structures (de Chiara et al., 2007; Huyton & Wolberger, 2007; Tochio et al., 2010; Zamora-Caballero et al., 2015). The LCAR is approximately 100 amino acids long and contains 60 - 75% either glutamic acid or aspartic acid residues. The nature of the low-complexity region makes sequence homology searches very difficult and the lack of hydrophobic residues prevents formation of tertiary structures, allowing the LCAR to remain flexible in solution and to interact with any positively charged surface. ANP32A does not possess

any homo-oligomerisation motifs and it thus thought to exist in a monomeric state (Pan et al., 2009).

The subcellular localisation of ANP32A is still controversial as different studies have suggested it is either predominantly nuclear, predominantly cytoplasmic, shuttling between nucleus and cytoplasm, present on the cell surface or secreted (Reilly et al., 2014). This probably points to the fact that ANP32A is a very flexible and mobile protein and that its multitude of interactions with other host proteins mediate a wide range of subcellular localisations. Several potential phosphorylation sites have been proposed in ANP32A and phosphorylation by Casein kinase II has been verified (Hong et al., 2004). Phosphorylation is predicted to regulate and affect localisation and function of ANP32A (Fries et al., 2007; Ulitzur et al., 1997). ANP32A has been ascribed a multitude of different functions in the host cell, including chromatin remodelling and modification, apoptotic caspase modulation, protein phosphatase inhibition and regulation of intracellular transport (Reilly et al., 2014).

It was reported that mammalian hosts lack an activating avian host factor for efficient avian RdRP activity (Moncorge et al., 2010). A recent study found that expression of the avian homologue of ANP32A in human cells fully restored avian RdRP activity, while expression of other avian host factors previously proposed to regulate host range restriction, e.g. avian DDX17 or importin- α , did not have a stimulatory effect on the avian RdRP (Long et al., 2016). The avian and mammalian homologues of ANP32A have a very high sequence identity with the exception of a 33-amino acid insertion in the C-terminal LCAR domain of avian ANP32A. This insert is based on a 27 amino acid-long duplication of residues 149 to 175 and a unique 6 amino acid-long sequence (Long et al., 2016). Later the 6 amino acid-long unique sequence was detected to contain a hydrophobic SUMO interaction motif (SIM)-like sequence, which appears to be critical for stimulation of avian RdRP activity (Domingues & Hale, 2017).

Silencing of ANP32A and ANP32B in both human and avian cells revealed that ANP32A not only has a host range restrictive function, but that it also is required for viral RdRP activity in general (Long et al., 2016; Sugiyama et al., 2015). Following up on this it was demonstrated that human ANP32A and ANP32B both are required for viral replication but not transcription and that *in vitro* both are strong enhancers of vRNA, but not cRNA synthesis (Sugiyama et al., 2015).

ANP32A and ANP32B have been shown to interact with the viral RdRP in several large-scale proteomic studies (Bradel-Tretheway et al., 2011; York et al., 2014). An interaction between ANP32B and NA was also reported (Watanabe et al., 2014). In co-purification experiments it was demonstrated that ANP32A interacts with the full heterotrimeric viral RdRP, but not with monomeric individual RdRP subunits or with RdRPs in the context of RNPs (Sugiyama et al., 2015). Moreover, avian ANP32A was shown to form a much stronger interaction with the viral RdRP than human ANP32A (Domingues & Hale, 2017). Interaction of ANP32A with the viral RdRP is mediated by the C-terminal disordered LCAR region of ANP32A, with a 14-amino acid-long sequence between residues 220 and 235 being crucial for this interaction. The immediate downstream region (235 to 250) as well as the 6-amino acid-long SIM-like sequence are also important for binding of the viral RdRP. The unique avian SIM-like sequence appears to mediate enhanced binding and enhanced stimulation of the avian RdRP. Substitution of this SIM-like sequence with other known SIMs also allowed partial restoration of avian RdRP activity (Domingues & Hale, 2017). This suggests that indirect SUMOylation, a process where a SUMOylated host factor interacts with the SIM-like sequence of avian ANP32A, could be responsible for the species specific of avian ANP32A on viral polymerase function.

1.7.3 Host adaptation and the PB1 polymerase subunit

Table 1.4 List of all described host adaptive mutations in the PB1 subunit

Subunit	Residue	Adaptive mutation	Reference
PB1	52	K52R	(Naffakh et al., 2000)
	179	M179I	(Hu et al., 2014)
	216	S216G	(Hu et al., 2014)
	298	L298I	(Hu et al., 2014)
	327	R327K	(Chen et al., 2006)
	336	V336I	(Miotto et al., 2010)
	361	S361R	(Hu et al., 2014)
	375	N375S	(Kawaoka et al., 1989)
	430	R430K	(Hu et al., 2014)
	473	L473V	(Xu et al., 2012)
	486	R486K	(Hu et al., 2014)
	517	I517V	(Tamuri et al., 2009)
	581	E581D	(Hu et al., 2014)
	584	R584Q	(Tamuri et al., 2009)
	598	L598P	(Xu et al., 2012)
	621	Q621R	(Hu et al., 2014)
	741	A741S	(Hu et al., 2014)

Table 1.5 List of all described host adaptive mutations in the PA subunit

Subunit	Residue	Adaptive mutation	Reference
PA	28	P28L	(Chen et al., 2006)
	36	A36T	(Zhu et al., 2012)
	38	I38M	(Tan et al., 2014)
	55	D55N	(Chen et al., 2006)
	57	R57Q	(Naffakh et al., 2000)
	65	S65L	(Naffakh et al., 2000)
	66	G66D	(Miotto et al., 2010)
	85	T85I	(Bussey et al., 2011)
	97	T97I	(Song et al., 2009)
	100	V100A	(Naffakh et al., 2000)
	142	K142N	(Kim et al., 2010)
	186	G186S	(Bussey et al., 2011)
	204	R204K	(Hu et al., 2014)
	225	S225C	(Chen et al., 2006)
	256	R256K	(Hu et al., 2014)
	268	L268I	(Chen et al., 2006)
	277	S277H	(Hu et al., 2014)
	295	L295P	(Ilyushina et al., 2010)
	312	K312R	(Naffakh et al., 2000)
	336	L336M	(Bussey et al., 2011)
	337	A337S	(Naffakh et al., 2000)
	356	K356R	(Chen et al., 2006)
	382	E382D	(Taubenberger et al., 2005)
	400	P400L	(Mehle et al., 2012)
	404	A404S	(Chen et al., 2006)
	409	S409N	(Naffakh et al., 2000)
	421	S421I	(Kim et al., 2010)
	423	M423I	(Mehle et al., 2012)
	476	V476A	(Mehle et al., 2012)
	552	T552S	(Chen et al., 2006)
	615	K615R	(Gabriel et al., 2005)
	630	V630E	(Mehle et al., 2012)

The PB1 subunit of the viral RdRP appears to be least susceptible to host adaptive mutations (Table 1.4 and Figure 1.8). A switch in residue 375 of PB1 from an asparagine, which preferentially is found in avian viruses, to a serine, predominantly found in mammalian-adapted viruses, was reported previously (Kawaoka et al., 1989; Taubenberger et al., 2005). This mutation was present in the 1918, 1957 and 1968 pandemic strains (Naffakh et al., 2008). Because this switch is not entirely conserved in mammalian-adapted viruses, it is likely that compensatory mutations can be introduced in PB1 or a different viral protein (Taubenberger et al., 2005). The role of PB1-375 in viral polymerase function and, hence, the effect of an asparagine to serine substitution is still unknown.

A leucine to valine substitution at residue 473, as well as a leucine to proline substitution at residue 598 was shown to significantly increase avian RdRP activity both in avian and mammalian host cells and has been observed in the pandemic pH1N1 virus (Xu et al., 2012).

Recently, various bioinformatics studies have identified several other host adaptive mutations in the PB1 subunit (Chen et al., 2006; Hu et al., 2014; Miotto et al., 2010; Tamuri et al., 2009).

1.7.4 Host adaptation and the PA polymerase subunit

The role of PA in host adaptation is not as well characterised as the role of PB2. However, an increasing number of studies have shown that PA plays a significant role in overcoming host restriction (Table 1.5 and Figure 1.8). One of the most predominant host adaptive mutations in the PA subunit is a threonine to isoleucine substitution in residue 97 (T97I), which has been reported several times. One study found that in LPAI H5N2 it was selected for after serial passage in mice but not chicken and increased RdRP activity and viral replication (Nam et al., 2017; Song et al., 2009). Serial passage of an avian H6N1

virus in mice gave rise to the PA-T97I mutation (Yu et al., 2014). If introduced into an H6N1 avian influenza virus it conferred enhanced viral replication and virulence in mice (Cheng et al., 2014). Serial passage of avian H10N7 and H7N9 viruses in mice also resulted in the T97I mutations and enhanced viral replication (Wu et al., 2016; Zhao et al., 2016). It was also identified in a large-scale screening approach for mutations in the avian H5N1 RdRP complex that confer efficient replication in mammalian cells (Taft et al., 2015). While it has been widely reported on, the biological role of the T97I mutation remains elusive.

In the PA subunit of pH1N1 viruses mutations T85I, G186S and L336M have also been implicated in host adaptation (Bussey et al., 2011). Further RdRP activity and pathogenicity enhancing mutations were found by serially passaging the pH1N1 virus in mice: A36T, I38M and L295P (Ilyushina et al., 2010; Tan et al., 2014; Zhu et al., 2012). The only study to date that investigated the effect of PA mutations in both human and avian cells identified several mutations (P400L, M423I, V476A, T552S, V630E) that only increased avian polymerase activity in human cells (Mehle et al., 2012). The T552S mutation is also present in highly adapted seasonal influenza viruses. In addition, bioinformatics analyses have identified several other PA residues as host range determinants (Chen et al., 2006; Finkelstein et al., 2007; Miotto et al., 2010; Taubenberger & Morens, 2006; Tamuri et al., 2009).

The molecular mechanisms of the host adaptive effect of these mutations are not known. However, it is apparent that there is a redundancy in host adaptive mutations, *i.e.* certain mutations can compensate for the lack of others (Bussey et al., 2011). Furthermore, it has also been revealed that although host adaptive mutations in PA and other RdRP subunits increase avian RdRP activity in mammalian hosts, these mutations do not necessarily correlate with increased pathogenicity or virulence in mice (Bussey et al., 2011; Gabriel et al., 2005).

1.7 Thesis objectives and outline

Influenza A viruses predominantly infect avian host species, but occasionally also spread to human hosts, encountering a different host cellular environment. One of the main obstacles of avian influenza A viruses when infecting mammalian cells is to overcome the low activity of the avian RdRP in mammalian cells. A single E627K mutations in the PB2 subunit of the viral RdRP can overcome this host restrictive effect on the avian RdRP, but numerous other host adaptive mutations have been observed and reported on. Several of these host adaptive mutations cluster in or around the PB2 627-domain, a structurally distinct region which forms one of the flexible appendices of the viral RdRP. While the other protruding appendices of the viral RdRP have been characterised in detail, it is still unknown what the role of the PB2 627-domain in viral RdRP function is. To address this, a truncated RdRP lacking the 627-domain was generated and characterised using *in vitro* cell-based activity assays, revealing that the 627-domain is not required for core RdRP catalytic functions, but is important for the recruitment process of free RdRP to replicating RNPs (Chapter 3).

Recently, the host factor ANP32A was identified to be the crucial determinant of host range in influenza A virus infections. Expression of the avian ANP32A host factor in mammalian host cells is capable of restoring avian RdRP. ANP32A and the related ANP32B host proteins were also shown to be generally important for viral RdRP activity. Using *in vitro* and cell-based activity assays, the role of ANP32A in viral RdRP activity was further analysed, revealing that avian viruses are specifically restricted in synthesising vRNA in mammalian cells and that ANP32A enhances vRNA replication initiation (Chapter 4).

The viral RdRP has been shown to form oligomeric structures. Interestingly, many host adaptive mutations are found in the regions where potential RdRP dimer interfaces are formed. In vitro and cell-based assays were used to gain a better understanding of the biological role of these potential RdRP dimers and characterising the role of the RdRP dimer interface 3 (Chapter 5).

RNA-binding proteins are essential components of RNA metabolic processes and, thus, are highly regulated. Because of their central role in the life cycle of the cell, changes in the RNA-binding proteome can be observed if the cell is exposed to different forms of biological stimuli, including viral infections. Capturing the RNA-binding proteome in influenza A virus infected cells, proteins that change their mRNA binding profile following influenza virus infection were determined. Using the RNA-binding proteome, cellular processes that are affected by influenza virus infection, and host proteins which could have potential pro- or antiviral functions, were identified (Chapter 6).

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 General reagents

All general chemical reagents used were AnalaR grade and purchased from New England Biolabs (NEB), Roche, GE Healthcare Life Sciences, Millipore, Invitrogen Life Technologies, Sigma-Aldrich, Thermo Scientific, Stratagene, Agilent Technologies, Geneflow, VWR International, QIAGEN, Bio-Rad and Promega. The origins of special reagents are indicated separately.

2.1.2 Enzymes

Restriction endonuclease were purchased from NEB, Roche and Thermo Scientific. T4 Polynucleotide Kinase (PNK) was purchased from Thermo Scientific, Pfu Turbo DNA polymerase from Stratagene, Tobacco Etch Virus (AcTEV) protease and SuperScript II reverse transcriptase from Invitrogen, RNase A, RNase T1 and Proteinase K from Sigma-Aldrich, Calf intestinal alkaline phosphatase (CIP) and RNasin Ribonuclease Inhibitor from Promega.

2.1.3 Radiochemicals

Radiochemicals were all obtained from PerkinElmer (UK) Ltd. at the following specific activities:

[α -32P] GTP ~ 3000 Ci/mmol (10 mCi/ml)

[γ -32P] ATP ~ 3000 Ci/mmol (10 mCi/ml)

2.1.4 Synthetic DNA and RNA

Synthetic oligodeoxyribonucleotides were purchased from Sigma-Aldrich and Invitrogen.

2.1.5 Antibodies

The rabbit polyclonal anti-NP antibody was a kind gift from Otto Haller (University of Freiburg, Germany). The rabbit polyclonal anti-PB2, the rabbit polyclonal anti-PA and rabbit polyclonal anti-3P antibodies were kind gifts from Simon Carr, Tao Deng and Mónica Martínez-Alonso (Sir William Dunn School of Pathology, University of Oxford), respectively. The rabbit polyclonal anti-NEP/M1 antibody was a kind gift from Rob Ruigrok (EMBL Grenoble, France).

The rabbit polyclonal anti-PA (GTX118991), anti-PB1(GTX125923), anti-PB2 (GTX125925), anti-NS1 (GTX125990), anti-M1 (GTX125928), anti-PABPC1 (GTX113954) and anti-ADAR (GTX101602) antibodies were purchased from GeneTex. The rabbit polyclonal anti-ANP32A antibody (AV40203) and the mouse monoclonal anti-FLAG M2 (F3165) and anti- β -Actin (A5441) antibodies were purchased from Sigma-Aldrich. The rabbit monoclonal anti-PHAPI2 (Ab184565), rabbit polyclonal anti-PHAP1 (Ab51013) and mouse monoclonal anti-PTBP1 (Ab30317) antibodies were purchased from Abcam. The rabbit polyclonal anti-

KPNA2 (IHC-00201) antibody was purchased from Bethyl Laboratories. The rabbit polyclonal anti-XRCC5 (CSB-PA026233LA01HU) and anti-HNRNPR (CSB-PA010614GA01HU) antibodies were purchased from CUSABIO.

The horseradish peroxidase (HRP) conjugated goat anti-Rabbit IgG (A8275) and goat anti-Mouse IgG (A4416) antibodies were purchased from Sigma-Aldrich. The Alexa Fluor 532 conjugated goat anti-Rabbit IgG (A-11009) was purchased from Thermo Scientific.

2.1.6 Bacteria and bacterial culture media

Escherichia coli strain DH5 α was purchased from Invitrogen. Escherichia coli strain BL21 (DE3) was purchased from NEB.

Bacterial culture media 2xYT (16 g bacto-tryptone, 10 g yeast extract and 5 g NaCl per litre), Lysogeny broth (LB) (10 g bacto-tryptone, 5 g yeast extract and 10 g NaCl per litre) and nutrient agar plates with 100 μ g/ml ampicillin were supplied by the departmental media service of the Sir William Dunn School of Pathology. Ampicillin was added to culture media at a final concentration of 100 μ g/ml.

2.1.7 Cell culture media

Minimal Essential Medium (MEM) and Dulbecco's Modified Eagle Medium (DMEM), Foetal Calf Serum (FCS), Phosphate Buffered Saline (PBS), L-Arginine and L-Lysine were purchased from Sigma-Aldrich. Dulbecco's Modified Eagle Medium (DMEM) for SILAC was purchased from Thermo Scientific. TrypLE Express dissociation reagent, Opti-MEM and L-Glutamine were purchased from Invitrogen. Stable isotope-labelled amino acids medium (¹³C₆-L-Arginine HCl) and heavy (¹³C₆, ¹⁵N₄-L-Arginine HCl) L-Arginine and

medium (4,4,5,5-D₄-L-Lysine 2HCl) and heavy (13C₆, 15N₂-L-Lysine HCl) L-Lysine were purchased from Silantes.

2.1.8 Cell lines

Human Embryonic Kidney 293T (HEK-293T) cells, Madin-Darby Bovine Kidney Epithelial (MDBK) cells, Madin-Darby Canine Epithelial (MDCK) cells, human adenocarcinoma alveolar basal epithelial (A549) cells and chicken embryo fibroblast (DF-1) cells were obtained from the cell bank of the Sir William Dunn School of Pathology, University of Oxford.

2.1.9 Viruses

Influenza A /WSN/33 (H1N1) virus (WSN) and mutant avian-like influenza / A /WSN/33-K627E (H1N1) virus (WSN-K627E) were generated using reverse genetics (Hoffmann et al., 2000).

2.1.10 Plasmids

Protein expression plasmids pcDNA-PB1, pcDNA-PB2, pcDNA-PA, pcDNA-NP and pcDNA3A (Fodor et al., 2002) as well as pcDNA-PB1a (Vreede et al., 2004), pcDNA-PA-TAP and pcDNA-PB2-TAP (Deng et al., 2005), pcDNA-PB2-K627E (Ng et al., 2012b), pcDNA-PA-C95A, pcDNA-PA-T97A, pcDNA-PA-T98A, pcDNA-PA-G99A, pcDNA-PA-D108A (Hara et al., 2006), pcDNA-PB2-F404A (Fechter et al., 2003) and pcDNA-PP7CP-Strep (York et al., 2013) have been described previously.

Protein expression plasmids pcDNA-PA Δ 51-72 was a kind gift from Narin Hengrung (Sir William Dunn School of Pathology). Protein expression plasmids pcDNA-PA-71A/72A/73A, pcDNA-317A/318A/319A, pcDNA-346A/347A/348A, pcDNA-PB1-565A/566A/567A, pcDNA-PB2-71A/72A/73A, pcDNA-PB2-273A/274A/275A, pcDNA-PB2-499A/500A/501A and pcDNA-PB2-627A/628A/629A were kind gifts from Jane Sharps and Ervin Fodor (Sir William Dunn School of Pathology, University of Oxford). Protein expression plasmids pcDNA-PB2-423A/424A/425A and pcDNA-PB2-427A/428A/429A were kind gifts from Mónica Martínez-Alonso (Sir William Dunn School of Pathology). Protein expression plasmids pCAGGS-FLAG-KPNA1, pCAGGS-FLAG-KPNA2 and pCAGGS-FLAG-KPNA6 were kind gifts from Othmar Engelhardt (Sir William Dunn School of Pathology, University of Oxford).

The vRNA transcription plasmids pPOLI-NA (Fodor et al., 1999) and pPRC425-NA and pPOLI-NA47 (Paterson et al., 2014) have been described before. The cRNA transcription plasmid pPOLI-cNA47 was a kind gift from Aartjan te Velhuis (Sir William Dunn School of Pathology, University of Oxford).

The bidirectional pHW2000 plasmids used for influenza A/WSN/33 (H1N1) virus reverse genetics have been described previously (Hoffmann et al., 2000) and were a kind gift from Gülsah Gabriel (University of Hamburg, Germany). The bidirectional pHW2000-PB2-K627E plasmid was a kind gift from Duncan Paterson (Sir William Dunn School of Pathology, University of Oxford).

Protein expression plasmids pCAGGS-chANP32A, pCAGGS-huANP32A and pCAGGS-huANP32B have been described previously (Long et al., 2016) and were kind gifts from Wendy Barclay (Imperial College London).

Table 2.1 Sequence of primers used in this study

Primer	Sequence (5' to 3')
PCR mutagenesis	
pcDNA-PB2Δ535-667-F	ATAACTTACTCATCGTCACTCGGAAAGGATGCTGGC
pcDNA-PB2Δ535-667-R	GCCAGCATCCTTTCCGAGTGACGATGAGTAAGTTAT
pcDNA-PB2Δ248-759-F	GGAACAGATGTACACTCCAGGATAGTGTCTGAATAGTTAAAAAC
pcDNA-PB2Δ248-759-R	GTTTTTAAACTATTTCGACACTATCCTGGAGTGACATCTGTTC
pcDNA-PB2Δ248-679-F	GATGTACACTCCAGGAGATGAAGGCACAGCTG
pcDNA-PB2Δ248-679-R	CAGCTGTGCCTTCATCTCCTGGAGTGACATC
pcDNA-PB2Δ213-679-F	GGAGAGAGAACTGGTCGATGAAGGCACAGCTG
pcDNA-PB2Δ213-679-R	CAGCTGTGCCTTCATCGACCAGTTCTCTCTCC
pcDNA-PB2Δ154-679-F	TCCTGGTCATGCAGATGATGAAGGCACAGCTG
pcDNA-PB2Δ154-679-R	CAGCTGTGCCTTCATCATCTGCATGACCAGGA
pcDNA-PB2Δ111-679-F	GTGACAAGTACAGTTCATGATGAAGGCACAGCTGGA
pcDNA-PB2Δ111-679-R	TCCAGCTGTGCCTTCATCATGAACTGTACTTGTCAC
pcDNA-PB2Δ55-679-F	GATGAAATGGATGATGGCAATGAAAGATGAAGGCACAGC
pcDNA-PB2Δ55-679-R	GCTGTGCCTTCATCTTTCATTGCCATCATCCATTTCATC
pcDNA-PB2-E361A-F	GAGTGCATGAGGGATATGCAGAGTTCACAATGGTTGG
pcDNA-PB2-E361A-R	CCAACCATGTGAACTCTGCATATCCCTCATGCACTC
Cloning	
NotI-Strep-F	GGCCGCATGTGGAGCCACCCCCAGTTCGAGAAGAGCGCC
NotI-Strep-R	GGCCGGCGCTCTTCTCGAACTGGGGGTGGCTCCACATGC
Primer extension	
NA+	TCCAGTATGGTTTTGATTTCGG
NA-	TGGA TAGTGGGAGCATCAT
NA47+	CATTAAACTCCTGCTTTCGC
NA47-	GTTCAAAAAACTCCTTGTTTC
5S rRNA	TCCCAGGCGGTCTCCCATCC

2.2 Cloning

2.2.1 PCR mutagenesis

Mutations were introduced into PB2 cDNA sequences in pcDNA-PB2 with PCR primers shown in Table 2.1. The PCR reactions were carried out using 100 ng template plasmid, 25 pmol of each primer, 200 μ M of each dNTP, 3% DMSO and Pfu Turbo DNA polymerase in 1 $\times$ Pfu Turbo reaction buffer. The PCR reaction mixture was incubated in a thermocycler at 95°C for 30 s denaturing, 50°C for 30 s annealing and 72°C for 16 min elongation for a total of 18 cycles. PCR products were digested with DpnI, column purified and transformed into E. coli DH5 α cells.

2.2.2 Plasmid construction

To construct N-terminally Strep-tagged ANP32A and ANP32B constructs, pCAGGS-chANP32A, pCAGGS-huANP32A and pCAGGS-huANP32B were digested with NotI and subsequently dephosphorylated with FastAP and column purified. Oligodeoxyribonucleotides containing terminal NotI overhang sequences, an N-terminal start codon, a single Strep-tag sequence and a SA-linker (Table 2.1) were phosphorylated with PNK and annealed. Annealed inserts were ligated into the digested vectors using Takara DNA Ligation Kit and transformed into E. coli DH5 α cells.

2.2.3 Transformation into competent *E. coli*

Chemically competent *E. coli* DH5 α cells were thawed on ice and 8 μ l bacterial cells were incubated with 2 μ l plasmid DNA for 5 min on ice. Cells were briefly heat-shocked at 42°C for 1 min before being incubated on ice for a further 5 min. 190 μ l of SOC Medium (Sigma-Aldrich) was added to transformed cells and incubated for 30 min at 37°C. Cells were spread onto nutrient agar plates containing 100 μ g/ml ampicillin and incubated at 37°C overnight.

2.2.4 Plasmid preparation and sequencing

Bacterial colonies containing the plasmid of interest were picked and incubated in 5 ml LB medium supplemented with 100 μ g/ml ampicillin in a shaking incubator overnight at 37°C. Cells were pelleted and plasmid DNA was extracted using QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturer's instructions. Plasmid DNA was eluted in 50 μ l ddH₂O.

For large scale plasmid DNA preparation, bacterial colonies were inoculated in 100 ml LB medium supplemented with 100 μ g/ml ampicillin in a shaking incubator overnight at 37°C. Cells were pelleted and plasmid DNA was extracted using Genopure Plasmid Maxi Kit (Roche) according to the manufacturer's instructions. Plasmid DNA was dissolved in ddH₂O to a final concentration of 1 mg/ml.

All plasmids were sequenced by Source Bioscience (Oxford) using Sanger sequencing. Sequence analysis was conducted using Clustal Omega software (<http://www.ebi.ac.uk/Tools/msa/clustalo>).

2.3 Eukaryotic cell culture

HEK-293T cells, MDCK cells, DF-1 cells and A549 cells were maintained in DMEM supplemented with 10% FCS. MDBK cells were maintained in MEM supplemented with 2 mM L-glutamine and 10% FCS. For stable isotope labelling with amino acids in cell culture (SILAC) of A549 cells, A549 cells were grown for >5 passages in DMEM for SILAC supplemented with 10% FCS. DMEM for SILAC was supplemented with 84 mg/l L-Arginine and 146 mg/l L-Lysine. For light labelling L-arginine and L-lysine was used, for medium labelling $^{13}\text{C}_6$ -L-Arginine HCl and 4,4,5,5- D_4 -L-Lysine 2HCl and for heavy labelling $^{13}\text{C}_6$, $^{15}\text{N}_4$ -L-Arginine HCl and $^{13}\text{C}_6$, $^{15}\text{N}_2$ -L-Lysine HCl. After 5 passages SILAC incorporation was determined by mass spectrometry and assessed for >99% incorporation rate. All cells were maintained in humidified incubators at 37°C and 5% CO_2 . Cells were passaged by detaching the cell monolayer from the cell culture flask surface with TrypLE Express and diluted 1:10 in the appropriate cell culture media.

2.4 Virus culture

2.4.1 Reverse genetics

Wild-type and mutant influenza A/WSN/33 (H1N1) viruses were rescued using the eight plasmid reverse genetics system (Hoffmann et al., 2000). HEK-293T cells were grown in 6 cm dishes and transfected with 1 µg of each of the eight bidirectional plasmid (pHW2000) containing the influenza A virus segment genomic and transcript sequences using Lipofectamine 2000 transfection reagent (Invitrogen) according to the manufacturer's specifications. Transfected cells were incubated at 37°C and 5% CO_2 in DMEM (10% FCS)

for 24 h and then rinsed with PBS and the medium was replaced with low serum DMEM (0.5% FCS) and the cells were incubated for 48 h at 37°C and 5% CO₂.

Amplification of virus stocks was carried out by applying virus-containing medium to 80% confluent MDBK or MDCK cells in 175 cm² cell culture flasks for 1 h at room temperature in low serum MEM or DMEM (0.5% FCS) and subsequent incubation at 37°C and 5% CO₂ for 72 h. Media containing virus was harvested and cleared from cellular debris by centrifugation at 4,000 × g for 10 min at 4°C. Supernatants were stored at -80°C. Viral titres were assessed by performing plaque assays.

2.4.2 Infection of mammalian cells with influenza A virus

Influenza A viruses were diluted in the appropriate cell culture media supplemented with 0.5% FCS to the desired multiplicity of infection (MOI). Cell monolayers were infected with diluted virus for 1 h at room temperature in a volume sufficient to cover the cell monolayer. Fresh low serum medium (0.5% FCS) was added after 1 h and cells were incubated at 37°C and 5% CO₂. Cells were harvested at the indicated time points. If transiently transfected cells were infected, infections were performed 24 to 48 h post-transfection.

2.4.3 Plaque assay

Influenza A viruses were serially diluted tenfold in a total volume of 250 µl MEM supplemented with 0.5% FCS and 2 mM L-glutamine. MDBK cells were rinsed with PBS and infected as monolayers in 6-well plates with 200 µl of serially diluted virus for 1 h at

room temperature. After infection, the inoculum was aspirated and 2 ml of agarose solution (2% agarose dissolved in PBS diluted 1:1 in MEM supplemented with 0.5% FCS and 2 mM L-glutamine) was overlaid and allowed to polymerise at room temperature. The plates were then incubated at 37°C and 5% CO₂ for 72 h, followed by removal of the agar overlay and staining of the plaques using 2 ml of 0.2% Coomassie Brilliant Blue dye in 7.5% glacial acetic acid and 50% methanol.

2.4.4 Growth curve analysis

Viral growth was analysed by infecting 80% confluent HEK-293T cells in 3.5 cm dishes at a MOI of 0.01 with influenza A viruses and harvesting the supernatant at 6 h, 24 h and 48 h post-infection. Viral titres were determined by plaque assay.

2.4.5 Purification of influenza A virus

Influenza A/WSN/33 virus was purified as described before (Hutchinson et al., 2014). Harvested virus was cleared from cellular debris by centrifugation at 4,000 × g and 4°C for 10 min. Cleared virus was transferred to thick-walled ultracentrifuge tubes and further clarified by ultracentrifugation at 17,000 × g and 4°C for 15 min in a SW 32 Ti Rotor and Optima XPN-80 ultracentrifuge (Beckman Coulter). Supernatants were layered onto a cushion of 30% sucrose in virus suspension buffer (VSB) containing 5 mM CaCl₂, 100 mM NaCl and 20 mM Tris-HCl [pH 7.4] and pelleted by ultracentrifugation at 112,000 × g and 4°C for 90 min. Supernatants were carefully aspirated and viral pellets were resuspended

in 1 ml VSB and incubated overnight at 4°C. Resuspended purified virus was harvested and stored at -80°C. Viral titres were determined by plaque assay.

2.5 Viral RNA analysis

2.5.1 RNP reconstitution assay

Viral RNPs were reconstituted in approximately 1×10^6 HEK-293T or DF-1 cells in 3.5 cm dishes in 1.5 ml DMEM (10% FCS) by transfection of 1 µg of each of the indicated plasmids using Lipofectamine 2000 transfection reagent (Invitrogen) according to the manufacturer's instructions. Cells were incubated at 37°C and 5% CO₂ and harvested 24 - 48 h post-transfection.

2.5.2 cRNA stabilisation assay

Approximately 1×10^6 HEK-293T were transfected in 3.5 cm dishes in 1.5 ml DMEM (10% FCS) with 1 µg of each of the RdRP subunit expressing plasmids and if indicated NP, ANP32A and ANP32B using Lipofectamine 2000 transfection reagent (Invitrogen) according to the manufacturer's instructions. 48 h post-transfection the media was aspirated and cells were infected with influenza A viruses at an MOI of 5 in the presence of 5 µg/ml of actinomycin D (ACTD) for 1 h at room temperature. Cells were incubated at 37°C for 6 h before cells were harvested and total RNA was extracted.

2.5.3 RNA extraction

Unless stated otherwise, total RNA was extracted from cells or samples using TRI Reagent (Sigma-Aldrich) according to the manufacturer's instructions. Samples were lysed in an appropriate volume of TRI Reagent (typically 1 ml) and incubated for 5 min at room temperature. 0.2 volumes of chloroform were added and samples were vigorously agitated, followed by centrifugation at $17,000 \times g$ and 4°C for 20 min. The clear upper aqueous phase was transferred to new tubes and supplemented with an equal volume of isopropanol. If indicated 20 μg of tRNA from *E. coli* MRE 600 (Roche) was added to aid precipitation of RNA. RNA was precipitated by centrifugation at $17,000 \times g$ and 4°C for 20 min. The RNA pellet was washed in 70% ethanol and dissolved in 10 μl ddH₂O and stored at -20°C .

2.5.4 Primer extension analysis

2.5.4.1 Primers

Influenza A virus segment 6 derived RNAs and cellular 5S rRNA were detected using specific oligodeoxyribonucleotides as presented in Table 2.1. Primer NA⁺ was used to detect mRNA and cRNA and primer NA⁻ was used to detect vRNA derived from viral segment 6. Primer NA47⁺ was used to detect mRNA and cRNA and primer NA47⁻ was used to detect vRNA derived from a 47-nucleotide long viral segment 6 as previously characterised (Paterson et al., 2014).

2.5.4.2 Radiolabelling of oligodeoxyribonucleotides

Oligodeoxyribonucleotides were labelled in a 10 µl reaction containing 10 pmol of relevant primer, 10 µCi [γ -32P] ATP and 2.5 U of PNK in 1× kinase buffer A. The reaction was performed for 1 h at 37°C. Excess unlabelled [γ -32P] ATP and reaction constituents were removed using the QIAquick Nucleotide Removal Kit (QIAGEN) according to the manufacturer's specifications. Radiolabelled oligodeoxyribonucleotides were eluted in 30 µl ddH₂O and stored at -20°C.

2.5.4.3 Primer extension

Primer extension assays were performed as previously described (Robb et al., 2009). Briefly, in a 5 µl reaction, RNA was mixed with an excess of radiolabelled primers complementary to positive and negative sense viral segment 6 RNA and complementary to cellular 5S rRNA which is used as an internal loading control. The mixture was denatured by heating to 95°C for 3 min and immediately cooled on ice. 5 µl of transcription mix (2× First Strand Buffer [Invitrogen], 20 mM dithiothreitol [DTT], 1 mM dNTP mix [Thermo Scientific] and 50 U Superscript III RT [Invitrogen]) was added and the reaction was heated to 50°C for 1 h, before heat inactivating the reaction at 70°C for 10 min.

Reactions were terminated by addition of 10 µl of 80% formamide, 1 mM EDTA, bromophenol blue and xylene cyan and incubation at 95°C for 3 min. Reaction products were resolved by 6% (for full-length NA segment template) or 12% (for 47-nucleotide NA segment template) denaturing PAGE containing 7 M urea in TBE buffer and visualised by autoradiography.

2.5.5 Polyacrylamide gel electrophoresis (PAGE)

DNA products from primer extension assays or RNA products from *in vitro* transcription assays were analysed by PAGE (acrylamide:bisacrylamide, 19:1) (ProtoGel, National Diagnostics) in 7M urea and 1× TBE (90 mM Tris, 90 mM orthoboric acid and 2.5 mM EDTA). Gels were polymerised by addition of ammonium persulfate (APS) to 0.1% and N,N,N',N'-Tetramethylethylenediamine (TEMED) to 0.1%. Gels were run at 1250 V in 1× TBE buffer. DNA or RNA products were detected by autoradiography and quantitated using phosphor Fujifilm imaging plates and a Fujifilm FLA-5000 fluorescent image analyser. Autoradiography data was analysed using ImageJ software.

2.6 Protein purification and analysis

2.6.1 Recombinant influenza A virus RdRP purification

Recombinant influenza A/WSN/33 virus RdRP was transiently expressed in mammalian cells by transfecting approximately 5.5×10^6 HEK-239T cells with 5 µg of each RdRP subunit expression plasmid (pcDNA-PA-TAP/pcDNA-PA, pcDNA-PB1 and pcDNA-PB2-TAP/pcDNA-PB2 or variants of these) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were harvested 48 h post-transfection by centrifugation at $200 \times g$ for 5 min at 4°C. Cells were washed with ice cold PBS and pelleted by centrifugation at $200 \times g$ for 5 min at 4°C and subsequently lysed in 1 ml Tris lysis buffer (50 mM Tris-HCl [pH 8.0], 200 mM NaCl, 33% glycerol, 0.5% Igepal CA-630, 1 mM DTT, 1× complete EDTA-free protease inhibitor cocktail) rotating for 1 h at 4°C. The insoluble and soluble fractions were separated by centrifugation at $17,000 \times g$ for 5 min at 4°C. The soluble fraction was diluted 1:5 with binding buffer (20 mM Tris-HCl [pH 8.0],

150 mM NaCl and 1× complete EDTA-free protease inhibitor cocktail) and incubated with 100 µl of washed IgG Sepharose (GE Healthcare Life Sciences) for 3 h by rotation at 4°C. IgG Sepharose was washed extensively with wash buffer (20 mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.1% Igepal CA-630, 10% glycerol and 1× phenylmethanesulfonyl fluoride solution [PMSF]). The bound recombinant RdRP was released by cleavage of the protein A tag overnight at 4°C with 20 U of AcTEV (Invitrogen) in 500 µl cleavage buffer (10 mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.1% Igepal CA-630, 10% glycerol, 1 mM DTT and 1× complete EDTA-free protease inhibitor cocktail). Released viral RdRP was separated from IgG Sepharose by centrifugation at $17,000 \times g$ for 5 min at 4°C and stored at -80°C. Purified RdRP was analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

2.6.2 Recombinant ANP32A and ANP32B purification

Recombinant chicken ANP32A and human ANP32A and ANP32B proteins were transiently expressed in mammalian cells by transfecting approximately 5.5×10^6 HEK-239T cells with 10 µg of expression plasmid (pCAGGS-Strep-chANP32A, pCAGGS-Strep-huANP32A, pCAGGS-Strep-huANP32B) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Cells were harvested 48 h post-transfection by centrifugation at $200 \times g$ for 5 min at 4°C. Cells were washed with ice cold PBS and pelleted by centrifugation at $200 \times g$ for 5 min at 4°C and subsequently lysed in 1 ml Tris lysis buffer (50 mM Tris-HCl [pH 8.0], 200 mM NaCl, 30% glycerol, 0.5% Igepal CA-630, 1 mM DTT, 1× complete EDTA-free protease inhibitor cocktail) rotating for 1 h at 4°C. The insoluble and soluble fractions were separated by centrifugation at $17,000 \times g$ for 5 min at 4°C. The soluble fraction was incubated with 100 µl of washed 50% suspension Strep-Tactin Superflow high-capacity resin (IBA GmbH) for 3 h by rotation at 4°C. The resin was

washed extensively with wash buffer (20 mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.1% Igepal CA-630, 10% glycerol and 1× PMSF). Proteins were eluted in 500 µl 1× Buffer BE Strep-Tactin Elution Buffer with Biotin (IBA GmbH) rotating overnight at 4°C. Released protein was separated from the resin by centrifugation at $17,000 \times g$ for 5 min at 4°C.

If proteins were to be analysed by mass spectrometry, samples were submitted to the Advanced Proteomics Facility (University of Oxford) at this stage.

If samples were to be used for *in vitro* activity assays, eluted proteins were concentrated in polymerase cleavage buffer (10 mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.1% Igepal CA-630, 10% glycerol, 1 mM DTT and 1× complete EDTA-free protease inhibitor cocktail) using Amicon Ultra-4 Centrifugal Filters (0.5 ml, 3 kDa MWCO) and stored at -80°C. Purified proteins were analysed by SDS-PAGE.

2.6.3 Protein gel electrophoresis

2.6.3.1 SDS-PAGE

Protein samples were mixed with an equal volume of 2× Lämmli buffer (125 mM Tris-HCl [pH 6.8], 4% SDS, 20% glycerol, 10% β-mercaptoethanol and 0.01% bromophenol blue) and heated at 95°C for 5 min. Samples were then resolved by electrophoresis using a discontinuous PAGE system (Bio-Rad). Resolving gels contained 6-14% acrylamide (acrylamide:bisacrylamide, 29:1) in 375 mM Tris-HCl (pH 8.8) and 0.1% SDS. Stacking gels contained 3.2% acrylamide (acrylamide:bisacrylamide, 29:1) in 125 mM Tris-HCl (pH 6.8) and 0.1% SDS. Polymerisation of the gels was induced by addition of 0.1% APS and 0.1% TEMED. Samples were run at 120 V in SDS-PAGE running buffer (25 mM Tris, 250 mM

glycine and 0.1% SDS) at room temperature. A prestained protein marker was run in parallel.

2.6.3.2 Western blotting

Proteins were transferred from SDS-PAGE gels onto polyvinylidene difluoride (PDVF) or nitrocellulose membranes in transfer buffer (25 mM Tris, 190 mM glycine, 20% methanol) at 10 V and room temperature using the Trans-Blot Turbo Transfer system (Bio-Rad). Membranes were then blocked in 5% non-fat milk in PBS for 1 h at room temperature or overnight at 4°C. Membranes were subsequently incubated with the indicated antibodies diluted in 5% non-fat milk in PBS for 1 h at room temperature on a rotating wheel. Membranes were washed three times in PBS (0.1% Tween-20) for 10 min each. Membranes were then incubated with an appropriate secondary antibody conjugated to horseradish peroxidase (HRP) for 1 h at room temperature. Membranes were washed three times in PBS (0.1% Tween-20) for 10 min each again. Membranes were developed by chemiluminescence using Immobilon Western Chemiluminescent HRP Substrate (Millipore). Membranes were incubated for 1 min with the chemiluminescent reagent, drained and exposed to X-ray film (Super RX Fuji Medical X-Ray film, Kodak).

2.6.3.3 Staining of protein gels

Gels were silver stained using SilverXpress kit (Invitrogen) according to the manufacturer's instructions. SDS-PAGE gels were transferred to a clean plastic dish and fixed in 50% methanol and 10% acetic acid for 10 min. The fixed gel was sensitised in 50% methanol and 2.5% sensitiser, followed by three washes in dH₂O. Gels were stained and developed according the manufacturer's instructions.

For Coomassie staining the Instant Blue (Expedeon) was used according to the manufacturer's specifications.

2.7 *In vitro* activity assays

2.7.1 Capped-oligonucleotide cleavage assay

The ability of RdRP to cleave capped-oligonucleotides was assessed in a capped-oligonucleotide cleavage assay as described before (Velthuis et al., 2016). Synthetic 20-nucleotide RNA (Table 2.2) with a 5' diphosphate (Chemgenes) was capped with a radiolabelled cap-1 structure in a 20 µl reaction containing 1 µM RNA, 0.25 µM [α -32P] GTP (3,000 Ci/mmol, Perkin-Elmer), 0.8 mM S-adenosylmethionine, 0.5 U/µl Vaccinia virus capping enzyme (NEB) and 2.5 U/µl 2'-O-methyltransferase (NEB) at 37°C for 1 h. The product was analysed by 16% denaturing PAGE, excised, eluted overnight in deionised water and desalted using NAP-10 columns (GE Healthcare).

Capped-RNA cleavage assays were performed in a reaction containing 5 mM MgCl₂, 1 mM DTT, 2 U/µl RNasin, 0.5 µM 5' vRNA promoter, 0.5 µM 3' vRNA promoter (Table 2.2), ~1,500 cpm capped RNA primer and 5 ng/µl RdRP in elution buffer. Reaction mixtures were incubated at 30°C for 1-4 h and terminated by addition of an equal volume of 80% formamide, 1 mM EDTA, bromophenol blue and xylene cyan and incubation at 95°C for 3 min. Reaction products were resolved by 20% denaturing PAGE containing 7 M urea in TBE buffer and visualised by autoradiography.

2.7.2 Capped-oligonucleotide extension assay

The ability of RdRP to extend capped-oligonucleotides was assessed in capped-oligonucleotide extension assays as described before (Velthuis et al., 2016). Synthetic 11-nucleotide RNA (Table 2.2) with a 5' diphosphate (Chemgenes) was capped with a

Table 2.2 Sequences of oligoribonucleotides used in this study

Oligo	Sequence (5' to 3')
5' vRNA promoter	pppAGUAGAAACAAGGCC
3' vRNA promoter	pppGGCCUGCUUUUGCU
5' cRNA promoter	pppAGCAAAAGCAGGCC
3' cRNA promoter	pppGGCCUUGUUUCUACU
3' vRNA 40-nt	pppGCAUGUCACUCUGUGAGUGAUUAUCUGGCCUGCUUUUGCU
3' cRNA 41-nt	pppGCAUGUCACUCUGUGAGUGAUUAUCUGGCCUUGUUUCUACU
11-nt RNA	ppGAAUACUCAAG
20-nt RNA	ppAAUCUAUAAUAGCAUUAUCC

radiolabelled cap-1 structure in a 20 μ l reaction containing 1 μ M RNA, 0.25 μ M [α -32P] GTP (3,000 Ci/mmol, Perkin-Elmer), 0.8 mM S-adenosylmethionine, 0.5 U/ μ l Vaccinia virus capping enzyme (NEB) and 2.5 U/ μ l 2'-O-methyltransferase (NEB) at 37°C for 1 h. The product was analysed by 16% denaturing PAGE, excised, eluted overnight in deionised water and desalted using NAP-10 columns (GE Healthcare).

Capped-RNA cleavage assays were performed in a reaction containing 5 mM MgCl₂, 1 mM DTT, 2 U/ μ l RNasin, 0.5 μ M 5' vRNA promoter, 0.5 μ M 3' vRNA promoter, 1 mM ATP, 0.5 mM UTP, 0.5 mM CTP, 0.5 mM GTP, ~1,500 cpm capped RNA primer (11-nucleotide or 20-nucleotide RNA) and 5 ng/ μ l RdRP in elution buffer. Reaction mixtures were incubated at 30°C for 1-4 h and terminated by addition of an equal volume of 80% formamide, 1 mM EDTA, bromophenol blue and xylene cyan and incubation at 95°C for 3 min. Reaction products were resolved by 20% denaturing PAGE containing 7 M urea in TBE buffer and visualised by autoradiography.

2.7.3 Globin mRNA transcription assay

The ability of RdRP to use β -globin mRNA as a primer for transcription was assessed as described previously (Kerry et al., 2008). Reaction mixtures containing 0.05 μ M [α -32P] GTP (3,000 Ci/mmol, Perkin-Elmer), 1 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 50 ng rabbit β -globin mRNA (Sigma-Aldrich), 5 mM MgCl₂, 1 mM DTT, 2 U/ μ l RNasin, 0.5 μ M 5' vRNA promoter, 0.5 μ M 3' vRNA promoter and 5 ng/ μ l RdRP in elution buffer were incubated at 30°C for 1 - 12 h, followed by heating to 95°C for 3 min. Reactions were terminated by addition of an equal volume of 80% formamide, 1 mM EDTA, bromophenol blue and xylene cyan and incubation at 95°C for 3 min. Reaction products were resolved by 20% denaturing PAGE containing 7 M urea in TBE buffer and visualised by autoradiography.

2.7.4 Dinucleotide synthesis assay

The ability of RdRP to synthesise a pppApG dinucleotide *de novo* was assessed as described before (Velthuis et al., 2016). For this, 0.05 μM [α -32P] GTP (3,000 Ci/mmol, Perkin-Elmer), 1 mM ATP, 5 mM MgCl_2 , 1 mM DTT, 2 U/ μl RNasin, 0.5 μM 5' vRNA or cRNA promoter, 0.5 μM 3' vRNA or cRNA promoter (Table 2.2) and 5 ng/ μl RdRP in elution buffer were incubated at 30°C for 1 - 12 h, followed by heating to 95°C for 3 min. Reactions were treated with 1 U FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific) for 1 h at 37°C. Reactions were terminated by addition of an equal volume of 80% formamide, 1 mM EDTA, bromophenol blue and xylene cyan and incubation at 95°C for 3 min. Reaction products were resolved by 20% denaturing PAGE containing 7 M urea in TBE buffer and visualised by autoradiography.

2.7.5 Primer-independent replication assay

The ability of RdRP to initiate replication *de novo* was assessed as described previously (Velthuis et al., 2016). Reaction mixtures containing 0.05 μM [α -32P] GTP (3,000 Ci/mmol, Perkin-Elmer), 1 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 5 mM MgCl_2 , 1 mM DTT, 2 U/ μl RNasin, 0.5 μM 5' vRNA or cRNA promoter, 0.5 μM 3' vRNA or cRNA promoter, 1 $\mu\text{g}/\mu\text{l}$ RdRP in elution buffer and if indicated 100 ng/ μl ANP32A or ANP32B were incubated at 30°C for 1 - 12 h, followed by heating to 95°C for 3 min. Reactions were terminated by addition of an equal volume of 80% formamide, 1 mM EDTA, bromophenol blue and xylene cyan and incubation at 95°C for 3 min. Reaction products were resolved by 20% denaturing PAGE containing 7 M urea in TBE buffer and visualised by autoradiography. Recombinant influenza A/Northern Territory/60/68 (H3N2) virus or influenza A/duck/Fujian/01/2002 (H5N1) virus RdRPs were expressed in *Spodoptera frugiperda* Sf9 cells

using a baculovirus expression system and were a kind gift from Itziar Serna Martin and Aartjan te Velhuis (Sir William Dunn School of Pathology, University of Oxford).

2.7.6 ApG extension assay

The ability of RdRP to extend an ApG dinucleotide was assessed as described previously (Deng et al., 2006b). Reaction mixtures containing 0.05 μ M [α -32P] GTP (3,000 Ci/mmol, Perkin-Elmer), 1 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 0.25 mM ApG (IBA Lifesciences), 5 mM MgCl₂, 1 mM DTT, 2 U/ μ l RNasin, 0.5 μ M 5' vRNA or cRNA promoter, 0.5 μ M 3' vRNA or cRNA promoter and 5 ng/ μ l RdRP in elution buffer were incubated at 30°C for 1 - 12 h, followed by heating to 95°C for 3 min.

In order to assess the ability of the RdRP to extend an ApG dinucleotide along a longer RNA template, reaction mixtures containing 0.05 μ M [α -32P] GTP (3,000 Ci/mmol, Perkin-Elmer), 1 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 0.25 mM ApG (IBA Lifesciences), 5 mM MgCl₂, 1 mM DTT, 2 U/ μ l RNasin, 0.5 μ M 5' v RNA or cRNA promoter, 0.5 μ M 3' 40-nt-long vRNA or 41-nt-long cRNA (Table 2.2) and 5 ng/ μ l RdRP in elution buffer were incubated at 30°C for 12 h, followed by heating to 95°C for 3 min.

If the effect of ANP32 on the ability of the RdRP to extend an ApG primer was assessed, reaction mixtures containing 0.05 μ M [α -32P] GTP (3,000 Ci/mmol, Perkin-Elmer), 1 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 0.25 mM ApG (IBA Lifesciences), 5 mM MgCl₂, 1 mM DTT, 2 U/ μ l RNasin, 0.5 μ M 5' RNA promoter, 0.5 μ M 3' RNA promoter, 1 μ g/ μ l RdRP (see Chapter 2.7.4) in elution buffer and 100 ng/ μ l ANP32A or ANP32B were incubated at 30°C for 1 - 12 h, followed by heating to 95°C for 3 min.

Reactions were terminated by addition of an equal volume of 80% formamide, 1 mM EDTA, bromophenol blue and xylene cyan and incubation at 95°C for 3 min. Reaction

products were resolved by 20% denaturing PAGE containing 7 M urea in TBE buffer and visualised by autoradiography.

2.7.7 RNP activity assay

RNP activity was assessed as described previously in 20 µl reactions containing 10 µl purified RNPs and 5 µl purified RdRP (York et al., 2013). Reactions were incubated for 16 h at 30°C with 1 mM ATP, 0.5 mM of CTP, 0.5 mM GTP, 0.5 mM UTP, 0.25 mM ApG, 5 mM MgCl₂, 2 mM DTT and 2 U/µl RNasin. Purified RNPs were a kind gift from Itziar Serna Martin (Sir William Dunn School of Pathology, University of Oxford). RNA reaction products were extracted using TRI Reagent and precipitated in the presence of 20 µg of tRNA from *E. coli* MRE 600 (Roche). Extracted RNA was analysed by primer extension.

2.8 Viral RNA binding assays

2.8.1 Co-purification of viral RNA with the viral RdRP

Viral RNA association with recombinant RdRP was analysed through co-immunoprecipitation of viral RdRP and viral RNA. Recombinant inactive RdRP expressing the active site mutant PB1a (PB1-D445A/D446A; (Vreede et al., 2004)) was transiently co-expressed with short viral RNA (pPOLI-vNA47 or pPOLI-cNA47) and purified as described in section 2.6.3. RNA was extracted from purified protein samples using TRI Reagent in the presence of 20 µg of tRNA from *E. coli* MRE 600 (Roche). Extracted RNA was analysed for the presence of viral RNA using primer extension.

2.8.2 smFRET promoter binding assay

Promoter binding of purified recombinant RdRP was analysed using a single molecule fluorescence resonance energy transfer (smFRET) assay as previously described (Robb et al., 2016; Velthuis et al., 2016). A Cy3B donor dye on U17 of the 3' strand and an Atto647N acceptor dye on U6 of the 5' strand of the vRNA promoter or a Cy3B donor dye on U1 of the 3' strand and an Atto647N acceptor dye on U5 of the 5' strand of the cRNA promoter (see Table 2.2 for promoter sequences) was used to minimise interference of the dyes with RdRP-promoter complex formation, while keeping the distance between dyes after promoter binding at ~ 10 Å. Approximately 100 nM RdRP was pre-incubated with 1 nM double-labelled promoter RNA in buffer A (50 mM Tris-HCl [pH 8.0], 5% glycerol, 500 mM NaCl, 10 mM MgCl₂ and 100 µg/ml bovine serum album) for 15 min at 28°C. Samples were then diluted 10-fold in buffer A. Measurements were taken on a custom-built confocal microscope that allows alternating-laser excitation (ALEX) of donor and acceptor dyes. Average excitation intensities were 250 µW at 532 nm and 60 µW at 635 nm. Signals were detected using a custom LabView code and analysed using a custom Matlab code as previously described (Tomescu et al., 2014). The E^* values were plotted as one-dimensional distributions and fitted with a single (vRNA) or double (cRNA) Gaussian curve to obtain the mean E^* and the associated standard deviations.

2.9 RNA sequencing

2.9.1 3' UTR mRNA sequencing

2.9.1.1 Sample preparation and sequencing

A confluent monolayer of 9.5×10^5 human A549 cells was infected with ultracentrifuged purified influenza A/WSN/33 (H1N1) virus at an MOI of 5 in low serum DMEM (0.5% FCS) for 1 h at room temperature. Infected cells were incubated for 3 h, 6 h, 9 h, 12 h or 18 h at 37°C. At the indicated time points, cell monolayers were washed $2 \times$ in ice cold PBS. Total RNA was extracted using TRI Reagent and Direct-zol MiniPrep kit (Zymo Research). RNA samples were treated with RQ1 RNase-free DNase (Promega). RNA quality was assessed on a NanoDrop 2000 Spectrophotometer and on a denaturing 1% agarose gel.

cDNA libraries were made using the QuantSeq 3' mRNA-Seq Library Prep Kit for Ion Torrent (Lexogen) with 2 µg RNA as input. The quality and molarity of cDNA libraries were assessed using the High Sensitivity DNA Kit on a 2100 Bioanalyzer instrument (Agilent). cDNA libraries were diluted to 350 pM.

The Ion PI Hi-Q Chef system was used to prepare samples and load multiplexed libraries onto a Ion PI Chip v3. Sequencing on an Ion Proton System was performed using the Ion PI Hi-Q Sequencing 200 Kit (Thermo Scientific).

2.9.1.2 RNA sequence analysis

RNA sequence analysis was performed with the help of Michael Tellier (Sir William Dunn School of Pathology, University of Oxford). Raw Ion Torrent Bam files were converted into Fastq files using the SamToFastq command from PicardTools. No adaptor contamination was detected with Fastqc, therefore reads were directly mapped to the human genome (hg19) as described previously using the bowtie2 command with the -M 8 --local option

(Neve et al., 2016). This mode allows the ends of the sequence not to be part of the alignment (soft clipped). This is useful for reads containing sequence corresponding to the poly(A) tail which are not present in the genomic sequence. Samtools was used to convert the files into Bam format, keeping only properly mapped reads and sorting the Bam files. The number of mapped reads per chromosome was returned and normalised into reads per millions (RPM) using Idxstats.

```
java -jar picard.jar SamToFastq I=File01.bam FASTQ=File01.fastq  
  
bowtie2 -M 8 -- local -x /HumanGenome -U File-1.fastq -S File01.sam  
  
samtools view -b -F 4 -o File01.bam File01.sam  
  
samtools sort -o File01.sorted.bam File01.bam  
  
samtools index File01.sorted.bam  
  
samtools idxstats File01.sorted.bam
```

To map the sequenced reads to the influenza virus genome, a Bowtie2 index was created with all the influenza virus genome segments. The reads were then mapped to the viral genome and analysed as described above. To determine the relative abundance of viral transcripts, the number of viral reads was divided by the number of total mapped reads (human + viral). To determine the number of reads mapped to each viral genome segment, Bedtools was used to calculate the coverage of the final 3' nucleotide per read in each segment, of which the sum equals the number of reads aligned to each viral genome segment.

```
bowtie2-build IAV.fa IAV  
  
bowtie2 -M 8 -- local -x /IAVgenome -U File-1.fastq -S File01_IAV.sam  
  
samtools view -b -F 4 -o File01_IAV.bam File01_IAV.sam
```

```
samtools sort -o File01_IAV.sorted.bam File01_IAV.bam

samtools index File01_IAV.sorted.bam

samtools idxstats File01_IAV.sorted.bam

bedtools genomecov -ibam File01.IAV.sorted.bam -d -3 > File01_IAV.bedGraph
```

For hierarchical clustering the Morpheus software (<https://software.broadinstitute.org/morpheus/>) was used with the Euclidean distance as Metric and Rows as Cluster settings applied. Heatmaps were generated in R using the Bioconductor package biocLite.

```
source("http://bioconductor.org/biocLite.R")

biocLite(c("RColorBrewer", "pheatmap"))

library("pheatmap")

library("RColorBrewer")

setwd("path to your directory")

basedir <- getwd()

drows1 <- "euclidean"

filename <- "Heatmap Genes.pdf"

outfile <- paste(basedir, filename, sep="/")

col.pal <- brewer.pal(9, "RdBu")

GeneList = as.matrix(read.table("GeneList_Exp.txt", header=TRUE))

hm.parameters <- list(GeneList, color=col.pal, cellwidth=NA, cellheight=NA,
scale="none", treeheight_row=50, kmeans_k=NA, show_rownames = T,
show_colnames = T, main = "Heatmap Genes", clustering_method = "average",
cluster_rows = TRUE, cluster_cols = FALSE, clustering_distance_rows =
drows1)

do.call("pheatmap", c(hm.parameters, filename=outfile))
```

Analysis of polyadenylation site usage was performed using the 3' READS protocol as has been described previously (Hoque et al., 2013; Neve et al., 2016) with the help of Alastair Louey (Department of Biochemistry, University of Oxford).

For GO enrichment analysis the GO Consortium enrichment analysis software was used (<http://www.geneontology.org/>).

2.9.2 Whole transcriptome sequencing

A confluent monolayer of 9.5×10^5 human A549 cells was infected with ultracentrifuged purified influenza A/WSN/33 (H1N1) virus at an MOI of 5 in low serum DMEM (0.5% FCS) for 1 h at room temperature. Infected cells were incubated for 6 h or 9 h at 37°C. At the indicated time points, cell monolayers were washed 2 × in ice cold PBS. Total RNA was extracted using TRI Reagent and Direct-zol MiniPrep kit (Zymo Research). RNA samples were treated with RQ1 RNase-free DNase (Promega). RNA quality was assessed on a NanoDrop 2000 Spectrophotometer. 5 µg RNA was used to deplete cytoplasmic (5S, 5.8S, 18S, 28S) and mitochondrial (12S, 16S) ribosomal RNA (rRNA) using the RiboMinus Eukaryote System v2 (Ambion). The quality of rRNA-depleted RNA was assessed using the RNA 6000 Nano Kit on a 2100 Bioanalyzer instrument (Agilent). cDNA libraries were made with 25 ng rRNA-depleted RNA and 252 pg SIRV Mix E0 (Lexogen) as input, using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB) and the NEBNext Multiplex Oligos for Illumina Set 1 (NEB). The quality and molarity of cDNA libraries were assessed using the High Sensitivity DNA Kit on a 2100 Bioanalyzer instrument (Agilent). cDNA libraries were pooled at a 2 nM concentration per library.

Samples were sequenced using single-end runs on a NextSeq 500 System using a NextSeq 500/550 High Output v2 Kit (150 cycles) (Illumina).

2.10 RBPome capture

2.10.1 Label-free RBPome capture

RBPome capture in A549 cells was adapted from previously established protocols (Castello et al., 2013). 7.6×10^7 human A549 cells were infected with ultracentrifuged purified influenza A/WSN/33 (H1N1) virus at MOI = 5 in low serum (0.5%) DMEM for 1 h at room temperature, before incubation at 37°C and 5% CO₂. 6 h or 9 h post-infection, cells were washed 2 × in PBS and all liquid was aspirated. Cells were irradiated with UV light at 254 nm (0.15 J/cm²) on ice. If cells were not infected, 7.6×10^7 human A549 cells were washed 2 × in PBS and all liquid was aspirated. Cells were irradiated with UV light at 254 nm (0.15 J/cm²) on ice.

Cells were lysed in 1 ml lysis buffer (20 mM Tris-HCl [pH 7.5], 0.5 M LiCl, 0.5 % Lithium dodecyl sulphate (LiDS), 1 mM EDTA, 5 mM DTT) per dish (5 ml per condition) and incubated for 10 min on ice. Lysates were harvested by scraping, pooled with corresponding lysates and homogenised by syringing through a narrow gauge needle (0.4 mm diameter). 0.5 ml input whole cell lysate was kept as input for protein analysis and 50 µl for RNA analysis.

Poly(A)⁺ mRNA was captured using 1 ml washed Oligo d(T)₂₅ Magnetic Beads (NEB) at 4°C for 1 h. Oligo d(T)₂₅ beads were captured with a magnet and cleared cell lysates were collected and kept on ice. Oligo d(T)₂₅ beads were washed 1 × in 10 ml lysis buffer, incubated for 5 min on ice, captured and the cleared buffer was aspirated. Oligo d(T)₂₅

beads were washed 2 × in 10 ml buffer 1 (20 mM Tris-HCl [pH 7.5], 0.5 M LiCl, 0.1 % LiDS, 1 mM EDTA, 5 mM DTT, 0.025 % Igepal CA-630), incubated for 5 min on ice, captured and the cleared buffer was aspirated. Oligo d(T)25 beads were washed 2 × in 10 ml buffer 2 (20 mM Tris-HCl [pH 7.5], 0.5 M LiCl, 1 mM EDTA, 5 mM DTT, 0.025 % Igepal CA-630), incubated for 5 min on ice, captured and the cleared buffer was aspirated. Oligo d(T)25 beads were washed 2 × in 10 ml buffer 3 (20 mM Tris-HCl [pH 7.5], 0.2 M LiCl, 1 mM EDTA, 5 mM DTT), captured and the cleared buffer was aspirated. Oligo d(T)25 beads were eluted in 0.5 ml elution buffer (20 mM Tris-HCl [pH 7.5], 1 mM EDTA) for 3 min at 55°C, captured and eluted samples were collected. Oligo d(T)25 beads were recycled in 1 ml 0.1 M NaOH for 5 min at 55°C, captured and washed 3 × in 10 ml lysis buffer and added back onto cell lysates. Capture of poly(A)⁺ mRNA was repeated twice. RNA concentrations were measured on a NanoDrop 2000 Spectrophotometer and eluates with more than 50 ng/μl RNA were pooled.

Protein contents were digested in 10 μl eluate or 50 μl whole cell lysate using 10 μg proteinase K (Roche) and incubated for 2 h at 50°C in Proteinase K buffer (10 mM Tris-HCl [pH 7.5], 150 mM NaCl, 0.2 % SDS, 10 mM EDTA, 0.5 mM DTT, 5 mM CaCl₂). RNA was extracted using TRI Reagent and a Direct-zol MiniPrep kit (Zymo Research) and measured on a NanoDrop 2000 Spectrophotometer.

From the remaining whole cell lysates and eluates RNA content was digested for 90 min at 37°C with RNase A (Sigma Aldrich) and RNase T1 (Thermo Scientific), followed by heat inactivation at 55°C for 15 min. Samples were analysed by mass spectrometry and SDS-PAGE. For analysis by SDS-PAGE, samples were concentrated using Amicon Ultra Centrifugal Filters (0.5 ml, 3 kDa MWCO) (Millipore) and a buffer exchange with buffer 4 (10 mM Tris-HCl [pH 7.5], 50 mM NaCl) was performed.

2.10.2 SILAC-labelled RBPome-capture

Approximately 7.6×10^7 human SILAC-labelled A549 cells were infected with ultracentrifuged purified influenza A/WSN/33 (H1N1) virus at MOI = 5 in low serum (0.5%) DMEM for SILAC for 1 h at room temperature, before incubation at 37°C and 5% CO₂. 6 h or 9 h post-infection, cells were washed 2 × in PBS and all liquid was aspirated. Cells were irradiated with UV light at 254 nm (0.15 J/cm²) on ice. Cells were lysed in 1 ml lysis buffer (20 mM Tris-HCl [pH 7.5], 0.5 M LiCl, 0.5 % LiDS, 1 mM EDTA, 5 mM DTT) per dish (5 ml per condition) and incubated for 10 min on ice. Lysates were harvested by scraping, pooled with corresponding lysates and homogenised by syringing through a narrow gauge needle (0.4 mm diameter). Samples were frozen at -80°C. Three replicates were generated with each condition (mock-infected, 6 h infection, 9 h infection) used once in each SILAC cell line ('light', 'medium', 'heavy').

Protein concentrations were measured using Pierce 660 nm Protein Assay (Thermo Scientific) with Ionic Detergent Compatibility Reagent for Pierce 660 nm Protein Assay Reagent (Thermo Scientific) and protein contents were normalised in samples within each replicate. A 0.5 ml aliquot of whole cell lysate was analysed by mass spectrometry.

Samples from same replicate were mixed and poly(A)⁺ mRNA was captured using 1 ml washed Oligo d(T)₂₅ Magnetic Beads (NEB) at 4°C for 1 h. Oligo d(T)₂₅ beads were captured with a magnet and cleared cell lysates were collected and kept on ice. Oligo d(T)₂₅ beads were washed 1 × in 10 ml lysis buffer, incubated for 5 min on ice, captured and the cleared buffer was aspirated. Oligo d(T)₂₅ beads were washed 2 × in 10 ml buffer 1 (20 mM Tris-HCl [pH 7.5], 0.5 M LiCl, 0.1 % LiDS, 1 mM EDTA, 5 mM DTT, 0.025 % Igepal CA-630), incubated for 5 min on ice, captured and the cleared buffer was aspirated. Oligo d(T)₂₅ beads were washed 2 × in 10 ml buffer 2 (20 mM Tris-HCl [pH 7.5], 0.5 M LiCl, 1 mM EDTA, 5 mM DTT, 0.025 % Igepal CA-630), incubated for 5 min on ice, captured and the cleared buffer was aspirated. Oligo d(T)₂₅ beads were washed 2 × in 10

ml buffer 3 (20 mM Tris-HCl [pH 7.5], 0.2 M LiCl, 1 mM EDTA, 5 mM DTT), captured and the cleared buffer was aspirated. Oligo d(T)25 beads were eluted in 0.5 ml elution buffer (20 mM Tris-HCl [pH 7.5], 1 mM EDTA) for 3 min at 55°C, captured and eluted samples were collected. Oligo d(T)25 beads were recycled in 1 ml 0.1 M NaOH for 5 min at 55°C, captured and washed 3 × in 10 ml lysis buffer and added back onto cell lysates. Capture of poly(A)⁺ mRNA was repeated twice. RNA concentrations were measured on a NanoDrop 2000 Spectrophotometer and eluates with more than 50 ng/μl RNA were pooled.

RNA was digested for 90 min at 37°C from eluted poly(A)⁺ mRNA-protein complexes with RNase A (Sigma Aldrich) and RNase T1 (Thermo Scientific), followed by heat inactivation at 55°C for 15 min.

2.11 Mass spectrometry

2.11.1 Filter Aided Sample Preparation (FASP)

Samples were submitted to the Advanced Proteomics Facility (University of Oxford) for tryptic digestion and mass spectrometry analysis. For dimethyl label-dependent analysis of the RBPome of uninfected A549 cells, three biologically independent replicates were submitted. For SILAC-labelled analysis of the RBPome of infected A549 cells, three biologically independent replicates were submitted. For dimethyl label-dependent analysis of ANP32A interactomes two replicates were submitted.

Filter Aided Sample Preparation (FASP) was used for sample preparation and digestion. Vivacon 500 10 kDA filters (Sartorius) were washed by centrifugation (15,000 × g, 15 min) in 0.1% trifluoroacetic acid (TFA) in 50% acetonitrile. Samples were loaded onto the filters

and denatured with 8 M urea in 100 mM triethylammonium bicarbonate buffer (TEAB) for 30 min at room temperature. Samples were reduced with 10 mM tris(2-carboxyethyl)phosphine (TCEP) for 30 min at room temperature. Samples were alkylated with 50 mM C-IAA for 30 min at room temperature in the dark. Filters were washed twice in 6 M urea in 50 mM TEAB (pH 8). Lys-C protease in 6 M urea and 50 mM TEAB was added at a 1:40 ratio to the sample (w/w) and incubated at 37°C overnight. Trypsin in 50 mM TEAB was added at a 1:40 ratio to the sample (w/w) and incubated at 37°C for 4 h. Samples were centrifuged (15,000 × g, 15 min) and filtrated samples were collected. Filters were washed with 0.1% TFA, followed by a wash with 50% ACN in 0.1% TFA. Acidified samples were dried down in a SpeedVac (Thermo Fisher Scientific).

2.11.2 Mass spectrometry

Peptides were resuspended in 5 % formic acid and 5 % DMSO. They were separated on a Ultimate 3000 UHPLC system (Thermo Fisher Scientific) and electrosprayed directly into an QExactive mass spectrometer (Thermo Fisher Scientific) through an EASY-Spray nano-electrospray ion source (Thermo Fisher Scientific). The peptides were trapped on a C18 PepMap100 pre-column (300 µm i.d. × 5 mm, 100 Å, Thermo Fisher Scientific) using solvent A (0.1% formic acid in water) at a pressure of 500 bar. The peptides were separated on an in-house packed analytical column (75 µm i.d. packed with ReproSil-Pur 120 C18-AQ, 1.9 µm, 120 Å, Dr. Maisch HPLC GmbH) using a linear gradient (120-180 min for RBPomes/60 min for ANP32A interactomes, 15% to 35% solvent B [0.1% formic acid in acetonitrile], flowrate of 200 nl/min).

The raw data was acquired on the mass spectrometer in a data-dependent mode (DDA). Full scan MS spectra were acquired in the Orbitrap (scan range 350-1500 m/z, resolution 70000, AGC target value of 3×10⁶, maximum injection time 50 ms). after the MS scans the

10 most intense peaks were selected for HCD fragmentation at 30% of normalised collision energy. HCD spectra were also acquired in the Orbitrap (resolution 17500, AGC target value of 5×10^4 , maximum injection time 120 ms) with first fixed mass at 180 m/z. Charge exclusion was selected for unassigned and 1+ ions. The dynamic exclusion was set to 20 s.

The raw data files generated were processed using MaxQuant (Version 1.5.0.35), integrated with the Andromeda search engine as previously described (Cox & Mann, 2008; Cox et al., 2011). For protein group identification, peak lists were searched against the human database (UPR_HomoeSapien_9606_160113.fasta), the currently known influenza A/WSN/33 (H1N1) data base (Appendix A) as well as a list of common contaminants by Andromeda. A maximum of two missed cleavage sites in trypsin digestion was selected. Acetylation (Protein N-term), Oxidation (M) were used as variable modifications while Carbamidomethylation (C) was set as fixed modification. Protein and peptide-spectrum match (PSM) false discovery rate (FDR) were set as 0.01 and “Match-between-Runs” (for dimethyl label-dependent quantitation) or “Quantitation” (SILAC quantitation) was applied. For SILAC quantitation medium label settings (R[6], K[4]) and heavy label settings (R[10], K[8]) were applied.

2.11.3 Analysis of mass spectrometry data

Analysis of the processed raw data files for the RBPome data sets was performed by Ni Shuai and Bernd Fischer (DKFZ, Germany) using the R scripts attached in Appendices B and C.

Analysis of ANP32A datasets was performed using Perseus (Version 1.5.5.5). Before further processing, common contaminants, keratins, serum albumin and decoy sequences

were removed from the data sets. Proteins that were not present in both biological replicates were discarded from further analysis. Protein abundance was normalised by using the mean for each protein of the ratio of protein intensity in the ANP32A sample compared to the protein intensity in the PP7CP sample. Proteins that were not at least 10-fold enriched in the huANP32A or chANP32A samples compared to the PP7CP control sample were excluded from further analysis. When comparing uninfected to infected samples, missing values for proteins that were enriched in either the uninfected or infected samples were replaced with 1 in order to calculate a fold change.

Gene Ontology analysis was performed using the GO Consortium enrichment analysis software (<http://www.geneontology.org/>) and the PANTHER classification system (Mi et al., 2013).

2.12 Microscopy

2.12.1 Immunofluorescence confocal microscopy

To analyse the subcellular localisation of wild-type or mutant PB2, HEK-293T cells were grown on sterilised 13 mm glass coverslips in 24-well plates. Cells were transiently transfected with 0.25 µg of RdRP subunit expression plasmids (pcDNA-PA, pcDNA-PB1 and pcDNA-PB2 or pcDNA-PB2-K627E or pcDNA-PB2Δ535-667) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. After 48 h of incubation at 37°C cells were fixed for 15 min at room temperature with 4% paraformaldehyde in 250 mM HEPES (pH 7.9) prior to three washes with PBS. Cells were then permeabilised for 15 min at room temperature in 0.25% Triton X-100 in PBS and washed three times with PBS. Cells were blocked in blocking PBS (10% normal goat serum, 0.5% Triton X-100 and 3%

bovine serum album) overnight at 4°C. Cells were stained using a rabbit polyclonal anti-PB2 antibody raised against the N-terminal 180 amino acids of PB2 at a 1:100 dilution in blocking PBS for 1 h at room temperature. Cells were washed three times with blocking PBS. Cells were stained using a goat anti-Rabbit IgG antibody conjugated to an Alexa532 dye at a dilution of 1:500 in blocking PBS for 1 h at room temperature. Cells were twice washed with blocking PBS and twice washed with PBS prior to mounting in Mowiol (Calbiochem) containing 1 µg/ml 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich). Images were obtained on a Olympus Fluoview FV1000 confocal laser scanning microscope. Image analysis was performed using ImageJ software.

2.12.2 Immunofluorescence microscopy

To analyse the infection efficiency of influenza A/WSN/33 (H1N1) virus in A549 cells, A549 cells were grown on sterilised 13 mm glass coverslips in 24-well plates. Cells were washed with PBS prior to infection with influenza A/WSN/33 at an MOI of 5. At 0, 3, 6, 9, 12 and 18 h post-infection cells were twice washed with PBS and fixed for 15 min at room temperature with 4% paraformaldehyde in 250 mM HEPES (pH 7.9) prior to three washes with PBS. Cells were then permeabilised for 15 min at room temperature in 0.25% Triton X-100 in PBS and washed three times with PBS. Cells were blocked in blocking PBS (10% normal goat serum, 0.5% Triton X-100 and 3% bovine serum album) overnight at 4°C. Cells were stained using a rabbit polyclonal anti-NP antibody at a 1:500 dilution in blocking PBS for 1 h at room temperature. Cells were washed three times with blocking PBS. Cells were stained using a goat anti-Rabbit IgG antibody conjugated to an Alexa532 dye at a dilution of 1:500 in blocking PBS for 1 h at room temperature. Cells were twice washed with blocking PBS and twice washed with PBS prior to mounting in Mowiol

(Calbiochem) containing 1 $\mu\text{g/ml}$ 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich). Images were obtained on a EVOS FLoid Cell Imaging Station (Life Technologies). Image analysis was performed using ImageJ software.

Chapter 3: The role of the PB2 627-domain in polymerase function

3.1 Introduction

Crystal structures of the influenza virus RdRP have shown that the enzyme consists of a central polymerase domain and several flexible peripheral appendices (Hengrung et al., 2015; Pflug et al., 2014; Reich et al., 2014; Thierry et al., 2016). The central domain is made up of PB1, the C-terminal domain of PA and the N-terminal domains of PB2 (N-terminus, lid, N1 and N2 linker domains). The peripheral appendices are formed by the N-terminal PA endonuclease domain (Dias et al., 2009; Yuan et al., 2009) and the C-terminal domains of PB2, including the cap-binding, mid-link, 627- and NLS domains (Guilligay et al., 2008; Tarendeau et al., 2008; Yamada et al., 2010). Two of these three major protruding peripheral domains, the cap-binding and endonuclease domains, are involved in cap-snatching by binding to and cleaving host capped RNA respectively, while the role of the so-called 627-domain in RdRP function remains unclear. Recently published crystal structures of influenza virus RNA polymerases have revealed that binding of the conserved vRNA or cRNA terminal ends can trigger significant conformational rearrangements in the peripheral domains (Hengrung et al., 2015; Pflug et al., 2014; Reich et al., 2014; Thierry et al., 2016). The conformation that the RdRP adopts following vRNA binding, the open form, is compatible with transcription, while the more closed form is not. However, the closed form could be consistent with replication.

Avian influenza virus RdRPs are severely restricted in their activity in mammalian hosts, resulting in impaired viral genome replication and mRNA synthesis and, thus, a reduced potential for acquiring beneficial mutations (Massin et al., 2001). For instance, avian influenza A virus RdRPs almost universally contain a glutamic acid (E) residue at position 627 of PB2, whereas this residue is frequently mutated to a lysine (K) in mammalian-adapted RdRPs (Subbarao et al., 1993). It has been shown that introducing the E627K mutation restores activity of an avian RdRP in mammalian cells (Cauldwell et al., 2013; Mehle & Doudna, 2008; Paterson et al., 2014). PB2-627 is located on the PB2 627-domain, which folds into a structurally distinct domain protruding from the RdRP core. In the conformation that the RdRP adopts following binding of vRNA, PB2-627 is surface-exposed and located near the presumed nascent RNA product exit channel (Pflug et al., 2014; Tarendeau et al., 2008; Yamada et al., 2010). In addition to E627K, several other adaptive mutations have been found to cluster on the surface or in close vicinity of the 627-domain and the nascent RNA product exit channel (Cauldwell et al., 2014; Gabriel & Fodor, 2014).

Although adaptive mutations have been demonstrated to enhance the activity of avian RdRPs in mammalian cells, there is disparity in the literature regarding the mechanism through which these mutations enhance RdRP activity. For instance, it has been proposed that PB2-627E could impair vRNP assembly in mammalian cells by destabilising RdRP interactions with NP (Labadie et al., 2007; Ng et al., 2012b; Rameix-Welti et al., 2009), but this view has been challenged in recent studies (Cauldwell et al., 2013; Paterson et al., 2014). PB2-627 has also been proposed to regulate viral promoter binding (Crescenzo-Chaigne et al., 2002; Paterson et al., 2014), interactions between the viral RdRP and importin- α (Gabriel et al., 2011; Hudjetz & Gabriel, 2012), the interaction between the viral RdRP and a cellular inhibitor of virus infections in human cells (Mehle & Doudna, 2008) or the interaction between the viral RdRP and an activating host factor (Moncorge et al.,

2010). However, no molecular mechanism has yet been presented that can explain how PB2-627 or in fact the entire 627-domain determines influenza A virus host range.

In this study, the aim is to characterise the role of the PB2 627-domain and other PB2 domains of the influenza A virus RdRP using a combination of *in vitro* RdRP activity assays and cell-based RNP reconstitution assays to gain a better understanding of the molecular mechanism that determines host range.

3.2 Results

3.2.1 Assembly of the heterotrimeric viral RdRP

In order to address the function of the flexible C-terminal PB2 domains and to identify the minimal PB2 region required for core RdRP functions, plasmids encoding PB2 genes derived from influenza A/WSN/33 (H1N1) with systematic internal or C-terminal truncations were generated using site-directed PCR mutagenesis (Fig. 3.1). The truncations were designed in flexible regions to minimise the structural impact on protein folding.

First, the ability of these PB2 mutants to form heterotrimeric RdRP complexes with PB1 and PA was determined by co-expression of PB2 mutants with PB1 and PA fused to a tandem affinity purification tag (PA-TAP) in human HEK-239T cells. An avian-like RdRP with a lysine to glutamic acid substitution at PB2-627 (K627E) was included in all studies as a control. Purification of these recombinant RdRPs from cell lysates using the protein A tag on PA and analysis of purified protein complexes by SDS-PAGE and silver staining showed that PA and PB1 form a stable dimer in the absence of PB2 or the presence of mutant PB2 as was reported previously (Fodor & Smith, 2004; Swale et al., 2016). Western

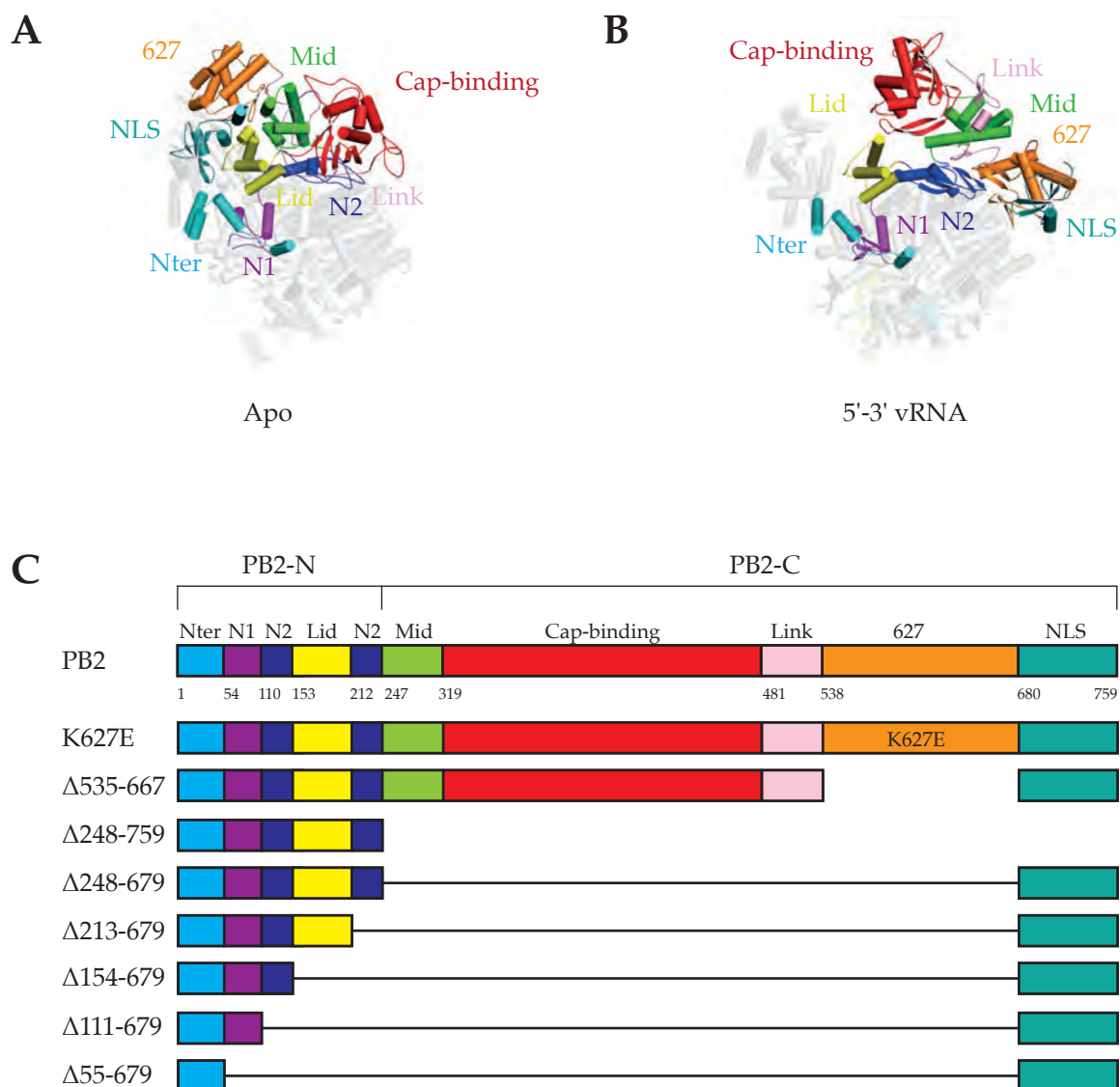


Fig. 3.1 Design of PB2 deletion mutants. (A) Crystal structure of the (apo) influenza C/Johannesburg/1/66 virus polymerase (PDB code 5D98). (B) Crystal structure of the (vRNA promoter bound) influenza A/Little yellow-shouldered bat/Guatemala/060/10 (H17N10) virus RdRP (PDB code 4WSB). PB2 domains are shown in different colours. (C) Schematic of PB2 and truncated PB2 mutants with PB2 domains coloured as in panel A and B.

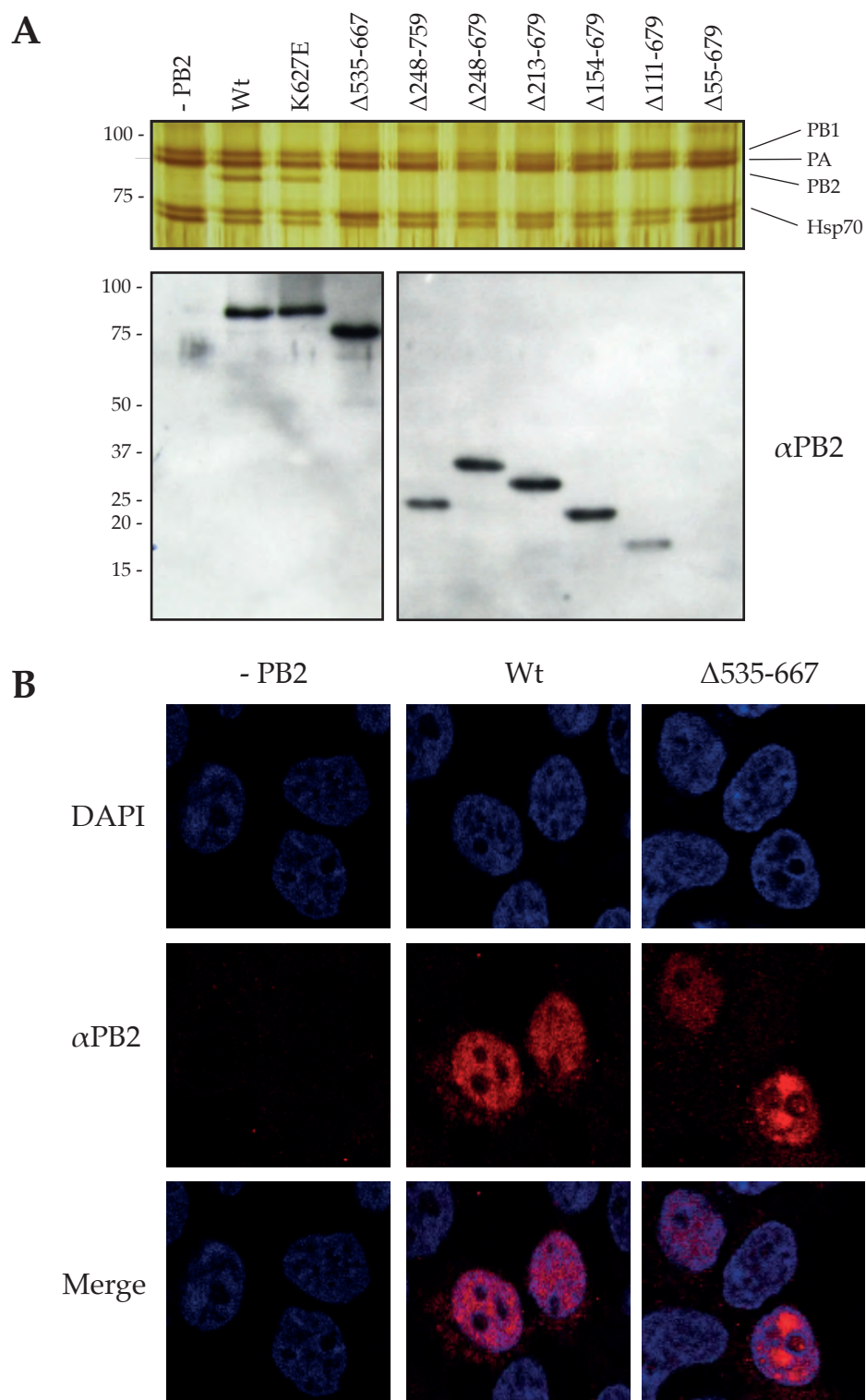


Fig. 3.2 Expression and subcellular localisation of PB2 deletion mutants. (A) Heterotrimeric polymerase complex formation of PB2 mutants with PB1 and PA. Polymerase subunits PB1 and PA were co-expressed with PB2 in HEK-293T cells and purified by IgG Sepharose chromatography. Purified recombinant polymerase was analysed by SDS-PAGE. PA, PB1 and full-length PB2 were detected by silver staining. Co-purification of PB2 with the PA-PB1 dimer was detected by Western blotting with a primary antibody targeting the N-terminal region (1-180) of PB2 (Carr et al., 2006). Molecular masses, indicated on the left, are depicted in kilodaltons. (B) Polymerase subunits PA, PB1 and PB2 were co-expressed in HEK-293T cells and PB2 was detected by immunofluorescence using a primary antibody targeting the N-terminal region (1-180) of PB2.

blotting with an antibody raised against the N-terminal 180 amino acids of PB2 revealed that all PB2 mutants assembled into heterotrimeric complexes with PB1 and PA, with the exception of PB2 Δ 55-679 (Fig. 3.2A). It is unclear however, if the lack of signal for PB2 Δ 55-679 was due to a defect in assembly or because the antibody was ineffective against the small N-terminal fragment of PB2. Interestingly, equal amounts of PB2 Δ 248-679 and PB2 Δ 248-759 were co-purified, even though PB2 assembly with the PA-PB1 dimer is supposed to occur in the nucleus and the NLS of PB2 (amino acids 680-759) is required for nuclear import of PB2. Therefore, PB2 Δ 248-759 and PB2 Δ 55-679 were excluded from further downstream analyses.

It is believed that the RdRP complex formation takes place in the nucleus after separate nuclear imports of the PB1-PA dimer and the PB2 monomer (Deng et al., 2005; 2006a; Fodor & Smith, 2004; Huet et al., 2010). To resolve whether truncation of the 627-domain affects subcellular localisation of PB2, PB2 was co-expressed with PA and PB2 in HEK-293T cells and its localisation was determined using immunofluorescence. Confocal microscopy revealed a predominantly nuclear localisation for the wild-type PB2 subunit and the PB2 subunit lacking the 627-domain (Fig. 3.2B).

Together, these results show that the PB2 N2, lid, cap-binding, mid-link, 627- and NLS domains are not required for stable heterotrimeric RdRP assembly with PA and PB1 and that the 627-domain does not affect nuclear accumulation of PB2.

3.2.2 The PB2 627-domain does not affect binding of viral RNA

Viral promoter RNA has been shown to take up significantly different structural arrangements depending on whether it exists in an RdRP-unbound or -bound state (Tomescu et al., 2014; Pflug et al., 2014; Robb et al., 2016). Utilising this change in viral promoter structure, the ability of the RdRP lacking the 627-domain to bind to vRNA or

cRNA promoter structures was analysed using single-molecule fluorescence resonance energy transfer (smFRET). Annealed vRNA or cRNA promoter was fluorescently labelled with a donor dye on the 5' strand and with an acceptor dye on the 3' strand and incubated with purified recombinant RdRP. Single-molecule FRET was measured on complexes diffusing in solution and apparent FRET efficiency values (E^*) reporting on the donor-acceptor distance between fluorophores on the RNA were depicted as histograms and fitted with Gaussian functions to determine the mean FRET efficiency of the distributions (Hohlbein et al., 2013; Robb et al., 2013). In the absence of viral RdRP, both the vRNA and cRNA promoter regions form a partially double-stranded structure producing a single FRET population (Fig. 3.3A and 3.3B, left panels). The addition of wild-type RdRP, K627E RdRP or RdRP lacking the 627-domain to vRNA promoter resulted in a shifted FRET distribution, representing the changed distance between donor and acceptor dye due to binding of the vRNA promoter by the viral RdRP (Fig. 3.3A). No differences in the mean FRET efficiency values was observed between wild-type or mutant RdRPs. Addition of wild-type RdRP, K627E RdRP or RdRP lacking the 627-domain to cRNA promoter resulted in a bimodally shifted FRET distribution (Fig. 3.3B), representing the pre-initiation state and the initiation state of the RdRP as reported previously (Robb et al., 2016). Mean FRET efficiency values did not differ significantly, even though the variation was higher compared to the vRNA promoter due to a lower number of molecules tested. To determine whether the RdRP lacking the 627-domain is also able to bind a longer vRNA or cRNA template in the cell, recombinant RdRP containing an active-site mutation in PB1 (PB1a), which allows RNA binding but no transcription or replication, was co-expressed with a 47-nucleotide-long vRNA or cRNA template and purified via a protein A tag on PA. RNA co-purified with the RdRP was subsequently extracted and analysed by primer extension. Only low levels of co-purifying vRNA and cRNA were associated with a PA-PB1a dimer, which is in line with previous observations that PB2 is required for

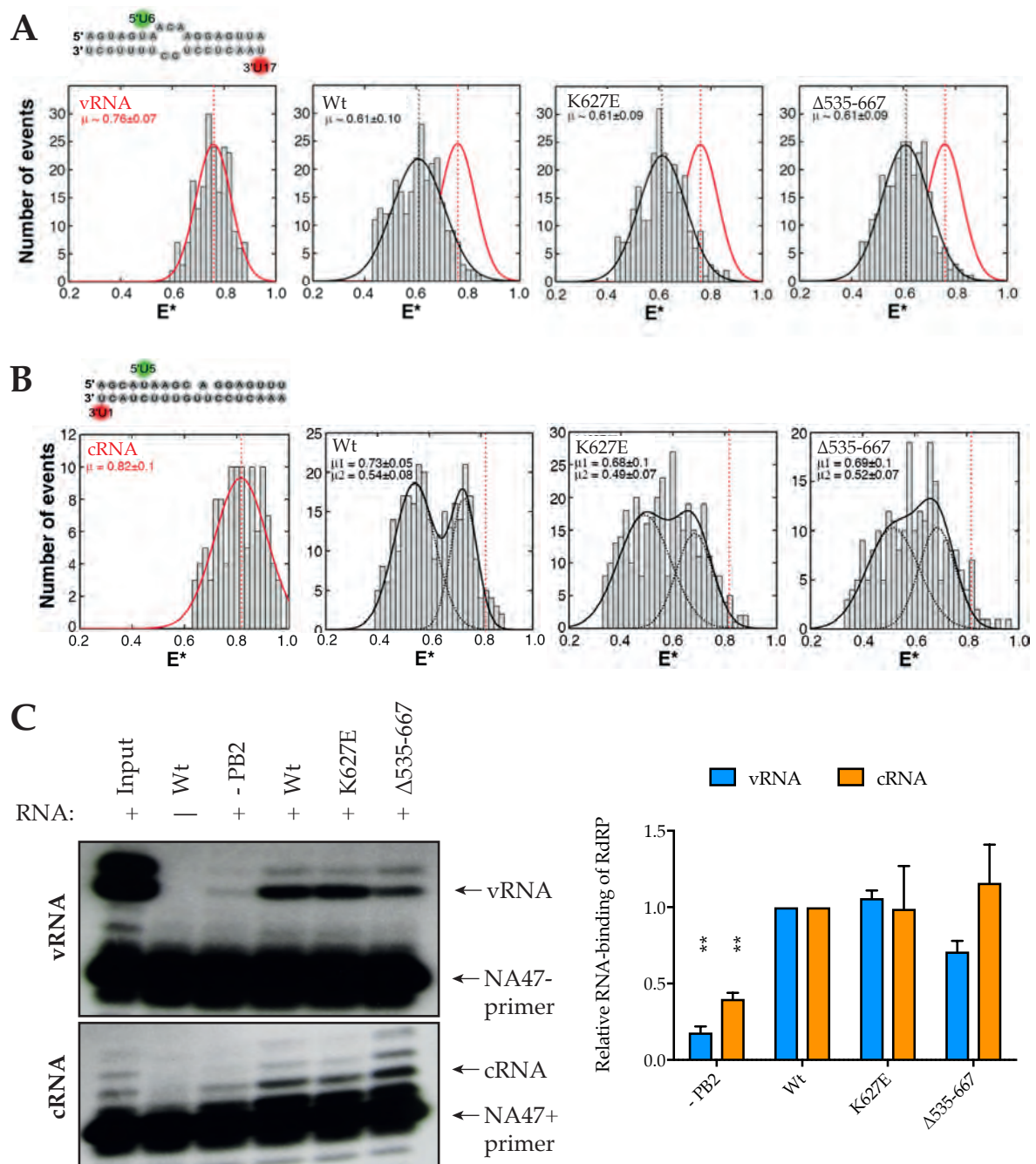


Fig. 3.3 The 627-domain is not required for binding of viral RNA. Annealed RNAs corresponding to the 5' and 3' ends of the (A) vRNA or (B) cRNA promoter were labelled with donor and acceptor fluorophores and were either analysed alone (left panel) or incubated with purified RdRP (middle and right panels) before single-molecule FRET spectroscopy on diffusing molecules was carried out. Ratio E^* represents the uncorrected FRET efficiency and curves were fitted with Gaussian functions to determine the centre of the distributions (μ). (C) RdRP subunits PA-TAP, PB1a and PB2 as well as segment 6-derived 47-nt-long vRNA or cRNA were co-expressed in HEK-293T cells and purified by IgG Sepharose chromatography. Levels of vRNA or cRNA that co-purified with RdRP were assessed by primer extension. The graph shows the mean intensity (subtracted by the mean intensity of a no RNA control) relative to wild-type RdRP from three independent biological replicates ($n = 3$). Error bars represent the standard error of the mean and the asterisks indicate a significant difference to wild-type (two-tailed one-sample t test) as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

efficient promoter binding (Swale et al., 2016). Co-purifying vRNA and cRNA levels were comparable between wild-type RdRP, K627E RdRP and RdRP lacking the 627-domain (Fig. 3.3B).

Together, this shows that the 627-domain does not interfere with viral promoter binding *in vitro* or in the cell.

3.2.3 Requirement of PB2 domains for RdRP *in vitro* activity

The RdRP needs to carry out several different actions in order to make mRNA, cRNA and vRNA. To address the question whether the PB2 N2, lid, cap-binding, mid-link and 627-domains are needed for different RdRP functions, *in vitro* activity assays were performed using purified recombinant RdRP.

First the ability of the mutant RdRP to carry out cap-snatching and capped RNA primer extension, which are required for viral RNA transcription, was evaluated. Purified RdRP was bound to vRNA promoter, comprising a 15-nucleotide-long RNA corresponding to the 5' end of vRNA and a 14-nucleotide-long RNA corresponding to the 3' end of vRNA, and incubated with a 20-nucleotide-long radiolabelled synthetic capped RNA in the absence (assessing cap-snatching) or presence (assessing cap-snatching followed by capped RNA primer extension) of nucleotides. In the absence of nucleotides, wild-type and K627E RdRP produced two major cleavage products (Fig. 3.4A). These major cleavage products were extended into two major transcription products in the presence of nucleotides (Fig. 3.4B). The RdRP lacking the 627-domain was also active in cap-snatching but produced only a single cleavage product and two extension products that were produced in unequal amounts. Furthermore, the extension products were one nucleotide shorter than the products made by the wild-type RdRP. This suggests that the 627-domain

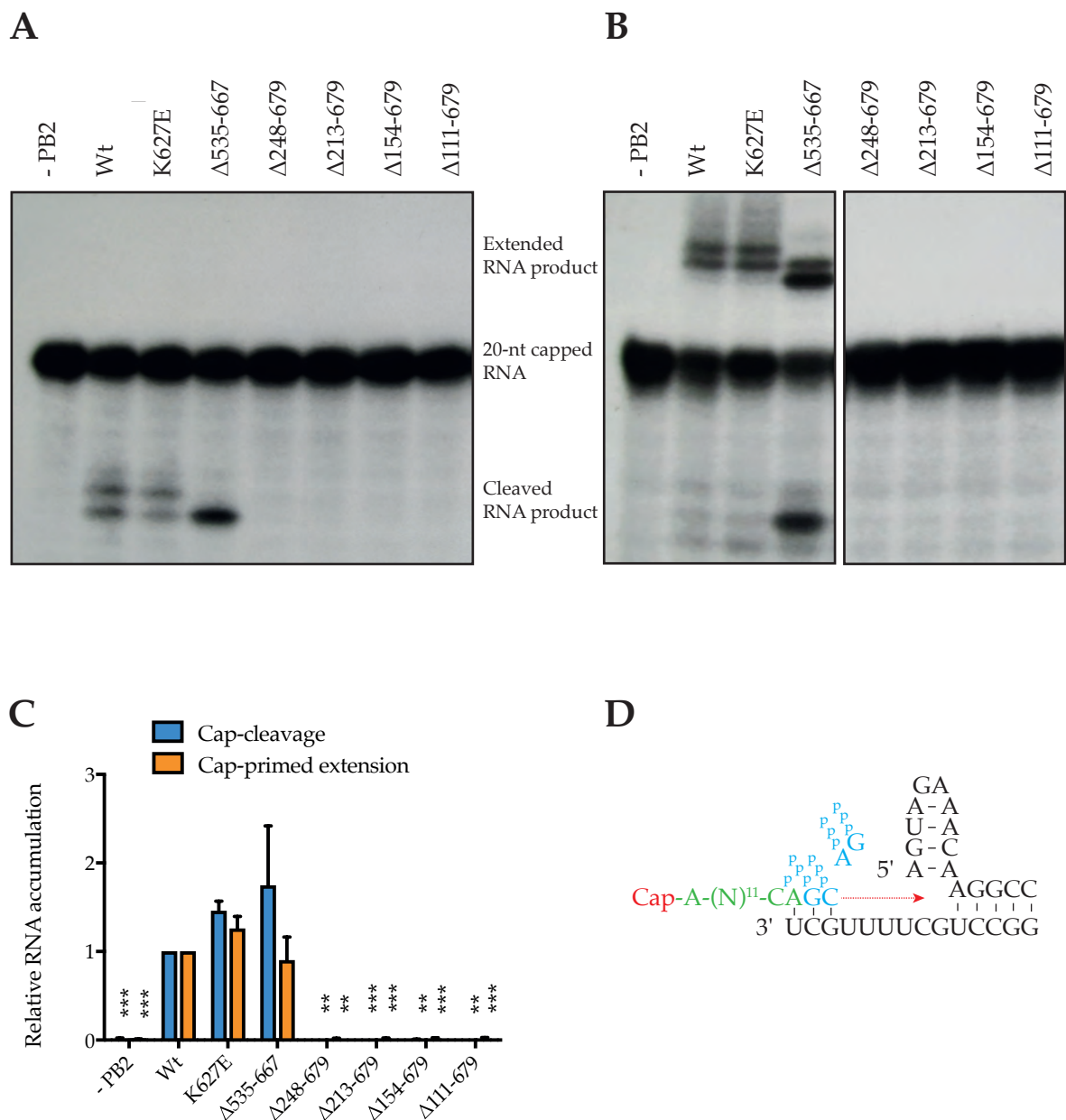


Fig. 3.4 Polymerase lacking the 627-domain is active in cap-dependent transcription *in vitro*. Purified recombinant RdRP was incubated in reaction mixtures containing radiolabelled capped RNA and vRNA template in the (A) absence or (B) presence of nucleotides. Reaction products were analysed by 20% PAGE and autoradiography. (C) The graph shows the mean intensity signal relative to wild-type RdRP from three independent biological replicates ($n = 3$), with error bars representing the standard error of the mean and the asteriks indicating a significant difference to wild-type (two-tailed one-sample t test) as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. (D) The schematic shows cap-dependent initiation on a vRNA promoter (adapted from te Velthuis et al., 2016). Unlabelled nucleotides are shown in blue and the capped RNA primer in green with the radiolabelled cap in red.

affects the relative positioning of the cap-binding and endonuclease domains of the RdRP and that deletion of the 627-domain results in cleavage and initiation taking place at alternative nucleotides of the capped RNA primer. Deletion of additional PB2 domains resulted in an RdRP inactive in cap-snatching, which is in line with the fact that these all lack the cap-binding domain.

Next, the ability of the RdRP to initiate viral replication *de novo* was analysed (Fig. 3.5). Purified RdRP was bound to vRNA promoter (see above) or cRNA promoter, comprising a 14-nucleotide-long RNA corresponding to the 5' end of cRNA and a 15-nucleotide-long RNA corresponding to the 3' end of cRNA, in the presence of ATP and GTP. Wild-type RdRP, K627E RdRP and RdRP lacking the 627-domain were all able to initiate replication *de novo*, producing a pppApG dinucleotide product on both the vRNA and cRNA promoter templates. The RdRP lacking the cap-binding, mid-link and 627-domains (PB2 Δ 248-679) was also able synthesise a pppApG dinucleotide on a vRNA promoter, but interestingly only had a reduced ability to do so on a cRNA promoter template. Further deletions into the nonflexible N-terminal domains of PB2 resulted in inhibition of pppApG synthesis on a vRNA template, while on a cRNA template dinucleotide synthesis was only slightly reduced in PB2 Δ 213-679 and not affected in PB2 Δ 154-679 and PB2 Δ 111-679.

Finally, the ability of the RdRP to replicate viral RNA, using the extension of an ApG dinucleotide primer on vRNA or cRNA templates as a readout, was analysed. Purified RdRP was bound to vRNA promoter or cRNA promoter in the presence of ApG primer and nucleotides (Fig. 3.6A). Wild-type RdRP, K627E RdRP and RdRP lacking the 627-domain were all able to elongate ApG primers, producing 14- and 15-nucleotide-long products on the vRNA or cRNA promoter templates respectively. The RdRP lacking the 627-, cap-binding and mid-link domains (PB2 Δ 248-679) was also able to elongate an ApG primer on a vRNA promoter template, but interestingly was unable to do so on a cRNA

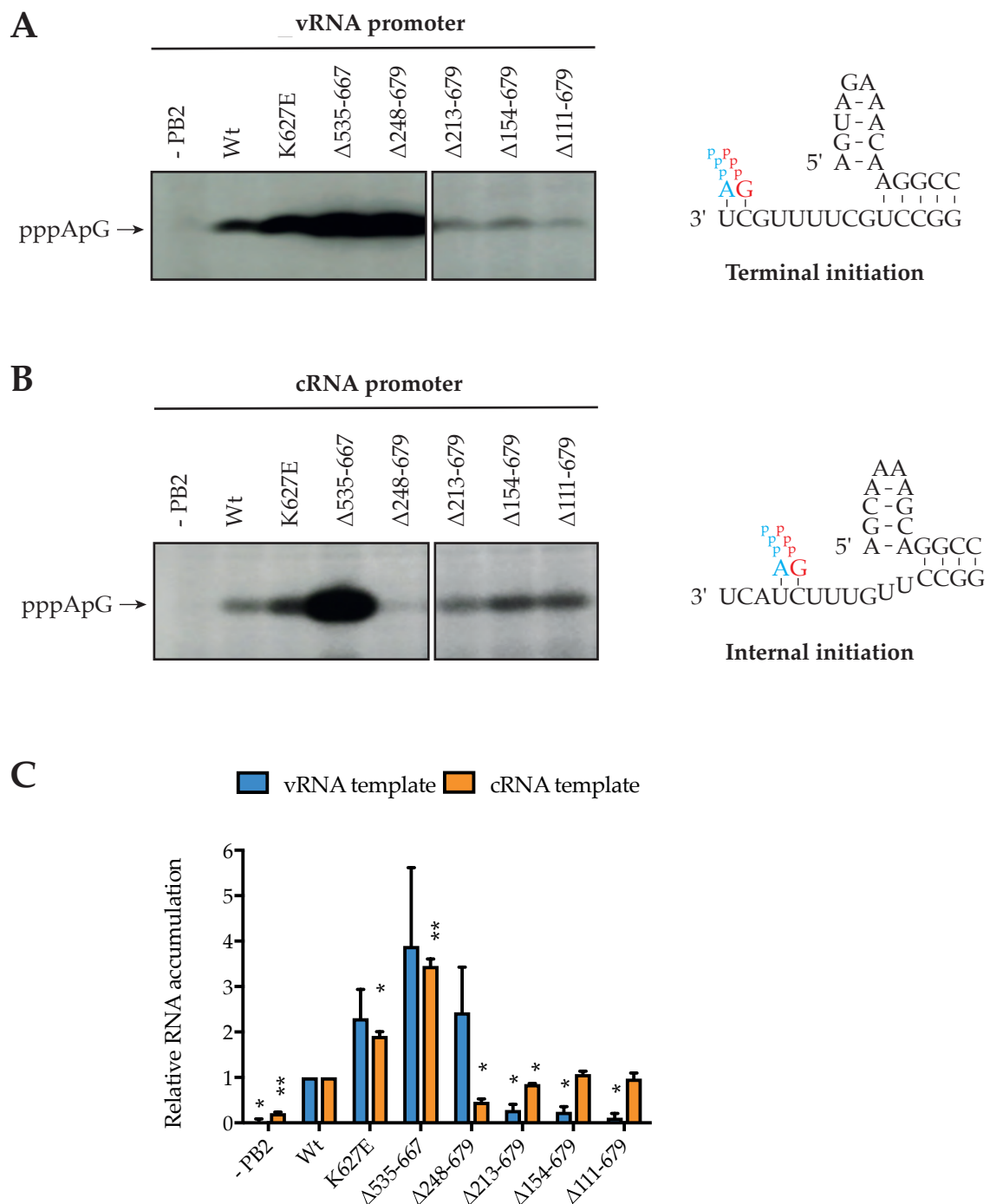


Fig. 3.5 Polymerase lacking the 627-domain is active in primer-independent initiation *in vitro*. Purified recombinant RdRP was incubated in reaction mixtures containing ATP, [α - ^{32} P] GTP and (A) either vRNA or (B) cRNA promoter template. Reaction products were treated with alkaline phosphatase and analysed by 20% PAGE and autoradiography. The schematics show (A) terminal initiation on a vRNA promoter and (B) internal initiation on a cRNA promoter (adapted from te Velthuis et al., 2016). Unlabelled nucleotides are shown in blue and radiolabelled GTP is shown in red. (C) The graph shows the mean intensity signal relative to wild-type RdRP from three independent biological replicates ($n = 3$), with error bars representing the standard error of the mean and the asteriks indicating a significant difference to wild-type (two-tailed one-sample t test) as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

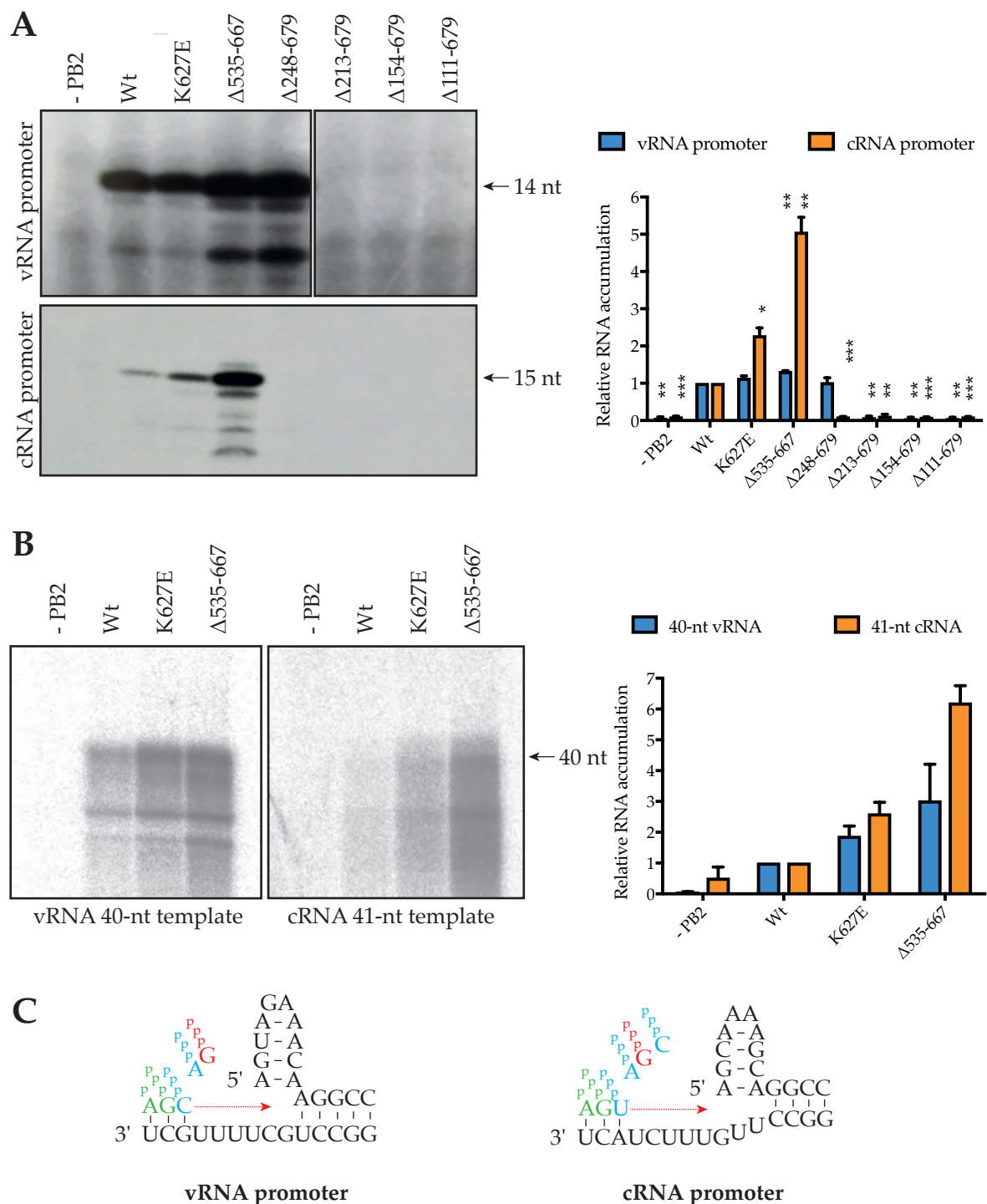


Fig. 3.6 Polymerase lacking the 627-domain is active in ApG-primed extension *in vitro*. Purified recombinant RdRP was incubated in reaction mixtures containing ATP, UTP, CTP, [α - ^{32} P] GTP, ApG primer and (A) either vRNA or cRNA promoter template or (B) a 47-nt-long vRNA or cRNA template. Reaction products were analysed by 20% PAGE and autoradiography. The graphs show the mean intensity signal relative to wild-type RdRP from three independent biological replicates ($n = 3$), with error bars representing the standard error of the mean and the asteriks indicating a significant difference to wild-type (two-tailed one-sample t test) as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. (C) The schematics show ApG-primed initiation on a vRNA or cRNA promoter structure (adapted from te Velhuis et al., 2016). The ApG primer is shown in green, unlabelled nucleotides are shown in blue and radiolabelled GTP is shown in red.

promoter template. Further deletions into the nonflexible N-terminal domains of PB2 resulted in a RdRP incapable of ApG elongation on either template. In order to determine whether the RdRP lacking the 627-domain is able to elongate RNA templates longer than the promoter regions, purified RdRP was bound to 5' vRNA or cRNA promoter and 40-nucleotide-long 3' vRNA or 41-nucleotide-long 3' cRNA in the presence of ApG and nucleotides (Fig. 3.6B). Wild-type RdRP, K627E RdRP and PB2 Δ 535-667 RdRP were all able to elongate an ApG primer on long vRNA and cRNA templates, however, the intensity of RNA products were much weaker with a long cRNA template.

Interestingly, PB2 Δ 535-667 RdRP exhibited an increased activity in *de novo* replication initiation and ApG primer extension on both vRNA and cRNA templates (Fig. 3.5 and 3.6).

Taken together, these results show that a polymerase lacking the 627-domain is active in cap-snatching, cap-dependent transcription initiation, *de novo* primer-independent replication initiation and ApG extension *in vitro*, indicating that the 627-domain is not required for core RdRP enzymatic functions. Furthermore, the data shows that an RdRP lacking not only the 627-domain but also the cap-binding and mid-link domains (PB2 Δ 248-679) is capable of primer-independent replication initiation and ApG primer extension on a vRNA template. However, on a cRNA template, primer-independent replication initiation is impaired and ApG primer extension fully inhibited, revealing differential requirements of the PB2 structure for viral replication initiation on the vRNA and cRNA promoters. The N-terminal nonflexible PB2 domains are essential for replication of the viral genome, even though they might only play a limited role in synthesis of the initial pppApG dinucleotide on a cRNA template.

3.2.4 The PB2 627-domain is required for cellular RdRP activity

Having shown that a RdRP lacking the 627-domain is active in transcription and replication *in vitro*, the role of the 627-domain in RdRP activity in a cellular context was assessed using mini-replicon assays. RNPs were reconstituted by co-expression of the three RdRP subunits, NP and segment 6 vRNA in human HEK-293T cells and the accumulation of positive- and negative-sense viral RNA was analysed by primer extension. In contrast to results obtained *in vitro* (Fig. 3.4 to 3.6), but in agreement with previous findings (Mehle & Doudna, 2008; Paterson et al., 2014), the avian-like K627E RdRP was significantly restricted in its ability to transcribe and replicate vRNA compared to wild-type RdRP (Fig. 3.7A). Interestingly, the RdRP lacking the 627-domain was unable to generate any viral RNAs.

To address the question, whether this observation is host specific or whether the 627-domain is also required in avian host cells, RNPs were reconstituted as before in chicken DF-1 cells. Wild-type and avian-like K627E RdRPs showed similar levels of viral RNA accumulation as has been reported previously (Paterson et al., 2014). However, PB2 Δ 535-667 RdRP was still incapable of viral RNA accumulation (Fig. 3.7B).

The nature of PB2-627 has been linked to affecting RNP assembly, putatively by recruiting or interacting with the first NP of the RNP (Labadie et al., 2007; Ng et al., 2012b; Rameix-Welti et al., 2009). To investigate whether the inability of the mutant RdRP to synthesise viral RNAs in a cellular context is associated with the disruption of RdRP-NP interactions, an alternative mini-replicon assay was utilised that relies on a segment 6 derived 47-nucleotide-long vRNA that can efficiently be transcribed and replicated in an NP-independent manner (Paterson et al., 2014; Turrell et al., 2013). Primer extension analysis showed significant levels of RNA accumulation by the avian-like K627E RdRP (Fig. 3.8), in

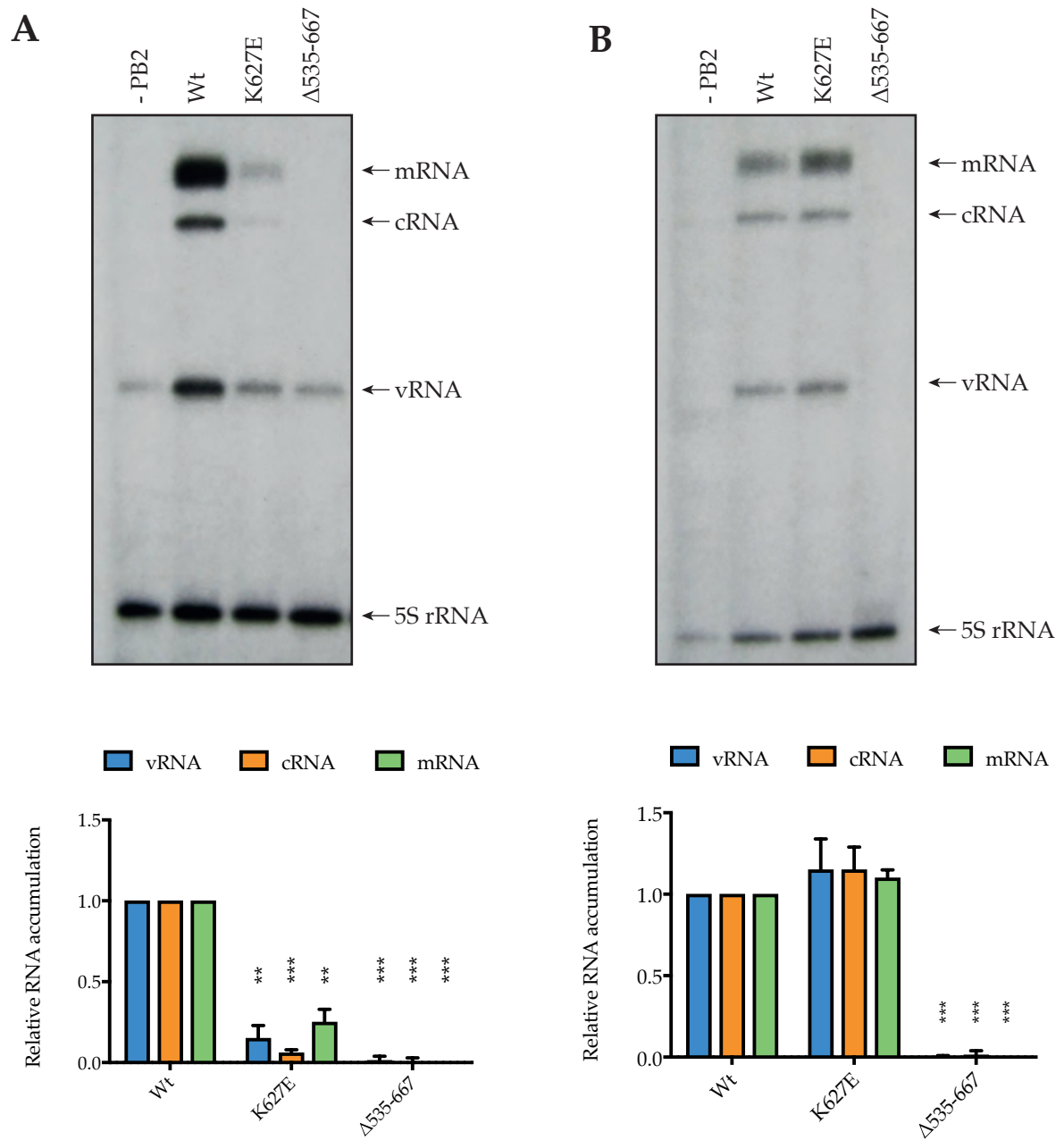


Fig. 3.7 **The PB2 627-domain is required for polymerase activity in the cell.** (A) Human HEK-293T cells or (B) chicken DF-1 cells were co-transfected with plasmids expressing PA, PB1, wild-type or mutant PB2, NP, and segment 6 vRNA. Accumulation of mRNA, cRNA, and vRNA was analysed by primer extension. The graphs show the mean intensity relative to wild-type RdRP from three independent biological replicates ($n = 3$). Mean intensities were subtracted by the mean intensity of the negative control (-PB2). Error bars represent the standard error of the mean and the asterisks indicate a significant difference to wild-type (two-tailed one-sample t test) as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

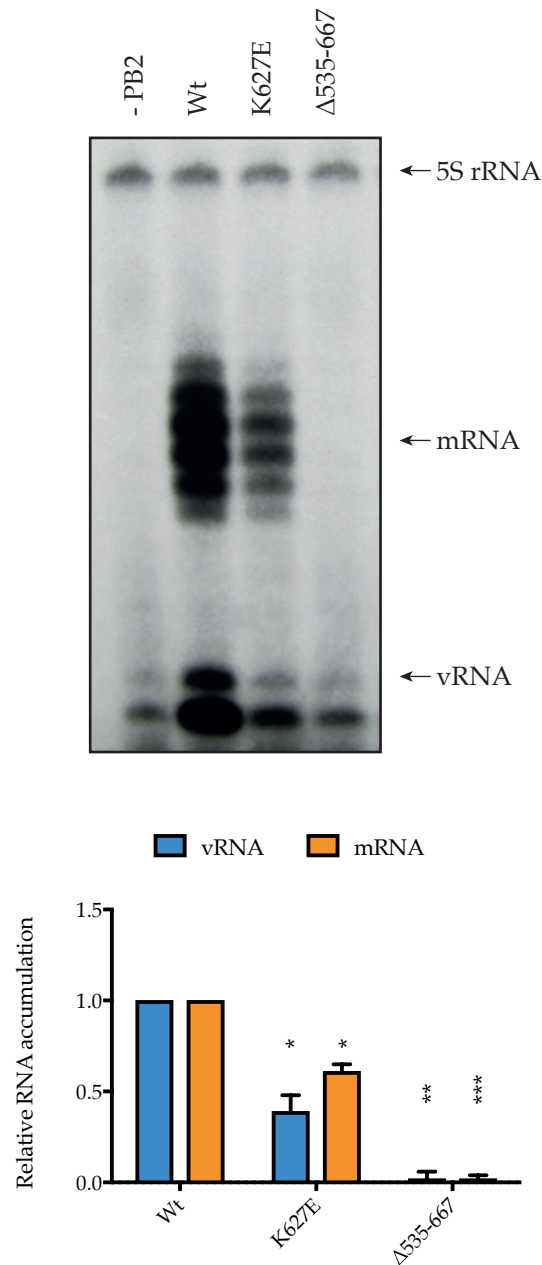


Fig. 3.8 The essential role of PB2 627-domain in viral replication is NP-independent. Human HEK-293T cells were co-transfected with plasmids expressing PA, PB1, wild-type or mutant PB2 and 47-nt-long segment 6-derived vRNA. Accumulation of mRNA and vRNA was analysed by primer extension. The graph shows the mean intensity relative to wild-type RdRP from three independent biological replicates ($n = 3$). Mean intensities were subtracted by the mean intensity of the negative control (-PB2). Error bars represent the standard error of the mean and the asterisks indicate a significant difference to wild-type (two-tailed one-sample t test) as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

agreement with previous findings that the PB2-627E-mediated restriction is diminished on short vRNA templates in mammalian host cells (Paterson et al., 2014). Nevertheless, PB2 Δ 535-667 RdRP remained incapable of RdRP activity on a short vRNA template. While it cannot be excluded that the 627-domain is required for RNP assembly on longer vRNA templates through an interaction with NP, its interaction with NP is not the primary cause for the observed defect in RNA synthesis in a cellular environment.

It was previously reported that avian RdRPs are successfully able to synthesise significant levels of mRNA in mammalian host cells, if *trans*-complemented with a transcription-deficient but replication-competent RdRP, suggesting that the restriction of avian RdRPs in mammalian hosts is replication specific and does not affect transcription itself (Mänz et al., 2012). Using this *trans*-complementation assay, the capability of PB2 Δ 535-667 RdRP for transcription was assessed. RNPs were reconstituted by co-expression of the three RdRP subunits, NP, segment 6 vRNA and as indicated *trans*-complementing transcription-deficient PB2 in human HEK-293T cells and the accumulation of viral RNA was analysed by primer extension (Fig. 3.9). In line with the original study, avian-like K627E RdRP was able to synthesise significantly enhanced levels of mRNA after *trans*-complementation with the transcription-deficient PB2 mutant. *Trans*-complementation of PB2 Δ 535-667 RdRP, however, did not stimulate mRNA synthesis, suggesting that PB2 Δ 535-667 RdRP is inherently defective in transcription in a cellular context.

Overall, the data shows that the 627-domain is essential for transcription and replication of vRNA in a cellular context, independent of the host species and the presence of NP and confirming the central role of PB2-627 in determining the ability of the RdRP to transcribe and replicate vRNA in mammalian host cells.

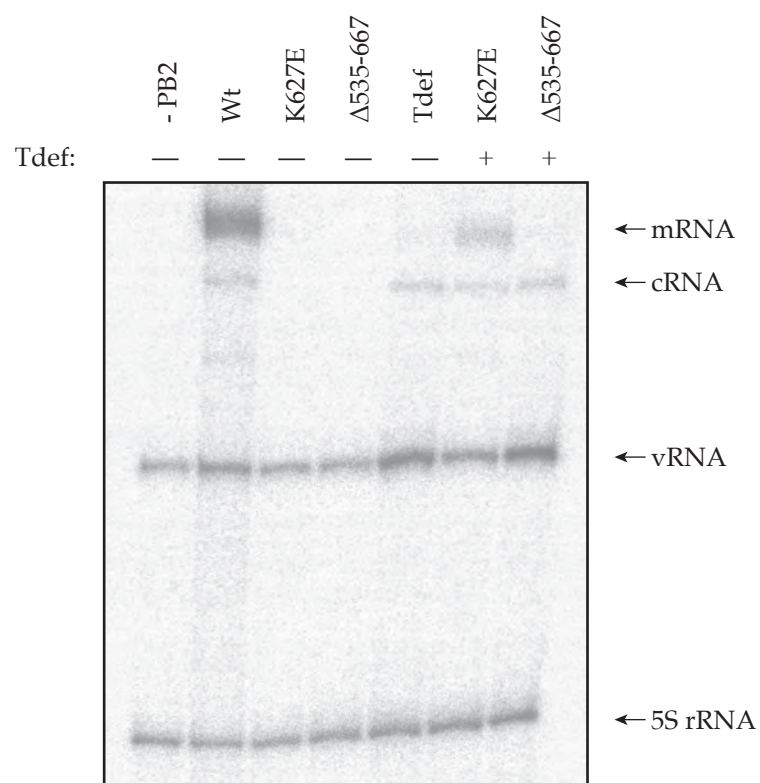


Fig. 3.9 PB2 627-domain is required for viral transcription in the cell. Human HEK-293T cells were co-transfected with plasmids expressing PA, PB1, PB2, NP, segment 6 vRNA and as indicated empty (-) vector or a transcription-deficient but replication-competent PB2-E361A/F404A (Tdef). Accumulation of mRNA and vRNA was analysed by primer extension, 6% PAGE and autoradiography. 5S rRNA was used as an internal loading control.

3.2.5 The PB2 627-domain mediates RdRP recruitment

The findings described above show that the PB2 627-domain is not required for core RdRP functions *in vitro*, but that it is essential for the accumulation of viral RNAs in the cell. In a cellular context, newly synthesised RNA-unbound RdRP needs to be imported into the nucleus and recruited to replicating vRNPs to stabilise newly synthesised cRNA by protecting it from degradation by host nucleases (Vreede et al., 2004). To investigate whether an RdRP lacking the 627-domain can stabilise cRNA in the cell following infection, the RdRP subunits PA, a PB1 active-site mutant (PB1a), which binds but does not replicate or transcribe vRNA, and wild-type or mutant PB2, were co-expressed in human HEK-293T cells, prior to infection with influenza A/WSN/33 (H1N1) virus in the presence of the cellular transcription-inhibitor actinomycin D (ACTD). Total RNA was extracted and viral RNAs were analysed by primer extension. Both in the presence of pre-expressed NP (Fig. 3.10A) and absence of pre-expressed NP (Fig. 3.10B), pre-expressed wild-type and K627E RdRPs were able to stabilise cRNA equally well, whereas a PA-PB1a dimer and the PB2 Δ 535-667 RdRP were unable to stabilise any cRNA. Stabilisation of cRNA was more efficient in the presence of NP than in the absence, as has been reported previously (Vreede et al., 2004), which could be attributed to the fact that while binding of the RdRP to the 5' and 3' cRNA promoter ends would confer protection of the nascent cRNA strand from host exonucleases, the internal unprotected region of the cRNA strand remain unprotected from host endonucleases, unless NP is present and able to form a cRNP-like structure.

The fact that the PB2 627-domain is required for stabilising nascent cRNA in infected cells (Fig. 3.10), but that it is located far away from the template entry channel and it is not required for viral promoter binding (Fig. 3.3), suggests that it could be involved in either facilitating activation of cRNA synthesis in the replicating vRNP (as a *trans*-activating RdRP) or in mediating recruitment of the free RdRP to the replicating vRNP.

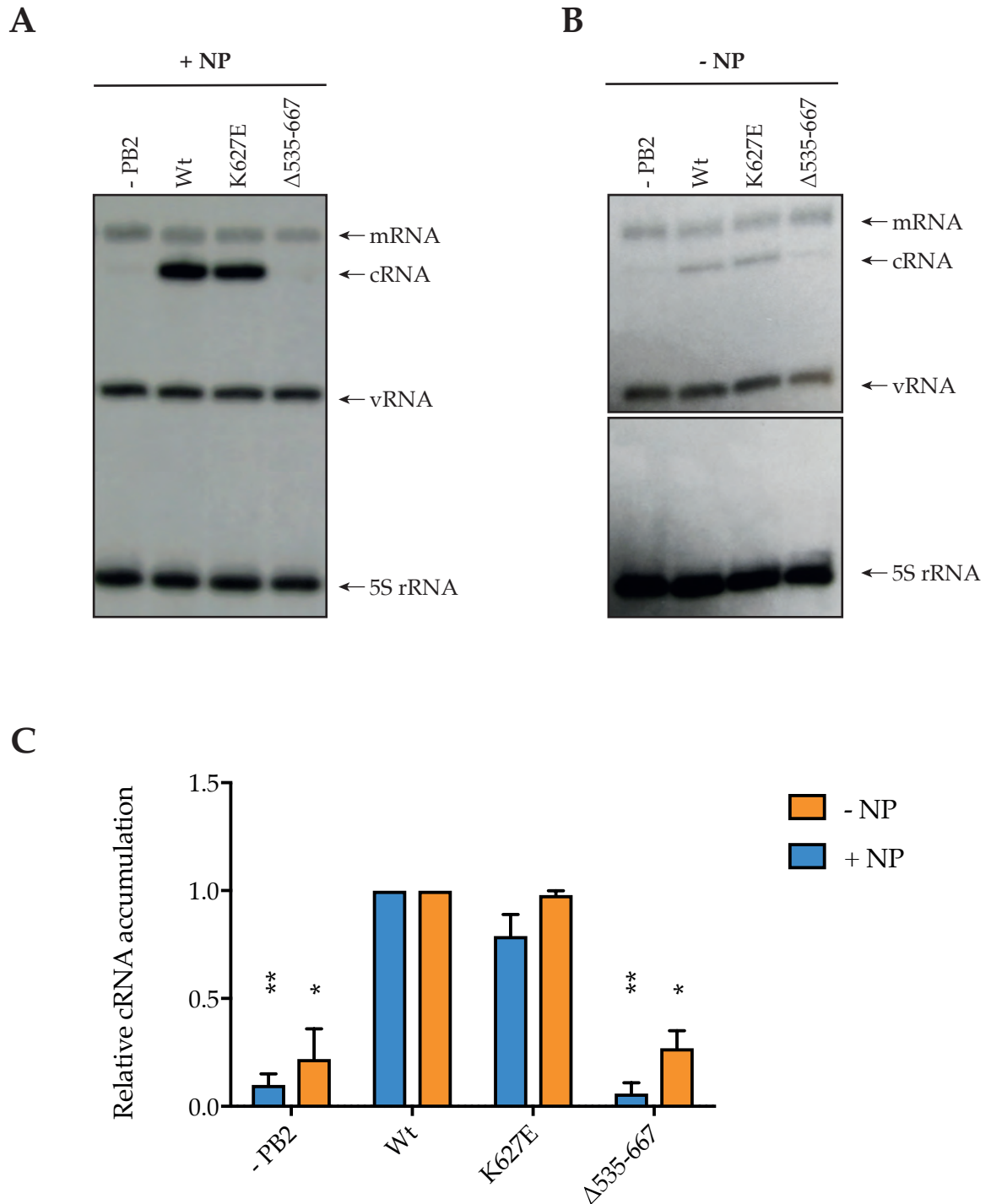


Fig. 3.10 The PB2 627-domain is required for cRNA accumulation in infected cells. Human HEK-293T cells were co-transfected with plasmids expressing PA, PB1a, wild-type or mutant PB2 (A) in the presence or (B) absence of NP. 48 h post-transfection, cells were infected with influenza A/WSN/33 (H1N1) virus at an MOI of 5 in the presence of actinomycin D. Accumulation of mRNA, cRNA, and vRNA 6 h.p.i. was analysed by primer extension. (C) The graph shows the mean intensity signal relative to wild-type RdRP from three independent biological replicates ($n = 3$). Error bars represent the standard error of the mean and the asterisks indicate a significant difference to wild-type (two-tailed one-sample t test) as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

A previous study revealed that purified cRNPs are unable to efficiently synthesise vRNA *in vitro*, unless additional RdRP, which does not require catalytic activity, is provided *in trans*, whereas purified vRNPs can efficiently synthesise cRNA in the absence of complementing RdRP, suggesting that free RdRP is needed to *trans*-activate cRNP through a direct RdRP-RdRP interaction (York et al., 2013). To assess the ability of PB2 Δ 535-667 RdRP to *trans*-activate RNP replication, purified vRNPs or cRNPs were incubated in the presence of ApG primers, nucleotides and purified RdRP. Accumulated viral RNA products were extracted and analysed by primer extension. In agreement with the original study, vRNPs were capable of cRNA synthesis in an RdRP-independent manner, while cRNPs required additional RdRP to be able to produce vRNA (Fig. 3.11). Both wild-type RdRP and RdRP lacking the 627-domain were equally able to stimulate vRNA synthesis, suggesting that the 627-domain is not involved in *trans*-activation of cRNPs and that its deletion does not interfere with polymerase-polymerase interaction. The original study was performed using recombinant polymerase expressed in a baculovirus system which allows for significantly higher concentrations of polymerase. This difference in polymerase concentration could explain the weak levels of *trans*-activation observed in Figure 3.11.

In order to determine whether in general assembly of oligomeric polymerase complexes is affected by the loss of the 627-domain, RdRP subunits PB1, PB2, PA as well as PB2-TAP were co-expressed as indicated in human HEK-293T cells and purified using the protein A tag on PB2. Analysis of purified protein complexes by SDS-PAGE and western blotting for PB2 revealed that PB2 monomers and PB2-PB1 and PB2-PA dimers did not interact with each other to form dimeric complexes. However, full heterotrimeric RdRPs did in fact form RdRP dimer structures and no difference could be observed between wild-type PB2, PB2-K627E or PB2 Δ 535-667 in their effect on co-purification of two RdRPs (Fig. 3.12). This shows that viral RdRPs can interact with each other and that PB2-627 and the entire 627-

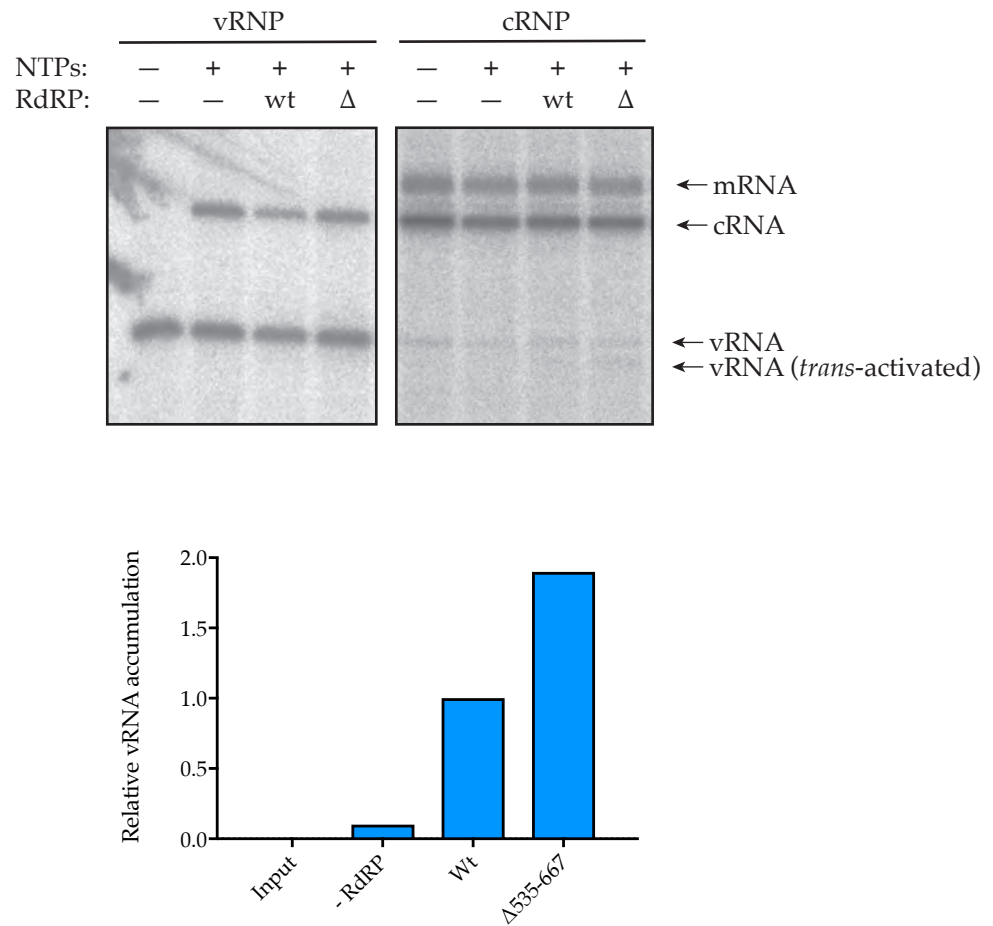


Fig. 3.11 **The 627-domain is not required for *trans*-activation of cRNPs.** Purified vRNPs or cRNPs were incubated in the presence of ApG primers, nucleotides and purified RdRP (wild-type = wt, PB2 Δ 535-667 = Δ) as indicated. The accumulation of vRNA and cRNA was analysed by primer extension. The graph shows the intensity of the vRNA accumulation from *trans*-activated cRNPs from a single experiment.

domain do not affect this interaction, strengthening the observation that the lack of the 627-domain does not interfere with cRNP *trans*-activation (Fig. 3.11). However, it is important to note that this result was observed solely in the presence of viral RdRP and in the absence of any viral RNA and NP.

Taken together, these results show that the PB2 627-domain is required for the stabilisation of nascent cRNA in infected cells, but that it isn't necessary for direct polymerase-polymerase interaction or for cRNP *trans*-activation. However, the nature of the amino acid residue of PB2-627 does not affect the ability of the RdRP to stabilise cRNA.

3.3 Discussion

The aim of this study was to gain insight into the function of the C-terminal domains of PB2 in general and the 627-domain of the influenza A virus RNA polymerase in particular. The study found that the 627-domain is not required for nuclear accumulation of PB2 or for heterotrimeric RdRP assembly. Recombinant RdRP lacking the 627-domain is able to carry out core RdRP functions such as capped RNA primer-dependent transcription and RNA synthesis *in vitro*. However, in a cellular context, it is unable to replicate or transcribe viral RNA, although it is still able to bind vRNA and cRNA templates. In infected cells, vRNPs do not stably produce cRNA if PB2 Δ 535-667 RdRP is provided in trans, even though the RdRP does not depend on the 627-domain for assembly of oligomeric polymerase complexes or for *trans*-activation of cRNPs, suggesting that the 627-domain may play a role in viral RNA genome replication through mediating RdRP recruitment.

The finding that the PB2 627-domain is not required for nuclear accumulation and RdRP assembly is consistent with previous data. PB2 is imported into the nucleus through its C-terminal NLS using the classical importin- α/β nuclear import pathway (Tarendeau et al.,

2007). The presented data suggests that the deletion of the 627-domain does not affect the interaction of the C-terminal NLS with importin- α . Furthermore, the finding that the 627-domain is not required for heterotrimeric RdRP assembly agrees with previous findings that N-terminal PB2 fragments can form stable complexes with the PB1-PA dimer (Chang et al., 2015; Swale et al., 2016). Interestingly though, these findings also suggest that the presence or absence of the C-terminal NLS on PB2 does not affect the amount of heterotrimeric RdRP in a cell, which conflicts with the classical model where heterotrimer assembly takes place in the nucleus. This suggests either that nuclear import of PB2 is not solely dependent on the C-terminal NLS of PB2 or that the RdRP in fact can assemble in the cytoplasm.

The 627-domain was also shown not to interfere with RNA synthesis *in vitro*, indicating that the 627-domain does not contribute to core RdRP enzymatic functions. This is consistent with the location of the 627-domain on the outside of the RdRP core, which is made up of PB1, the C-terminal domain of PA and the N-terminal PB2 Nter, N1, N2 and lid (Hengrung et al., 2015; Pflug et al., 2014; Reich et al., 2014). The 627-domain is also not required for RNA template binding, as the RNA-binding functions reside in the RdRP core (Pflug et al., 2014; Reich et al., 2014). The RdRP lacking the 627-domain was also found to be able to cleave capped RNA and extend the resulting capped RNA primer. Interestingly though, it produced a shorter major capped RNA product in the extension reaction than the wild-type RdRP and only a single cleavage product. This suggests that deletion of the 627-domain affects the relative positioning of the cap-binding domain, the endonuclease domain and the RdRP active site and that the 627-domain plays a minor structural role in influenza virus transcription.

A differential requirement for the cap-binding and mid-link domains of PB2 in primer-independent *de novo* and ApG-primed RNA synthesis on vRNA and cRNA templates was also revealed. While an RdRP lacking the 627-domain could initiate replication and extend

an ApG primer on both vRNA and cRNA templates, truncation of the cap-binding and mid-link domains resulted in a RdRP that was able to initiate and extend on a vRNA template, but was unable to do so efficiently on a cRNA template. This differential requirement of PB2 domains for RdRP activity could be linked to the different modes of initiation of the RdRP. While the initiation of replication on the vRNA template occurs at the terminal positions 1 and 2 of the 3' end, initiation on the cRNA template takes place internally at positions 4 and 5, resulting in a pppApG dinucleotide that is used as a primer for full-length vRNA synthesis only after backtracking on the cRNA template to positions 1 and 2 in the RdRP active site (Deng et al., 2006b). Furthermore, initiation on the vRNA template requires a priming loop, a β -hairpin protruding into the RdRP active site, while for initiation on the cRNA template the priming loop needs to be retracted (Velthuis et al., 2016). Thus, it can be speculated that deletion of the mid-link and cap-binding domains might affect the ability of the template to backtrack or the priming loop to retract, leading to the inhibition on the cRNA but not on the vRNA template. Further deletions of PB2 resulted in an RdRP with no detectable activity in ApG extension and reduced initiation on a vRNA template, demonstrating that the stable N-terminal domains (Nter, N1, N2, lid; amino acids 1 to 247) are not only structurally but also functionally part of the RdRP core as has been suggested in other studies (Hara et al., 2017).

Although the 627-domain is not important for RNA polymerase activity *in vitro*, it is absolutely essential for viral RNA accumulation in RNP reconstitution assays in a cellular context. Previous studies suggested that the nature of the amino acid at position 627 affects the assembly of viral RNPs (Labadie et al., 2007; Ng et al., 2012b; Rameix-Welti et al., 2009). However, we found that deletion of the 627-domain also prevents replication of short vRNA templates that can be replicated in the absence of NP (Paterson et al., 2014; Turrell et al., 2013). This suggests that the inhibition of RdRP function by the deletion of the 627-domain is not simply due to a disruption of an interaction between the RdRP and

NP, even if it possibly may play a role. How, then, does the PB2 627-domain contribute to viral RNA replication in a cellular context? My data shows that vRNPs do not produce cRNA in infected cells if the RdRP lacking the PB2 627-domain is provided in trans. Previous studies suggested that vRNPs produce cRNA early in infection but that cRNA gets degraded unless RdRP and NP reach sufficient levels to bind and stabilise it (Vreede et al., 2004). However, these studies did not exclude the possibility that, in a cellular context, the stabilising RdRP also needs to associate with the replicating RdRP and possibly *trans*-activate it before cRNA is produced.

It was found that vRNA and cRNA expressed through plasmid transfection could be co-purified with the RdRP lacking the 627-domain, suggesting that RNA binding is not affected, in agreement with the *in vitro* data described above. If the 627-domain lacking RdRP can bind cRNA, it follows that it must be deficient in binding and/or *trans*-activating the vRNP-resident RdRP. Oligomerisation of the RdRP heterotrimer has been proposed previously (Chang et al., 2015; Jorba et al., 2008; Swale et al., 2016) and a *trans*-activating or *trans*-acting RdRP has been implicated in replication (Jorba et al., 2009; York et al., 2013). It would therefore be tempting to speculate that the PB2 627-domain might be involved in the interaction between the RNP-resident and the *trans*-acting or *trans*-activating RdRP. However, the data described above suggests that the RdRP lacking the 627-domain can *trans*-activate cRNPs *in vitro* and form RdRP heterotrimer dimers or oligomers. It is important to remember though, that the latter result was obtained in the absence of viral RNA, NP and RNPs and that the RdRP dimer observed could be irrelevant for a vRNP-RdRP interaction. Therefore, one could hypothesise that the 627-domain might be involved in an interaction with a cellular host factor that could be important for the recruitment of the *trans*-acting or *trans*-activating RdRP to RNPs in a cellular environment.

Recently, ANP32A was identified as a host factor underlying the host restriction mediated by PB2 amino acid residue 627 (Long et al., 2016). Avian-like (PB2-627E) and mammalian (PB2-627K) RdRPs are equally able to stabilise cRNA, suggesting that while the 627-domain as a whole might be important for RdRP recruitment, the nature of position 627 itself does not affect this recruitment of RdRP to replicating vRNPs. Although not specifically investigated in this study, one could assume that the 627-domain is also required for stabilising vRNA and recruitment of RdRPs to replicating cRNPs.

3.4 Summary

In summary, it was demonstrated here that the PB2 627-domain of the influenza A virus RNA polymerase is not involved in core catalytic functions of the RdRP but that it is essential for the replication of RNPs in a cellular context. We propose that the 627-domain in RNA-free RdRPs is necessary for mediating recruitment of the RdRP to replicating RNPs through a host factor in a cellular environment. These findings further our understanding of the role of the PB2 627-domain in transcription and replication of the influenza A virus RNA genome by the viral polymerase.

Chapter 4: The role of ANP32A in viral polymerase function

4.1 Introduction

Avian influenza virus RdRPs have long been known to be restricted in their activity in mammalian hosts and this host range restriction was initially traced to the PB2 segment of the viral genome (Almond, 1977). However, a single mutation in PB2-627 is able to restore avian influenza virus RdRP activity in mammalian host cells (Subbarao et al., 1993). Similar host adaptive mutations have been found in all three RdRP subunits. Nevertheless, the mechanism of why avian RdRPs are restricted in mammalian cells is not fully understood yet. It has been shown though, that avian RdRPs have a defect in vRNA synthesis, which the authors attribute to avian cRNPs being unstable (Mänz et al., 2012). Primary transcription by avian influenza viruses was shown to be unimpaired (Mänz et al., 2012).

Recently, the avian host factor ANP32A was identified as the missing host factor that avian RdRPs rely on for efficient viral replication. Expression of the avian homologue of ANP32A was able to fully restore avian RdRP activity in human host cells (Long et al., 2016). ANP32A is not only involved in host range restriction of influenza viruses, but is also a necessary host factor for viral replication, but not transcription, in general for both avian and mammalian hosts (Long et al., 2016; Sugiyama et al., 2015). Interestingly, human ANP32A and ANP32B have both been characterised to be enhancers of vRNA synthesis by a human-adapted viral RdRP *in vitro* (Sugiyama et al., 2015).

The avian and mammalian homologues of ANP32A have retained a high sequence identity except for a 33 amino acid insert in the avian ANP32A. This insert consists of a 27 amino acid-long duplication of an upstream sequence and a 6 amino acid-long unique sequence, which contains a SUMO interaction motif (SIM)-like sequence (Domingues & Hale, 2017; Long et al., 2016).

It is still unknown how ANP32A regulates the viral RdRP and how it exerts its species-specific function. Interactions between ANP32A and the viral RdRP have been reported in several studies and it has been suggested that ANP32A interacts with the full heterotrimeric RdRP which is in the viral RNA-free apo form (Bradel-Tretheway et al., 2011; Domingues & Hale, 2017; Sugiyama et al., 2015; York et al., 2014). The C-terminal LCAR of ANP32A has been reported to be mainly responsible for interacting with the viral RdRP (Domingues & Hale, 2017). Interestingly, a differential ability for avian and mammalian ANP32A homologues to interact with the viral RdRP has been reported recently, suggesting that the species-specific effect is facilitated by the unique 6 amino acid-long sequence which is specific to the avian ANP32A host factor, suggesting that another SUMOylated host factor potentially could be involved in stimulating the avian RdRP (Domingues & Hale, 2017).

The aim of this study is to gain a better understanding of how ANP32A regulates viral replication by using a combination of mass spectrometry analyses of the ANP32A interactome as well as cellular and *in vitro* activity assays. An improved understanding of the role of ANP32A in viral RdRP function will address many of the remaining questions concerning the molecular mechanisms of influenza virus host range restriction.

4.2 Results

4.2.1 The ANP32A interactome

Although ANP32A and ANP32B have been identified as co-purifying host factors with the viral RdRP in previous experiments (Bradel-Tretheway et al., 2011; York et al., 2014), no comprehensive analysis of the interactome of ANP32A has been presented yet. While it is known that ANP32A plays a central role in viral replication, the mechanism of how it regulates this process is unknown. In vitro studies have suggested that ANP32A directly exert their stimulatory effect on the viral RdRP in vRNA synthesis, but it does not exclude the possibility that other ANP32A-interacting host factors are also involved in this process.

In order to characterise the full ANP32A interactome, chicken ANP32A (chANP32A) and human ANP32A (huANP32A) fused to an N-terminal Strep-tag were expressed in human HEK-293T cells. A Strep-tagged coat protein of Pseudomonas bacteriophage PP7 (PP7CP-Strep) was included as a negative control. Protein complexes were purified using Strep-Tactin Sepharose chromatography and analysed by mass spectrometry (Fig. 4.1A). Mass spectrometry analyses and protein identifications were performed by the Advanced Proteomics Facility (University of Oxford) using MaxQuant and the Andromeda softwares. Identified protein groups were further analysed using the Perseus software. Common contaminants, serum albumin and keratins were excluded from the analysis. Proteins that were not present in both biological replicates were dismissed from further analysis (Fig. 4.1B). Protein abundance was normalised by using the mean ratio of protein intensity between ANP32A interactome and PP7CP interactome. Proteins that were not at least 10-fold enriched ($FC \geq 10$) in their abundance in the chANP32A or huANP32A interactomes compared to the PP7CP interactome were excluded from the analysis.

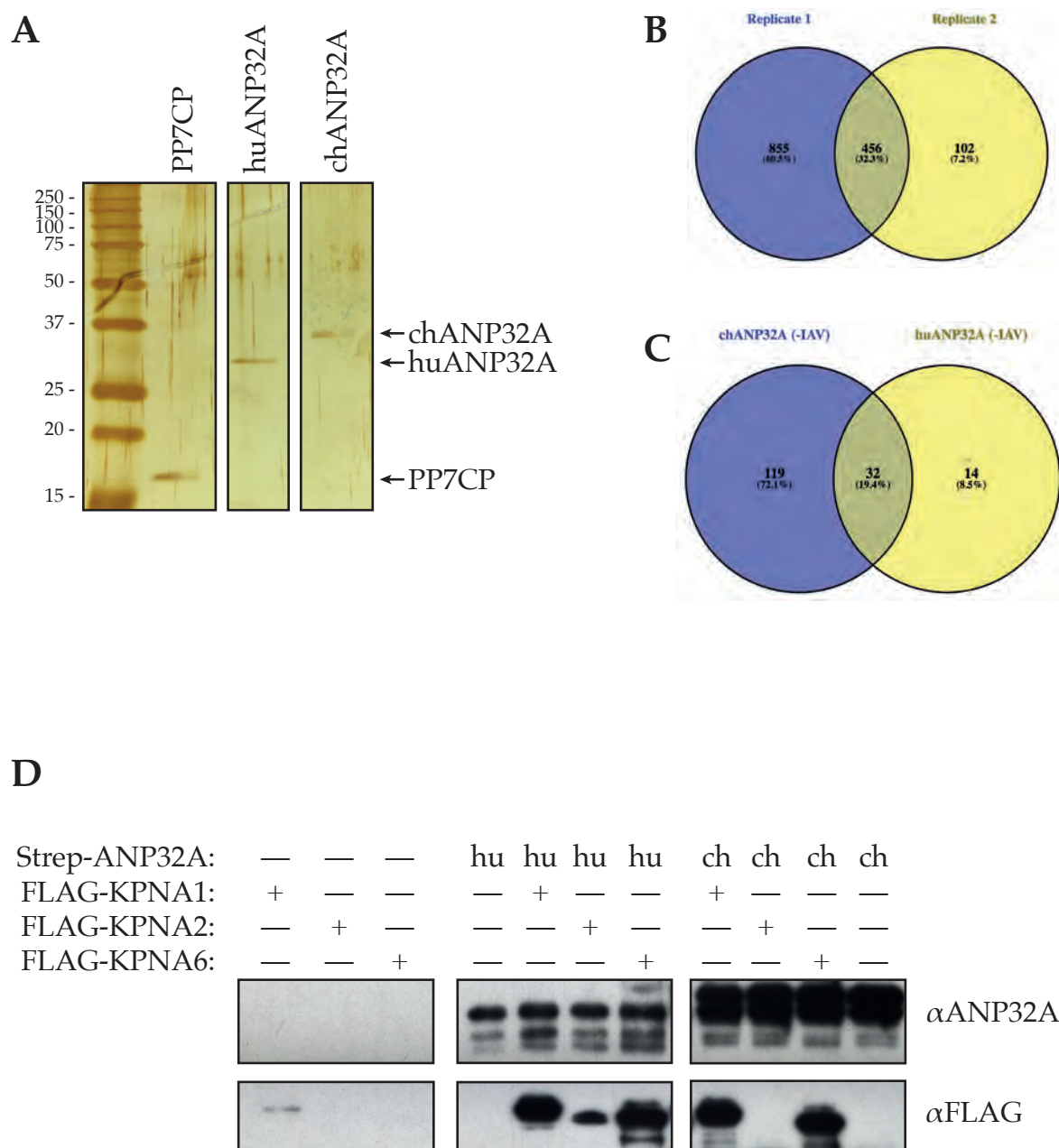


Fig. 4.1 Isolation of ANP32A and the associated interactomes from uninfected cells. (A) Avian and human ANP32A was expressed in and purified from human HEK-239T cells. Purified protein complexes were analysed by SDS-PAGE and silver staining. Molecular masses, indicated on the left, are depicted in kilodaltons. (B) 456 common proteins that were identified in both biologically independent replicates were analysed. (C) 32 common proteins were detected in chicken and human ANP32A interactomes. (D) Strep-tagged human (hu) or chicken (ch) ANP32A were co-expressed with FLAG-tagged KPNA1, KPNA2 or KPNA6 in human HEK-293T cells. Protein complexes were purified by Strep-Tactin chromatography and analysed by SDS-PAGE and western blotting.

Using these conditions a total of 151 host proteins were identified to interact with chANP32A in human HEK-293T cells (Appendix D), while 46 proteins were interacting with huANP32A (Appendix E). Between both data sets 32 common proteins were identified (Fig. 4.1C). Interestingly, both chANP32A and huANP32A were found to strongly interact with KPNA1 and KPNA6, while only weakly interacting with KPNA2 and KPNA5. The importin- α family plays a major role in nuclear import of influenza virus protein complexes and has been linked to host adaptation before (Gabriel et al., 2011).

When comparing the enrichment of proteins in the chANP32A or huANP32A samples to each other, 86 proteins were found to be more than 10-fold enriched in chANP32A samples compared to huANP32A samples and only 7 proteins were found to be more than 10-fold enriched in huANP32A samples compared to chANP32A samples (Appendix F). Interestingly, human ANP32B was found to strongly interact with human ANP32A, while chicken ANP32A did not show any interaction with human ANP32B. No difference in enrichment for any of the importin- α family members was detected.

In order to validate importin- α binding to ANP32A, ANP32A fused to an N-terminal Strep tag was co-expressed with KPNA1, KPNA2 and KPNA6 fused to an N-terminal FLAG tag in human HEK-239T cells. Protein complexes were purified from cell lysates by Strep-Tactin Sepharose chromatography and analysed by SDS-PAGE and western blot (Fig. 4.1D). Strong interaction of both human and chicken ANP32A with human KPNA1 and KPNA6 is detected. No interaction between chicken ANP32A and KPNA2 and a strongly reduced interaction between human ANP32A and KPNA2 was observed, confirming a much weaker binding of ANP32A to KPNA2 compared to KPNA1 and KPNA6. In addition, a slightly stronger interaction between human ANP32A and KPNA2 was also detected by mass spectrometry.

4.2.2 The ANP32A interactome in influenza A virus infection

ANP32A has been shown to be a crucial host factor required for viral infection. In order to determine whether infection with influenza A virus alters the ANP32A interactome, chicken ANP32A and human ANP32A fused to an N-terminal Strep-tag were expressed in human HEK-293T cells. A Strep-tagged coat protein of Pseudomonas bacteriophage PP7 was included as a negative control. 48 h post-transfection cells were infected with influenza A/WSN/33 (H1N1) virus at a high MOI of 5 for 6 h. Protein complexes were purified using Strep-Tactin Sepharose chromatography and analysed by mass spectrometry (Fig. 4.2A). Using the same analysis as before, only host proteins that were detected in both biological replicates were included into the analysis (Fig. 4.2B). In total 196 host proteins were found to interact with chANP32A, while 69 interacted with huANP32A with an overlap of 64 host proteins (Fig. 4.2C). Five proteins (ANP32B, AZGP1, TXN, PIP and HAGH) were significantly (> 10-fold) enriched in the huANP32A samples compared to the chANP32A samples. Vice versa, there were 106 proteins that were significantly enriched in the chANP32A samples compared to the huANP32A samples (Appendix G).

All three viral RdRP subunits PA, PB1 and PB2 were consistently shown to interact with both human and chicken ANP32A, although a stronger interaction was detected with chANP32A, which is consistent with a previous report (Domingues & Hale, 2017). Notably, significantly more viral RdRP was detected than viral NP. Considering that NP is much more abundant in viral RNPs than viral RdRP, this suggests that ANP32A does not interact with RNPs, but instead with RNA-free RdRP. chANP32A also appeared to strongly interact with the viral membrane glycoproteins HA and NA (Table 4.1). An interaction between human ANP32B and NA was reported previously (Watanabe et al., 2014).

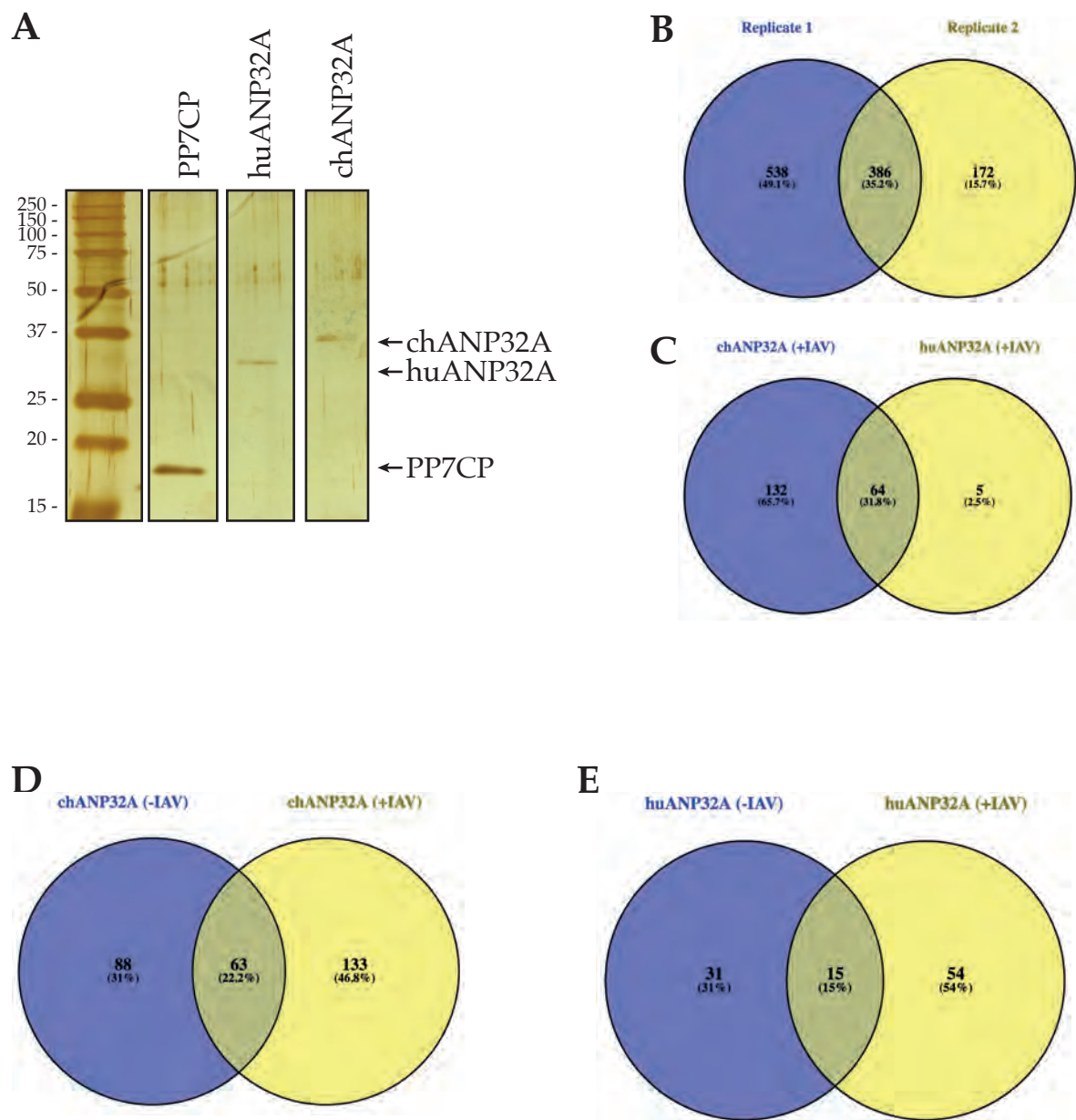


Fig. 4.2 Isolation of ANP32A and the associated interactomes from infected cells. (A) Avian and human ANP32A was expressed in human HEK-239T cells. Cells were infected with influenza A/WSN/33 (H1N1) virus for 6 h before protein complexes were purified by Strep-Tactin chromatography and analysed by SDS-PAGE and silver staining. Molecular masses, indicated on the left, are depicted in kilodaltons. (B) 386 common proteins that were identified in both biologically independent replicates were analysed. (C) 64 common proteins were detected in chicken and human ANP32A interactomes. (D) Infection with influenza A virus induced significant changes in the (D) chicken and (E) human ANP32A interactomes.

Table 4.1 Interactions of chicken ANP32A and human ANP32A with influenza A/WSN/33 (H1N1) virus

#	Viral protein	FC (chANP32A)	FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
1	PB2	1.69E+02	2.34E+01	7.22E+00
2	PB1	5.74E+01	1.62E+01	3.54E+00
3	PA	1.69E+02	2.34E+01	7.94E+00
4	HA	4.50E+01	6.73E-01	6.69E+01
5	NP	3.15E+00	1.11E+00	2.83E+00
6	NA	1.77E+01	5.12E-01	3.46E+01
7	M1	3.20E+00	9.88E-01	3.24E+00
8	NS1	1.71E+00	8.27E-01	2.07E+00

Comparison of the ANP32A interactomes in uninfected and infected cells revealed that viral infection induced a significant change in the ANP32A interactome. Only 22% of the significantly enriched chANP32A interactome was found both in uninfected and infected cells (Fig. 4.2D). In the huANP32A interactome only 15% common proteins were identified (Fig. 4.2E). 103 proteins were more than 10-fold enriched in their interaction with chANP32A after infection (Appendix H), while 44 proteins were more than 10-fold enriched in their interaction with huANP32A after viral infection compared to mock infected cells (Appendix I). Interestingly, KPNA2, which exhibited a weak interaction with ANP32A in the absence of a viral infection, was detected at much higher levels after infection (chANP32A: 70-fold enrichment compared to uninfected cells; huANP32A: 37-fold enrichment compared to uninfected cells).

Together, there are 300 unique proteins that were more than 10-fold enriched in their interaction with either human or chicken ANP32A in mock-infected or infected samples compared to PP7CP. Gene ontology (GO) analysis using the PANTHER classification system showed that the cumulative interactome of ANP32A is linked to a wide range of biological processes with the majority being classified as cellular processes (mainly associated with cell cycle and cell communication processes) and metabolic processes (mainly associated with the primary metabolic process and nitrogen compound metabolic process) as presented in Figure 4.3. GO analysis for molecular functions of the ANP32A interactome demonstrated a strong association with nucleic acid binding, various catalytic activities and structural function in the ribosome (Fig. 4.4).

Taken together, these findings represent the first comprehensive interactome study of avian and human ANP32A and the dynamic effect influenza A virus infection has on the interactome. The viral RdRP was found to interact with ANP32A regardless of host species and a weaker interaction with both HA and NA was detected as well. The strong enrichment of viral RdRP over NP suggests an interaction of ANP32A with RNA-free

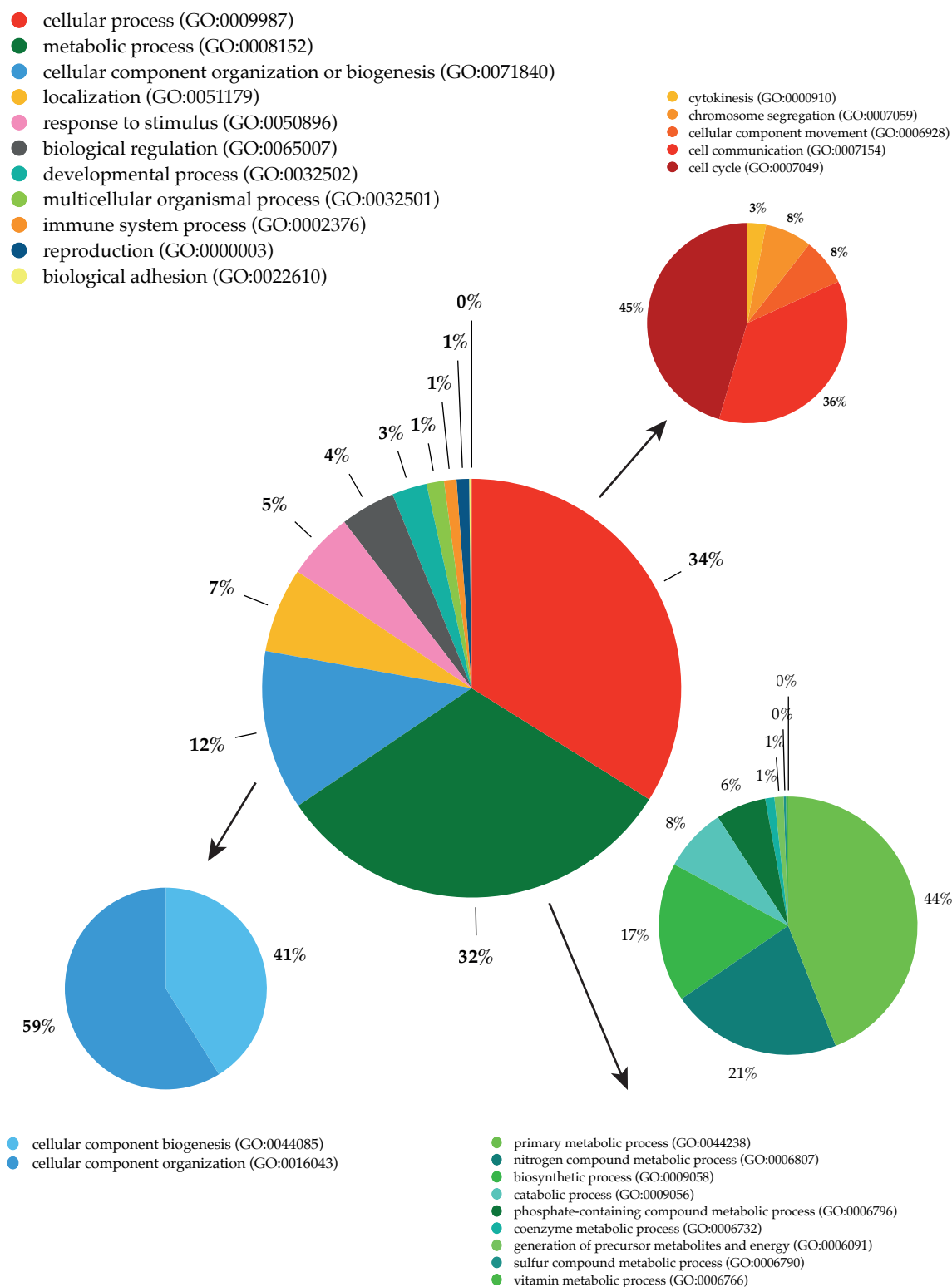


Fig. 4.3 Gene ontology (GO) analysis for biological processes of the ANP32A interactome. All 300 host proteins that were found to interact with either huANP32A or chANP32A (10-fold enrichment over PP7CP) were subjected to PANTHER gene list analysis for biological processes associated with them. Note that a single protein can be associated with several GO terms.

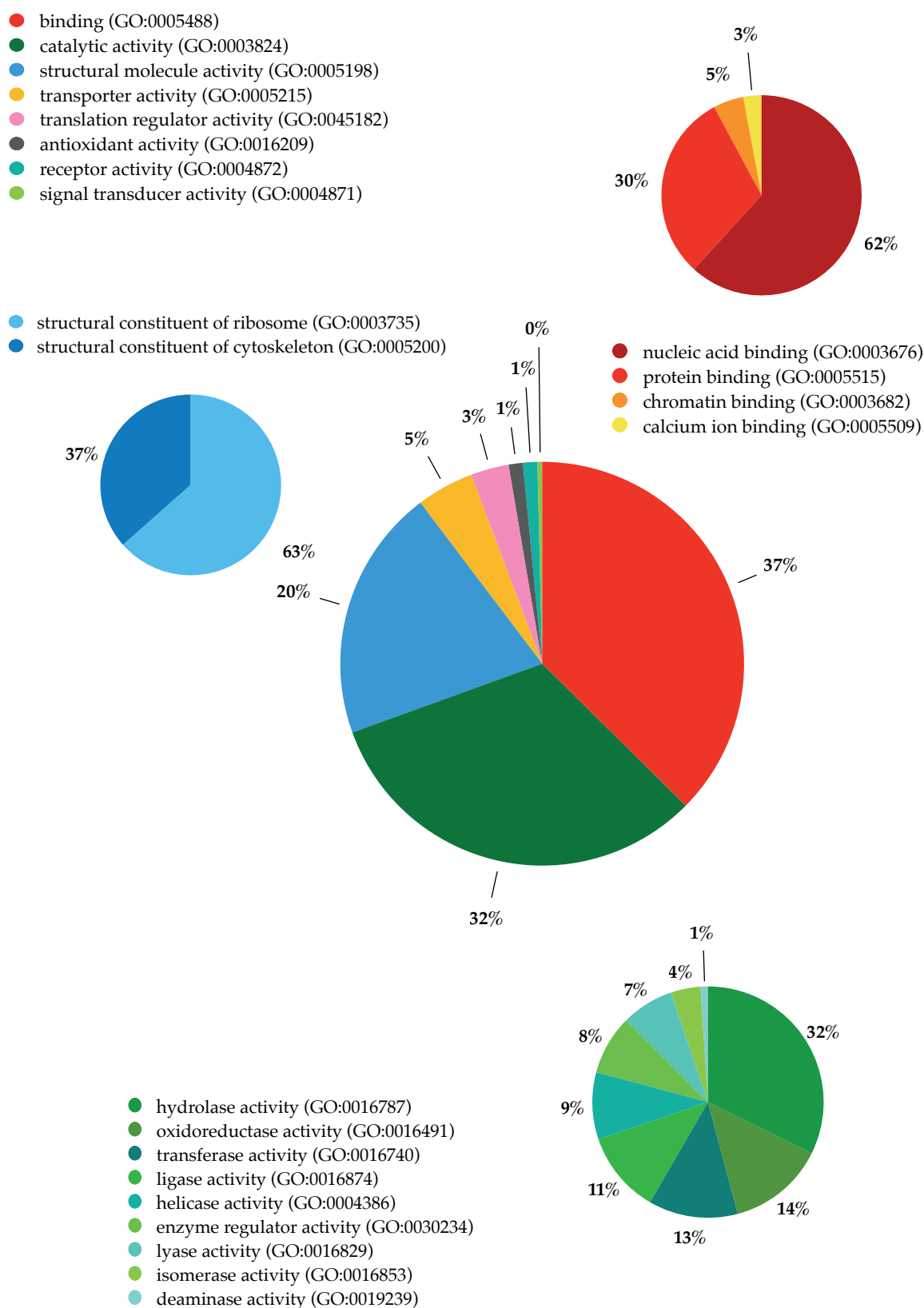


Fig. 4.4 Gene ontology (GO) analysis for biological processes of the ANP32A interactome. All 300 host proteins that were found to interact with either huANP32A or chANP32A (10-fold enrichment over PP7CP) were subjected to PANTHER gene list analysis for molecular functions associated with them. Note that a single protein can be associated with several GO terms.

RdRP and not RNPs. The importin- α family, which is an essential component of the nuclear import mechanism of influenza virus protein complexes, was established to be among the main interacting partners of ANP32A. Moreover, increased binding to KPNA2 was observed after infection.

4.2.3 Avian ANP32A is required for avian IAV RdRP activity in mammalian host cells

Previously, it had been suggested that avian influenza A viruses miss a positive host factor in mammalian host cells for efficient RdRP activity (Moncorge et al., 2010). In agreement with this, a different study showed that the host factor ANP32A is generally required for efficient viral replication in a cellular environment, but also, that it determines the host range of avian influenza A viruses (Long et al., 2016). In order to confirm these results, segment 6 vRNA, NP, PA, PB1 and either a wild-type mammalian-adapted (627K) or an avian-like (K627E) PB2 subunit were co-expressed in the presence of either huANP32A, chANP32A or the related huANP32B in human HEK-293T cells. The accumulation of positive- and negative-sense viral RNA was analysed by primer extension. In line with the previously reported findings, in the absence of avian ANP32A the avian-like RdRP was significantly restricted compared to the human-adapted RdRP and overexpression of huANP32A did not affect RdRP activity. However, overexpression of chANP32A significantly enhanced vRNA, cRNA and mRNA synthesis by the avian-like RdRP in mammalian host cells, while not stimulating the human-adapted RdRP (Fig. 4.5).

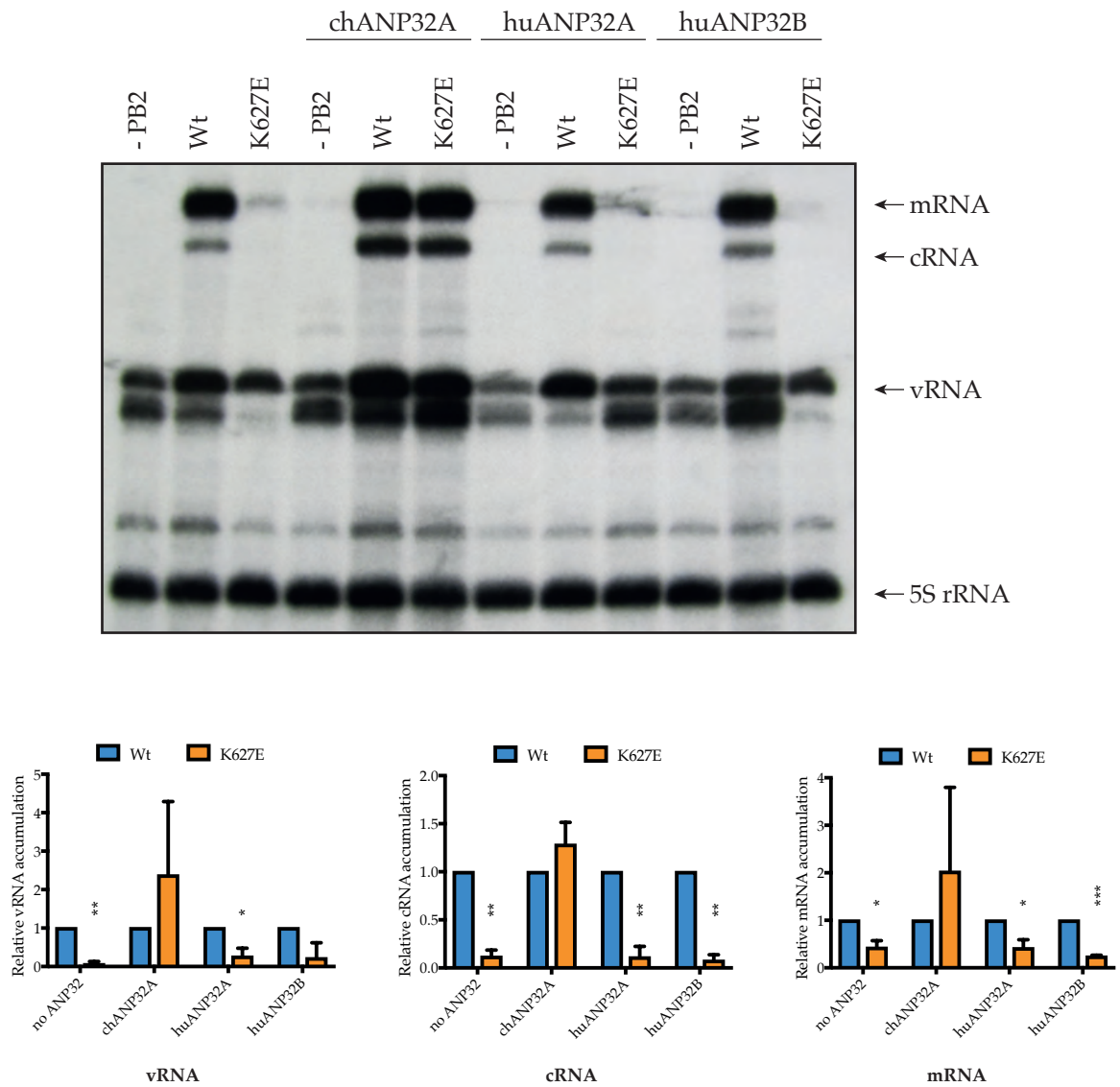


Fig. 4.5 Avian ANP32A rescues avian RdRP activity in mammalian cells in mini-replicon assays. Human HEK-293T cells were co-transfected with plasmids expressing PA, PB1, wild-type or avian-like PB2, NP, segment 6 vRNA and if indicated chANP32A, huANP32A or huANP32B. Accumulation of mRNA, cRNA, and vRNA was analysed by primer extension. 5S rRNA was used as an internal loading control. The graphs show the mean intensity signal relative to wild-type RdRP from three independent biological replicates ($n = 3$). Mean intensities were subtracted by the mean intensity of the negative control (-PB2). Error bars represent the standard error of the mean and the asterisks indicate a significant difference to wild-type (two-tailed one-sample t test) as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

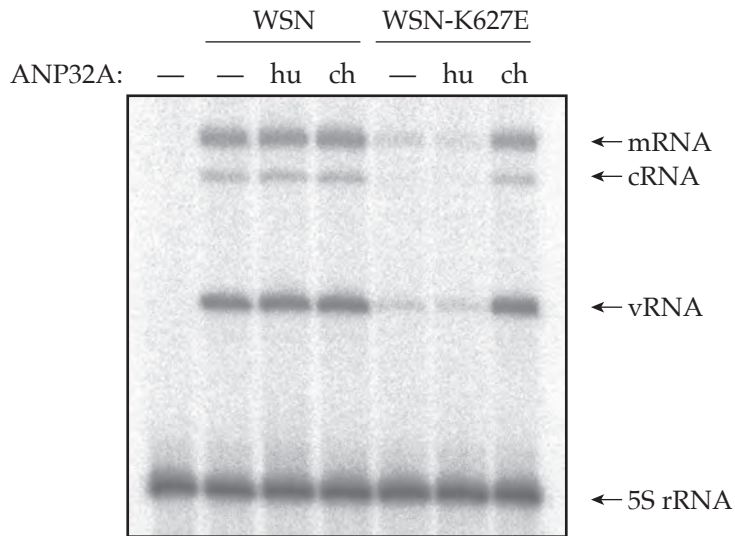
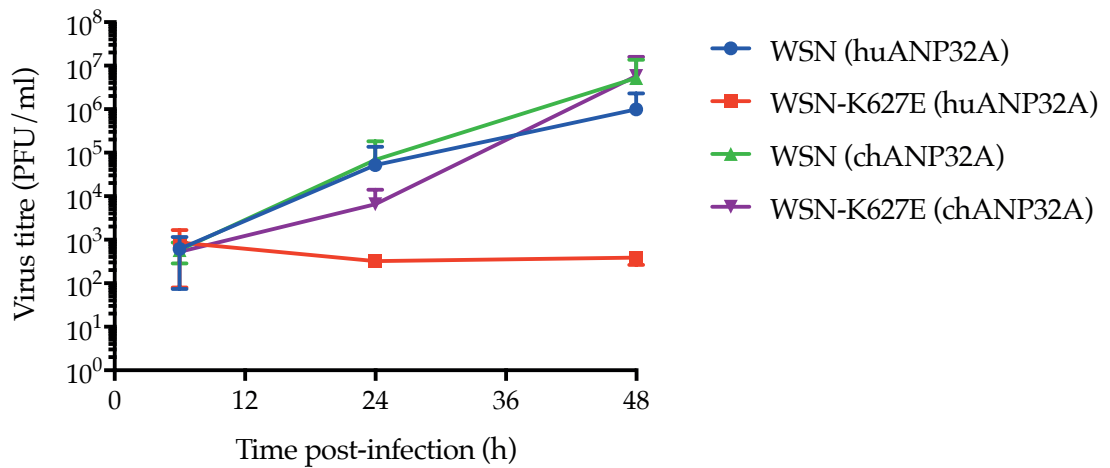
A**B**

Fig. 4.6 Avian ANP32A rescues avian influenza A virus replication and growth in infected mammalian cells. (A) chANP32A or huANP32A were expressed in human HEK-293T cells. 48 h post-transfection, cells were infected with WSN or WSN-K627E virus (MOI = 1). 6 h.p.i. the accumulation of mRNA, cRNA, and vRNA was analysed by primer extension. 5S rRNA was used as an internal loading control. (B) Growth curve of avian-like influenza A virus in the absence or presence of avian ANP32A. Human HEK-293T cells were transfected with plasmids expressing chANP32A or huANP32A. 24 h post-transfection cells were infected with WSN or WSN-K627E virus (MOI = 0.01). At the indicated time points, viral supernatants were collected and viral titres were determined using plaque assays. Error bars show the standard error of the mean (n = 3).

Similarly, infection of human HEK-239T cells with the mammalian-adapted WSN virus or the avian-like WSN-K627E virus resulted in a significantly enhanced viral RNA synthesis (Fig. 4.6A) and viral titres (Fig. 4.6B) by the avian-like influenza A virus if chANP32A had been overexpressed.

Overall, the data shows that avian ANP32A is a missing positive host factor required for efficient replication and transcription by avian influenza A viruses in mammalian host cells.

4.2.4 Avian ANP32A is required for efficient vRNA synthesis by the avian RdRP in mammalian hosts

It has been proposed in a previous study that avian influenza A viruses are unable to generate functional cRNPs that can act as bona fide templates for vRNA synthesis and that these cRNPs are more unstable than cRNPs generated by mammalian influenza A viruses (Mänz et al., 2012). In order to address the question whether avian influenza A viruses are restricted in cRNA or vRNA synthesis and whether avian ANP32A is able to restore efficient vRNA synthesis by avian influenza viruses in human host cells, the RdRP subunits and NP were co-expressed in human HEK-293T cells, followed by infection with either WSN or the avian-like WSN-K627E virus in the presence of the cellular transcription-inhibitor ACTD. Total RNA was extracted and viral RNAs were analysed by primer extension (Fig. 4.7). If no viral proteins were pre-expressed, neither a mammalian virus (lane 1) nor an avian-like virus (lane 2) was able to replicate or transcribe. If either a mammalian (lane 3) or avian-like RdRP (lane 5) with an active site mutation allowing RNA binding but no RNA synthesis (PB1a) was expressed, both viruses were able to stabilise cRNA equally well. This finding is supported by the studies showing that avian and mammalian vRNPs can stabilise similar amounts of cRNA independent of the nature

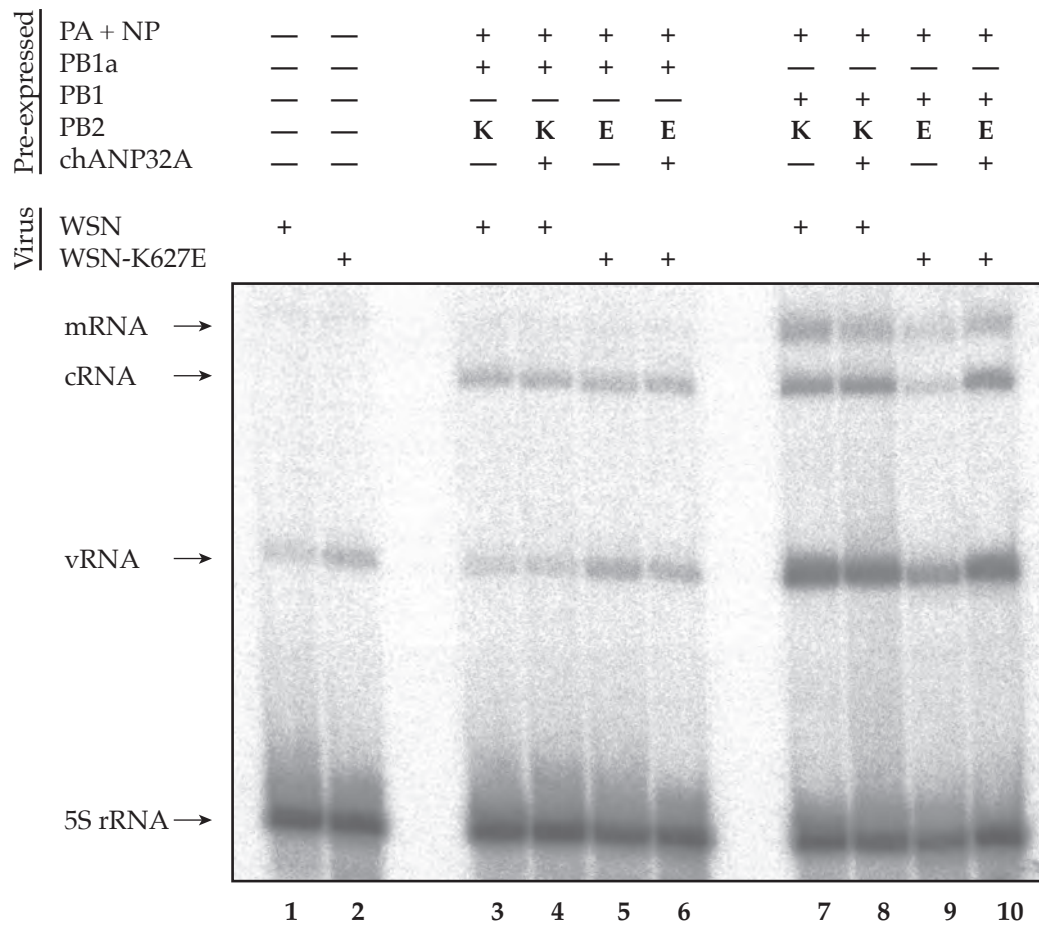


Fig. 4.7 Avian ANP32A rescues specifically avian influenza A virus vRNA synthesis in infected mammalian cells. Human HEK-293T cells were co-transfected with plasmids expressing NP, PA, PB1, wild-type (K) or avian-like (E) PB2 and chANP32A as indicated. A PB1 subunit with an active site mutation (PB1a) which allows RNA-binding but prevents RNA synthesis was included. 48 h post-transfection, cells were infected with influenza A/WSN/33 or influenza A/WSN/33-K627E (H1N1) virus at an MOI of 5 in the presence of actinomycin D. Accumulation of mRNA, cRNA, and vRNA 6 h post-infection was assessed by primer extension. 5S rRNA was used as an internal loading control.

of the pre-expressed RdRP (Mänz et al., 2012; Nilsson et al., 2017). Additional expression of chANP32A did not affect cRNA stabilisation of either virus (lanes 4 and 6). If a catalytically active mammalian RdRP was expressed in the absence (lane 7) or presence (lane 8) of chANP32A, high levels of vRNA, cRNA and mRNA were observed following infection with WSN. However, if a catalytically active avian-like RdRP was expressed, infection with the avian-like WSN-K627E virus resulted in a reduced ability to replicate and transcribe compared to the mammalian-adapted virus (lane 9). This effect was reversed by additional expression of chANP32A (lane 10), which resulted in vRNA, cRNA and mRNA levels comparable to infection with wild-type IAV (lanes 7 and 8).

Overall, this data shows that avian influenza A viruses are not restricted in synthesis and encapsidation of cRNA, but are defective in generating vRNA in mammalian host cells. However, expression of avian ANP32A enabled the avian influenza A virus to overcome its restriction and replicate efficiently.

4.2.5 ANP32A enhances vRNA synthesis *in vitro*

While the avian-like RdRP is not restricted in its *in vitro* activity compared to the mammalian RdRP (Fig. 3.5 and 3.6), those findings do not exclude the possibility that vRNA synthesis in general is inefficient and requires ANP32A to stimulate this process. Indeed, a recent study showed that both human ANP32A and ANP32B play a role in influenza virus replication by upregulating vRNA synthesis (Sugiyama et al., 2015).

In order to determine whether ANP32A stimulates vRNA synthesis *in vitro* and whether ANP32A exhibits a species specificity towards the viral RdRP, chANP32A, huANP32A or huANP32B fused to an N-terminal Strep tag were expressed in human HEK-293T cells and purified by Strep-Tactin Sepharose chromatography. Purified proteins were analysed by SDS-PAGE followed by silver staining and western blot with an antibody raised

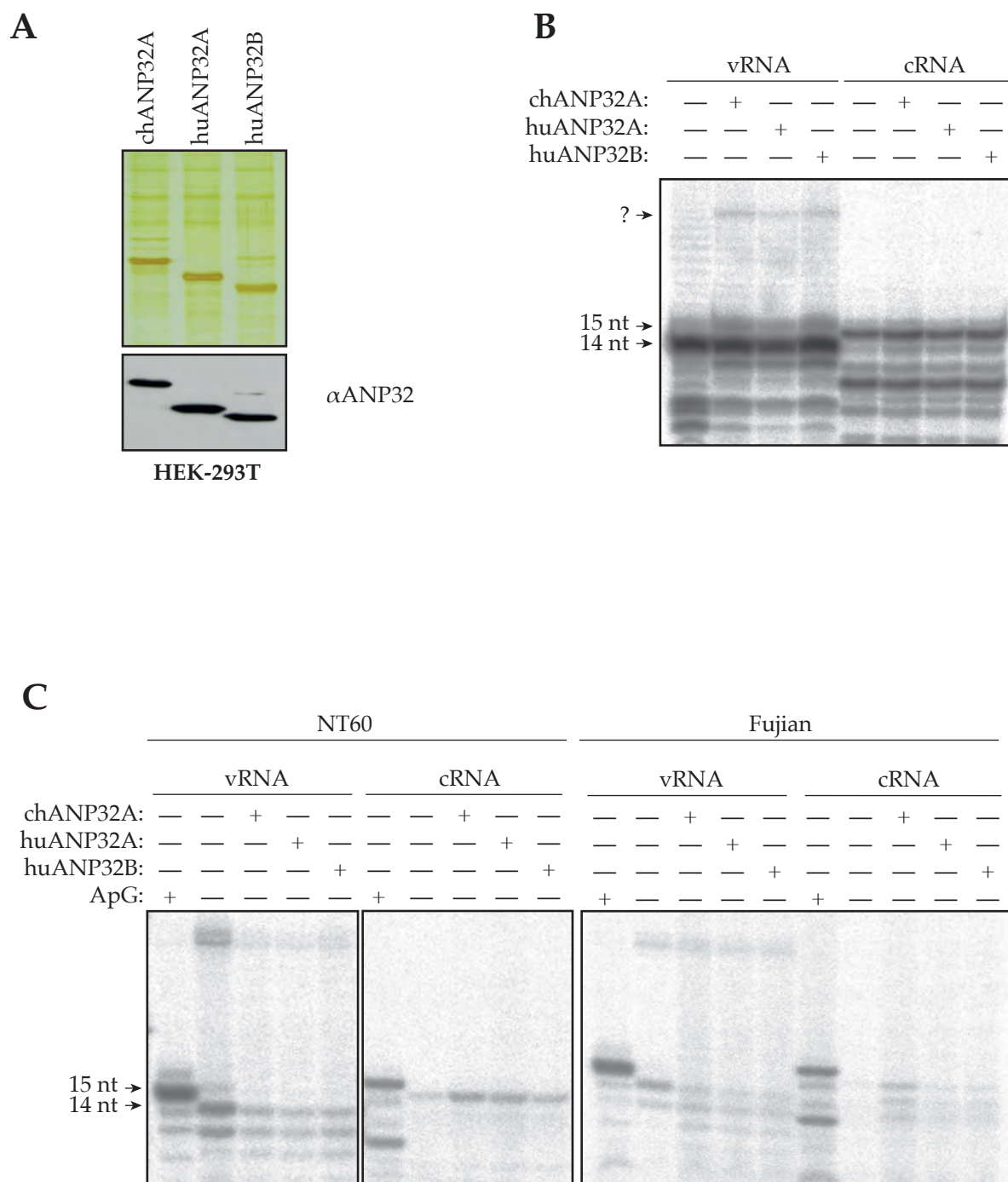


Fig. 4.8 ANP32A and ANP32B enhance *de novo* vRNA synthesis *in vitro*. (A) ANP32A and ANP32B were expressed in human HEK-293T cells. Proteins were purified using Step-Tactin Sepharose chromatography and analysed by SDS-PAGE and silver staining or western blotting. (B) Purified recombinant viral RdRP (influenza A/Northern Territory/60/1968) was incubated in the presence of ApG primers, nucleotides and ANP32A or ANP32B. Reaction products were analysed by 20% PAGE and autoradiography. (C) Purified recombinant human influenza virus RdRP (influenza A/Northern Territory/60/1968) or avian influenza virus RdRP (influenza A/duck/Fujian/01/2002) was incubated in the presence of ApG primers (as indicated), nucleotides and ANP32A or ANP32B. Reaction products were analysed by 20% PAGE and autoradiography.

against the ANP32 family, revealing successful isolation of ANP32A and ANP32B from transfected cells (Fig. 4.8A).

Purified recombinant ANP32A or ANP32B was then incubated with baculovirus-expressed purified viral RdRP (influenza A/Northern Territory/60/1968 (H3N2) virus) bound to vRNA or cRNA promoter in the presence of ApG primers and nucleotides (Fig. 4.8B). No enhancement or inhibition of ApG-primed initiation was observed in the presence of ANP32A or ANP32B.

In order to determine whether ANP32A enhances primer-independent initiation, purified recombinant ANP32A or ANP32B was incubated with baculovirus-expressed purified human-adapted RdRP (influenza A/Northern Territory/60/1968 (H3N2) virus) or avian RdRP (influenza A/duck/Fujian/01/2002 (H5N1) virus) bound to vRNA or cRNA promoter in the presence of nucleotides (Fig. 4.8C). On a vRNA promoter template addition of ANP32A or ANP32B appeared to decrease both primer-independent RdRP initiation and cap-dependent transcription. Capped RNA is a known contaminant in viral RdRP purification and these capped RNAs subsequently are able to act as capped primers for viral transcription in vRNA-bound RdRPs. However, on a cRNA promoter template human RdRP activity was significantly enhanced by the addition of avian and human ANP32A and ANP32B, while the avian RdRP activity was only enhanced by avian ANP32A. This is in agreement with published data, suggesting that both ANP32A and ANP32B enhance cRNA-bound human-adapted RdRP *de novo* initiation (Sugiyama et al., 2015).

Taken together, these findings are in agreement with previously published studies, that the host factors ANP32A and ANP32B are enhancers of viral RdRP activity that specifically stimulate *de novo* initiation of vRNA synthesis. Furthermore, avian ANP32A exerts a significantly stronger effect than human ANP32A which can be linked to its stronger affinity for the viral RdRP compared to human ANP32A (Fig. 4.3).

4.3 Discussion

In this study, the interactomes of human and chicken ANP32A were characterised for the first time using mass spectrometry. Importin- α was found to be one of the main binding partners of ANP32A and a significant increase in binding of KPNA2 was observed after influenza A virus infection. ANP32A was shown to bind to the viral heterotrimeric RdRP and the avian isoform appeared to have a stronger affinity for the viral RdRP than the human ANP32A isoform. In mammalian host cells it was shown that expression of avian ANP32A was able to restore avian RdRP activity both in mini-replicon assays and in infection. Avian influenza viruses were shown to be restricted in vRNA but not cRNA synthesis in mammalian host cells and this defect could be reversed in the presence of avian ANP32A. Correspondingly, a stimulatory effect on vRNA synthesis was observed *in vitro* by both human and chicken ANP32A. Notably though, stimulation of the avian RdRP was only observed in the presence of chANP32A.

Importin- α is a well-known host factor required for the nuclear import of viral protein complexes. Nuclear import of NP and vRNP complexes is mediated by KPNA1, KPNA2 and KPNA4 (Cros et al., 2005; Melen et al., 2003; O'Neill & Palese, 1995; Wang et al., 1997). The viral RdRP subunits PA and PB1 form a dimer in the cytoplasm and are imported as a dimer into the nucleus through an interaction with RanBP5 (Deng et al., 2006a). Nuclear import of PB2 has been shown to take place separately from PA and PB1 and has been linked to host adaptation of the viral RdRP (Deng et al., 2005; Gabriel et al., 2008). PB2 subunits from avian influenza viruses appear to preferentially interact with KPNA4 (Gabriel et al., 2011), whereas human adaptation induced preferential binding of PB2 to KPNA2 and KPNA6 (Gabriel et al., 2008; 2011; Hudjetz & Gabriel, 2012; Resa-Infante et al., 2014). In this study it was demonstrated that ANP32A strongly interacts with KPNA1 and KPNA6. This agrees with yeast-two-hybrid experiments showing interaction between ANP32A and KPNA1, KPNA5 and KPNA6 (Lim et al., 2006). The fact that these

interactions remain unchanged after infection with influenza A virus suggests no specific function of these interactions in viral infection and suggest that ANP32A is imported into the nucleus through these nuclear import factors. However, a strong increase in binding to KPNA2 was seen after infection with a human-adapted influenza A virus. Importin- α has been proposed to be required for more than nuclear import of viral proteins in viral infections and that it could be required for recruitment of host factors needed for viral replication (Hudjetz & Gabriel, 2012). In light of all these findings it is tempting to speculate that association of PB2 with KPNA2 is needed to recruit ANP32A to the viral RdRP. It would be highly informative to repeat this experiment with an avian-adapted influenza A virus in order to see whether a similar upregulation of importin- α binding can be seen and whether there is a specificity for a different importin- α isoform.

In this chapter it was also demonstrated that ANP32A interacts with the viral heterotrimeric RdRP, which has been reported previously in large-scale proteomic studies (Bradel-Tretheway et al., 2011; York et al., 2014) as well as co-expression studies (Domingues & Hale, 2017; Sugiyama et al., 2015). One study observed an interaction between ANP32A and an avian H5N1 RdRP only if treated with nucleases (Bradel-Tretheway et al., 2011). Supporting this, another study found that ANP32A only interacts with the full heterotrimeric RdRP and only if the RdRP is not in the context of an RNP (Sugiyama et al., 2015). Seemingly in contrast to this, ANP32A was also found to interact with RNPs, however this study was not able to differentiate between isolated RNPs, heterotrimeric RdRPs and monomeric PB2 subunits, it is therefore likely that the interaction with ANP32A was only detected when the RdRP was isolated and not RNPs or monomeric PB2 (York et al., 2014). From structural studies it is known that PB2 and the viral RdRP can exist in various different conformations depending on their interaction partners and RNA substrates (Tarendeau et al., 2007; 2008; Hengrunga et al., 2015; Pflug et al., 2014; Thierry et al., 2016). It is therefore possible that the interaction between the viral

RdRP and ANP32A is facilitated only when the C-terminal domains of PB2 are found in the specific structural arrangement which it exists in in the RNA-free heterotrimeric RdRP. Due to the similar structures of the RNA-free apo RdRP and the cRNA-bound RdRP it could be possible that ANP32A also interacts with the resident RdRP of cRNPs. In this chapter mass spectrometry analysis showed a strong enrichment of the viral RdRP over NP in the ANP32A interactome which suggests an interaction of ANP32A with RNA-free RdRP as opposed to vRNPs. One could speculate that this result still allows for an interaction of ANP32A with cRNPs, due to the similar conformational arrangements of the viral RdRP in the RNA-free and cRNA-bound forms and the fact that cRNPs are significantly less abundant in the cell than vRNPs. Regardless, this binding specificity points to a specific function of ANP32A in viral replication linked to RNA-free RdRP and possibly also cRNPs.

Furthermore, it was also shown here that chANP32A forms a stronger interaction with the viral RdRP than huANP32A, which is in line with previous findings (Domingues & Hale, 2017). The interaction of ANP32A with the viral RdRP in general takes place in the disordered acidic C-terminal LCAR region of ANP32A. The enhanced binding ability of avian ANP32A though is specifically linked to the unique 6 amino acid-long SIM-like sequence in the avian isoform, which also is responsible for the enhancement of the avian RdRP activity (Domingues & Hale, 2017). One could speculate, that a prolonged LCAR region helps stabilising an interaction between ANP32A and the viral RdRP. However, a potential interaction with a SUMOylated host factor through the SIM-like sequence in avian ANP32A isoforms could also be involved in facilitating the enhanced binding of avian ANP32A to the viral RdRP.

A recent study has suggested that both human ANP32A and ANP32B play a role in influenza virus replication by up regulating vRNA synthesis (Sugiyama et al., 2015). Silencing of ANP32A and ANP32B has been shown to significantly reduce viral

replication and transcription both in human and avian cells (Long et al., 2016; Sugiyama et al., 2015). These observations together with the findings presented in this chapter suggest that ANP32A and ANP32B are in general required for viral RdRP activity and that avian ANP32A specifically is required for avian RdRP activity.

Several studies have shown that the avian RdRP is not restricted in its activity to make vRNA or cRNA in comparison to a mammalian-adapted RdRP *in vitro* (Aggarwal et al., 2011; Nilsson et al., 2017; Paterson et al., 2014). Indeed, the avian RdRP even displays a higher activity at higher temperatures compared to a mammalian RdRP, which concurs with the fact that the avian intestinal tract, the site of viral replication in birds, has a higher internal temperature than the respiratory tract of mammals, where viral replication takes place in mammalian hosts (Aggarwal et al., 2011). Nevertheless, in mammalian hosts avian influenza viruses are critically restricted as has been shown before as well as in this chapter. This defect has been linked to an inability to efficiently synthesise vRNA (Mänz et al., 2012). The authors concluded from their results that avian influenza viruses produce defective cRNP structures in mammalian cells, which cannot act as bona fide templates for vRNA synthesis. In this study, it was confirmed that avian influenza viruses are restricted in vRNA synthesis, but that this defect can be reversed by expression of avian ANP32A. Taking the newest findings about the viral RdRP and viral replication into account, we propose here in an alternative explanation for the observed defect in vRNA synthesis that avian cRNPs can in fact act as bona fide templates and that avian cRNA-bound RdRPs instead are inherently inefficient in generating vRNA and have evolved to utilise the avian ANP32A host factor to support this process. In mammalian hosts, the shorter mammalian isoform of ANP32A is not able to efficiently fulfil the same role as avian ANP32A in viral replication, resulting in evolutionary pressure for adaptive mutations which compensate for the lack of ANP32A in vRNA synthesis.

This is supported by evidence showing that the PB2-627E-mediated host restrictive effect can be overcome if mutations in the 3' vRNA promoter are introduced, resulting in a 5' cRNA terminus that is identical to the 5' terminus of the vRNA template (Paterson et al., 2014). This suggests that avian influenza virus RdRPs can replicate efficiently in mammalian host cells, if they do not generate and bind to viral RNAs that resemble the 5' cRNA terminus. In other words, if these mutations are introduced the avian RdRP can replicate efficiently because it always binds a 5' terminus that is identical to the vRNA promoter. Considering that the influenza virus RdRP exists in significantly different conformations, depending on the nature of the viral RNA template (Hengrung et al., 2015; Pflug et al., 2014; Thierry et al., 2016), and considering that it is known that vRNA and cRNA initiation is carried out differently (Velthuis & Fodor, 2016), it is likely that avian influenza virus RdRPs require ANP32A for a process which is specific to vRNA synthesis, e.g. internal initiation, realignment of the initial dinucleotide or for *trans*-activation of the viral RdRP. A previous study already described ANP32A as an enhancing factor for vRNA synthesis *in vitro* (Sugiyama et al., 2015). In this chapter it was shown that the stimulatory effect of ANP32A is linked to *de novo* replication initiation on a cRNA template specifically, but not ApG-primed extension or replication initiation on a vRNA template. Additionally, vRNA synthesis by an avian RdRP was only significantly enhanced by avian ANP32A. This finding is supported by the previous findings that avian ANP32A interacts more strongly with the viral RdRP than human ANP32A, which could allow it to be a stronger enhancer of avian RdRPs.

Together this suggests that while influenza virus RdRPs can synthesise vRNA to a certain extent *in vitro*, this process is inefficient in a cellular environment and requires the presence of ANP32A. Both human and avian ANP32A interact with the avian and human-adapted RdRPs, possibly through some sort of interaction mediated by KPNA2. However, avian ANP32A binds more strongly to the viral RdRP than human ANP32A, possibly

through a SUMOylated host factor. Avian RdRPs have evolved to utilise the longer avian isoform of ANP32A and a possible SUMOylated host factor for *de novo* vRNA synthesis and therefore are inefficient in vRNA synthesis in mammalian host cells, leading to evolutionary pressure for host adaptive mutations to occur in the avian RdRP that can compensate for the lack of avian ANP32A in mammalian host cells. It is known that cRNA synthesis and vRNA synthesis are two distinct processes in influenza virus replication. Synthesis of vRNA requires internal initiation and a *trans*-activating RdRP, as described before. A role of ANP32A in realignment of the internal initiation product is unlikely, since no major RNA products are observed in the absence of ANP32A that could correspond to failed realignment. However, a role for ANP32A in *trans*-activation could be imagined, as it has been shown to interact with RNA-free RdRP and also possibly also could interact with cRNA-bound RdRPs.

4.4 Summary

In summary, the first comprehensive ANP32A interactome was described in this chapter, revealing a strong association of ANP32A with members of the importin- α family. After infection a significantly enhanced interaction with KPNA2 was described for both human and chicken ANP32A. ANP32A was also confirmed to interact with the viral RdRP. Avian ANP32A was confirmed to be the missing positive host factor that avian influenza viruses require for efficient viral replication in mammalian host cells. Specifically, it was shown that avian influenza A viruses were not restricted in their ability to generate cRNA in mammalian host cells, but that avian ANP32A is required for efficient vRNA synthesis. Furthermore, the role of ANP32A was narrowed down to *de novo* initiation on a cRNA template, suggesting that ANP32A could play a role in *trans*-activation of cRNPs.

Chapter 5: Characterising the viral polymerase dimer interface

5.1 Introduction

It has been reported that influenza virus RdRPs can form oligomeric complexes (Jorba et al., 2008). While the question of the biological function of these RdRP dimers or oligomers has not been fully answered yet, it is tempting to speculate that one or more polymerase-polymerase interactions are required for viral replication. For example, during replication nascent RNA strands need to be bound by a free RdRP (Vreede et al., 2004), a process which could be mediated by an interaction between the resident RdRP of the replicating RNP and an RNA-free RdRP. Supporting the hypothesis that RNA-free RdRPs interact with RNPs, it has been shown that purified cRNPs require additional RdRP *in vitro* for efficient vRNA synthesis, suggesting a *trans*-activating interaction (York et al., 2013). Recent structural studies have reported that an influenza A virus RdRP complex consisting of PA, PB1 and the N-terminus of PB2 forms dimeric structures in solution and that its association with vRNA and cRNA regulates its transition into monomeric or tetrameric states (Chang et al., 2015). Unpublished work from our group by Haitian Fan supports these findings that the RdRP can exist in a dimeric state. Crystal structures of the avian influenza A/duck/Fujian/01/2002 (H5N1) virus RdRP and the human influenza A/Northern Territory/60/1968 (H3N2) virus RdRP reveal dimeric complexes in the absence of any viral promoter RNA (Fig. 5.1).

The study observing dimeric and tetrameric RdRP structures (Chang et al., 2015) proposes two different interaction interfaces. One stable interface (from now on called interface 1) was described between the C-terminus of PA (approximately residues 293 to 355) and the N-terminus of PB2 (approximately residues 86 to 130) of two RdRP monomers. Furthermore, a weaker interface (from now on called interface 2) was determined to be formed by the finger domains of PB1 (approximately residues 180 to 205 and residues 355 to 375). In addition to these published interaction surfaces, the structural studies from our group also reveal two alternative RdRP interactions. One matches interface 1 from the published study (Fig. 5.1.A), and one interface (from now on called interface 3) is formed between the PA N-terminus (approximately residues 51 to 115) and parts of the PB2 mid-link, cap-binding, NLS and 627-domains of two RdRP monomers (Fig. 5.1B). The ability of viral heterotrimeric RdRPs to form oligomeric RdRP complexes was presented in Figure 3.12.

Interestingly, many residues which have previously been reported to be able to harbour host adaptations, including PB2-627, are found to cluster around the potential RdRP dimer interfaces (Fig. 5.2 and Table 5.1). While it is still unknown, whether these potential dimer interfaces are biologically relevant, one could speculate that host range restriction is linked to the role of RdRP dimers in viral replication. This is especially tempting, considering that in Chapter 4 it was shown that avian influenza viruses are restricted specifically in vRNA synthesis and that it had already been shown before that cRNPs require *trans*-activation by a second RdRP for efficient vRNA synthesis (York et al., 2013).

In order to test the role of potential RdRP dimer interfaces in polymerase function, point mutations and an internal truncation targeting these interfaces were introduced into the viral RdRP and their effect was analysed using a combination of *in vitro* activity and cellular mini-replicon assays.

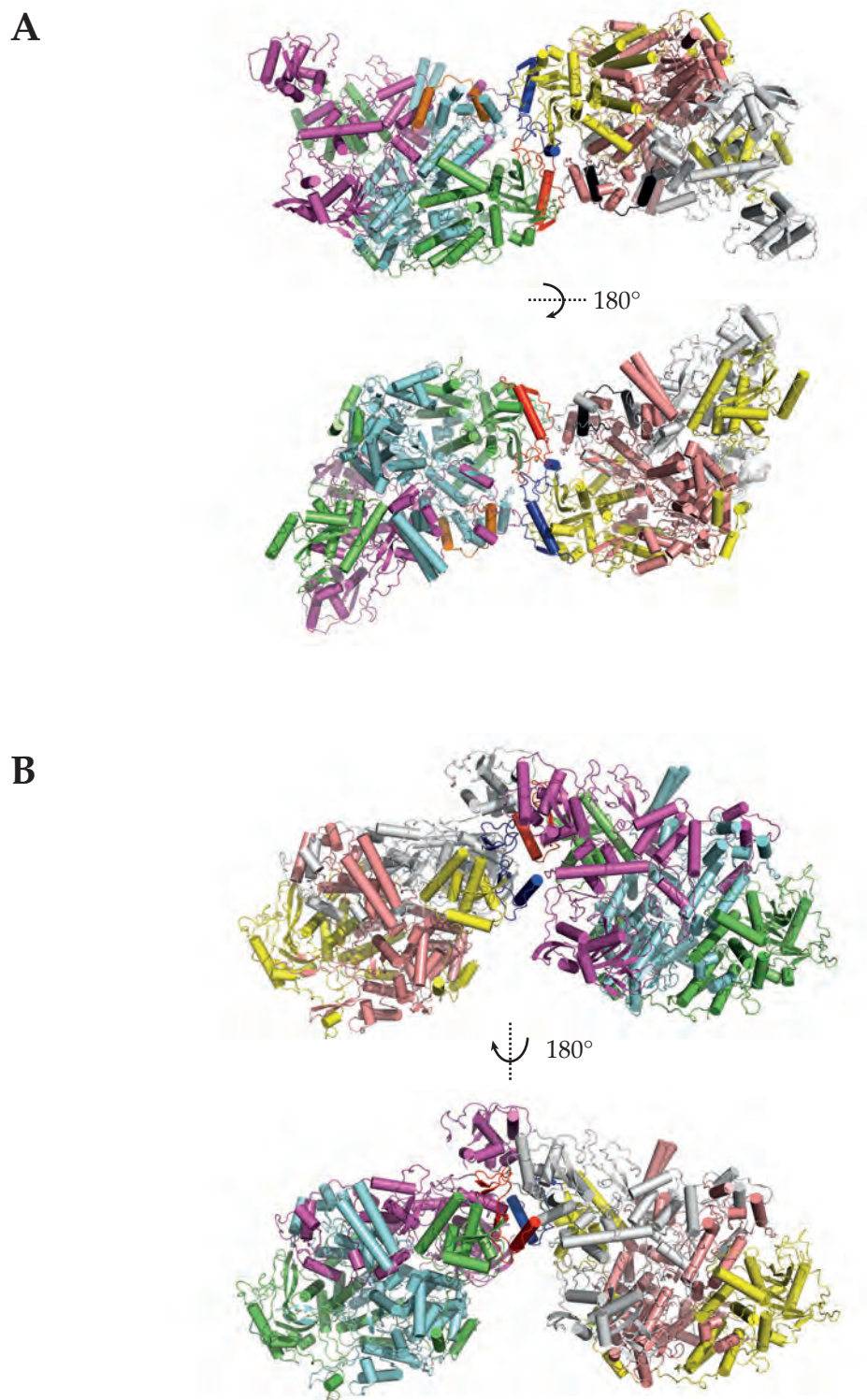
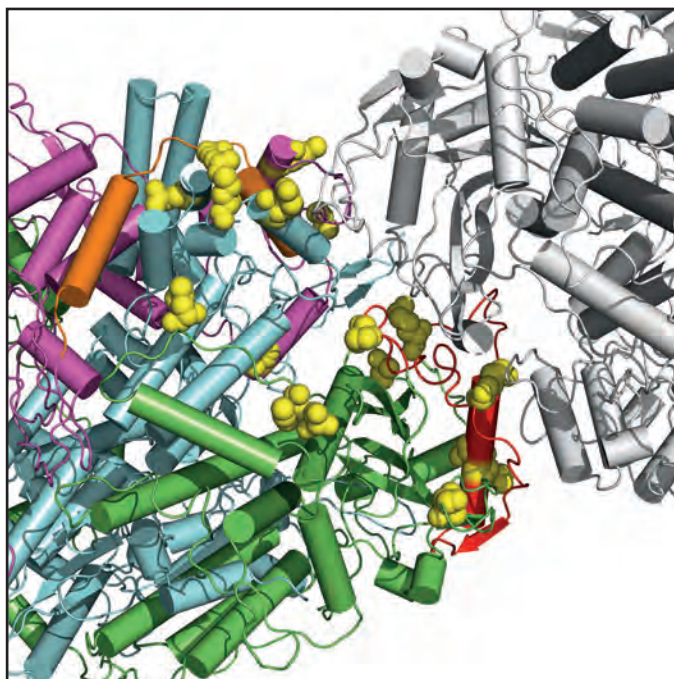


Fig. 5.1 Crystal structure of the influenza A virus RdRP dimer interface. Crystal structures of two influenza A/Northern Territory/60/1968 (H3N2) virus heterotrimeric RdRP monomers forming a dimeric complex through (A) dimer interface 1 and (B) dimer interface 3. PA (green/yellow), PB1 (cyan/grey) and PB2 (purple/grey) are displayed in the same colour scheme throughout this figure. (A) PA residues 29-355 are highlighted in red and dark blue and PB2 residues 8-130 in orange and black in interface 1. (B) PA residues 51-115 are highlighted in red and dark blue in interface 3. This work is unpublished work by Haitian Fan (Sir William Dunn School of Pathology, University of Oxford).

A



B

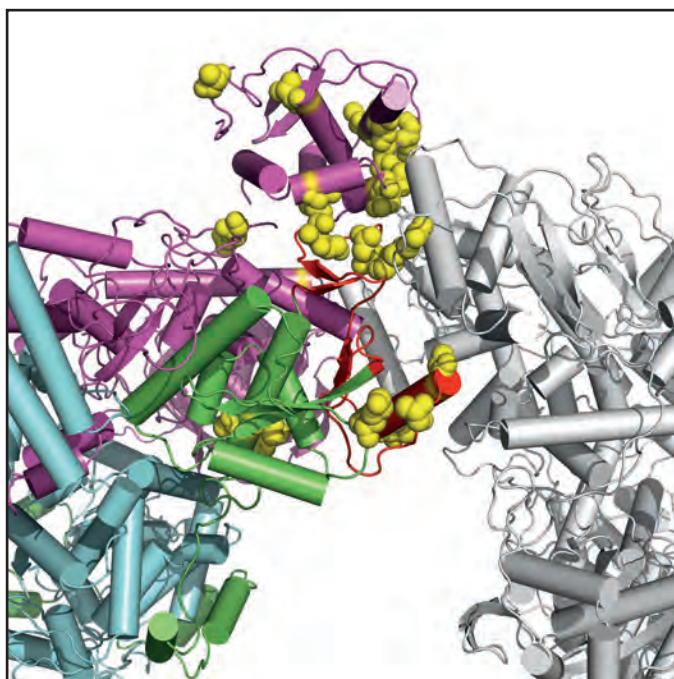


Fig. 5.2 Host adaptive mutations in the influenza A virus RdRP dimer interface. Zoomed in view of the influenza A/Northern Territory/60/1968 (H3N2) virus RdRP dimer (A) interface region 1 and (B) interface region 3 as shown in Figure 5.1. PA (green), PB1 (cyan) and PB2 (purple) or highlighted in colours on one RdRP monomer, while the other RdRP monomer is shown in grey. Known host adaptive mutations in the dimer interface regions (see Table 5.1) are highlighted as yellow spheres.

Table 5.1 Host adaptive mutations at the RdRP dimer interfaces

Subunit	Residue	Adaptive mutation (avian to human)	RdRP dimer interface
PA	55	D55N	2
PA	57	R57Q	2
PA	65	S65L	2
PA	66	G66D	2
PA	85	T85I	2
PA	97	T97I	2
PA	100	V100A	2
PA	142	K142N	2
PA	295	L295P	1
PA	312	K312R	1
PA	336	L336M	1
PA	337	A337S	1
PA	356	K356R	1
PA	421	S421I	1
PA	423	M423I	1
PA	476	V476A	1
PA	552	T552S	1
PB1	581	E581D	1
PB1	584	R584Q	1
PB1	598	L598P	1
PB1	621	Q621R	1
PB2	44	A44S	1
PB2	64	M64T	1
PB2	81	T81M	1
PB2	105	T105V	1
PB2	271	T271A	2
PB2	559	T559I	2

Subunit	Residue	Adaptive mutation (avian to human)	RdRP dimer interface
PB2	567	D567N	2
PB2	569	T569A	2
PB2	588	A588I	2
PB2	590	G590S	2
PB2	591	Q591R	2
PB2	598	V598T	2
PB2	613	V613T	2
PB2	627	E627K	2
PB2	674	A674T	2
PB2	676	T676I	2
PB2	682	G682S	2
PB2	684	A684S	2

5.2 Results

5.2.1 The N-terminus of PA is required for vRNA synthesis

In order to address whether the observed dimer interface 3 plays a biological role in polymerase function, a plasmid encoding the PA gene derived from influenza A/WSN/33 (H1N1) with an internal truncation in the N-terminal region associated with part of dimer interface 3 (51-72) was generated using site-directed PCR mutagenesis and a GGS-linker was inserted instead (Fig. 5.3A-B). Recombinant RdRPs were expressed in human HEK-293T cells and purified from cell lysates using a protein A tag on PB2. Purified protein complexes were analysed by SDS-PAGE and Western blotting (Fig. 5.3C).

First, the ability of this mutant RdRP to carry out core RdRP functions *in vitro* was assessed. To assess cap-snatching and cap-dependent primer extension, purified RdRP was bound to vRNA promoter and incubated with β -globin mRNA and nucleotides. Both wild-type and PA Δ 51-72 RdRP produced a cap-dependent transcription product, showing that transcription is not affected *in vitro* by deletion of PA residues 52 to 73 (Fig. 5.3D). Next, purified RdRP was bound to vRNA or cRNA promoter and incubated in the presence of ApG primers and nucleotides in order to analyse its ability to extend an ApG primer as a measure for viral replication. On a vRNA promoter template, both wild-type and PA Δ 51-72 RdRP were able to produce a 14-nucleotide-long RNA product (Fig. 5.3E, left panel). On a cRNA promoter template however, ApG-primed vRNA synthesis by PA Δ 51-72 RdRP was strongly inhibited compared to wild-type RdRP (Fig. 5.3E, right panel). These results suggest that residues 52 to 73 of PA are required for cRNA-to-vRNA replication, while not being important for vRNA-to-cRNA replication or cap-snatching *in vitro*.

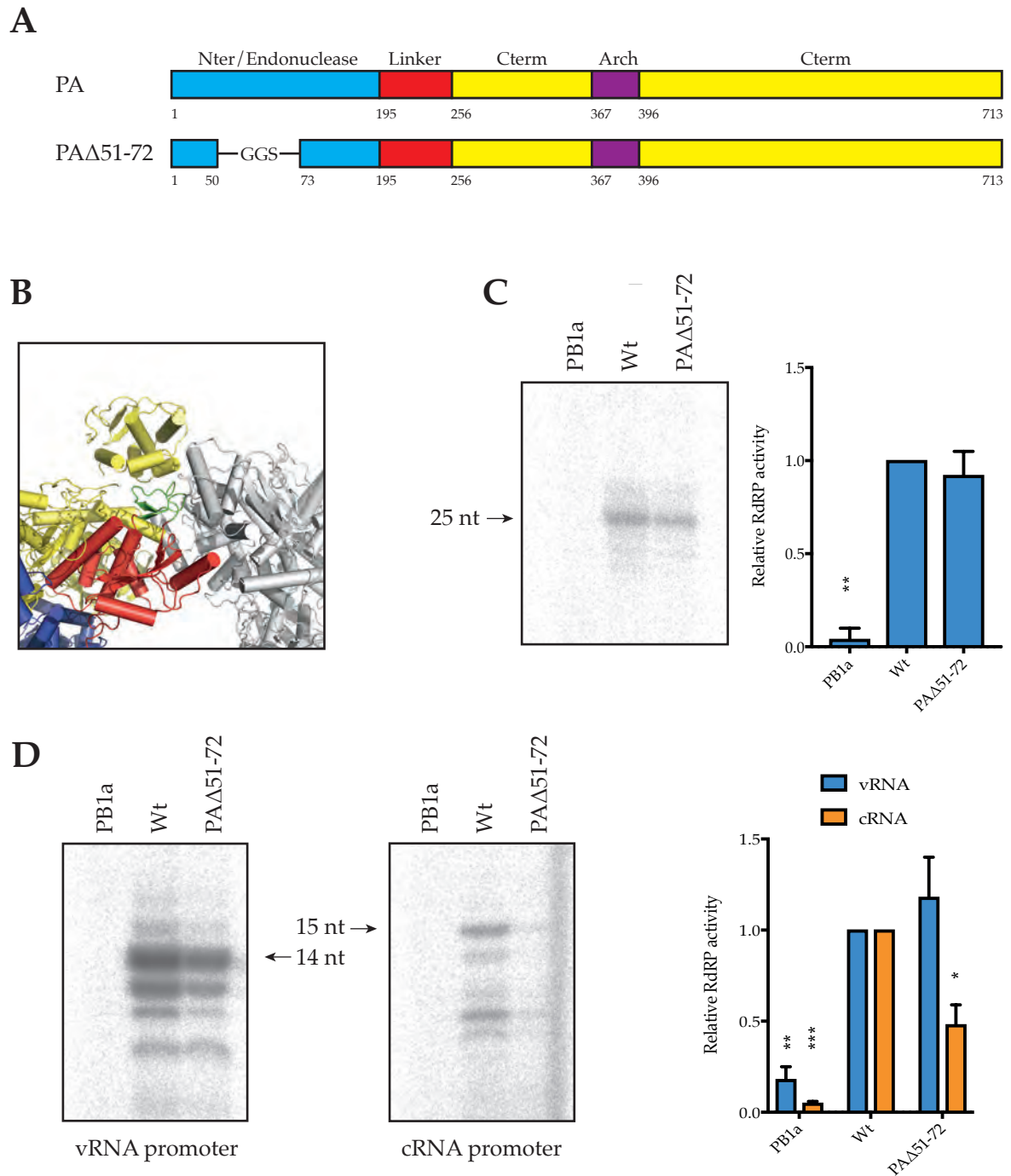


Fig. 5.3 Polymerase lacking residues 51 to 72 of PA is defective in ApG-primed vRNA synthesis *in vitro*. (A) Schematic of the wild-type and mutant PA subunit. (B) Zoomed in view of the RdRP dimer interface 3 as shown in Figure 5.1. One RdRP monomer is shown in grey, the other RdRP is shown in red (PA), blue (PB1) and yellow (PB2). PA residues 51-72 are highlighted in green. (C-D) Purified recombinant RdRP was incubated in reaction mixtures containing ATP, UTP, CTP, [α - ^{32} P] GTP and (C) β -globin mRNA and vRNA promoter template or (D) ApG primer and either vRNA or cRNA promoter template. Reaction products were analysed by 20% PAGE and autoradiography. The graphs show the mean intensity signal relative to wild-type RdRP from three independent biological replicates ($n = 3$), with error bars representing the standard error of the mean and the asterisks indicating a significant difference to wild-type (two-tailed one-sample t test) as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Having shown that PA Δ 51-72 RdRP is active in transcription and vRNA-to-cRNA replication *in vitro*, mini-replicon assays were used to investigate whether similar effects could be observed in a cellular environment. RNPs were reconstituted by co-expression of the three RdRP subunits, NP and segment 6 vRNA in human HEK-293T cells and the accumulation of positive- and negative-sense viral RNA was analysed by primer extension. RNPs containing PA Δ 51-72 were fully restricted in these mini-replicon assays, in contrast to the findings observed *in vitro*, where PA Δ 51-72 RdRP was still able to function in cap-snatching and transcription (Fig. 5.4A). The lack of mRNA signal in the mini-replicon assay could be caused by the defect in replication though. To determine whether transcription could be restored, RNPs were *trans*-complemented with either a transcription-deficient (PA-D108A) or replication-deficient (PA-C95A) PA subunit in a mini-replicon assay (Fig. 5.4B). When *trans*-complemented with a transcription-deficient PA subunit, PA Δ 51-72 RdRP was able to generate significant levels of mRNA, while *trans*-complementation with a replication-deficient PA subunit did not enhance vRNA or cRNA signals.

All these findings suggest that PA residues 52 to 73 are involved in viral replication but not transcription. As presented in Figure 5.3B, PA residues 52 to 73 are predicted to be involved in formation of heterotrimeric RdRP dimer structures. To test this hypothesis, wild-type or mutant PA, PB1, PB2 and PB2-TAP were co-expressed in human HEK-293T cells and purified using the protein A tag on PB2 as already presented in Figure 3.12. Analysis of purified protein complexes by SDS-PAGE and western blotting for PB2 revealed that PB2 monomers or PB2/PA dimers cannot form a stable interaction, therefore preventing co-purification of PB2-PB2 or PB2/PA-PB2/PA complexes. However, heterotrimeric RdRPs containing both wild-type PA or PA Δ 51-72 were able to form stable interactions allowing co-purification of a dimeric RdRP complex, as seen by the presence of two PB2 bands, one containing the remains of a cleaved TAP tag and one being

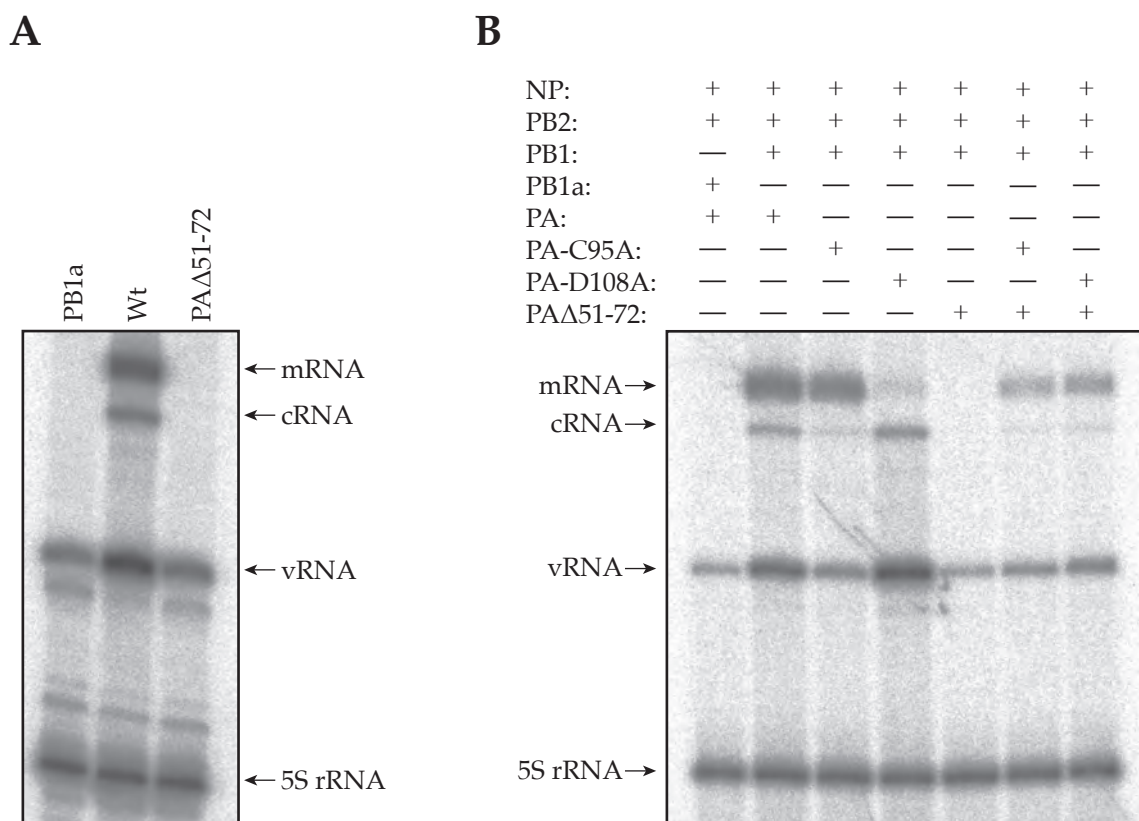


Fig. 5.4 PA residues 51 to 72 are required for replication but not transcription in the cell. (A) The effect of PA Δ 51-72 on RNP activity in human cells was assessed by co-expressing wild-type or mutant PA, PB1, PB2, NP and segment 6 vRNA in HEK-293T cells. (B) The effect of PA Δ 51-72 on replication or transcription specifically was analysed by co-expressing wild-type or mutant PA, PB1, PB2, NP and segment 6 vRNA in HEK-293T cell. Mutant PA was transcomplemented with replication-deficient (PA-C95A) or transcription-deficient (PA-D108A) PA subunit as indicated. The accumulation of viral RNAs was analysed by primer extension. 5S rRNA was used as an internal loading control.

untagged (Fig. 5.5A). These results confirm that RdRPs can form stable dimeric structures in the absence of NP and viral RNA. Furthermore, PA residues 52 to 73, implicated in the RdRP dimer interface 3, are not critically required to form this interaction.

RdRP dimers are thought to occur when free RdRP interacts with RNPs in order to *trans*-activate RNPs or to encapsidate the nascent product strands (Vreede et al., 2004; York et al., 2013). While the findings described above suggest that PA Δ 51-72 does not interfere with formation of RdRP dimers in a NP- and viral RNA-free environment in general, it is possible that in the context of a viral infection RdRP dimers are formed and regulated in a more nuanced manner. In order to assess whether PA Δ 51-72 RdRP can bind to nascent cRNA strands generated by replicating vRNPs, RdRP subunits (containing the active-site mutant PB1a) and NP were co-expressed in human HEK-239T cells, prior to infection with WSN virus in the presence of the cellular transcription-inhibitor ACTD. The accumulation of cRNA was analysed by primer extension. Pre-expressed wild-type and PA Δ 51-72 RdRP were both able to encapsidate and stabilise cRNA in an infection, suggesting that residues 51-72 of PA, and therefore possibly RdRP dimer interface 3, are not involved in the encapsidation process of cRNAs (Fig. 5.5B).

In a previous study it was shown that purified cRNPs are unable to efficiently synthesise vRNA *in vitro*, unless additional RdRP is provided, pointing to a *trans*-activating function for RNA-free RdRP in cRNP activity (York et al., 2013). *In vitro*, my findings suggested that PA residues 52 to 73 are required for cRNA-to-vRNA replication. In order to determine whether the observed defect in vRNA synthesis was due to an inability of the mutant RdRP to *trans*-activate the cRNA-bound RdRP, purified RNPs were incubated in the presence of ApG primers, nucleotides and purified RdRP containing PA Δ 51-72. Accumulation of viral RNA products were analysed by primer extension. As shown before, purified vRNPs were able to synthesise cRNA independently of any added RdRP, while purified cRNPs required presence of free RdRP to generate weak levels of vRNA

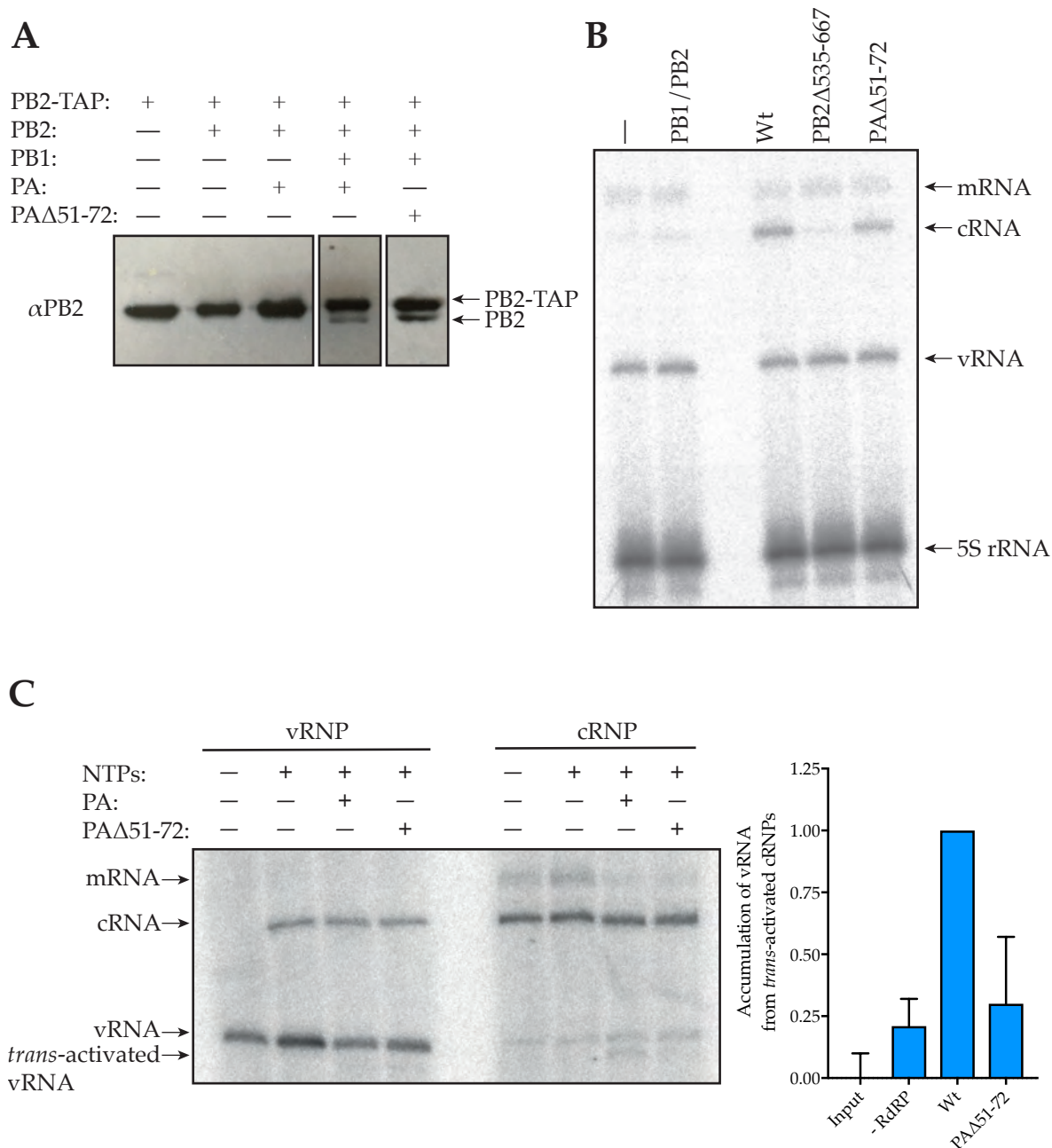


Fig. 5.5 PA residues 51-72 are not required for cRNA encapsidation or RdRP interactions. (A) RdRP subunits were co-expressed as indicated in HEK-293T cells and purified by IgG Sepharose chromatography using the TAP tag on PB2. Purified RdRP complexes were analysed by SDS-PAGE. Co-purification of two PB2 subunits was detected by Western blotting utilising the shift in molecular weight conferred by the TAP tag. (B) Human HEK-293T cells were co-transfected with plasmids expressing RdRP subunits (containing the active-site mutant PB1a) and NP as indicated. 48 h post-transfection, cells were infected with WSN virus at an MOI of 5 in the presence of actinomycin D. Accumulation of mRNA, cRNA, and vRNA 6 h.p.i. was analysed by primer extension. 5S rRNA was used as an internal loading control. (C) Purified vRNPs or cRNPs were incubated in the presence of ApG primers, nucleotides and purified RdRP as indicated. Accumulation of vRNA and cRNA was analysed by primer extension. The graph shows the mean intensity of accumulated vRNA from *trans*-activated cRNPs (n = 2, error bars = standard error of the mean).

(Fig. 5.5C). However, in the absence of PA residues 51-72, cRNPs failed to be *trans*-activated efficiently.

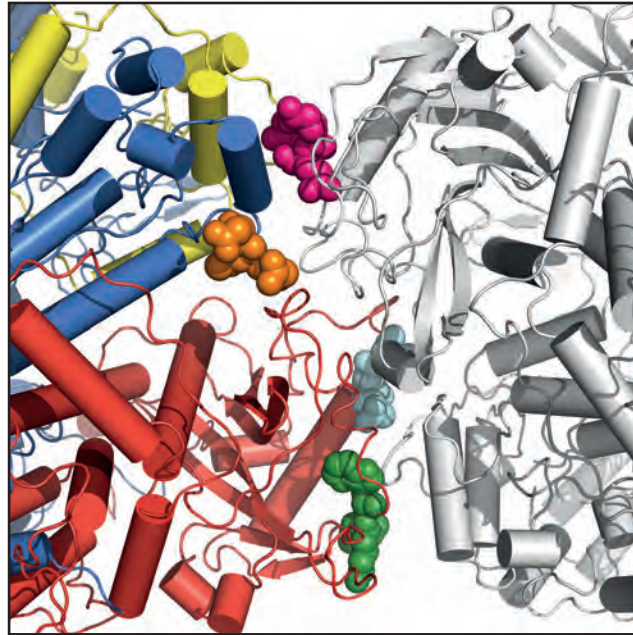
Taken together, these findings, show that PA residues 52 to 73 do not interfere with transcription or cRNA synthesis, while being essential for *trans*-activation of cRNPs, pointing to a role of RdRP dimer interface 3 in *trans*-activation.

5.2.2 RdRP dimer interfaces are important for viral replication

A previous unrelated study systematically introduced alanine mutations into the PA subunit (Hara et al., 2006). Some of these mutations, especially at residue PA-95, but also at residues 45, 98, 99 and 157 amongst others, specifically reduced accumulation of vRNA and cRNA while not affecting mRNA levels significantly. Like PA residues 52 to 73 described in Chapter 5.2.1, these residues are all found in or close to the N-terminal region of PA which is associated with dimer interface 3. The findings described in Chapter 5.2.1 suggest that a hypothetical dimer formed through the interface 3 could have a functional role in viral replication, specifically in *trans*-activation of cRNPs. It is possible to assume that all of these PA residues are critical for the same mechanism in viral replication.

Considering these results and to get a better understanding of the function of different potential RdRP dimer interfaces in viral replication, triple alanine mutations were introduced in all three RdRP subunits in regions associated with dimer interfaces 1 and 3. Their effect on RNP activity was tested using mini-replicon assays as described before and the accumulation of viral RNA was analysed by primer extension. Four different mutants targeting RdRP dimer interface 1 (Fig. 5.6A) and six mutants targeting RdRP dimer interface 3 were tested in addition to the previously described mutants PB2-K627E, PA-C95A, PA-T97A, PA-T98A and PA-G99A which are all associated with dimer interface 3 (Fig. 5.7A). At the RdRP dimer interface 1, mutants targeting PA residues 317-319 resulted

A



B

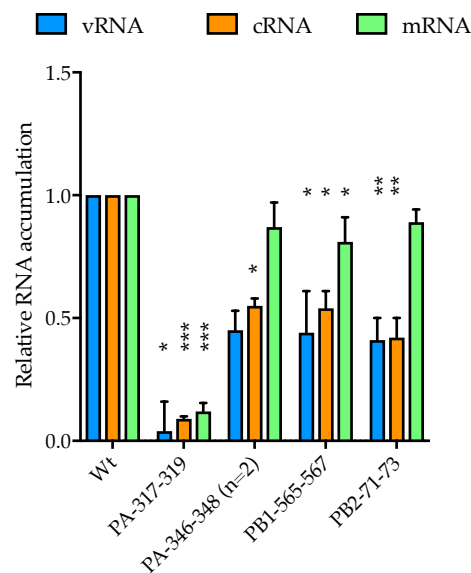
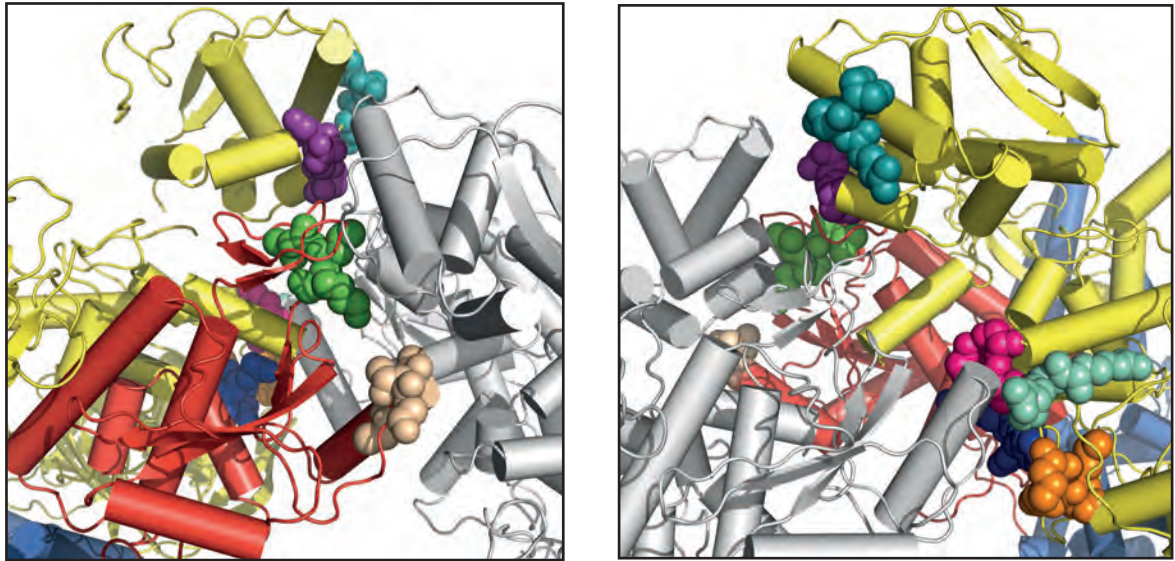


Fig. 5.6 The role of RdRP dimer interface 1 in viral replication. (A) Zoomed in view of the RdRP dimer interface 1 as shown in Figure 5.1. One RdRP monomer is shown in grey, the other RdRP is shown in red (PA), blue (PB1) and yellow (PB2). Triple alanine mutations were introduced in PA 317-319 (green), PA 346-348 (cyan), PB1 565-567 (orange) and PB2 71-73 (magenta). (B) Human HEK-293T cells were co-transfected with plasmids expressing segment 6 vRNA, NP, PA, PB1 and PB2 with mutations in RdRP dimer interface 1 as indicated. Accumulation of mRNA, cRNA and vRNA was analysed by primer extension. The graph shows the mean intensity signal relative to wild-type RdRP. 5S rRNA was used as an internal loading control. Results were obtained from three independent biological replicates if not indicated otherwise. Error bars represent the standard error of the mean and the asterisks indicate a significant difference to wild-type (two-tailed one-sample *t* test) as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

A



B

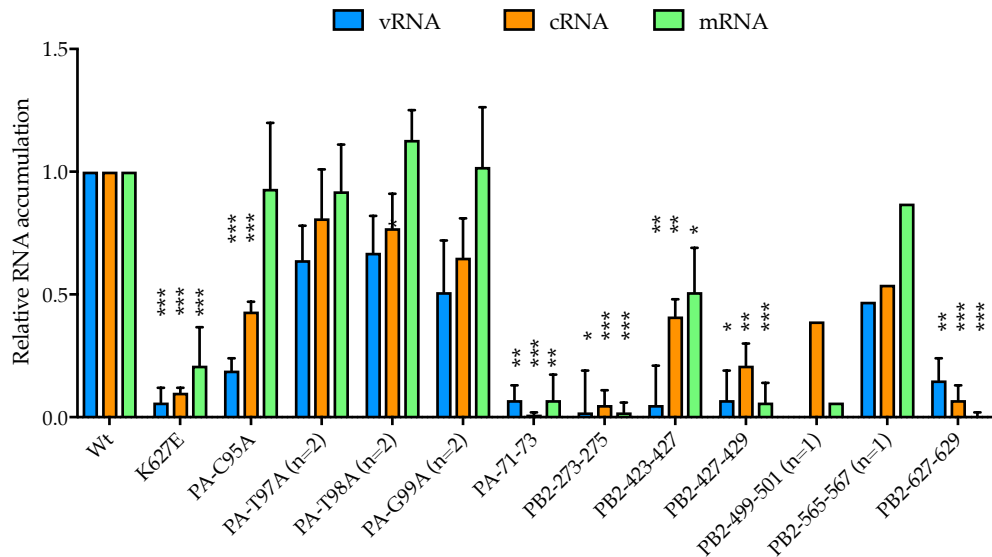


Fig. 5.7 The role of the RdRP dimer interfaces 3 in viral replication. (A) Zoomed in view of the RdRP dimer interface 3 as shown in Figure 5.1. One RdRP monomer is shown in grey, the other RdRP is shown in red (PA), blue (PB1) and yellow (PB2). Triple alanine mutations were introduced in PA 71-73 (green), PB2 273-275 (magenta), PB3 423-425 (cyan), PB2 427-429 (orange), PB2 499-501 (dark blue), PB2 565-567 (purple) and PB2 627-629 (teal). Single alanine mutations as described by Hara *et al.* (2006) were introduced in PA-95, PA-97, PA-98 and PA-99 (beige). (B) Human HEK-293T cells were co-transfected with plasmids expressing segment 6 vRNA, NP, PA, PB1 and PB2 with mutations in RdRP dimer interface 3 as indicated. Accumulation of mRNA, cRNA and vRNA was analysed by primer extension. The graph shows the mean intensity signal relative to wild-type RdRP. 5S rRNA was used as an internal loading control. Results were obtained from three independent biological replicates if not indicated otherwise. Error bars represent the standard error of the mean and the asterisks indicate a significant difference to wild-type (two-tailed one-sample *t* test) as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

in full restriction of viral replication and transcription, while mutants targeting PA residues 346-348, PB1 residues 565-567 or PB2 residues 71-73 only had a very modest or no significant impact on mRNA transcription, but a significant impact on replication (Fig. 5.6B). At the RdRP dimer interface 3, all triple alanine as well as the PA-C95A and PB2-K627E mutants significantly inhibited viral replication, while the remaining single alanine mutations somewhat but not significantly reduced replication. Viral transcription was inhibited in all triple alanine mutants while not being affected by the single alanine mutations (Fig. 5.7B).

All in all, introduction of mutations in both RdRP dimer interface 1 and 3 significantly reduced viral replication in mini-replicon assays. Interestingly, some mutations specifically affected viral replication, while other mutations had a broad effect on both replication and transcription.

5.2.3 Avian ANP32A can restore replication by PA-C95A

The findings in Chapter 5.2.2 agree with a previous study, showing that introducing a single alanine substitution in PA-C95 has a strong inhibitory effect on viral replication, while not affecting viral transcription (Hara et al., 2006). PA-95 is located on the surface of the RdRP proposed to be part of interface 3. Interestingly, the well-known host determinant residue PB2-627 is also found on the interface 3 of the same RdRP monomer and on the opposing RdRP monomer immediately opposite of PA-95 (Fig. 5.8A). Moreover, the nearby residue PA-97 is also known to harbour host adaptive mutations (Table 1.5). The avian ANP32A host factor is known to restore replication of avian-like RdRPs containing PB2-627E. This raises the question, whether PA-95 also underlies the restrictive effect as PB2-627 and whether avian ANP32A can restore replication in an RdRP containing the mutant PA-C95A.

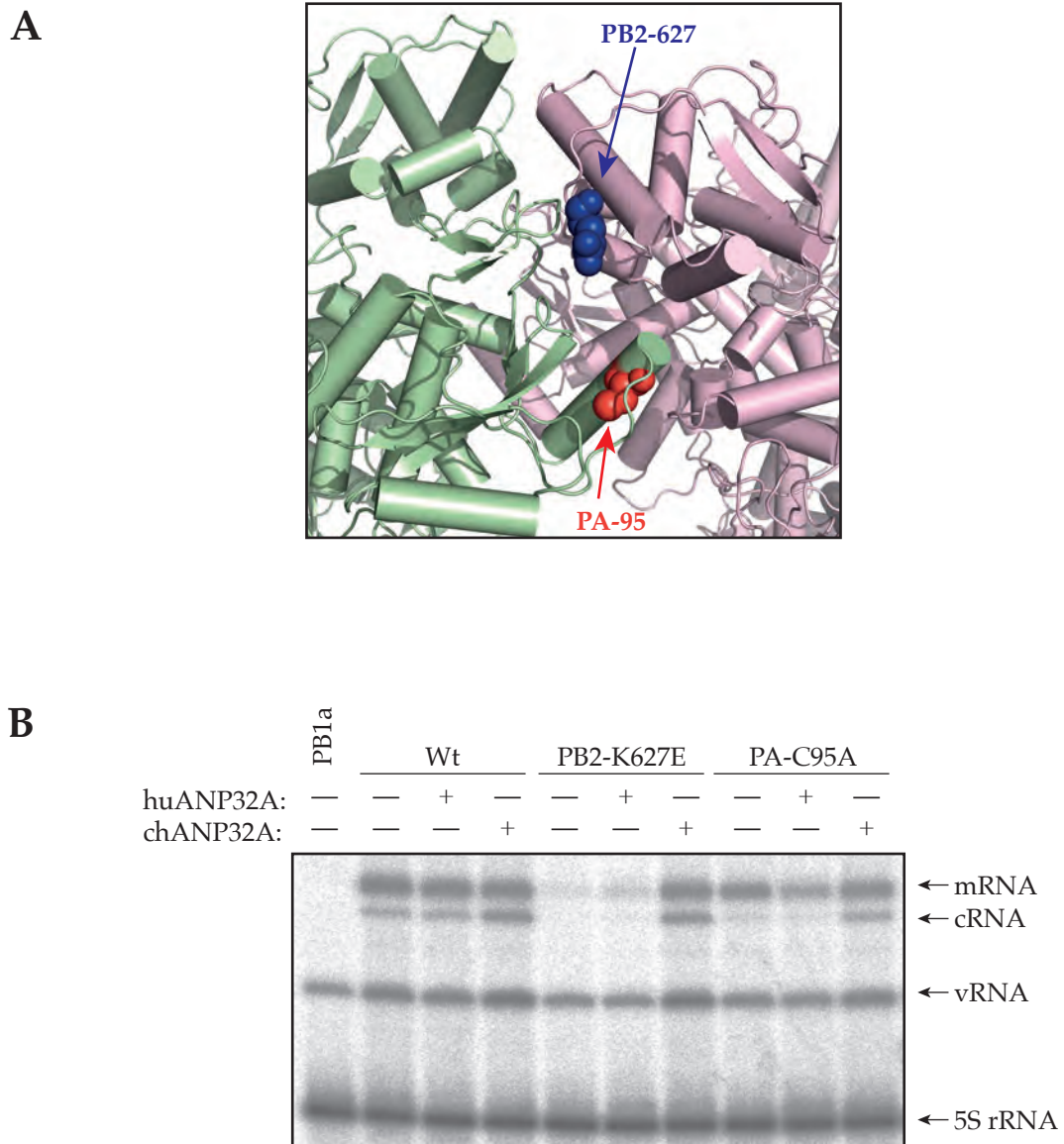


Fig. 5.8 Avian ANP32A stimulates RdRP activity through PA-95 and PB2-627. (A) Zoomed in view of the RdRP dimer interface 3 as shown in Figure 5.1. One RdRP monomer is shown in green, the other RdRP is shown in pink. PA-95 (red) and PB2-627 (blue) are highlighted as spheres. (B) Human HEK-293T cells were co-transfected with plasmids expressing RdRP subunits, NP, segment 6 vRNA and human or chicken ANP32A as indicated. Accumulation of mRNA, cRNA and vRNA was analysed by primer extension. 5S rRNA was used as an internal loading control.

To analyse the effect of avian ANP32A on PA-C95A activity, RNPs were reconstituted by co-expression of the three RdRP subunits, NP and segment 6 vRNA together with human or chicken ANP32A in human HEK-239T cells and the accumulation of viral RNA was analysed by primer extension. As expected, expression of huANP32A (which is already present in human HEK-239T cells) did not have any effect on viral RNA accumulation. Expression of chANP32A however, strongly enhanced vRNA and cRNA levels of the PA-C95A mutant RdRP, showing a similar effect as on PB2-627E (Fig. 5.8B). This proves that ANP32A does not exclusively stimulate the viral RdRP through PB2-627, but also at least through PA-95. Considering that both residues are located on the same proposed RdRP interface and that they in fact face each other on opposing RdRP monomers, it raises the question whether the role of ANP32A in viral replication can be linked to RdRP dimer interface 3 or RdRP dimer formation in general.

5.3 Discussion

The aim of this study was to determine whether the RdRP dimers observed in different structural studies have a biological function and to characterise their role in viral replication. This study identified that residues 52 to 73 of the PA subunit are not required for transcription both *in vitro* and in a cellular context, but that they play a critical role in replication, specifically in *trans*-activating cRNA-bound RdRPs to initiate vRNA synthesis. Nevertheless, these residues were not needed for general RdRP interactions or encapsidation of nascent cRNA. Furthermore, introducing alanine mutations into the RdRP dimer interfaces 1 and 3 revealed that both interfaces are critical for viral replication. Avian ANP32A was shown to be able to stimulate RdRP activity of both of the replication-deficient PA-C95A and PB2-K627E polymerases, suggesting that the action of avian ANP32A is linked to RdRP dimers.

In Chapter 3.2.3 it was shown that the flexible appendices of the viral RdRP are not necessarily required for all core enzymatic functions. Here, a 22-amino acid truncation (51-72) was introduced into the N-terminal endonuclease domain of the PA subunit. Both in a cellular environment and *in vitro* this truncation did not interfere with cap-snatching and cap-dependent transcription, which fits the fact that the catalytic site of the viral endonuclease is found around residue PA-108 (Yuan et al., 2009). Residues 52 to 73 were also shown to be superfluous for cRNA synthesis, which is consistent with the location of the PA endonuclease on the outside of the RdRP core (Hengrung et al., 2015; Pflug et al., 2014; Reich et al., 2014). Interestingly though, there seemed to be a dependency on this PA region for ApG-primed vRNA synthesis. It was previously shown that purified cRNPs require *trans*-activation by RNA-free RdRP to initiate vRNA synthesis (York et al., 2013). Purified RdRPs with an internal truncation in residues 51-72 of the PA subunit were unable to *trans*-activate purified cRNPs efficiently. Together with its inability to initiate ApG-primed vRNA synthesis, this suggests that the N-terminal PA region is required for the *trans*-activating interaction of RNA-free RdRP with cRNA-bound RdRPs. PA residues 52 to 73 are however not needed for the encapsidation of nascent cRNA products by an RNA-free RdRP. This suggests that encapsidation and cRNP *trans*-activation are facilitated through different polymerase-polymerase interactions, which fits the fact that several different dimer interfaces have been described so far.

These findings suggest that the RdRP dimer interface 3 is responsible for forming the interaction between an RNA-free RdRP and the resident RdRP of a cRNP during *trans*-activation. In the RdRP dimer 3 the interacting interface is formed by a direct interaction of the N-terminal region of PA with the C-terminal domains of PB2, especially the 627-domain, on an opposing RdRP. In Chapter 3 it was shown that the 627-domain is not required for *trans*-activation of cRNPs, which stands in direct contrast to the results for the PA N-terminus described in this chapter. It is not immediately obvious how these findings

could be reconciled, but one could speculate that the PA region is more critical for this RdRP interaction and that in the absence of the 627-domain the PA N-terminus is still able to form an alternative interaction with the opposing RdRP. However, it is also feasible that the *trans*-activating interaction is mediated through a different dimer structure, which has not been identified yet. More work is needed to characterise the role of different RdRP regions in different viral functions.

Another interesting difference between truncations in the PA N-terminus and the PB2 627-domain is their ability to stabilise and encapsidate cRNA. While the 627-domain plays a critical role in encapsidation, PA residues 51-72 were dispensable for this process. Again, one explanation of this phenomenon could be that the 627-domain is more critical for this interaction. However, an interaction through the PB2 C-terminal domains and the PA N-terminus of the encapsidating RdRP appears unlikely, because the template entry channel, which needs to be aligned with the product exit channel of the replicating RdRP, is located on the opposite end of the RdRP. Moreover, recently the host factor ANP32A was shown to be important for viral replication related to PB2-627 (Long et al., 2016; Sugiyama et al., 2015). Therefore, one could speculate that ANP32A is involved in facilitating encapsidation of nascent cRNA and presumably also vRNA by RNA-free RdRP. Whether this is mediated by ANP32A through a direct interaction between two RdRPs or whether no direct interaction is in fact needed with ANP32A present, remains unknown.

What is clear though is that *trans*-activation of cRNA-bound RdRP and encapsidation of nascent cRNA is mediated through different RdRP interactions. It appears logical that the viral RdRP needs to form several RdRP-RdRP interactions in viral replication. RNA-free RdRPs need to encapsidate nascent cRNA and vRNA products (Vreede et al., 2004), as well as bind the 3' template strand in the *trans*-acting replication model (Jorba et al., 2009) and *trans*-activate cRNA-bound RdRPs (York et al., 2013), all of which presumably require different direct interactions between the two RdRPs. Furthermore, it has been shown that

vRNA-bound, cRNA-bound or RNA-unbound RdRPs exist in different conformational arrangements (Hengrung et al., 2015; Pflug et al., 2014; Thierry et al., 2016). Together, this suggests that for different RdRP functions, different RdRP interactions can be formed, mediated by the different conformational arrangements of the RdRP.

The findings that mutations targeting both dimer interface 1 and interface 3 inhibit cRNA and vRNA synthesis, fits the proposed roles of RdRP dimers in viral replication. While it appears clear that viral replication is inhibited by these mutations, viral transcription seems to be differentially affected. However, awaiting further data gathering, one could speculate that viral transcription is only reduced in these cases because of a lack of template amplification as is the case of for the avian-like PB2-627E mutation. Mutations in PA residues 95, 97, 98 and 99 confirm previously established findings that viral replication is specifically reduced (Hara et al., 2006). The most dramatic restriction of RdRP activity is observed when mutating PA 71-73 and PB2 273-275. The results for PA residues 71-73 correspond very well with the deleterious effect observed when residues 51-72 were truncated. Interestingly, the well-known host adaptive site PB2-271 is found adjacent to the mutated residues 273-275. PB2-271 is found on the top of an alpha helix which makes contact with another alpha helix containing PA-95 and PA-97, also known to harbour a host adaptive mutation, in the opposing RdRP in the dimer structure 3. Several more host adaptive mutations can be found on both dimer interfaces and close to the mutations analysed in this study. Host adaptive mutations in PB1-581/584, PA-295/312/336/337/356 and PB2-81/105 are all found close to the introduced alanine mutations targeting dimer interface 1. PB2-271/493/559/567/569/588/590/591/627 and PA-55/57/66 are all located close to the introduced alanine mutations targeting dimer interface 3. This clearly points to a role of RdRP dimers in host range restriction. With the recent finding linking ANP32A to host range restriction mediated through PB2-627, it is tempting to speculate that ANP32A could have a similar effect on mutants in the RdRP

dimer interface. This is supported by the finding that replication by the PA-C95A mutant RdRP is restored by co-expression of chicken ANP32A, but more data needs to be gathered to draw a more comprehensive conclusion on the effect of ANP32A on RdRP dimers.

5.4 Summary

All in all, this study presented that the PA N-terminus is involved in vRNA synthesis by the viral RdRP through a potentially *trans*-activating interaction with a replicating RdRP, but that it is not required for encapsidation of nascent cRNA products in a cellular environment. We propose that at least two different RdRP dimer structures are formed during viral replication. It was also shown that mutations in the proposed RdRP dimer interfaces interfere with viral replication, supporting the hypothesis that RdRP dimers are necessary for viral replication. These dimer interfaces were linked to host range restriction and avian ANP32A was shown to exhibit a similar effect on PA-95 as on PB2-627. We therefore propose that ANP32A is involved in mediating formation of RdRP dimer structures and that host range restriction is linked to differential abilities to form RdRP dimers between avian and mammalian influenza virus RdRPs.

Chapter 6: The mRNA interactome in influenza virus infection

6.1 Introduction

The repertoire of RNA-binding proteins (RBPs) as well as RNA-binding domains (RBDs) has been significantly expanded in recent years (Beckmann et al., 2016). Nevertheless, novel RBPs and RBDs are still being discovered, suggesting a much wider role of RBPs in cell physiology than classically proposed. RBPs play a critical role in RNA metabolism ranging from transcription, mRNA processing, export and transport to translation and mRNA degradation (Glisovic et al., 2008). Consequently, changes in the RBPome reflect the dynamic cellular response to different internal and external stimuli.

RNA viruses, including influenza viruses, are a major threat to global health and cause significant morbidity and mortality. RNA-binding proteins have key pro- and antiviral functions in infections with these RNA viruses. Viral RNA genomes can be detected by the host innate immune system through a range of RNA-sensing receptors that activate the antiviral immune response. On the other hand, viruses have evolved to hijack host RNA-binding proteins for their biological cycle. RNA viruses typically have a relatively small genome with a restricted coding capacity for viral proteins. Therefore, these viruses have evolved to utilise and depend on cellular host proteins for various aspects of viral replication, translation and packaging. Influenza viruses rely on several host interactions in their replication cycle, in particular during transcription and translation of their genome as reviewed in Chapter 1.6.3. Therefore, it is likely that RNA-binding proteins

and their interactions with viral and host mRNA are highly regulated in viral infections, reflecting the need for the virus to monopolise the transcriptional and translational host machinery for viral gene expression.

It is known that infection with influenza virus induces a global host shut off as reviewed in Chapters 1.6.4 to 1.6.6, induced by the viral NS1, PA-X and RdRP proteins. Host shut off results in strongly reduced gene expression and cellular protein synthesis and preferential translation of viral proteins (Yángüez et al., 2011). Changes in the host cellular proteome in influenza virus infections have been extensively studied before (Coombs et al., 2010; Ding et al., 2016; Dove et al., 2012; Kummer et al., 2014). However, no studies have so far ventured to analyse the RNA-binding proteome in influenza virus infected cells specifically or to determine whether viral infection alters RNA-binding behaviour of host RBPs.

The aim of this study is to determine the dynamic response of the RNA-binding proteome to influenza A viruses in human host cells and to assess how this plastic response is achieved. Adopting a previously established RNA-binding interactome capture technique, RBPs were isolated from human cells after *in vivo* crosslinking of nucleic acids and proteins and analysed by mass spectrometry.

6.2 Results

6.2.1 Capture of human RBPs in A549 cells

In order to analyse potential changes to the RBPome caused by influenza virus infections it is important to understand the complexity of the RBPome of the cell line. To assemble a complete catalogue of proteins that directly interact with polyadenylated RNA in

immortalised human alveolar basal epithelial (A549) cells, proteins were covalently bound to mRNA by UV crosslinking (254 nm). Polyadenylated RNA was subsequently captured on oligo(dT) magnetic beads from cell lysates. Protein-RNA complexes were washed under highly stringent conditions and treated with RNase before isolated RNA-binding proteins were characterised by SDS-PAGE and mass spectrometry (Fig. 6.1A).

Isolated proteins were analysed by SDS-PAGE and silver staining, revealing that protein complexes were indeed co-purified with polyadenylated RNA only when cells were irradiated with UV light (Fig. 6.1B). Western blotting analyses showed that the well-established RBP polypyrimidine tract-binding protein 1 (PTBP1) was successfully detected in the irradiated sample but not in the control sample that had not been crosslinked, whereas the negative control ACTB was not detected in either sample (Fig. 6.1C). This demonstrates that the RBPome capture technique was successfully applied in A549 cells and that RNA-binding proteins were selectively isolated using this technique.

In follow up experiments proteins were digested with trypsin and analysed by dimethyl label-dependent quantification after tandem mass spectrometry (MS/MS) by the Advanced Proteomics Facility (University of Oxford). Identified peptides were analysed by Ni Shuai and Dr Bernd Fischer using a custom made pipeline utilising the Limma software and R (DKFZ, Germany). In the UV-crosslinked samples a total of 396 proteins were confidently identified as RNA-binding proteins ($p < 0.05$). Gene Ontology (GO) analysis of these proteins showed a significant association with RNA binding (Fig. 6.2A). The majority of these proteins were already reported as RBPs in other RBPome captures and most identified proteins that failed the statistical threshold ($p < 0.05$) were not part of these previous RBPomes, suggesting that the A549 RBPome is comparable to the RBPomes of other cell lines like HEK-293 and HeLa (Fig. 6.2B). Nevertheless, many proteins that failed the statistical threshold had already been described as RBPs in other interactome captures, suggesting that the pool of RBPs in A549 cells is significantly larger

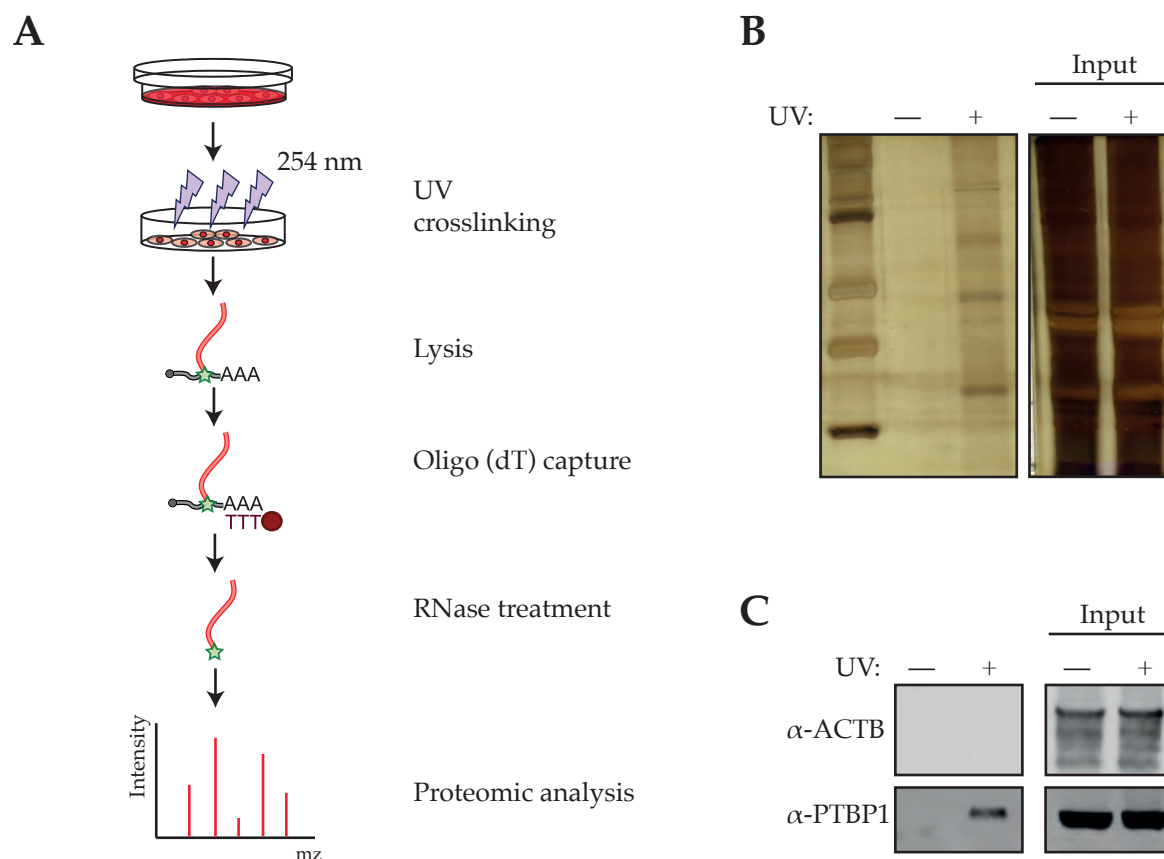


Fig. 6.1 *In vivo* capture of mRNA-binding proteins. (A) Schematic of RBPome capture. Proteins and nucleic acids are crosslinked in human A549 cells by UV irradiation. RNA-binding proteins are isolated from cell lysates using oligo(dT) magnetic beads. Captured RNA-binding proteins are treated with RNase and analysed by mass spectrometry. (B-C) Validation of successful isolation of mRNA-binding proteins. Isolated proteins were analysed by SDS-PAGE and (B) silver staining and (C) western blotting for a known RNA-binding protein (PTBP1) and a negative control (ACTB).

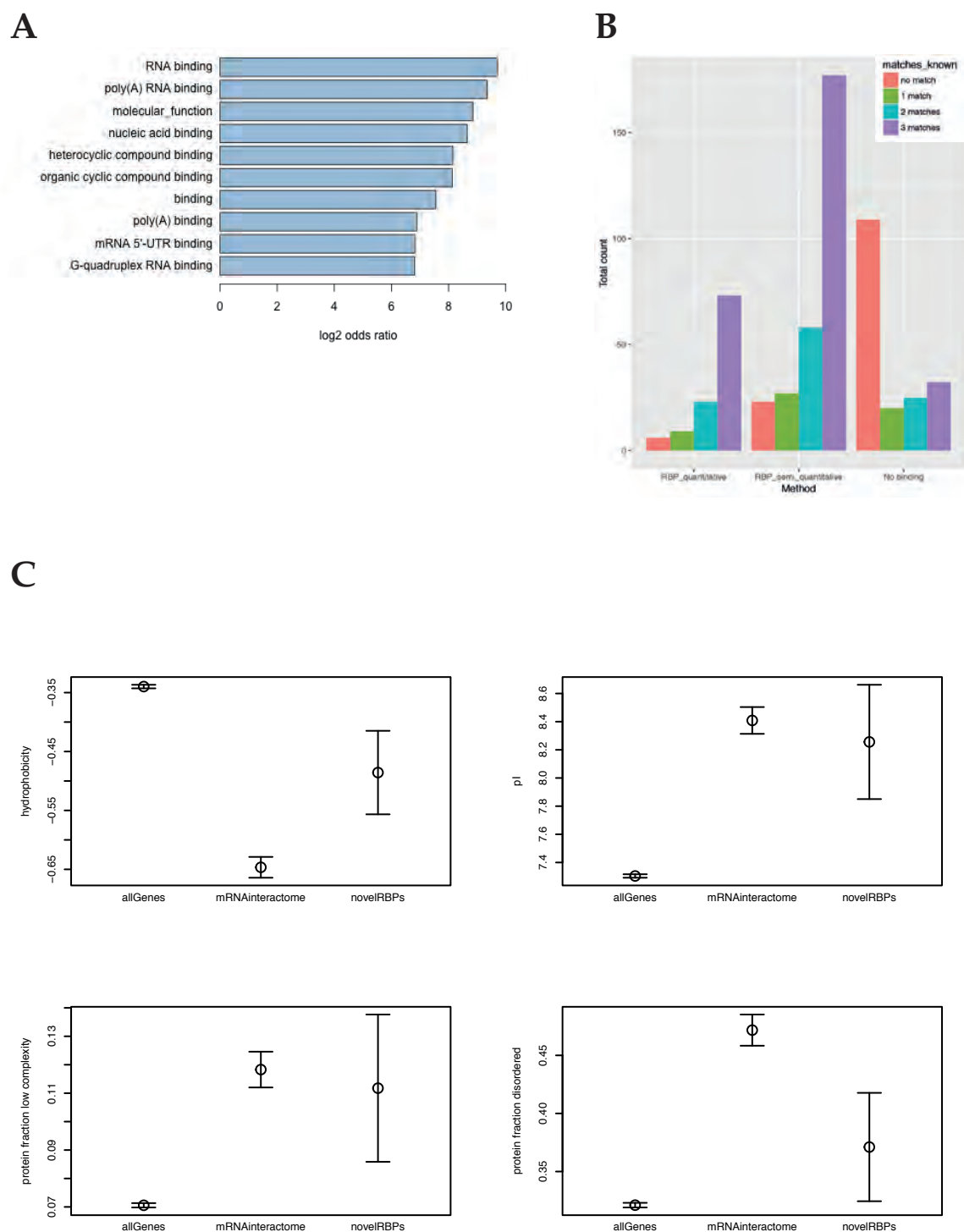


Fig. 6.2 The RBPome of A549 cells. (A) The ten of the most significant overrepresented molecular function GO terms of the A549 mRNA interactome. (B) Number of proteins identified by MS that have been identified in previous mRNA interactomes (HeLa, HEK-293 and mouse embryonic stem cells). (C) Biochemical characteristics of the mRNA interactome compared to the total proteome. Analysis performed by Ni Shuai and Dr Bernd Fischer (DKFZ, Germany).

than presented here. Comparing the biochemical characteristics of high-confidence RBPs, some interesting differences between previously known RBPs and novel RBPs identified in this study could be observed. As a whole, novel RBPs were more hydrophilic and less disordered than previously known RBPs, while their fraction of low complexity regions and their isoelectric point were similar to known RBPs (Fig. 6.2C). A complete list of identified RBPs is attached in Appendix J.

6.2.2 The dynamics of viral gene expression

Once the RNA-interactome capture was established in A549 cells, the effect of influenza WSN virus infection on the repertoire of RBPs in A549 cells was investigated. First however, the kinetics of influenza virus infection in A549 cells were determined in order to optimise viral infections.

WSN virus was grown in bovine MDBK cells and 2 days after infection, viral supernatants were harvested and cleaned from cellular debris by low-speed centrifugation. In order to clean viral preparations from the growth media containing an abundance of host factors that could have a profound impact on viral infections, virions were purified by ultracentrifugation through a sucrose gradient and viral titres were determined using plaque assays. Analysis of equivalent amounts of virions (as measured by PFU) by SDS-PAGE and silver staining (Fig. 6.3A) showed a significantly reduced smear of proteins present in purified virions and protein banding pattern corresponding to viral proteins identified in other studies of isolated virions (Hutchinson et al., 2014; Shaw et al., 2008).

To determine the dynamics of viral protein synthesis, A549 cells were infected with WSN at a high MOI of 5 and at the indicated time cells were lysed. Analysis of whole cell lysates (WCL) revealed a clear accumulation of a virus-induced protein with a molecular weight of approximately 26 kDa, which corresponds to the viral M1 protein (Fig. 6.3B).

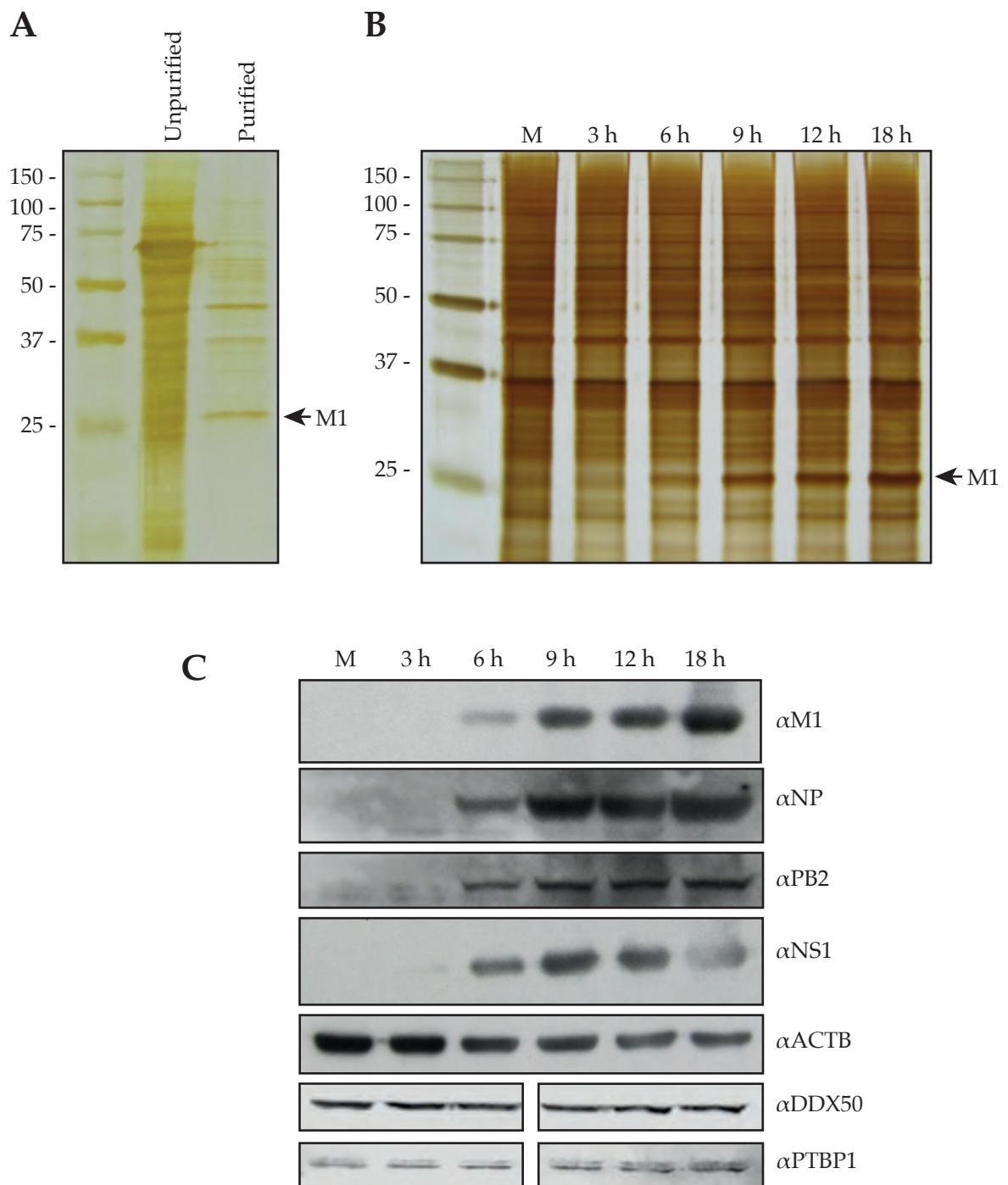


Fig. 6.3 Influenza A virus protein expression in human A549 cells. (A) Influenza A/WSN/33 (H1N1) virus harvested from viral supernatant was purified by ultracentrifugation through a sucrose gradient. Equivalent amounts of infectious particles, as determined by PFU, were analysed by SDS-PAGE and silver staining. Molecular masses, indicated on the left, are depicted in kilodaltons. (B-C) Human A549 cells were infected with WSN virus at an MOI of 5 and WCLs were analysed by SDS-PAGE and (B) silver staining or (C) western blotting.

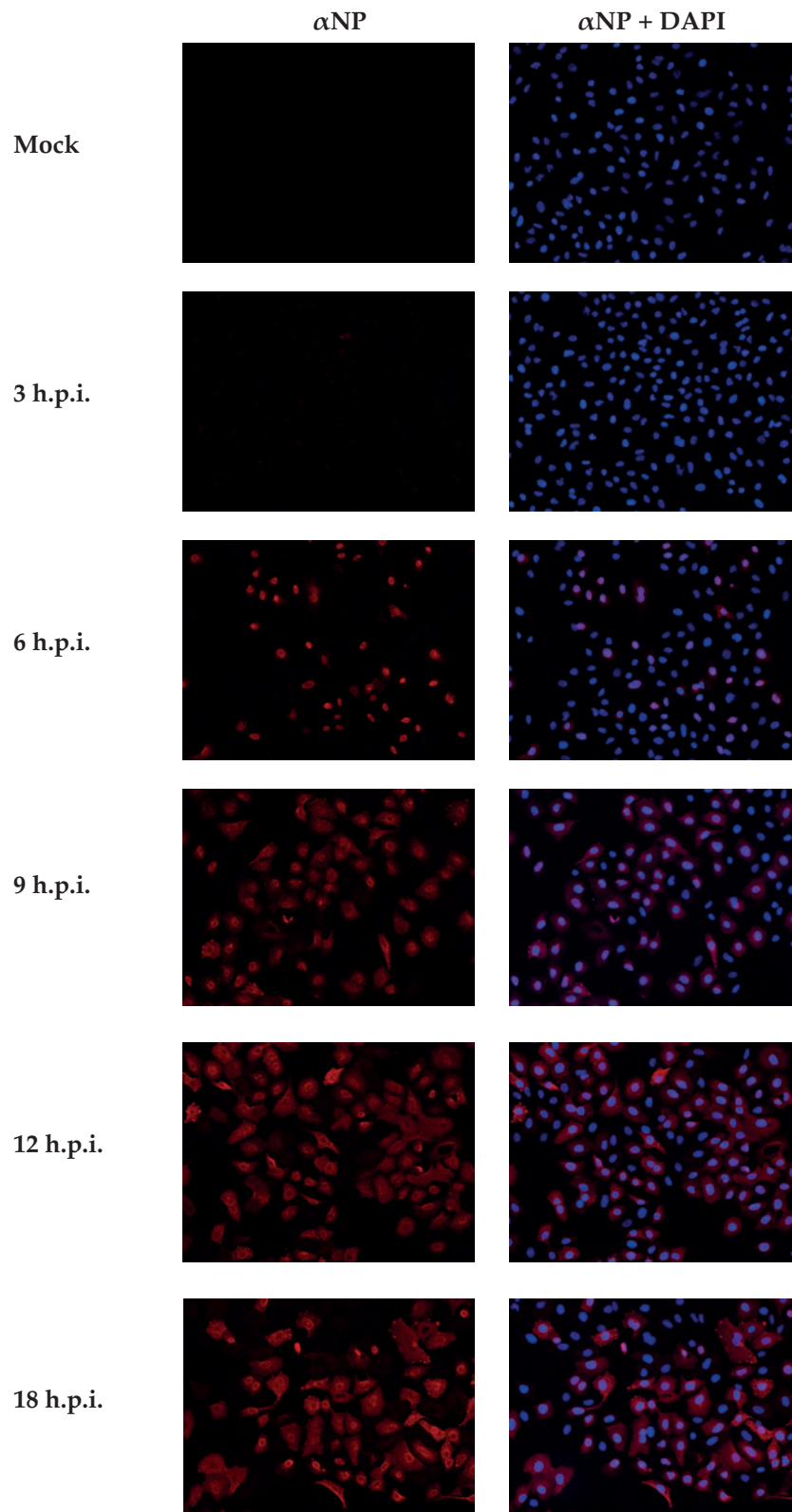


Fig. 6.4 **Infection efficiency of A549 cells with influenza A virus.** Human A549 cells were infected with influenza A/WSN/33 (H1N1) virus at an MOI of 5. At the indicated time points post-infection cells were fixed and cell nuclei (blue) and viral NP (red) were stained.

Proteins detected here do not solely correspond to proteins being synthesised during the infection due to the half-life of most cellular proteins being longer than 18 h. Previous studies have shown that influenza A viruses hijack the host translational machinery, leading to a strong preference in translation for viral transcripts after 6 h of infection (Yángüez et al., 2011). Analysis of protein samples for viral and host proteins by western blot, showed that 6 h.p.i. accumulation of viral proteins can be observed. 9 h.p.i. levels of NP, PB2 and NS1 were at or close to their maximum, while M1 expression was continuously increasing. Expression of host proteins DDX50 and PTBP1 remained stable, while ACTB decreased slightly, most likely reflecting increasing amounts of cell death late in viral infection (Fig. 6.3C).

Because this RNA-interactome capture technique surveys the RBPome of all cells in a sample, it is important to ensure that a high proportion of cells are successfully infected at the relevant time points. Therefore, A549 cells were infected with WSN virus at a high MOI of 5 and at the indicated time points fixed for immunofluorescent staining. Staining of infected cells with an antibody targeting the viral NP protein showed that 6 h.p.i. weak signals of NP were mainly detected in the nuclei of infected cells (Fig. 6.4). From 9 h.p.i. NP was detected in most cells and was also recognised in the cytoplasm.

Together, these results indicate that viral protein synthesis significantly increases between 3 h and 6 h.p.i. and that expression of most viral proteins reaches a peak after 9 h in A549 cells.

6.2.3 The dynamics of viral RNA synthesis

RNA-interactome capture relies on the capture of polyadenylated RNA. It is therefore critical to know the composition of the host and viral transcriptomes. Superinfection of human HEK-239T cells with WSN virus has shown before that viral mRNA levels peak

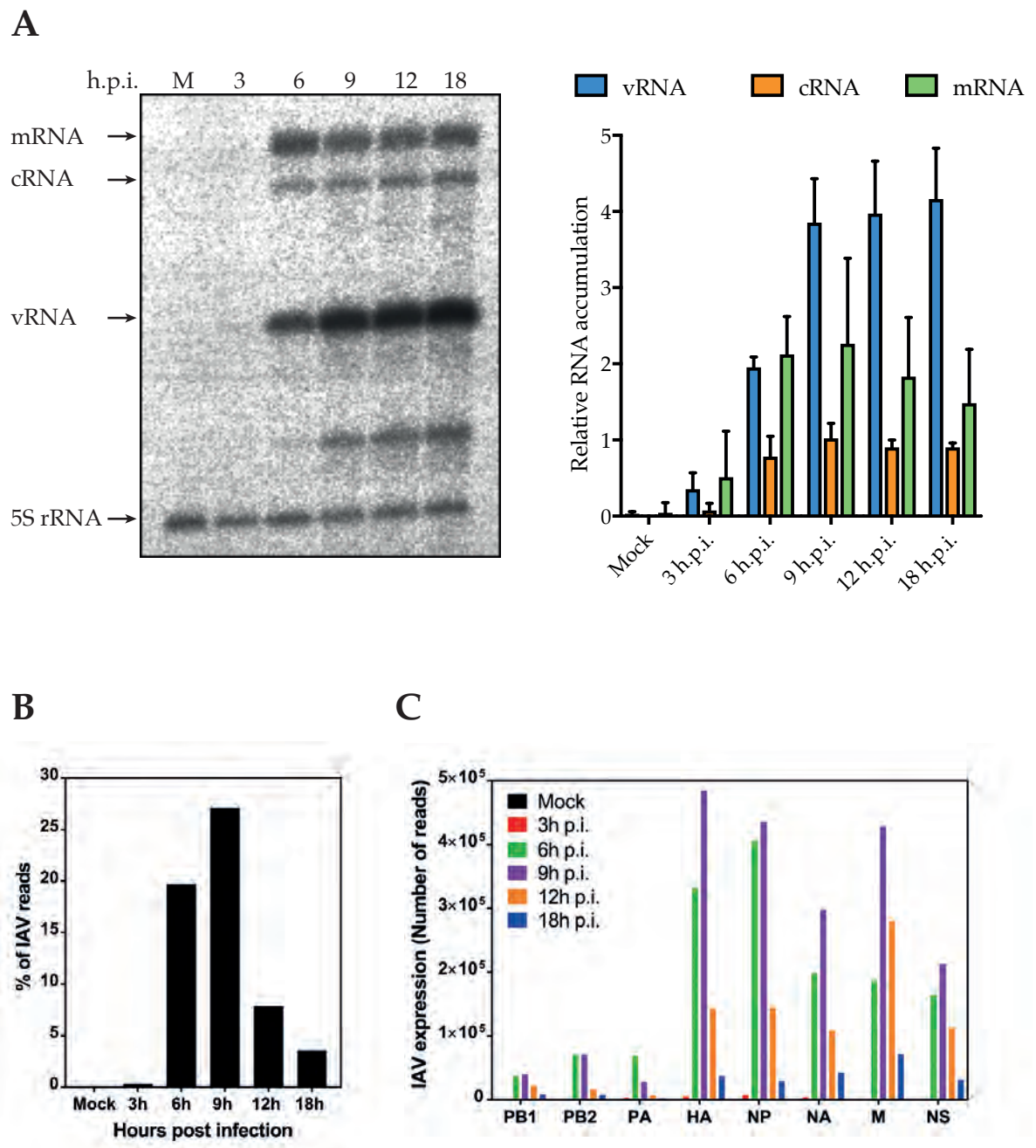


Fig. 6.5 Influenza A/WSN/33 (H1N1) gene expression in A549 cells. (A) Human A549 cells were infected with WSN virus at an MOI of 5. At the indicated time point post-infection total RNA was extracted and the accumulation of viral RNA was analysed by primer extension and 6% PAGE. The graph shows the mean accumulation of viral RNA from three independent biological replicates with error bars depicting the standard error of the mean. 5S rRNA was used as an internal loading control. (B) 3' UTRs of poly(A) RNA from infected cells were sequenced and reads were mapped to the human and viral genomes. The amount of viral mRNA at the indicated time points post-infection is depicted as percentage of all mRNA. (C) Reads mapped to individual viral gene segments at the indicated time points post-infection. RNA sequence analysis was performed by Michael Tellier (Sir William Dunn School of Pathology, University of Oxford).

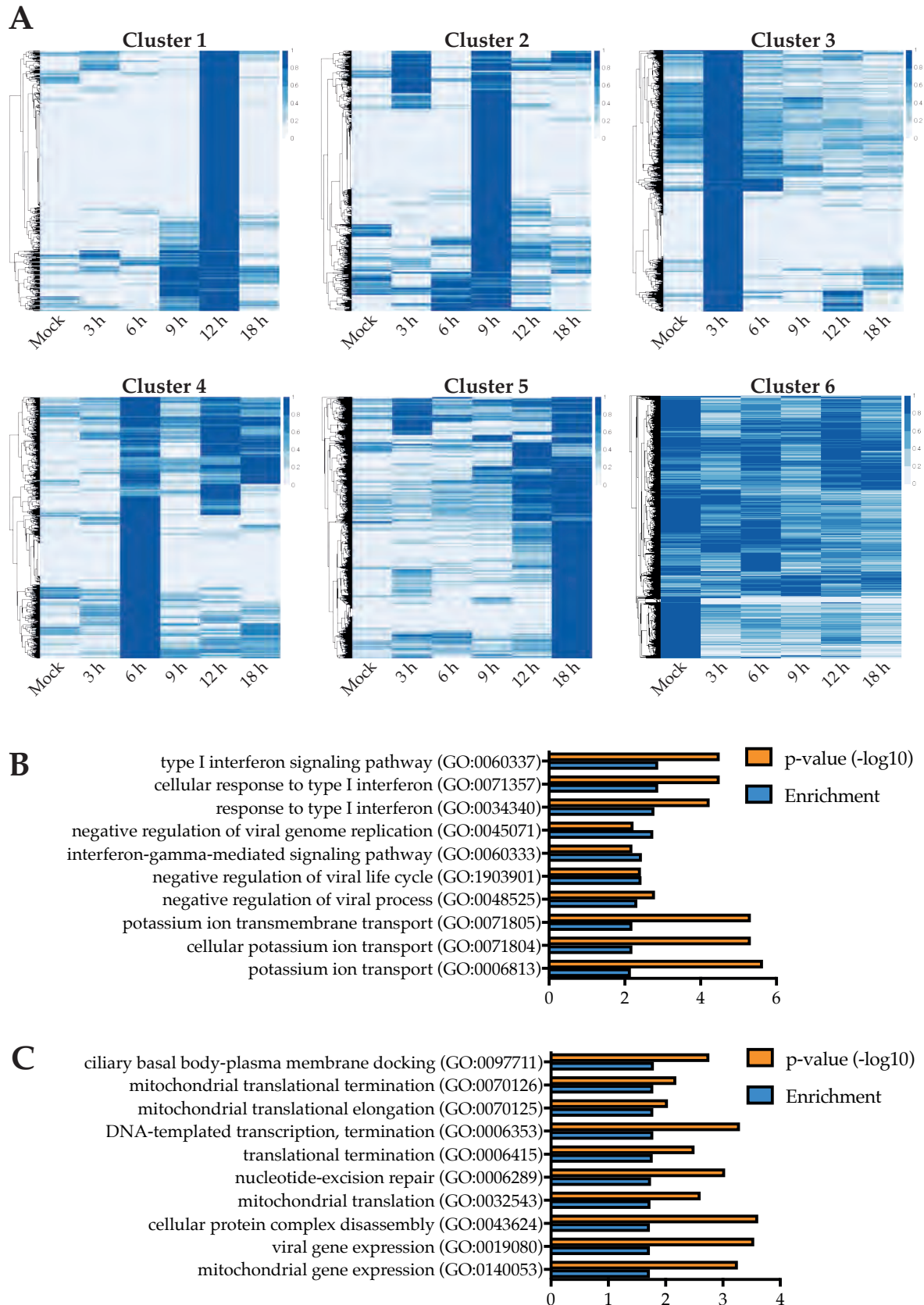


Fig. 6.6 Host gene expression in response to influenza A virus infection. (A) Host genes were hierarchically clustered into six clusters according to their gene expression during viral infection. Clustering was performed by Michael Tellier (Sir William Dunn School of Pathology, University of Oxford). (B) The ten most enriched GO terms for biological processes and their statistical significance associated with cluster 5. (C) The ten most enriched GO terms for biological processes and their statistical significance associated with cluster 6.

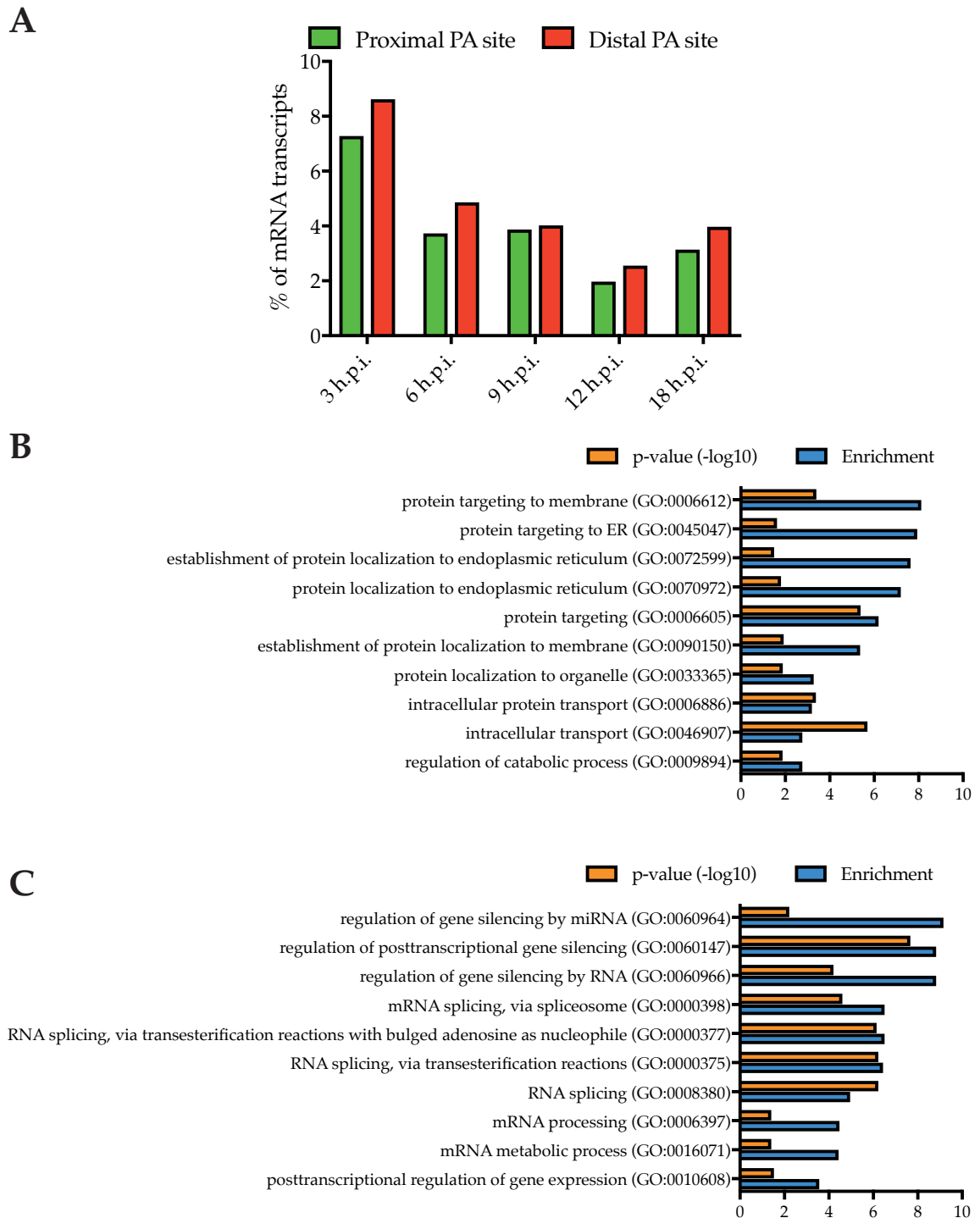


Fig. 6.7 **Alternative polyadenylation of host genes in response to influenza A virus infection.** (A) Changes in preferred polyadenylation (PA) site usage of host genes induced by influenza A virus infection. The percentage of all mRNA transcripts, that are alternatively polyadenylated by switching to a proximal or distal PA site in infected compared to uninfected cells, is depicted. Changes are only classified as significant if $p < 0.01$ and the abundance change is $> 5\%$ (B) The ten most enriched GO terms for biological processes and their statistical significance associated with shortened transcripts (proximal PA site). (C) The ten most enriched GO terms for biological processes and their statistical significance associated with lengthened transcripts (distal PA site).

approximately 4 to 6 h.p.i. (Robb et al., 2009; Vreede et al., 2011). To determine the dynamics of viral RNA synthesis in A549 cells, A549 cells were infected with WSN virus at a high MOI of 5 and at the indicated time points total RNA was extracted and analysed by primer extension with primers targeting positive- and negative-sense viral RNA, followed by 6% PAGE. 5S rRNA was used as an internal loading control. Very low levels of mRNA, cRNA and vRNA were detected at 3 h.p.i., while a rapid increase in all viral RNA species was seen between 3 and 9 h post-infection. After 9 h the amount of vRNA and cRNA present in infected cells remained stable, while a decrease in mRNA was observed (Fig. 6.5A). A decrease in mRNA levels late in infection has been observed before for WSN virus (Robb et al., 2009; Vreede et al., 2011).

The association of RBPs with RNA depends on the state of the cellular mRNA pool. In order to analyse the mRNA transcript pool in A549 cells infected with WSN virus, A549 cells were infected at an MOI of 5 with WSN virus for the indicated time points. Total RNA was extracted and treated with DNase before polyadenylated RNA was selected for cDNA library preparation. Using an Ion Proton System, the 3' UTRs of mRNA transcripts were sequenced and mapped to the human as well as influenza A/WSN/33 (H1N1) genome. Sequence alignment showed that 3 h.p.i. viral mRNA transcripts constituted only 1% of all identified mRNA transcripts. Until 9 h.p.i. the proportion of viral mRNA transcripts rose rapidly to approximately 28%. Late in infection a significant decrease in viral mRNA was observed (Fig. 6.5B). Notably, a much sharper decline in viral mRNA can be seen in the sequencing data compared to the segment 6 mRNA presented in Figure 6.5A.

Interestingly, viral gene segments appear to be differentially expressed. Polymerase segments PB1, PB2 and PA were generally transcribed at lower levels compared to the remaining segments (Fig. 6.5C). HA and NP segments appear to be preferentially transcribed early in infection (3 and 6 h), whereas the M segment was preferentially

transcribed late in infection (12 and 18 h). Because sequencing of the 3' UTR does not allow the differentiation between alternatively spliced transcripts, the number of reads per segment does not give any information about gene expression of alternatively spliced segment 1, segment 7 and segment 8 transcripts.

Infection with influenza virus induced a change in host gene expression. Hierarchical clustering of gene expression profiles over the time course of infection resulted in six host gene clusters (Fig. 6.6A). GO term analysis of clusters 1 to 4 did not show a strong enrichment for biological processes. Genes in cluster 5 were strongly upregulated over the time course of infection and were significantly associated with antiviral defence and ion transport (Fig. 6.6B). Genes in cluster 6 were downregulated over the time course of infection and were strongly linked to cellular metabolic processes and gene expression (Fig. 6.6C).

Host mRNAs can form transcripts with differing 3' UTRs by alternative polyadenylation (APA), resulting in mRNA isoforms with shortened or lengthened 3' UTRs and proteins with alternative C-termini (reviewed by Tian & Manley, 2017). Here, the use of the proximal and distal polyadenylation (PA) sites in influenza virus infection was determined and compared to PA site usage in mock infected cells. Viral infection induced APA compared to mock infected cells (Fig. 6.7A). Generally, APA was more pronounced early in infection and a reversion to the PA site preferentially used in mock infected cells was observed late in infection. No preference for shortening or lengthening in APA induced by influenza virus was seen. GO term analysis of shortened transcripts revealed a strong association with intracellular protein localisation and transport (Fig. 6.7B). Lengthened transcripts however were strongly linked to RNA processing (Fig. 6.7C).

Taken together, this data suggests a strong transcriptional activity of the virus early in infection with highest viral mRNA levels approximately 9 h post-infection. Host gene expression shows an upregulation of antiviral and ion transport-related host genes and a

down regulation of metabolic pathways. Alternative polyadenylation is induced by viral infection and might differentially affect different host genes.

6.2.4 The effect of influenza virus infection on the RBPome

To characterise any effects that influenza virus infections might have on the repertoire of A549 RBPs, SILAC-labeled A549 cells were infected at a high MOI of 5 with purified WSN virus. Cells were infected for 6 or 9 h, representing time points where viral transcription is in its exponential phase and viral proteins can be detected (6 h) and where viral mRNA levels are at their peak (9 h). Infected or mock infected cells were irradiated with UV and polyadenylated RNA was captured as described in Chapter 6.2.1. Samples were treated with RNase and identified and quantified by quantitative mass spectrometry (Fig. 6.8A).

Analysis of isolated proteins by SDS-PAGE and silver staining showed successful crosslinking of RBPs. Furthermore, late in infection differences in the banding patterns could be observed, pointing towards a changing RBPome after viral infection (Fig. 6.8B).

Quantitative analysis of isolated proteins by mass spectrometry identified a total of 651 RBPs, of which 521 had been described in other RNA interactomes before and of which 130 were novel RBPs (Fig. 6.9A). Of these 651 RBPs, 18 (1% false discovery rate, FDR), 22 (10% FDR) and 28 (20% FDR) proteins were enriched or depleted after 9 h of infection compared to mock infected cells (Fig. 6.9B).

Protein intensities of captured RBPomes were compared between mock-infected and infected samples and the fold change was determined. The dynamic range of alterations induced by infection was very narrow 6 h.p.i., *i.e.* viral infection did not induce many large changes in captured protein abundance (Fig. 6.9C). This is supported by the fact that only few proteins show a statistically significant enrichment or depletion in abundance

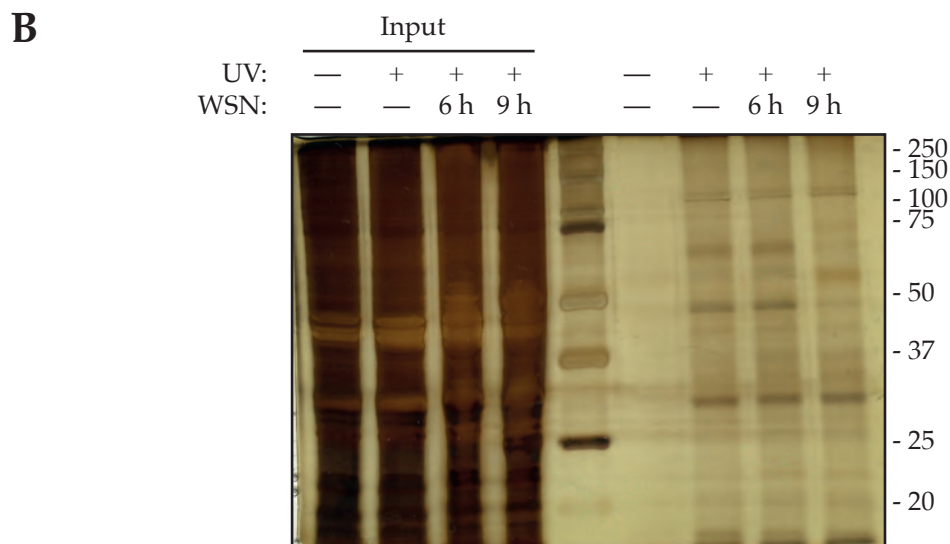
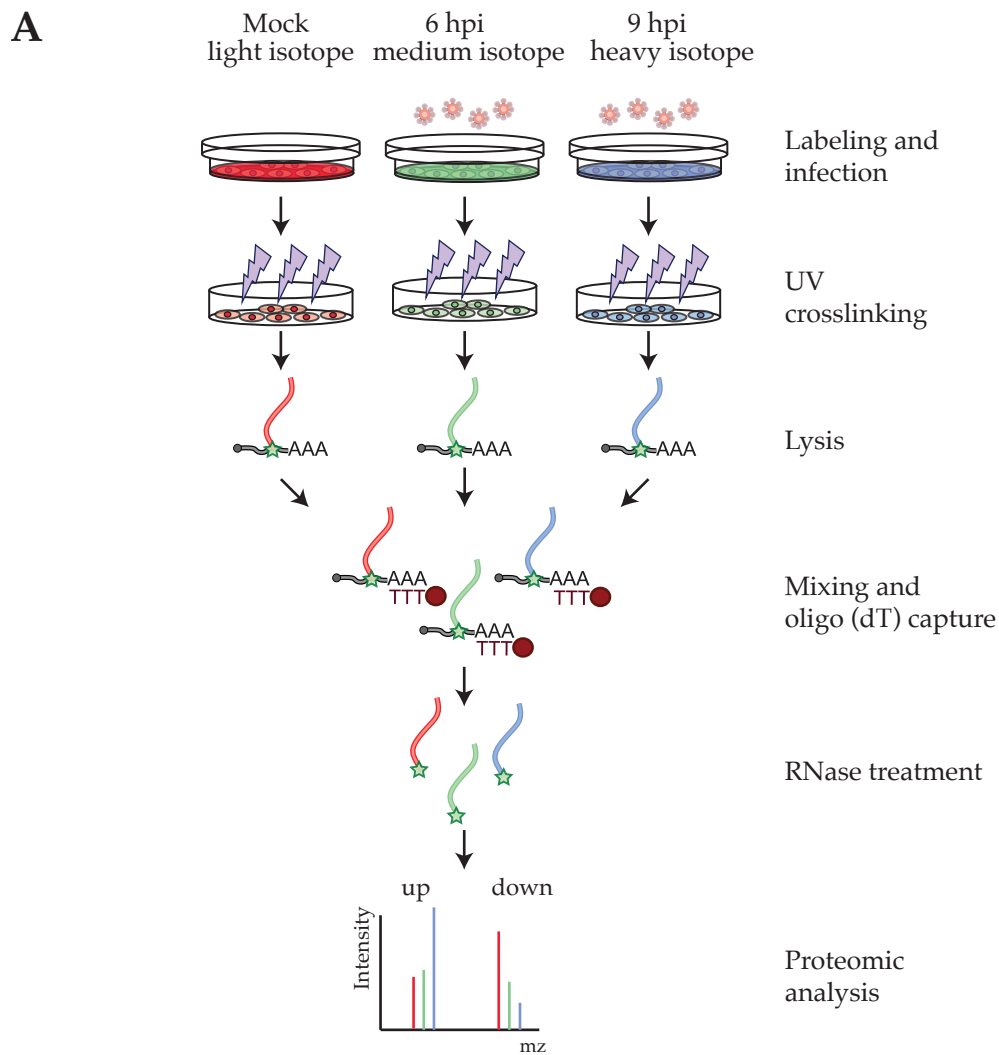


Fig. 6.8 RBPome capture of infected A549 cells. (A) Schematic of RBPome capture in infected cells. SILAC-labelled A549 cells are infected with influenza A/WSN/33 (H1N1) virus for 6h or 9h. Proteins and nucleic acids are crosslinked by UV irradiation. Samples are mixed and mRNA-binding proteins are isolated from mixed cell lysates using oligo(dT) magnetic beads. Captured RNA-binding proteins are treated with RNase and analysed by mass spectrometry. (B) Isolated proteins were analysed by SDS-PAGE and silver staining. Molecular masses, indicated on the right, are depicted in kilodaltons.

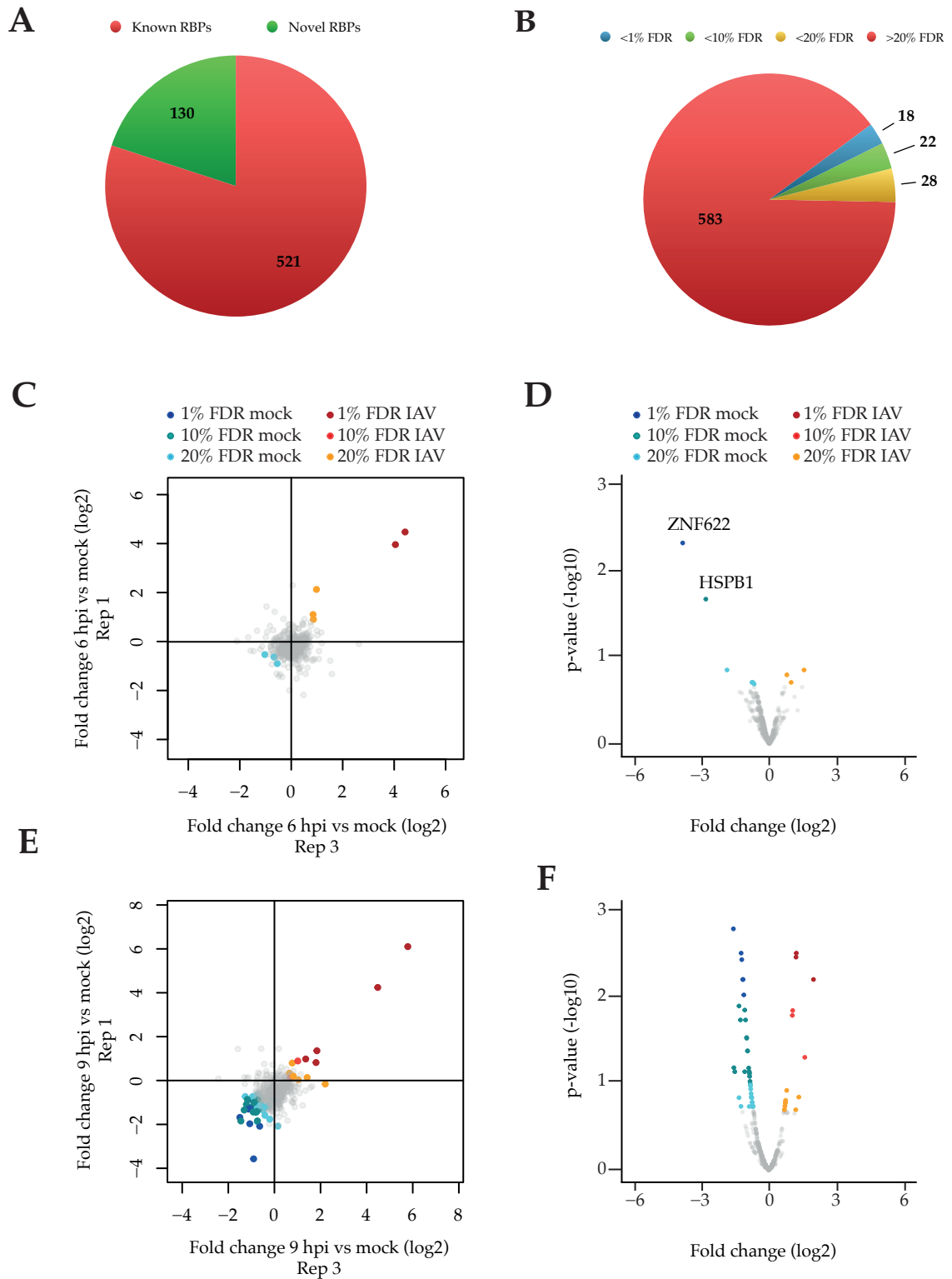


Fig. 6.9 The dynamic response of the RBPome to influenza A virus infection. (A) Known vs. novel RBPs identified 9 h.p.i. in human A549 cells. (B) False discovery rate (FDR) and number of identified RNA-binding proteins changing their RNA-binding behaviour in response to influenza A virus infection (9 h.p.i.). (C) Fold change of identified RBP intensities 6 h.p.i. compared to mock-infected cells. Here replicates 1 and 3 are displayed as representatives of all three independent biological replicates. (D) Statistical significance of the fold change displayed in C. (E) Fold change of identified RBP intensities 9 h.p.i. compared to mock-infected cells. Here replicates 1 and 3 are displayed as representatives of all three independent biological replicates. (F) Statistical significance of the fold change displayed in E. Analysis performed by Ni Shuai and Dr Bernd Fischer (DKFZ, Germany).

(Fig. 6.9D). 9 h.p.i. larger fold changes in protein abundance were observed (Fig. 6.9E) which also correlated with higher statistical significance of changes in abundance (Fig. 6.9F). The statistical significance was determined using t-tests and p-values were adjusted using Benjamini-Hochberg corrections.

RBP that were significantly enriched or depleted ($p < 0.1$) in cells infected 9 h.p.i. compared to 6 h.p.i. or uninfected samples were analysed by hierarchical clustering and presented as a heat map (Fig. 6.10A). 9 h.p.i. 68 proteins were significantly ($p < 0.2$) enriched or depleted compared to mock, 19 of which exhibited increased and 49 decreased RNA-binding activity. Two viral proteins (NP and NS1) were identified as RBPs, however semi-quantitative analysis also revealed association of HA, M1, PB1 and PB2 with RNA. A complete list of all identified RBPs and their binding features in infected cells is attached in Appendix K.

GO term analysis of significantly enriched and depleted RBPs shows association of enriched RBPs with ribonucleoprotein complex biogenesis and RNA processing (Fig. 6.10B) of depleted proteins with RNA processing, splicing and stability (Fig. 6.10C).

6.2.5 Protein abundance does not drive changes in the RBPome

Interactome capture revealed a small number of RBPs that were significantly enriched or depleted in infected cells. A change in association with RNA could be due to a change in protein levels or an altered RNA-binding behaviour. In order to understand whether protein abundance is affected by viral infection, the WCL from the RNA-interactome experiments was analysed by mass spectrometry as detailed before. No changes in protein abundance were detected at 6 h.p.i. (Fig. 6.11A) or 9 h.p.i. (Fig. 6.11B), except for viral proteins and the host protein HIST1H3J (Table 6.1).

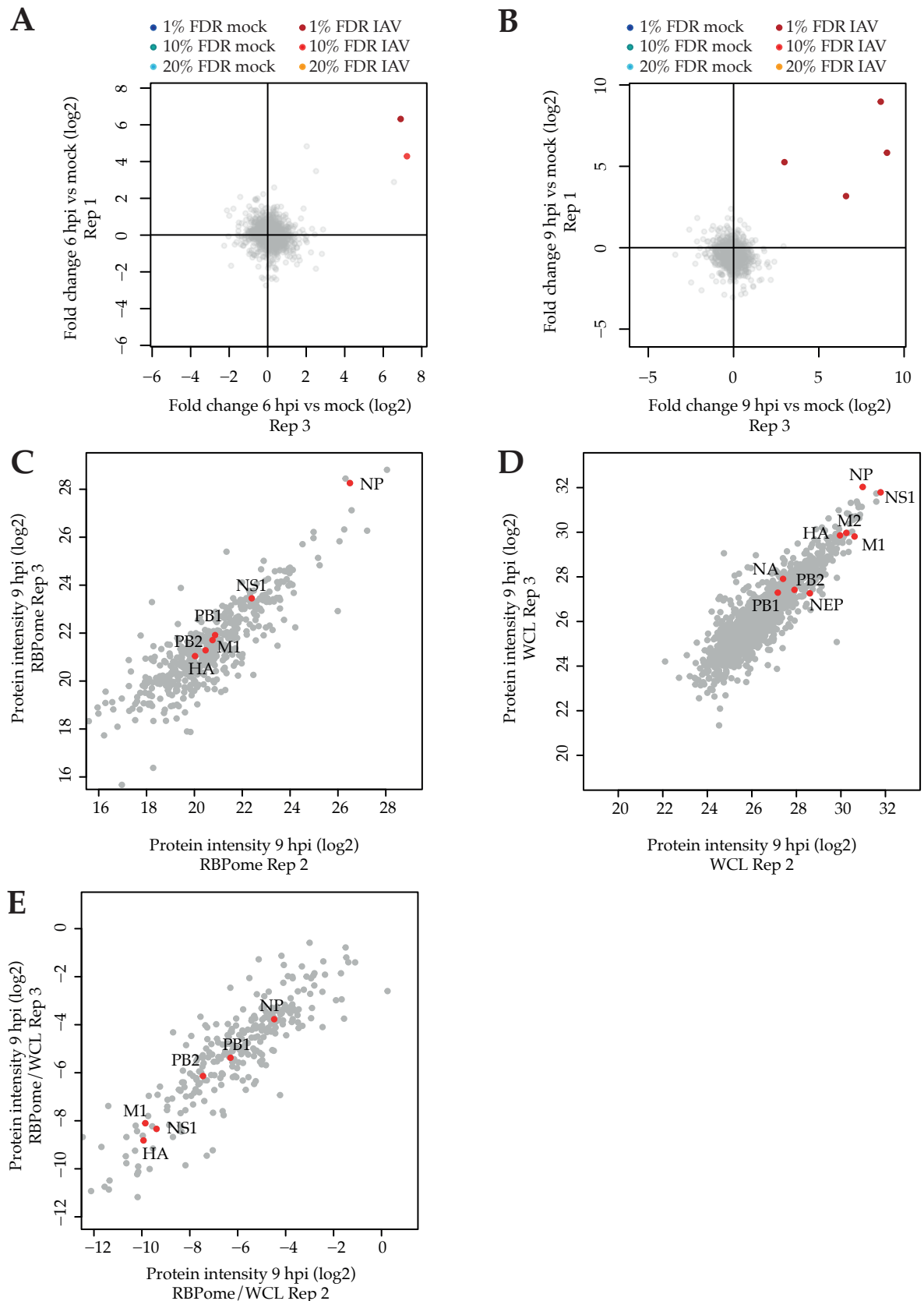


Fig. 6.11 **Changes in RBPome are not driven by protein abundance in the WCL.** (A) Fold change of protein intensities in the WCL 6 h.p.i. compared to mock-infected cells. (B) Fold change of protein intensities in the WCL 9 h.p.i. compared to mock-infected cells. In (A) and (B) replicates 1 and 3 are displayed as representatives of all three independent biological replicates. (C) Protein intensities in the RBPome 9 h.p.i. (D) Protein intensities in the WCL 9 h.p.i. (E) Ratio of protein intensities between the RBPome and WCL. Viral proteins are highlighted. Replicates 2 and 3 are used as representatives of all three independent biological replicates. Analysis performed by Ni Shuai and Dr Bernd Fischer (DKFZ, Germany).

Table 6.1 Proteins that are significantly ($p < 0.2$) altered in their abundance 9 h.p.i. in the WCL

ENSGid	Gene	Known RBP	FC (log2)	Adjusted p -value	Up/Down
Inf_Q20MH3_NS1	Inf_NS1	no	8.076	5.223E-06	up
Inf_P15682_NCAP	Inf_NP	no	6.943	9.963E-05	up
Inf_P03454_HEMA	Inf_HA	no	4.217	5.981E-04	up
Inf_Q67183_M1	Inf_M1	no	5.363	9.290E-04	up
Inf_P05780_M2	Inf_M2	no	5.978	1.394E-03	up
ENSG00000197153	HIST1H3J	no	-4.999	4.572E-03	down

On a global level this suggests that a change in enrichment in the RBPome is not due to a change in protein abundance and suggests that the RNA-binding of these proteins is altered. The ratio between the fold change of protein intensities in infected (9 h) and mock infected cells of individual proteins in the RBPome and WCL can be used as a determinant for altered RNA-binding behaviour. If the fold change of enriched proteins in the infected RBPome is larger than the fold change in the WCL, then viral infection induces an increase in RNA-binding which is independent of protein abundance. Accordingly, if the fold change of depleted proteins in the infected RBPome is larger than the fold change in the WCL, then these proteins exhibit a decrease in RNA-binding after viral infection. As seen in Table 6.2, all significantly enriched host RBPs exhibited an increase in RNA-binding which was independent of increased protein abundance. Similarly, all significantly depleted host RBPs, except for ZC3H7B, showed decreased RNA-binding independently of protein abundance, as shown in Table 6.3. This shows, that influenza virus infection has profound effects on the host cell, regulating the RNA-binding behaviour of a range of RBPs. However, it is not possible to tell whether the altered RNA-binding behaviour is induced by an altered mRNA abundance or by an altered binding-specificity.

In Chapter 6.2.4 the viral proteins NP, NS1, PB1, PB2, HA and M1 are identified as RBPs. Comparison of proteome intensities and the relative abundance of viral proteins in the proteome in the captured RBPome (Fig. 6.11C) and the WCL (Fig. 6.11D) revealed a different relative abundance of viral proteins in the RBPome compared to the WCL. The fraction of a viral protein recovered in the interactome compared to its expression in the WCL was taken as a measure for the ability of viral proteins to interact with RNA. According to this measure, NP was most efficient at binding RNA, followed by the viral RdRP subunits PB1 and PB2, whereas M1, NS1 and HA were inefficient at binding of RNA (Fig. 6.11E).

Table 6.2 Comparison of RNA-binding activity of significantly enriched RBPs in the RBPome and the WCL 9 h.p.i.

Gene	RBPome FC (log2)	WCL FC (log2)	RBPome FC / WCL FC	Increased RNA-binding
Inf_NP	6.068	6.943	-0.875	no
Inf_NS1	4.228	8.076	-3.848	no
VEZT	7.217	-	-	-
GNL3L	1.765	0.269	1.496	yes
PABPC1	1.229	-0.221	1.450	yes
PABPC4	1.216	0.255	0.961	yes
NMD3	1.995	-	-	-
WDR43	1.066	-	-	-
ADAR	1.052	-0.272	1.323	yes
SF1	1.607	-0.175	1.782	yes
ATXN2	0.799	0.668	0.131	yes
KPNA2	1.342	-0.696	2.038	yes
NIFK	0.752	-0.654	1.406	yes
KNOP1	0.744	0.133	0.611	yes
NOL11	0.735	-0.279	1.014	yes
SNRPF	0.721	-0.584	1.305	yes
PWP2	0.697	-	-	-
DHX29	1.203	-0.315	1.518	yes
ZFR	0.693	-0.484	1.177	yes

Table 6.3 Comparison of RNA-binding activity of significantly depleted RBPs in the RBPome and the WCL 9 h.p.i.

Gene	RBPome FC (log2)	WCL FC (log2)	RBPome FC / WCL FC	Decreased RNA-binding
ZNF622	-3.185	-	-	-
SF3B4	-1.660	-0.065	-1.595	yes
TMEM94	-2.005	-	-	-
HSPB1	-2.684	0.154	-2.837	yes
UPF1	-1.534	-0.066	-1.469	yes
ZC3H14	-1.579	-0.257	-1.321	yes
SKIV2L2	-1.234	-0.230	-1.004	yes
CELF1	-1.205	0.361	-1.566	yes
KRT18	-1.159	-0.302	-0.857	yes
R3HDM1	-1.150	-	-	-
ALYREF	-1.113	-0.133	-0.980	yes
PCBP3	-1.328	-0.022	-1.306	yes
DRG1	-1.071	-0.143	-0.929	yes
SRP9	-1.266	0.354	-1.620	yes
TIAL1	-1.036	-0.289	-0.747	yes
MBNL1	-0.990	-	-	-
STAU2	-0.984	-0.478	-0.507	yes
R3HDM2	-0.949	-	-	-
ELAVL1	-0.947	0.069	-1.016	yes
CSTF2	-1.556	-0.634	-0.922	yes
CHTOP	-0.895	-0.443	-0.451	yes
HSPA6	-1.516	-0.189	-1.327	yes
HERC5	-1.073	-0.167	-0.906	yes
PUM1	-0.878	-0.361	-0.517	yes
TIA1	-0.873	-0.059	-0.814	yes
FUBP3	-0.865	0.015	-0.880	yes
RGAG4	-0.864	1.201	-2.065	yes
RBM47	-0.853	-0.103	-0.750	yes

PTBP3	-0.837	-0.412	-0.425	yes
MAP4	-0.826	-0.118	-0.709	yes
CPEB2	-0.822	0.036	-0.857	yes
RBPMS	-0.817	-	-	-
MSN	-0.805	-0.519	-0.286	yes
RBM10	-0.788	-	-	-
TKT	-0.773	-0.055	-0.718	yes
EEF1G	-0.770	-0.010	-0.760	yes
CLUH	-1.327	-0.429	-0.898	yes
EIF2D	-0.763	-	-	-
SRSF10	-0.762	0.173	-0.936	yes
SERBP1	-0.742	-0.134	-0.608	yes
DAZAP1	-0.740	-0.238	-0.502	yes
EDF1	-0.727	-0.462	-0.265	yes
FAM120A	-0.721	0.214	-0.935	yes
ZNFX1	-1.240	-	-	-
HNRNPDL	-0.718	-0.198	-0.520	yes
PCBP2	-0.714	-0.022	-0.693	yes
CCDC9	-0.872	-	-	-
SUMO1	-0.709	-0.341	-0.368	yes
ZC3H7B	-0.708	-1.501	0.793	no

To validate the altered RNA-binding activity of identified RBPs upon viral infection, the RNA-interactome capture was repeated and isolated proteins were analysed by SDS-PAGE and western blotting. Of the six viral proteins identified as RBPs by mass spectrometry, four (PB1, PB2, NP and NS1) could be validated by western blot. PABPC1, ADAR and KPNA2 were all confirmed to increase RNA-binding, while expression levels in the WCL remained stable (PABPC1 had reduced expression 9 h post-infection in the WCL). XRCC5, HNRNPR and PTBP1 exhibited reduced RNA-binding, as uncovered by the proteomics analysis. ACTB did not interact with RNA as shown previously (Fig. 6.12). Taken together, proteomic analysis of the WCL shows that protein abundance is not significantly affected by influenza virus infection and that the change in the RBPome induced by infection is mediated by altered RNA-binding behaviour and not protein expression. Viral proteins were also shown to have differential efficiencies in RNA-binding.

6.3 Discussion

The aim of this study was to investigate the global response of RNA-binding proteins to influenza virus infection. This study characterised for the first time the RBPome of the human lung epithelial A549 cell line, revealing a RBPome which resembled previously established RBPomes (Baltz et al., 2012; Castello et al., 2012; Kwon et al., 2013). The kinetics of influenza A virus gene expression and the host cell response was analysed over a time course of infection, revealing a peak at 9 h.p.i. of viral mRNA abundance. Furthermore, upregulation of genes linked to the innate immune response and ion transport and downregulation of genes associated with RNA processing was seen after viral infection. Influenza virus infection also induced alternative polyadenylation of host genes, mainly in genes linked to intracellular transport and RNA processing. Infection

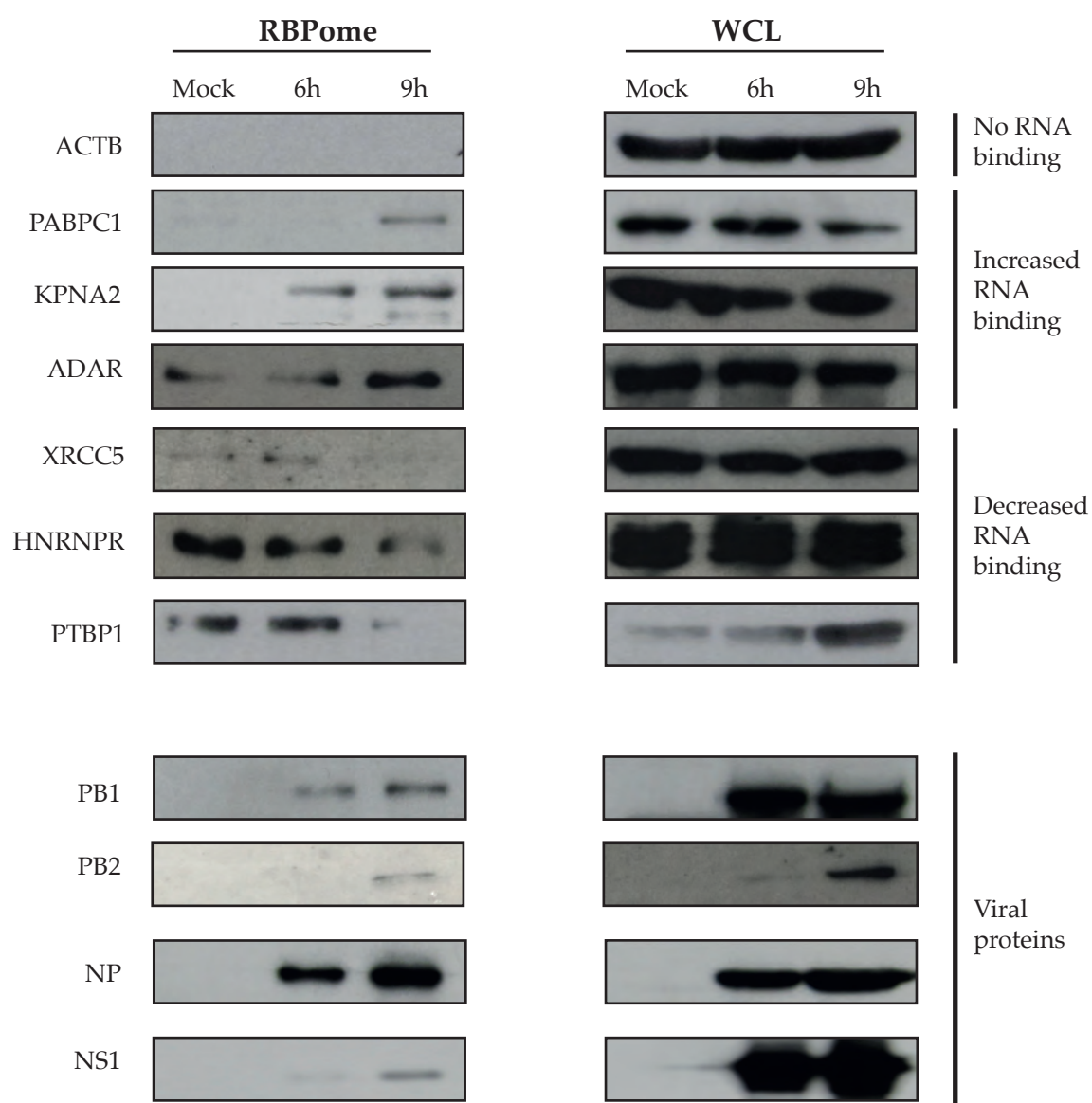


Fig. 6.12 Validation of mass spectrometry data. RBP and WCL samples were analysed by SDS-PAGE and western blotting with antibodies against the indicated proteins.

with influenza A virus induced a dynamic response of the host RBPome which was independent of protein abundance. The viral proteins, NP, NS1, PB1, PB2, HA and M1 were all shown to be RNA-binding proteins. Of these viral proteins, NP and the RdRP subunits PB1 and PB2 appear to be particularly efficient mRNA binders.

The kinetics of influenza A/WSN/33 (H1N1) virus RNA synthesis appears to be dependent on the host cell line. Previous studies in human HEK-239T cells showed a similar vRNA, mRNA and cRNA synthesis profile as presented here, although the peak of mRNA levels was observed earlier in infection (4 to 6 h.p.i.) compared to A549 cells (Robb et al., 2009; Vreede et al., 2011). Further analysis by RNA sequencing revealed that 9 h.p.i. a significant fraction of all mRNA in the cell was viral (28%). This is in line with previous work, showing an even higher fraction of viral mRNA (57%) 8 h.p.i. with influenza A/Puerto Rico/8/1934 (H1N1) virus (Bercovich-Kinori et al., 2016). Surprisingly, after 9 h.p.i. a sharp decrease in mRNA abundance could be seen in WSN infection. Unfortunately, no data is available for the influenza A/Puerto Rico/8/1934 virus at these late time points. However, a reduction in viral mRNA levels has also been observed by primer extension in HEK-239T (Robb et al., 2009; Vreede et al., 2011) as well as by pulse-labelling in BHK-21 cells (Shapiro et al., 1987). A reduction of viral mRNA late in infection could be explained by a host shut off and reduced host transcription and therefore also viral transcription. Host transcripts seen late in infection could be highly stable mRNAs with long half lives. Together with a reduction in viral and host transcription, but enrichment for host mRNAs with a long half live, this would increase the relative amount of host mRNA compared to viral mRNA in the cell. Nascent RNA sequence analysis could address the question whether the host transcriptional rate is reduced at a similar rate as viral transcription.

It was found that infection with influenza A virus induced APA of host transcripts. No preference for a shortening or lengthening of 3' UTRs was seen, however, a strong association of shortened transcripts with cellular localisation and transport and of

lengthened transcripts with RNA processing was detected. APA can have profound effects on post-transcriptional gene regulation by affecting mRNA stability, translation, nuclear export and cellular localisation and can be induced by a wide range of biological stimuli (Neve et al., 2017). A systematic shortening of 3' UTRs has been reported in cancer cells, cells infected by vesicular stomatitis virus (VSV) and in proliferating cells (Jia et al., 2017; Mayr & Bartel, 2009; Sandberg et al., 2008). A systematic lengthening of 3' UTRs was reported in differentiating cells and nerve cells (Ji et al., 2010; Miura et al., 2013; Shepard et al., 2011). In VSV infections, APA is primarily linked to an upregulation of the antiviral host response, but also to intracellular transport and localisation of proteins (Jia et al., 2017). Infection with HSV-1 was shown to induce APA in genes associated with chromatin modification, intracellular transport and regulation of transcription (Hu et al., 2016). Here, influenza virus induced APA is not linked to the innate immune response, but instead to intracellular transport on the one hand and RNA processing on the other hand. Analysis of differential gene expression revealed that influenza virus induced a strong up regulation of antiviral pathways, translation and intracellular transport, as well as a down regulation of the RNA processing machinery. Together, this suggests that influenza virus infection induces enhanced gene expression related to intracellular transport through shortening of 3' UTRs and reduced gene expression related to RNA processing through lengthening of 3' UTRs. RNA processing is essential for viral replication and a shut down of the host transcriptional machinery could be induced by the host as an antiviral response. Interestingly, a common feature of viral infections appears to be APA of genes related to intracellular transport. Together with an increased expression of these genes, this points to a general mechanism of how the host transport machinery is utilised in response to viral infections.

Several hundreds of RBPs were identified in A549 cells, 396 in uninfected and 651 in infected A549 cells. Viral infection induced a dynamic response of the RBPome, resulting

in a relatively small set of RBPs that were significantly enriched or depleted after infection. Enriched RBPs were mainly associated with ribosome biogenesis, while depleted RBPs were associated with mRNA splicing and processing. This response of the RBPome was independent of protein abundance, but instead, was linked to a change in RNA-binding behaviour. It remains to be established whether increased RNA binding corresponds to increased expression of certain mRNA transcripts. Whole transcriptome analysis by RNA sequencing will be performed to address this question.

Among the RBPs that are enriched in viral infection is ADAR1. ADAR1 is an RNA editing enzyme that deaminates adenosine (A) into inosine (I) in double-stranded RNA. ADAR1 has two alternative gene transcripts, ADAR1-L and ADAR1-S, of which the longer ADAR1-L transcript can be induced by IFN- $\alpha/\beta/\gamma$ (Samuel, 2001). Although it is an interferon-stimulated gene, it has been shown to suppress IFN signalling (Hartner et al., 2009) and can possess both proviral or antiviral functions in a multitude of viral infections, including hepatitis delta virus (Wong & Lazinski, 2002), measles virus (Cattaneo et al., 1988), vesicular stomatitis virus (Jia et al., 2017), human parainfluenza virus (Murphy et al., 1991), respiratory syncytial virus (Martínez et al., 1997), HIV-1 (Doria et al., 2009), lymphocytic choriomeningitis virus (Zahn et al., 2007), Rift Valley fever virus (Suspene et al., 2008), mumps virus (Chambers et al., 2009), dengue virus (de Chassey et al., 2013) and hepatitis C (Taylor et al., 2005) as well as the DNA virus mouse polyoma virus (Kumar & Carmichael, 1997).

Interestingly, ADAR1 has been shown to be significantly enriched during influenza A virus infections and change from a mainly cytoplasmic localisation to being enriched in the nucleolus (Emmott et al., 2010). Hyperediting of influenza virus RNAs has been observed and attributed to ADAR1, without conclusively showing that ADAR1 is responsible for the RNA editing (Suspene et al., 2011; tenOever et al., 2007; van Wielink et al., 2012). In these studies hyperediting was demonstrated in the HA, M and NS segments,

both in negative-sense and positive-sense viral RNAs. ADAR1 has been shown to interact with the viral NS1 protein, which is also a double-stranded RNA binding protein involved in post-transcriptional RNA regulation (de Chassey et al., 2013; Ngamurulert et al., 2009). Initially ADAR1 was thought to have an antiviral role in influenza virus infection due to its induction by interferon and increased susceptibility to influenza virus infection and cytopathic effects in mice when ADAR1 was silenced (tenOever et al., 2007; Ward et al., 2011). Increased hyperediting was observed in the absence of NS1, suggesting that NS1 has an inhibitory function on ADAR1 (van Wielink et al., 2012). However, in another study inhibition of ADAR1 led to a decrease in viral protein synthesis and viral replication, suggesting a proviral role of ADAR1 in influenza virus infection (de Chassey et al., 2013). This is thought to be mediated by an interaction with NS1, possibly by inhibiting PKR signalling (de Chassey et al., 2013; Li et al., 2006; Toth et al., 2009).

Its pro- and antiviral function appears to be both RNA editing-dependent and editing-independent. The role of ADAR1 in suppressing IFN stimulation in influenza virus infection needs to be investigated further. Furthermore, analysis of the effect of hyperediting of the influenza virus genome needs to be done to understand whether these mutations are beneficial to the virus. It is tempting to speculate that ADAR also could have a potential antiviral role in influenza virus infections mediated by its RNA editing ability. Influenza viruses rely on their RNA structure and sequence for packaging (Fournier et al., 2012b; 2012a; Gavazzi et al., 2013; Hutchinson et al., 2010). Editing of viral RNA sequences required for efficient genome packaging could lead to incorrect packaging and defective interfering viral particles. RNA sequencing should be performed and analysed for RNA editing. The effect of silencing or over expression of ADAR on viral genome replication, protein expression, viral growth and generation of infectious particles could be assessed as well.

Another unregulated RBPs in viral infection is KPNA2, which is a well-known host factor important for nuclear import of the viral RdRP subunit PB2 (Gabriel et al., 2008; Resa-Infante et al., 2008). Structural studies of the viral RdRP have shown that the RdRP exists in different conformations (Hengrung et al., 2015; Pflug et al., 2014; Thierry et al., 2016). One could speculate that association of KPNA2 with PB2 allows the RdRP to remain in the transcriptionally active conformation and that KPNA2 therefore could form some interaction with host or viral mRNA through its close proximity to the cap-binding domain and product exit channels.

The most significantly enriched RBP is vezatin (VEZT), which is a transmembrane protein located in adherens junctions and shown to bind myosin VIIA. VEZT has been observed in the cell nucleus as well (Hyenne et al., 2005). Upregulation of VEZT has been observed in patients infected with dengue virus exhibiting symptoms of the dengue shock syndrome (Loke et al., 2010).

Among the viral proteins, NP, NS1, PB1, PB2, M1 and HA were shown to be binding mRNA. Of these NP in particular, but also PB1 and PB2 were able to efficiently bind mRNA. NS1 has been shown to bind to host mRNA before (Marión et al., 1997; Qiu & Krug, 1994) and its function in host shut off has been studied extensively. NP, which appears to be the most abundant RNA-binding protein in influenza virus infection, has not been linked to host or viral mRNA binding before. However, NP is the major component of viral RNP complexes. Viral RNA is wound around a double helical NP structure in viral RNPs mediated through a basic groove on the surface of NP monomers (Arranz et al., 2012). It is therefore possible that viral NP monomers or oligomers form RNP structures with host mRNAs as well. One could even speculate that these NP interactions with host mRNA could help induce host shut off by engaging host mRNAs and thereby reducing the amount of translatable host transcripts. The viral RdRP interacts with capped host RNAs through the cap-binding domain of PB2, but cap-snatching

results in cleavage of host mRNAs before they are polyadenylated. However, the viral RdRP is subsequently associated with viral mRNA which eventually also is polyadenylated (Velthuis & Fodor, 2016). Viral M1 has been shown to be able to specifically interact with single-stranded RNA (Wakefield & Brownlee, 1989). This interaction is known to be mediated through the NLS region of M1 (Elster et al., 1997). No interaction of HA with RNA has been reported before.

6.4 Summary

In summary, the RBPome of A549 cells was characterised and 130 novel RBPs were confidently identified in this study. Analysis of the human transcriptome after influenza virus infection demonstrated the importance of RNA metabolism in response virus infection and highlighted underlying similarities to infections with other RNA viruses. A dynamic response of the RBPome which was independent of protein abundance was described in response to influenza virus infection, revealing a pool of candidate proteins that could have potential pro- or antiviral functions. Further analysis of these candidate RBPs is currently being performed.

REFERENCES

- Aggarwal, S., Dewhurst, S., Takimoto, T. & Kim, B. (2011). Biochemical impact of the host adaptation-associated PB2 E627K mutation on the temperature-dependent RNA synthesis kinetics of influenza A virus polymerase complex. *The Journal of biological chemistry*. 286 (40). pp. 34504–34513.
- Air, G.M. & Laver, W.G. (1989). The Neuraminidase of Influenza-Virus. *Proteins-Structure Function and Genetics*. 6 (4). pp. 341–356.
- Ali, A., Avalos, R.T., Ponimaskin, E. & Nayak, D.P. (2000). Influenza virus assembly: Effect of influenza virus glycoproteins on the membrane association of M1 protein. *Journal of Virology*. 74 (18). pp. 8709–8719.
- Almond, J.W. (1977). A single gene determines the host range of influenza virus. *Nature*. 270 (5638). pp. 617–618.
- Amorim, M.J., Bruce, E.A., Read, E.K.C., Foeglein, A., Mahen, R., Stuart, A.D. & Digard, P. (2011). A Rab11-and Microtubule-Dependent Mechanism for Cytoplasmic Transport of Influenza A Virus Viral RNA. *Journal of Virology*. 85 (9). pp. 4143–4156.
- Amorim, M.J., Read, E.K., Dalton, R.M., Medcalf, L. & Digard, P. (2007). Nuclear export of influenza A virus mRNAs requires ongoing RNA polymerase II activity. *Traffic*. 8 (1). pp. 1–11.
- Anantharaman, V., Koonin, E.V. & Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic acids research*. 30 (7). pp. 1427–1464.
- Area, E., Martin-Benito, J., Gastaminza, P., Torreira, E., Valpuesta, J.M., Carrascosa, J.L. & Ortín, J. (2004). 3D structure of the influenza virus polymerase complex: Localization of subunit domains. *Proceedings of the National Academy of Sciences of the United States of America*. 101 (1). pp. 308–313.
- Arranz, R., Coloma, R., Chichón, F.J., Conesa, J.J., Carrascosa, J.L., Valpuesta, J.M., Ortín, J. & Martín-Benito, J. (2012). The structure of native influenza virion ribonucleoproteins. *Science (New York, N.Y.)*. 338 (6114). pp. 1634–1637.
- Ayora-Talavera, G., Shelton, H., Scull, M.A., Ren, J., Jones, I.M., Pickles, R.J. & Barclay, W.S. (2009). Mutations in H5N1 influenza virus hemagglutinin that confer binding to human tracheal airway epithelium. C. A. Stoddart (ed.). *Plos One*. 4 (11). p. e7836.
- Azzeh, M., Flick, R. & Hobom, G. (2001). Functional analysis of the influenza A virus cRNA promoter and construction of an ambisense transcription system. *Virology*. 289 (2). pp. 400–410.
- Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C. & Landthaler, M. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular cell*. 46 (5). pp. 674–690.

- Baudin, F., Bach, C., Cusack, S. & Ruigrok, R.W.H. (1994). Structure of Influenza-Virus Rnp .1. Influenza-Virus Nucleoprotein Melts Secondary Structure in Panhandle Rna and Exposes the Bases to the Solvent. *Embo Journal*. 13 (13). pp. 3158–3165.
- Baudin, F., Petit, I., Weissenhorn, W. & Ruigrok, R.W. (2001). In vitro dissection of the membrane and RNP binding activities of influenza virus M1 protein. *Virology*. 281 (1). pp. 102–108.
- Bavagnoli, L., Cucuzza, S., Campanini, G., Rovida, F., Paolucci, S., Baldanti, F. & Maga, G. (2015). The novel influenza A virus protein PA-X and its naturally deleted variant show different enzymatic properties in comparison to the viral endonuclease PA. *Nucleic acids research*. 43 (19). pp. 9405–9417.
- Beaton, A.R. & Krug, R.M. (1981). Selected Host-Cell Capped Rna Fragments Prime Influenza Viral-Rna Transcription In vivo. *Nucleic acids research*. 9 (17). pp. 4423–4436.
- Beckmann, B.M., Castello, A. & Medenbach, J. (2016). The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflugers Archiv : European journal of physiology*. 468 (6). pp. 1029–1040.
- Beckmann, B.M., Horos, R., Fischer, B., Castello, A., Eichelbaum, K., Alleaume, A.-M., Schwarzl, T., Curk, T., Foehr, S., Huber, W., Krijgsveld, J. & Hentze, M.W. (2015). The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nature Communications*. 6. p. 10127.
- Bercovich-Kinori, A., Tai, J., Gelbart, I.A., Shitrit, A., Ben-Moshe, S., Drori, Y., Itzkovitz, S., Mandelboim, M. & Stern-Ginossar, N. (2016). A systematic view on influenza induced host shutoff. *eLife*. 5. p. 600.
- Bergmann, M. & Muster, T. (1995). The Relative Amount of an Influenza-a Virus Segment Present in the Viral Particle Is Not Affected by a Reduction in Replication of That Segment. *The Journal of general virology*. 76 (12). pp. 3211–3215.
- Bier, K., York, A. & Fodor, E. (2011). Cellular cap-binding proteins associate with influenza virus mRNAs. *The Journal of general virology*. 92 (Pt 7). pp. 1627–1634.
- Biswas, S.K., Boutz, P.L. & Nayak, D.P. (1998). Influenza virus nucleoprotein interacts with influenza virus polymerase proteins. *Journal of Virology*. 72 (7). pp. 5493–5501.
- Blaas, D., Patzelt, E. & Kuechler, E. (1982). Identification of the Cap Binding-Protein of Influenza-Virus. *Nucleic acids research*. 10 (15). pp. 4803–4812.
- Bogs, J., Kalthoff, D., Veits, J., Pavlova, S., Schwemmle, M., Mänz, B., Mettenleiter, T.C. & Stech, J. (2011). Reversion of PB2-627E to -627K during replication of an H5N1 Clade 2.2 virus in mammalian hosts depends on the origin of the nucleoprotein. *Journal of Virology*. 85 (20). pp. 10691–10698.
- Bortz, E., Westera, L., Maamary, J., Steel, J., Albrecht, R.A., Manicassamy, B., Chase, G., Martínez-Sobrido, L., Schwemmle, M. & García-Sastre, A. (2011). Host- and strain-specific regulation of influenza virus polymerase activity by interacting cellular proteins. *Mbio*. 2 (4). pp. e00151–11–e00151–11.

- Boulay, F., Doms, R.W., Wilson, I. & Helenius, A. (1987). The influenza hemagglutinin precursor as an acid-sensitive probe of the biosynthetic pathway. *Embo Journal*. 6 (9). pp. 2643–2650.
- Bradel-Tretheway, B.G., Mattiaccio, J.L., Krasnoselsky, A., Stevenson, C., Purdy, D., Dewhurst, S. & Katze, M.G. (2011). Comprehensive proteomic analysis of influenza virus polymerase complex reveals a novel association with mitochondrial proteins and RNA polymerase accessory factors. *Journal of Virology*. 85 (17). pp. 8569–8581.
- Brooke, C.B., Ince, W.L., Wrammert, J., Ahmed, R., Wilson, P.C., Bennink, J.R. & Yewdell, J.W. (2013). Most Influenza A Virions Fail To Express at Least One Essential Viral Protein (vol 87, pg 3155, 2013). *Journal of Virology*. 87 (14). pp. 8267–8267.
- Bruce, E.A., Digard, P. & Stuart, A.D. (2010). The Rab11 Pathway Is Required for Influenza A Virus Budding and Filament Formation. *Journal of Virology*. 84 (12). pp. 5848–5859.
- Brunotte, L., Flies, J., Bolte, H., Reuther, P., Vreede, F. & Schwemmler, M. (2014). The Nuclear Export Protein of H5N1 Influenza A Viruses Recruits Matrix 1 (M1) Protein to the Viral Ribonucleoprotein to Mediate Nuclear Export. *The Journal of biological chemistry*. 289 (29). pp. 20067–20077.
- Bui, M., Whittaker, G. & Helenius, A. (1996). Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins. *Journal of Virology*. 70 (12). pp. 8391–8401.
- Bui, M., Wills, E.G., Helenius, A. & Whittaker, G.R. (2000). Role of the influenza virus M1 protein in nuclear export of viral ribonucleoproteins. *Journal of Virology*. 74 (4). pp. 1781–1786.
- Bullough, P.A., Hughson, F.M., Skehel, J.J. & Wiley, D.C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature*. 371 (6492). pp. 37–43.
- Bussey, K.A., Bousse, T.L., Desmet, E.A., Kim, B. & Takimoto, T. (2010). PB2 residue 271 plays a key role in enhanced polymerase activity of influenza A viruses in mammalian host cells. *Journal of Virology*. 84 (9). pp. 4395–4406.
- Bussey, K.A., Desmet, E.A., Mattiaccio, J.L., Hamilton, A., Bradel-Tretheway, B., Bussey, H.E., Kim, B., Dewhurst, S. & Takimoto, T. (2011). PA residues in the 2009 H1N1 pandemic influenza virus enhance avian influenza virus polymerase activity in mammalian cells. *Journal of Virology*. 85 (14). pp. 7020–7028.
- Calder, L.J., Wasilewski, S., Berriman, J.A. & Rosenthal, P.B. (2010). Structural organization of a filamentous influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America*. 107 (23). pp. 10685–10690.
- Carmody, S.R. & Wente, S.R. (2009). mRNA nuclear export at a glance. *Journal of cell science*. 122 (Pt 12). pp. 1933–1937.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., Krijgsvelde, J. & Hentze, M.W. (2012). Insights into RNA biology from an atlas of mammalian RNA-binding proteins. *Cell*. 149 (6). pp. 1393–1406.

- Castello, A., Horos, R., Strein, C., Fischer, B., Eichelbaum, K., Steinmetz, L.M., Krijgsveld, J. & Hentze, M.W. (2013). System-wide identification of RNA-binding proteins by interactome capture. *Nature Protocols*. 8 (3). pp. 491–500.
- Cattaneo, R., Schmid, A., Eschle, D., Bacsko, K., Meulen, ter, V. & Billeter, M.A. (1988). Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell*. 55 (2). pp. 255–265.
- Cauldwell, A.V., Long, J.S., Moncorge, O. & Barclay, W.S. (2014). Viral determinants of influenza A virus host range. *The Journal of general virology*. 95 (Pt 6). pp. 1193–1210.
- Cauldwell, A.V., Moncorge, O. & Barclay, W.S. (2013). Unstable polymerase-nucleoprotein interaction is not responsible for avian influenza virus polymerase restriction in human cells. *Journal of Virology*. 87 (2). pp. 1278–1284.
- Chambers, P., Rima, B.K. & Duprex, W.P. (2009). Molecular differences between two Jeryl Lynn mumps virus vaccine component strains, JL5 and JL2. *The Journal of general virology*. 90 (Pt 12). pp. 2973–2981.
- Chan, A.Y., Vreede, F.T., Smith, M., Engelhardt, O.G. & Fodor, E. (2006). Influenza virus inhibits RNA polymerase II elongation. *Virology*. 351 (1). pp. 210–217.
- Chan, W.-H., Ng, A.K.-L., Robb, N.C., Lam, M.K.-H., Chan, P.K.-S., Au, S.W.-N., Wang, J.-H., Fodor, E. & Shaw, P.-C. (2010). Functional analysis of the influenza virus H5N1 nucleoprotein tail loop reveals amino acids that are crucial for oligomerization and ribonucleoprotein activities. *Journal of Virology*. 84 (14). pp. 7337–7345.
- Chang, S., Sun, D., Liang, H., Wang, J., Li, J., Guo, L., Wang, X., Guan, C., Boruah, B.M., Yuan, L., Feng, F., Yang, M., Wang, L., Wang, Y., Wojdyla, J., Li, L., Wang, J., Wang, M., Cheng, G., Wang, H.-W. & Liu, Y. (2015). Cryo-EM structure of influenza virus RNA polymerase complex at 4.3 Å resolution. *Molecular cell*. 57 (5). pp. 925–935.
- Chen, B.J., Leser, G.P., Jackson, D. & Lamb, R.A. (2008). The Influenza Virus M2 Protein Cytoplasmic Tail Interacts with the M1 Protein and Influences Virus Assembly at the Site of Virus Budding. *Journal of Virology*. 82 (20). pp. 10059–10070.
- Chen, B.J., Leser, G.P., Morita, E. & Lamb, R.A. (2007a). Influenza virus hemagglutinin and neuraminidase, but not the matrix protein, are required for assembly and budding of plasmid-derived virus-like particles. *Journal of Virology*. 81 (13). pp. 7111–7123.
- Chen, G.-W., Chang, S.-C., Mok, C.-K., Lo, Y.-L., Kung, Y.-N., Huang, J.-H., Shih, Y.-H., Wang, J.-Y., Chiang, C., Chen, C.-J. & Shih, S.-R. (2006). Genomic signatures of human versus avian influenza A viruses. *Emerging infectious diseases*. 12 (9). pp. 1353–1360.
- Chen, H., Bright, R.A., Subbarao, K., Smith, C., Cox, N.J., Katz, J.M. & Matsuoka, Y. (2007b). Polygenic virulence factors involved in pathogenesis of 1997 Hong Kong H5N1 influenza viruses in mice. *Virus Research*. 128 (1-2). pp. 159–163.
- Chen, J., Lee, K.H., Steinhauer, D.A., Stevens, D.J., Skehel, J.J. & Wiley, D.C. (1998). Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell*. 95 (3). pp. 409–417.

- Chen, T.H., Brody, J.R., Romantsev, F.E., Yu, J.G., Kayler, A.E., Voneiff, E., Kuhajda, F.P. & Pasternack, G.R. (1996). Structure of pp32, an acidic nuclear protein which inhibits oncogene-induced formation of transformed foci. *Molecular biology of the cell*. 7 (12). pp. 2045–2056.
- Chen, W., Calvo, P.A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., Basta, S., O'Neill, R., Schickli, J., Palese, P., Henklein, P., Bennink, J.R. & Yewdell, J.W. (2001). A novel influenza A virus mitochondrial protein that induces cell death. *Nature Medicine*. 7 (12). pp. 1306–1312.
- Chen, Z.Y. & Krug, R.M. (2000). Selective nuclear export of viral mRNAs in influenza-virus-infected cells. *Trends in Microbiology*. 8 (8). pp. 376–383.
- Chen, Z.Y., Li, Y.Z. & Krug, R.M. (1999). Influenza A virus NS1 protein targets poly(A)-binding protein II of the cellular 3'-end processing machinery. *Embo Journal*. 18 (8). pp. 2273–2283.
- Cheng, K., Yu, Z., Chai, H., Sun, W., Xin, Y., Zhang, Q., Huang, J., Zhang, K., Li, X., Yang, S., Wang, T., Zheng, X., Wang, H., Qin, C., Qian, J., Chen, H., Hua, Y., Gao, Y. & Xia, X. (2014). PB2-E627K and PA-T97I substitutions enhance polymerase activity and confer a virulent phenotype to an H6N1 avian influenza virus in mice. *Virology*. 468-470. pp. 207–213.
- Chou, Y.-Y., Vafabakhsh, R., Doğanay, S., Gao, Q., Ha, T. & Palese, P. (2012). One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis. *Proceedings of the National Academy of Sciences of the United States of America*. 109 (23). pp. 9101–9106.
- Chua, M.A., Schmid, S., Perez, J.T., Langlois, R.A. & tenOever, B.R. (2013). Influenza A Virus Utilizes Suboptimal Splicing to Coordinate the Timing of Infection. *Cell Reports*. 3 (1). pp. 23–29.
- Chutinimitkul, S., Herfst, S., Steel, J., Lowen, A.C., Ye, J., van Riel, D., Schrauwen, E.J.A., Bestebroer, T.M., Koel, B., Burke, D.F., Sutherland-Cash, K.H., Whittleston, C.S., Russell, C.A., Wales, D.J., Smith, D.J., Jonges, M., Meijer, A., Koopmans, M., Rimmelzwaan, G.F., Kuiken, T., Osterhaus, A.D.M.E., García-Sastre, A., Perez, D.R. & Fouchier, R.A.M. (2010). Virulence-Associated Substitution D222G in the Hemagglutinin of 2009 Pandemic Influenza A(H1N1) Virus Affects Receptor Binding. *Journal of Virology*. 84 (22). pp. 11802–11813.
- Cianci, C., Tiley, L. & Krystal, M. (1995). Differential Activation of the Influenza-Virus Polymerase via Template Rna-Binding. *Journal of Virology*. 69 (7). pp. 3995–3999.
- Claas, E.C., Osterhaus, A.D., van Beek, R., De Jong, J.C., Rimmelzwaan, G.F., Senne, D.A., Krauss, S., Shortridge, K.F. & Webster, R.G. (1998). Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet (London, England)*. 351 (9101). pp. 472–477.
- Coloma, R., Valpuesta, J.M., Arranz, R., Carrascosa, J.L., Ortín, J. & Martín-Benito, J. (2009). The Structure of a Biologically Active Influenza Virus Ribonucleoprotein Complex F. A. Rey (ed.). *PLoS pathogens*. 5 (6).

- Compans, R.W., Content, J. & Duesberg, P.H. (1972). Structure of Ribonucleoprotein of Influenza Virus. *Journal of Virology*. 10 (4). pp. 795–&.
- Connor, R.J., Kawaoka, Y., Webster, R.G. & Paulson, J.C. (1994). Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology*. 205 (1). pp. 17–23.
- Cook, P.R. (2010). A Model for all Genomes: The Role of Transcription Factories. *Journal of molecular biology*. 395 (1). pp. 1–10.
- Coombs, K.M., Berard, A., Xu, W., Krokhin, O., Meng, X., Cortens, J.P., Kobasa, D., Wilkins, J. & Brown, E.G. (2010). Quantitative proteomic analyses of influenza virus-infected cultured human lung cells. *Journal of Virology*. 84 (20). pp. 10888–10906.
- Cooper, T.A., Wan, L. & Dreyfuss, G. (2009). RNA and Disease. *Cell*. 136 (4). pp. 777–793.
- Couceiro, J.N., Paulson, J.C. & Baum, L.G. (1993). Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus Research*. 29 (2). pp. 155–165.
- Cox, J. & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*. 26 (12). pp. 1367–1372.
- Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V. & Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research*. 10 (4). pp. 1794–1805.
- Crescenzo-Chaigne, B., van der Werf, S. & Naffakh, N. (2002). Differential effect of nucleotide substitutions in the 3' arm of the influenza A virus vRNA promoter on transcription/replication by avian and human polymerase complexes is related to the nature of PB2 amino acid 627. *Virology*. 303 (2). pp. 240–252.
- Cros, J.F., García-Sastre, A. & Palese, P. (2005). An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein. *Traffic*. 6 (3). pp. 205–213.
- Crow, M., Deng, T., Addley, M. & Brownlee, G.G. (2004). Mutational analysis of the influenza virus cRNA promoter and identification of nucleotides critical for replication. *Journal of Virology*. 78 (12). pp. 6263–6270.
- Darnell, R.B. (2010). RNA Regulation in Neurologic Disease and Cancer. *Cancer Research and Treatment*. 42 (3). pp. 125–129.
- Dawood, F.S., Iuliano, A.D., Reed, C., Meltzer, M.I., Shay, D.K., Cheng, P.-Y., Bandaranayake, D., Breiman, R.F., Brooks, W.A., Buchy, P., Feikin, D.R., Fowler, K.B., Gordon, A., Hien, N.T., Horby, P., Huang, Q.S., Katz, M.A., Krishnan, A., Lal, R., Montgomery, J.M., Mølbak, K., Pebody, R., Presanis, A.M., Razuri, H., Steens, A., Tinoco, Y.O., Wallinga, J., Yu, H., Vong, S., Bresee, J. & Widdowson, M.-A. (2012). Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *The Lancet. Infectious diseases*. 12 (9). pp. 687–695.

- de Chasse, B., Aublin-Gex, A., Ruggieri, A., Meyniel-Schicklin, L., Pradezynski, F., Davoust, N., Chantier, T., Tafforeau, L., Mangeot, P.-E., Cancia, C., Perrin-Cocon, L., Bartenschlager, R., André, P. & Lotteau, V. (2013). The interactomes of influenza virus NS1 and NS2 proteins identify new host factors and provide insights for ADAR1 playing a supportive role in virus replication. *PLoS pathogens*. 9 (7). p. e1003440.
- de Chiara, C., Kelly, G., Frenkiel, T.A. & Pastore, A. (2007). NMR assignment of the leucine-rich repeat domain of LANP / Anp32a. *Journal of biomolecular NMR*. 38 (2). pp. 177–177.
- de Jong, M.D., Simmons, C.P., Thanh, T.T., Hien, V.M., Smith, G.J.D., Chau, T.N.B., Hoang, D.M., Chau, N.V.V., Khanh, T.H., Dong, V.C., Qui, P.T., Cam, B.V., Ha, D.Q., Guan, Y., Peiris, J.S.M., Chinh, N.T., Hien, T.T. & Farrar, J. (2006). Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nature Medicine*. 12 (10). pp. 1203–1207.
- Deng, T., Engelhardt, O.G., Thomas, B., Akoulitchiev, A.V., Brownlee, G.G. & Fodor, E. (2006a). Role of ran binding protein 5 in nuclear import and assembly of the influenza virus RNA polymerase complex. *Journal of Virology*. 80 (24). pp. 11911–11919.
- Deng, T., Sharps, J., Fodor, E. & Brownlee, G.G. (2005). In vitro assembly of PB2 with a PB1-PA dimer supports a new model of assembly of influenza A virus polymerase subunits into a functional trimeric complex. *Journal of Virology*. 79 (13). pp. 8669–8674.
- Deng, T., Vreede, F.T. & Brownlee, G.G. (2006b). Different *de novo* initiation strategies are used by influenza virus RNA polymerase on its cRNA and viral RNA promoters during viral RNA replication. *Journal of Virology*. 80 (5). pp. 2337–2348.
- Desselberger, U., Racaniello, V.R., Zazra, J.J. & Palese, P. (1980). 3'-Terminal and 5'-Terminal Sequences of Influenza-a, Influenza-B and Influenza-C Virus-Rna Segments Are Highly Conserved and Show Partial Inverted Complementarity. *Gene*. 8 (3). pp. 315–328.
- Dias, A., Bouvier, D., Crepin, T., McCarthy, A.A., Hart, D.J., Baudin, F., Cusack, S. & Ruigrok, R.W.H. (2009). The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*. 458 (7240). pp. 914–918.
- Ding, X., Lu, J., Yu, R., Wang, X., Wang, T., Dong, F., Peng, B., Wu, W., Liu, H., Geng, Y., Zhang, R., Ma, H., Cheng, J., Yu, M. & Fang, S. (2016). Preliminary Proteomic Analysis of A549 Cells Infected with Avian Influenza Virus H7N9 and Influenza A Virus H1N1. S. Sambhara (ed.). *Plos One*. 11 (5). p. e0156017.
- Domingues, P. & Hale, B.G. (2017). Functional Insights into ANP32A-Dependent Influenza A Virus Polymerase Host Restriction. *Cell Reports*. 20 (11). pp. 2538–2546.
- Doria, M., Neri, F., Gallo, A., Farace, M.G. & Michienzi, A. (2009). Editing of HIV-1 RNA by the double-stranded RNA deaminase ADAR1 stimulates viral infection. *Nucleic acids research*. 37 (17). pp. 5848–5858.
- Dove, B.K., Surtees, R., Bean, T.J.H., Munday, D., Wise, H.M., Digard, P., Carroll, M.W., Ajuh, P., Barr, J.N. & Hiscox, J.A. (2012). A quantitative proteomic analysis of lung epithelial (A549) cells infected with 2009 pandemic influenza A virus using stable isotope labelling with amino acids in cell culture. *Proteomics*. 12 (9). pp. 1431–1436.

- Dubois, J., Terrier, O. & Rosa-Calatrava, M. (2014). Influenza Viruses and mRNA Splicing: Doing More with Less. *Mbio*. 5 (3). pp. –14.
- Duesberg, P.H. (1969). Distinct subunits of the ribonucleoprotein of influenza virus. *Journal of molecular biology*. 42 (3). pp. 485–499.
- Eccles, R. (2005). Understanding the symptoms of the common cold and influenza. *The Lancet. Infectious diseases*. 5 (11). pp. 718–725.
- Edinger, T.O., Pohl, M.O. & Stertz, S. (2014). Entry of influenza A virus: host factors and antiviral targets. *The Journal of general virology*. 95 (Pt 2). pp. 263–277.
- Elster, C., Larsen, K., Gagnon, J., Ruigrok, R.W. & Baudin, F. (1997). Influenza virus M1 protein binds to RNA through its nuclear localization signal. *The Journal of general virology*. 78 (Pt 7) (7). pp. 1589–1596.
- Elton, D., Medcalf, L., Bishop, K., Harrison, D. & Digard, P. (1999). Identification of amino acid residues of influenza virus nucleoprotein essential for RNA binding. *Journal of Virology*. 73 (9). pp. 7357–7367.
- Elton, D., Simpson-Holley, M., Archer, K., Medcalf, L., Hallam, R., McCauley, J. & Digard, P. (2001). Interaction of the influenza virus nucleoprotein with the cellular CRM1-mediated nuclear export pathway. *Journal of Virology*. 75 (1). pp. 408–419.
- Emmott, E., Wise, H., Loucaides, E.M., Matthews, D.A., Digard, P. & Hiscox, J.A. (2010). Quantitative proteomics using SILAC coupled to LC-MS/MS reveals changes in the nucleolar proteome in influenza A virus-infected cells. *Journal of proteome research*. 9 (10). pp. 5335–5345.
- Enami, M., Fukuda, R. & Ishihama, A. (1985). Transcription and replication of eight RNA segments of influenza virus. *Virology*. 142 (1). pp. 68–77.
- Engelhardt, O.G. & Fodor, E. (2006). Functional association between viral and cellular transcription during influenza virus infection. *Reviews in medical virology*. 16 (5). pp. 329–345.
- Engelhardt, O.G., Smith, M. & Fodor, E. (2005). Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II. *Journal of Virology*. 79 (9). pp. 5812–5818.
- Fechter, P., Mingay, L., Sharps, J., Chambers, A., Fodor, E. & Brownlee, G.G. (2003). Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding. *The Journal of biological chemistry*. 278 (22). pp. 20381–20388.
- Fields, S. & Winter, G. (1982). Nucleotide sequences of influenza virus segments 1 and 3 reveal mosaic structure of a small viral RNA segment. *Cell*. 28 (2). pp. 303–313.
- Fields, S., Winter, G. & Brownlee, G.G. (1981). Structure of the Neuraminidase Gene in Human Influenza-Virus a-Pr-8-34. *Nature*. 290 (5803). pp. 213–217.
- Finkelstein, D.B., Mukatira, S., Mehta, P.K., Obenauer, J.C., Su, X., Webster, R.G. & Naeve, C.W. (2007). Persistent host markers in pandemic and H5N1 influenza viruses. *Journal of Virology*. 81 (19). pp. 10292–10299.

- Firth, A.E., Jagger, B.W., Wise, H.M., Nelson, C.C., Parsawar, K., Wills, N.M., Napthine, S., Taubenberger, J.K., Digard, P. & Atkins, J.F. (2012). Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC_UUU_CGU and is in the +1 direction. *Open Biology*. 2 (10).
- Flick, R. & Hobom, G. (1999). Interaction of influenza virus polymerase with viral RNA in the 'corkscrew' conformation. *The Journal of general virology*. 80 (10). pp. 2565–2572.
- Flick, R., Neumann, G., Hoffmann, E., Neumeier, E. & Hobom, G. (1996). Promoter elements in the influenza vRNA terminal structure. *Rna-a Publication of the Rna Society*. 2 (10). pp. 1046–1057.
- Fodor, E. & Smith, M. (2004). The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex. *Journal of Virology*. 78 (17). pp. 9144–9153.
- Fodor, E., Crow, M., Mingay, L.J., Deng, T., Sharps, J., Fechter, P. & Brownlee, G.G. (2002). A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *Journal of Virology*. 76 (18). pp. 8989–9001.
- Fodor, E., Devenish, L., Engelhardt, O.G., Palese, P., Brownlee, G.G. & García-Sastre, A. (1999). Rescue of influenza A virus from recombinant DNA. *Journal of Virology*. 73 (11). pp. 9679–9682.
- Fodor, E., Pritlove, D.C. & Brownlee, G.G. (1995). Characterization of the RNA-fork model of virion RNA in the initiation of transcription in influenza A virus. *Journal of Virology*. 69 (7). pp. 4012–4019.
- Fodor, E., Pritlove, D.C. & Brownlee, G.G. (1994). The influenza virus panhandle is involved in the initiation of transcription. *Journal of Virology*. 68 (6). pp. 4092–4096.
- Foeglein, A., Loucaides, E.M., Mura, M., Wise, H.M., Barclay, W.S. & Digard, P. (2011). Influence of PB2 host-range determinants on the intranuclear mobility of the influenza A virus polymerase. *The Journal of general virology*. 92 (Pt 7). pp. 1650–1661.
- Fournier, E., Moules, V., Essere, B., Paillart, J.-C., Sirbat, J.-D., Cavalier, A., Rolland, J.-P., Thomas, D., Lina, B., Isel, C. & Marquet, R. (2012a). Interaction network linking the human H3N2 influenza A virus genomic RNA segments. *Vaccine*. 30 (51). pp. 7359–7367.
- Fournier, E., Moules, V., Essere, B., Paillart, J.-C., Sirbat, J.-D., Isel, C., Cavalier, A., Rolland, J.-P., Thomas, D., Lina, B. & Marquet, R. (2012b). A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic acids research*. 40 (5). pp. 2197–2209.
- Fournier, G., Chiang, C., Munier, S., Tomoiu, A. & Demeret, C. (2014). Recruitment of RED-SMU1 Complex by Influenza A Virus RNA Polymerase to Control Ciral mRNA Splicing (vol 10, e1004164, 2014). *PLoS pathogens*. 10 (7).

- Fries, B., Heukeshoven, J., Hauber, I., Grüttner, C., Stocking, C., Kehlenbach, R.H., Hauber, J. & Chemnitz, J. (2007). Analysis of nucleocytoplasmic trafficking of the HuR ligand APRIL and its influence on CD83 expression. *The Journal of biological chemistry*. 282 (7). pp. 4504–4515.
- Gabriel, G. & Fodor, E. (2014). Molecular determinants of pathogenicity in the polymerase complex. *Current topics in microbiology and immunology*. 385 (Chapter 386). pp. 35–60.
- Gabriel, G., Dauber, B., Wolff, T., Planz, O., Klenk, H.-D. & Stech, J. (2005). The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proceedings of the National Academy of Sciences of the United States of America*. 102 (51). pp. 18590–18595.
- Gabriel, G., Herwig, A. & Klenk, H.-D. (2008). Interaction of polymerase subunit PB2 and NP with importin alpha1 is a determinant of host range of influenza A virus. *PLoS pathogens*. 4 (2). p. e11.
- Gabriel, G., Klingel, K., Otte, A., Thiele, S., Hudjetz, B., Arman-Kalcek, G., Sauter, M., Schmidt, T., Rother, F., Baumgarte, S., Keiner, B., Hartmann, E., Bader, M., Brownlee, G.G., Fodor, E. & Klenk, H.-D. (2011). Differential use of importin- α isoforms governs cell tropism and host adaptation of influenza virus. *Nature Communications*. 2. p. 156.
- Gao, Q., Chou, Y.-Y., Doğanay, S., Vafabakhsh, R., Ha, T. & Palese, P. (2012). The Influenza A Virus PB2, PA, NP, and M Segments Play a Pivotal Role during Genome Packaging. *Journal of Virology*. 86 (13). pp. 7043–7051.
- Gao, Y., Zhang, Y., Shinya, K., Deng, G., Jiang, Y., Li, Z., Guan, Y., Tian, G., Li, Y., Shi, J., Liu, L., Zeng, X., Bu, Z., Xia, X., Kawaoka, Y. & Chen, H. (2009). Identification of amino acids in HA and PB2 critical for the transmission of H5N1 avian influenza viruses in a mammalian host. B. Sherry (ed.). *PLoS pathogens*. 5 (12). p. e1000709.
- Garaigorta, U. & Ortín, J. (2007). Mutation analysis of a recombinant NS replicon shows that influenza virus NS1 protein blocks the splicing and nucleo-cytoplasmic transport of its own viral mRNA. *Nucleic acids research*. 35 (14). pp. 4573–4582.
- Gastaminza, P., Perales, B., Falcon, A.M. & Ortín, J. (2003). Mutations in the N-terminal region of influenza virus PB2 protein affect virus RNA replication but not transcription. *Journal of Virology*. 77 (9). pp. 5098–5108.
- Gavazzi, C., Isel, C., Fournier, E., Moules, V., Cavalier, A., Thomas, D., Lina, B. & Marquet, R. (2013). An *in vitro* network of intermolecular interactions between viral RNA segments of an avian H5N2 influenza A virus: comparison with a human H3N2 virus. *Nucleic acids research*. 41 (2). pp. 1241–1254.
- Gerber, M., Moules, V., Gavazzi, C., Marquet, R. & Isel, C. (2014). Selective packaging of influenza A virus genome. *Virologie*. 18 (5). pp. 278–296.
- Gerstberger, S., Hafner, M. & Tuschl, T. (2014a). A census of human RNA-binding proteins. *Nature Reviews Genetics*. 15 (12). pp. 829–845.
- Gerstberger, S., Hafner, M., Ascano, M. & Tuschl, T. (2014b). Evolutionary Conservation and Expression of Human RNA-Binding Proteins and Their Role in Human Genetic Disease. *Systems Biology of Rna Binding Proteins*. 825 (Chapter 1). pp. 1–55.

- Glisovic, T., Bachorik, J.L., Yong, J. & Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters*. 582 (14). pp. 1977–1986.
- Gonzalez, S. & Ortín, J. (1999). Distinct regions of influenza virus PB1 polymerase subunit recognize vRNA and cRNA templates. *Embo Journal*. 18 (13). pp. 3767–3775.
- Gu, W., Gallagher, G.R., Dai, W., Liu, P., Li, R., Trombly, M.I., Gammon, D.B., Mello, C.C., Wang, J.P. & Finberg, R.W. (2015). Influenza A virus preferentially snatches noncoding RNA caps. *Rna-a Publication of the Rna Society*. 21 (12). pp. 2067–2075.
- Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W.H., Ortín, J., Hart, D.J. & Cusack, S. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Structural & Molecular Biology*. 15 (5). pp. 500–506.
- Hale, B.G., Randall, R.E., Ortín, J. & Jackson, D. (2008). The multifunctional NS1 protein of influenza A viruses. *The Journal of general virology*. 89 (Pt 10). pp. 2359–2376.
- Hamilton, B.S., Whittaker, G.R. & Daniel, S. (2012). Influenza virus-mediated membrane fusion: determinants of hemagglutinin fusogenic activity and experimental approaches for assessing virus fusion. *Viruses*. 4 (7). pp. 1144–1168.
- Hao, L., Sakurai, A., Watanabe, T., Sorensen, E., Nidom, C.A., Newton, M.A., Ahlquist, P. & Kawaoka, Y. (2008). Drosophila RNAi screen identifies host genes important for influenza virus replication. *Nature*. 454 (7206). pp. 890–U46.
- Hara, K., Kashiwagi, T., Hamada, N. & Watanabe, H. (2017). Basic amino acids in the N-terminal half of the PB2 subunit of influenza virus RNA polymerase are involved in both transcription and replication. *The Journal of general virology*. 98 (5). pp. 900–905.
- Hara, K., Schmidt, F.I., Crow, M. & Brownlee, G.G. (2006). Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. *Journal of Virology*. 80 (16). pp. 7789–7798.
- Harris, A., Cardone, G., Winkler, D.C., Heymann, J.B., Brecher, M., White, J.M. & Steven, A.C. (2006). Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proceedings of the National Academy of Sciences of the United States of America*. 103 (50). pp. 19123–19127.
- Hartner, J.C., Walkley, C.R., Lu, J. & Orkin, S.H. (2009). ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. *Nature immunology*. 10 (1). pp. 109–115.
- Hatada, E., Hasegawa, M., Mukaigawa, J., Shimizu, K. & Fukuda, R. (1989). Control of influenza virus gene expression: quantitative analysis of each viral RNA species in infected cells. *Journal of biochemistry*. 105 (4). pp. 537–546.
- Hatta, M., Gao, P., Halfmann, P. & Kawaoka, Y. (2001). Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science (New York, N.Y.)*. 293 (5536). pp. 1840–1842.

- Hatta, M., Hatta, Y., Kim, J.H., Watanabe, S., Shinya, K., Lien, T.N.P.S., Le, Q.M. & Kawaoka, Y. (2007). Growth of H5N1 influenza A viruses in the upper respiratory tracts of mice. *PLoS pathogens*. 3 (10). pp. 1374–1379.
- Hayashi, T., Wills, S., Bussey, K.A. & Takimoto, T. (2015). Identification of Influenza A Virus PB2 Residues Involved in Enhanced Polymerase Activity and Virus Growth in Mammalian Cells at Low Temperatures. D. S. Lyles (ed.). *Journal of Virology*. 89 (15). pp. 8042–8049.
- He, X., Zhou, J., Bartlam, M., Zhang, R., Ma, J., Lou, Z., Li, X., Li, J., Joachimiak, A., Zeng, Z., Ge, R., Rao, Z. & Liu, Y. (2008). Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature*. 454 (7208). pp. 1123–U51.
- Heggeness, M.H., Smith, P.R., Ulmanen, I., Krug, R.M. & Choppin, P.W. (1982). Studies on the Helical Nucleocapsid of Influenza-Virus. *Virology*. 118 (2). pp. 466–470.
- Hengrung, N., Omari, El, K., Serna Martin, I., Vreede, F.T., Cusack, S., Rambo, R.P., Vonnrhein, C., Bricogne, G., Stuart, D.I., Grimes, J.M. & Fodor, E. (2015). Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. *Nature*. 527 (7576). pp. 114–117.
- Herfst, S., Chutinimitkul, S., Ye, J., de Wit, E., Munster, V.J., Schrauwen, E.J.A., Bestebroer, T.M., Jonges, M., Meijer, A., Koopmans, M., Rimmelzwaan, G.F., Osterhaus, A.D.M.E., Perez, D.R. & Fouchier, R.A.M. (2010). Introduction of Virulence Markers in PB2 of Pandemic Swine-Origin Influenza Virus Does Not Result in Enhanced Virulence or Transmission. *Journal of Virology*. 84 (8). pp. 3752–3758.
- Herfst, S., Schrauwen, E.J.A., Linster, M., Chutinimitkul, S., de Wit, E., Munster, V.J., Sorrell, E.M., Bestebroer, T.M., Burke, D.F., Smith, D.J., Rimmelzwaan, G.F., Osterhaus, A.D.M.E. & Fouchier, R.A.M. (2012). Airborne transmission of influenza A/H5N1 virus between ferrets. *Science (New York, N.Y.)*. 336 (6088). pp. 1534–1541.
- Hershey, J.W.B., Sonenberg, N. & Mathews, M.B. (2012). Principles of Translational Control: An Overview. *Cold Spring Harbor Perspectives in Biology*. 4 (12). pp. –a011528.
- Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G. & Webster, R.G. (2000). A DNA transfection system for generation of influenza A virus from eight plasmids. *Proceedings of the National Academy of Sciences of the United States of America*. 97 (11). pp. 6108–6113.
- Hofmann, P., Sprenger, H., Kaufmann, A., Bender, A., Hasse, C., Nain, M. & Gems, D. (1997). Susceptibility of mononuclear phagocytes to influenza A virus infection and possible role in the antiviral response. *Journal of leukocyte biology*. 61 (4). pp. 408–414.
- Hohlbein, J., Aigrain, L., Craggs, T.D., Bermek, O., Potapova, O., Shoolizadeh, P., Grindley, N.D.F., Joyce, C.M. & Kapanidis, A.N. (2013). Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion. *Nature Communications*. 4. p. 2131.
- Hong, R., Macfarlan, T., Kutney, S.N., Seo, S.-B., Mukai, Y., Yelin, F., Pasternack, G.R. & Chakravarti, D. (2004). The identification of phosphorylation sites of pp32 and biochemical purification of a cellular pp32-kinase. *Biochemistry*. 43 (31). pp. 10157–10165.

- Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., Park, J.Y., Yehia, G. & Tian, B. (2013). Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nature methods*. 10 (2). pp. 133–139.
- Hsu, M.T., Parvin, J.D., Gupta, S., Krystal, M. & Palese, P. (1987). Genomic RNAs of Influenza-Viruses Are Held in a Circular Conformation in Virions and in Infected-Cells by a Terminal Panhandle. *Proceedings of the National Academy of Sciences of the United States of America*. 84 (22). pp. 8140–8144.
- Hu, B., Li, X., Huo, Y., Yu, Y., Zhang, Q., Chen, G., Zhang, Y., Fraser, N.W., Wu, D. & Zhou, J. (2016). Cellular responses to HSV-1 infection are linked to specific types of alterations in the host transcriptome. *Scientific reports*. 6 (1). p. 28075.
- Hu, M., Yuan, S., Zhang, K., Singh, K., Ma, Q., Zhou, J., Chu, H. & Zheng, B.-J. (2017). PB2 substitutions V598T/I increase the virulence of H7N9 influenza A virus in mammals. *Virology*. 501. pp. 92–101.
- Hu, Y.-J., Tu, P.-C., Lin, C.-S. & Guo, S.-T. (2014). Identification and chronological analysis of genomic signatures in influenza A viruses. Y. Wang (ed.). *Plos One*. 9 (1). p. e84638.
- Huang, I.-C., Li, W., Sui, J., Marasco, W., Choe, H. & Farzan, M. (2008). Influenza A virus neuraminidase limits viral superinfection. *Journal of Virology*. 82 (10). pp. 4834–4843.
- Huang, T.S., Palese, P. & Krystal, M. (1990). Determination of Influenza-Virus Proteins Required for Genome Replication. *Journal of Virology*. 64 (11). pp. 5669–5673.
- Huang, X., Liu, T., Muller, J., Levandowski, R.A. & Ye, Z. (2001). Effect of influenza virus matrix protein and viral RNA on ribonucleoprotein formation and nuclear export. *Virology*. 287 (2). pp. 405–416.
- Hudjetz, B. & Gabriel, G. (2012). Human-like PB2 627K influenza virus polymerase activity is regulated by importin- α 1 and - α 7. *PLoS pathogens*. 8 (1). p. e1002488.
- Huet, S., Avilov, S.V., Ferbitz, L., Daigle, N., Cusack, S. & Ellenberg, J. (2010). Nuclear import and assembly of influenza A virus RNA polymerase studied in live cells by fluorescence cross-correlation spectroscopy. *Journal of Virology*. 84 (3). pp. 1254–1264.
- Hutchinson, E.C. & Fodor, E. (2013). Transport of the influenza virus genome from nucleus to nucleus. *Viruses*. 5 (10). pp. 2424–2446.
- Hutchinson, E.C., Charles, P.D., Hester, S.S., Thomas, B., Trudgian, D., Martínez-Alonso, M. & Fodor, E. (2014). Conserved and host-specific features of influenza virion architecture. *Nature Communications*. 5. p. 4816.
- Hutchinson, E.C., Kirchbach, von, J.C., Gog, J.R. & Digard, P. (2010). Genome packaging in influenza A virus. *The Journal of general virology*. 91 (Pt 2). pp. 313–328.
- Huyton, T. & Wolberger, C. (2007). The crystal structure of the tumor suppressor protein pp32 (Anp32a): structural insights into Anp32 family of proteins. *Protein science : a publication of the Protein Society*. 16 (7). pp. 1308–1315.

- Hyenne, V., Louvet-Vallée, S., El-Amraoui, A., Petit, C., Maro, B. & Simmler, M.-C. (2005). Vezatin, a protein associated to adherens junctions, is required for mouse blastocyst morphogenesis. *Developmental biology*. 287 (1). pp. 180–191.
- Ilyushina, N.A., Khalenkov, A.M., Seiler, J.P., Forrest, H.L., Bovin, N.V., Marjuki, H., Barman, S., Webster, R.G. & Webby, R.J. (2010). Adaptation of pandemic H1N1 influenza viruses in mice. *Journal of Virology*. 84 (17). pp. 8607–8616.
- Imai, M., Watanabe, T., Hatta, M., Das, S.C., Ozawa, M., Shinya, K., Zhong, G., Hanson, A., Katsura, H., Watanabe, S., Li, C., Kawakami, E., Yamada, S., Kiso, M., Suzuki, Y., Maher, E.A., Neumann, G. & Kawaoka, Y. (2012). Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/ H1N1 virus in ferrets. *Nature*. 486 (7403). pp. 420–428.
- Ito, T., Couceiro, J.N., Kelm, S., Baum, L.G., Krauss, S., Castrucci, M.R., Donatelli, I., Kida, H., Paulson, J.C., Webster, R.G. & Kawaoka, Y. (1998). Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. *Journal of Virology*. 72 (9). pp. 7367–7373.
- Jagger, B.W., Memoli, M.J., Sheng, Z.-M., Qi, L., Hrabal, R.J., Allen, G.L., Dugan, V.G., Wang, R., Digard, P., Kash, J.C. & Taubenberger, J.K. (2010). The PB2-E627K mutation attenuates viruses containing the 2009 H1N1 influenza pandemic polymerase. *Mbio*. 1 (1). pp. e00067–10–e00067–10.
- Jagger, B.W., Wise, H.M., Kash, J.C., Walters, K.-A., Wills, N.M., Xiao, Y.-L., Dunfee, R.L., Schwartzman, L.M., Ozinsky, A., Bell, G.L., Dalton, R.M., Lo, A., Efstathiou, S., Atkins, J.F., Firth, A.E., Taubenberger, J.K. & Digard, P. (2012). An Overlapping Protein-Coding Region in Influenza A Virus Segment 3 Modulates the Host Response. *Science (New York, N.Y.)*. 337 (6091). pp. 199–204.
- Ji, G., Wu, X., Shen, Y., Huang, J. & Quinn Li, Q. (2010). A classification-based prediction model of messenger RNA polyadenylation sites. *Journal of theoretical biology*. 265 (3). pp. 287–296.
- Jia, J., Arif, A., Ray, P.S. & Fox, P.L. (2008). WHEP domains direct noncanonical function of glutamyl-prolyl tRNA synthetase in translational control of gene expression. *Molecular cell*. 29 (6). pp. 679–690.
- Jia, X., Yuan, S., Wang, Y., Fu, Y., Ge, Y., Ge, Y., Lan, X., Feng, Y., Qiu, F., Li, P., Chen, S. & Xu, A. (2017). The role of alternative polyadenylation in the antiviral innate immune response. *Nature Communications*. 8. p. 14605.
- Jorba, N., Area, E. & Ortín, J. (2008). Oligomerization of the influenza virus polymerase complex in vivo. *The Journal of general virology*. 89 (Pt 2). pp. 520–524.
- Jorba, N., Coloma, R. & Ortín, J. (2009). Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. S. P. J. Whelan (ed.). *PLoS pathogens*. 5 (5). p. e1000462.
- Katahira, J. (2012). mRNA export and the TREX complex. *Biochimica Et Biophysica Acta- Gene Regulatory Mechanisms*. 1819 (6). pp. 507–513.

- Kawaoka, Y., Krauss, S. & Webster, R.G. (1989). Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. *Journal of Virology*. 63 (11). pp. 4603–4608.
- Kelly, S.M. & Corbett, A.H. (2009). Messenger RNA Export from the Nucleus: A Series of Molecular Wardrobe Changes. *Traffic*. 10 (9). pp. 1199–1208.
- Kerry, P.S., Willsher, N. & Fodor, E. (2008). A cluster of conserved basic amino acids near the C-terminus of the PB1 subunit of the influenza virus RNA polymerase is involved in the regulation of viral transcription. *Virology*. 373 (1). pp. 202–210.
- Khapersky, D.A., Emara, M.M., Johnston, B.P., Anderson, P., Hatchette, T.F. & McCormick, C. (2014). Influenza A Virus Host Shutoff Disables Antiviral Stress-Induced Translation Arrest A. Pekosz (ed.). *PLoS pathogens*. 10 (7).
- Khapersky, D.A., Schmalig, S., Larkins-Ford, J., McCormick, C. & Gaglia, M.M. (2016). Selective Degradation of Host RNA Polymerase II Transcripts by Influenza A Virus PA-X Host Shutoff Protein. P. Palese (ed.). *PLoS pathogens*. 12 (2). p. e1005427.
- Kim, H.J., Fodor, E., Brownlee, G.G. & Seong, B.L. (1997). Mutational analysis of the RNA-fork model of the influenza A virus vRNA promoter in vivo. *The Journal of general virology*. 78 (Pt 2). pp. 353–357.
- Kim, J.H., Hatta, M., Watanabe, S., Neumann, G., Watanabe, T. & Kawaoka, Y. (2010). Role of host-specific amino acids in the pathogenicity of avian H5N1 influenza viruses in mice. *The Journal of general virology*. 91 (Pt 5). pp. 1284–1289.
- Kirui, J., Bucci, M.D., Poole, D.S. & Mehle, A. (2014). Conserved features of the PB2 627 domain impact influenza virus polymerase function and replication. *Journal of Virology*. 88 (11). pp. 5977–5986.
- Koehler, A. & Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nature Reviews Molecular Cell Biology*. 8 (10). pp. 761–773.
- Koppstein, D., Ashour, J. & Bartel, D.P. (2015). Sequencing the cap-snatching repertoire of H1N1 influenza provides insight into the mechanism of viral transcription initiation. *Nucleic acids research*. 43 (10). pp. 5052–5064.
- Krug, R.M. (2015). Functions of the influenza A virus NS1 protein in antiviral defense. *Current Opinion in Virology*. 12. pp. 1–6.
- Kumar, M. & Carmichael, G.G. (1997). Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. *Proceedings of the National Academy of Sciences of the United States of America*. 94 (8). pp. 3542–3547.
- Kummer, S., Flöttmann, M., Schwanhäusser, B., Sieben, C., Veit, M., Selbach, M., Klipp, E. & Herrmann, A. (2014). Alteration of protein levels during influenza virus H1N1 infection in host cells: a proteomic survey of host and virus reveals differential dynamics. P. Digard (ed.). *Plos One*. 9 (4). p. e94257.
- Kuss, S.K., Mata, M.A., Zhang, L. & Fontoura, B.M.A. (2013). Nuclear imprisonment: viral strategies to arrest host mRNA nuclear export. *Viruses*. 5 (7). pp. 1824–1849.

- Kuzuhara, T., Kise, D., Yoshida, H., Horita, T., Murazaki, Y., Nishimura, A., Echigo, N., Utsunomiya, H. & Tsuge, H. (2009). Structural basis of the influenza A virus RNA polymerase PB2 RNA-binding domain containing the pathogenicity-determinant lysine 627 residue. *The Journal of biological chemistry*. 284 (11). pp. 6855–6860.
- Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, J., Hentze, M.W. & Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. *Nature Structural & Molecular Biology*. 20 (9). pp. 1122–1130.
- la Luna, de, S., Fortes, P., Beloso, A. & Ortín, J. (1995). Influenza virus NS1 protein enhances the rate of translation initiation of viral mRNAs. *Journal of Virology*. 69 (4). pp. 2427–2433.
- Labadie, K., Santos Afonso, Dos, E., Rameix-Welti, M.-A., van der Werf, S. & Naffakh, N. (2007). Host-range determinants on the PB2 protein of influenza A viruses control the interaction between the viral polymerase and nucleoprotein in human cells. *Virology*. 362 (2). pp. 271–282.
- Lakdawala, S.S., Wu, Y., Wawrzusin, P., Kabat, J., Broadbent, A.J., Lamirande, E.W., Fodor, E., Altan-Bonnet, N., Shroff, H. & Subbarao, K. (2014). Influenza A virus assembly intermediates fuse in the cytoplasm. M. T. Heise (ed.). *PLoS pathogens*. 10 (3). p. e1003971.
- Lamb, R.A. & Choppin, P.W. (1981). Identification of a 2nd Protein (M2) Encoded by Rna Segment-7 of Influenza-Virus. *Virology*. 112 (2). pp. 729–737.
- Lamb, R.A. & Horvath, C.M. (1991). Diversity of Coding Strategies in Influenza-Viruses. *Trends in genetics : TIG*. 7 (8). pp. 261–266.
- Lamb, R.A. & Lai, C.J. (1980). Sequence of Interrupted and Uninterrupted Messenger-Rnas and Cloned Dna Coding for the 2 Overlapping Nonstructural Proteins of Influenza-Virus. *Cell*. 21 (2). pp. 475–485.
- Lamb, R.A., Choppin, P.W., Chanock, R.M. & Lai, C.J. (1980). Mapping of the 2 Overlapping Genes for Polypeptides Ns1 and Ns2 on Rna Segment-8 of Influenza-Virus Genome. *Proceedings of the National Academy of Sciences of the United States of America*. 77 (4). pp. 1857–1861.
- Lamb, R.A., Lai, C.J. & Choppin, P.W. (1981). Sequences of Messenger-Rnas Derived From Genome Rna Segment 7 of Influenza-Virus - Colinear and Interrupted Messenger-Rnas Code for Overlapping Proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 78 (7). pp. 4170–4174.
- Leahy, M.B., Dobbyn, H.C. & Brownlee, G.G. (2001a). Hairpin loop structure in the 3 ' arm of the influenza A virus virion RNA promoter is required for endonuclease activity. *Journal of Virology*. 75 (15). pp. 7042–7049.
- Leahy, M.B., Pritlove, D.C., Poon, L. & Brownlee, G.G. (2001b). Mutagenic analysis of the 5 ' arm of the influenza A virus virion RNA promoter defines the sequence requirements for endonuclease activity. *Journal of Virology*. 75 (1). pp. 134–142.
- Lee, I. & Hong, W.J. (2004). RAP - a putative RNA-binding domain. *Trends in Biochemical Sciences*. 29 (11). pp. 567–570.

- Lee, K.H. & Seong, B.L. (1998). The position 4 nucleotide at the 3' end of the influenza virus neuraminidase vRNA is involved in temporal regulation of transcription and replication of neuraminidase RNAs and affects the repertoire of influenza virus surface antigens. *The Journal of general virology*. 79 (Pt 8) (8). pp. 1923–1934.
- Lee, M., Bishop, K., Medcalf, L., Elton, D., Digard, P. & Tiley, L. (2002). Definition of the minimal viral components required for the initiation of unprimed RNA synthesis by influenza virus RNA polymerase. *Nucleic acids research*. 30 (2). pp. 429–438.
- Lee, M.-K., Bae, S.-H., Park, C.-J., Cheong, H.-K., Cheong, C. & Choi, B.-S. (2003). A single-nucleotide natural variation (U4 to C4) in an influenza A virus promoter exhibits a large structural change: implications for differential viral RNA synthesis by RNA-dependent RNA polymerase. *Nucleic acids research*. 31 (4). pp. 1216–1223.
- Lee, N., Le Sage, V., Nanni, A.V., Snyder, D.J., Cooper, V.S. & Lakdawala, S.S. (2017). Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic acids research*. 45 (15). pp. 8968–8977.
- Leser, G.P. & Lamb, R.A. (2005). Influenza virus assembly and budding in raft-derived microdomains: A quantitative analysis of the surface distribution of HA, NA and M2 proteins. *Virology*. 342 (2). pp. 215–227.
- Li, M.L., Ramirez, B.C. & Krug, R.M. (1998). RNA-dependent activation of primer RNA production by influenza virus polymerase: different regions of the same protein subunit constitute the two required RNA-binding sites. *Embo Journal*. 17 (19). pp. 5844–5852.
- Li, S., Min, J.-Y., Krug, R.M. & Sen, G.C. (2006). Binding of the influenza A virus NS1 protein to PKR mediates the inhibition of its activation by either PACT or double-stranded RNA. *Virology*. 349 (1). pp. 13–21.
- Li, Z., Chen, H., Jiao, P., Deng, G., Tian, G., Li, Y., Hoffmann, E., Webster, R.G., Matsuoka, Y. & Yu, K. (2005). Molecular basis of replication of duck H5N1 influenza viruses in a mammalian mouse model. *Journal of Virology*. 79 (18). pp. 12058–12064.
- Liao, Y., Castello, A., Fischer, B., Leicht, S., Foehr, S., Frese, C.K., Ragan, C., Kurscheid, S., Pagler, E., Yang, H., Krijgsveld, J., Hentze, M.W. & Preiss, T. (2016). The Cardiomyocyte RNA-Binding Proteome: Links to Intermediary Metabolism and Heart Disease. *Cell Reports*. 16 (5). pp. 1456–1469.
- Liepelt, A., Naarmann-de Vries, I.S., Simons, N., Eichelbaum, K., Föhr, S., Archer, S.K., Castello, A., Usadel, B., Krijgsveld, J., Preiss, T., Marx, G., Hentze, M.W., Ostareck, D.H. & Ostareck-Lederer, A. (2016). Identification of RNA-binding Proteins in Macrophages by Interactome Capture. *Molecular & cellular proteomics : MCP*. 15 (8). pp. 2699–2714.
- Lim, J., Hao, T., Shaw, C., Patel, A.J., Szabó, G., Rual, J.-F., Fisk, C.J., Li, N., Smolyar, A., Hill, D.E., Barabási, A.-L., Vidal, M. & Zoghbi, H.Y. (2006). A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell*. 125 (4). pp. 801–814.

- Liu, W., Yang, K., Qi, X., Xu, K., Ji, H., Ai, J., Ge, A., Wu, Y., Li, Y., Dai, Q., Liang, Q., Bao, C., Bergquist, R., Tang, F. & Zhu, Y. (2013). Spatial and temporal analysis of human infection with avian influenza A(H7N9) virus in China, 2013. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*. 18 (47).
- Lobley, A., Swindells, M.B., Orengo, C.A. & Jones, D.T. (2007). Inferring function using patterns of native disorder in proteins. *Plos Computational Biology*. 3 (8). pp. 1567–1579.
- Loke, P., Hammond, S.N., Leung, J.M., Kim, C.C., Batra, S., Rocha, C., Balmaseda, A. & Harris, E. (2010). Gene expression patterns of dengue virus-infected children from nicaragua reveal a distinct signature of increased metabolism. A. M. Powers (ed.). *PLoS neglected tropical diseases*. 4 (6). p. e710.
- Long, J.S., Giotis, E.S., Moncorge, O., Frise, R., Mistry, B., James, J., Morisson, M., Iqbal, M., Vignal, A., Skinner, M.A. & Barclay, W.S. (2016). Species difference in ANP32A underlies influenza A virus polymerase host restriction. *Nature*. 529 (7584). pp. 101–104.
- Long, J.S., Howard, W.A., Núñez, A., Moncorge, O., Lycett, S., Banks, J. & Barclay, W.S. (2013). The effect of the PB2 mutation 627K on highly pathogenic H5N1 avian influenza virus is dependent on the virus lineage. *Journal of Virology*. 87 (18). pp. 9983–9996.
- Lukarska, M., Fournier, G., Pflug, A., Resa-Infante, P., Reich, S., Naffakh, N. & Cusack, S. (2017). Structural basis of an essential interaction between influenza polymerase and Pol II CTD. *Nature*. 541 (7635). pp. 117–121.
- Lukong, K.E., Chang, K.-W., Khandjian, E.W. & Richard, S. (2008). RNA-binding proteins in human genetic disease. *Trends in genetics : TIG*. 24 (8). pp. 416–425.
- Lunde, B.M., Moore, C. & Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nature Reviews Molecular Cell Biology*. 8 (6). pp. 479–490.
- Manzoor, R., Sakoda, Y., Nomura, N., Tsuda, Y., Ozaki, H., Okamatsu, M. & Kida, H. (2009). PB2 protein of a highly pathogenic avian influenza virus strain A/chicken/Yamaguchi/7/2004 (H5N1) determines its replication potential in pigs. *Journal of Virology*. 83 (4). pp. 1572–1578.
- Marión, R.M., Aragón, T., Beloso, A., Nieto, A. & Ortín, J. (1997). The N-terminal half of the influenza virus NS1 protein is sufficient for nuclear retention of mRNA and enhancement of viral mRNA translation. *Nucleic acids research*. 25 (21). pp. 4271–4277.
- Marondedze, C., Thomas, L., Serrano, N.L., Lilley, K.S. & Gehring, C. (2016). The RNA-binding protein repertoire of Arabidopsis thaliana. *Scientific reports*. 6 (1).
- Martin, K. & Helenius, A. (1991a). Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. *Cell*. 67 (1). pp. 117–130.
- Martin, K. & Helenius, A. (1991b). Transport of incoming influenza virus nucleocapsids into the nucleus. *Journal of Virology*. 65 (1). pp. 232–244.

- Martínez, I., Dopazo, J. & Melero, J.A. (1997). Antigenic structure of the human respiratory syncytial virus G glycoprotein and relevance of hypermutation events for the generation of antigenic variants. *The Journal of general virology*. 78 (Pt 10) (10). pp. 2419–2429.
- Martínez-Alonso, M., Hengrung, N. & Fodor, E. (2016). RNA-Free and Ribonucleoprotein-Associated Influenza Virus Polymerases Directly Bind the Serine-5-Phosphorylated Carboxyl-Terminal Domain of Host RNA Polymerase II. *Journal of Virology*. 90 (13). pp. 6014–6021.
- Massin, P., van der Werf, S. & Naffakh, N. (2001). Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. *Journal of Virology*. 75 (11). pp. 5398–5404.
- Matera, A.G. & Wang, Z. (2014). A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology*. 15 (2). pp. 108–121.
- Matia-Gonzalez, A.M., Laing, E.E. & Gerber, A.P. (2015). Conserved RNA-binding proteomes in eukaryotic organisms. *Nature Structural & Molecular Biology*. 22 (12). pp. 1027–1033.
- Matilla, A. & Radrizzani, M. (2005). The Anp32 family of proteins containing leucine-rich repeats. *Cerebellum (London, England)*. 4 (1). pp. 7–18.
- Matsuoka, Y., Matsumae, H., Katoh, M., Eisfeld, A.J., Neumann, G., Hase, T., Ghosh, S., Shoemaker, J.E., Lopes, T.J.S., Watanabe, T., Watanabe, S., Fukuyama, S., Kitano, H. & Kawaoka, Y. (2013). A comprehensive map of the influenza A virus replication cycle. *BMC systems biology*. 7 (1). p. 97.
- Mayr, C. & Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*. 138 (4). pp. 673–684.
- Mänz, B., Brunotte, L., Reuther, P. & Schwemmle, M. (2012). Adaptive mutations in NEP compensate for defective H5N1 RNA replication in cultured human cells. *Nature Communications*. 3. p. 802.
- McCown, M.F. & Pekosz, A. (2006). Distinct domains of the influenza A virus M2 protein cytoplasmic tail mediate binding to the M1 protein and facilitate infectious virus production. *Journal of Virology*. 80 (16). pp. 8178–8189.
- Mehle, A. & Doudna, J.A. (2009). Adaptive strategies of the influenza virus polymerase for replication in humans. *Proceedings of the National Academy of Sciences of the United States of America*. 106 (50). pp. 21312–21316.
- Mehle, A. & Doudna, J.A. (2008). An inhibitory activity in human cells restricts the function of an avian-like influenza virus polymerase. *Cell host & microbe*. 4 (2). pp. 111–122.
- Mehle, A., Dugan, V.G., Taubenberger, J.K. & Doudna, J.A. (2012). Reassortment and mutation of the avian influenza virus polymerase PA subunit overcome species barriers. *Journal of Virology*. 86 (3). pp. 1750–1757.

- Melen, K., Fagerlund, R., Franke, J., Kohler, M., Kinnunen, L. & Julkunen, I. (2003). Importin alpha nuclear localization signal binding sites for STAT1, STAT2, and influenza A virus nucleoprotein. *The Journal of biological chemistry*. 278 (30). pp. 28193–28200.
- Mercer, J., Schelhaas, M. & Helenius, A. (2010). Virus entry by endocytosis. *Annual review of biochemistry*. 79 (1). pp. 803–833.
- Miotto, O., Heiny, A., Tan, T.W., August, J.T. & Brusic, V. (2008). Identification of human-to-human transmissibility factors in PB2 proteins of influenza A by large-scale mutual information analysis. *BMC bioinformatics*. 9 Suppl 1 (Suppl 1). p. S18.
- Miotto, O., Heiny, A.T., Albrecht, R., García-Sastre, A., Tan, T.W., August, J.T. & Brusic, V. (2010). Complete-proteome mapping of human influenza A adaptive mutations: implications for human transmissibility of zoonotic strains. A. F. Y. Poon (ed.). *Plos One*. 5 (2). p. e9025.
- Mitchell, S.F., Jain, S., She, M. & Parker, R. (2013). Global analysis of yeast mRNPs. *Nature Structural & Molecular Biology*. 20 (1). pp. 127–U161.
- Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J.O. & Lai, E.C. (2013). Widespread and extensive lengthening of 3' UTRs in the mammalian brain. *Genome research*. 23 (5). pp. 812–825.
- Moeller, A., Kirchdoerfer, R.N., Potter, C.S., Carragher, B. & Wilson, I.A. (2012). Organization of the influenza virus replication machinery. *Science (New York, N.Y.)*. 338 (6114). pp. 1631–1634.
- Mok, C.K.P., Yen, H.-L., Yu, M.Y.M., Yuen, K.M., Sia, S.F., Chan, M.C.W., Qin, G., Tu, W.W. & Peiris, J.S.M. (2011). Amino acid residues 253 and 591 of the PB2 protein of avian influenza virus A H9N2 contribute to mammalian pathogenesis. *Journal of Virology*. 85 (18). pp. 9641–9645.
- Momose, F., Sekimoto, T., Ohkura, T., Jo, S., Kawaguchi, A., Nagata, K. & Morikawa, Y. (2011). Apical Transport of Influenza A Virus Ribonucleoprotein Requires Rab11-positive Recycling Endosome J.-L. E. P. H. Darlix (ed.). *Plos One*. 6 (6).
- Moncorge, O., Mura, M. & Barclay, W.S. (2010). Evidence for avian and human host cell factors that affect the activity of influenza virus polymerase. *Journal of Virology*. 84 (19). pp. 9978–9986.
- Muramoto, Y., Noda, T., Kawakami, E., Akkina, R. & Kawaoka, Y. (2013). Identification of novel influenza A virus proteins translated from PA mRNA. *Journal of Virology*. 87 (5). pp. 2455–2462.
- Murphy, D.G., Dimock, K. & Kang, C.Y. (1991). Numerous transitions in human parainfluenza virus 3 RNA recovered from persistently infected cells. *Virology*. 181 (2). pp. 760–763.
- Naffakh, N., Massin, P., Escriou, N., Crescenzo-Chaigne, B. & van der Werf, S. (2000). Genetic analysis of the compatibility between polymerase proteins from human and avian strains of influenza A viruses. *The Journal of general virology*. 81 (Pt 5). pp. 1283–1291.

- Naffakh, N., Tomoiu, A., Rameix-Welti, M.-A. & van der Werf, S. (2008). Host restriction of avian influenza viruses at the level of the ribonucleoproteins. *Annual review of microbiology*. 62 (1). pp. 403–424.
- Nakazono, Y., Hara, K., Kashiwagi, T., Hamada, N. & Watanabe, H. (2012). The RNA polymerase PB2 subunit of influenza A/HongKong/156/1997 (H5N1) restricts the replication of reassortant ribonucleoprotein complexes [corrected]. E. G. Brown (ed.). *Plos One*. 7 (2). p. e32634.
- Nam, J.-H., Shim, S.-M., Song, E.-J., España, E., Jeong, D.-G., Song, D. & Kim, J.-K. (2017). Rapid virulence shift of an H5N2 avian influenza virus during a single passage in mice. *Archives of virology*. 162 (10). pp. 3017–3024.
- Nandan, D., Thomas, S.A., Nguyen, A., Moon, K.-M., Foster, L.J. & Reiner, N.E. (2017). Comprehensive Identification of mRNA Binding Proteins of *Leishmania donovani* by Interactome Capture T. Preiss (ed.). *Plos One*. 12 (1).
- Nemeroff, M.E., Barabino, S., Li, Y.Z., Keller, W. & Krug, R.M. (1998). Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3' end formation of cellular pre-mRNAs. *Molecular cell*. 1 (7). pp. 991–1000.
- Neumann, G. & Hobom, G. (1995). Mutational analysis of influenza virus promoter elements in vivo. *The Journal of general virology*. 76 (Pt 7) (7). pp. 1709–1717.
- Neumann, G., Macken, C.A. & Kawaoka, Y. (2014). Identification of amino acid changes that may have been critical for the genesis of A(H7N9) influenza viruses. *Journal of Virology*. 88 (9). pp. 4877–4896.
- Neumann, G. & Kawaoka, Y. (2015). Transmission of influenza A viruses. *Virology*. 479–480. pp. 234–246.
- Neumann, G., Hughes, M.T. & Kawaoka, Y. (2000). Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *Embo Journal*. 19 (24). pp. 6751–6758.
- Neumann, G., Noda, T. & Kawaoka, Y. (2009). Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature*. 459 (7249). pp. 931–939.
- Neve, J., Burger, K., Li, W., Hoque, M., Patel, R., Tian, B., Gullerova, M. & Furger, A. (2016). Subcellular RNA profiling links splicing and nuclear DICER1 to alternative cleavage and polyadenylation. *Genome research*. 26 (1). pp. 24–35.
- Neve, J., Patel, R., Wang, Z., Louey, A. & Furger, A.M. (2017). Cleavage and polyadenylation: Ending the message expands gene regulation. *RNA biology*. 14 (7). pp. 865–890.
- Ng, A.K.-L., Chan, W.-H., Choi, S.-T., Lam, M.K.-H., Lau, K.-F., Chan, P.K.-S., Au, S.W.-N., Fodor, E. & Shaw, P.-C. (2012a). Influenza polymerase activity correlates with the strength of interaction between nucleoprotein and PB2 through the host-specific residue K/E627. E. G. Brown (ed.). *Plos One*. 7 (5). p. e36415.

- Ng, A.K.-L., Chan, W.-H., Choi, S.-T., Lam, M.K.-H., Lau, K.-F., Chan, P.K.-S., Au, S.W.-N., Fodor, E. & Shaw, P.-C. (2012b). Influenza polymerase activity correlates with the strength of interaction between nucleoprotein and PB2 through the host-specific residue K/E627. *Plos One*. 7 (5). p. e36415.
- Ng, A.K.-L., Zhang, H., Tan, K., Li, Z., Liu, J.-H., Chan, P.K.-S., Li, S.-M., Chan, W.-Y., Au, S.W.-N., Joachimiak, A., Walz, T., Wang, J.-H. & Shaw, P.-C. (2008). Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *Faseb Journal*. 22 (10). pp. 3638–3647.
- Ngamurulert, S., Limjindaporn, T. & Auewaraku, P. (2009). Identification of cellular partners of Influenza A virus (H5N1) non-structural protein NS1 by yeast two-hybrid system. *Acta Virologica*. 53 (3). pp. 153–159.
- Niessing, D., Huttelmaier, S., Zenklusen, D., Singer, R.H. & Burley, S.K. (2004). She2p is a novel RNA binding protein with a basic helical hairpin motif. *Cell*. 119 (4). pp. 491–502.
- Nilsson, B.E., Velthuis, te, A.J.W. & Fodor, E. (2017). The role of the PB2 627-domain in influenza A virus polymerase function. *Journal of Virology*. pp. JVI.02467–16.
- Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R.H. & Kawaoka, Y. (2006). Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature*. 439 (7075). pp. 490–492.
- Noton, S.L., Medcalf, E., Fisher, D., Mullin, A.E., Elton, D. & Digard, P. (2007). Identification of the domains of the influenza A virus M1 matrix protein required for NP binding, oligomerization and incorporation into virions. *The Journal of general virology*. 88 (Pt 8). pp. 2280–2290.
- O'Neill, R.E. & Palese, P. (1995). NPI-1, the human homolog of SRP-1, interacts with influenza virus nucleoprotein. *Virology*. 206 (1). pp. 116–125.
- O'Neill, R.E., Jaskunas, R., Blobel, G., Palese, P. & Moroianu, J. (1995). Nuclear import of influenza virus RNA can be mediated by viral nucleoprotein and transport factors required for protein import. *The Journal of biological chemistry*. 270 (39). pp. 22701–22704.
- O'Neill, R.E., Talon, J. & Palese, P. (1998). The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *Embo Journal*. 17 (1). pp. 288–296.
- Obayashi, E., Yoshida, H., Kawai, F., Shibayama, N., Kawaguchi, A., Nagata, K., Tame, J.R.H. & Park, S.-Y. (2008). The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature*. 454 (7208). pp. 1127–U57.
- Oberstrass, F.C., Auweter, S.D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L. & Allain, F. (2005). Structure of PTB bound to RNA: Specific binding and implications for splicing regulation. *Science (New York, N.Y.)*. 309 (5743). pp. 2054–2057.
- Oishi, K., Yamayoshi, S. & Kawaoka, Y. (2015). Mapping of a Region of the PA-X Protein of Influenza A Virus That Is Important for Its Shutoff Activity T. S. Dermody (ed.). *Journal of Virology*. 89 (16). pp. 8661–8665.

- Ortega, J., Martin-Benito, J., Zurcher, T., Valpuesta, J.M., Carrascosa, J.L. & Ortín, J. (2000). Ultrastructural and functional analyses of recombinant influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification. *Journal of Virology*. 74 (1). pp. 156–163.
- Palese, P. & Shaw, M.L. (2013). Orthomyxoviridae. In: D. M. Knipe & P. M. Howley (eds.). *Fields Virology*. Philadelphia, pp. 1151–1185.
- Pan, W., da Graca, L.S., Shao, Y., Yin, Q., Wu, H. & Jiang, X. (2009). PHAPI/pp32 suppresses tumorigenesis by stimulating apoptosis. *The Journal of biological chemistry*. 284 (11). pp. 6946–6954.
- Paterson, D. & Fodor, E. (2012). Emerging roles for the influenza A virus nuclear export protein (NEP). *PLoS pathogens*. 8 (12). p. e1003019.
- Paterson, D., Velthuis, te, A.J.W., Vreede, F.T. & Fodor, E. (2014). Host restriction of influenza virus polymerase activity by PB2 627E is diminished on short viral templates in a nucleoprotein-independent manner. *Journal of Virology*. 88 (1). pp. 339–344.
- Perez, D.R. & Donis, R.O. (1998). The matrix 1 protein of influenza A virus inhibits the transcriptase activity of a model influenza reporter genome in vivo. *Virology*. 249 (1). pp. 52–61.
- Perez, J.T., Varble, A., Sachidanandam, R., Zlatev, I., Manoharan, M., García-Sastre, A. & tenOever, B.R. (2010). Influenza A virus-generated small RNAs regulate the switch from transcription to replication. *Proceedings of the National Academy of Sciences of the United States of America*. 107 (25). pp. 11525–11530.
- Pflug, A., Guilligay, D., Reich, S. & Cusack, S. (2014). Structure of influenza A polymerase bound to the viral RNA promoter. *Nature*. 516 (7531). pp. 355–360.
- Phan, A.T., Kuryavyi, V., Darnell, J.C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D., Darnell, R.B. & Patel, D.J. (2011). Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nature Structural & Molecular Biology*. 18 (7). pp. 796–U73.
- Pinto, L.H., Holsinger, L.J. & Lamb, R.A. (1992). Influenza virus M2 protein has ion channel activity. *Cell*. 69 (3). pp. 517–528.
- Pleschka, S., Wolff, T., Ehrhardt, C., Hobom, G., Planz, O., Rapp, U.R. & Ludwig, S. (2001). Influenza virus propagation is impaired by inhibition of the Raf/MEK/ERK signalling cascade. *Nature Cell Biology*. 3 (3). pp. 301–305.
- Plotch, S.J., Bouloy, M., Ulmanen, I. & Krug, R.M. (1981). A Unique Cap(M7gpppxm)-Dependent Influenza Virion Endonuclease Cleaves Capped Rnas to Generate the Primers That Initiate Viral-Rna Transcription. *Cell*. 23 (3). pp. 847–858.
- Pons, M.W., Schulze, I.T., Hirst, G.K. & Hauser, R. (1969). Isolation and Characterization of Ribonucleoprotein of Influenza Virus. *Virology*. 39 (2). pp. 250–&.

- Poole, E., Elton, D., Medcalf, L. & Digard, P. (2004). Functional domains of the influenza A virus PB2 protein: identification of NP- and PB1-binding sites. *Virology*. 321 (1). pp. 120–133.
- Potter, C.W. (2001). A history of influenza. *Journal of applied microbiology*. 91 (4). pp. 572–579.
- Pritlove, D.C., Poon, L., Devenish, L.J., Leahy, M.B. & Brownlee, G.G. (1999). A hairpin loop at the 5' end of influenza a virus virion RNA is required for synthesis of poly(A) (+) mRNA *in vitro*. *Journal of Virology*. 73 (3). pp. 2109–2114.
- Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. *Genes & Development*. 25 (17). pp. 1770–1782.
- Proudfoot, N.J. & Brownlee, G.G. (1976). 3' Non-Coding Region Sequences in Eukaryotic Messenger-Rna. *Nature*. 263 (5574). pp. 211–214.
- Qi, X., Pan, Y., Qin, Y., Zu, R., Tang, F., Zhou, M., Wang, H. & Song, Y. (2012). Molecular characterization of avian-like H1N1 swine influenza a viruses isolated in Eastern China, 2011. *Virologica Sinica*. 27 (5). pp. 292–298.
- Qiu, Y. & Krug, R.M. (1994). The influenza virus NS1 protein is a poly(A)-binding protein that inhibits nuclear export of mRNAs containing poly(A). *Journal of Virology*. 68 (4). pp. 2425–2432.
- Rameix-Welti, M.-A., Tomoiu, A., Santos Afonso, Dos, E., van der Werf, S. & Naffakh, N. (2009). Avian Influenza A virus polymerase association with nucleoprotein, but not polymerase assembly, is impaired in human cells during the course of infection. *Journal of Virology*. 83 (3). pp. 1320–1331.
- Rammelt, C., Bilen, B., Zavolan, M. & Keller, W. (2011). PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. *Rna-a Publication of the Rna Society*. 17 (9). pp. 1737–1746.
- Read, E.K.C. & Digard, P. (2010). Individual influenza A virus mRNAs show differential dependence on cellular NXF1/TAP for their nuclear export. *The Journal of general virology*. 91 (Pt 5). pp. 1290–1301.
- Reed, R. (2003). Coupling transcription, splicing and mRNA export. *Current Opinion in Cell Biology*. 15 (3). pp. 326–331.
- Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crepin, T., Hart, D., Lunardi, T., Nanao, M., Ruigrok, R.W.H. & Cusack, S. (2014). Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature*. 516 (7531). pp. 361–366.
- Reichel, M., Liao, Y., Rettel, M., Ragan, C., Evers, M., Alleaume, A.-M., Horos, R., Hentze, M.W., Preiss, T. & Millar, A.A. (2016). In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedlings. *Plant Cell*. 28 (10). pp. 2435–2452.
- Reilly, P.T., Yu, Y., Hamiche, A. & Wang, L. (2014). Cracking the ANP32 whips: important functions, unequal requirement, and hints at disease implications. *BioEssays : news and reviews in molecular, cellular and developmental biology*. 36 (11). pp. 1062–1071.

- Resa-Infante, P., Angeles Recuero-Checa, M., Zamarreño, N., Llorca, O. & Ortín, J. (2010). Structural and Functional Characterization of an Influenza Virus RNA Polymerase-Genomic RNA Complex. *Journal of Virology*. 84 (20). pp. 10477–10487.
- Resa-Infante, P., Jorba, N., Zamarreño, N., Fernández, Y., Juárez, S. & Ortín, J. (2008). The host-dependent interaction of alpha-importins with influenza PB2 polymerase subunit is required for virus RNA replication. W.-S. Ryu (ed.). *Plos One*. 3 (12). p. e3904.
- Resa-Infante, P., Thieme, R., Ernst, T., Arck, P.C., Ittrich, H., Reimer, R. & Gabriel, G. (2014). Importin- α 7 is required for enhanced influenza A virus replication in the alveolar epithelium and severe lung damage in mice. A. García-Sastre (ed.). *Journal of Virology*. 88 (14). pp. 8166–8179.
- Richardson, J.C. & Akkina, R.K. (1991). NS2 protein of influenza virus is found in purified virus and phosphorylated in infected cells. *Archives of virology*. 116 (1-4). pp. 69–80.
- Robb, N.C. & Fodor, E. (2012). The accumulation of influenza A virus segment 7 spliced mRNAs is regulated by the NS1 protein. *The Journal of general virology*. 93 (Pt 1). pp. 113–118.
- Robb, N.C., Cordes, T., Hwang, L.C., Gryte, K., Duchi, D., Craggs, T.D., Santoso, Y., Weiss, S., Ebright, R.H. & Kapanidis, A.N. (2013). The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. *Journal of molecular biology*. 425 (5). pp. 875–885.
- Robb, N.C., Jackson, D., Vreede, F.T. & Fodor, E. (2010). Splicing of influenza A virus NS1 mRNA is independent of the viral NS1 protein. *The Journal of general virology*. 91 (Pt 9). pp. 2331–2340.
- Robb, N.C., Smith, M., Vreede, F.T. & Fodor, E. (2009). NS2/NEP protein regulates transcription and replication of the influenza virus RNA genome. *The Journal of general virology*. 90 (Pt 6). pp. 1398–1407.
- Robb, N.C., Velthuis, te, A.J.W., Wieneke, R., Tampé, R., Cordes, T., Fodor, E. & Kapanidis, A.N. (2016). Single-molecule FRET reveals the pre-initiation and initiation conformations of influenza virus promoter RNA. *Nucleic acids research*.
- Robertson, J.S. (1979). 5' and 3' Terminal Nucleotide-Sequences of the Rna Genome Segments of Influenza-Virus. *Nucleic acids research*. 6 (12). pp. 3745–3757.
- Robertson, J.S., Schubert, M. & Lazzarini, R.A. (1981). Polyadenylation Sites for Influenza-Virus Messenger-Rna. *Journal of Virology*. 38 (1). pp. 157–163.
- Rodriguez Boulan, E. & Sabatini, D.D. (1978). Asymmetric budding of viruses in epithelial monolayers: a model system for study of epithelial polarity. *Proceedings of the National Academy of Sciences of the United States of America*. 75 (10). pp. 5071–5075.
- Rodriguez, A., Perez-Gonzalez, A. & Nieto, A. (2007). Influenza virus infection causes specific degradation of the largest subunit of cellular RNA polymerase II. *Journal of Virology*. 81 (10). pp. 5315–5324.

- Rossman, J.S. & Lamb, R.A. (2011). Influenza virus assembly and budding. *Virology*. 411 (2). pp. 229–236.
- Rossman, J.S., Jing, X., Leser, G.P., Balannik, V., Pinto, L.H. & Lamb, R.A. (2010). Influenza Virus M2 Ion Channel Protein Is Necessary for Filamentous Virion Formation. *Journal of Virology*. 84 (10). pp. 5078–5088.
- Roux, P.P. & Topisirovic, I. (2012). Regulation of mRNA Translation by Signaling Pathways. *Cold Spring Harbor Perspectives in Biology*. 4 (11). pp. –a012252.
- Ruigrok, R.W.H. & Baudin, F. (1995). Structure of Influenza-Virus Ribonucleoprotein-Particles .2. Purified Rna-Free Influenza-Virus Ribonucleoprotein Forms Structures That Are Indistinguishable From the Intact Influenza-Virus Ribonucleoprotein-Particles. *The Journal of general virology*. 76 (4). pp. 1009–1014.
- Ruigrok, R.W.H., Crepin, T., Hart, D.J. & Cusack, S. (2010). Towards an atomic resolution understanding of the influenza virus replication machinery. *Current Opinion in Structural Biology*. 20 (1). pp. 104–113.
- Samuel, C.E. (2001). Antiviral actions of interferons. *Clinical microbiology reviews*. 14 (4). pp. 778–809– table of contents.
- Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A. & Burge, C.B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science (New York, N.Y.)*. 320 (5883). pp. 1643–1647.
- Satterly, N., Tsai, P.-L., van Deursen, J., Nussenzveig, D.R., Wang, Y., Faria, P.A., Levay, A., Levy, D.E. & Fontoura, B.M.A. (2007). Influenza virus targets the mRNA export machinery and the nuclear pore complex. *Proceedings of the National Academy of Sciences of the United States of America*. 104 (6). pp. 1853–1858.
- Scheiffele, P., Rietveld, A., Wilk, T. & Simons, K. (1999). Influenza viruses select ordered lipid domains during budding from the plasma membrane. *The Journal of biological chemistry*. 274 (4). pp. 2038–2044.
- Scherrer, T., Mittal, N., Janga, S.C. & Gerber, A.P. (2010). A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. J. Bähler (ed.). *Plos One*. 5 (11). p. e15499.
- Scholtissek, C., Rohde, W., Hoyningen, von, V. & Rott, R. (1978). Origin of Human Influenza-Virus Subtypes H2n2 and H3n2. *Virology*. 87 (1). pp. 13–20.
- Schrauwen, E.J.A. & Fouchier, R.A.M. (2014). Host adaptation and transmission of influenza A viruses in mammals. *Emerging Microbes & Infections*. 3 (2).
- Scull, M.A., Gillim-Ross, L., Santos, C., Roberts, K.L., Bordonali, E., Subbarao, K., Barclay, W.S. & Pickles, R.J. (2009). Avian Influenza Virus Glycoproteins Restrict Virus Replication and Spread through Human Airway Epithelium at Temperatures of the Proximal Airways R. A. M. Fouchier (ed.). *PLoS pathogens*. 5 (5).
- Selman, M., Dankar, S.K., Forbes, N.E., Jia, J.-J. & Brown, E.G. (2012). Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. *Emerging Microbes & Infections*. 1 (11).

- Seo, S.H., Webby, R. & Webster, R.G. (2004). No apoptotic deaths and different levels of inductions of inflammatory cytokines in alveolar macrophages infected with influenza viruses. *Virology*. 329 (2). pp. 270–279.
- Shandilya, J. & Roberts, S.G.E. (2012). The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. *Biochimica Et Biophysica Acta- Gene Regulatory Mechanisms*. 1819 (5). pp. 391–400.
- Shapiro, G.I., Gurney, T. & Krug, R.M. (1987). Influenza virus gene expression: control mechanisms at early and late times of infection and nuclear-cytoplasmic transport of virus-specific RNAs. *Journal of Virology*. 61 (3). pp. 764–773.
- Shaw, M.L., Stone, K.L., Colangelo, C.M., Gulcicek, E.E. & Palese, P. (2008). Cellular proteins in influenza virus particles. K. Früh (ed.). *PLoS pathogens*. 4 (6). p. e1000085.
- Shaw, M.W. & Lamb, R.A. (1984). A Specific Sub-Set of Host-Cell Messenger-Rnas Prime Influenza-Virus Messenger-Rna Synthesis. *Virus Research*. 1 (6). pp. 455–467.
- Shepard, P.J., Choi, E.-A., Lu, J., Flanagan, L.A., Hertel, K.J. & Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *Rna-a Publication of the Rna Society*. 17 (4). pp. 761–772.
- Shih, S.R. & Krug, R.M. (1996). Novel exploitation of a nuclear function by influenza virus: The cellular SF2/ASF splicing factor controls the amount of the essential viral M2 ion channel protein in infected cells. *Embo Journal*. 15 (19). pp. 5415–5427.
- Shih, S.R., Nemeroff, M.E. & Krug, R.M. (1995). The Choice of Alternative 5' Splice Sites in Influenza-Virus M1 Messenger-Rna Is Regulated by the Viral Polymerase Complex. *Proceedings of the National Academy of Sciences of the United States of America*. 92 (14). pp. 6324–6328.
- Shih, S.R., Suen, P.C., Chen, Y.S. & Chang, S.C. (1998). A novel spliced transcript of influenza A/WSN/33 virus. *Virus Genes*. 17 (2). pp. 179–183.
- Shinya, K., Hamm, S., Hatta, M., Ito, H., Ito, T. & Kawaoka, Y. (2004). PB2 amino acid at position 627 affects, replicative efficiency, but not cell tropism, of Hong Kong H5N1 influenza A viruses in mice. *Virology*. 320 (2). pp. 258–266.
- Sikora, D., Rocheleau, L., Brown, E.G. & Pelchat, M. (2014). Deep sequencing reveals the eight facets of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process. *Scientific reports*. 4 (1).
- Simons, K. & Ikonen, E. (1997). Functional rafts in cell membranes. *Nature*. 387 (6633). pp. 569–572.
- Skehel, J.J. & Hay, A.J. (1978). Nucleotide-Sequences at 5'-Termini of Influenza-Virus RNAs and Their Transcripts. *Nucleic acids research*. 5 (4). pp. 1207–1219.
- Smed-Sörensen, A., Chalouni, C., Chatterjee, B., Cohn, L., Blattmann, P., Nakamura, N., Delamarre, L. & Mellman, I. (2012). Influenza A virus infection of human primary dendritic cells impairs their ability to cross-present antigen to CD8 T cells. C. F. Basler (ed.). *PLoS pathogens*. 8 (3). p. e1002572.

- Smith, D.B. & Inglis, S.C. (1985). Regulated Production of an Influenza-Virus Spliced Messenger-Rna Mediated by Virus-Specific Products. *Embo Journal*. 4 (9). pp. 2313–2319.
- Smith, G.L. & Hay, A.J. (1982). Replication of the influenza virus genome. *Virology*. 118 (1). pp. 96–108.
- Sonenberg, N. & Hinnebusch, A.G. (2009). Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell*. 136 (4). pp. 731–745.
- Song, M.-S., Pascua, P.N.Q., Lee, J.H., Baek, Y.H., Lee, O.-J., Kim, C.-J., Kim, H., Webby, R.J., Webster, R.G. & Choi, Y.K. (2009). The polymerase acidic protein gene of influenza A virus contributes to pathogenicity in a mouse model. *Journal of Virology*. 83 (23). pp. 12325–12335.
- Song, W., Wang, P., Mok, B.W.-Y., Lau, S.-Y., Huang, X., Wu, W.-L., Zheng, M., Wen, X., Yang, S., Chen, Y., Li, L., Yuen, K.-Y. & Chen, H. (2014). The K526R substitution in viral protein PB2 enhances the effects of E627K on influenza virus replication. *Nature Communications*. 5. p. 5509.
- Spooner, L. & Barry, R.D. (1977). Participation of Dna-Dependent Rna Polymerase-Ii in Replication of Influenza-Viruses. *Nature*. 268 (5621). pp. 650–652.
- Steel, J., Lowen, A.C., Mubareka, S. & Palese, P. (2009). Transmission of influenza virus in a mammalian host is increased by PB2 amino acids 627K or 627E/701N. R. S. Baric (ed.). *PLoS pathogens*. 5 (1). p. e1000252.
- Stieneke-Gröber, A., Vey, M., Angliker, H., Shaw, E., Thomas, G., Roberts, C., Klenk, H.-D. & Garten, W. (1992). Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *Embo Journal*. 11 (7). pp. 2407–2414.
- Subbarao, E.K., London, W. & Murphy, B.R. (1993). A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *Journal of Virology*. 67 (4). pp. 1761–1764.
- Sugita, Y., Sagara, H., Noda, T. & Kawaoka, Y. (2013). Configuration of Viral Ribonucleoprotein Complexes within the Influenza A Virion. *Journal of Virology*. 87 (23). pp. 12879–12884.
- Sugiyama, K., Kawaguchi, A., Okuwaki, M. & Nagata, K. (2015). pp32 and APRIL are host cell-derived regulators of influenza virus RNA synthesis from cRNA. *eLife*. 4. p. 4682.
- Suspene, R., Petit, V., Puyraimond-Zemmour, D., Aynaud, M.-M., Henry, M., Guetard, D., Rusniok, C., Wain-Hobson, S. & Vartanian, J.-P. (2011). Double-Stranded RNA Adenosine Deaminase ADAR-1-Induced Hypermutated Genomes among Inactivated Seasonal Influenza and Live Attenuated Measles Virus Vaccines. *Journal of Virology*. 85 (5). pp. 2458–2462.
- Suspene, R., Renard, M., Henry, M., Guetard, D., Puyraimond-Zemmour, D., Billecocq, A., Bouloy, M., Tangy, F., Vartanian, J.-P. & Wain-Hobson, S. (2008). Inversing the natural hydrogen bonding rule to selectively amplify GC-rich ADAR-edited RNAs. *Nucleic acids research*. 36 (12). pp. e72–e72.

- Swale, C., Monod, A., Tengo, L., Labaronne, A., Garzoni, F., Bourhis, J.-M., Cusack, S., Schoehn, G., Berger, I., Ruigrok, R.W.H. & Crepin, T. (2016). Structural characterization of recombinant IAV polymerase reveals a stable complex between viral PA-PB1 heterodimer and host RanBP5. *Scientific reports*. 6. p. 24727.
- Sysoev, V.O., Fischer, B., Frese, C.K., Gupta, I., Krijgsveld, J., Hentze, M.W., Castello, A. & Ephrussi, A. (2016). Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in *Drosophila*. *Nature Communications*. 7. p. 12128.
- Taft, A.S., Ozawa, M., Fitch, A., Depasse, J.V., Halfmann, P.J., Hill-Batorski, L., Hatta, M., Friedrich, T.C., Lopes, T.J.S., Maher, E.A., Ghedin, E., Macken, C.A., Neumann, G. & Kawaoka, Y. (2015). Identification of mammalian-adapting mutations in the polymerase complex of an avian H5N1 influenza virus. *Nature Communications*. 6. p. 7491.
- Takeda, M., Leser, G.P., Russell, C.J. & Lamb, R.A. (2003). Influenza virus hemagglutinin concentrates in lipid raft microdomains for efficient viral fusion. *Proceedings of the National Academy of Sciences of the United States of America*. 100 (25). pp. 14610–14617.
- Tamuri, A.U., Reis, Dos, M., Hay, A.J. & Goldstein, R.A. (2009). Identifying changes in selective constraints: host shifts in influenza. C. Fraser (ed.). *Plos Computational Biology*. 5 (11). p. e1000564.
- Tan, L., Su, S., Smith, D.K., He, S., Zheng, Y., Shao, Z., Ma, J., Zhu, H. & Zhang, G. (2014). A combination of HA and PA mutations enhances virulence in a mouse-adapted H6N6 influenza A virus. *Journal of Virology*. 88 (24). pp. 14116–14125.
- Tarendeau, F., Boudet, J., Guilligay, D., Mas, P.J., Bougault, C.M., Boulo, S., Baudin, F., Ruigrok, R.W.H., Daigle, N., Ellenberg, J., Cusack, S., Simorre, J.-P. & Hart, D.J. (2007). Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nature Structural & Molecular Biology*. 14 (3). pp. 229–233.
- Tarendeau, F., Crepin, T., Guilligay, D., Ruigrok, R.W.H., Cusack, S. & Hart, D.J. (2008). Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit Y. Kawaoka (ed.). *PLoS pathogens*. 4 (8).
- Taubenberger, J.K. & Kash, J.C. (2010). Influenza virus evolution, host adaptation, and pandemic formation. *Cell host & microbe*. 7 (6). pp. 440–451.
- Taubenberger, J.K. & Morens, D.M. (2006). 1918 Influenza: the mother of all pandemics. *Emerging infectious diseases*. 12 (1). pp. 15–22.
- Taubenberger, J.K., Reid, A.H., Lourens, R.M., Wang, R.X., Jin, G.Z. & Fanning, T.G. (2005). Characterization of the 1918 influenza virus polymerase genes. *Nature*. 437 (7060). pp. 889–893.
- Taylor, D.R., Puig, M., Darnell, M.E.R., Mihalik, K. & Feinstone, S.M. (2005). New antiviral pathway that mediates hepatitis C virus replicon interferon sensitivity through ADAR1. *Journal of Virology*. 79 (10). pp. 6291–6298.
- tenOever, B.R., Ng, S.-L., Chua, M.A., McWhirter, S.M., García-Sastre, A. & Maniatis, T. (2007). Multiple functions of the IKK-related kinase IKKepsilon in interferon-mediated antiviral immunity. *Science (New York, N.Y.)*. 315 (5816). pp. 1274–1278.

- Thierry, E., Guilligay, D., Kosinski, J., Bock, T., Gaudon, S., Round, A., Pflug, A., Hengrung, N., Omari, El, K., Baudin, F., Hart, D.J., Beck, M. & Cusack, S. (2016). Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2 Domains. *Molecular cell*. 61 (1). pp. 125–137.
- Tiley, L.S., Hagen, M., Matthews, J.T. & Krystal, M. (1994). Sequence-Specific Binding of the Influenza-Virus Rna-Polymerase to Sequences Located at the 5' Ends of the Viral Rnas. *Journal of Virology*. 68 (8). pp. 5108–5116.
- Tochio, N., Umehara, T., Munemasa, Y., Suzuki, T., Sato, S., Tsuda, K., Koshiba, S., Kigawa, T., Nagai, R. & Yokoyama, S. (2010). Solution structure of histone chaperone ANP32B: interaction with core histones H3-H4 through its acidic concave domain. *Journal of molecular biology*. 401 (1). pp. 97–114.
- Tokunaga, K., Shibuya, T., Ishihama, Y., Tadakuma, H., Ide, M., Yoshida, M., Funatsu, T., Ohshima, Y. & Tani, T. (2006). Nucleocytoplasmic transport of fluorescent mRNA in living mammalian cells: nuclear mRNA export is coupled to ongoing gene transcription. *Genes to Cells*. 11 (3). pp. 305–317.
- Tomescu, A.I., Robb, N.C., Hengrung, N., Fodor, E. & Kapanidis, A.N. (2014). Single-molecule FRET reveals a corkscrew RNA structure for the polymerase-bound influenza virus promoter. *Proceedings of the National Academy of Sciences of the United States of America*. 111 (32). pp. E3335–E3342.
- Topisirovic, I., Svitkin, Y.V., Sonenberg, N. & Shatkin, A.J. (2011). Cap and cap-binding proteins in the control of gene expression. *Wiley Interdisciplinary Reviews-Rna*. 2 (2). pp. 277–298.
- Toth, A.M., Li, Z., Cattaneo, R. & Samuel, C.E. (2009). RNA-specific adenosine deaminase ADAR1 suppresses measles virus-induced apoptosis and activation of protein kinase PKR. *The Journal of biological chemistry*. 284 (43). pp. 29350–29356.
- Tsai, P.-L., Chiou, N.-T., Kuss, S., García-Sastre, A., Lynch, K.W. & Fontoura, B.M.A. (2013). Cellular RNA Binding Proteins NS1-BP and hnRNP K Regulate Influenza A Virus RNA Splicing A. Pekosz (ed.). *PLoS pathogens*. 9 (6).
- Tsvetanova, N.G., Klass, D.M., Salzman, J. & Brown, P.O. (2010). Proteome-Wide Search Reveals Unexpected RNA-Binding Proteins in *Saccharomyces cerevisiae* R. Aramayo (ed.). *Plos One*. 5 (9).
- Turrell, L., Lyall, J.W., Tiley, L.S., Fodor, E. & Vreede, F.T. (2013). The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. *Nature Communications*. 4. p. 1591.
- Ulitzur, N., Rancaño, C. & Pfeffer, S.R. (1997). Biochemical characterization of mapmodulin, a protein that binds microtubule-associated proteins. *The Journal of biological chemistry*. 272 (48). pp. 30577–30582.
- Umbach, J.L., Yen, H.-L., Poon, L.L.M. & Cullen, B.R. (2010). Influenza A Virus Expresses High Levels of an Unusual Class of Small Viral Leader RNAs in Infected Cells. *Mbio*. 1 (4). pp. –e00204–10.

- van Wielink, R., Harmsen, M.M., Martens, D.E., Peeters, B.P.H., Wijffels, R.H. & Moormann, R.J.M. (2012). Mutations in the M-Gene Segment Can Substantially Increase Replication Efficiency of NS1 Deletion Influenza A Virus in MDCK Cells. *Journal of Virology*. 86 (22). pp. 12341–12350.
- Vasin, A.V., Temkina, O.A., Egorov, V.V., Klotchenko, S.A., Plotnikova, M.A. & Kiselev, O.I. (2014). Molecular mechanisms enhancing the proteome of influenza A viruses: an overview of recently discovered proteins. *Virus Research*. 185. pp. 53–63.
- Velthuis, te, A.J.W. (2014). Common and unique features of viral RNA-dependent polymerases. *Cellular and molecular life sciences : CMLS*. 71 (22). pp. 4403–4420.
- Velthuis, te, A.J.W. & Fodor, E. (2016). Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nature reviews. Microbiology*. 14 (8). pp. 479–493.
- Velthuis, te, A.J.W., Robb, N.C., Kapanidis, A.N. & Fodor, E. (2016). The role of the priming loop in Influenza A virus RNA synthesis. *Nature microbiology*. 1 (5). p. nmicrobiol201629.
- Vidic, J., Noiray, M., Bagchi, A. & Slama-Schwok, A. (2016). Identification of a Novel Complex between the Nucleoprotein and PA(1-27) of Influenza A Virus Polymerase. *Biochemistry*. 55 (31). pp. 4259–4262.
- Vreede, F.T. & Brownlee, G.G. (2007). Influenza virion-derived viral ribonucleoproteins synthesize both mRNA and cRNA *in vitro*. *Journal of Virology*. 81 (5). pp. 2196–2204.
- Vreede, F.T. & Fodor, E. (2010). The role of the influenza virus RNA polymerase in host shut-off. *Virulence*. 1 (5). pp. 436–439.
- Vreede, F.T., Chan, A.Y., Sharps, J. & Fodor, E. (2010). Mechanisms and functional implications of the degradation of host RNA polymerase II in influenza virus infected cells. *Virology*. 396 (1). pp. 125–134.
- Vreede, F.T., Gifford, H. & Brownlee, G.G. (2008). Role of initiating nucleoside triphosphate concentrations in the regulation of influenza virus replication and transcription. *Journal of Virology*. 82 (14). pp. 6902–6910.
- Vreede, F.T., Jung, T.E. & Brownlee, G.G. (2004). Model suggesting that replication of influenza virus is regulated by stabilization of replicative intermediates. *Journal of Virology*. 78 (17). pp. 9568–9572.
- Vreede, F.T., Ng, A.K.-L., Shaw, P.-C. & Fodor, E. (2011). Stabilization of influenza virus replication intermediates is dependent on the RNA-binding but not the homooligomerization activity of the viral nucleoprotein. *Journal of Virology*. 85 (22). pp. 12073–12078.
- Wakefield, L. & Brownlee, G.G. (1989). RNA-binding properties of influenza A virus matrix protein M1. *Nucleic acids research*. 17 (21). pp. 8569–8580.
- Walden, W.E., Selezneva, A.I., Dupuy, J., Volbeda, A., Fontecilla-Camps, J.C., Theil, E.C. & Volz, K. (2006). Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science (New York, N.Y.)*. 314 (5807). pp. 1903–1908.

- Wang, P., Palese, P. & O'Neill, R.E. (1997). The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza A virus nucleoprotein NP is a nonconventional nuclear localization signal. *Journal of Virology*. 71 (3). pp. 1850–1856.
- Wang, W., Cui, Z.-Q., Han, H., Zhang, Z.-P., Wei, H.-P., Zhou, Y.-F., Chen, Z. & Zhang, X.-E. (2008). Imaging and characterizing influenza A virus mRNA transport in living cells. *Nucleic acids research*. 36 (15). pp. 4913–4928.
- Ward, S.V., George, C.X., Welch, M.J., Liou, L.-Y., Hahm, B., Lewicki, H., la Torre, de, J.C., Samuel, C.E. & Oldstone, M.B. (2011). RNA editing enzyme adenosine deaminase is a restriction factor for controlling measles virus replication that also is required for embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 108 (1). pp. 331–336.
- Watanabe, K., Handa, H., Mizumoto, K. & Nagata, K. (1996). Mechanism for inhibition of influenza virus RNA polymerase activity by matrix protein. *Journal of Virology*. 70 (1). pp. 241–247.
- Watanabe, T., Kawakami, E., Shoemaker, J.E., Lopes, T.J.S., Matsuoka, Y., Tomita, Y., Kozuka-Hata, H., Gorai, T., Kuwahara, T., Takeda, E., Nagata, A., Takano, R., Kiso, M., Yamashita, M., Sakai-Tagawa, Y., Katsura, H., Nonaka, N., Fujii, H., Fujii, K., Sugita, Y., Noda, T., Goto, H., Fukuyama, S., Watanabe, S., Neumann, G., Oyama, M., Kitano, H. & Kawaoka, Y. (2014). Influenza virus-host interactome screen as a platform for antiviral drug development. *Cell host & microbe*. 16 (6). pp. 795–805.
- Wessels, H.-H., Imami, K., Baltz, A.G., Kolinski, M., Beldovskaya, A., Selbach, M., Small, S., Ohler, U. & Landthaler, M. (2016). The mRNA-bound proteome of the early fly embryo. *Genome research*. 26 (7). pp. 1000–1009.
- WHO (2017a). *Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003-2017*. [Online]. Available from: http://www.who.int/influenza/human_animal_interface/2017_07_25_tableH5N1.pdf?ua=1.
- WHO (2016). *Influenza (Seasonal)*. [Online]. Available from: <http://www.who.int/mediacentre/factsheets/fs211/en/>.
- WHO (2017b). *Influenza at the human-animal interface*. [Online]. Available from: http://www.who.int/influenza/human_animal_interface/Influenza_Summary_IRA_HA_interface_07_25_2017.pdf?ua=1.
- Winter, G. & Fields, S. (1980). Cloning of influenza cDNA into M13: the sequence of the RNA segment encoding the A/PR/8/34 matrix protein. *Nucleic acids research*. 8 (9). pp. 1965–1974.
- Winter, G. & Fields, S. (1982). Nucleotide sequence of human influenza A/PR/8/34 segment 2. *Nucleic acids research*. 10 (6). pp. 2135–2143.
- Winter, G. & Fields, S. (1981). The Structure of the Gene Encoding the Nucleoprotein of Human Influenza-Virus A-Pr-8-34. *Virology*. 114 (2). pp. 423–428.
- Winter, G., Fields, S. & Brownlee, G.G. (1981a). Nucleotide-Sequence of the Hemagglutinin Gene of a Human Influenza-Virus H-1 Subtype. *Nature*. 292 (5818). pp. 72–75.

- Winter, G., Fields, S., GAIT, M.J. & Brownlee, G.G. (1981b). The Use of Synthetic Oligodeoxynucleotide Primers in Cloning and Sequencing Segment 8 of Influenza-Virus (a-Pr-8-34). *Nucleic acids research*. 9 (2). pp. 237–245.
- Wise, H.M., Foeglein, A., Sun, J., Dalton, R.M., Patel, S., Howard, W., Anderson, E.C., Barclay, W.S. & Digard, P. (2009). A complicated message: Identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. *Journal of Virology*. 83 (16). pp. 8021–8031.
- Wise, H.M., Hutchinson, E.C., Jagger, B.W., Stuart, A.D., Kang, Z.H., Robb, N., Schwartzman, L.M., Kash, J.C., Fodor, E., Firth, A.E., Gog, J.R., Taubenberger, J.K. & Digard, P. (2012a). Identification of a Novel Splice Variant Form of the Influenza A Virus M2 Ion Channel with an Antigenically Distinct Ectodomain A. Pekosz (ed.). *PLoS pathogens*. 8 (11).
- Wise, H.M., Hutchinson, E.C., Jagger, B.W., Stuart, A.D., Kang, Z.H., Robb, N., Schwartzman, L.M., Kash, J.C., Fodor, E., Firth, A.E., Gog, J.R., Taubenberger, J.K. & Digard, P. (2012b). Identification of a Novel Splice Variant Form of the Influenza A Virus M2 Ion Channel with an Antigenically Distinct Ectodomain A. Pekosz (ed.). *PLoS pathogens*. 8 (11).
- Wong, S.K. & Lazinski, D.W. (2002). Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proceedings of the National Academy of Sciences of the United States of America*. 99 (23). pp. 15118–15123.
- Wright, P.F., Neumann, G. & Kawaoka, Y. (2013). Orthomyxoviridae. In: D. M. Knipe & P. M. Howley (eds.). *Fields Virology*. Philadelphia: Lippincott Williams and Wilkins, pp. 1186–1243.
- Wu, H., Peng, X., Peng, X., Cheng, L., Jin, C., Lu, X., Xie, T., Yao, H. & Wu, N. (2016). Multiple amino acid substitutions involved in the adaptation of avian-origin influenza A (H10N7) virus in mice. *Archives of virology*. 161 (4). pp. 977–980.
- Wurzer, W.J., Planz, O., Ehrhardt, C., Giner, M., Silberzahn, T., Pleschka, S. & Ludwig, S. (2003). Caspase 3 activation is essential for efficient influenza virus propagation. *Embo Journal*. 22 (11). pp. 2717–2728.
- Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Uversky, V.N. & Obradovic, Z. (2007). Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *Journal of proteome research*. 6 (5). pp. 1882–1898.
- Xu, C., Hu, W.-B., Xu, K., He, Y.-X., Wang, T.-Y., Chen, Z., Li, T.-X., Liu, J.-H., Buchy, P. & Sun, B. (2012). Amino acids 473V and 598P of PB1 from an avian-origin influenza A virus contribute to polymerase activity, especially in mammalian cells. *The Journal of general virology*. 93 (Pt 3). pp. 531–540.
- Yamada, S., Hatta, M., Staker, B.L., Watanabe, S., Imai, M., Shinya, K., Sakai-Tagawa, Y., Ito, M., Ozawa, M., Watanabe, T., Sakabe, S., Li, C., Kim, J.H., Myler, P.J., Phan, I., Raymond, A., Smith, E., Stacy, R., Nidom, C.A., Lank, S.M., Wiseman, R.W., Bimber, B.N., O'Connor, D.H., Neumann, G., Stewart, L.J. & Kawaoka, Y. (2010). Biological and structural characterization of a host-adapting amino acid in influenza virus. *PLoS pathogens*. 6 (8). p. e1001034.

- Yamanaka, K., Ishihama, A. & Nagata, K. (1990). Reconstitution of Influenza-Virus Rna-Nucleoprotein Complexes Structurally Resembling Native Viral Ribonucleoprotein Cores. *The Journal of biological chemistry*. 265 (19). pp. 11151–11155.
- Yamayoshi, S., Watanabe, M., Goto, H. & Kawaoka, Y. (2016). Identification of a Novel Viral Protein Expressed from the PB2 Segment of Influenza A Virus T. S. Dermody (ed.). *Journal of Virology*. 90 (1). pp. 444–456.
- Yángüez, E. & Nieto, A. (2011). So similar, yet so different: selective translation of capped and polyadenylated viral mRNAs in the influenza virus infected cell. *Virus Research*. 156 (1-2). pp. 1–12.
- Yángüez, E., Castello, A., Welnowska, E., Carrasco, L., Goodfellow, I. & Nieto, A. (2011). Functional impairment of eIF4A and eIF4G factors correlates with inhibition of influenza virus mRNA translation. *Virology*. 413 (1). pp. 93–102.
- Ye, Q., Krug, R.M. & Tao, Y.J. (2006). The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature*. 444 (7122). pp. 1078–1082.
- York, A. & Fodor, E. (2013). Biogenesis, assembly, and export of viral messenger ribonucleoproteins in the influenza A virus infected cell. *RNA biology*. 10 (8). pp. 1274–1282.
- York, A., Hengrung, N., Vreede, F.T., Huiskonen, J.T. & Fodor, E. (2013). Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. *Proceedings of the National Academy of Sciences of the United States of America*. 110 (45). pp. E4238–45.
- York, A., Hutchinson, E.C. & Fodor, E. (2014). Interactome analysis of the influenza A virus transcription/replication machinery identifies protein phosphatase 6 as a cellular factor required for efficient virus replication. *Journal of Virology*. 88 (22). pp. 13284–13299.
- Yu, M., Liu, X., Cao, S., Zhao, Z., Zhang, K., Xie, Q., Chen, C., Gao, S., Bi, Y., Sun, L., Ye, X., Gao, G.F. & Liu, W. (2012). Identification and characterization of three novel nuclear export signals in the influenza A virus nucleoprotein. *Journal of Virology*. 86 (9). pp. 4970–4980.
- Yu, Z., Cheng, K., Xin, Y., Sun, W., Li, X., Huang, J., Zhang, K., Yang, S., Wang, T., Zheng, X., Wang, H., Hua, Y., Chai, H., Qin, C., Qian, J., Gao, Y. & Xia, X. (2014). Multiple amino acid substitutions involved in the adaptation of H6N1 avian influenza virus in mice. *Veterinary microbiology*. 174 (3-4). pp. 316–321.
- Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., Lv, Z., Ge, R., Li, X., Deng, T., Fodor, E., Rao, Z. & Liu, Y. (2009). Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature*. 458 (7240). pp. 909–913.
- Zahn, R.C., Schelp, I., Utermöhlen, O. & Laer, von, D. (2007). A-to-G hypermutation in the genome of lymphocytic choriomeningitis virus. *Journal of Virology*. 81 (2). pp. 457–464.

- Zalfa, F., Adinolfi, S., Napoli, I., Kuhn-Holsken, E., Urlaub, H., Achsel, T., Pastore, A. & Bagni, C. (2005). Fragile X mental retardation protein (FMRP) binds specifically to the brain cytoplasmic RNAs BC1/BC200 via a novel RNA-binding motif. *The Journal of biological chemistry*. 280 (39). pp. 33403–33410.
- Zamora-Caballero, S., Šiaučiunaite-Gaubard, L. & Bravo, J. (2015). High-resolution crystal structure of the leucine-rich repeat domain of the human tumour suppressor PP32A (ANP32A). *Acta crystallographica. Section F, Structural biology communications*. 71 (Pt 6). pp. 684–687.
- Zhang, J., Leser, G.P., Pekosz, A. & Lamb, R.A. (2000). The cytoplasmic tails of the influenza virus spike glycoproteins are required for normal genome packaging. *Virology*. 269 (2). pp. 325–334.
- Zhang, Y., Zhang, Q., Gao, Y., He, X., Kong, H., Jiang, Y., Guan, Y., Xia, X., Shu, Y., Kawaoka, Y., Bu, Z. & Chen, H. (2012). Key molecular factors in hemagglutinin and PB2 contribute to efficient transmission of the 2009 H1N1 pandemic influenza virus. *Journal of Virology*. 86 (18). pp. 9666–9674.
- Zhang, Y., Zhang, Q., Kong, H., Jiang, Y., Gao, Y., Deng, G., Shi, J., Tian, G., Liu, L., Liu, J., Guan, Y., Bu, Z. & Chen, H. (2013). H5N1 hybrid viruses bearing 2009/H1N1 virus genes transmit in guinea pigs by respiratory droplet. *Science (New York, N.Y.)*. 340 (6139). pp. 1459–1463.
- Zhang, Z., Boonen, K., Ferrari, P., Schoofs, L., Janssens, E., van Noort, V., Rolland, F. & Geuten, K. (2016). UV crosslinked RNA-binding proteins captured from leaf mesophyll protoplasts. *Plant Methods*. 12 (1).
- Zhao, Y., Yu, Z., Liu, L., Wang, T., Sun, W., Wang, C., Xia, Z., Gao, Y., Zhou, B., Qian, J. & Xia, X. (2016). Adaptive amino acid substitutions enhance the virulence of a novel human H7N9 influenza virus in mice. *Veterinary microbiology*. 187. pp. 8–14.
- Zhirnov, O.P., Ikizler, M.R. & Wright, P.F. (2002). Cleavage of influenza A virus hemagglutinin in human respiratory epithelium is cell associated and sensitive to exogenous antiproteases. *Journal of Virology*. 76 (17). pp. 8682–8689.
- Zhirnov, O.P., Poyarkov, S.V., Vorob'eva, I.V., Safonova, O.A., Malyshev, N.A. & Klenk, H.-D. (2007). Segment NS of influenza A virus contains an additional gene NSP in positive-sense orientation. *Doklady. Biochemistry and biophysics*. 414. pp. 127–133.
- Zhou, B., Li, Y., Halpin, R., Hine, E., Spiro, D.J. & Wentworth, D.E. (2011). PB2 residue 158 is a pathogenic determinant of pandemic H1N1 and H5 influenza A viruses in mice. *Journal of Virology*. 85 (1). pp. 357–365.
- Zhou, B., Pearce, M.B., Li, Y., Wang, J., Mason, R.J., Tumpey, T.M. & Wentworth, D.E. (2013). Asparagine substitution at PB2 residue 701 enhances the replication, pathogenicity, and transmission of the 2009 pandemic H1N1 influenza A virus. A. Pekosz (ed.). *Plos One*. 8 (6). p. e67616.
- Zhu, H., Wang, J., Wang, P., Song, W., Zheng, Z., Chen, R., Guo, K., Zhang, T., Peiris, J.S.M., Chen, H. & Guan, Y. (2010). Substitution of lysine at 627 position in PB2 protein does not change virulence of the 2009 pandemic H1N1 virus in mice. *Virology*. 401 (1). pp. 1–5.

Zhu, W., Zhu, Y., Qin, K., Yu, Z., Gao, R., Yu, H., Zhou, J. & Shu, Y. (2012). Mutations in polymerase genes enhanced the virulence of 2009 pandemic H1N1 influenza virus in mice. P. Digard (ed.). *Plos One*. 7 (3). p. e33383.

APPENDICES

Appendix A Influenza A /WSN/33 (H1N1) nucleotide and protein sequences

Protein sequences

>PB2

MERIKELRNLMSSQSRTEILTKTTVDHMAIIKKYTSGRQEKNPALRMKMMMAMKYPITADKRITEMIPERNEQ
GQTLWSKMNDAGSDRMVSPPLAVTWNNRNGPVTSTVHYPKIYKTYFEKVERLKHGTFGPVHFRNQVKIRRRVD
INPGHADLSAKEAQDVIMEVVPNEVGARILTSQSLTTTKEKKEELQGCKISPLMVAYMLERELVRKTRFLP
VAGGTSSVYIEVLHLTQGTCEWQMYTPGGEARNDVDQSLIIAARNIVRRATVSADPLASLLEMCHSTQIGGI
RMVNILRQNPTEEQAVIDICKAAMGLRISSSFSGGFTFKRTSGSSVKREEEVLGTNLQTLKIRVHEGYEEFTM
VGRRATAILRKATRRLIQLIVSGRDEQSIAEAIIVAMVFSQEDCMIKAVRGDLNLFVNRRANQRLNPMHQLLRHF
QKDAKVLFFQNWGIESIDNVMGMIGILPDMTPSTEMSMRGVRIKMGVDEYSSAEKIVVSIIDRFLVRDQRGV
LLSPPEEVSETQGTETLTITYSSSMWEINGPESVLVNTYQWIIIRNWETVKIQWSQNPTMLYNKMEFEPFQSLV
PKAVRGQYSGFVRTLFQQMRDVLGTFDTAQIIKLLPFAAAPPKQSGMQFSSLTINVRGSGMRILVRGNSPVFN
YNKTTKRLTVLGKDAGPLTEDPDEGTAGVESAVLRGFLILGKEDRRYGPALSINELSNLAKGEKANVLIQGD
VVLVMKRKRNSILTDSQTATKRIRMAIN

>PB2-S1

MERIKELRNLMSSQSRTEILTKTTVDHMAIIKKYTSGRQEKNPALRMKMMMAMKYPITADKRITEMIPERNEQ
GQTLWSKMNDAGSDRMVSPPLAVTWNNRNGPVTSTVHYPKIYKTYFEKVERLKHGTFGPVHFRNQVKIRRRVD
INPGHADLSAKEAQDVIMEVVPNEVGARILTSQSLTTTKEKKEELQGCKISPLMVAYMLERELVRKTRFLP
VAGGTSSVYIEVLHLTQGTCEWQMYTPGGEARNDVDQSLIIAARNIVRRATVSADPLASLLEMCHSTQIGGI
RMVNILRQNPTEEQAVIDICKAAMGLRISSSFSGGFTFKRTSGSSVKREEEVLGTNLQTLKIRVHEGYEEFTM
VGRRATAILRKATRRLIQLIVSGRDEQSIAEAIIVAMVFSQEDCMIKAVRGDLNLFVNRRANQRLNPMHQLLRHF
QKDAKVLFFQNWGIESIDNVMGMIGILPDMTPSTEMSMRGVRIKMGVDEYSSAEKIVPLHQSKVECCSPH

>PB1

MDVNPTLLFLKVPQAIAISTTFPYTGDPYPYSHGTGTGYTMDTVNRTHQYSEGRWTTNTETGAPQLNPIDGPL
PEDNEPSGYAQTDCVLEAMAFLEESHPIGFETSCLETMEVVQQTRVDKLTQGRQTYDWTNLNRNQAATALANT
IEVFRSNGLTANESGRLIDFLKDVESMNKEEMEITTHFQRKRVRDNMTKKMVTQRTIGKRKQRLNKRSLI
RALTLNMTKDAERGKLRRAIATPGMQIRGFVYFVETLARSICEKLEQSGLPVGGNEKKAKLANVVRKMMTN
SQDTEISFTITGDNTKWNENQNPRMFLAMITYITRNQPEWFRNVLSIAPIMFSNKMARLGKGYMFESKSMKLR
TQIPAEMLASIDLKYFNDSTRKKIEKIRPLLDGTASLSPGMMGMFNMLSTVLGVSILNLGQKRHTKTTYWW
DGLQSSDDFALIVNAPNHEGIQAGVNRFYRTCKLLGINMSKKKSYINRTGTTEFTSFFYRYGFVANFSMELPS
FGVSGINESADMSIGVTVIKNMINNDLGPATAQMALQLFIKDYRYTYRCHRGDTQIQTRRSFEIKKLWEQTH
SKAGLLVSDGGPNLYNIRNLHIPEVCLKWELMEDDYQGRLCNPLNPFVNHKDIESVNNAVIMPAHGPAKNMEY
DAVATTHSWIPKRNRISLNTSQRGILEDEQMYQKCCNLFKFFPSSSYRRPVGISSMVEAMVSRARIDARIDF
ESGRIKKEEFTEIMKICSTIEELRRQK

>PB1-N40

MDTVNRTHQYSEGRWTTNTETGAPQLNPIDGPLPEDNEPSGYAQTDCVLEAMAFLEESHPIGFETSCLETME
VVQQTRVDKLTQGRQTYDWTNLNRNQAATALANTIEVFRSNGLTANESGRLIDFLKDVESMNKEEMEITTHF
QRKRVRDNMTKKMVTQRTIGKRKQRLNKRSLIRALTLNMTKDAERGKLRRAIATPGMQIRGFVYFVETL
ARSICEKLEQSGLPVGGNEKKAKLANVVRKMMTN
SQDTEISFTITGDNTKWNENQNPRMFLAMITYITRNQPEWFRNVLSIAPIMFSNKMARLGKGYMFESKSMKLR
TQIPAEMLASIDLKYFNDSTRKKIEKIRPLLDGTASLSPGMMGMFNMLSTVLGVSILNLGQKRHTKTTYWW
DGLQSSDDFALIVNAPNHEGIQAGVNRFYRTCKLLGINMSKKKSYINRTGTTEFTSFFYRYGFVANFSMELPS
FGVSGINESADMSIGVTVIKNMINNDLGPATAQMALQLFIKDYRYTYRCHRGDTQIQTRRSFEIKKLWEQTH
SKAGLLVSDGGPNLYNIRNLHIPEVCLKWELMEDDYQGR

LCNPLNPFVNHKDIESVNNAVIMPAHGPAKNMEYDAVATTHSWIPKRNRSILNTSQRGILEDEQMYQKCCNLF
EKFFPSSSYRRPVGISSMVEAMVSRARIDARIDFESGRIKKEEFTEIMKICSTIEELRRQK

>PB1-F2

MGQEQTDPWILSTGHISTQKGEDGQQTPKLEHRNSTRLMGHCQKTMNQVVMKPQIVYWKQWPSLRNPILVSLR
PRVLKRWRLFSSKHEWTS

>PA

MEDFVRQCFNPMIVELAEKAMKEYGEDLKIETNKFAAICTHLEVCFMYSDFHFIDEQGESIVVELGDPNALLK
HRFEIIIEGRDRTIAWTVINSICNTTGAEKPKFLPDLYDYKKNRFIEIGVTRREVHIYYLEKANKIKSEKTHIH
IFSFTGEEMATKADYTLDEESRARIKTRLFTIRQEMASRGLWDSFRQSERGEETIEERFEITGTMRKLADQSL
PPNFSSLENFRAYVDGFEPNGYIEGKLSQMSKEVNARIEPFLKSTPRPLRLPDGPPCSQORSKFLLMDALKLSI
EDPSHEGEGIPLYDAIKCMRTFFGWKEPNVVKPHEKGINPNYLLSWKQVLAELQDIENEEKIPRTKNMKKTSQ
LKWALGENMAPEKVDFFDCKDVGDLKQYDSDEPELRSLSAWIQNEFNKACELTDSSWIELDEIGEDAAPIEHI
ASMRNYFTAEVSHCRATEYIMKGVYINTALLNASCAAMDDFQLIPMISKCRTKEGRKTNLYGFIIKGRSHL
RNDTDVNVFVSMEFSLTDPRLEPHKWEKYCVLEVGDMLLRSAGHVS RPMFLYVRTNGTSKIKMKWGMEMRRC
LLQSLQQIESMIEAESSVKEKDMTKEFFENKSETWPGVESPKEVEEGSIGKVCRTLLAKSVFNLSLYAS PQLEG
FSAESRKL LLLIVQALRDNLPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALR

>PA-N155

MATKADYTLDEESRARIKTRLFTIRQEMASRGLWDSFRQSERGEETIEERFEITGTMRKLADQSLPPNFSSLE
NFRAYVDGFEPNGYIEGKLSQMSKEVNARIEPFLKSTPRPLRLPDGPPCSQORSKFLLMDALKLSIEDPSHEGE
GIPLYDAIKCMRTFFGWKEPNVVKPHEKGINPNYLLSWKQVLAELQDIENEEKIPRTKNMKKTSQ LKWALGEN
MAPEKVDFFDCKDVGDLKQYDSDEPELRSLSAWIQNEFNKACELTDSSWIELDEIGEDAAPIEHIASMRNYF
TAEVSHCRATEYIMKGVYINTALLNASCAAMDDFQLIPMISKCRTKEGRKTNLYGFIIKGRSHLRNDTDVNV
FVSMEFSLTDPRLEPHKWEKYCVLEVGDMLLRSAGHVS RPMFLYVRTNGTSKIKMKWGMEMRRC LLQSLQQI
ESMIEAESSVKEKDMTKEFFENKSETWPGVESPKEVEEGSIGKVCRTLLAKSVFNLSLYAS PQLEGFSAESRKL
LLIVQALRDNLPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALR

>PA-N182

MASRGLWDSFRQSERGEETIEERFEITGTMRKLADQSLPPNFSSLENFRAYVDGFEPNGYIEGKLSQMSKEVN
ARIEPFLKSTPRPLRLPDGPPCSQORSKFLLMDALKLSIEDPSHEGEGIPLYDAIKCMRTFFGWKEPNVVKPHE
KGINPNYLLSWKQVLAELQDIENEEKIPRTKNMKKTSQ LKWALGENMAPEKVDFFDCKDVGDLKQYDSDEPEL
RSLASWIQNEFNKACELTDSSWIELDEIGEDAAPIEHIASMRNYFTAEVSHCRATEYIMKGVYINTALLNAS
CAAMDDFQLIPMISKCRTKEGRKTNLYGFIIKGRSHLRNDTDVNVFVSMEFSLTDPRLEPHKWEKYCVLEV
GDMLLRSAGHVS RPMFLYVRTNGTSKIKMKWGMEMRRC LLQSLQQIESMIEAESSVKEKDMTKEFFENKSETW
PGVESPKEVEEGSIGKVCRTLLAKSVFNLSLYAS PQLEGFSAESRKL LLLIVQALRDNLPGTFDLGGLYEAIEE
CLINDPWVLLNASWFNSFLTHALR

>PA-X

MEDFVRQCFNPMIVELAEKAMKEYGEDLKIETNKFAAICTHLEVCFMYSDFHFIDEQGESIVVELGDPNALLK
HRFEIIIEGRDRTIAWTVINSICNTTGAEKPKFLPDLYDYKKNRFIEIGVTRREVHIYYLEKANKIKSEKTHIH
IFSFTGEEMATKADYTLDEESRARIKTRLFTIRQEMASRGLWDSF
VSPREAKRQLKKDLKSQEQCASLPTKVSRTSPALKILEPMWMDSNRTATLRASFLKCPKK

>HA

MKAFVLVLLYAFVATDADTICIGYHANNSTDTVDTIFEKNVAVTHSVNLLDRHNGKLCKLKGIAPLQLGKCN
ITGWLLGNPECD SLLPARSWSYIVETPNSENGACYPGDFIDYEELREQLSSVSSLERFEIFPKESSWP NHTFN
GVTVSCSHRGKSSFYRNLLWLTKKGDSYPKLTNSYVNKGKEVLVLWG VHPSSSDEQQSLYSNGNAYVSVAS
SNYNRRFTPEIAARPKVKDQHGRMNYYWTLLEPGDTII FEATGNLIAPWYAFALS RGFESGIITSNASMHECN
TKCQTPQGSINSNLPFQNIHPVTIGECPKYVRSTKL RMTGLRNIPSIQYRGLFGA IAGFIEGGWTGMIDGWY
GYHHQNEQSGYAADQKSTQNAINGITNKVNSVIEKMNTQFTAVGKEFNNLEKRMENLNKKVDDGFLDIWTYN
AELLVLENERLTDFHDLNVKNLYEKVKSQ LKNNAKEIGNGCFEFYHKCDNECMESVRNGTYDYPKYSEESKL
NREKIDGVKLES MG VYQILAIYSTVASSLVLLVSLGAISFWMC SNGSLQCRICI

>NP

MATKGTKRSYEQMETDGERQNATEIRASVGK MIDGIGRFYIQMCTELKLS DYEGRLIQNSLTIERMVLSAFDE
RRNKYLEEHPSAGKDPKKTGGPIYRRVDGKWRRELILYDKEEIRRIWRQANNGDDATAGLTHMMIWHSNL NDA
TYQRTALVRTGMDPRMCSLMQGSTLPRRSGAAGAAVKG VGTMMELIRMIKRGINDRNFWRGENGRRTRIAY
ERMCNII LKGKFQTAAQRTMVDQVRESRNP GNAEFEDLI FLARSALILRGSAVHKSCLPACVYGS AVASGYDFE
REGYSLVGIDPFRL LQNSQVYSLIRPNENPAHKSQ LVMMACHSAAFEDLRVSSFIRGTKVVPRGKLSTRGVQI
ASNENMETMESSTLELRSRYWAIRTRSGGNTNQQRASSGQISIQPTFSVQRNLPFDRPTIMAAFTGNTEGRTS
DMRTEIIRLMESARPEDVSFQGRGVFELSDEKATSPIVPSFDM SNEG SYFFGDNAEEYDN

>NA
MNPNQKIITIGSICMVVGIIISLILQIGNIISIWISHSIQTGNQNHTGICNQGSITYKVAVAGQDSTSVILTGNS
SLCPIRGWAIHSDKNGIRIGSKGDVVFVIREPFISCSHLECRFTFFLTQGALLNDKHSRGTFKDRSPYRALMSCP
VGEAPSPYNSRFESVAWSASACHDGMGWLITIGISGPDDGAVAVLKYNGIITETIKSWRKNILRTQESECTCVN
GSCFTIMTDGPSDGLASYKIFKIEKGKVTKSIELNAPNSHYEECSYPTDGKVMCVCRDNWHGNSRNPWVSFDQ
NLDYKIGYICSGVFGDNPRPKDGTGSCGPVSADGANGVKGSYKYGNVWIGRTKSDSSRHGFEMIWDPNGWT
ETDSRFSMRQDVVAMTDRSGYSGSFVQHPELTGLDCMRPCFWVELIRGLPEEDAIWTSGSIISFCGVNGD TVD
WSWPDGAELPFTIDK

>M1
MSLLTEVETYVLSIVPSGPLKAEIAQRLEDVFAGKNTDLEVLMEWLKTRPILSPLTKGILGFVFTLTVP SERG
LQRRRFVQNALNGNDPNNMDKAVKLYRKLKREITFHGAKEIALSYSAGALASCMGLIYNRMGAVTTEVAFGL
VCATCEQIADSQHRSHRQMVT TTNPLIRHENRMVLASTTAKAMEQMAGSSEQAAEAMDIASQARQMVQAMRTI
GTHPSSSAGLKDDLLENLQAYQKRMGVQMQRFK

>M2
MSLLTEVET
PIRNEWGCRCNDSSDPLVIAANIIIEILHLILWILDRLFFKCIYRRFKYGLKRG PSTEGVPESMREEYRKEQQN
AVDVDDGHFVNIELE

>M3
MGVQMQRFK

>M4
MSLLTEVETYVLSIVPSGPLKAEIAQRLEDVFAGKNTDLE
AYQKRMGVQMQRFK

>M42
MSLQGRTPILR
PIRNEWGCRCNDSSDPLVIAANIIIEILHLILWILDRLFFKCIYRRFKYGLKRG PSTEGVPESMREEYRKEQQN
AVDVDDGHFVNIELE

>NS1
MDPNTVSSSFQVDCFLWHVRKR VADQELGDAPFLDRLRRDQKSLRGRGSTLGLDIETATRAGKQIVERILKEES
DEALKMTMASVPASRYLTDMTLEEMSRHWFMLMPKQKVAGPLCIRMDQAIMDKNIILKANFSVIFDRLET LIL
LRAFTEEGTIVGEISPLPSLPGHTDEDVKNAGVVLIGGLEWNNNTVRVSETLQRF AWRSSNENGRPPLTPKQK
RKMAGTIRSEV

>NS3
MDPNTVSSSFQVDCFLWHVRKR VADQELGDAPFLDRLRRDQKSLRGRGSTLGLDIETATRAGKQIVERILKEES
DEALKMTMASVPASRYLTDMTLEEMSRHWFMLMPKQKVAGPLCIRMDQAIMD
HTDEDVKNAGVVLIGGLEWNNNTVRVSETLQRF AWRSSNENGRPPLTPKQKRKMAGTIRSEV

>NEP
MDPNTVSSSFQ
DILMRMSKMQLGSSSEDNGIITQFESLKLYRDSLGEAVMRMGDLHSLQNRNGKWREQLGQKFEEIRW LIEEV
RHRLKITENSFEQITFMQALQLLLEVEQEIRTF SFQLI

>NEG8
MLFAQNYSLSSVCVSL LQSTILFLQTS DLIVPAIFRFCFGVSGGLPFSLLLLQANLCRVSETRTVLLFHSSP
PMRTPTAFLTSSSVCPGREGSGEISPTIVPSSVKALS NIRVSSRSKITLKFAFSMMFLSMIAWSILIQ RGPAT
FC
LGMSMNQCLDISSRVMSVR

Nucleotide sequences

>PB2
AGCGAAAGCAGGTCAATTATATTCAATATGGAAAGAATAAAAGAACTAAGGAATCTAATGTGCGAGTCTC
GCACTCGCGAGATACTCACAAAACCACCGTGGACCATATGGCCATAATCAAGAAGTACACATCAGGAAG
ACAGGAGAAGAACCCAGCACTTAGGATGAAATGGATGATGGCAATGAAATATCCAATTACAGCAGACAAG
AGGATAACGGAAATGATTCCTGAGAGAAATGAGCAGGGACAACTTTATGGAGTAAAATGAATGACGCCG

GATCAGACCGAGTGATGGTATCACCTCTGGCTGTGACATGGTGGAATAGGAATGGACCAGTGACAAGTAC
AGTTTCATTATCCAAAAATCTACAAAACCTTATTTTGAAGGTTGAAAGGTTAAACATGGAACCTTTGGC
CCTGTCCATTTTAGAAACCAAGTCAAATACGTCGAAGAGTTGACATAAATCCTGGGTCAATGCAGATCTCA
GTGCCAAAGAGGCACAGGATGTAATCATGGAAGTTGTTTTCCCTAACGAAGTGGGAGCCAGGATACTAAC
ATCGGAATCGCAACTAACGACAACCAAAGAGAAGAAAGAAGAACTCCAGGGTTGCAAAATTTCTCCTCTG
ATGGTGGCATAACATGTTGGAGAGAGAACTGGTCCGCAAAACGAGATTCCTCCCAGTGGCTGGTGGAACAA
GCAGTGTGTACATTGAAGTGTTCATTTGACCCAAGGAACATGCTGGGAACAGATGTACACTCCAGGAGG
GGAGGCGAGGAATGATGATGTTGATCAAAGCTTAATTATTGCTGCTAGAAACATAGTAAGAAGAGCCACA
GTaTCAGCAGATCCACTAGCATCTTTATTGGAGATGTGCCACAGCACGCAGATTGGTGGAATAAGGATGG
TAAACATCCTTAGGCAGAACCCAAACAGAAGAGCAAGCCGTGGATATTTGCAAGGCTGCAATGGGACTGAG
AATTAGCTCATCCTTCAGTTTTGGTGGATTACATTTAAGAGAACAAGCGGATCATCAGTCAAGAGAGAG
GAAGAGGTGCTTACGGGCAATCTTCAGACATTGAAGATAAGAGTGCATGAGGGATATGAAGAGTTCACAA
TGGTTGGGAGAAGAGCAACAGCTATACTCAGAAAAGCAACCAGGAGATTGATTTCAGCTGATAGTGAGTGG
GAGaGACGAACAGTCGATTGCCGAAGCAATAATTGTGGCCATGGTATTTTACAAAGAGGATTGTATGATA
AAAGCAGTTAGAGGTGACCTGAATTTCTGCAATAGGGCGAATCAGCGATTGAATCCCATGCACCAACTTT
TGAGACATTTTCAGAAGGATGCAAAGGtGCTCTTTCAAATTTGGGAATTGAATCCATCGACAAATGTGAT
GGGAATGATCGGGATATTGCCCGACATGACTCCAAGCACCGAGATGTCAATGAGAGGAGTGAGAATCAGC
AAAATGGGGGTAGATGAGTATTCCAGCGCGGAGAAGATAGTGGTGAGCATTGACCGTTTTTTTGAGAGTTA
GGGACCAACGTGGGAATGTACTACTGTCTCCCGAGGAGGTGAGTGAACACAGGGAACAGAGAAACTGAC
AATAACTTACTCATCGTCAATGATGTGGGAGATTAATGGTCCTGAATCAGTGTGGTCAATACCTATCAG
TGGATtATCAGAAACTGGGAAACTGTTAAATTCAGTGGTCCCAGAATCCTACAATGCTGTACAATAAAAA
TGGAATTTGAGCCATTTTCAGTCTTTAGTTCCAAAGGCCGTTAGAGGCCAATACAGTGGGTTTGTGAGAAC
TCTGTTCCAACAAATGAGGGATGTGCTTGGGACATTTGATACCGCTCAGATAATAAAACTTCTTCCCTTC
GCAGCCGCTCCACCAAAGCAAAGTGAATGCAGTTCTCCTCATTGACTATAAATGTGAGGGGATCAGGAA
TGAGAATACTTGTAAAGGGGCAATTCTCCAgTATTCAACTACAACAAGACCACTAAAAGACTCACAGTTCT
CGGAAAGGATGCTGGCCCTTAACTGAAGACCCAGATGAAGGCACAGCTGGAGTTGAGTCCGCAGTTCTG
AGAGGATTCCTCATCTGGGCAAAGAAGCAGGAGATATGGACCAGCATTAAAGCATAAATGAAGTGAAGCA
ACCTTGCGAAAGGAGAGAAGGCTAATGTGCTAATTGGGCAAGGAGACGTGGTGTGGTAATGAAACGGAA
ACGGAACCTTAGCATACTTACTGACAGCCAGACGCGACCAAAAGAATTCGGATGGCCATCAATTAGTGT
CGAATAGTTTAAAAACGACCTTGTCTTACT

>PB1

AGCGAAAGCAGGCAAACCATTTGAATGGATGTCAATCCGACTTTACTTTTCTTAAAGTGCCAGCACAAA
ATGCTATAAGCACAACTTTCCCTTATACTGGAGACCCTCCTTACAGCCATGGGACAGGAACAGGATACAC
CATGGATACTGTCAACAGGACACATCAGTACTCAGAAAGGGGAAGATGGACAACAAACACCGAAACTGGA
GCACCGCAACTCAACCCGATTGATGGGCCACTGCCAGAAGACAATGAACCAAGTGGTTATGCCCAAACAG
ATTGTGTATTGGAAGCAATGGCCTTCCTTGAGGAATCCCATCCTGGTATCTTTGAGACCTCGTGTCTTGA
AACGATGGAGGTTGTTTCAGCAAACACGAGTGGACAAGCTGACACAAGGCCGACAGACCTATGACTGGACT
CTAAATAGGAACCAGCCTGCTGCAACAGCATTGGCCAACACAATAGAAGTGTTCAGATCAAATGGCCCTCA
CGGCCAATGAATCtGGAAGGCTCATAGACTTCTTAAAGATGTAATGGAGTCAATGAACAAAGAAGAAAT
GGAGATCACAACTCATTTTCAGAGAAAGAGACGAGTGAGAGACAATATGACTAAGAAAATGGTGACACAG
AGAACAATAGGTAAAAGGAAGCAGAGATTGAACAAAAGGAGTTATCTAATTAGGGCATTaACCCTGAACA
CAATGACCAAAGATGCTGAGAGAGGGAAGCTAAAACGGAGAGCAATTGCAACCCCAGGGATGCAAATAAG
GGGGTTTGTATACTTTGTTGAGACACTAGCAAGGAGTATATGTGAGAACTTGAACAATCAGGATTGCCA
GTTGGAGGCAATGAGAAGAAAGCAAAGTTGGCAAATGTTGTAAGGAAGATGATGACCAATTCTCAGGACA
CTGAAATTTCTTTTACCATCACTGGAGATAACACCAAATGGAACGAAAATCAGAACCCTCGGATGTTTTT
GGCCATGATCACATATATAACCAGAAATCAGCCCGAATGGTTTCAGAAATGTTCTAAGTATTGCTCCAATA
ATGTTCTCAAACAAAATGGCGAGACTGGGAAAGGGGTACATGTTTGAGAGCAAGAGTATgAAAcTTAGAA
CTCAAATACCTGCAGAAATGCTAGCAAGCATCGATTGAAATACTTCAATGATTCAACTAGAAAGAAGAT
TGAAAAAATCCGGCCGCTCTTAATAGATGGGACTGCATCTTGGACAAAAGAGACACACCAAGACTACTT
ACTGGTGGGATGGTCTTCAATCTTCTGATGATTTTGCTCTGATTGTGAATGCACCCAATCATGAAGGGAT
TCAAGCCGGAGTCAACAGGTTTTATCGAACCTGTAAGCTACTTGGAATTAATATGAGCAAGAAAAAGTCT
TACATAAACAGAACAGGTACATTTGAATTCACAAGTTTTTTCTATCGTTATGGGTTTGTGGCAATTTCA
GCATGGAGCTTCCCAGCTTTGGGGTGTCTGGGATCAACGAGTCTGCGGACATGAGTATTGGAGTTACTGT
CATCAAAAACAATATGATAAACAATGATCTTGGTCCAGCAACCGCTCAAATGGCCCTTCAGCTGTTTCATC
AAAGATTACAGGTACACGTACCGGTGCCATAGAGGTGACACACAAATACAAACCCGAAGATCATTTGAAA
TAAAGAAACTGTGGGAGCAAACCCATTCCAAAGCTGGACTGCTGGTCTCCGACGGAGGCCCAAATTTATA
CAACATTAGAAATCTCCACATTCCTGAAGTCTGCTTGAATGGGAATTAATGGATGAGGATTACCAGGGG
CGTTTATGCAACCCACTGAACCCATTTGTCAACCATAAAGACATTGAATCAGTGAACAATGCAGTGATAA
TGCCAGCACATGGTCCAGCCAAAACATGGAGTATGATGCTGTTGCAACAACACAtTCCTGGATCCCCAA
AAGAAATCGATCCATCTTGAATACAAGCCAAAGAGGAATACTTGAAGATGAACAAATGTACCAAAAGTGC
TGCAACTTATTTGAAAAATCTTCCCCAGCAGTTTCATACAGAAGACAGTCGGGATATCCAGTATGGTGG
AGGCTATGGTTTTCCAGAGCCCGAATTGATGACGCAATTTGATTTTCAATCTGGAAGGATAAAGAAAGAGGA
GTTCACTGAGATCATGAAGATCTGTTCCACCATTGAAGAGCTCAGACGGCAAAAATAGTGAATTTAGCTT
GTCTTCATGAAAAAATGCCTTGTCTTACT

>PA

AGCGAAAGCAGGTACTGATTCAAAATGGAAGATTTTGTGCGACAATGCTTCAATCCGATGATTGTGCGAGC
TTGCGGAAAAGGCAATGAAAGAGTATGGAGAGGACCTGAAAATCGAAACAAACAAATTTGCAGCAATATG
CACTCACTTGGGAAGTGTGCTTCATGTATTTCAGATTTTCACTTCATCGATGAGCAGAGTCATCAATAGTC
GTAGAACTTGGCGATCCAAATGCACCTTTTGAAGCACAGATTTGAAATAATCGAGGGAAGAGATCGCACAA
TAGCCTGGACAGTAATAAACAGTATTTGCAACACTACAGGGGCTGAGAAACCAAAGTTTCTACCAGATTT
GTATGATTACAAGAAGAATAGATTCATCGAAATTGGAGTAACAAGGAGAGAAGTTCACATATACTATCTG
GAAAAGGCCAATAAAATTAATCTGAGAAGACACACATCCACATTTTCTCATTCACTGGGGAGGAAATGG
CCACAAAGGCCGACTACACTCTCGATGAAGAAAGCAGGGCTAGGATCAAAACCAGGCTATTCCACCATAAG
ACAAGAAATGGCTAGCAGAGGCCCTCTGGGATTCTTTCTCGTCAGTCCGAGAGAGGGCGAAGAGACAATTGAA
GAAAGATTTGAAATCACAGGAACAATGCGCAAGCTTGCCGACCAAAGTCTCCCGCCAAACTTCTCCAGCC
TTGAAAATTTTAGAGCCTATGTGGATGGATTCTGAACCGAACGGCTACATTGAGGGCAAGCTTTTCTCAAAAT
GTCCAAAGAAGTAAATGCTAGAAATTGAACCTTTTTTGAATCAACACCACGACCACCTTAGACTTCCGGAT
GGGCCTCCCTGTTCTCAGCGGTCCAAATTCCTGCTGATGGATGCCTTAAATTAAGCATTGAGGACCCAA
GTCATGAGGGAGAGGGGATACCGCTATATGATGCAATCAAATGCATGAGAACATTCTTTGGATGGAAGGA
ACCAATGTTGTAAACCACACGAAAAGGGAATAAATCCAAATTATCTTCTGTCTGATGGAAGCAAGTACTG
GCAGAACTGCAGGACATTGAGAATGAGGAGAAAATTCCAAGGACTAAAAATATGAAGAAAAACGAGTCAGT
TAAAGTGGGCACCTGGTGAGAACATGGCACCAGAAAAGGTAGACTTTGACGATTGTAAAGATGTAGGCGA
TTTGAAGCAATATGATAGTGATGAACCAGAATTGAGGTGCTTGCAAGTTGGATTGAGAATGAGTTCAAC
AAGGCATGTGAAGTACCGATTCAAGCTGGATAGAGCTCGATGAGATTGGAGAAGATGCGGCTCCAATTG
AACACATTGCAAGCATGAGAAGGAATTATTTACAGCAGAGGTGTCTCATTGCAGAGCCACAGAATACAT
AATGAAGGGGTGTACATCAATACTGCCTTGCTTAATGCATCCTGTGCAGCAATGGATGATTTCCAATTg
ATTTCCAATGATAAGCAAGTGTAGAACTAAGGAGGGAAGGCGAAAGACCAATTTGTACGGTTTCATCATAA
AAGGAAGATCCCACTTAAGGAATGACACCGATGTGGTAAACTTTGTGAGCATGGAGTTTTCCCTCACTGA
CCCAAGACTTGAACCACACAAATGGGAGAAGTACTGTGTTCTTGAGGTAGGAGATATGCTTCTAAGAAGT
GCCATAGGCCATGTGTCAAGGCCTATGTTCTTGTATGTGAGGACAAATGGAACCTCAAAAATTAATAATGA
AATGGGGGATGGAATGAGGCGTTGCCTCCTTCAGTCACTTCAACAAATCGAGAGTATGATTGAAGCTGA
GTCCTCTGTCAAGGAGAAAGACATGACCAAGAGTTCTTTGAAAACAAATCAGAAACATGGCCCGTTGGA
GAGTCCCCCAAAGGAGTGAGGAAGGTTCCATTGGGAAGGTCTGCAGAACTTTATTGGCAAAGTCGGTAT
TCAACAGCTTGTATGCATCTCCACAACCTGGAAGGATTTTCAGCTGAATCAAGAAAACCTGCTTCTTATCGT
TCAGGCTCTTAGGGACAACCTGGAACCTGGGACCTTTGATCTTGGGGGGCTATATGAAGCAATTGAGGAG
TGCTTATTAATGATCCCTGGGTTTTGCTTAATGCTTCTTGGTTCAACTCCTTCTCACACATGCATTGA
GATAGTTGTGGCAATGCTACTATTTGCTATCCATACTGTCCAAAAAGTACCTTGTCTTCTACT

>HA

AGCAAAAGCAGGGGAAAATAAAAAACAACCAAAATGAAGGCTtttTgTACTgGTCTGTTATATGCATTTGT
AGCTACAGATGCAGACACAATATGTATAGGCTACCATGCGAACAACCTCAACCGACACTGTTGACACAATA
TTCGAGAAGAATGTGGCAGTGACACATTCTGTAAACCTGCTCGAAGACAGACACAACGGGAAACTATGTA
AATTAAGGAATAGCCCCACTACAATTGGGGAAATGTAACATCACCGGATGGCTCTTGGGAAATCCAGA
ATGCGACTCACTGCTTCCAGCGAGATCATGGTCTTACATTGTAGAAACACCAAACTCTGAGAATGGAGCA
TGTATCCAGGAGATTTTCATCGACTATGAGGAACCTGAGGAGCAATTGAGCTCAGTATCATATTAGAAA
GATTTCGAAATATTTCCCAAGGAAAGTTTCATGGCCCCAACACACATTCAACGGAGTAACAGTATCATGCTC
CCATAGGGGAAAAAGCAGTTTTTACAGAAATTTGCTATGGCTGACGAAGAAGGGGGATTATACCCAAAG
CTGACCAATTCTATGTGAACAATAAAGGGAAAGAAGTCCTTGTACTATGGGGTGTTCATCACCCGCTcA
GCAGTGATGAGCAACAGAGTCTCTATAGTAATGGAATGCTTATGTCTCTGTAGCGTCTTCAAATTATAA
CAGGAGATTACCCCGGAAATAGCTGCAAGGCCCAAAGTAAAGATCAACATGGGAGGATGAACATTATAC
TGGACCTTGCTAGAACCCGGAGACACAATAATATTTGAGGCAACTGGTAATCTAATAGCACCATGGTATG
CTTTCGCACTGAGTAGAGGGTTTGTAGTCCGGCATCATCACTCAAACGCGTCAATGCATGAGTGTAACAC
GAAGTGTCAAACACCCCAAGGATCTATAAACAGCAATCTCCCTTTCCAGAATATACACCCAGTCACAATA
GGAGAGTGCCCAAAATATGTCAGGAGTACCAATTGAGGATGGTTACAGGACTAAGAAACATCCCATCCA
TTCAATACAGAGGTCTATTTGGAGCCATTGCTGGTTTTATTGAGGGGGGATGGACTGGAATGATAGATGG
ATGGTATGGTTATCATCATCAGAATGAACAGGGATCAGGCTATGCAGCGGATCAAAAAAGCACACAGAAT
GCCATTAACGGGATTACAAACAAGGTGAACCTGTTATCGAGAAAATGAACACTCAATTCACAGCTGTGG
GTAAAGAATTCAACAACCTTAGAAAAAAGGATGGAATAATTAATAAAAAAGTTGATGATGGGTTCTGGA
CATTTGGACATATAATGCAGAATTGTTAGTTCTACTGGAAAATGAAAGAACCTTTGGATTTCCATGACTTA
AATGTGAAGAATCTGTACGAGAAAGTAAAAAGCCAATTAAAGAATAATGCCAAAGAAATCGGAAATGGGT
GTTTTGAGTTCTACCACAAGTGTGACAATGAATGCATGGAAGTGTAAGAAATGGGACTTATGATTATCC
AAAATATTCAGAAGAATCAAAGTTGAACAGGGAAAAGATAGATGGAGTGAAATTGGAATCAATGGGGGTG
TATCAGATTCTGGCGATCTACTCAACTGTCGCCAGTTCACTGGTGCTTTTGGTCTCCCTGGGGGCAATCA
GTTTCTGGATGTGTTCTAATGGGTCTTTGCAGTGCAGAATATGCATCTGAGATTAGGATTTTCAGAAATAT
AAGGAAAAACACCTTGTCTTCTACT

>NP

AGCAAAAGCAGGGTAGATAATCACTCACAGAGTGACATCGAAATCATGGCGACCAAAGGCACCAAACGAT
CTTACGAACAGATGGAGACTGATGGAGAACGCCAGAATGCCACTGAAATCAGAGCATCTGTGCGGAAAAAT
GATTGaTGGAAATTGGACGATTCTACATCCAAATGTGCACCGAAGCTTAAACTCAGTGATTATGAGGGACGG
CTGATTTCAGAACAGCTTAACAATAGAGAGAATGGTGCTCTCTGCTTTTGACGAGAGGAGGAATAAATATC
TAGAAGAACATCCAGTGCGGGGAAAGATCCTAAGAAAAGCTGGAGGACCTATATACAGGAGAGTAGATGG
AAAGTGGAGGAGAGAACTCATCCTTTATGACAAAGAAGAAATAAGACGAATCTGGCGCCAAGCTAATAAT
GGTGACGATGCAACGGCTGGTCTGACTCACATGATGATCTGGCACTCCAATTTGAATGATGCAACTTACC
AGAGGACAAGAGCTCTTGTTTCGCACAGGAATGGATCCCAGGATGTGCTCACTGATGCAGGGTTCAACCCCT
CCCTAGGAGGTCTGGGGCCGCAGGTGCTGCAGTCAAAGGAGTTGGAACAATGGTGATGGAATTGATCAGA
ATGATCAAACGTGGGATCAATGATCGGAACCTTCTGGAGGGGTGAGAATGGACGGGAGAACAGGATTGCTT
ATGAAAGAATGTGCAACATTCTCAAAGGGAATTTCAAACAGCTGCACAAAGAAaCAATGGTGGATCAAGT
GAGAGAGAGCCGGAATCCAGGAAATGCTGAGTTCGAAGATCTCATCTTTtTAGCACGGTCTGCACTCATA
TTGAGAGGGTCAGTTGCTCACAAGTCCTGCCTGCCTGCCTGTGTGTATGGATctGCCGTAGCCAGTGGAT
ACGACTTTGAAAGAGAGGGGATACTCTCTAGTCGGAATAGACCCTTTCAGACTGCTTCAAAACAGCCAAGT
ATACAGCCTAATCAGACCAAATGAGAATCCAGCACACAAGAGTCAACTGGTGTGGATGGCATGCCATTCT
GCTGCATTTGAAGATCTAAGAGTATCAAGCTTCATCAGAGGGACGAAAGTGGTCCCAAGAGGGGAAGCTTT
CCACTAGAGGAGTTCAAATTGCTTCCAATGAAAACATGGAGACTATGGAATCAAGTACCCTTGAAGTGA
AAGCAGATACTGGGCCATAAGGACCAGAAGTGGAGGGAACACCAATCAACAGAGGGCTTCTCGGGCCAA
ATCAGCATAACAACCTACGTTCTCAGTACAGAGAAATCTCCCTTTTGACAGACCAACCATTATGGCAGCAT
TCACTGGGAATACAGAGGGGAGAACATCTGACATGAGAACCAGAAATCATAAGGCTGATGGAAAGTGAAG
ACCAGAAGATGTGTCTTTCCAGGGGCGGGGAGTCTTCGAGCTCTCGGACGAAAAGGCAaCGAGCCCGATC
GTGCCCTCCTTTGACATGAGTAATGAAGGATCTTATTTCTTCGGAGACAATGCAGAGGAGTACGACAATT
AAAGAAAAATACCCCTTGTTTCTACT

>NA

AGCGAAAGCAGGAGTTTAAATGAATCCAAACCAGAAAAATAATAACCATTGGaTCgATCTGTATGGTAGTC
GGAATAATTAGCCTAATATTGCAAATAGGAAATATAATCTCAATATGGATTAGCCATTCAATTCAAACCG
GAAATCAAACCATACCTGGAATATGCAACCAAGGCAGCATTACCTATAAAGTTGTTGCTGGGCAGGACTC
AAGTTTCAGTGATATTAAACCGCAATTTCATCTCTTTGTCCCATCCGTGGGTGGGTGATACACAGCAAAGAC
AATGGCATAAGAATTGGTTCCAAAGGAGACGTTTTTGTGATAAGAGAGCCTTTTATTTTCATGTTCTCACT
TGGAATGCAGGACCTTTTTTCTGACTCAAGGCGCCTTACTGAATGACAAGCATTCAAGGGGGACCTTTAA
GGACAGAAGCCCTTATAGGGCCTTAATGAGCTGCCCTGTGCGTGAAGCTCCGTCCCCGTACAATTCAGG
TTTGAATCGGTTGCTTGGTCAGCAAGTGCATGTCATGATGGAaTGGGCTGGCTAACAATCGGAATTTCTG
GTCCAGATGATGGAGCAGTGGCTGTATTAAAATACAACgGCATAATAACTGAAACCATAAAAAAGTTGGAG
GAAGAATATATTGAGAACACAAGAGTCTGAATGTACCTGTGTAAATGGTTTCATGTTTTACCATAATGACC
GATGGCCCAAGTGATGGGCTGGCCTCGTACAAAATTTTCAAGATCGAGAAGGGGAAGGTTACTAAATCAA
TAGAGTTGAATGCACCTAATTTCTCACTACGAGGAATGTTCTGTACCCTGATACCGGCAAAGTGATGTG
TGTGTGCAGAGACAATTGGCACGGTTCGAACCGACCATGGGTGTCTTCGACCAAAACCTAGATTATAAA
ATAGGATACATCTGCAGTGGGGTTTTCGGTGACAACCCGCTCCCAAGATGGAACAGGCAGCTGTGGCC
CAGTGTCTGCTGATGGAGCAAACGGAGTAAAGGGATTTTCATATAAGTATGGtAATGGTGTTTGGATAGG
AAGGACTAAAAGTGACAGTTCCAGACATGGGTTTGAGATGATTTGGGATCCTAATGGATGGACAGAGACT
GATAGTAGGTTCTCTATGAGACAAGATGTTGTGGCAATgACTgATCGGTGAGGGTACAGCGGAAGTTTCG
TTCAACATCCTGAGCTAACAGGGCTAGACTGTATGAGGCCTTGCTTCTGGGTTGAATTAATCAGGGGGCT
ACCTGAGGAGGACGCAATCTGGACTAGTGGGAGCATATTTCTTTTTGTGGTGTGAATgGTGATACTGTA
GATTGGTCTTGGCCAGACGGTGTGAGcTcCGGTTACCATTGACAAGTAGTTTGTTCAAAAAACTCCTT
GTTTCTACT

>M

AGCAAAAGCAGGTAGATATTGAAAGATGAGTCTTCTAACCGAGGTCGAAACGTACGTTCTCTCTATCGTC
CCGTACAGGCCCCCTCAAAGCCGAGATCGCACAGAGACTTGAAGATGTCTTTGCAGGGAAGAACACCGATC
TTGAGGTTCTCATGGAATGGCTAAAGACAAGACCAATCCTGTCACCTCTGACTAAGGGGATTTTAGGATT
TGTGTTACGCTCACCCTGCCCAGTGAGCGGGGACTGCAGCGTAGACGCTTTGTCCAAAATGCTCTTAAT
GGGAACGGAGATCCAAATAACATGGACAAAGCAGTTAAACTGTATAGGAAGCTTAAGAGGGAGATAACAT
TCCATGGGGCCAAAGAAATAGCACTCAGTTATTCTGCTGGTGCATTTGCCAGTTGTATGGGCCTCATATA
CAACAGGATGGGGGCTGTGACCACTGAAGTGGCATTGTCGCTGGTATGCGCAACCTGTGAACAGATTGCT
GACTCCCAGCATCGGTCTCATAGGCAAATGGTGACAACAACCAATCCACTAATCAGACATGAGAACAGAA
TGGTTCTAGCCAGCACTACAGCTAAGGCTATGGAGCAAATGGCTGGATCGAGTGAGCAAGCAGCAGAGGC
CATGGATATTGCTAGTCAGGCCAGGCAAATGGTGCAGGCGATGAGAACCATTGGGACTCATCTAGCTCC
AGTGCTGGTCTAAAAGATGATCTTCTTGAAAATTTGCAGGCCTATCAGAAACGAATGGGGGTGCAGATGC
AACGATTCAAGTGATCCTCTCGTCATTGCAGCAAATATCATTGAGATCTTGCACTTGATATTGTGGATTCT
TTGATCGTCTTTTTTTCAAATGCATTTATCGTCGCTTTAAATACGGTTTGAAAAGAGGGCCTTCTACGGA
AGGAGTGCCAGAGTCTATGAGGGAAGAATATCGAAAGGAACAGCAGAATGCTGTGGATGTTGACGATGGT
CATTTTGTCAACATAGAGCTGGAGTAAAAAACTA

>NS
AGCAAAAGCAGGGTGACAAAGACATAATGGATCCAAACACTGTGTCAAGCTTTCAGGTAGATTGCTTTCT
TTGGCATGTCCGCAAAAGAGTTGCAGACCAAGAACTAGGTGATGCCCCATTCTTGATCGGCTTCGCCGA
GATCAGAAAGTCCCTAAGAGGAAGAGGCAGCACTCTCGGTCTGGACATCGAAACAGCCACCCGTGCTGGAA
AGCAAATAGTGGAGCGGATTCTGAAGGAAGAATCTGATGAGGCACTCAAAATGACCATGGCCTCTGTACC
TGCATCGCGCTACCTAACTGACATGACTCTTGAGGAAATGTCAAGGCACTGGTTCATGCTCATGCCCAAG
CAGAAAGTGGCAGGCCCTCTTTGTATCAGAATGGACCAGGCGATCATGGATAAGAACATCATACTGAAAG
CGAACTTCAGTGTGATTTTTGACCGGCTGGAGACTCTAATATTACTAAGGGCCTTCACCGAAGAGGGGAC
AATTGTTGGCGAAATTTCACTACTGCCCTCTCTCCAGGACATACTGATGAGGATGTCAAAAATGCAGTT
GGGGTCCTCATCGGAGGACTTGAATGGAATAATAACACAGTTCGAGTCTCTGAAACTCTACAGAGATTCTG
CTTGAGAGAAGCAGTAATGAGAATGGGAGACCTCCACTCACTCCAAAACAGAAACGGAAAATGGCGGGAAC
AATTAGGTCAGAAGTTTGAAGAAATAAGATGGTTGATTGAAGAAGTGAGACACAGACTGAAGATAACAGA
GAATAGTTTTGAGCAAATAACATTTATGCAAGCCTTACAACATTTGCTTGAAGTGGAGCAAGAGATAAGA
ACTTTCTCGTTTCAGCTTATTTAATAATAAAAAACACCCTTGTTTCTACT

mRNA interactome capture of A549

Shuai Ni and Bernd Fischer

German Cancer Research Center (DKFZ),
Heidelberg, Germany
n.shuai@dkfz.de

September 26, 2017

Contents

1	Read data	1
1.1	load packages and datasets	1
1.2	Read peptide files	1
2	Map peptides on proteins	2
2.1	Consider only peptides matched uniquely to the longest protein for downstream analysis	4
3	Differential protein abundance	5
3.1	Quantitative method (moderated t-test)	5
3.2	semi-quantitative method	6
3.3	Overview of the number of detected peptides with intensity values in noCL and CL over 3 replicates.	6
3.4	semi-quantitative comparison from the aggregated peptides for each protein. Comparing peptide counts between noCL and CL in 3 replicates)	7
4	Comparison to known RBDs	18
4.1	Find the corresponding match by ENSEMBL gene IDs for single match peptides in human cell lines	18
5	Gene set enrichment analysis	19
6	Sequence features	20

1 Read data

1.1 load packages and datasets

```
library(RBDmap)
library(Biobase)
library(Biostrings)

## Warning: package 'S4Vectors' was built under R version 3.3.3
## Warning: package 'IRanges' was built under R version 3.3.3
## Warning: package 'XVector' was built under R version 3.3.3

library(limma)

## Warning: package 'limma' was built under R version 3.3.3

library(ggplot2)
library(knitr)
```

```
library(hwriter)
library(RColorBrewer)
library(org.Hs.eg.db)
load(file.path('inputdata', 'Annotation', 'ENSG2category.rda'))
load(file.path('inputdata', 'Annotation', 'ENSGannotation.rda'))
source(file.path('R', 'doGSEA.R'))
source(file.path('R', 'hwriteSidebar.R'))
###import database of all proteins from uniprot
load(file.path('inputdata', 'Annotation', 'ProtFeatures.rda'))
load(file.path('inputdata', 'Annotation', 'Index.rda'))
```

1.2 Read peptide files

```
Peptides = read.table(file.path("inputdata", "MSfiles", pep.filename),
                      sep="\t", comment.char="", quote="", header=TRUE,
                      stringsAsFactors=FALSE)
```

Remove contaminant and decoy peptide sequences.

```
Peptides=Peptides[Peptides$Potential.contaminant==' ' & Peptides$Reverse==' ',]
```

Remove unnecessary columns from MaxQuant peptide table.

```
Peptides = Peptides[,c(grep("Sequence", colnames(Peptides)),
                      grep("Intensity.", colnames(Peptides)))]
```

A function to map back the protein IDs to Ensembl gene IDs.

```
MapProt2Ensg=function(protein_list){
  ###MapProt2Ensg maps the peptide-mapping proteins to genes
  ###and return a list of mapped genes
  lapply(protein_list,function(x)
    { y=unname(ProtFeatures$GeneName[x])
      unique(y[!is.na(y)])})
}
```

A function to return a match status of a list of mapped ENSEMBL gene id from Uniprot ID.

```
match_summary=function(mapped_protein_list,mapped_ensg_list){
  ##returns a vector of match types from list of mapped Ensg_id
  ll=listLen(mapped_ensg_list)
  k=ifelse(ll>1, 'multimatch',
          ifelse(ll==0, 'no_match', 'single_match'))
  ####Find the corresponding longest isoform
  ####for the whole list of ensembl gene IDs
  Longest_prot=data.frame(ensg_id=ProtFeatures$GeneName,
                          len_prot=width(ProtFeatures$ProtSeq))
  Longest_prot=
    Longest_prot[with(Longest_prot, order(ensg_id, -len_prot)),]
  Longest_prot=Longest_prot[!duplicated(Longest_prot$ensg_id),]
  Longest_prot=Longest_prot[!is.na(Longest_prot$ensg_id),]
  ####Find peptides which mapped to the longest isoforms
  single_match_ensgid=sapply(mapped_ensg_list,
    function(x) ifelse(length(x)==1, x, NA))
  longest_protein=names(Longest_prot$ensg_id)[
    match(single_match_ensgid, Longest_prot$ensg_id)]
  k[maply( function(x,y) {x%in%y},
           longest_protein, mapped_protein_list)]= 'longest_match'
  data.frame(matches=k, prot_id=longest_protein)
```

```
}
```

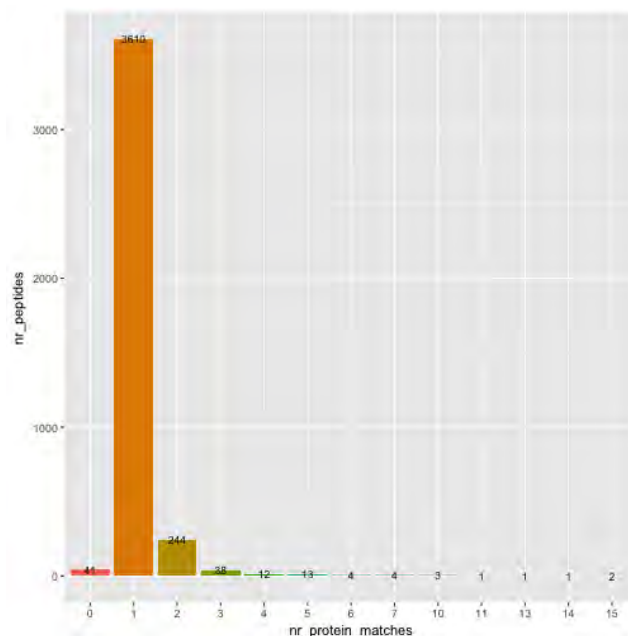
2 Map peptides on proteins

Map the peptides onto the protein sequence database.

```
Pep2Uniprot = mapPeptides(PeptideSet=Peptides$Sequence,
                          ProtFeatures$ProtSeq, Index=Index, verbose=FALSE)
```

Summary of number of proteins where single peptides are mapped to.

```
df=as.data.frame(table(listLen(Pep2Uniprot)))
colnames(df)=c('nr_protein_matches', 'nr_peptides')
p=ggplot(df, aes(x=nr_protein_matches, y=nr_peptides,
                 fill=factor(nr_protein_matches)))
p+geom_bar(stat='identity')+theme(legend.position = "none") +
geom_text(aes(label = nr_peptides), size = 3)
```

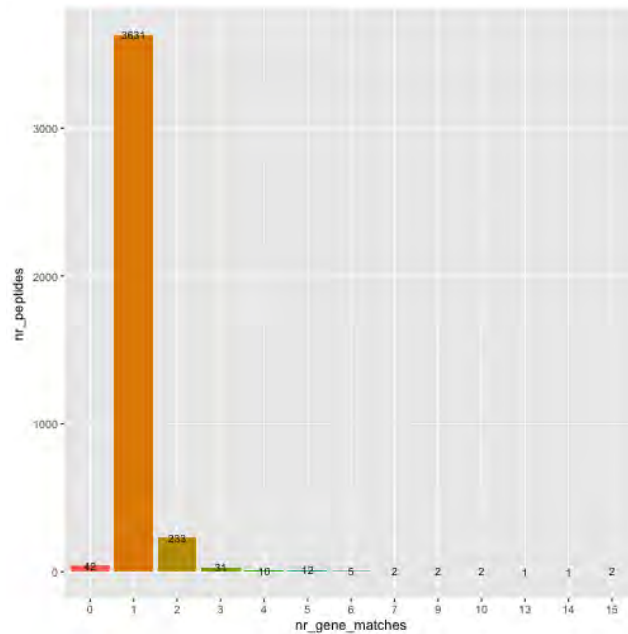


Map protein id to ENSG id.

```
Pep2Ensg=MapProt2Ensg(Pep2Uniprot)
```

Summary of number of gene where single peptides are mapped to.

```
df=as.data.frame(table(listLen(Pep2Ensg)))
colnames(df)=c('nr_gene_matches', 'nr_peptides')
p=ggplot(df, aes(x=nr_gene_matches, y=nr_peptides,
                 fill=factor(nr_gene_matches)))
p+geom_bar(stat='identity')+theme(legend.position = "none") +
  geom_text(aes(label = nr_peptides), size = 3)
```

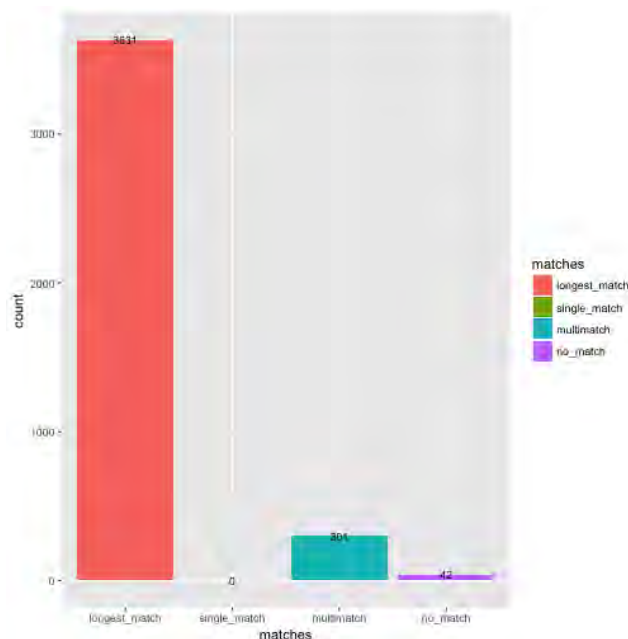



Find the mapping status of each peptide.

```
pep_match=match_summary(Pep2Uniprot,Pep2Ensg)
```

Draw statistics for 4 peptide matching levels (single match, no match, multiple match and not on the longest protein isoform).

```
df=as.data.frame(table(factor(pep_match$matches,
  levels=c('longest_match','single_match','multimatch','no_match'))))
colnames(df)=c('matches','count')
p=ggplot(df, aes(x=matches,y=count,
  fill=matches))+geom_bar(stat='identity')
p + geom_text(aes(label = count), size = 3, hjust = 0.5)
```



2.1 Consider only peptides matched uniquely to the longest protein for downstream analysis

```

Peptides$prot_id = pep_match$prot_id
Peptides$matches = pep_match$matches
Peptides$ENSG = ProtFeatures$GeneName[as.character(Peptides$prot_id)]
Peptides$Symbol = ProtFeatures$Symbol[as.character(Peptides$prot_id)]
Peptides=Peptides[pep_match$matches=='longest_match',]
head(Peptides)

##              Sequence Intensity.L Intensity.H Intensity.A Intensity.L.A
## 1      AAAAAAAAAAGAAGGR          0          0      7285200          0
## 2      AAAAAAALQAK      51853000      656190000      125200000      8801400
## 3 AAAMLPTVGEYGYGPESELSQASAATR          0      3673000      3673000          0
## 4      AACQCVEPIK          0      4244400      1690000          0
## 5      AADPPAENSSAPEAEQGGAE          0      60060000      18763000          0
## 6      AAEGVSAADMAK          0      140630000      37192000          0
## Intensity.H.A Intensity.B Intensity.L.B Intensity.H.B Intensity.C Intensity.L.C
## 1          0      4151900          0          0          0          0
## 2      116400000      287250000      25909000      261340000      295600000      17142000
## 3      3673000          0          0          0          0          0
## 4      1690000      2554400          0      2554400          0          0
## 5      18763000          0          0          0      41296000          0
## 6      37192000      64247000          0      64247000      39188000          0
## Intensity.H.C prot_id      matches      ENSG Symbol
## 1          0 Q86U42 longest_match ENSG00000100836 PABPN1
## 2      278460000 P36578 longest_match ENSG00000174444 RPL4
## 3          0 Q9BQ04 longest_match ENSG00000173914 RBM4B
## 4          0 Q99575 longest_match ENSG00000104356 POP1
## 5      41296000 P67809 longest_match ENSG00000065978 YBX1
## 6      39188000 Q9NW13 longest_match ENSG00000106344 RBM28

PeptideIntensity = as.data.frame(Peptides[,c("Intensity.L.A","Intensity.H.A",
      "Intensity.L.B","Intensity.H.B",
      "Intensity.L.C","Intensity.H.C")])
colnames(PeptideIntensity) =
  c( "A_light", 'A_heavy', 'B_light', 'B_heavy', 'C_light', 'C_heavy')
PeptideIntensity[PeptideIntensity < 1] = NA
PeptideIntensity = log2(PeptideIntensity)
Peptides=cbind(Peptides[c(1, 13,14,15,16)], PeptideIntensity)

```

3 Differential protein abundance

3.1 Quantitative method (moderated t-test)

unpaired intensity values from peptides are removed for enrichment over a non-crosslinking control by a moderated t-test for proteins, the left intensity values are aggregated to protein level by their mapping origin.

```

# rm_unpaired=function(vec, pos) {
#   for (i in pos) {
#     if (sum(is.na(vec[c(i,i+1)]))!=0 ) {
#       vec[i]=NA
#       vec[i+1]=NA
#     }
#   }
#   return(vec)
# }
#
# Intensity_values=as.data.frame(
#   t(apply(PeptideIntensity[,1:6], 1, function(x) {

```

```
# rm_unpaired(x,c(1,3,5))
# })))
```

Paired intensity values of peptides are aggregated to protein level by their mapping origin.

```
ProteinIntensity=aggregate(Peptides[,6:11],
                           by=list(Peptides$ENSG),mean, na.rm=TRUE)
colnames(ProteinIntensity)[1]='ENSG'

ProteinIntensity$Symbol = Peptides$Symbol[match(ProteinIntensity$ENSG, Peptides$ENSG)]
ProteinIntensity$prot_id= Peptides$prot_id[match(ProteinIntensity$ENSG, Peptides$ENSG)]
ProteinIntensity=ProteinIntensity[,c(9,2:7, 1, 8 )]
head(ProteinIntensity)
```

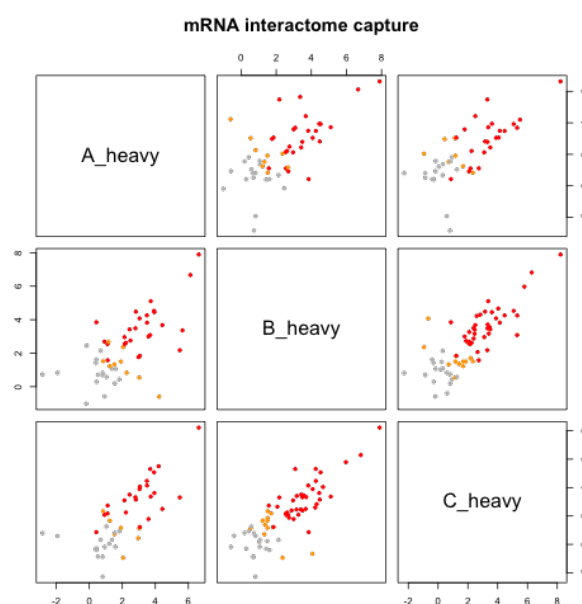
##	prot_id	A_light	A_heavy	B_light	B_heavy	C_light	C_heavy	ENSG	Symbol
## 1	Q9Y4W2	NaN	21.69290	NaN	21.37042	NaN	21.10930	ENSG000000001497	LAS1L
## 2	Q92900	NaN	21.96673	NaN	22.37417	NaN	22.34096	ENSG000000005007	UPF1
## 3	Q9H6R0	NaN	NaN	NaN	20.50603	NaN	19.90626	ENSG000000005100	DHX33
## 4	P60866	NaN	22.75326	NaN	22.48381	NaN	22.94878	ENSG000000008988	RPS20
## 5	O75534	NaN	23.39307	NaN	23.75295	NaN	23.40701	ENSG000000009307	CSDE1
## 6	Q9ULX6	NaN	NaN	NaN	NaN	22.86797	NaN	ENSG000000011243	AKAP8L

The proteins are tested for enrichment over a non-crosslinking control by a moderated t-test as implemented in the Bioconductor-package limma. p-values are corrected for multiple testing by the method of Benjamini-Hochberg.

```
fit=eBayes(lmFit(ProteinIntensity[,c(3,5,7)]-ProteinIntensity[,c(2,4,6)]))
fit$p.adj=p.adjust(fit$p.value,method="BH")
ProteinIntensity$limma_padj=fit$p.adj
```

Scatter plots of peptides comparing the two replicates are plotted.

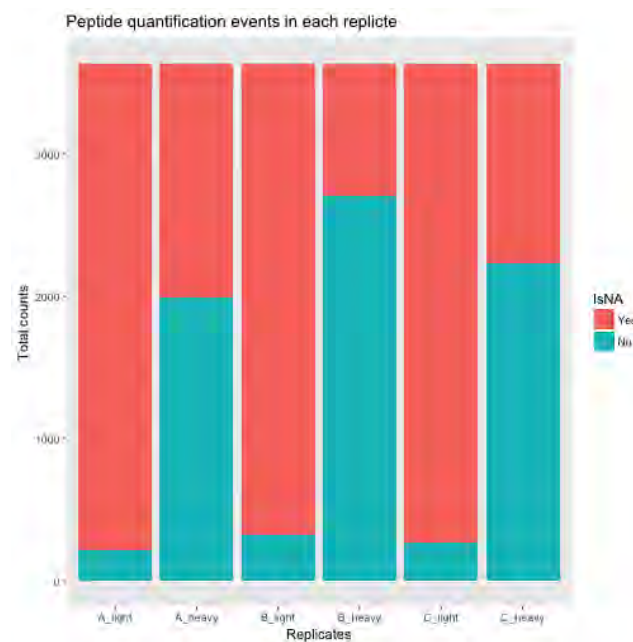
```
col = ifelse(fit$p.adj <= 0.1, "orange", "gray")
col[which(fit$p.adj <= 0.01)] = "red"
col[which(fit$p.adj <= 0.1 & fit$coefficients < 0)] = "cyan"
col[which(fit$p.adj <= 0.01 & fit$coefficients < 0)] = "blue"
pairs(ProteinIntensity[,c(3,5,7)]-ProteinIntensity[,c(2,4,6)],
      col=col, pch=20,
      main="mRNA interactome capture")
```



3.2 semi-quantitative method

3.3 Overview of the number of detected peptides with intensity values in noCL and CL over 3 replicates.

```
peptide_counts=t(apply(PeptideIntensity[,1:6],2,
                      function(x) table(is.na(x))))
replicates=rownames(peptide_counts,2)
peptide_counts=as.data.frame(as.vector(peptide_counts))
peptide_counts=cbind(peptide_counts,rep(c('No','Yes'),
                                         each=length(replicates)))
colnames(peptide_counts)=c('count','replicates','IsNA')
peptide_counts$IsNA=factor(peptide_counts$IsNA,levels=c('Yes','No'))
peptide_counts$replicates = factor(peptide_counts$replicates,
                                   levels=c('A_light','A_heavy','B_light',
                                             'B_heavy','C_light','C_heavy'))
cp_protein_abundency=
  ggplot(data=peptide_counts,aes(x=replicates,y=count,fill=IsNA))
cp_protein_abundency+geom_bar(stat='identity',position='stack')+
  xlab("Replicates") + ylab("Total counts") +
  ggtitle("Peptide quantification events in each replicate")
```



3.4 semi-quantitative comparison from the aggregated peptides for each protein. Comparing peptide counts between noCL and CL in 3 replicates)

```
na_count=aggregate(Peptides[,6:11],
                   by=list(Peptides$prot_id),function(x) sum(!is.na(x)))
na_count=na_count[,c(2,3,4,5,6,7,1)]
names(na_count)[c(7)]=c('prot_id')
na_count[na_count$prot_id=='P15121',]

##      A_light A_heavy B_light B_heavy C_light C_heavy prot_id
## 96         0        2        0        0        0        0 P15121
```

If there are no values in all three light channels but all values in heavy channels, We say it is semi-quantified (definition 1)

```

n.noCL = rowSums(na_count[,c(1,3,5)] > 0)
n.CL = rowSums(na_count[,c(2,4,6)] > 0)
na_count$semi = (n.CL - n.noCL > 1)
na_count

```

##	A_light	A_heavy	B_light	B_heavy	C_light	C_heavy	prot_id	semi
## 1	0	10	0	10	0	9	AOAV96	TRUE
## 2	0	1	0	0	0	0	A6NC98	FALSE
## 3	1	0	0	0	0	0	A6NF34	FALSE
## 4	0	1	0	0	0	1	A6NHQ2	TRUE
## 5	0	0	0	0	0	0	A9YQ3	FALSE
## 6	0	1	0	0	0	0	HOYAA0	FALSE
## 7	0	0	1	0	0	0	I3L1E1	FALSE
## 8	0	0	0	0	1	1	000410	FALSE
## 9	0	3	0	5	0	5	000411	TRUE
## 10	0	6	0	2	0	4	000425	TRUE
## 11	0	8	0	9	0	9	000541	TRUE
## 12	1	2	1	4	0	5	000566	FALSE
## 13	0	9	0	16	0	11	000567	TRUE
## 14	0	4	0	5	0	2	000571	TRUE
## 15	0	0	0	0	0	0	014497	FALSE
## 16	0	0	0	0	0	1	014933	FALSE
## 17	0	5	1	5	0	5	014979	TRUE
## 18	0	6	0	10	0	7	015213	TRUE
## 19	0	6	1	9	0	9	015226	TRUE
## 20	0	2	0	9	0	5	015371	TRUE
## 21	0	0	0	3	0	0	015381	FALSE
## 22	0	0	0	4	0	3	043143	TRUE
## 23	0	9	0	11	0	10	043159	TRUE
## 24	0	0	0	0	0	0	043242	FALSE
## 25	1	0	0	0	0	0	043264	FALSE
## 26	0	11	0	13	2	11	043390	TRUE
## 27	0	0	0	0	0	0	043766	FALSE
## 28	0	0	0	1	0	0	043818	FALSE
## 29	0	0	1	1	0	0	060282	FALSE
## 30	0	15	2	28	0	23	060287	TRUE
## 31	1	1	0	0	0	0	060499	FALSE
## 32	0	16	2	17	0	14	060506	TRUE
## 33	1	1	0	0	0	0	060763	FALSE
## 34	0	2	0	4	0	3	060832	TRUE
## 35	0	0	0	2	0	1	060841	TRUE
## 36	0	1	0	2	0	1	060869	TRUE
## 37	0	0	0	2	0	0	075153	FALSE
## 38	0	1	0	0	0	0	075179	FALSE
## 39	0	1	0	1	0	1	075494	TRUE
## 40	0	1	0	1	0	0	075533	TRUE
## 41	0	17	0	21	0	10	075534	TRUE
## 42	0	1	0	2	0	1	075616	TRUE
## 43	0	2	0	4	0	3	075683	TRUE
## 44	0	17	2	44	0	30	075691	TRUE
## 45	0	0	0	2	0	1	075821	TRUE
## 46	0	1	0	0	0	0	075822	FALSE
## 47	0	9	5	13	2	14	076021	FALSE
## 48	0	0	0	1	0	0	076094	FALSE
## 49	0	0	0	0	1	1	094956	FALSE
## 50	0	0	0	1	0	0	095453	FALSE
## 51	0	0	0	3	0	2	095478	TRUE
## 52	0	0	0	0	0	0	095490	FALSE

## 53	0	4	0	7	0	2	095758	TRUE
## 54	0	0	0	0	1	0	095759	FALSE
## 55	0	0	0	1	0	1	095777	TRUE
## 56	0	4	0	4	0	0	095793	TRUE
## 57	0	0	0	3	0	1	095900	TRUE
## 58	0	0	0	1	0	0	095985	FALSE
## 59	0	3	0	0	0	0	P00352	FALSE
## 60	2	3	2	3	1	1	P01040	FALSE
## 61	2	2	2	2	2	2	P02792	FALSE
## 62	0	0	0	1	0	0	P04080	FALSE
## 63	0	1	0	1	0	0	P04083	TRUE
## 64	0	0	0	1	0	0	P04406	FALSE
## 65	0	9	0	13	1	11	P05455	TRUE
## 66	0	2	0	0	0	0	P05783	FALSE
## 67	0	0	0	0	0	0	P05813	FALSE
## 68	0	4	0	0	0	1	P06733	TRUE
## 69	0	6	4	9	0	9	P06748	TRUE
## 70	0	4	0	4	1	2	P07355	TRUE
## 71	0	1	0	1	0	0	P07437	TRUE
## 72	0	2	0	1	0	1	P07737	TRUE
## 73	0	2	0	0	0	0	P07900	FALSE
## 74	0	4	0	4	0	4	P07910	TRUE
## 75	1	0	0	0	0	0	P08235	FALSE
## 76	0	1	0	0	0	0	P08238	FALSE
## 77	1	3	0	2	1	2	P08708	FALSE
## 78	1	5	1	5	0	3	P08865	FALSE
## 79	1	1	0	0	1	1	P08962	FALSE
## 80	0	1	0	3	0	1	P09001	TRUE
## 81	0	0	0	1	0	0	P09429	FALSE
## 82	1	0	0	0	1	0	P09467	FALSE
## 83	0	0	0	0	0	1	P09493	FALSE
## 84	0	7	0	5	2	6	P09651	TRUE
## 85	0	1	0	1	0	0	P11142	TRUE
## 86	0	0	0	0	0	1	P11387	FALSE
## 87	1	1	1	4	1	2	P11388	FALSE
## 88	0	1	0	3	0	3	P11940	TRUE
## 89	0	7	0	9	0	10	P12956	TRUE
## 90	0	9	0	7	0	5	P13010	TRUE
## 91	0	7	0	7	0	6	P13639	TRUE
## 92	0	0	0	0	0	0	P14136	FALSE
## 93	0	1	0	1	0	1	P14618	TRUE
## 94	0	8	0	7	0	5	P14866	TRUE
## 95	0	0	1	7	5	4	P14923	FALSE
## 96	0	2	0	0	0	0	P15121	FALSE
## 97	0	13	0	15	1	14	P15880	TRUE
## 98	2	2	0	9	5	7	P15924	FALSE
## 99	0	3	0	6	0	5	P16989	TRUE
## 100	0	0	0	1	0	0	P17022	FALSE
## 101	0	0	0	2	0	1	P17480	TRUE
## 102	0	12	0	14	0	17	P17844	TRUE
## 103	3	9	4	9	4	6	P18124	FALSE
## 104	0	2	0	4	0	5	P18621	TRUE
## 105	2	28	3	31	1	35	P19338	FALSE
## 106	0	1	0	1	0	0	P20290	TRUE
## 107	0	13	0	9	0	2	P21333	TRUE
## 108	0	1	0	0	0	0	P21980	FALSE
## 109	0	2	0	5	0	5	P22087	TRUE
## 110	0	14	0	15	0	14	P22626	TRUE

## 111	0	15	0	17	1	12	P23246	TRUE
## 112	1	11	1	12	0	9	P23396	FALSE
## 113	0	1	0	0	0	0	P23528	FALSE
## 114	0	5	0	7	0	5	P23588	TRUE
## 115	0	1	0	2	0	2	P25398	TRUE
## 116	0	0	0	0	0	0	P25788	FALSE
## 117	0	0	0	2	0	1	P26196	TRUE
## 118	0	0	1	0	0	0	P26358	FALSE
## 119	0	7	0	6	0	4	P26368	TRUE
## 120	0	1	0	5	1	3	P26373	TRUE
## 121	1	7	0	7	0	6	P26599	TRUE
## 122	0	0	0	2	0	1	P26641	TRUE
## 123	0	0	0	2	0	2	P27635	TRUE
## 124	0	0	0	1	0	1	P27695	TRUE
## 125	0	4	0	5	0	3	P27816	TRUE
## 126	0	3	0	2	0	1	P29401	TRUE
## 127	0	3	0	4	1	3	P30086	TRUE
## 128	0	0	0	0	0	0	P31350	FALSE
## 129	1	2	0	1	0	0	P31483	FALSE
## 130	0	7	0	8	0	6	P31942	TRUE
## 131	1	10	0	8	0	8	P31943	TRUE
## 132	0	1	0	1	0	0	P33240	TRUE
## 133	0	4	0	3	0	3	P35268	TRUE
## 134	0	6	0	6	0	5	P35637	TRUE
## 135	0	1	1	1	1	1	P35908	FALSE
## 136	13	17	19	21	17	20	P36578	FALSE
## 137	0	1	0	0	0	0	P37802	FALSE
## 138	0	1	0	1	0	1	P38159	TRUE
## 139	0	5	1	7	1	5	P39023	FALSE
## 140	3	3	4	4	1	4	P40429	FALSE
## 141	0	0	0	1	0	0	P42166	FALSE
## 142	0	4	0	5	0	6	P42285	TRUE
## 143	0	2	0	0	0	0	P42330	FALSE
## 144	0	7	0	8	0	9	P42696	TRUE
## 145	3	41	3	49	4	45	P42704	FALSE
## 146	1	1	1	1	0	1	P42766	FALSE
## 147	5	10	9	10	9	8	P43243	FALSE
## 148	0	0	0	0	0	0	P45984	FALSE
## 149	0	30	5	57	0	57	P46013	TRUE
## 150	1	10	2	17	2	17	P46087	FALSE
## 151	0	1	1	2	0	2	P46776	TRUE
## 152	0	8	0	10	0	9	P46777	TRUE
## 153	0	1	0	1	0	1	P46778	TRUE
## 154	1	1	1	2	0	2	P46779	FALSE
## 155	1	3	0	2	0	2	P46781	TRUE
## 156	0	2	0	5	1	4	P46782	TRUE
## 157	0	0	1	1	0	0	P47869	FALSE
## 158	0	1	1	1	0	1	P47914	TRUE
## 159	0	1	0	2	0	1	P48634	TRUE
## 160	0	2	0	0	0	1	P49207	TRUE
## 161	0	1	0	2	0	0	P50579	TRUE
## 162	0	3	0	3	1	3	P50914	TRUE
## 163	0	1	0	4	0	1	P51114	TRUE
## 164	0	1	0	1	0	0	P51116	TRUE
## 165	0	0	0	1	0	0	P51648	FALSE
## 166	0	0	0	0	0	1	P51795	FALSE
## 167	0	5	0	7	0	4	P51991	TRUE
## 168	1	19	0	13	0	13	P52272	TRUE

## 227	1	0	0	0	1	0	P78371	FALSE
## 228	0	1	0	11	0	11	P78527	TRUE
## 229	2	2	2	2	2	2	P81605	FALSE
## 230	0	0	0	1	0	0	P82933	FALSE
## 231	0	1	0	2	0	2	P83731	TRUE
## 232	0	2	2	3	1	3	P84098	FALSE
## 233	0	3	0	3	0	3	P84103	TRUE
## 234	1	2	0	0	0	0	P98175	FALSE
## 235	0	4	0	4	0	1	P98179	TRUE
## 236	0	9	0	15	0	5	Q00341	TRUE
## 237	0	12	5	11	2	13	Q00839	FALSE
## 238	0	4	1	5	1	2	Q01085	FALSE
## 239	0	2	0	3	0	3	Q01130	TRUE
## 240	0	0	0	2	0	0	Q01469	FALSE
## 241	0	5	1	14	0	9	Q01780	TRUE
## 242	0	2	0	2	0	2	Q01844	TRUE
## 243	1	1	2	4	4	6	Q02413	FALSE
## 244	1	5	0	8	0	8	Q02543	TRUE
## 245	2	5	6	7	6	8	Q02878	FALSE
## 246	0	0	0	0	0	0	Q02880	FALSE
## 247	1	13	0	26	0	22	Q03701	TRUE
## 248	0	1	0	2	0	0	Q04637	TRUE
## 249	0	0	2	3	2	3	Q04837	FALSE
## 250	0	0	0	0	0	1	Q06265	FALSE
## 251	2	2	2	2	1	2	Q07020	FALSE
## 252	1	7	2	11	1	6	Q07065	FALSE
## 253	0	1	0	0	0	1	Q07666	TRUE
## 254	0	6	0	5	0	6	Q07955	TRUE
## 255	2	21	6	25	7	28	Q08211	FALSE
## 256	0	0	0	2	1	1	Q08554	FALSE
## 257	0	0	0	1	0	0	Q08945	FALSE
## 258	0	6	0	10	0	9	Q08J23	TRUE
## 259	0	17	0	8	0	0	Q09666	TRUE
## 260	0	0	0	0	1	1	Q12770	FALSE
## 261	0	0	0	1	0	0	Q12797	FALSE
## 262	0	2	0	6	0	1	Q12849	TRUE
## 263	0	1	0	0	0	0	Q12874	FALSE
## 264	0	5	0	8	0	6	Q12905	TRUE
## 265	0	10	1	14	3	16	Q12906	FALSE
## 266	0	8	0	5	0	6	Q13148	TRUE
## 267	0	5	1	5	1	3	Q13151	FALSE
## 268	0	5	0	8	0	8	Q13206	TRUE
## 269	0	0	0	1	0	0	Q13239	FALSE
## 270	0	5	0	2	0	3	Q13242	TRUE
## 271	0	1	0	1	0	1	Q13243	TRUE
## 272	0	2	0	2	0	2	Q13247	TRUE
## 273	0	7	0	7	0	5	Q13283	TRUE
## 274	0	3	1	6	0	1	Q13310	TRUE
## 275	0	3	1	2	0	2	Q13501	TRUE
## 276	0	0	0	0	0	0	Q13572	FALSE
## 277	0	2	0	0	0	0	Q13595	FALSE
## 278	0	2	0	4	0	3	Q13601	TRUE
## 279	0	11	1	16	0	15	Q13823	TRUE
## 280	0	0	0	1	0	0	Q13835	FALSE
## 281	1	0	0	0	0	0	Q13867	FALSE
## 282	1	2	0	4	0	1	Q13868	TRUE
## 283	0	4	0	5	0	6	Q13895	TRUE
## 284	0	0	1	1	0	0	Q14005	FALSE

## 285	1	3	0	4	0	3	Q14011	TRUE
## 286	0	6	0	5	0	7	Q14103	TRUE
## 287	1	3	0	3	0	1	Q14137	TRUE
## 288	0	2	0	2	0	1	Q14146	TRUE
## 289	1	11	1	17	0	13	Q14152	FALSE
## 290	0	2	0	3	0	0	Q14157	TRUE
## 291	0	0	0	4	0	3	Q14258	TRUE
## 292	0	3	1	4	1	3	Q14444	FALSE
## 293	0	2	0	1	0	0	Q14498	TRUE
## 294	1	6	0	7	0	5	Q14671	TRUE
## 295	0	8	0	8	0	7	Q14684	TRUE
## 296	63	64	72	76	71	76	Q14690	FALSE
## 297	0	20	1	25	1	27	Q14692	FALSE
## 298	0	2	1	3	0	3	Q14694	TRUE
## 299	0	2	0	3	0	2	Q14978	TRUE
## 300	0	4	0	6	0	7	Q15050	TRUE
## 301	0	7	1	9	1	5	Q15056	FALSE
## 302	0	2	0	2	0	1	Q15061	TRUE
## 303	0	11	1	10	0	1	Q15149	TRUE
## 304	1	17	0	14	0	14	Q15233	TRUE
## 305	0	1	0	2	0	2	Q15269	TRUE
## 306	0	0	0	0	0	0	Q15326	FALSE
## 307	0	1	0	0	0	0	Q15361	FALSE
## 308	0	7	0	10	0	8	Q15365	TRUE
## 309	0	3	0	5	0	3	Q15366	TRUE
## 310	0	13	0	14	0	12	Q15397	TRUE
## 311	0	1	0	0	0	1	Q15427	TRUE
## 312	0	3	0	2	0	1	Q15434	TRUE
## 313	0	2	0	0	0	0	Q15637	FALSE
## 314	0	9	0	7	1	7	Q15717	TRUE
## 315	0	0	0	0	0	0	Q16533	FALSE
## 316	0	4	0	2	0	2	Q16629	TRUE
## 317	0	1	0	0	0	0	Q16650	FALSE
## 318	0	0	0	1	0	0	Q16678	FALSE
## 319	1	0	0	0	0	0	Q16890	FALSE
## 320	0	0	0	1	0	2	Q1ED39	TRUE
## 321	0	1	0	0	0	0	Q1KMD3	FALSE
## 322	0	0	0	0	1	0	Q3MJ16	FALSE
## 323	0	3	0	4	0	4	Q4G0J3	TRUE
## 324	0	0	0	1	0	0	Q52LJ0	FALSE
## 325	0	0	0	1	0	0	Q53EP0	FALSE
## 326	0	3	0	4	0	2	Q5BKZ1	TRUE
## 327	0	1	0	3	0	0	Q5C9Z4	TRUE
## 328	0	11	1	22	0	17	Q5JTH9	TRUE
## 329	1	7	0	6	0	7	Q5QJE6	TRUE
## 330	0	1	0	4	1	5	Q5SY16	TRUE
## 331	0	1	0	1	0	1	Q5T280	TRUE
## 332	0	2	0	2	0	2	Q5T3I0	TRUE
## 333	1	2	1	3	1	2	Q5T653	FALSE
## 334	0	3	0	7	0	3	Q5T6F2	TRUE
## 335	0	0	0	2	2	2	Q5T749	FALSE
## 336	0	0	0	0	0	0	Q5TDH0	FALSE
## 337	0	0	1	0	0	0	Q5TID7	FALSE
## 338	0	0	0	4	0	1	Q5VYS8	TRUE
## 339	0	0	0	0	1	0	Q5XPI4	FALSE
## 340	0	0	0	1	0	1	Q68D10	TRUE
## 341	0	3	0	3	0	2	Q6DKI1	TRUE
## 342	0	0	0	5	0	1	Q6P158	TRUE

## 343	1	1	0	0	0	0	Q6P1N9	FALSE
## 344	3	6	4	4	3	3	Q6P2Q9	FALSE
## 345	0	0	0	0	0	2	Q6P5R6	FALSE
## 346	0	4	2	5	0	3	Q6PK04	TRUE
## 347	1	6	1	8	0	4	Q6PKG0	FALSE
## 348	0	1	0	0	0	0	Q6Y7W6	FALSE
## 349	0	0	0	1	0	1	Q6ZN17	TRUE
## 350	0	0	0	0	0	0	Q719H9	FALSE
## 351	0	1	0	1	0	0	Q71RC2	TRUE
## 352	0	4	0	7	0	3	Q76FK4	TRUE
## 353	0	11	0	19	0	18	Q7KZF4	TRUE
## 354	0	1	1	5	1	3	Q7L2E3	FALSE
## 355	1	0	0	0	0	0	Q7L2Z9	FALSE
## 356	0	2	0	3	0	3	Q7Z2T5	TRUE
## 357	0	11	0	15	0	10	Q7Z2W4	TRUE
## 358	0	2	0	2	0	0	Q7Z417	TRUE
## 359	0	0	0	2	0	0	Q7Z478	FALSE
## 360	0	0	0	1	0	2	Q86TB9	TRUE
## 361	0	8	0	11	0	12	Q86U38	TRUE
## 362	0	0	1	2	0	0	Q86U42	FALSE
## 363	0	0	0	2	0	0	Q86UE4	FALSE
## 364	0	4	0	4	0	2	Q86V81	TRUE
## 365	0	1	0	1	0	3	Q86W50	TRUE
## 366	0	0	0	2	0	1	Q86WX3	TRUE
## 367	0	0	0	0	1	0	Q86YP4	FALSE
## 368	0	0	0	0	0	0	Q8IUA7	FALSE
## 369	0	0	0	0	1	0	Q8IVG5	FALSE
## 370	0	17	0	25	1	23	Q8IWA0	TRUE
## 371	0	2	0	3	0	2	Q8IWR0	TRUE
## 372	0	1	0	1	0	1	Q8IXM3	TRUE
## 373	0	0	0	2	0	1	Q8IXT5	TRUE
## 374	0	1	0	7	0	6	Q8IY37	TRUE
## 375	0	1	0	0	0	0	Q8IY67	FALSE
## 376	5	13	12	19	5	22	Q8IY81	FALSE
## 377	0	0	0	0	0	0	Q8N159	FALSE
## 378	0	0	0	1	0	0	Q8N1G4	FALSE
## 379	0	3	0	3	0	2	Q8N3Z6	TRUE
## 380	0	0	0	2	0	1	Q8N5C6	TRUE
## 381	0	1	0	1	0	0	Q8N5P1	TRUE
## 382	0	0	2	4	0	5	Q8N8A6	FALSE
## 383	1	3	2	2	1	1	Q8N983	FALSE
## 384	0	0	0	2	0	1	Q8N9T8	TRUE
## 385	0	0	0	0	0	0	Q8NBF1	FALSE
## 386	0	8	0	8	0	6	Q8NC51	TRUE
## 387	0	2	0	2	0	2	Q8NCA5	TRUE
## 388	0	3	0	2	0	1	Q8ND56	TRUE
## 389	0	0	0	0	0	0	Q8ND76	FALSE
## 390	0	0	0	3	0	1	Q8NDF8	TRUE
## 391	0	2	0	2	0	0	Q8NDT2	TRUE
## 392	0	0	0	1	0	0	Q8NDV7	FALSE
## 393	0	0	0	5	0	3	Q8NE71	TRUE
## 394	0	2	0	2	0	0	Q8NEF9	TRUE
## 395	0	0	0	3	0	1	Q8NEJ9	TRUE
## 396	0	5	1	8	0	5	Q8NHQ9	TRUE
## 397	0	1	0	2	0	2	Q8NI36	TRUE
## 398	1	0	0	0	0	0	Q8NI51	FALSE
## 399	0	1	0	2	0	0	Q8TB72	TRUE
## 400	0	1	0	1	0	0	Q8TBF4	TRUE

## 401	1	17	1	26	1	26	Q8TDD1	FALSE
## 402	0	2	0	7	0	7	Q8TDN6	TRUE
## 403	0	12	0	14	0	13	Q8TED0	TRUE
## 404	1	0	0	0	0	0	Q8WTQ4	FALSE
## 405	0	8	0	13	1	12	Q8WTT2	TRUE
## 406	0	0	0	1	0	0	Q8WU90	FALSE
## 407	0	4	0	3	0	0	Q8WVV9	TRUE
## 408	0	0	0	1	0	0	Q8WWM7	FALSE
## 409	1	6	0	5	0	1	Q8WXF1	TRUE
## 410	0	0	0	1	0	1	Q92499	TRUE
## 411	0	1	0	1	0	1	Q92523	TRUE
## 412	0	5	0	4	0	3	Q92804	TRUE
## 413	0	0	0	0	0	0	Q92824	FALSE
## 414	0	14	0	11	0	10	Q92841	TRUE
## 415	0	5	0	6	0	3	Q92879	TRUE
## 416	0	6	0	10	0	7	Q92900	TRUE
## 417	0	17	0	12	1	10	Q92945	TRUE
## 418	0	0	0	1	0	1	Q969S3	TRUE
## 419	0	1	0	5	0	3	Q969X6	TRUE
## 420	0	2	0	4	0	6	Q969Z0	TRUE
## 421	0	7	1	2	0	1	Q96AE4	TRUE
## 422	0	3	0	4	0	1	Q96AG4	TRUE
## 423	0	3	0	3	0	1	Q96EP5	TRUE
## 424	0	3	0	5	0	3	Q96EU6	TRUE
## 425	0	2	0	8	0	2	Q96EY7	TRUE
## 426	2	15	1	25	1	19	Q96GQ7	FALSE
## 427	0	18	1	14	0	10	Q96I24	TRUE
## 428	0	0	0	0	0	1	Q96I51	FALSE
## 429	0	0	0	0	0	0	Q96JB1	FALSE
## 430	1	0	0	0	0	0	Q96KP4	FALSE
## 431	0	1	0	5	0	5	Q96KR1	TRUE
## 432	1	9	0	13	0	9	Q96P11	TRUE
## 433	1	12	1	7	0	9	Q96PK6	FALSE
## 434	0	0	0	0	0	0	Q96PX9	FALSE
## 435	0	1	1	2	0	0	Q96QR8	FALSE
## 436	1	1	0	0	0	0	Q96S79	FALSE
## 437	0	0	1	1	1	1	Q96S97	FALSE
## 438	1	1	1	1	1	1	Q96SC8	FALSE
## 439	0	0	0	1	0	0	Q96SU4	FALSE
## 440	0	4	0	5	0	2	Q96T37	TRUE
## 441	0	1	0	0	0	1	Q96T58	TRUE
## 442	1	6	1	10	1	7	Q99575	FALSE
## 443	0	1	0	3	0	4	Q99729	TRUE
## 444	0	4	0	8	0	6	Q99848	TRUE
## 445	0	3	0	2	0	0	Q9BQ04	TRUE
## 446	0	1	0	5	0	6	Q9BQ39	TRUE
## 447	0	0	0	1	0	1	Q9BQ48	TRUE
## 448	0	0	0	2	0	2	Q9BQ67	TRUE
## 449	0	0	0	0	0	0	Q9BQ69	FALSE
## 450	0	1	0	1	0	1	Q9BQC6	TRUE
## 451	0	19	0	34	0	32	Q9BQG0	TRUE
## 452	0	0	0	0	0	1	Q9BQQ7	FALSE
## 453	0	0	0	1	0	0	Q9BRJ2	FALSE
## 454	0	1	0	2	0	1	Q9BRT6	TRUE
## 455	0	0	0	2	0	2	Q9BRU9	TRUE
## 456	0	2	0	0	0	0	Q9BRZ2	FALSE
## 457	0	2	0	7	0	7	Q9BSC4	TRUE
## 458	0	1	0	0	0	1	Q9BSH4	TRUE

## 459	0	1	0	0	0	0	Q9BTD8	FALSE
## 460	0	4	1	7	1	7	Q9BUJ2	FALSE
## 461	0	0	0	3	0	3	Q9BVI4	TRUE
## 462	0	13	2	16	1	12	Q9BVJ6	FALSE
## 463	0	5	0	9	0	10	Q9BVP2	TRUE
## 464	0	5	0	4	0	3	Q9BWF3	TRUE
## 465	0	0	0	0	0	1	Q9BXJ9	FALSE
## 466	0	0	0	0	0	0	Q9BXS1	FALSE
## 467	0	4	0	4	0	3	Q9BYD1	TRUE
## 468	0	0	0	2	0	1	Q9BYD2	TRUE
## 469	0	0	0	1	0	2	Q9BYD3	TRUE
## 470	1	5	0	5	0	4	Q9BYG3	TRUE
## 471	1	3	2	4	0	3	Q9BZE1	FALSE
## 472	1	11	0	17	0	16	Q9BZE4	TRUE
## 473	0	0	0	0	0	1	Q9COH6	FALSE
## 474	0	1	0	4	0	3	Q9GZR2	TRUE
## 475	0	13	0	17	0	13	Q9GZR7	TRUE
## 476	0	3	0	3	0	3	Q9GZT3	TRUE
## 477	0	2	0	10	0	8	Q9H0A0	TRUE
## 478	0	2	0	3	0	6	Q9H0D6	TRUE
## 479	1	0	0	0	1	0	Q9H0F6	FALSE
## 480	0	4	0	3	0	3	Q9H0S4	TRUE
## 481	0	2	0	5	0	2	Q9H2U1	TRUE
## 482	0	1	0	1	0	2	Q9H2Y7	TRUE
## 483	0	1	0	0	0	1	Q9H361	TRUE
## 484	0	4	0	4	0	6	Q9H501	TRUE
## 485	0	0	0	6	0	4	Q9H583	TRUE
## 486	0	0	0	0	0	0	Q9H5I5	FALSE
## 487	0	0	0	2	0	1	Q9H6R0	TRUE
## 488	0	5	0	6	0	7	Q9H6R4	TRUE
## 489	0	3	0	5	0	4	Q9H7B2	TRUE
## 490	0	9	1	10	0	5	Q9H8H0	TRUE
## 491	0	8	0	14	0	17	Q9H8H2	TRUE
## 492	0	2	0	3	0	3	Q9H9L3	TRUE
## 493	0	2	0	5	0	4	Q9H9Y2	TRUE
## 494	0	0	0	2	0	2	Q9HAV4	TRUE
## 495	0	0	0	2	0	0	Q9HC36	FALSE
## 496	0	4	0	5	0	4	Q9HCE1	TRUE
## 497	0	1	0	2	0	0	Q9NQZ2	TRUE
## 498	2	29	1	36	2	36	Q9NR30	FALSE
## 499	0	1	0	0	0	0	Q9NRJ5	FALSE
## 500	0	1	0	0	0	0	Q9NRX1	FALSE
## 501	0	1	0	1	0	1	Q9NSD9	TRUE
## 502	0	1	0	4	0	2	Q9NUD5	TRUE
## 503	0	2	0	2	0	0	Q9NUL3	TRUE
## 504	0	4	0	4	0	3	Q9NUQ6	TRUE
## 505	1	8	1	13	0	11	Q9NV06	FALSE
## 506	0	1	0	5	0	3	Q9NV31	TRUE
## 507	0	6	0	6	0	4	Q9NVN8	TRUE
## 508	0	6	1	10	2	10	Q9NVP1	FALSE
## 509	0	2	0	3	0	0	Q9NVS2	TRUE
## 510	0	1	0	7	0	4	Q9NVU7	TRUE
## 511	0	0	0	0	0	1	Q9NVV5	FALSE
## 512	0	18	2	22	0	17	Q9NW13	TRUE
## 513	0	2	0	0	0	0	Q9NW64	FALSE
## 514	0	0	0	1	0	0	Q9NX05	FALSE
## 515	0	0	0	0	0	1	Q9NX20	FALSE
## 516	0	2	0	1	0	2	Q9NX24	TRUE

## 517	0	0	1	4	0	6	Q9NXF1	FALSE
## 518	0	0	1	0	0	0	Q9NXG2	FALSE
## 519	0	3	0	4	0	5	Q9NY12	TRUE
## 520	0	0	1	0	0	0	Q9NY28	FALSE
## 521	0	3	0	5	0	4	Q9NY61	TRUE
## 522	0	12	2	17	1	13	Q9NY93	FALSE
## 523	0	1	0	1	0	0	Q9NYK5	TRUE
## 524	0	4	0	3	0	3	Q9NYY8	TRUE
## 525	1	16	4	26	2	16	Q9NZB2	FALSE
## 526	0	10	0	11	0	8	Q9NZI8	TRUE
## 527	0	3	1	2	1	5	Q9NZM5	FALSE
## 528	1	1	0	2	0	2	Q9P015	TRUE
## 529	1	6	0	2	1	4	Q9P031	FALSE
## 530	0	0	0	0	1	1	Q9P0S2	FALSE
## 531	0	1	0	5	0	4	Q9P275	TRUE
## 532	2	3	8	12	5	8	Q9P2E9	FALSE
## 533	0	3	0	1	0	0	Q9P2K5	TRUE
## 534	0	0	0	0	0	0	Q9UBE0	FALSE
## 535	0	0	0	0	0	0	Q9UBI4	FALSE
## 536	0	0	0	1	0	1	Q9UBU9	TRUE
## 537	0	0	0	1	0	0	Q9UG63	FALSE
## 538	0	0	0	0	0	0	Q9UGI8	FALSE
## 539	0	1	0	0	0	0	Q9UGP8	FALSE
## 540	0	1	0	1	0	0	Q9UGR2	TRUE
## 541	1	1	1	1	1	1	Q9UHD1	FALSE
## 542	0	0	0	0	0	0	Q9UHD8	FALSE
## 543	0	4	1	4	1	4	Q9UHX1	FALSE
## 544	0	1	0	0	0	0	Q9UK53	FALSE
## 545	0	6	0	4	0	5	Q9UKM9	TRUE
## 546	1	1	1	3	1	3	Q9ULW3	FALSE
## 547	0	0	0	0	1	0	Q9ULX6	FALSE
## 548	0	1	0	0	0	0	Q9UMY1	FALSE
## 549	0	2	0	1	0	1	Q9UN86	TRUE
## 550	0	0	1	1	0	0	Q9UNL2	FALSE
## 551	1	4	2	5	1	5	Q9UNQ2	FALSE
## 552	0	4	0	10	0	9	Q9UNX4	TRUE
## 553	0	10	1	11	1	10	Q9UQ80	FALSE
## 554	0	1	0	1	0	0	Q9Y221	TRUE
## 555	1	0	0	0	0	0	Q9Y272	FALSE
## 556	0	0	0	2	0	0	Q9Y295	FALSE
## 557	0	0	1	0	0	0	Q9Y2D0	FALSE
## 558	0	0	0	0	0	0	Q9Y2J8	FALSE
## 559	0	0	0	2	0	0	Q9Y2K5	FALSE
## 560	0	0	0	1	0	1	Q9Y2L1	TRUE
## 561	0	1	0	5	0	3	Q9Y2R4	TRUE
## 562	0	4	0	0	0	0	Q9Y2W1	FALSE
## 563	0	9	0	15	0	9	Q9Y2X3	TRUE
## 564	0	0	0	1	0	2	Q9Y2Z4	TRUE
## 565	0	0	0	1	0	1	Q9Y324	TRUE
## 566	0	0	0	1	1	2	Q9Y333	FALSE
## 567	0	3	0	4	0	4	Q9Y3A2	TRUE
## 568	0	0	0	0	0	2	Q9Y3A4	FALSE
## 569	1	0	0	0	0	0	Q9Y3C0	FALSE
## 570	1	3	1	4	2	4	Q9Y3C1	FALSE
## 571	0	2	0	7	0	5	Q9Y3T9	TRUE
## 572	0	1	0	0	0	1	Q9Y3U8	TRUE
## 573	0	1	0	0	0	0	Q9Y3Y2	FALSE
## 574	0	25	2	24	2	23	Q9Y4C8	FALSE

```
## 575      0      1      0      3      0      2 Q9Y4W2 TRUE
## 576      0      0      0      1      0      0 Q9Y4Y9 FALSE
## 577      1      3      2     10      2      8 Q9Y520 FALSE
## 578      0      1      0      1      0      1 Q9Y5B9 TRUE
## 579      0      2      1      7      0      2 Q9Y5J1 TRUE
## 580      0      2      0      1      0      1 Q9Y6G3 TRUE
## 581      0      1      0      0      0      0 Q9Y6M1 FALSE
## 582      0      0      0      1      0      2 Q9Y6V7 TRUE

ProteinIntensity$semi_quant=na_count$semi[match(ProteinIntensity$prot_id, na_count$prot_id)]
ProteinIntensity[ProteinIntensity$prot_id=='P15121',]

##      prot_id A_light  A_heavy B_light B_heavy C_light C_heavy      ENSG Symbol
## 64  P15121      NaN 20.58149      NaN      NaN      NaN      NaN ENSG00000085662 AKR1B1
##      limma_padj semi_quant
## 64           NA      FALSE
```

Record RBPs classified by both quantitative and semi-quantitative methods

```
ProteinIntensity$RBP_type='no_evidence'
ProteinIntensity$RBP_type[ProteinIntensity$semi_quant]='semi_quantified'
ProteinIntensity$RBP_type[ProteinIntensity$limma_padj<=0.05]='quantified'
ProteinIntensity[[cellline]]='no_evidence'
ProteinIntensity[[cellline]][ProteinIntensity$RBP_type %in%
                             c("semi_quantified","quantified")] = 'mRNA interactome'
```

4 Comparison to known RBDs

4.1 Find the corresponding match by ENSEMBL gene IDs for single match peptides in human cell lines

```
load(file.path('inputdata','Annotation','enigmRBP.rda'))
m = match(ProteinIntensity$ENSG,enigmRBP$Ensembl.gene.ID)
ProteinIntensity = cbind(ProteinIntensity,enigmRBP[m,c("HeLa","HEK293","HuH7")])

save(ProteinIntensity, file=file.path("differentialProteinAbundance","ProteinIntensity.rda"))
write.csv(ProteinIntensity, file=file.path("differentialProteinAbundance","ProteinIntensity.csv"),row.names=NA)

#
# map_summary$ensg_id=unlist(Pep2Ensg[pep_match$matches=='longest_match'])
# l=sapply(map_summary$ensg_id,
#          function(x) which(enigmRBP$Ensembl.gene.ID==x))
# l[l$length==0]=NA
# l=unlist(l)
# map_summary=cbind(map_summary,
#                   enigmRBP[l,c("HeLa","HEK293","HuH7")])
# map_summary=map_summary[order(map_summary$prot_id),]
# row.names(map_summary)=NULL
```

We define a protein as known RBP if it has been previously described as RBP in at least 1 cell lines.

```
known_RBP=apply(ProteinIntensity[,c('HeLa','HEK293','HuH7')],1,
                function(x) sum(x=='mRNA interactome',na.rm=TRUE))
quant_count=table(factor(known_RBP
                        [ProteinIntensity$RBP_type=='quantified'],levels=c(0,1,2,3)))
semi_count=table(factor(known_RBP
                       [ProteinIntensity$RBP_type=='semi_quantified'],levels=c(0,1,2,3)))
```

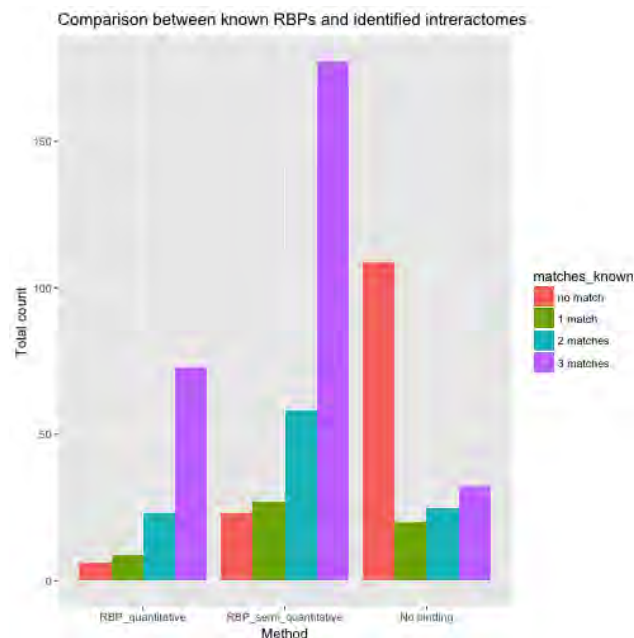
```

none_count=table(factor(known_RBP
                        [ProteinIntensity$RBP_type=='no_evidence'],levels=c(0,1,2,3)))

Intersect=data.frame(method=rep(c('RBP_quantitative',
                                  'RBP_semi_quantitative', 'No binding'),each=4),
                     count=c(quant_count,semi_count,none_count),
                     matches_known=rep(c('no match',
                                          '1 match','2 matches','3 matches'),3))
Intersect$matches_known=factor(Intersect$matches_known,
                                levels=c('no match',
                                          '1 match','2 matches','3 matches'))
Intersect$method=factor(Intersect$method,levels=
                        c('RBP_quantitative','RBP_semi_quantitative','No binding'))

overview=ggplot(data=Intersect,
                aes(x=method,y=count,fill=matches_known))
overview+geom_bar(stat='identity',
                  position='dodge')+xlab("Method") + ylab('Total count') +
ggtitle("Comparison between known RBPs and identified intreractomes")

```



5 Gene set enrichment analysis

Gene set enrichment analysis is performed by applying Fishers exact test to categories from GO, Interpro, Pfam, and Uniprot annotations.

```

resultdir = 'GSEA'

GSEA = list()
RBP_protid=ProteinIntensity$prot_id[ProteinIntensity$RBP_type!='no_evidence']
RBP_Ensg = ProteinIntensity$ENSG[ ProteinIntensity[[cellline]] == "mRNA interactome" ]
GSEA=doGSEA(RBP_Ensg, unique(ENSGannotation$ensembl_gene_id), ENSG2category, k=3)
dir.create(file.path(resultdir))
save(GSEA, file=file.path(resultdir,'GSEA.rda'))

```

The tables of the GSEA are written.


```
#for (n in names(GSEA)) {
dir.create(file.path(resultdir,"table"),showWarnings=FALSE,recursive = TRUE)
file.copy(system.file("images","hwriter.css",package="hwriter"),
           file.path(resultdir,"hwriter.css"))

## [1] TRUE

hwriteGSEA(GSEA, dir = file.path(resultdir,"table"),
           GeneAnnotation=ENSGannotation)
hwriteGSEAtext(GSEA, file = file.path(resultdir,"GSEA.txt"),
              GeneAnnotation=ENSGannotation)
hwriteGSEAreviso(GSEA, file = file.path(resultdir,"GSEAreviso.txt"),
                 GeneAnnotation=ENSGannotation,padj=0.4)
```

A webpage summarizing the GSEA and respective barplots is generated.

```
Sources = unique(GSEA$T$source[GSEA$T$padj <= 0.05])
page = openPage(file.path(resultdir, "index.html"),link.css = "hwriter.css")
hwriteSidebar(page=page, link= "../mainmenu.html")
hwrite(sprintf("Gene set enrichment analysis"),heading=1,page=page)
hwrite("table of enriched gene sets",link=file.path("table","index.html"),
       target="_GSEA",page=page,br=TRUE)
hwrite("[Download as text file]",link="GSEA.txt",page=page,br=TRUE)
hwrite("[Download text file for revigo GO enrichment visualization]",
       link="GSEAreviso.txt",page=page,br=FALSE)
hwrite(" Upload the text file on revigo",link="http://revigo.irb.hr",page=page,br=FALSE)
M = NULL
for (s in Sources) {
  GSEAnew = GSEA
  GSEAnew$T = GSEAnew$T[GSEAnew$T$source == s,]
  pdf(width=10,height=7,file=file.path(resultdir,sprintf("GSEA-%s.pdf",s)))
  x = plotGSEA(GSEAnew, main = s, maxp = 0.05,nrOfSets = 0, cex=0.6)
  dev.off()
  h = x / 20 + 0.25
  pdf(width=9,height=h*7,file=file.path(resultdir,sprintf("GSEA-%s.pdf",s)))
  plotGSEA(GSEAnew, main = s, maxp = 0.05,nrOfSets = 0, cex=1)
  dev.off()
  png(width=685,height=h*480,file=file.path(resultdir,sprintf("GSEA-%s.png",s)))
  x = plotGSEA(GSEAnew, main = s, maxp = 0.05,nrOfSets = 0, cex=1)
  dev.off()
  if (x > 0) {
    M = rbind(M, c(sprintf("GSEA-%s.%s",s,c("png","pdf"))))
    row.names(M)[nrow(M)] = s
  }
}
if (!is.null(M)) {
  hwrite("Bar plots of enriched gene sets (FDR 0.05)",heading=1,page=page)
  hwriteImage(M[,1,drop=FALSE],link=M[,1,drop=FALSE],page=page)
} else {
  hwrite("no enriched categories",page=page,heading=3)
}
closePage(page,splash=FALSE)
#}
# page = openPage(file.path(resultdir, "index.html"),link.css = "hwriter.css")
# hwriteSidebar(page=page, link= "../mainmenu.html")
# hwrite("Gene set enrichment analysis",heading=1,page=page)
# M = matrix(names(GSEA),ncol=1)
# L = matrix(sprintf("%s/index.html",names(GSEA)),ncol=1)
# hwrite(M, link=L, page=page)
```

```
# closePage(page,splash=FALSE)
#}
# page = openPage(file.path(resultdir, "index.html"),link.css = "hwriter.css")
# hwriteSidebar(page=page, link=../mainmenu.html")
# hwrite("Gene set enrichment analysis",heading=1,page=page)
# M = matrix(names(GSEA),ncol=1)
# L = matrix(sprintf("%s/index.html",names(GSEA)),ncol=1)
# hwrite(M, link=L, page=page)
# closePage(page,splash=FALSE)
```

6 Sequence features

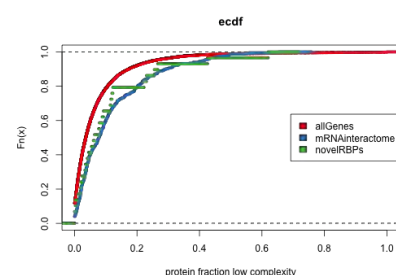
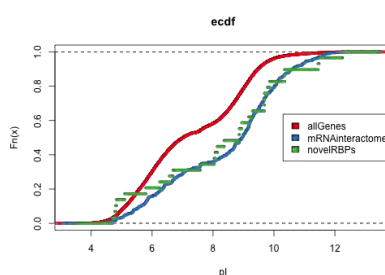
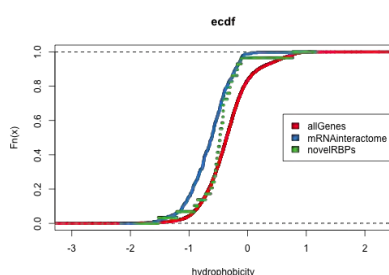
```
library(geneplotter)

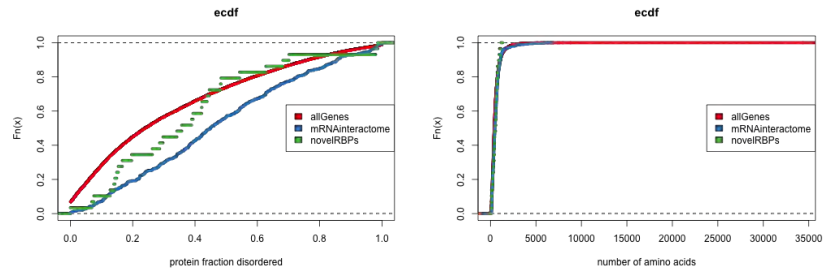
## Loading required package: lattice
## Loading required package: annotate
## Loading required package: XML

n = width(ProtFeatures$ProtSeq)
names(n) = names(ProtFeatures$ProtSeq)
ENSG = ENSGannotation$ensembl_gene_id
df = data.frame(seq_along(ProtFeatures$GeneName),ProtFeatures$GeneName, n)
df = df[!is.na(df[[2]]),]
df = df[order(df[[3]],decreasing=TRUE),]
df = df[!duplicated(df[[2]]),]
ENSG2prot = tapply(df[[1]], factor(df[[2]],levels=ENSG), function(x) { x })
```

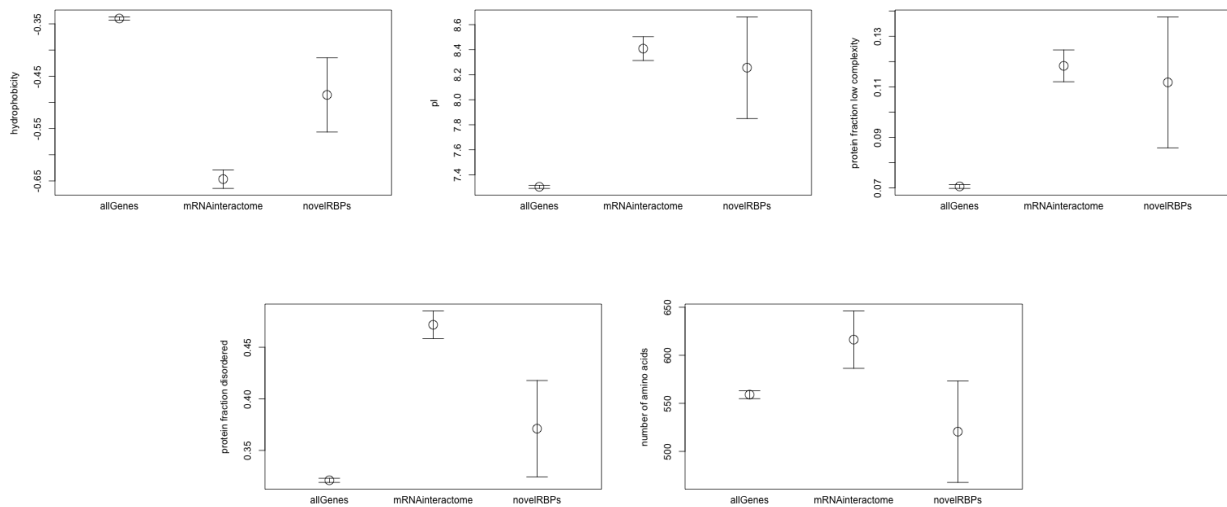
```
N = sapply(n[ENSG2prot], mean, na.rm=TRUE)
LowCompl = ProtFeatures$fracLowComplex[ENSG2prot]
Disorder = ProtFeatures$fracDisorder[ENSG2prot]
h = ProtFeatures$h[ENSG2prot]
pI = ProtFeatures$pI[ENSG2prot]
names(h) = names(pI) = names(LowCompl) =
  names(Disorder) = names(N) = ENSG
FT = list(h, pI, LowCompl, Disorder, N)
names(FT) = c("hydrophobicity","pI","protein fraction low complexity",
  "protein fraction disordered", "number of amino acids")
```

```
L = list(allGenes = ENSGannotation$ensembl_gene_id,
  mRNAInteractome = ProteinIntensity$ENSG[ ProteinIntensity[[cellline]] == "mRNA interactome"],
  novelRBPs = ProteinIntensity$ENSG[ ProteinIntensity[[cellline]] == "mRNA interactome" & known_
for (i in seq_along(FT)) {
  L1 = lapply(L, function(x) { FT[[i]][x] })
  L1 = lapply(L1, function(x) { x[is.finite(x)]})
  multiecdf(L1,lwd=5, xlab=names(FT)[i],subsample = FALSE)
}
```





```
for (i in seq_along(FT)) {
  L1 = lapply(L, function(x) { FT[[i]][x] })
  L1 = lapply(L1, function(x) { x[is.finite(x)] })
  m = sapply(L1, mean)
  s = sapply(L1, sd)
  n = sapply(L1, length)
  sem = s / sqrt(n)
  ylim=range(c(m+sem,m-sem))
  plot(1:3+0.5,m,xlim=c(1,4),xlab="",xaxt="n",ylab=names(FT)[i],cex=2,ylim=ylim)
  segments(x0 = 1:3+0.5, x1=1:3+0.5, y0=m-sem, y1=m+sem)
  segments(x0 = 1:3+0.5-0.1, x1=1:3+0.5+0.1, y0=m-sem, y1=m-sem)
  segments(x0 = 1:3+0.5-0.1, x1=1:3+0.5+0.1, y0=m+sem, y1=m+sem)
  p = par(xpd=NA)
  text(1:3+0.5,ylim[1]-diff(ylim)/10,names(L1))
  par(p)
}
```



Comparison of protein intensity in cells infected with Influenza virus

Shuai Ni and Bernd Fischer

German Cancer Research Center (DKFZ),
Heidelberg, Germany
n.shuai@dkfz.de

April 3, 2017

Contents

1	Read data	1
2	Map the sequences to the protein sequence database	1
2.1	Aggregate mean intensity values for each protein	3
3	Differential t test in protein level	4
3.1	Moderated t-test for set enrichment	4
3.2	Volcano plots	6
3.3	A semi-quantified analysis for protein enrichment	8
4	Heatmaps, line charts and barplots	10
4.1	Bar plots	10
4.2	Line charts for protein intensities with no missing intensities	11
4.3	Heatmat for protein intensity profiles	12
5	Gene set enrichment analysis for differentially expressed proteins	13

1 Read data

Import database of all proteins from uniprot

Read peptide sequences and intensity values from two datasets.

```
peptides= read.table(  
  file.path("inputdata/peptides2.txt"),  
  sep="\t", comment.char="",quote="",  
  header=TRUE, stringsAsFactors=FALSE)
```

Remove contaminant and decoy peptide sequences.

```
peptides=peptides[peptides$Potential.contaminant==' '  
  & peptides$Reverse==' ',]
```

Specify sample names and remove unused columns from peptide table.

```
sample_names=c('hour9','hour6','mock')  
peptides =  
  peptides[,c(grep("Sequence",colnames(peptides)),  
    grep("Intensity.",colnames(peptides)))]
```

```

peptides=peptides[,c(1, 6:8, 10:12, 14:16)]
peptides[,c(2:10)][peptides[,c(2:10)]<1]=NA
peptides[,c(2:10)]=log2(peptides[,c(2:10)])

peptides=peptides[,c(1, 3,2,4, 5,7,6, 10,9,8)]

names(peptides)=
  c('sequences',paste(sample_names,rep(1:3,each=3),sep='_'))

```

2 Map the sequences to the protein sequence database

A function to map protein IDs to Ensembl gene IDs

```

MapProt2Ensg=function(protein_list){
  ###MapProt2Ensg maps the peptide-mapping proteins to genes
  ###and return a list of mapped genes
  lapply(protein_list,function(x)
    { y=unnname(ProtFeatures$GeneName[x])
      unique(y[!is.na(y)])})
}

```

Extract single matched peptides (peptides can only be mapped to a single gene). The name of the viral proteins is kept in the conversion from protein id to ENSEMBL gene id.

```

ProtIDs=mapPeptides(PeptideSet=peptides$sequence,
                    c(ProtFeatures$ProtSeq,IV_seq),
                    verbose=FALSE)
ENSGid=MapProt2Ensg(ProtIDs)
ENSGid[grep('Inf', ProtIDs)] = ProtIDs[grep('Inf', ProtIDs)]
ENSGid[grep('Inf', ProtIDs)]=lapply(ENSGid[grep('Inf', ProtIDs)], function(x) x[1])
####Peptides mapped to

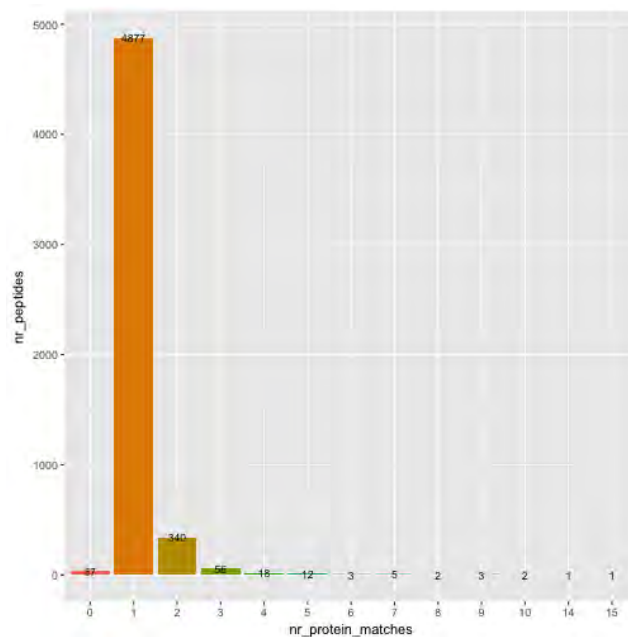
```

Mapping summary.

```

df=as.data.frame(table(listLen(ENSGid)))
colnames(df)=c('nr_protein_matches','nr_peptides')
p = ggplot(df, aes(x=nr_protein_matches,y=nr_peptides,
                  fill=factor(nr_protein_matches)))
p = p + geom_bar(stat='identity')+theme(legend.position = "none") +
  geom_text(aes(label = nr_peptides), size = 3)
p

```



Peptides mapping to multiple genes are not considered. For peptides mapped to more than one gene in influenza, the first one is counted.

```
ENSGidUnique = ENSGid
ENSGidUnique[listLen(ENSGidUnique)!= 1]=NA
ENSGidUnique=unlist(ENSGidUnique)
peptides=cbind(peptides,ENSGid=ENSGidUnique,stringsAsFactors=FALSE)
peptides$symbol = unname(ProtFeatures$Symbol[match(peptides$ENSGid,
                                                    ProtFeatures$GeneName,
                                                    incomparables = NA)])

peptides$symbol[grep('Inf',peptides$ENSGid)]=
  as.character(peptides$ENSGid[grep('Inf',peptides$ENSGid)])
```

The table of peptide intensities is saved.

```
dir.create("peptideMapping", showWarnings = FALSE)
save(peptides, file=file.path("peptideMapping/peptides.rda"))
```

The x axis represents the number protein(s) a peptide mapped to. The y axis represents the total number of peptides. Peptides mapped to more than one gene are discarded for downstream analysis.

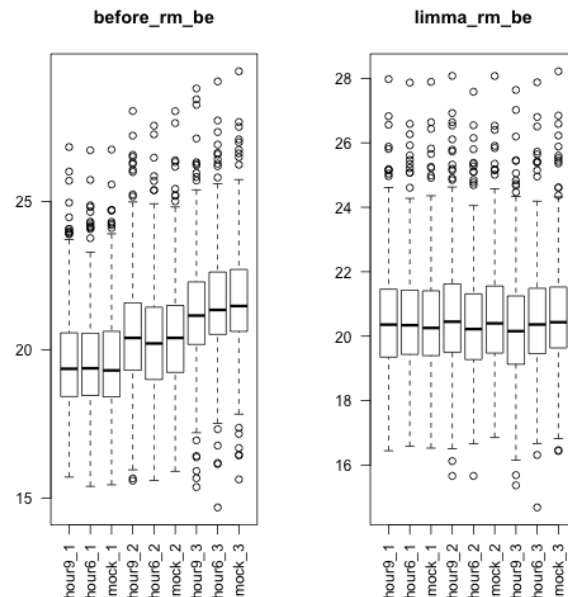
2.1 Aggregate mean intensity values for each protein

The mean intensity values of peptides mapped to the same gene are calculated to represent the intensity of that protein. For different batches, remove batch effects by fitting a linear model to the data with 'removeBatchEffect' in limma package.

```
par(mfrow=c(1,2))
proteins=aggregate(peptides[,2:10],
  by=list(peptides$ENSGid), mean, na.rm=TRUE)
names(proteins)[1]='ENSGid'
proteins$symbol = peptides$symbol[match(proteins$ENSGid, peptides$ENSGid)]

proteins$Know_RBP=ifelse(
  proteins$ENSGid %in% enigmRBP$Ensembl.gene.ID, 'known_RBP', 'no')
proteins=proteins[,c(1,11,12,2:10)]
proteins=data.frame(proteins,stringsAsFactors = FALSE )
boxplot(proteins[,4:12], las=2, main='before_rm_be')
batch <- c("A","A","A","B","B","B","C","C","C")
```

```
proteins[,4:12]=removeBatchEffect(proteins[,4:12], batch)
## Warning: Partial NA coefficients for 149 probe(s)
boxplot(proteins[,4:12], las=2, main='limma_rm_be')
```



```
head(proteins[,4:12])
```

	hour9_1	hour6_1	mock_1	hour9_2	hour6_2	mock_2	hour9_3	hour6_3	mock_3
## 1	NaN	NaN	NaN	NaN	NaN	NaN	19.33049	19.43274	19.73408
## 2	17.89934	17.79686	18.27341	NaN	NaN	NaN	17.59473	18.04519	18.32969
## 3	19.53525	20.24571	21.03399	19.44123	20.50825	20.86547	19.25870	20.61763	20.93862
## 4	NaN	18.61872	NaN	18.74992	18.19436	18.91189	18.34495	18.74934	18.76188
## 5	22.47857	22.14278	22.45854	22.50070	22.07163	22.50757	22.12579	22.31877	22.63534
## 6	22.40569	22.63386	22.41255	22.52634	22.40486	22.52090	22.10802	22.52970	22.81439

The table of protein intensities is saved.

```
save(proteins, file=file.path("peptideMapping/proteins.rda"))
```

3 Differential t test in protein level

3.1 Moderated t-test for set enrichment

The aggregated mean intensity values for each protein are tested for enrichment between different states of infection, using a moderated t-test as implemented in the Bioconductor-package limma. p-values are corrected for multiple testing by the method of Benjamini-Hochberg.

```
dirTtest='t_test'
dir.create(file.path(dirTtest), showWarnings = FALSE)

Intensities = as.matrix(proteins[,grepl("hour",colnames(proteins)) |
                        grepl("mock",colnames(proteins))])
cond = sapply(strsplit(colnames(Intensities), split="_"),
              function(x) { x[1] })

sample_combi=combn(sample_names,2,simplify=TRUE)
colnames(sample_combi) =
```

```

apply(sample_combi,2,
      function(x) {
        paste('diff',paste(x,collapse='_'),sep='_')
      })

diff_table = diff_intensities = list()
for (s in colnames(sample_combi)) {
  sample1 = Intensities[,cond == sample_combi[1,s]]
  sample2 = Intensities[,cond == sample_combi[2,s]]
  X =sample1-sample2
  diff_intensities[[s]] = X
  fit=eBayes(lmFit(X))
  fit$p.adj=p.adjust(fit$p.value,method="BH")
  dt = data.frame(ENSGid = proteins$ENSGid,
                  symbol = proteins$symbol,
                  Know_RBP = proteins$Know_RBP,
                  log2FC = fit$coefficients[,1],
                  p.value = fit$p.value[,1],
                  p.adj = fit$p.adj,
                  stringsAsFactors=FALSE)

  dt$sig=''
  sum(dt$sig=='')
  dt$sig[dt$p.adj<0.1]='*'
  dt$sig[dt$p.adj<0.01]='**'
  dt=dt[,c(1,2,3,4,5,6,7)]
  diff_table[[s]] = dt
}

```

The t-test result tables are saved.

```
save(diff_table, file=file.path("t_test/diff_table.rda"))
```

And tables are written as text files.

```

for (s in colnames(sample_combi)) {
  comparison=diff_table[[s]]
  col = ifelse(comparison$p.adj <= 0.2, "orange", "gray")
  col[which(comparison$p.adj <= 0.1)] = "firebrick1"
  col[which(comparison$p.adj <= 0.01)] = "firebrick"
  col[which(comparison$p.adj <= 0.2 & comparison$log2FC < 0)] = "turquoise1"
  col[which(comparison$p.adj <= 0.1 & comparison$log2FC < 0)] = "turquoise4"
  col[which(comparison$p.adj <= 0.01 & comparison$log2FC < 0)] = "blue"
  col[col=='gray']=adjustcolor('gray', alpha=0.2)
  my_line <- function(x,y,...){
    points(x,y,col=col, ...)
    abline(h=0 ,v=0 , ...)
  }

  Int=diff_intensities[[s]]

  ##Add dummy data as a control of max and min range of the plot
  MX=max(Int, na.rm=TRUE)+0.5
  MN=min(Int, na.rm=TRUE)-0.5
  nonset=matrix(rep(NA, 18), ncol=3)
  nonset[1:2,1]=c(MX, MN)
  nonset[3:4,2]=c(MX, MN)
  nonset[5:6,3]=c(MX, MN)
  colnames(nonset)=colnames(Int)
  Int=rbind(Int, nonset)
}

```



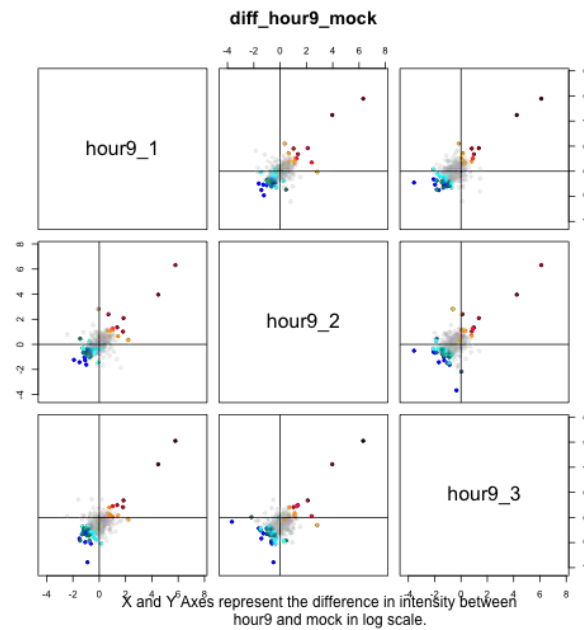
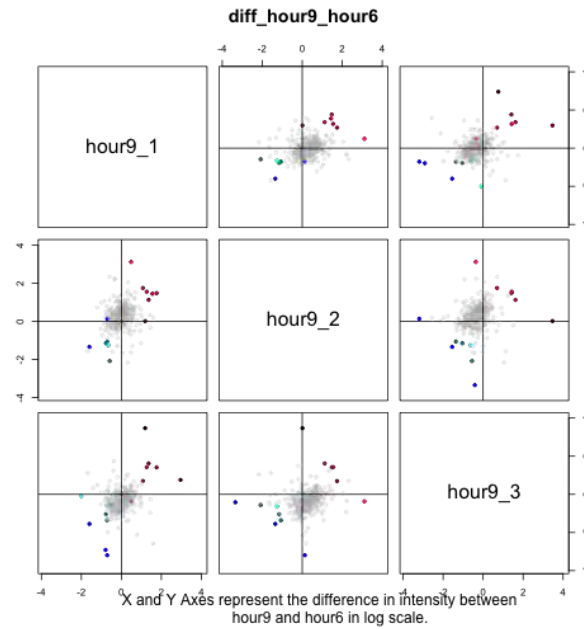
```

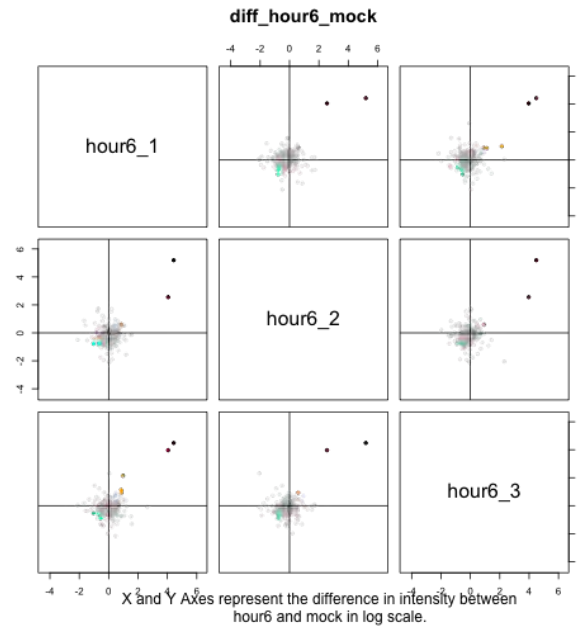
pairs(Int,pch=20, main=s,
panel=my_line)
title(sub = paste0('X and Y Axes represent the difference in intensity between
', strsplit(s, '_')[[1]][2], ' and ', strsplit(s, '_')[[1]][3], ' in log scale.'))

```

```

}
```





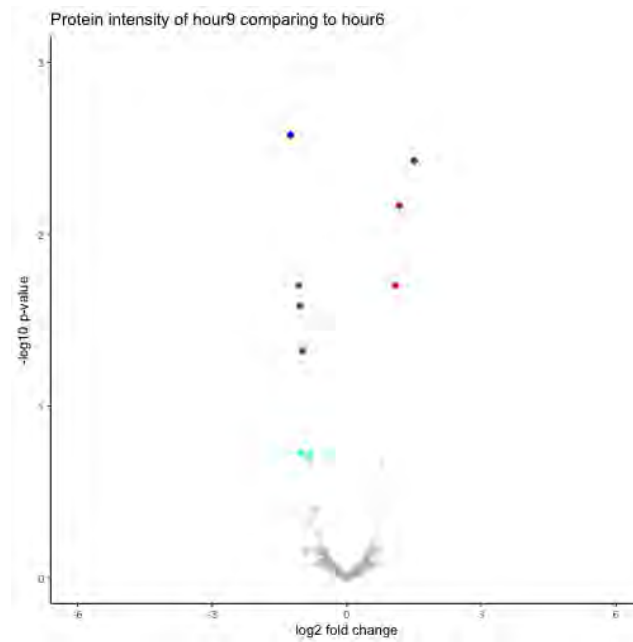
3.2 Volcano plots

```
voldir=file.path('t_test')
dir.create(voldir, showWarnings = FALSE)
for (s in colnames(sample_combi)) {
  volData = diff_table[[s]]
  volData = volData[!is.na(volData$p.adj),]
  col = ifelse(volData$p.adj <= 0.2, "orange", "gray")
  col[which(volData$p.adj <= 0.1)] = "firebrick1"
  col[which(volData$p.adj <= 0.01)] = "firebrick"
  col[which(volData$p.adj <= 0.2 & volData$log2FC < 0)] = "turquoise1"
  col[which(volData$p.adj <= 0.1 & volData$log2FC < 0)] = "turquoise4"
  col[which(volData$p.adj <= 0.01 & volData$log2FC < 0)] = "blue"
  col[col=="gray"]=adjustcolor('gray', alpha=0.2)

  g = ggplot(volData, aes(log2FC,-log10(p.adj)))+
    geom_point(size=2, color=col)+
    theme(panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          panel.background = element_blank(),
          axis.line = element_line(colour = "black"))+
    geom_text(data=subset(volData, abs(log2FC) > 2 ),
              aes(log2FC,-log10(p.adj),label=symbol),
              angle=90,hjust = 0,nudge_y = 0.05) +
    xlim(c(-6, 6)) + ylim(c(0, 3)) +
    xlab("log2 fold change") + ylab("-log10 p-value") +
    labs(title = paste('Protein intensity of ',
                      sample_combi[1,s], ' comparing to ',
                      sample_combi[2,s],sep=''))

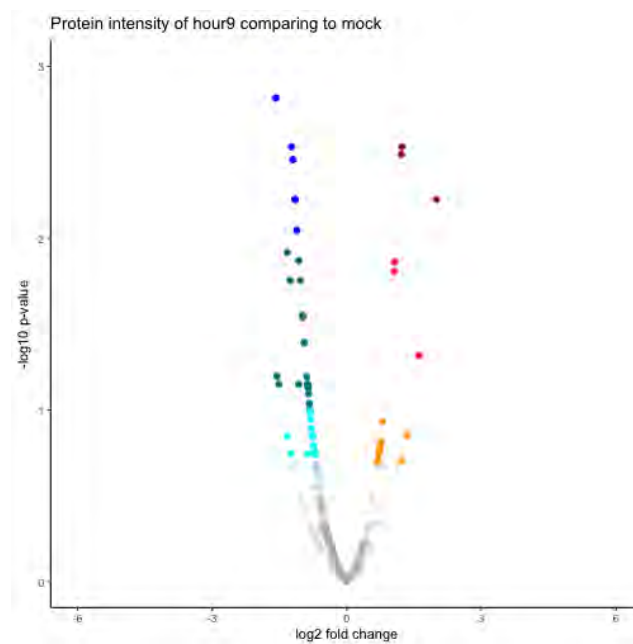
  print(g)
}

## Warning: Removed 8 rows containing missing values (geom_point).
```



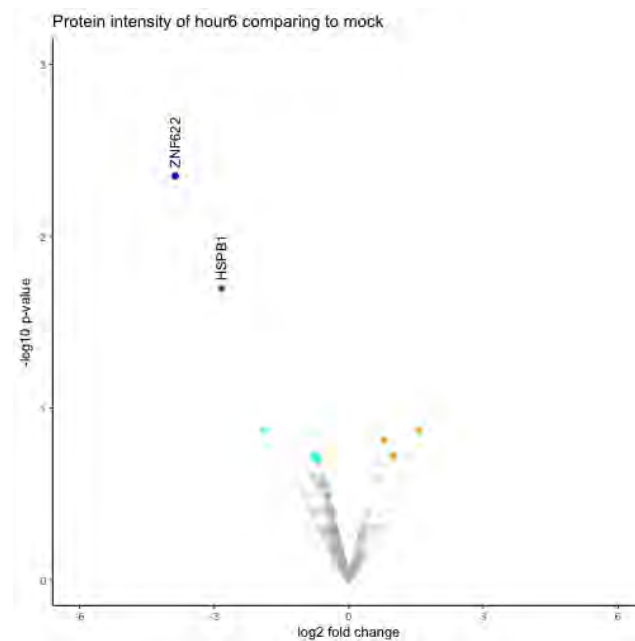
```
## Warning: Removed 9 rows containing missing values (geom_point).
```

```
## Warning: Removed 6 rows containing missing values (geom_text).
```



```
## Warning: Removed 3 rows containing missing values (geom_point).
```

```
## Warning: Removed 3 rows containing missing values (geom_text).
```



3.3 A semi-quantified analysis for protein enrichment

A semi-quantified analysis of the protein enrichment in different states of infection. Generating tables contain proteins which have detected intensity values in 3 or 2 of the replications in a time point but have only 1 or 0 detected intensity in another.

```
semi_dir=file.path('semi_quant')
dir.create(semi_dir, showWarnings = FALSE)
add_summary=function(proteinset){
  hour9_index=grep('9',colnames(proteinset))
  hour6_index=grep('6',colnames(proteinset))
  mock_index=grep('mock',colnames(proteinset))
  proteinset=cbind(proteinset,
    hour9_total=apply(proteinset,1,
      function(x) {
        sum(!is.na(x[hour9_index]))
      })
  )
  proteinset=cbind(proteinset,
    hour6_total=apply(proteinset,1,
      function(x) {
        sum(!is.na(x[hour6_index]))
      })
  )
  proteinset=cbind(proteinset,
    mock_total=apply(proteinset,1,
      function(x) {
        sum(!is.na(x[mock_index]))
      })
  )
  proteinset
}

semi_hour6_mock=apply(is.na(Intensities[,cond == "mock"]) -
  is.na(Intensities[,cond == "hour6"]), 1,
  function(x) { sum(x) } )
semi_hour6_mock_up = proteins[semi_hour6_mock>=2,]
semi_hour6_mock_up =
  semi_hour6_mock_up[,c(1,2,3,6,9,12,5,8,11,4,7,10)]
semi_hour6_mock_up=add_summary(semi_hour6_mock_up)
```

```

write.table(semi_hour6_mock_up, file=
            file.path(semi_dir, 'semi_hour6_mock_up.txt'),
            row.names=FALSE, quote=FALSE)

semi_hour6_mock_down=proteins[semi_hour6_mock<=-2,]
semi_hour6_mock_down=
    semi_hour6_mock_down[,c(1,2,3,6,9,12,5,8,11,4,7,10)]
semi_hour6_mock_down=add_summary(semi_hour6_mock_down)
write.table(semi_hour6_mock_down, file=
            file.path(semi_dir, 'semi_hour6_mock_down.txt'),
            row.names=FALSE, quote=FALSE)

semi_hour9_mock=apply(is.na(Intensities[,cond == "mock"]) -
                      is.na(Intensities[,cond == "hour9"]), 1,
                      function(x) { sum(x) } )
semi_hour9_mock_up=proteins[semi_hour9_mock>=2,]
semi_hour9_mock_up=
    semi_hour9_mock_up[,c(1,2,3,6,9,12,5,8,11,4,7,10)]
semi_hour9_mock_up=add_summary(semi_hour9_mock_up )
write.table(semi_hour9_mock_up, file=
            file.path(semi_dir, 'semi_hour9_mock_up.txt'),
            row.names=FALSE, quote=FALSE)

semi_hour9_mock_down=proteins[semi_hour9_mock<=-2,]
semi_hour9_mock_down=
    semi_hour9_mock_down[,c(1,2,3,6,9,12,5,8,11,4,7,10)]
semi_hour9_mock_down=add_summary(semi_hour9_mock_down)
write.table(semi_hour9_mock_down,
            file=file.path(semi_dir, 'semi_hour9_mock_down.txt'),
            row.names=FALSE, quote=FALSE)

semi_hour9_hour6=apply(is.na(Intensities[,cond == "hour6"]) -
                      is.na(Intensities[,cond == "hour9"]), 1,
                      function(x) { sum(x) } )
semi_hour9_hour6_up=proteins[semi_hour9_hour6>=2,]
semi_hour9_hour6_up=
    semi_hour9_hour6_up[,c(1,2,3,6,9,12,5,8,11,4,7,10)]
semi_hour9_hour6_up=add_summary(semi_hour9_hour6_up)
write.table(semi_hour9_hour6_up,
            file=file.path(semi_dir, 'semi_hour9_hour6_up.txt'),
            row.names=FALSE, quote=FALSE)

semi_hour9_hour6_down=proteins[semi_hour9_hour6<=-2,]
semi_hour9_hour6_down=
    semi_hour9_hour6_down[,c(1,2,3,6,9,12,5,8,11,4,7,10)]
semi_hour9_hour6_down=add_summary(semi_hour9_hour6_down)
write.table(semi_hour9_hour6_down,
            file=file.path(semi_dir, 'semi_hour9_hour6_down.txt'),
            row.names=FALSE, quote=FALSE)

```

4 Heatmaps, line charts and barplots

4.1 Bar plots

```

bardir=file.path('BarPlot')
dir.create(bardir, showWarnings = FALSE)
a = data.frame(hour6_mock = diff_table[["diff_hour6_mock"]]$log2FC,
               hour9_mock = diff_table[["diff_hour9_mock"]]$log2FC,
               protein = diff_table[["diff_hour9_mock"]]$symbol)

isFinite = !is.na(a$hour6_mock) & !is.na(a$hour9_mock)
sig_exist=matrix(0, 3, 3)
row.names(sig_exist)=colnames(sig_exist)=sample_names

for (i in seq_len(ncol(sample_combi))) {
  for (j in c("up","down")) {
    if (j == "up") {
      I = which(diff_table[[i]]$p.adj <= 0.1 & diff_table[[i]]$log2FC > 0 & isFinite)
      if (length(I)>1){
        sig_exist[row.names(sig_exist)==sample_combi[,i][1],
                  colnames(sig_exist)==sample_combi[,i][2]]=1
      }
    } else {
      I = which(diff_table[[i]]$p.adj <= 0.1 & diff_table[[i]]$log2FC < 0 & isFinite)
      if (length(I)>1){
        sig_exist[row.names(sig_exist)==sample_combi[,i][2],
                  colnames(sig_exist)==sample_combi[,i][1]]=1
      }
    }
  }
  if (length(I) >1) {
    b=a[I,]
    aa=data.frame(time=rep(c('hour6_mock','hour9_mock'),
                           each=dim(b)[1]),
                  protein=rep(b$protein, 2),
                  values=c(b$hour6_mock,b$hour9_mock))
    g = ggplot(data=aa, aes(x=factor(protein, levels=b$protein[order(b$hour9_mock)]),
                           y=values,
                           fill=factor(time, levels=c('hour6_mock','hour9_mock')))) +
      geom_bar(stat='identity', position=position_dodge()) +
      scale_fill_hue(name="Time after infection") +
      xlab("Protein") + ylab("Intensity level") +
      ggtitle(paste0(j,"-regulated proteins in ",sample_combi[1,i],
                     " compare to ",sample_combi[2,i]," (padj=0.1)")) +
      theme_bw()+
      theme(legend.key.size = unit(0.5, "cm"),
            axis.text.x=element_text(size=5,angle=90,hjust=1,vjust=0.5))
    ggsave(plot = g, filename = file.path("BarPlot",paste0(
      "BarPlot-",paste(sample_combi[,i],collapse = '_'),"-",j,".pdf")))
    ggsave(plot = g, filename = file.path("BarPlot",paste0(
      "BarPlot-",paste(sample_combi[,i], collapse = '_'),"-",j,".png")))
  }
}

```

4.2 Line charts for protein intensities with no missing intensities

The ion intensities of each protein (sum of all peptides) at each time point are plotted for proteins with no missing values in all 3 replicates of all 3 modes. The background shows the overall intensity change over 3 modes.

```

dirlinechart='line_chart'
dir.create(file.path(dirlinechart), showWarnings = FALSE)
a = data.frame(hour6_mock = diff_table[["diff_hour6_mock"]]$log2FC,
               hour9_mock = diff_table[["diff_hour9_mock"]]$log2FC,
               protein = diff_table[["diff_hour9_mock"]]$symbol)
isFinite = !is.na(a$hour6) & !is.na(a$hour9)
files = c()
for (i in seq_len(ncol(sample_combi))) {
  for (j in c("up", "down")) {
    if (j == "up") {
      sig = diff_table[[i]]$p.adj <= 0.1 & diff_table[[i]]$log2FC > 0 & isFinite
    } else {
      sig = diff_table[[i]]$p.adj <= 0.1 & diff_table[[i]]$log2FC < 0 & isFinite
    }
    sig[is.na(sig)] = FALSE

    I = order(diff_table[[i]]$p.adj, decreasing = TRUE)
    gene = a$protein
    gene = factor(gene, levels=gene[I])

    dat <- data.frame(time = "00:00",
                     gene = factor(a$protein[isFinite], levels=gene),
                     proteinLevel = 0.0,
                     sig=sig[isFinite])
    dat <- rbind(dat,
                 data.frame(time = "06:00",
                           gene = factor(a$protein[isFinite], levels=gene),
                           proteinLevel = a$hour6_mock[isFinite],
                           sig=sig[isFinite]))
    dat <- rbind(dat,
                 data.frame(time = "9:00",
                           gene = factor(a$protein[isFinite], levels=gene),
                           proteinLevel = a$hour9_mock[isFinite],
                           sig=sig[isFinite]))

    #dat$flwd = ifelse(dat$sig == "sig", 1, 1)
    if (sum(sig) > 0) {
      g = ggplot(dat, aes(time, proteinLevel, group=gene)) +
        geom_line(col="gray") +
        geom_line(data=dat[which(dat$sig),], aes(colour="blue"), size=2) +
        geom_point(data=dat[which(dat$sig & dat$time != "00:00"),], shape=19, size=2, colour="red")
      pdf(file.path(dirlinechart, paste0("LinePlot-", i, "-", j, ".pdf")))
      print(g)
      svg = gridSvgLinePlot(g)
      dev.off()
      files = c(files, paste0("LinePlot-", i, "-", j, ".html"))
      page = openPage(file.path(dirlinechart, paste0("LinePlot-", i, "-", j, ".html")),
                      head = "<script src='\"http://d3js.org/d3.v3.js\"'></script>",
                      link.css = "hwriter.css")
      hwriteSidebar(position = c("left"),
                   link = c("../mainmenu.html"),
                   size = 15, margin = 1, page = page)
      hwrite(paste0(j, "-regulated proteins in ", sample_combi[1, i],
                   " compare to ", sample_combi[2, i], " (padj=0.1)",
                   page=page, heading=1)
      hwrite(svg, table=FALSE, page=page)
      closePage(page, splash=FALSE)
    }
  }
}

```

```
}
}
## Warning: closing unused connection 7 (t_test/diff_hour6_mock.html)
```

4.3 Heatmap for protein intensity profiles

Heatmap for proteins have no missing measurements, normalized by replicates

```
dirheatmap='heatmap'
dir.create(file.path(dirheatmap), showWarnings = FALSE)
a = data.frame(hour6_mock = diff_table[["diff_hour6_mock"]]$log2FC,
               hour9_mock = diff_table[["diff_hour9_mock"]]$log2FC,
               protein = diff_table[["diff_hour9_mock"]]$symbol)
isFinite = !is.na(a$hour6) & !is.na(a$hour9)
isSig = apply(sapply(diff_table, function(x) { x$sig != "" } ),1,any)

a = a[isFinite & isSig,]
X = as.matrix(a[,1:2])
row.names(X) = a$protein
X[X > 3] = 3
X[X < -3] = -3
r = max(abs(X))
pheatmap(t(X), border_color = NA, cluster_rows = FALSE, fontsize_col = 6,
          color = colorRampPalette(rev(brewer.pal(n = 7, name = 'RdBu')))(256),
          breaks = seq(-r,r,length.out=257),cellwidth = 5, cellheight = 20,
          filename = file.path(dirheatmap, "heatmap.pdf"))
pheatmap(t(X), border_color = NA, cluster_rows = FALSE, fontsize_col = 6,
          color = colorRampPalette(rev(brewer.pal(n = 7, name = 'RdBu')))(256),
          breaks = seq(-r,r,length.out=257),cellwidth = 5, cellheight = 20,
          filename = file.path(dirheatmap, "heatmap.png"))
```

5 Gene set enrichment analysis for differentially expressed proteins

Gene set enrichment analysis is performed by applying Fishers exact test to categories from GO, Interpro, Pfam, and Uniprot annotations.

Gene set enrichment analysis for enriched genes in sets of comparisons.

```
GSEAdir='GSEA'
dir.create(file.path(GSEAdir), showWarnings = FALSE)
file.copy(system.file("images", "hwriter.css", package="hwriter"),
          file.path(GSEAdir, "hwriter.css"))

## [1] FALSE

GSEA = list()
for (s in colnames(sample_combi)) {
  enrich_table = diff_table[[s]]
  enrich_table = enrich_table[which(enrich_table$p.adj<0.1),]
  for (j in 1:2) {
    if (sample_combi[j,s]==sample_combi[1,s]) {
      Test_Geneset=enrich_table[enrich_table$log2FC>0,]
    } else {
      Test_Geneset=enrich_table[enrich_table$log2FC<0,]
    }
    n = paste('Proteins enriched in ',
```



```

        sample_combi[j,s], ' comparing to ',
        sample_combi[3-j,s], sep='')
GSEA[[n]] = doGSEA(GenesSelected = as.character(Test_Geneset$ENSGid),
                  GenesBackground = as.character(proteins$ENSGid),
                  ENSG2category, k=3)
    }
}

save(GSEA, file=file.path(GSEAdir, "GSEA.rda"))

```

The tables of the GSEA are written.

```

for (n in names(GSEA)) {
  dir.create(file.path(GSEAdir, n, "table"),
             showWarnings=FALSE, recursive = TRUE)
  file.copy(system.file("images", "hwriter.css",
                        package="hwriter"),
            file.path(GSEAdir, n, "hwriter.css"))
  if (length(GSEA[[n]]$GenesSelected)>0){
    hwriteGSEA(GSEA[[n]], dir = file.path(GSEAdir, n, "table"),
              GeneAnnotation=ENSGannotation)
    hwriteGSEAtext(GSEA[[n]],
                  file = file.path(GSEAdir, n, "GSEA.txt"),
                  GeneAnnotation=ENSGannotation)
    hwriteGSEAreviso(GSEA[[n]],
                    file = file.path(GSEAdir, n, "GSEAreviso.txt"),
                    GeneAnnotation=ENSGannotation, padj=0.2)

    page=openPage(file.path(GSEAdir, n, 'table', 'index.html'),
                  link.css='hwriter.css')
    sample=sample_names[sample_names %in% unlist(strsplit(n, split=' '))]
    for (names in names(diff_table)){
      if (any(grep(sample[1], names)) & any(grep(sample[2], names))){
        index=diff_table[[names]]$ENSGid %in% GSEA[[n]]$GenesSelected
        Enriched_genes=diff_table[[names]][index,]
        hwrite(as.matrix(Enriched_genes), page=page, br=TRUE )
        closePage(page, splash=FALSE)
      }
    }
    write.table(GSEA[[n]]$T[GSEA[[n]]$T$nENR>=2,], file=file.path(GSEAdir, n, 'GSEA.txt'))
  }
}

```

```

sessionInfo()

## R version 3.3.2 (2016-10-31)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: OS X El Capitan 10.11.6
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats4      parallel    grid        stats       graphics    grDevices   utils       datasets
## [9] methods     base
##
## other attached packages:
## [1] pheatmap_1.0.8      gridSVG_1.5-0      Biostrings_2.42.1  XVector_0.14.0
## [5] IRanges_2.8.1       S4Vectors_0.12.1   RColorBrewer_1.1-2 hwriter_1.3.2

```

```
## [9] RBDmap_0.0.17      XML_3.98-1.5      ggplot2_2.2.1      limma_3.30.7
## [13] Biobase_2.34.0     BiocGenerics_0.20.0 VennDiagram_1.6.17 futile.logger_1.4.3
## [17] knitr_1.15.1
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.8        plyr_1.8.4        highr_0.6          futile.options_1.0.0
## [5] bitops_1.0-6       tools_3.3.2       zlibbioc_1.20.0    biomaRt_2.30.0
## [9] digest_0.6.11      memoise_1.0.0     evaluate_0.10      RSQLite_1.1-1
## [13] tibble_1.2         gtable_0.2.0      DBI_0.5-1          stringr_1.1.0
## [17] AnnotationDbi_1.36.0 RJSONIO_1.3-0     lambda.r_1.1.9     magrittr_1.5
## [21] codetools_0.2-15   scales_0.4.1      assertthat_0.1     BiocStyle_2.2.1
## [25] colorspace_1.3-2   labeling_0.3       stringi_1.1.2      RCurl_1.95-4.8
## [29] lazyeval_0.2.0     munsell_0.4.3     rjson_0.2.15
```

Comparison of whole cell lysate protein intensity in cells infected with Influenza virus in different states

Shuai Ni and Bernd Fischer

German Cancer Research Center (DKFZ),
Heidelberg, Germany
n.shuai@dkfz.de

July 25, 2017

Contents

1	Read data	1
1.1	Read peptide seqnecs and intensity values from two datasets	1
1.2	Remove comtaminant and decoy peptide sequences	1
2	Map the sequences to the protein sequence database	2
2.1	Aggregate mean intensity values for each ENSEMBL gene ID	2
3	Differential t test in protein level	8
3.1	Moderated t-test for set enrichment	8
3.2	Volcano plots	17
3.3	Line charts for protein intensities with no missing intensities	21
3.4	Stistical power of increased number of simulated replicates	22
3.5	Heatmat for protein intensity profiles	23
3.6	conficence interval comparison	24
3.7	Bar plots	26
3.8	A semi-quantified analysis for protein enrichment	27
4	Gene set enrichment analysis for differentially expressed proteins	31

1 Read data

Load all datasets and packages required.

```
## Warning: package 'limma' was built under R version 3.3.3
## Warning: package 'S4Vectors' was built under R version 3.3.3
## Warning: package 'IRanges' was built under R version 3.3.3
## Warning: package 'XVector' was built under R version 3.3.3
```

1.1 Read peptide seqnecs and intensity values from two datasets

```
input_raw_who= read.table(
  file.path("inputdata", 'flu_wholecell_peptides.txt'),
  sep="\t", comment.char="", quote="",
  header=TRUE, stringsAsFactors=FALSE)
```

```
input_raw= read.table(
  file.path("inputdata", 'flu_interactome_peptides.txt'),
  sep="\t", comment.char="", quote="",
  header=TRUE, stringsAsFactors=FALSE)
```

1.2 Remove contaminant and decoy peptide sequences

```
input_raw_who=input_raw_who[input_raw_who$Potential.contaminant==' '
  &input_raw_who$Reverse==' ',]
###
input_raw=input_raw[input_raw$Potential.contaminant==' '
  &input_raw$Reverse==' ',]
```

Specify sample names and remove unused columns from peptide table.

```
sample_names=c('hour9', 'hour6', 'mock')

input_raw_who =
  input_raw_who[,c(grep("Sequence", colnames(input_raw_who)),
    grep("Intensity.", colnames(input_raw_who)))]

input_raw_who=input_raw_who[,c(1,6:8, 10:12, 14:16)]
input_raw_who[,c(2:10)][input_raw_who[,c(2:10)]<1]=NA
input_raw_who[,c(2:10)]=log2(input_raw_who[,c(2:10)])
input_raw_who=input_raw_who[,c(1,3,2,4,5,7,6,10,9,8)]
names(input_raw_who)=
  c('Sequence', paste(sample_names, rep(1:3, each=3), sep='_'))

input_raw =
  input_raw[,c(grep("Sequence", colnames(input_raw)),
    grep("Intensity.", colnames(input_raw)))]
input_raw=input_raw[,c(1,6:8, 10:12, 14:16)]
input_raw[,c(2:10)][input_raw[,c(2:10)]<1]=NA
input_raw[,c(2:10)]=log2(input_raw[,c(2:10)])
input_raw=input_raw[,c(1,3,2,4,5,7,6,10,9,8)]
names(input_raw)=
  c('Sequence', paste(sample_names, rep(1:3, each=3), sep='_'))
```

2 Map the sequences to the protein sequence database

A function to map protein IDs to Ensembl gene IDs

```
MapProt2Ensg=function(protein_list){
  ###MapProt2Ensg maps the peptide-mapping proteins to genes
  ###and return a list of mapped genes
  lapply(protein_list,function(x)
    { y=unname(ProtFeatures$GeneName[x])
      unique(y[!is.na(y)])})
}
```

Extract single matched peptides (peptides can only be mapped to a single gene). The name of the viral proteins is kept in the conversion from protein id to ENSEMBL gene id.

```
ProtIDs_who=mapPeptides(PeptideSet=input_raw_who$Sequence,
  c(ProtFeatures$ProtSeq, Inf_seq),
  verbose=FALSE)
```

```

Ensgid_who=MapProt2Ensg(ProtIDs_who)
Ensgid_who[grep('Inf_', ProtIDs_who)]=
  ProtIDs_who[grep('Inf_', ProtIDs_who)]

ProtIDs=mapPeptides(PeptideSet=input_raw$Sequence,
                    c(ProtFeatures$ProtSeq, Inf_seq),
                    verbose=FALSE)

Ensgid=MapProt2Ensg(ProtIDs)
Ensgid[grep('Inf_', ProtIDs)]=ProtIDs[grep('Inf_', ProtIDs)]
Ensgid[grep('Inf_', ProtIDs)]=lapply(Ensgid[grep('Inf_', ProtIDs)], function(x) x[1])

```

2.1 Aggregate mean intensity values for each ENSEMBL gene ID

The mean intensity values of peptides mapped to the same gene are calculated to represent the intensity of that protein.

```

Ensgid_who[listLen(Ensgid_who)!= 1]=NA
Ensgid_who=unlist(Ensgid_who)
input_raw_who=cbind(input_raw_who, Ensgid_who=Ensgid_who)
input_raw_who$symbol = unname(ProtFeatures$Symbol[match(input_raw_who$Ensgid,
                                                    ProtFeatures$GeneName,
                                                    incomparables = NA)])

input_raw_who$symbol[grep('Inf_', input_raw_who$Ensgid)]=
  as.character(input_raw_who$Ensgid[grep('Inf_', input_raw_who$Ensgid)])

proteins_who=aggregate(input_raw_who[,2:10],
  by=list(input_raw_who$Ensgid_who), mean, na.rm=TRUE)
names(proteins_who)[1]='ENSGid'
proteins_who$symbol = input_raw_who$symbol[match(proteins_who$ENSGid, input_raw_who$Ensgid)]

proteins_who$Know_RBP=ifelse(
  proteins_who$ENSGid %in% enigmRBP$Ensembl.gene.ID,
  'known_RBP', 'no')
proteins_who=proteins_who[,c(1,11,12,2:10)]
proteins_who=data.frame(proteins_who, stringsAsFactors = FALSE )
row.names(proteins_who)=NULL

####interactome
Ensgid[listLen(Ensgid)!= 1]=NA
Ensgid=unlist(Ensgid)
input_raw=cbind(input_raw, Ensgid=Ensgid, stringsAsFactors=FALSE)
input_raw$symbol = unname(ProtFeatures$Symbol[match(input_raw$Ensgid,
                                                    ProtFeatures$GeneName,
                                                    incomparables = NA)])

input_raw$symbol[grep('Inf_', input_raw$Ensgid)]=
  as.character(input_raw$Ensgid[grep('Inf_', input_raw$Ensgid)])

proteins=aggregate(input_raw[,2:10],
  by=list(input_raw$Ensgid), mean, na.rm=TRUE)

names(proteins)[1]='ENSGid'
proteins$symbol =input_raw$symbol[match(proteins$ENSGid, input_raw$Ensgid)]
proteins$Know_RBP=ifelse(
  proteins$ENSGid %in% enigmRBP$Ensembl.gene.ID,
  'known_RBP', 'no')
proteins=proteins[,c(1,11,12,2:10)]
proteins=data.frame(proteins, stringsAsFactors = FALSE )

```

Comparison of whole cell lysate protein intensity in cells infected with Influenza virus in different states 4

```
###Proteins above is interactome, below is interactome subtract whole cell lysate as background
proteins_int=proteins
```

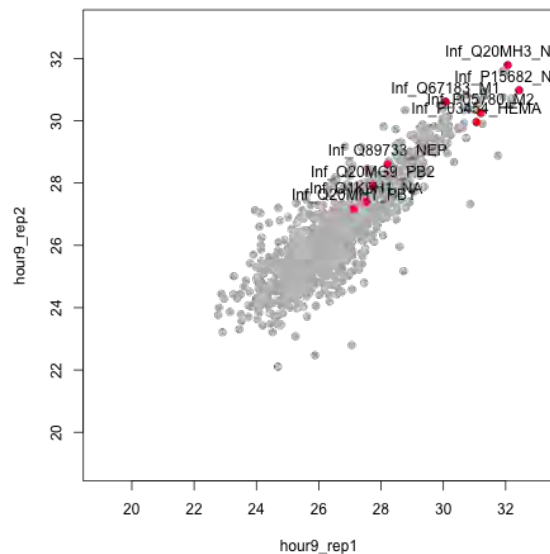
```
proteins=merge(proteins_who, proteins,
               by=c('ENSGid', 'symbol', 'Know_RBP'))
```

```
proteins[,4:12]=proteins[,4:12]-proteins[,13:21]
proteins=proteins[,1:12]
saveRDS(proteins, 'Proteins.rds')
saveRDS(proteins_who, 'Proteins_who.rds')
saveRDS(proteins_int, 'Proteins_int.rds')
```

```
####ploting SVs
```

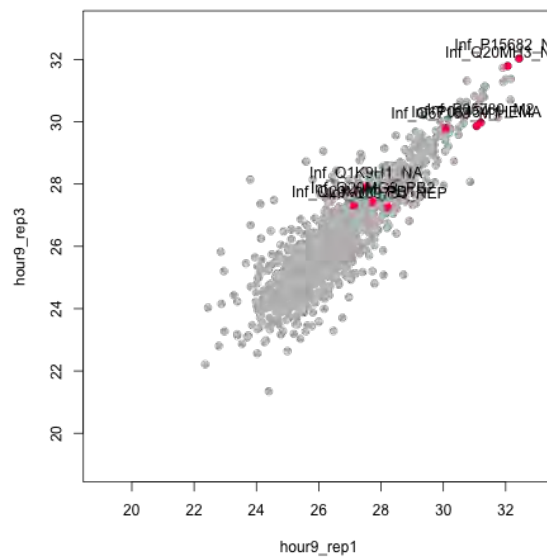
```
proteins_who=readRDS('proteins_who.rds')
grepl('Inf_',proteins_who$ENSGid)->I
par(pty="s")
plot(proteins_who[,4], proteins_who[,7], col='grey',
     pch=19,xlim=c(19,33), ylim=c(19,33),
     xlab='hour9_rep1', ylab='hour9_rep2')
points(proteins_who[I,4], proteins_who[I,7], col='red', pch=19)
text(proteins_who[I,4], proteins_who[I,7],
     labels=proteins_who$ENSGid[I], cex= 1, pos=3)
title(main='Protein abundance 9 hours after infection (rep1 Vs rep2), whole cell lysate' )
```

Protein abundance 9 hours after infection (rep1 Vs rep2), whole cell lysate



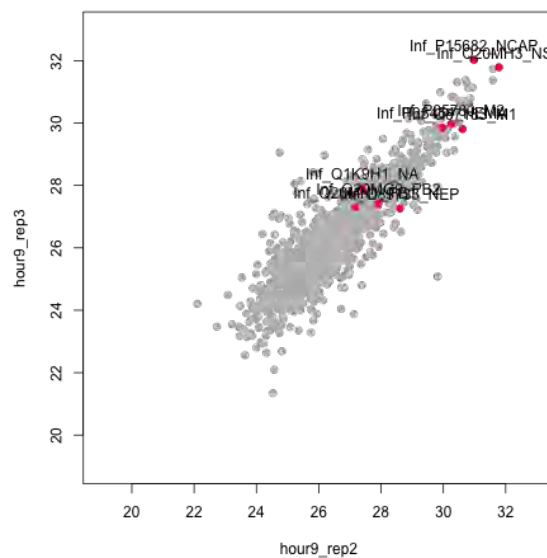
```
plot(proteins_who[,4], proteins_who[,10],
     col='grey', pch=19,xlim=c(19,33), ylim=c(19,33),
     xlab='hour9_rep1', ylab='hour9_rep3')
points(proteins_who[I,4], proteins_who[I,10], col='red', pch=19)
text(proteins_who[I,4], proteins_who[I,10],
     labels=proteins_who$ENSGid[I], cex= 1, pos=3)
title(main='Protein abundance 9 hours after infection (rep1 Vs rep3), whole cell lysate' )
```

Protein abundance 9 hours after infection (rep1 Vs rep3), whole cell lysate



```
plot(proteins_who[,7], proteins_who[,10],
     col='grey', pch=19, xlim=c(19,33), ylim=c(19,33),
     xlab='hour9_rep2', ylab='hour9_rep3')
points(proteins_who[I,7], proteins_who[I,10], col='red', pch=19)
text(proteins_who[I,7], proteins_who[I,10],
     labels=proteins_who$ENSGid[I], cex= 1, pos=3)
title(main='Protein abundance 9 hours after infection (rep2 Vs rep3), whole cell lysate' )
```

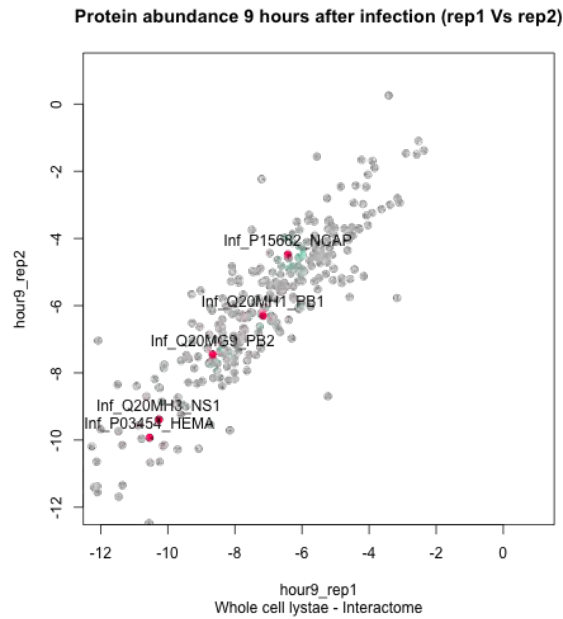
Protein abundance 9 hours after infection (rep2 Vs rep3), whole cell lysate



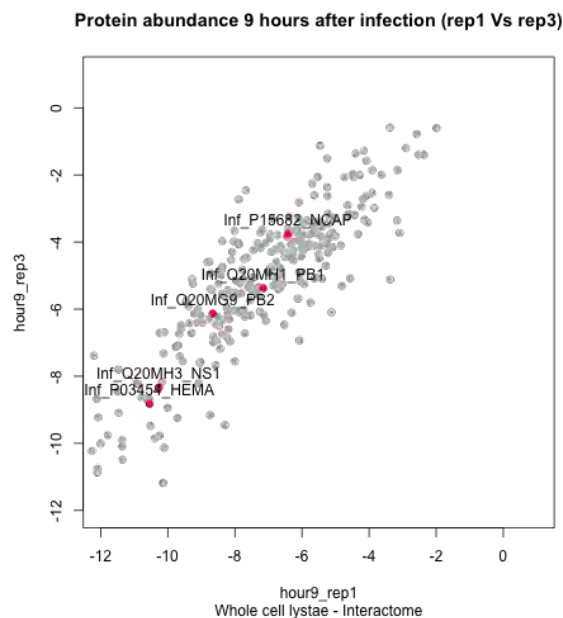
```
####
proteins=readRDS('proteins.rds')
grepl('Inf_',proteins$ENSGid)->I

plot(proteins[,4], proteins[,7], col='grey', pch=19, xlim=c(-12,1), ylim=c(-12,1),
     xlab='hour9_rep1', ylab='hour9_rep2')
points(proteins[I,4], proteins[I,7], col='red', pch=19)
text(proteins[I,4], proteins[I,7], labels=proteins$ENSGid[I], cex= 1, pos=3)
title(main='Protein abundance 9 hours after infection (rep1 Vs rep2)',
```

```
sub='Whole cell lystae - Interactome')
```

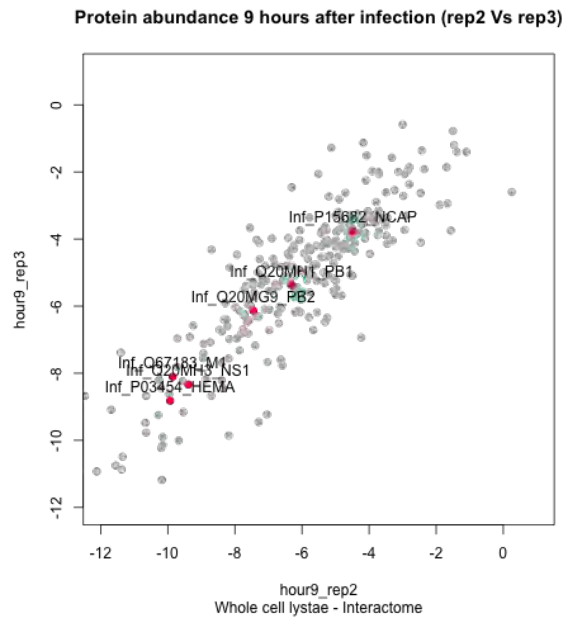


```
plot(proteins[,4], proteins[,10], col='grey', pch=19, xlim=c(-12,1), ylim=c(-12,1),
      xlab='hour9_rep1', ylab='hour9_rep3')
points(proteins[I,4], proteins[I,10], col='red', pch=19)
text(proteins[I,4], proteins[I,10], labels=proteins$ENSGid[I], cex= 1, pos=3)
title(main='Protein abundance 9 hours after infection (rep1 Vs rep3)' ,
      sub='Whole cell lystae - Interactome' )
```



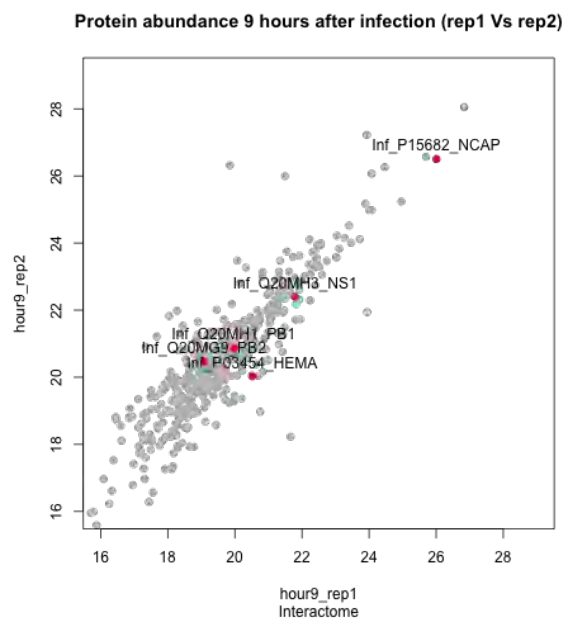
```
plot(proteins[,7], proteins[,10], col='grey', pch=19, xlim=c(-12,1), ylim=c(-12,1),
      xlab='hour9_rep2', ylab='hour9_rep3')
points(proteins[I,7], proteins[I,10], col='red', pch=19)
text(proteins[I,7], proteins[I,10], labels=proteins$ENSGid[I], cex= 1, pos=3)
title(main='Protein abundance 9 hours after infection (rep2 Vs rep3)' ,
      sub='Whole cell lystae - Interactome' )
```


Comparison of whole cell lysate protein intensity in cells infected with Influenza virus in different states 7

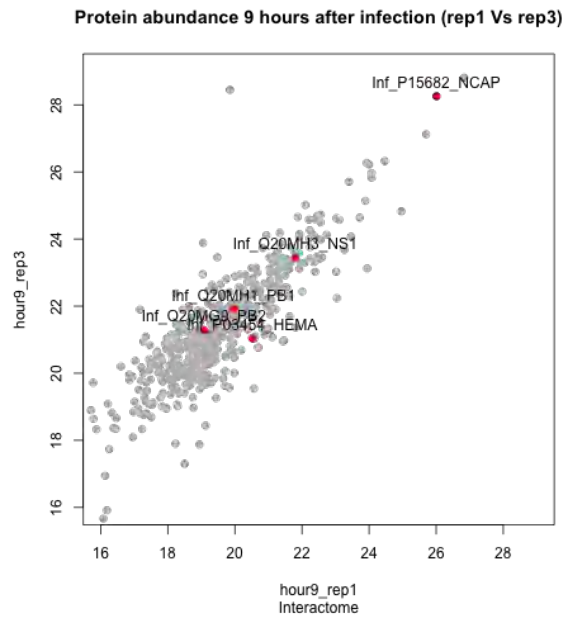


```
#####
proteins_int=readRDS('proteins_int.rds')
grepl('Inf_',proteins_int$ENSGid)->I

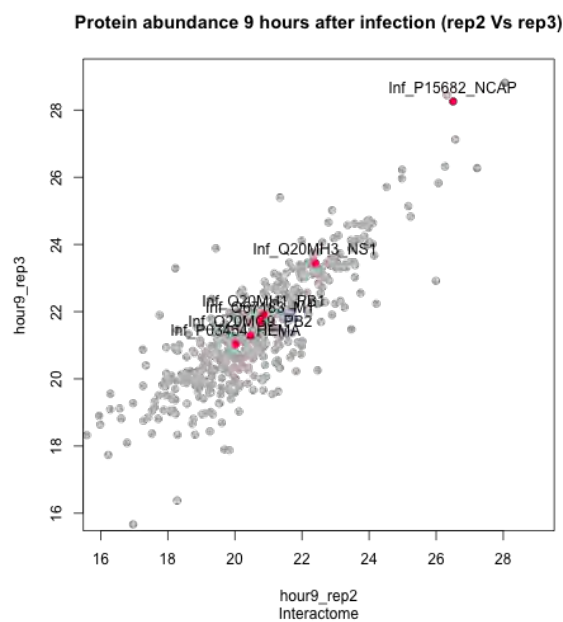
plot(proteins_int[,4], proteins_int[,7], col='grey', pch=19,xlim=c(16,29), ylim=c(16,29),
      xlab='hour9_rep1', ylab='hour9_rep2')
points(proteins_int[I,4], proteins_int[I,7], col='red', pch=19)
text(proteins_int[I,4], proteins_int[I,7], labels=proteins_int$ENSGid[I], cex= 1, pos=3)
title(main='Protein abundance 9 hours after infection (rep1 Vs rep2)' ,sub='Interactome' )
```



```
plot(proteins_int[,4], proteins_int[,10], col='grey', pch=19, xlim=c(16,29), ylim=c(16,29),
      xlab='hour9_rep1', ylab='hour9_rep3')
points(proteins_int[I,4], proteins_int[I,10], col='red', pch=19)
text(proteins_int[I,4], proteins_int[I,10], labels=proteins_int$ENSGid[I], cex= 1, pos=3)
title(main='Protein abundance 9 hours after infection (rep1 Vs rep3)' ,sub='Interactome' )
```



```
plot(proteins_int[,7], proteins_int[,10], col='grey', pch=19, xlim=c(16,29), ylim=c(16,29),
      xlab='hour9_rep2', ylab='hour9_rep3')
points(proteins_int[I,7], proteins_int[I,10], col='red', pch=19)
text(proteins_int[I,7], proteins_int[I,10], labels=proteins_int$ENSGid[I], cex= 1, pos=3)
title(main='Protein abundance 9 hours after infection (rep2 Vs rep3)' ,sub='Interactome' )
```



3 Differential t test in protein level

3.1 Moderated t-test for set enrichment

The aggregated mean intensity values for each protein are tested for enrichment between different states of infection, using a moderated t-test as implemented in the Bioconductor-package limma. p-values are corrected for multiple testing by the method of Benjamini-Hochberg.

```

dirTtest='t_test'
dir.create(file.path(dirTtest), showWarnings = FALSE)

##Interactome
Intensities = as.matrix(proteins_int[,grepl("hour",colnames(proteins_int)) |
                        grepl("mock",colnames(proteins_int))])

cond = sapply(strsplit(colnames(Intensities), split="_"),
              function(x) { x[1] })

sample_combi=combn(sample_names,2,simplify=TRUE)
colnames(sample_combi) =
  apply(sample_combi,2,
        function(x) {
          paste('diff',paste(x,collapse='_'),sep='_')
        })

diff_table_int = diff_intensities_int = list()
for (s in colnames(sample_combi)) {

  sample1 = Intensities[,cond == sample_combi[1,s]]
  sample2 = Intensities[,cond == sample_combi[2,s]]
  X =sample1-sample2
  diff_intensities_int[[s]] = X
  fit=eBayes(lmFit(X))
  fit$p.adj=p.adjust(fit$p.value,method="BH")
  dt = data.frame(ENSGid = proteins_int$ENSGid,
                  symbol = proteins_int$symbol,
                  Know_RBP = proteins_int$Know_RBP,
                  log2FC = fit$coefficients[,1],
                  p.value = fit$p.value[,1],
                  p.adj = fit$p.adj,
                  stringsAsFactors=FALSE)

  dt$sig=''
  sum(dt$sig=='')
  dt$sig[dt$p.adj<0.1]='*'
  dt$sig[dt$p.adj<0.01]='**'
  dt=dt[,c(1,2,3,4,5,6,7)]
  dt$CI_L=as.numeric(fit$coefficients-
                    dt(0.975, fit$df.prior+fit$df.residual) *
                    fit$stdev.unscaled* sqrt(fit$s2.post))
  dt$CI_H=as.numeric(fit$coefficients+
                    dt(0.975, fit$df.prior+fit$df.residual) *
                    fit$stdev.unscaled* sqrt(fit$s2.post))
  diff_table_int[[s]] = dt
}

####wholecell
Intensities = as.matrix(proteins_who[,grepl("hour",colnames(proteins_who)) |
                        grepl("mock",colnames(proteins_who))])

diff_table_who = diff_intensities_who = list()
for (s in colnames(sample_combi)) {
  sample1 = Intensities[,cond == sample_combi[1,s]]
  sample2 = Intensities[,cond == sample_combi[2,s]]
  X =sample1-sample2
  diff_intensities_who[[s]] = X
  fit=eBayes(lmFit(X))
  fit$p.adj=p.adjust(fit$p.value,method="BH")

```

```

dt = data.frame(ENSGid = proteins_who$ENSGid,
                symbol = proteins_who$symbol,
                Know_RBP = proteins_who$Know_RBP,
                log2FC = fit$coefficients[,1],
                p.value = fit$p.value[,1],
                p.adj = fit$p.adj,
                stringsAsFactors=FALSE)

dt$sig=''
sum(dt$sig=='')
dt$sig[dt$p.adj<0.1]='*'
dt$sig[dt$p.adj<0.01]='**'
dt=dt[,c(1,2,3,4,5,6,7)]
dt$CI_L=as.numeric(fit$coefficients-
                  dt(0.975, fit$df.prior+fit$df.residual) *
                  fit$stdev.unscaled* sqrt(fit$s2.post))
dt$CI_H=as.numeric(fit$coefficients+
                  dt(0.975, fit$df.prior+fit$df.residual) *
                  fit$stdev.unscaled* sqrt(fit$s2.post))
diff_table_who[[s]] = dt
}

#####interactome and whole cell lysate
Intensities = as.matrix(proteins[,grepl("hour",colnames(proteins)) |
                        grepl("mock",colnames(proteins))])

diff_table = diff_intensities = list()
for (s in colnames(sample_combi)) {
  sample1 = Intensities[,cond == sample_combi[1,s]]
  sample2 = Intensities[,cond == sample_combi[2,s]]
  X =sample1-sample2
  diff_intensities[[s]] = X
  fit=eBayes(lmFit(X))
  fit$p.adj=p.adjust(fit$p.value,method="BH")
  dt = data.frame(ENSGid = proteins$ENSGid,
                  symbol = proteins$symbol,
                  Know_RBP = proteins$Know_RBP,
                  log2FC = fit$coefficients[,1],
                  p.value = fit$p.value[,1],
                  p.adj = fit$p.adj,
                  stringsAsFactors=FALSE)

  dt$sig=''
  dt$sig[dt$p.adj<0.1]='*'
  dt$sig[dt$p.adj<0.01]='**'
  dt=dt[,c(1,2,3,4,5,6,7)]
  dt$CI_L=as.numeric(fit$coefficients-
                    dt(0.975, fit$df.prior+fit$df.residual) *
                    fit$stdev.unscaled* sqrt(fit$s2.post))
  dt$CI_H=as.numeric(fit$coefficients+
                    dt(0.975, fit$df.prior+fit$df.residual) *
                    fit$stdev.unscaled* sqrt(fit$s2.post))
  diff_table[[s]] = dt
}

```

The t-test result tables are saved.

```

saveRDS(diff_table_int, file=file.path("t_test/diff_table_int.rda"))
saveRDS(diff_table_who, file=file.path("t_test/diff_table_who.rda"))
saveRDS(diff_table, file=file.path("t_test/diff_table.rda"))
saveRDS(diff_intensities_int, file=file.path("t_test/diff_intensities_int.rda"))

```

Comparison of whole cell lysate protein intensity in cells infected with Influenza virus in different states11

```

saveRDS(diff_intensities_who, file=file.path("t_test/diff_intensities_who.rda"))
saveRDS(diff_intensities, file=file.path("t_test/diff_intensities.rda"))

for (s in names(diff_table_int)) {
  write.table(diff_table_int[[s]][order(diff_table_int[[s]]$p.adj,
                                       diff_table_int[[s]]$log2FC),],
             file=file.path("t_test",paste0(s,'_int.txt')),
             sep="\t",row.names =FALSE, quote=FALSE)
  page = openPage(file.path("t_test",paste0(s,"_int.html")),link.css="hwriter.css")
  hwrite(paste0("Result list ",s),heading=1,page=page)
  table=diff_table_int[[s]][order(diff_table_int[[s]]$p.adj, diff_table_int[[s]]$log2FC),]
  row.names(table)=NULL
  hwrite(format(table,digits=4),
        page=page)
  closePage(page,splash=FALSE)
}

for (s in names(diff_table_who)) {
  write.table(diff_table_who[[s]][order(diff_table_who[[s]]$p.adj,
                                       diff_table_who[[s]]$log2FC),],
             file=file.path("t_test",paste0(s,'_who.txt')),
             sep="\t",row.names =FALSE, quote=FALSE)
  page = openPage(file.path("t_test",paste0(s,"_who.html")),link.css="hwriter.css")
  hwrite(paste0("Result list ",s),heading=1,page=page)
  table=diff_table_who[[s]][order(diff_table_who[[s]]$p.adj, diff_table_who[[s]]$log2FC),]
  row.names(table)=NULL
  hwrite(format(table,digits=4),
        page=page)
  closePage(page,splash=FALSE)
}

for (s in names(diff_table)) {
  write.table(diff_table[[s]][order(diff_table[[s]]$p.adj,
                                   diff_table[[s]]$log2FC),],
             file=file.path("t_test",paste0(s,'.txt')),
             sep="\t",row.names =FALSE, quote=FALSE)
  page = openPage(file.path("t_test",paste0(s,".html")),link.css="hwriter.css")
  hwrite(paste0("Result list ",s),heading=1,page=page)
  table=diff_table[[s]][order(diff_table[[s]]$p.adj, diff_table[[s]]$log2FC),]
  row.names(table)=NULL
  hwrite(format(table,digits=4),
        page=page)
  closePage(page,splash=FALSE)
}

diff_table_int=readRDS( file=file.path("t_test/diff_table_int.rda"))
diff_table_who=readRDS(file=file.path("t_test/diff_table_who.rda"))
diff_table=readRDS(file=file.path("t_test/diff_table.rda"))
diff_intensities_int=readRDS( file=file.path("t_test/diff_intensities_int.rda"))
diff_intensities_who=readRDS( file=file.path("t_test/diff_intensities_who.rda"))
diff_intensities=readRDS( file=file.path("t_test/diff_intensities.rda"))

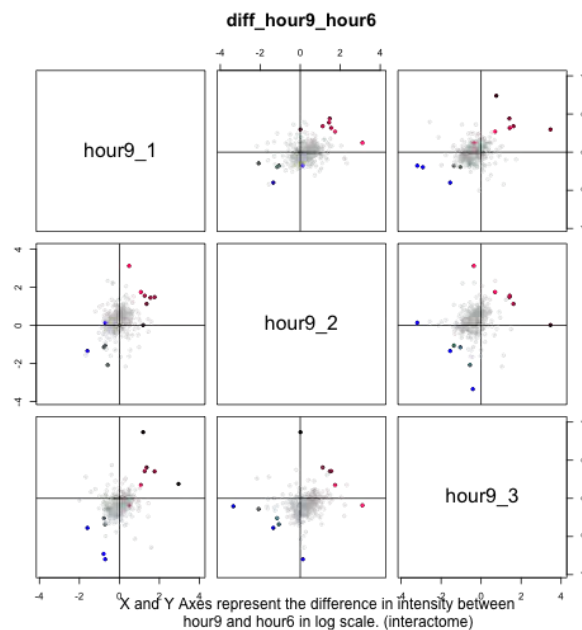
for (s in colnames(sample_combi)) {
  comparison=diff_table_int[[s]]
  col = ifelse(comparison$p.adj <= 0.2, "orange", "gray")
  col[which(comparison$p.adj <= 0.1)] = "firebrick1"
}

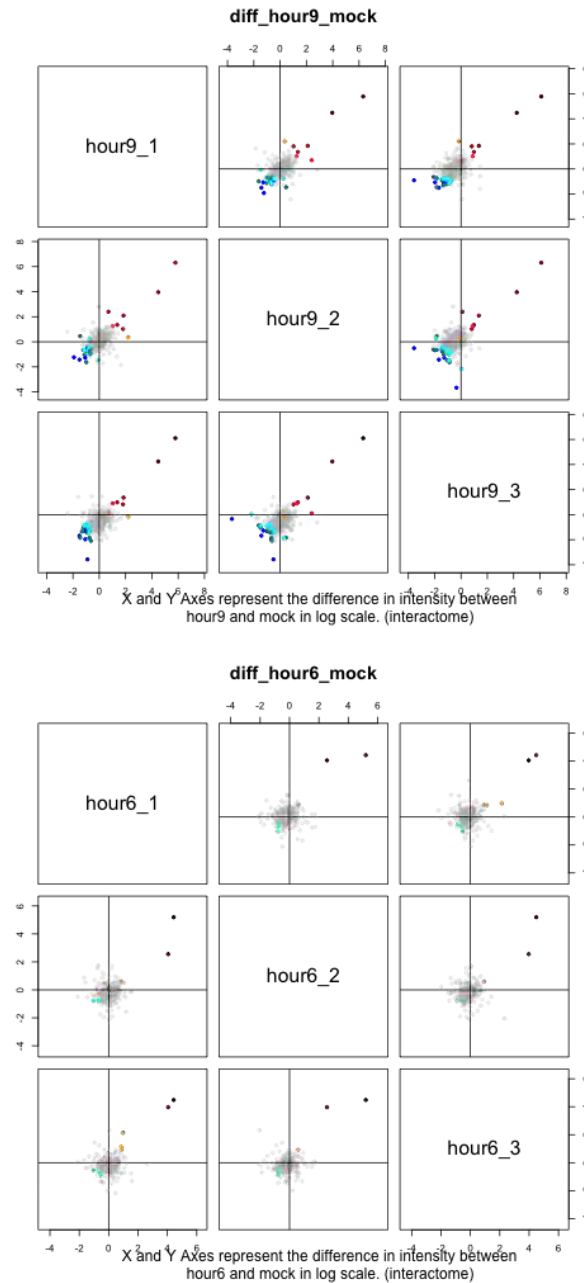
```

Comparison of whole cell lysate protein intensity in cells infected with Influenza virus in different states12

```
col[which(comparison$p.adj <= 0.01)] = "firebrick"
col[which(comparison$p.adj <= 0.2 & comparison$log2FC < 0)] = "turquoise1"
col[which(comparison$p.adj <= 0.1 & comparison$log2FC < 0)] = "turquoise4"
col[which(comparison$p.adj <= 0.01 & comparison$log2FC < 0)] = "blue"
col[col=="gray"]=adjustcolor('gray', alpha=0.2)
my_line <- function(x,y,...){
  points(x,y,col=col, ...)
  abline(h=0 ,v=0 , ...)
}

Int=diff_intensities_int[[s]]
##Add dummy data as a control of max and min range of the plot
MX=max(Int, na.rm=TRUE)+0.5
MN=min(Int, na.rm=TRUE)-0.5
nonset=matrix(rep(NA, 18), ncol=3)
nonset[1:2,1]=c(MX, MN)
nonset[3:4,2]=c(MX, MN)
nonset[5:6,3]=c(MX, MN)
colnames(nonset)=colnames(Int)
Int=rbind(Int, nonset)
pairs(Int,pch=20, main=s,
panel=my_line)
title(sub = paste0('X and Y Axes represent the difference in intensity between
', strsplit(s, '_')[[1]][2], ' and ', strsplit(s, '_')[[1]][3], ' in log scale. (interactome)'))
}
```





```
for (s in colnames(sample_combi)) {
  comparison=diff_table_who[[s]]
  col = ifelse(comparison$p.adj <= 0.2, "orange", "gray")
  col[which(comparison$p.adj <= 0.1)] = "firebrick1"
  col[which(comparison$p.adj <= 0.01)] = "firebrick"
  col[which(comparison$p.adj <= 0.2 & comparison$log2FC < 0)] = "turquoise1"
  col[which(comparison$p.adj <= 0.1 & comparison$log2FC < 0)] = "turquoise4"
  col[which(comparison$p.adj <= 0.01 & comparison$log2FC < 0)] = "blue"
  col[col=="gray"]=adjustcolor('gray', alpha=0.2)
  my_line <- function(x,y,...){
    points(x,y,col=col, ...)
    abline(h=0 ,v=0 , ...)
  }

  Int=diff_intensities_who[[s]]

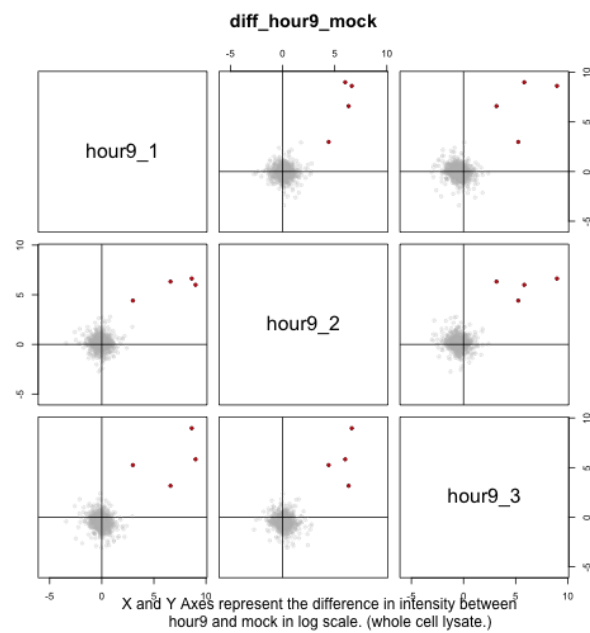
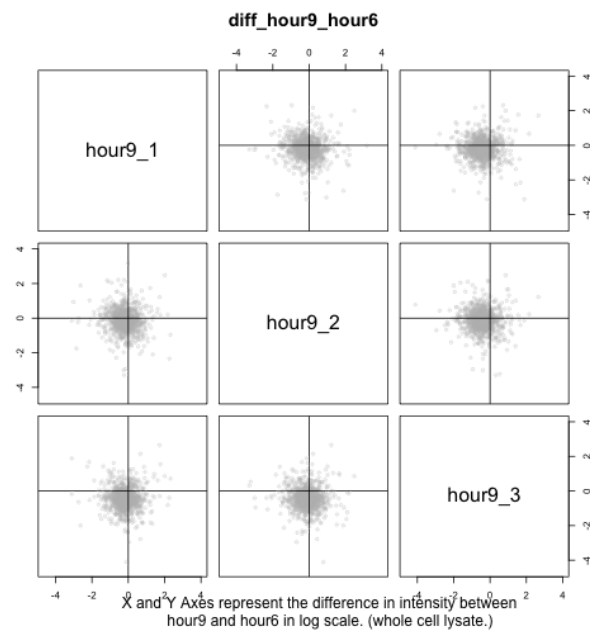
  ##Add dummy data as a control of max and min range of the plot
```

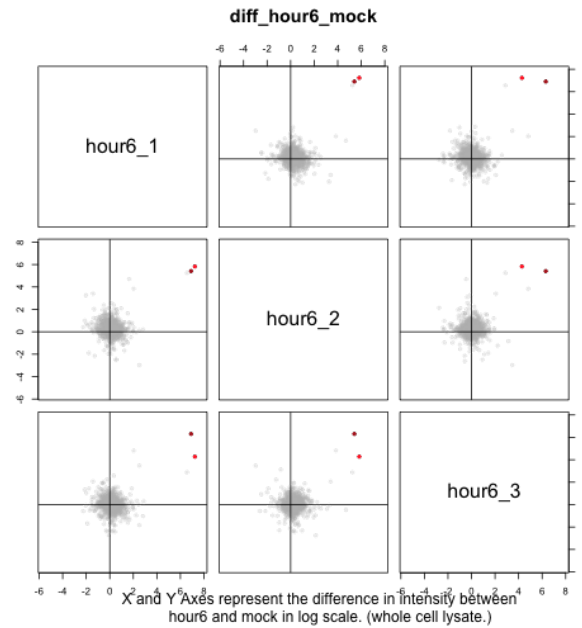
Comparison of whole cell lysate protein intensity in cells infected with Influenza virus in different states14

```

MX=max(Int, na.rm=TRUE)+0.5
MN=min(Int, na.rm=TRUE)-0.5
nonset=matrix(rep(NA, 18), ncol=3)
nonset[1:2,1]=c(MX, MN)
nonset[3:4,2]=c(MX, MN)
nonset[5:6,3]=c(MX, MN)
colnames(nonset)=colnames(Int)
Int=rbind(Int, nonset)
pairs(Int,pch=20, main=s,
panel=my_line)
title(sub = paste0('X and Y Axes represent the difference in intensity between
', strsplit(s, '_')[[1]][2], ' and ', strsplit(s, '_')[[1]][3], ' in log scale. (whole cell lysate.)

```





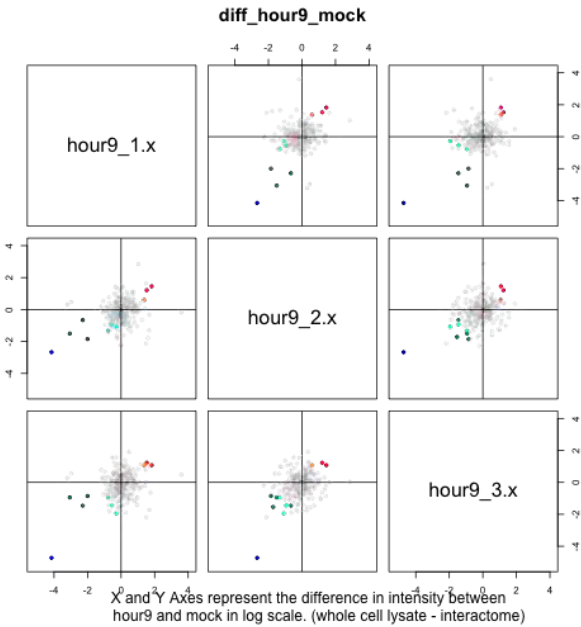
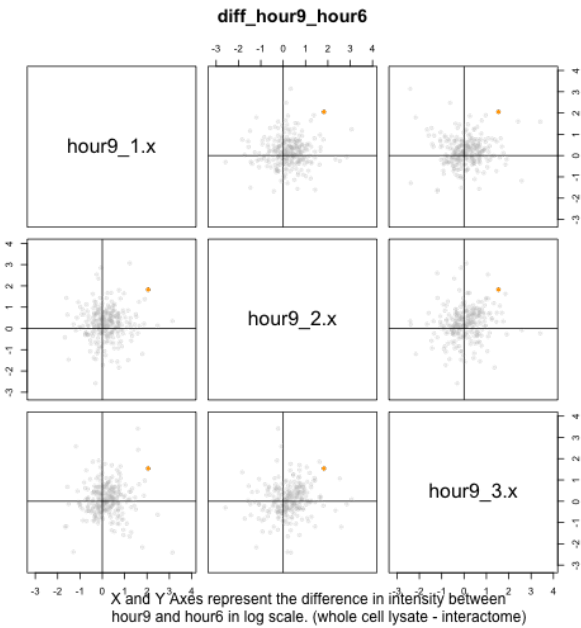
```

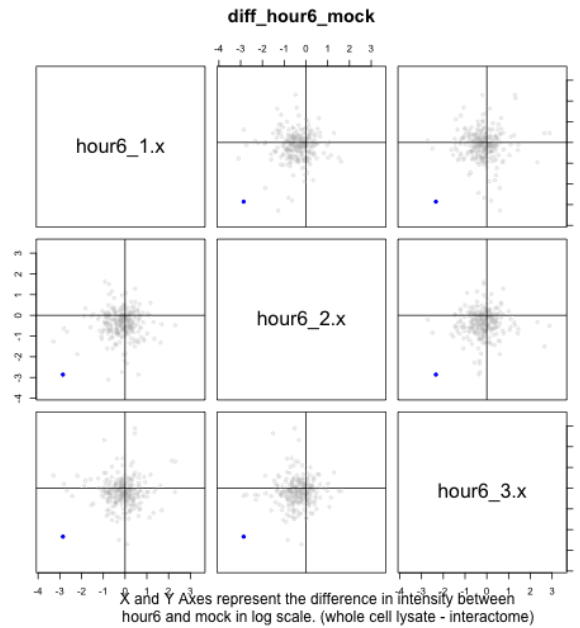
for (s in colnames(sample_combi)) {
  comparison=diff_table[[s]]
  col = ifelse(comparison$p.adj <= 0.2, "orange", "gray")
  col[which(comparison$p.adj <= 0.1)] = "firebrick1"
  col[which(comparison$p.adj <= 0.01)] = "firebrick"
  col[which(comparison$p.adj <= 0.2 & comparison$log2FC < 0)] = "turquoise1"
  col[which(comparison$p.adj <= 0.1 & comparison$log2FC < 0)] = "turquoise4"
  col[which(comparison$p.adj <= 0.01 & comparison$log2FC < 0)] = "blue"
  col[col=="gray"]=adjustcolor('gray', alpha=0.2)
  my_line <- function(x,y,...){
    points(x,y,col=col, ...)
    abline(h=0 , v=0 , ...)
  }

  Int=diff_intensities[[s]]

  ##Add dummy data as a control of max and min range of the plot
  MX=max(Int, na.rm=TRUE)+0.5
  MN=min(Int, na.rm=TRUE)-0.5
  nonset=matrix(rep(NA, 18), ncol=3)
  nonset[1:2,1]=c(MX, MN)
  nonset[3:4,2]=c(MX, MN)
  nonset[5:6,3]=c(MX, MN)
  colnames(nonset)=colnames(Int)
  Int=rbind(Int, nonset)
  pairs(Int,pch=20, main=s,
  panel=my_line)
  title(sub = paste0('X and Y Axes represent the difference in intensity between
  ', strsplit(s, '_')[[1]][2], ' and ', strsplit(s, '_')[[1]][3], ' in log scale. (whole cell lysate -
  }

```





3.2 Volcano plots

```
voldir=file.path('VolcanoPlot')
dir.create(voldir, showWarnings = FALSE)

for (s in colnames(sample_combi)) {
  volData = diff_table_int[[s]]
  volData = volData[!is.na(volData$p.adj),]
  col = ifelse(volData$p.adj <= 0.2, "orange", "gray")
  col[which(volData$p.adj <= 0.1)] = "firebrick1"
  col[which(volData$p.adj <= 0.01)] = "firebrick"
  col[which(volData$p.adj <= 0.2 & volData$log2FC < 0)] = "turquoise1"
  col[which(volData$p.adj <= 0.1 & volData$log2FC < 0)] = "turquoise4"
  col[which(volData$p.adj <= 0.01 & volData$log2FC < 0)] = "blue"
  col[col=='gray']=adjustcolor('gray', alpha=0.2)

  g = ggplot(volData, aes(log2FC,-log10(p.adj)))+
    geom_point(size=2, color=col)+
    theme(panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          panel.background = element_blank(),
          axis.line = element_line(colour = "black"))+
    geom_text(data=subset(volData, abs(log2FC) > 2 ),
              aes(log2FC,-log10(p.adj),label=symbol),
              angle=90,hjust = 0,nudge_y = 0.05) +
    xlim(c(-6, 6)) + ylim(c(0, 3)) +
    xlab("log2 fold change") + ylab("-log10 p-value") +
    labs(title = paste('Protein intensity of ',
                      sample_combi[1,s], ' comparing to ',
                      sample_combi[2,s], '(interactome)',sep=''))
  ggsave(paste0(voldir,'/', s, '.pdf'))
  ggsave(paste0(voldir,'/', s, '.png'))
}

## Saving 7 x 7 in image
## Warning: Removed 7 rows containing missing values (geom_point).
```

```

## Saving 7 x 7 in image
## Warning: Removed 7 rows containing missing values (geom_point).
## Saving 7 x 7 in image
## Warning: Removed 9 rows containing missing values (geom_point).
## Warning: Removed 6 rows containing missing values (geom_text).
## Saving 7 x 7 in image
## Warning: Removed 9 rows containing missing values (geom_point).
## Warning: Removed 6 rows containing missing values (geom_text).
## Saving 7 x 7 in image
## Warning: Removed 3 rows containing missing values (geom_point).
## Warning: Removed 3 rows containing missing values (geom_text).
## Saving 7 x 7 in image
## Warning: Removed 3 rows containing missing values (geom_point).
## Warning: Removed 3 rows containing missing values (geom_text).

#
# for (s in colnames(sample_combi)) {
#   volData = diff_table_who[[s]]
#   volData = volData[!is.na(volData$p.adj),]
#   col = ifelse(volData$p.adj <= 0.2, "orange", "gray")
#   col[which(volData$p.adj <= 0.1)] = "firebrick1"
#   col[which(volData$p.adj <= 0.01)] = "firebrick"
#   col[which(volData$p.adj <= 0.2 & volData$log2FC < 0)] = "turquoise1"
#   col[which(volData$p.adj <= 0.1 & volData$log2FC < 0)] = "turquoise4"
#   col[which(volData$p.adj <= 0.01 & volData$log2FC < 0)] = "blue"
#   col[col=="gray"]=adjustcolor('gray', alpha=0.2)
#
#   g = ggplot(volData, aes(log2FC, -log10(p.adj)))+
#     geom_point(size=2, color=col)+
#     theme(panel.grid.major = element_blank(),
#           panel.grid.minor = element_blank(),
#           panel.background = element_blank(),
#           axis.line = element_line(colour = "black"))+
#     geom_text(data=subset(volData, abs(log2FC) > 2),
#               aes(log2FC, -log10(p.adj), label=symbol),
#               angle=90, hjust = 0, nudge_y = 0.05) +
#     xlim(c(-6, 6)) + ylim(c(0, 3)) +
#     xlab("log2 fold change") + ylab("-log10 p-value") +
#     labs(title = paste('Protein intensity of ',
#                       sample_combi[1,s], ' comparing to ',
#                       sample_combi[2,s], '(whole cell lysate)', sep=''))
#
#   ggsave(paste0(voldir, '/', s, '.pdf'))
#   ggsave(paste0(voldir, '/', s, '.png'))
# }
#
# for (s in colnames(sample_combi)) {
#   volData = diff_table[[s]]
#   volData = volData[!is.na(volData$p.adj),]
#   col = ifelse(volData$p.adj <= 0.2, "orange", "gray")
#   col[which(volData$p.adj <= 0.1)] = "firebrick1"

```

```
# col[which(volData$p.adj <= 0.01)] = "firebrick"
# col[which(volData$p.adj <= 0.2 & volData$log2FC < 0)] = "turquoise1"
# col[which(volData$p.adj <= 0.1 & volData$log2FC < 0)] = "turquoise4"
# col[which(volData$p.adj <= 0.01 & volData$log2FC < 0)] = "blue"
# col[col=='gray']=adjustcolor('gray', alpha=0.2)
#
# g = ggplot(volData, aes(log2FC,-log10(p.adj)))+
#   geom_point(size=2, color=col)+
#   theme(panel.grid.major = element_blank(),
#         panel.grid.minor = element_blank(),
#         panel.background = element_blank(),
#         axis.line = element_line(colour = "black"))+
#   geom_text(data=subset(volData, abs(log2FC) > 2 ),
#             aes(log2FC,-log10(p.adj),label=symbol),
#             angle=90,hjust = 0,nudge_y = 0.05) +
#   xlim(c(-6, 6)) + ylim(c(0, 3)) +
#   xlab("log2 fold change") + ylab("-log10 p-value") +
#   labs(title = paste('Protein intensity of ',
#                     sample_combi[1,s], ' comparing to ',
#                     sample_combi[2,s], sep=' '))
# ggsave(paste0(voldir, '/', s, '.pdf'))
# ggsave(paste0(voldir, '/', s, '.png'))
# }
```

```
page=openPage(file.path('t_test','int_index.html'),
              link.css='hwriter.css')
file.copy(system.file('images','hwriter.css', package='hwriter'),
          file.path('t_test','hwriter.css'))

## [1] TRUE

hwriteSidebar(position = c("left"),
              link = c("../mainmenu.html"),
              size = 15, margin = 1, page = page)

hwrite('Protein abundance in interactome', heading=1,page=page)

hwrite('hour 9 to hour 6', heading=3,page=page, br=TRUE)
hwrite("Hitlist ", link="diff_hour9_hour6_int.html", page=page)
hwrite(paste("[Download text file:]"),
      link='diff_hour9_hour6_int.txt', page=page, br=TRUE)
M = matrix('t_testPairs-1.png',ncol=1)
L = matrix('t_testPairs-1.pdf',ncol=1)
hwriteImage(M, link=L, width=700, height=700, page=page)

hwrite('hour 9 to mock', heading=3,page=page, br=TRUE)
hwrite("Hitlist ", link="diff_hour9_mock_int.html", page=page)
hwrite(paste("[Download text file:]"),
      link='diff_hour9_mock_int.txt', page=page, br=TRUE)
M = matrix('t_testPairs-2.png',ncol=1)
L = matrix('t_testPairs-2.pdf',ncol=1)
hwriteImage(M, link=L, width=700, height=700, page=page)

hwrite('hour 6 to mock', heading=3,page=page)
hwrite("Hitlist ", link="diff_hour6_mock_int.html", page=page)
hwrite(paste("[Download text file:]"),
      link='diff_hour6_mock_int.txt', page=page, br=TRUE)
M = matrix('t_testPairs-3.png',ncol=1)
L = matrix('t_testPairs-3.pdf',ncol=1)
```

```

hwriteImage(M, link=L, width=700, height=700, page=page)
closePage(page,splash=FALSE)

###Interactome - whole cell lysate
page=openPage(file.path('t_test','dif_index.html'),
               link.css='hwriter.css')
file.copy(system.file('images','hwriter.css', package='hwriter'),
           file.path('t_test','hwriter.css'))

## [1] FALSE

hwriteSidebar(position = c("left"),
              link = c("../mainmenu.html"),
              size = 15, margin = 1, page = page)

hwrite('Protein abundance in interactome substracted by whole cell lysate', heading=1,page=page)
hwrite('hour 9 to hour 6', heading=3,page=page, br=TRUE)
hwrite("Hitlist ", link="dif_hour9_hour6.html", page=page)
hwrite(paste("[Download text file:]"),
       link='diff_hour9_hour6.txt', page=page, br=TRUE)
M = matrix('t_testPairs-7.png',ncol=1)
L = matrix('t_testPairs-7.pdf',ncol=1)
hwriteImage(M, link=L, width=700, height=700, page=page, br=TRUE)
hwrite('hour 9 to mock', heading=3,page=page)
hwrite("Hitlist ", link="dif_hour9_mock.html", page=page)
hwrite(paste("[Download text file:]"),
       link='diff_hour9_mock.txt', page=page, br=TRUE)
M = matrix('t_testPairs-8.png',ncol=1)
L = matrix('t_testPairs-8.pdf',ncol=1)
hwriteImage(M, link=L, width=700, height=700, page=page, br=TRUE)
hwrite('hour 6 to mock', heading=3,page=page)
hwrite("Hitlist ", link="dif_hour6_mock.html", page=page)
hwrite(paste("[Download text file:]"),
       link='diff_hour6_mock.txt', page=page, br=TRUE)
M = matrix('t_testPairs-9.png',ncol=1)
L = matrix('t_testPairs-9.pdf',ncol=1)
hwriteImage(M, link=L, width=700, height=700, page=page)
closePage(page,splash=FALSE)

page=openPage(file.path('t_test','who_index.html'),
               link.css='hwriter.css')
file.copy(system.file('images','hwriter.css', package='hwriter'),
           file.path('t_test','hwriter.css'))

## [1] FALSE

hwriteSidebar(position = c("left"),
              link = c("../mainmenu.html"),
              size = 15, margin = 1, page = page)

hwrite('Protein abundance in whole cell lysate', heading=1,page=page)
hwrite('hour 9 to hour 6', heading=3,page=page, br=TRUE)
hwrite("Hitlist ", link="dif_hour9_hour6_who.html", page=page)
hwrite(paste("[Download text file:]"),
       link='diff_hour9_hour6_who.txt', page=page, br=TRUE)
M = matrix('t_testPairs-4.png',ncol=1)
L = matrix('t_testPairs-4.pdf',ncol=1)
hwriteImage(M, link=L, width=700, height=700, page=page)

```



```

#           proteinLevel = a$hour6_mock[isFinite],
#           sig=sig[isFinite]))
#   dat <- rbind(dat,
#               data.frame(time = "09:00",
#                           gene = factor(a$protein[isFinite],
#                                         levels=gene),
#                           proteinLevel = a$hour9_mock[isFinite],
#                           sig=sig[isFinite]))
#   #dat$flwd = ifelse(dat$sig == "sig",1,1)
#   if (sum(sig) > 0) {
#     g = ggplot(dat, aes(time, proteinLevel, group=gene)) +
#       geom_line(col="gray") +
#       geom_line(data=dat[which(dat$sig),], aes(colour="blue"), size=2) +
#       geom_point(data=dat[which(dat$sig & dat$time != "00:00"),],
#                 shape=19, size=2, colour="red")
#     pdf(file.path(dirlinechart, paste0("LinePlot-", i, "-", j, ".pdf")))
#     print(g)
#     svg = gridSvgLinePlot(g)
#     gridSvgLinePlot
#     dev.off()
#     files = c(files, paste0("LinePlot-", i, "-", j, ".html"))
#     page = openPage(file.path(
#       dirlinechart, paste0("LinePlot-", i, "-", j, ".html")),
#                     head =
#                       "<script src=\"http://d3js.org/d3.v3.js\"></script>",
#                       link.css = "hwriter.css")
#     hwriteSidebar(position = c("left"),
#                   link = c("../mainmenu.html"),
#                   size = 15, margin = 1, page = page)
#     hwrite(paste0(j, "-regulated proteins in ", sample_combi[1, i],
#                   " compare to ", sample_combi[2, i], " (padj=0.1)",
#                   page=page, heading=1)
#     hwrite(svg, table=FALSE, page=page)
#     closePage(page, splash=FALSE)
#   }
# }
# }
#
# page=openPage(file.path(dirlinechart, 'index.html'),
#               link.css='hwriter.css')
# file.copy(system.file('images', 'hwriter.css', package='hwriter'),
#           file.path(dirlinechart, 'hwriter.css'))
# hwriteSidebar(position = c("left"),
#               link = c("../mainmenu.html"),
#               size = 15, margin = 1, page = page)
#
# hwrite('Line charts', heading=1, page=page)
# M = matrix(files, ncol=1)
# L = matrix(files, ncol=1)
# hwrite(M, link=L, page=page)
#
# closePage(page, splash=FALSE)

```

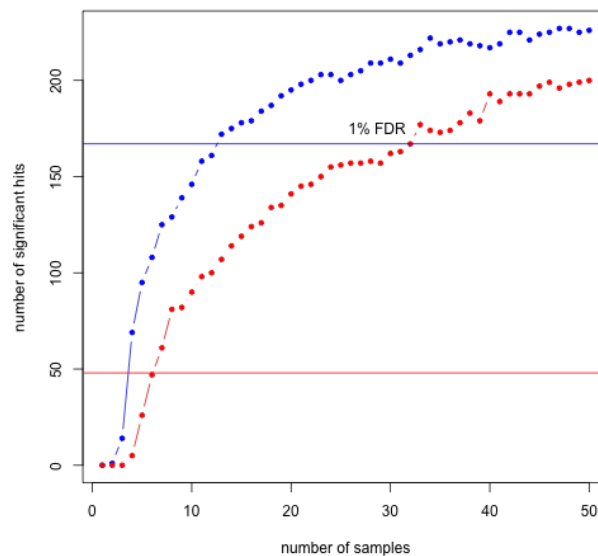
3.4 Stistical power of increased number of simulated replicates


```

X = proteins[,c(4,7,10)] - proteins[,c(6,9,12)]
m = apply(X, 1, mean, na.rm=TRUE)
s = apply(X, 1, sd, na.rm=TRUE)
I = which(!is.na(m) & !is.na(s))
m = m[I]
s = s[I]
X = matrix(NA, nrow=length(m), ncol=50)
for (j in seq_along(m)) {
  X[j,] = rnorm(50, mean=m[j], sd=s[j])
}

res = matrix(NA, nrow=50, ncol=2)
for (i in 1:50) {
  fit = eBayes(lmFit(X[,1:i]))
  padj = p.adjust(fit$p.value, method="BH")
  res[i,1] = sum(padj <= 0.01)
  res[i,2] = sum(padj <= 0.1)
}
plot(res[,2], col="blue", type="b", pch=20,
      xlab="number of samples", ylab="number of significant hits")
lines(res[,1], col="red", type="b", pch=20)
legend(10, 265, legend = '10% FDR', col='red',
       border = FALSE, bty='n')
legend(23, 185, legend = '1% FDR', col='red',
       border = FALSE, bty='n')
abline(h=48, col="red")
abline(h=167, col="blue")

```



```
## [1] FALSE
```

3.5 Heatmat for protein intensity profiles

Heatmap for proteins have no missing measurements, normalized by replicates

```

dirheatmap='heatmap'
dir.create(file.path(dirheatmap), showWarnings = FALSE)
a = data.frame(hour6_mock = diff_table[["diff_hour6_mock"]])$log2FC,

```

```

        hour9_mock = diff_table[["diff_hour9_mock"]]$log2FC,
        protein = diff_table[["diff_hour9_mock"]]$symbol)
isFinite = !is.na(a$hour6) & !is.na(a$hour9)
isSig = apply(sapply(diff_table, function(x) { x$sig != "" })),1,any)

a = a[isFinite & isSig,]
X = as.matrix(a[,1:2])
row.names(X) = a$protein
X[X > 3] = 3
X[X < -3] = -3
r = max(abs(X))
pheatmap(t(X), cluster_rows = FALSE, fontsize_col = 6,
         color = colorRampPalette(rev(brewer.pal(n = 7, name = 'RdBu')))(256),
         breaks = seq(-r,r,length.out=257),cellwidth = 5, cellheight = 20,
         filename = file.path(dirheatmap, "heatmap.pdf"))
pheatmap(t(X), cluster_rows = FALSE, fontsize_col = 6,
         color = colorRampPalette(rev(brewer.pal(n = 7, name = 'RdBu')))(256),
         breaks = seq(-r,r,length.out=257),cellwidth = 5, cellheight = 20,
         filename = file.path(dirheatmap, "heatmap.png"))

```

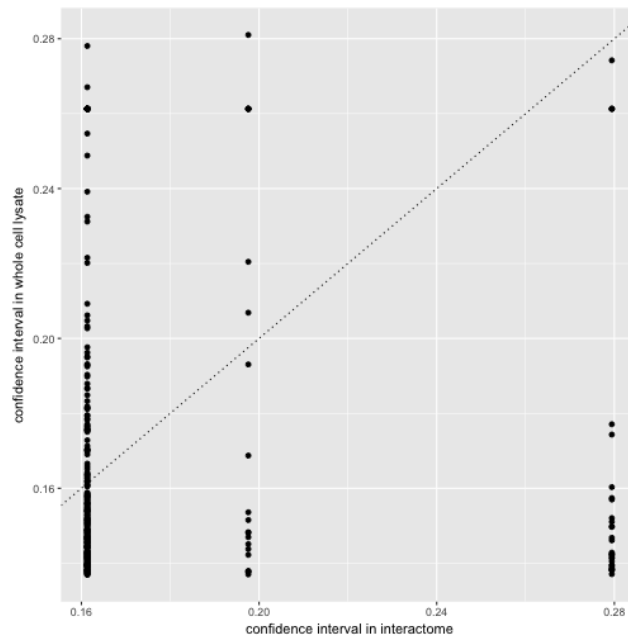
3.6 confidence interval comparison

```

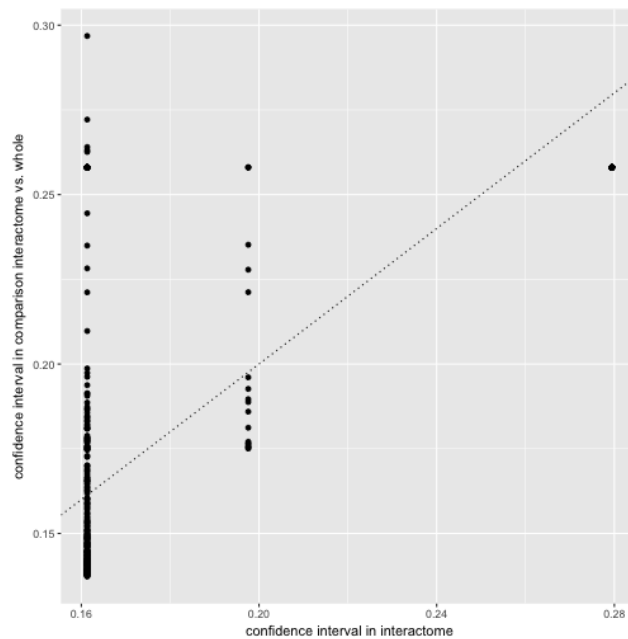
A = read.table("t_test/diff_hour9_mock_int.txt",
              sep="\t", header=TRUE, stringsAsFactors = FALSE)
B = read.table("t_test/diff_hour9_mock_who.txt",
              sep="\t", header=TRUE, stringsAsFactors = FALSE)
C = read.table("t_test/diff_hour9_mock.txt",
              sep="\t", header=TRUE, stringsAsFactors = FALSE)
A$CI = A$CI_H - A$CI_L
B$CI = B$CI_H - B$CI_L
C$CI = C$CI_H - C$CI_L

m = match(A$ENSGid, B$ENSGid)
qplot(A$CI, B$CI[m], xlab="confidence interval in interactome",
      ylab="confidence interval in whole cell lysate")+
geom_abline(slope = 1, intercept = 0, linetype='dotted')
## Warning: Removed 299 rows containing missing values (geom_point).

```

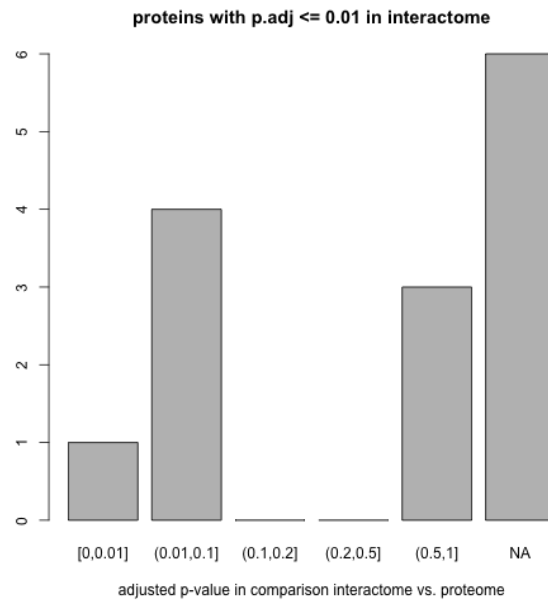


```
m = match(A$ENSGid, C$ENSGid)
qplot(A$CI, C$CI[m], xlab="confidence interval in interactome",
      ylab="confidence interval in comparison interactome vs. whole")+
geom_abline(slope = 1, intercept = 0, linetype='dotted')
## Warning: Removed 303 rows containing missing values (geom_point).
```

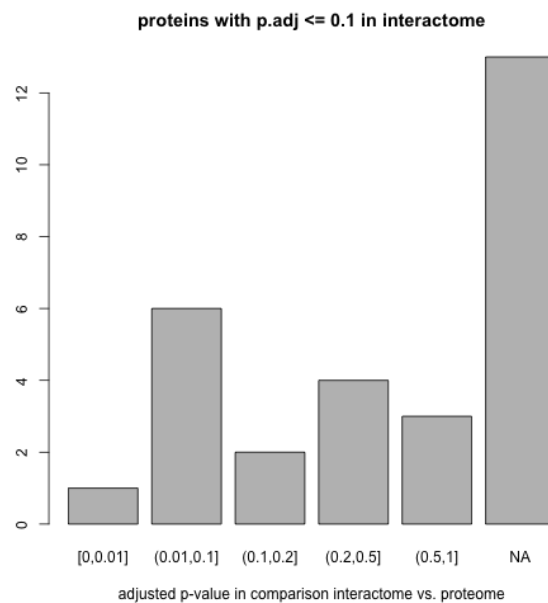


```
I = which(A$p.adj <= 0.01)
J = which(A$p.adj <= 0.1)

t = table(cut(C$p.adj[m][I], breaks = c(0, 0.01, 0.1, 0.2, 0.5, 1),
             include.lowest = TRUE))
t["NA"] = sum(is.na(C$p.adj[m][I]))
barplot(as.vector(t), names.arg = names(t),
       xlab="adjusted p-value in comparison interactome vs. proteome",
       main="proteins with p.adj <= 0.01 in interactome")
```



```
t = table(cut(C$p.adj[m][J], breaks = c(0, 0.01, 0.1, 0.2, 0.5, 1),
             include.lowest = TRUE))
t["NA"] = sum(is.na(C$p.adj[m][J]))
barplot(as.vector(t), names.arg = names(t),
        xlab="adjusted p-value in comparison interactome vs. proteome",
        main="proteins with p.adj <= 0.1 in interactome")
```



3.7 Bar plots

```
bardir=file.path('BarPlot')
dir.create(bardir, showWarnings = FALSE)
a = data.frame(hour6_mock = diff_table_int[["diff_hour6_mock"]][log2FC],
               hour9_mock = diff_table_int[["diff_hour9_mock"]][log2FC],
               protein = diff_table_int[["diff_hour9_mock"]][symbol])
isFinite = !is.na(a$hour6) & !is.na(a$hour9)
```

```

for (i in seq_len(ncol(sample_combi))) {
  for (j in c("up", "down")) {
    if (j == "up") {
      I = which(diff_table_int[[i]]$p.adj <= 0.1 & diff_table_int[[i]]$log2FC > 0 & isFinite)
    } else {
      I = which(diff_table_int[[i]]$p.adj <= 0.1 & diff_table_int[[i]]$log2FC < 0 & isFinite)
    }
    b=a[I,]
    aa=data.frame(time=rep(c('hour6_mock', 'hour9_mock'),
                          each=dim(b)[1]),
                  protein=rep(b$protein, 2),
                  values=c(b$hour6_mock, b$hour9_mock))
    g = ggplot(data=aa, aes(x=factor(protein, levels=b$protein[order(b$hour9_mock)]),
                          y=values,
                          fill=factor(time, levels=c('hour6_mock', 'hour9_mock')))) +
      geom_bar(stat='identity', position=position_dodge()) +
      scale_fill_hue(name="Time after infection") +
      xlab("Protein") + ylab("Intensity level") +
      ggtitle(paste0(j, "-regulated proteins in ", sample_combi[1, i],
                    " compare to ", sample_combi[2, i], " (padj=0.1)")) +
      theme_bw() +
      theme(legend.key.size = unit(0.5, "cm"),
            axis.text.x=element_text(size=5, angle=90, hjust=1, vjust=0.5))
    ggsave(plot = g, filename = file.path("BarPlot", paste0("BarPlot_", i, "_", j, ".pdf")),
           width = max(3+length(I)*5/60, 5))
    ggsave(plot = g, filename = file.path("BarPlot", paste0("BarPlot_", i, "_", j, ".png")),
           width = max(3+length(I)*5/60, 5))
  }
}

```

3.8 A semi-quantified analysis for protein enrichment

A semi-quantified analysis of the protein enrichment in different states of infection. Generating tables contain proteins which have detected intensity values in 3 or 2 of the replications in a time point but have only 1 or 0 detected intensity in another.

```

add_summary=function(proteinset){
  h18_index=grep('9', colnames(proteinset))
  h4_index=grep('6', colnames(proteinset))
  mock_index=grep('mock', colnames(proteinset))
  proteinset=cbind(
    proteinset, mock_total=apply(
      proteinset, 1, function(x) sum(!is.na(x[mock_index]))))
  proteinset=cbind(
    proteinset, hour6_total=apply(
      proteinset, 1, function(x) sum(!is.na(x[h4_index]))))
  proteinset=cbind(
    proteinset, hour9_total=apply(
      proteinset, 1, function(x) sum(!is.na(x[h18_index]))))
  proteinset
}

semi_dir=file.path('int_semi_quant')
dir.create(semi_dir, showWarnings = FALSE)
proteins_int=add_summary(proteins_int)
semi_hour9_mock_up=proteins_int[proteins_int$hour9_total-proteins_int$mock_total>=2,]
write.table(semi_hour9_mock_up,

```

```

        file=file.path(semi_dir,'semi_hour9_mock_up.txt'),
        row.names=FALSE)
semi_hour9_mock_down=proteins_int[proteins_int$mock_total-proteins_int$hour9_total>=2,]
write.table(semi_hour9_mock_down,
            file=file.path(semi_dir,'semi_hour9_mock_down.txt'),
            row.names=FALSE)
semi_hour6_mock_up=proteins_int[proteins_int$hour6_total-proteins_int$mock_total>=2,]
write.table(semi_hour6_mock_up,
            file=file.path(semi_dir,'semi_hour6_mock_up.txt'),
            row.names=FALSE)
semi_hour6_mock_down=proteins_int[proteins_int$mock_total-proteins_int$hour6_total>=2,]
write.table(semi_hour6_mock_down,
            file=file.path(semi_dir,'semi_hour6_mock_down.txt'),
            row.names=FALSE)
semi_hour9_hour6_up=proteins_int[proteins_int$hour9_total-proteins_int$hour6_total>=2,]
write.table(semi_hour9_hour6_up,
            file=file.path(semi_dir,'semi_hour9_hour6_up.txt'),
            row.names=FALSE)
semi_hour9_hour6_down=proteins_int[proteins_int$hour6_total-proteins_int$hour9_total>=2,]
write.table(semi_hour9_hour6_down,
            file=file.path(semi_dir,'semi_hour9_hour6_down.txt'),
            row.names=FALSE)

page=openPage(file.path(semi_dir,'index.html'),
              link.css='hwriter.css')
file.copy(system.file('images','hwriter.css',
                      package='hwriter'),
          file.path(semi_dir,'hwriter.css'))

## [1] FALSE

hwriteSidebar(position = c("left"),
              link = c("../mainmenu.html"),
              size = 15, margin = 1, page = page)
listedfiles=list.files(path=file.path(semi_dir))
hwrite("semi-quantified analysis of the proteints with missing values",
      heading=1,page=page)
for (i in 1:dim(sample_combi)[2]){
  for (j in c('down','up')) {
    hwrite(paste("Semi-quantative analysis of proteins in ",
                j,'-relagulated', sample_combi[1,i], ' to ',
                sample_combi[2,i],': ', sep=''),
          link=paste('semi_',sample_combi[1,i],'_',
                    sample_combi[2,i],'_', j, '.txt', sep=''),
          page=page, br=TRUE)
    semi_table=read.table(
      file.path(semi_dir,paste('semi_',sample_combi[1,i],'_',
                              sample_combi[2,i],'_', j, '.txt', sep='')),
      header=TRUE)
    semi_table=as.matrix(semi_table)
    hwrite(semi_table, page=page, br=TRUE )
  }
}
closePage(page,splash=FALSE)

semi_dir=file.path('who_semi_quant')
dir.create(semi_dir, showWarnings = FALSE)
proteins_who=add_summary(proteins_who)
semi_hour9_mock_up=proteins_who[proteins_who$hour9_total-proteins_who$mock_total>=2,]

```

```

write.table(semi_hour9_mock_up,
            file=file.path(semi_dir, 'semi_hour9_mock_up.txt'),
            row.names=FALSE)
semi_hour9_mock_down=proteins_who[proteins_who$mock_total-proteins_who$hour9_total>=2,]
write.table(semi_hour9_mock_down,
            file=file.path(semi_dir, 'semi_hour9_mock_down.txt'),
            row.names=FALSE)
semi_hour6_mock_up=proteins_who[proteins_who$hour6_total-proteins_who$mock_total>=2,]
write.table(semi_hour6_mock_up,
            file=file.path(semi_dir, 'semi_hour6_mock_up.txt'),
            row.names=FALSE)
semi_hour6_mock_down=proteins_who[proteins_who$mock_total-proteins_who$hour6_total>=2,]
write.table(semi_hour6_mock_down,
            file=file.path(semi_dir, 'semi_hour6_mock_down.txt'),
            row.names=FALSE)
semi_hour9_hour6_up=proteins_who[proteins_who$hour9_total-proteins_who$hour6_total>=2,]
write.table(semi_hour9_hour6_up,
            file=file.path(semi_dir, 'semi_hour9_hour6_up.txt'),
            row.names=FALSE)
semi_hour9_hour6_down=proteins_who[proteins_who$hour6_total-proteins_who$hour9_total>=2,]
write.table(semi_hour9_hour6_down,
            file=file.path(semi_dir, 'semi_hour9_hour6_down.txt'),
            row.names=FALSE)

page=openPage(file.path(semi_dir, 'index.html'),
              link.css='hwriter.css')
file.copy(system.file('images', 'hwriter.css',
                      package='hwriter'),
          file.path(semi_dir, 'hwriter.css'))

## [1] FALSE

hwriteSidebar(position = c("left"),
              link = c("../mainmenu.html"),
              size = 15, margin = 1, page = page)
listedfiles=list.files(path=file.path(semi_dir))
hwrite("semi-quantified analysis of the proteints with missing values",
      heading=1,page=page)
for (i in 1:dim(sample_combi)[2]){
  for (j in c('down','up')) {
    hwrite(paste("Semi-quantative analysis of proteins in ",
                j, '-relagulated', sample_combi[1,i], ' to ',
                sample_combi[2,i], ': ', sep=''),
          link=paste('semi_', sample_combi[1,i], '_',
                    sample_combi[2,i], '_', j, '.txt', sep=''),
          page=page, br=TRUE)
    semi_table=read.table(
      file.path(semi_dir, paste('semi_', sample_combi[1,i], '_',
                              sample_combi[2,i], '_', j, '.txt', sep='')),
      header=TRUE)
    semi_table=as.matrix(semi_table)
    hwrite(semi_table, page=page, br=TRUE )
  }
}
closePage(page, splash=FALSE)

####
semi_dir=file.path('dif_semi_quant')
dir.create(semi_dir, showWarnings = FALSE)

```

```

proteins=add_summary(proteins)
semi_hour9_mock_up=proteins[proteins$hour9_total-proteins$mock_total>=2,]
write.table(semi_hour9_mock_up,
            file=file.path(semi_dir, 'semi_hour9_mock_up.txt'),
            row.names=FALSE)
semi_hour9_mock_down=proteins[proteins$mock_total-proteins$hour9_total>=2,]
write.table(semi_hour9_mock_down,
            file=file.path(semi_dir, 'semi_hour9_mock_down.txt'),
            row.names=FALSE)
semi_hour6_mock_up=proteins[proteins$hour6_total-proteins$mock_total>=2,]
write.table(semi_hour6_mock_up,
            file=file.path(semi_dir, 'semi_hour6_mock_up.txt'),
            row.names=FALSE)
semi_hour6_mock_down=proteins[proteins$mock_total-proteins$hour6_total>=2,]
write.table(semi_hour6_mock_down,
            file=file.path(semi_dir, 'semi_hour6_mock_down.txt'),
            row.names=FALSE)
semi_hour9_hour6_up=proteins[proteins$hour9_total-proteins$hour6_total>=2,]
write.table(semi_hour9_hour6_up,
            file=file.path(semi_dir, 'semi_hour9_hour6_up.txt'),
            row.names=FALSE)
semi_hour9_hour6_down=proteins[proteins$hour6_total-proteins$hour9_total>=2,]
write.table(semi_hour9_hour6_down,
            file=file.path(semi_dir, 'semi_hour9_hour6_down.txt'),
            row.names=FALSE)

page=openPage(file.path(semi_dir, 'index.html'),
              link.css='hwriter.css')
file.copy(system.file('images', 'hwriter.css',
                      package='hwriter'),
          file.path(semi_dir, 'hwriter.css'))

## [1] FALSE

hwriteSidebar(position = c("left"),
              link = c("../mainmenu.html"),
              size = 15, margin = 1, page = page)
listedfiles=list.files(path=file.path(semi_dir))
hwrite("semi-quantified analysis of the proteints with missing values",
      heading=1,page=page)
for (i in 1:dim(sample_combi)[2]){
  for (j in c('down','up')) {
    hwrite(paste("Semi-quantative analysis of proteins in ",
                j, '-relagulated', sample_combi[1,i], ' to ',
                sample_combi[2,i], ': ', sep=''),
          link=paste('semi_', sample_combi[1,i], '_',
                    sample_combi[2,i], '_', j, '.txt', sep=''),
          page=page, br=TRUE)
    semi_table=read.table(
      file.path(semi_dir, paste('semi_', sample_combi[1,i], '_',
                              sample_combi[2,i], '_', j, '.txt', sep='')),
      header=TRUE)
    semi_table=as.matrix(semi_table)
    hwrite(semi_table, page=page, br=TRUE )
  }
}
closePage(page, splash=FALSE)

```


4 Gene set enrichment analysis for differentially expressed proteins

Gene set enrichment analysis is performed by applying Fishers exact test to categories from GO, Interpro, Pfam, and Uniprot annotations.

Gene set enrichment analysis for enriched genes in sets of comparisons.

```
GSEAdir='GSEA'
dir.create(file.path(GSEAdir), showWarnings = FALSE)
file.copy(system.file("images", "hwriter.css", package="hwriter"),
          file.path(GSEAdir, "hwriter.css"))

## [1] FALSE

GSEA = list()
for (s in colnames(sample_combi)) {
  enrich_table = diff_table[[s]]
  enrich_table = enrich_table[which(enrich_table$p.adj<0.1),]
  for (j in 1:2) {
    if (sample_combi[j,s]==sample_combi[1,s]) {
      Test_Geneset=enrich_table[enrich_table$log2FC>0,]
    } else {
      Test_Geneset=enrich_table[enrich_table$log2FC<0,]
    }
    n = paste('GSEA of proteins enriched in ',
              sample_combi[j,s], ' comparing to ',
              sample_combi[3-j,s], sep='')
    GSEA[[n]] = doGSEA(GenesSelected = as.character(Test_Geneset$ENSGid),
                      GenesBackground = as.character(proteins$ENSGid),
                      ENSG2category, k=3)
  }
}
save(GSEA, file=file.path(GSEAdir, "GSEA.rda"))
```

The tables of the GSEA are written.

```
for (n in names(GSEA)) {
  dir.create(file.path(GSEAdir, n, "table"),
            showWarnings=FALSE, recursive = TRUE)
  file.copy(system.file("images", "hwriter.css",
                        package="hwriter"),
            file.path(GSEAdir, n, "hwriter.css"))
  hwriteGSEA(GSEA[[n]], dir = file.path(GSEAdir, n, "table"),
            GeneAnnotation=ENSGannotation)
  hwriteGSEAtext(GSEA[[n]],
                file = file.path(GSEAdir, n, "GSEA.txt"),
                GeneAnnotation=ENSGannotation)
  hwriteGSEAreviso(GSEA[[n]],
                  file = file.path(GSEAdir, n, "GSEAreviso.txt"),
                  GeneAnnotation=ENSGannotation, padj=0.2)
}
```

Appendix D The interactome of chicken ANP32A in human HEK-239T cells

#	Gene	Mean protein intensity	Mean FC over PP7CP
1	KPNA1	2.20E+09	4.19E+03
2	KPNA6	2.63E+09	1.55E+03
3	ANP32A	7.91E+10	1.28E+03
4	TFAM	6.42E+08	5.21E+02
5	HNRNPH1	4.17E+08	2.45E+02
6	HAGH	1.84E+08	2.31E+02
7	PHB	2.26E+08	2.27E+02
8	SRP14	2.35E+08	2.16E+02
9	RPS7	5.84E+08	2.06E+02
10	SNRNP200	1.25E+08	1.75E+02
11	HMG2;HMG3	5.10E+07	1.65E+02
12	RRP1B	1.25E+08	1.59E+02
13	TOP1	1.61E+08	1.57E+02
14	GTF2I	1.26E+08	1.47E+02
15	SF3B2	2.58E+08	1.30E+02
16	SHMT2	1.33E+08	1.17E+02
17	LANCL1	3.96E+07	1.12E+02
18	NONO	1.66E+08	1.11E+02
19	STAT3	5.41E+07	1.05E+02
20	SFN	1.74E+07	1.02E+02
21	CDC5L	5.36E+07	1.01E+02
22	CBX3	2.87E+07	9.26E+01
23	SMC2	3.96E+07	8.93E+01
24	DACH1	1.46E+07	8.72E+01
25	CKAP4	7.16E+07	8.71E+01
26	SNRPD2	1.46E+08	8.34E+01
27	MYBBP1A	6.64E+07	7.97E+01
28	DDX6	2.20E+08	7.88E+01
29	HNRNPM	5.61E+07	7.54E+01

#	Gene	Mean protein intensity	Mean FC over PP7CP
30	RPS21	4.42E+07	7.53E+01
31	UQCRC2	2.73E+07	7.43E+01
32	NACA	1.04E+08	7.34E+01
33	EMD	3.90E+07	6.66E+01
34	PSMD2	3.21E+07	6.35E+01
35	NAP1L1	1.05E+08	6.27E+01
36	HADHB	5.89E+07	6.21E+01
37	ATP5D	5.25E+07	6.20E+01
38	RPS19	2.85E+07	6.20E+01
39	KPNA5	3.16E+07	5.88E+01
40	TCEB2	5.69E+07	5.67E+01
41	DYNLL1;DYNLL2	7.77E+07	5.42E+01
42	PRDX4	6.03E+07	5.36E+01
43	MRPL4	3.49E+07	5.20E+01
44	EIF3L	1.99E+07	5.13E+01
45	H1FX	4.23E+07	4.88E+01
46	MRPS16	1.52E+07	4.85E+01
47	SLIRP	3.17E+07	4.70E+01
48	RAVER1	6.11E+07	4.66E+01
49	FHL1	2.25E+07	4.64E+01
50	CUL5	5.56E+07	4.59E+01
51	PSMC4	1.82E+07	4.53E+01
52	HCFC1	5.66E+07	4.47E+01
53	DYNC1H1	1.00E+09	4.36E+01
54	QARS	7.58E+07	4.29E+01
55	RFC1	2.61E+07	4.28E+01
56	AIMP2	4.26E+07	4.25E+01
57	PSMD3	2.32E+07	4.19E+01
58	RPS20	7.48E+07	4.12E+01
59	TPR	8.18E+07	4.10E+01
60	MRPS7	4.74E+07	4.09E+01

61	PRPF6	5.19E+07	4.04E+01
#	Gene	Mean protein intensity	Mean FC over PP7CP
62	MRPL50	2.66E+07	3.98E+01
63	RPS14	7.61E+07	3.94E+01
64	MRPS22	4.37E+07	3.91E+01
65	ILF2	1.83E+07	3.89E+01
66	KARS	1.19E+08	3.84E+01
67	FAU	3.98E+07	3.80E+01
68	EIF3E	2.18E+07	3.74E+01
69	SLTM	5.68E+08	3.72E+01
70	RPS24	4.37E+07	3.71E+01
71	PHGDH	1.17E+08	3.67E+01
72	EIF3B	1.56E+07	3.62E+01
73	TCEB1	7.65E+07	3.51E+01
74	PTPLAD1	2.69E+07	3.50E+01
75	VARs	2.27E+07	3.49E+01
76	CHD4	2.07E+07	3.47E+01
77	FAM195B	5.57E+07	3.47E+01
78	RPS3A	3.12E+08	3.43E+01
79	SF3B4	6.41E+07	3.37E+01
80	MRPL37	3.42E+07	3.30E+01
81	MRPS14	9.89E+06	3.28E+01
82	MRPS18B	3.14E+07	3.27E+01
83	WASF2	3.52E+07	3.27E+01
84	PRPF3	2.14E+07	3.17E+01
85	ACP1	1.22E+07	3.00E+01
86	PSMD7	1.58E+07	2.89E+01
87	UBR4	1.03E+08	2.85E+01
88	LSM12	1.20E+08	2.80E+01
89	HNRNPH2	8.84E+06	2.79E+01
90	CAD	1.08E+08	2.77E+01
91	NME1	3.22E+07	2.74E+01

92	HTATSF1	5.74E+07	2.68E+01
93	TP53	8.88E+07	2.65E+01
#	Gene	Mean protein intensity	Mean FC over PP7CP
94	SMC4	2.83E+07	2.55E+01
95	SART1	2.89E+07	2.54E+01
96	DYNC1LI1	7.46E+07	2.47E+01
97	CDKN2AIP	1.01E+07	2.42E+01
98	RPL19	4.98E+07	2.42E+01
99	NHP2L1	1.67E+07	2.39E+01
100	MRPL53	2.77E+07	2.34E+01
101	RUVBL2	2.59E+07	2.27E+01
102	HNRNPR	1.18E+07	2.26E+01
103	U2SURP	1.28E+07	2.18E+01
104	SF3A3	5.68E+07	2.10E+01
105	MEPCE	7.99E+06	2.10E+01
106	PTBP1	3.23E+07	2.09E+01
107	RPL7A	9.77E+07	2.07E+01
108	MAD2L1	5.16E+07	2.03E+01
109	PSMA6	9.68E+06	1.94E+01
110	XP32	1.70E+07	1.93E+01
111	HNRNPH3	2.40E+07	1.92E+01
112	SMARCA4;SMARCA2	2.28E+07	1.89E+01
113	RPA2	5.93E+08	1.86E+01
114	PPIB	2.21E+07	1.84E+01
115	PRPF4	1.31E+07	1.79E+01
116	NUP93	7.25E+06	1.75E+01
117	SLC25A3	7.41E+07	1.70E+01
118	HIGD1A	2.37E+07	1.70E+01
119	YME1L1	2.33E+07	1.68E+01
120	SNRPA	2.04E+07	1.68E+01
121	VDAC2	2.26E+07	1.63E+01
122	PSMD1	4.71E+06	1.62E+01

123	RAN	1.57E+07	1.62E+01
124	YWHAB	5.99E+07	1.60E+01
125	RBBP7	3.04E+07	1.55E+01
#	Gene	Mean protein intensity	Mean FC over PP7CP
126	SMARCB1	6.78E+06	1.53E+01
127	EIF4A3	9.94E+06	1.48E+01
128	DDX23	1.58E+07	1.42E+01
129	MRPL48	1.34E+07	1.40E+01
130	PYCR1	2.34E+07	1.33E+01
131	ARCN1	6.20E+06	1.27E+01
132	PRDX6	4.64E+07	1.26E+01
133	DYNC1I2	1.05E+08	1.24E+01
134	TUBB2A	1.35E+07	1.24E+01
135	MRPS28	1.83E+07	1.23E+01
136	MRPS2	1.52E+07	1.21E+01
137	ACTL6A	1.57E+07	1.19E+01
138	PRPF8	1.55E+08	1.18E+01
139	RFC4	2.18E+07	1.17E+01
140	PSMA3	3.52E+06	1.17E+01
141	P4HB	4.16E+07	1.16E+01
142	CDV3	1.89E+07	1.14E+01
143	KPNA2	4.14E+08	1.14E+01
144	RNPS1	8.04E+06	1.13E+01
145	POLR1A	5.18E+06	1.12E+01
146	DDB1	2.06E+08	1.11E+01
147	BYSL	1.70E+07	1.10E+01
148	RPN2	1.83E+07	1.08E+01
149	ATP5O	3.11E+07	1.05E+01
150	MYL12B;MYL12A	9.78E+06	1.01E+01
151	SCFD1	4.97E+06	1.00E+01

Appendix E The interactome of human ANP32A in human HEK-239T cells

#	Gene	Mean protein intensity	Mean FC over PP7CP
1	ANP32B	1.65E+10	6.49E+03
2	ANP32A	1.32E+11	2.18E+03
3	KPNA1	1.64E+09	1.36E+03
4	KPNA6	2.89E+09	7.10E+02
5	SFN	2.94E+07	1.78E+02
6	AZGP1	3.82E+07	9.99E+01
7	XP32	3.00E+07	9.09E+01
8	RPS7	4.20E+08	7.42E+01
9	DACH1	1.07E+07	7.37E+01
10	PHB	2.29E+08	7.25E+01
11	SF3B2	2.49E+08	5.53E+01
12	KPNA5	5.43E+07	5.46E+01
13	SLTM	8.45E+08	5.10E+01
14	LGALS7	2.30E+07	4.75E+01
15	CALML3	1.29E+07	4.74E+01
16	NME1	4.50E+07	4.49E+01
17	HAGH	1.21E+08	4.45E+01
18	TGM3	3.11E+07	4.41E+01
19	S100A9	1.71E+08	4.27E+01
20	TFAM	5.45E+08	3.97E+01
21	HNRNPH1	4.67E+08	3.73E+01
22	RPS21	4.03E+07	3.65E+01
23	MRPL4	2.91E+07	2.76E+01
24	LSM12	5.11E+07	2.74E+01
25	PSMC4	1.54E+07	2.50E+01
26	HIGD1A	2.15E+07	2.43E+01
27	YWHAH	2.82E+07	2.06E+01
28	SF3B4	3.93E+07	1.77E+01
29	KPNA2	7.84E+08	1.75E+01

#	Gene	Mean protein intensity	Mean FC over PP7CP
30	NACA	1.36E+08	1.74E+01
31	SNRPD2	1.46E+08	1.71E+01
32	P4HB	5.10E+07	1.66E+01
33	HMGN2;HMGN3	1.61E+07	1.65E+01
34	SNRPA	1.60E+07	1.59E+01
35	SMC2	1.85E+07	1.54E+01
36	SHMT2	6.52E+07	1.46E+01
37	POF1B	5.06E+06	1.31E+01
38	PRDX2	1.80E+07	1.30E+01
39	EIF3E	1.66E+07	1.29E+01
40	COPA	1.22E+07	1.26E+01
41	HTATSF1	7.62E+07	1.23E+01
42	PKP1	2.52E+07	1.18E+01
43	FHL1	5.30E+06	1.14E+01
44	IMPDH2	2.61E+07	1.12E+01
45	SLC25A11	9.80E+06	1.02E+01
46	PRSS3	1.02E+07	1.00E+01

Appendix F Differential interactions of chicken and human ANP32A

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
1	UQCRC2	2.73E+07	6.31E+05	1.35E+02
2	STAT3	5.41E+07	1.24E+07	1.00E+02
3	TOP1	1.61E+08	5.56E+07	8.72E+01
4	SNRNP200	1.25E+08	9.59E+07	8.58E+01
5	GTF2I	1.26E+08	8.66E+07	8.05E+01
6	RRP1B	1.25E+08	4.00E+07	7.74E+01
7	CBX3	2.87E+07	4.02E+06	7.05E+01
8	LANCL1	3.96E+07	1.04E+07	6.81E+01
9	CKAP4	7.16E+07	4.55E+07	6.07E+01
10	RPS19	2.85E+07	1.19E+07	6.00E+01
11	ATP5D	5.25E+07	3.37E+07	5.94E+01
12	MYBBP1A	6.64E+07	3.79E+07	5.45E+01
13	DDX6	2.20E+08	2.39E+08	4.93E+01
14	TCEB2	5.69E+07	3.56E+07	4.90E+01
15	NONO	1.66E+08	1.77E+08	4.65E+01
16	CUL5	5.56E+07	4.30E+07	4.50E+01
17	HNRNPM	5.61E+07	4.92E+07	4.39E+01
18	HNRNPR	1.18E+07	6.31E+05	3.91E+01
19	DYNLL1;DYNLL2	7.77E+07	9.27E+07	3.75E+01
20	TCEB1	7.65E+07	2.92E+07	3.54E+01
21	EMD	3.90E+07	2.38E+07	3.46E+01
22	CDC5L	5.36E+07	2.30E+07	3.29E+01
23	FAU	3.98E+07	3.37E+07	3.27E+01
24	HCFC1	5.66E+07	4.28E+07	3.27E+01
25	RAVER1	6.11E+07	6.50E+07	3.11E+01
26	H1FX	4.23E+07	2.49E+07	3.10E+01
27	ACP1	1.22E+07	3.62E+06	3.10E+01
28	PTPLAD1	2.69E+07	1.71E+07	3.06E+01

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
29	SRP14	2.35E+08	1.98E+08	3.03E+01
30	RFC1	2.61E+07	1.65E+07	3.02E+01
31	RPS20	7.48E+07	6.82E+07	2.96E+01
32	SLC25A3	7.41E+07	8.93E+06	2.91E+01
33	PSMD3	2.32E+07	1.25E+07	2.91E+01
34	MRPS16	1.52E+07	2.77E+06	2.80E+01
35	CHD4	2.07E+07	1.47E+07	2.78E+01
36	SLIRP	3.17E+07	2.93E+07	2.78E+01
37	ILF2	1.83E+07	1.47E+07	2.77E+01
38	PRPF3	2.14E+07	1.39E+07	2.69E+01
39	PRPF6	5.19E+07	3.31E+07	2.61E+01
40	NHP2L1	1.67E+07	1.12E+07	2.58E+01
41	PSMD2	3.21E+07	3.30E+07	2.55E+01
42	PSMD1	4.71E+06	6.31E+05	2.53E+01
43	WASF2	3.52E+07	7.01E+07	2.48E+01
44	UBR4	1.03E+08	7.49E+07	2.39E+01
45	TP53	8.88E+07	7.01E+07	2.34E+01
46	MRPL50	2.66E+07	4.90E+07	2.20E+01
47	CAD	1.08E+08	1.00E+08	2.18E+01
48	PSMD7	1.58E+07	6.82E+06	2.14E+01
49	MRPS7	4.74E+07	3.65E+07	2.11E+01
50	MEPCE	7.99E+06	6.31E+05	2.10E+01
51	MRPS22	4.37E+07	5.06E+07	2.10E+01
52	KARS	1.19E+08	1.00E+08	2.08E+01
53	U2SURP	1.28E+07	6.18E+06	2.06E+01
54	MRPS14	9.89E+06	1.27E+07	2.05E+01
55	NAP1L1	1.05E+08	2.04E+08	2.00E+01
56	VAR5	2.27E+07	3.11E+07	1.99E+01
57	PRDX4	6.03E+07	6.51E+07	1.98E+01
58	DYNC1LI1	7.46E+07	6.19E+07	1.97E+01
59	RPS3A	3.12E+08	3.49E+08	1.96E+01

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
60	TPR	8.18E+07	5.82E+07	1.90E+01
61	FAM195B	5.57E+07	5.57E+07	1.79E+01
62	RUVBL2	2.59E+07	1.78E+07	1.76E+01
63	EIF3L	1.99E+07	1.31E+07	1.74E+01
64	VDAC2	2.26E+07	1.35E+07	1.73E+01
65	PSMA6	9.68E+06	4.17E+06	1.66E+01
66	RPL7A	9.77E+07	1.32E+08	1.61E+01
67	SART1	2.89E+07	2.58E+07	1.59E+01
68	SF3A3	5.68E+07	8.47E+07	1.51E+01
69	QARS	7.58E+07	7.68E+07	1.50E+01
70	SMARCA4;SMARCA2	2.28E+07	1.86E+07	1.47E+01
71	SMC4	2.83E+07	3.07E+07	1.46E+01
72	PHGDH	1.17E+08	1.62E+08	1.38E+01
73	MAD2L1	5.16E+07	4.82E+07	1.37E+01
74	HNRNPH3	2.40E+07	2.94E+07	1.34E+01
75	DYNC1H1	1.00E+09	1.05E+09	1.34E+01
76	RPS24	4.37E+07	7.62E+07	1.32E+01
77	TFAM	6.42E+08	5.45E+08	1.31E+01
78	RPL19	4.98E+07	7.34E+07	1.28E+01
79	YWHAB	5.99E+07	5.63E+07	1.21E+01
80	RAN	1.57E+07	1.43E+07	1.16E+01
81	YME1L1	2.33E+07	2.38E+07	1.16E+01
82	RNPS1	8.04E+06	6.77E+06	1.13E+01
83	NUP93	7.25E+06	1.05E+07	1.05E+01
84	MRPS2	1.52E+07	6.67E+06	1.03E+01
85	TUBB2A	1.35E+07	7.93E+06	1.02E+01
86	HMGN2;HMGN3	5.10E+07	1.61E+07	1.00E+01
87	BYSL	1.70E+07	1.72E+07	9.15E+00
88	PRPF4	1.31E+07	2.41E+07	8.96E+00
89	ATP5O	3.11E+07	2.45E+07	8.52E+00
90	CDKN2AIP	1.01E+07	9.66E+06	8.32E+00

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
91	RFC4	2.18E+07	2.71E+07	8.24E+00
92	MRPL48	1.34E+07	1.52E+07	8.07E+00
93	PRPF8	1.55E+08	9.83E+07	8.03E+00
94	SHMT2	1.33E+08	6.52E+07	8.00E+00
95	DDX23	1.58E+07	1.49E+07	7.82E+00
96	PRDX6	4.64E+07	6.29E+07	7.75E+00
97	ACTL6A	1.57E+07	1.10E+07	7.74E+00
98	DYNC1I2	1.05E+08	1.00E+08	7.71E+00
99	ARCN1	6.20E+06	3.49E+06	7.64E+00
100	MRPS28	1.83E+07	1.70E+07	7.59E+00
101	MYL12B;MYL12A	9.78E+06	7.92E+06	7.51E+00
102	PYCR1	2.34E+07	2.35E+07	7.51E+00
103	EIF4A3	9.94E+06	8.00E+06	7.15E+00
104	PTBP1	3.23E+07	4.00E+07	7.08E+00
105	CDV3	1.89E+07	2.53E+07	6.97E+00
106	SCFD1	4.97E+06	3.49E+06	6.84E+00
107	HNRNPH1	4.17E+08	4.67E+08	6.57E+00
108	HADHB	5.89E+07	2.74E+07	6.29E+00
109	SMC2	3.96E+07	1.85E+07	5.80E+00
110	RPN2	1.83E+07	3.70E+07	5.30E+00
111	HAGH	1.84E+08	1.21E+08	5.18E+00
112	SNRPD2	1.46E+08	1.46E+08	4.89E+00
113	POLR1A	5.18E+06	2.03E+06	4.68E+00
114	DDB1	2.06E+08	1.93E+08	4.57E+00
115	SMARCB1	6.78E+06	3.02E+06	4.52E+00
116	MRPS18B	3.14E+07	3.63E+07	4.48E+00
117	PSMA3	3.52E+06	8.42E+06	4.46E+00
118	AIMP2	4.26E+07	2.59E+07	4.44E+00
119	HNRNPH2	8.84E+06	5.93E+06	4.43E+00
120	RPA2	5.93E+08	4.69E+08	4.25E+00
121	NACA	1.04E+08	1.36E+08	4.21E+00

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
122	EIF3B	1.56E+07	8.35E+06	4.15E+00
123	RPS14	7.61E+07	6.72E+07	4.14E+00
124	FHL1	2.25E+07	5.30E+06	4.09E+00
125	MRPL53	2.77E+07	2.18E+07	4.06E+00
126	MRPL37	3.42E+07	3.39E+07	3.75E+00
127	PHB	2.26E+08	2.29E+08	3.14E+00
128	KPNA1	2.20E+09	1.64E+09	3.09E+00
129	RBBP7	3.04E+07	4.56E+07	3.07E+00
130	EIF3E	2.18E+07	1.66E+07	2.91E+00
131	RPS7	5.84E+08	4.20E+08	2.77E+00
132	SF3B2	2.58E+08	2.49E+08	2.35E+00
133	HTATSF1	5.74E+07	7.62E+07	2.18E+00
134	KPNA6	2.63E+09	2.89E+09	2.18E+00
135	RPS21	4.42E+07	4.03E+07	2.06E+00
136	PPIB	2.21E+07	4.20E+07	2.00E+00
137	SF3B4	6.41E+07	3.93E+07	1.91E+00
138	MRPL4	3.49E+07	2.91E+07	1.89E+00
139	PSMC4	1.82E+07	1.54E+07	1.81E+00
140	DACH1	1.46E+07	1.07E+07	1.18E+00
141	KPNA5	3.16E+07	5.43E+07	1.08E+00
142	SNRPA	2.04E+07	1.60E+07	1.06E+00
143	LSM12	1.20E+08	5.11E+07	1.02E+00
144	SLC25A11	2.53E+06	9.80E+06	8.12E-01
145	SLTM	5.68E+08	8.45E+08	7.28E-01
146	HIGD1A	2.37E+07	2.15E+07	7.01E-01
147	P4HB	4.16E+07	5.10E+07	6.99E-01
148	KPNA2	4.14E+08	7.84E+08	6.50E-01
149	NME1	3.22E+07	4.50E+07	6.09E-01
150	ANP32A	7.91E+10	1.32E+11	5.84E-01
151	SFN	1.74E+07	2.94E+07	5.72E-01
152	PRDX2	1.24E+07	1.80E+07	5.46E-01

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
153	IMPDH2	1.09E+07	2.61E+07	4.18E-01
154	COPA	4.39E+06	1.22E+07	3.78E-01
155	XP32	1.70E+07	3.00E+07	2.12E-01
156	PKP1	1.21E+07	2.52E+07	1.49E-01
157	S100A9	7.01E+06	1.71E+08	1.41E-01
158	PRSS3	1.15E+07	1.02E+07	1.01E-01
159	POF1B	6.31E+05	5.06E+06	7.67E-02
160	YWHAH	6.57E+06	2.82E+07	4.02E-02
161	TGM3	1.16E+07	3.11E+07	3.45E-02
162	AZGP1	1.16E+07	3.82E+07	1.95E-02
163	CALML3	6.31E+05	1.29E+07	1.31E-02
164	LGALS7	6.41E+05	2.30E+07	1.27E-02
165	ANP32B	1.22E+07	1.65E+10	3.36E-04

Appendix G Differential interactions of chicken and human ANP32A in HEK-239T cells
after infection with influenza A/WSN/33 (H1N1) virus

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
1	TFAM	1.88E+08	9.83E+05	7.95E+02
2	SRP14	9.13E+07	9.83E+05	3.87E+02
3	MARCKSL1	5.44E+07	9.83E+05	3.27E+02
4	EEF1B2	4.50E+07	9.83E+05	2.59E+02
5	HMG1	4.77E+07	9.83E+05	2.38E+02
6	SHMT2	4.06E+07	9.83E+05	2.38E+02
7	YBX1	1.38E+08	5.30E+07	2.09E+02
8	SNRNP200	4.41E+07	9.83E+05	2.00E+02
9	DDB1	9.33E+07	6.53E+07	1.82E+02
10	SF3B3	9.41E+07	1.68E+07	1.79E+02
11	HNRPDL	3.50E+07	9.83E+05	1.74E+02
12	ABI1	7.99E+07	9.83E+05	1.66E+02
13	SRP9	2.65E+07	9.83E+05	1.31E+02
14	PTCD3	2.81E+07	9.83E+05	1.28E+02
15	CCT4	2.50E+07	9.83E+05	1.23E+02
16	BCLAF1	2.48E+07	9.83E+05	1.22E+02
17	PABPN1	6.55E+07	9.83E+05	1.14E+02
18	NAP1L1	4.87E+07	1.42E+07	1.13E+02
19	VCP	3.03E+07	4.16E+08	1.12E+02
20	MRPS7	2.27E+07	9.83E+05	1.09E+02
21	EFTUD2	2.45E+07	9.83E+05	1.07E+02
22	RRP1B	2.04E+07	9.83E+05	1.06E+02
23	EIF3G	1.85E+07	9.83E+05	9.66E+01
24	DDX6	2.81E+07	1.53E+07	8.93E+01
25	ACTR1A	1.81E+07	9.83E+05	8.78E+01
26	NACA	3.89E+07	9.19E+07	8.75E+01
27	MRPS34	2.80E+07	1.32E+07	8.56E+01

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
28	DLST	6.48E+07	2.26E+07	8.33E+01
29	HNRNPA3	7.68E+07	1.88E+07	8.30E+01
30	RPS3A	9.45E+07	2.18E+07	8.29E+01
31	SF3A3	2.98E+07	9.83E+05	7.56E+01
32	Inf_HA	4.38E+08	2.25E+08	6.69E+01
33	RPS14	5.44E+07	9.83E+05	6.62E+01
34	YWHAE	2.08E+07	8.49E+07	6.47E+01
35	MRPL37	3.49E+07	9.83E+05	6.00E+01
36	DDX21	8.16E+07	1.40E+07	5.74E+01
37	WASF2	2.32E+07	9.83E+05	5.60E+01
38	MYBBP1A	2.59E+07	9.83E+05	5.55E+01
39	RPS5	1.40E+07	9.83E+05	5.36E+01
40	WRN	1.97E+07	9.83E+05	5.33E+01
41	HADHA	3.13E+07	9.03E+06	5.29E+01
42	SKP1	1.09E+07	9.83E+05	5.16E+01
43	DDOST	1.34E+07	1.91E+07	4.89E+01
44	CKAP4	1.35E+07	4.30E+07	4.87E+01
45	SLIRP	1.28E+07	9.83E+05	4.81E+01
46	FLYWCH2	3.28E+07	9.83E+05	4.65E+01
47	IARS	7.08E+07	4.19E+07	4.47E+01
48	LANCL1	9.58E+06	9.83E+05	4.47E+01
49	RCC1	1.19E+07	9.83E+05	4.42E+01
50	TCEB2	8.78E+06	9.83E+05	4.06E+01
51	RFC4	1.15E+07	9.83E+05	4.06E+01
52	CASP14	1.13E+07	4.09E+09	3.99E+01
53	CYFIP2	1.71E+07	9.83E+05	3.93E+01
54	PHGDH	8.33E+06	9.83E+05	3.83E+01
55	DNAJA1	6.28E+07	2.81E+07	3.82E+01
56	TUBA1C	1.01E+07	9.52E+06	3.69E+01
57	LARS	3.95E+07	2.98E+07	3.62E+01
58	Inf_NA	3.81E+07	9.83E+05	3.46E+01

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
59	HMGA1	1.16E+07	9.83E+05	3.46E+01
60	RPS3	1.14E+08	1.09E+08	3.44E+01
61	EIF3B	9.30E+06	9.87E+06	3.36E+01
62	FAU	7.34E+06	9.83E+05	3.33E+01
63	TMPO	1.32E+07	9.83E+05	3.31E+01
64	DNLZ	1.02E+07	9.83E+05	3.30E+01
65	QARS	1.20E+07	9.83E+05	3.16E+01
66	EMD	1.53E+07	9.83E+05	3.01E+01
67	RARS	4.24E+07	7.21E+06	2.96E+01
68	UBAP2L	2.37E+07	9.83E+05	2.96E+01
69	CAD	1.17E+07	9.83E+05	2.88E+01
70	EXOSC7	9.02E+06	9.83E+05	2.72E+01
71	KARS	1.78E+07	9.83E+05	2.71E+01
72	DACH1	1.28E+07	9.83E+05	2.66E+01
73	RPL7A	5.91E+07	3.48E+07	2.52E+01
74	RPL4	4.29E+07	4.76E+07	2.48E+01
75	SMC4	5.66E+06	9.83E+05	2.48E+01
76	MRPL17	6.96E+06	9.83E+05	2.40E+01
77	RPN1	7.06E+06	1.18E+08	2.38E+01
78	LDHB	5.41E+06	9.83E+05	2.35E+01
79	ACACA	2.98E+10	2.58E+10	2.33E+01
80	RPP30	7.09E+06	9.83E+05	2.30E+01
81	RPL6	8.06E+06	1.59E+07	2.28E+01
82	RPL29	3.19E+07	2.98E+07	2.27E+01
83	PAICS	6.56E+06	1.87E+07	2.23E+01
84	ACTL6A	7.31E+06	9.83E+05	1.95E+01
85	NOLC1	4.55E+06	9.83E+05	1.92E+01
86	RPS19	7.83E+06	9.83E+05	1.87E+01
87	RPS6	9.98E+07	5.61E+07	1.85E+01
88	CBX5	4.39E+06	9.83E+05	1.83E+01
89	PCBP1	2.24E+07	2.07E+07	1.74E+01

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
90	RPLP0;RPLP0P6	1.28E+08	1.13E+08	1.68E+01
91	MRPS35	3.96E+06	9.83E+05	1.61E+01
92	API5	3.81E+06	9.83E+05	1.54E+01
93	DYNC1H1	2.98E+08	1.65E+08	1.47E+01
94	YWHAQ	5.16E+06	9.95E+06	1.45E+01
95	PSMA6	4.56E+06	2.60E+08	1.42E+01
96	PRC1	3.54E+06	9.83E+05	1.40E+01
97	NPM1	2.53E+08	2.13E+08	1.35E+01
98	AIFM1	2.60E+08	6.37E+07	1.21E+01
99	RPA2	4.53E+08	1.43E+08	1.20E+01
100	MRPS2	3.75E+06	5.77E+06	1.19E+01
101	MSH3	8.39E+07	6.94E+06	1.11E+01
102	RPN2	3.67E+06	1.84E+07	1.10E+01
103	LRPPRC	2.44E+07	2.11E+07	1.06E+01
104	P4HB	3.55E+06	2.79E+07	1.04E+01
105	SF3B2	6.50E+07	1.40E+07	1.02E+01
106	SF3B1	1.14E+08	2.33E+07	1.00E+01
107	EIF3A	1.62E+07	2.35E+06	9.82E+00
108	PRPF8	5.69E+07	6.03E+06	9.19E+00
109	CBX3	4.69E+07	5.24E+06	8.58E+00
110	FLNA	4.79E+07	3.37E+07	8.44E+00
111	KPNB1	3.88E+07	8.90E+06	8.31E+00
112	ATXN2L	3.48E+08	9.28E+07	8.21E+00
113	Inf_PA	2.98E+08	1.25E+08	7.94E+00
114	SNRPF	1.34E+07	2.23E+06	7.83E+00
115	TPR	3.76E+07	2.40E+06	7.40E+00
116	MRPS23	1.97E+07	9.65E+06	7.25E+00
117	Inf_PB2	4.66E+08	1.85E+08	7.22E+00
118	VIM	4.16E+08	7.82E+07	7.19E+00
119	PHB	5.34E+07	2.87E+07	7.15E+00
120	SFPQ	8.99E+07	1.01E+07	7.10E+00

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
121	RPS28	1.22E+08	4.65E+07	7.09E+00
122	SF3A1	6.20E+07	2.13E+07	7.06E+00
123	RPS12	9.24E+07	7.25E+07	6.87E+00
124	RPLP2	3.13E+08	1.47E+08	6.82E+00
125	DYNLL1;DYNLL2	2.09E+07	4.05E+06	6.18E+00
126	DAP3	3.97E+07	1.06E+07	6.15E+00
127	KPRP	6.39E+07	1.35E+10	6.08E+00
128	CYFIP1	2.99E+08	9.73E+07	5.86E+00
129	SSBP1	2.54E+08	1.02E+08	5.59E+00
130	HIST1H4A	1.28E+08	3.08E+08	5.22E+00
131	STRAP	2.93E+07	3.08E+06	5.04E+00
132	HSP90AA1	1.09E+08	3.03E+08	4.95E+00
133	RPL23A	2.86E+07	3.92E+07	4.86E+00
134	FECH	8.51E+07	1.50E+07	4.85E+00
135	ALYREF	1.12E+08	2.23E+07	4.79E+00
136	YBX3	6.34E+08	2.40E+08	4.44E+00
137	PRH1	1.14E+08	2.60E+07	4.38E+00
138	KPNA6	1.16E+09	5.98E+07	4.28E+00
139	HSP90AB1	1.07E+08	3.77E+07	4.16E+00
140	STAT3	2.41E+07	6.07E+07	4.13E+00
141	ENO1	4.46E+07	2.04E+09	4.10E+00
142	PRPF40A	4.47E+07	7.53E+06	4.04E+00
143	DHX15	1.03E+07	4.92E+06	4.02E+00
144	LDHA	1.17E+07	3.33E+08	3.98E+00
145	Inf_PB1	3.65E+08	9.48E+07	3.54E+00
146	CFL1	2.13E+07	7.45E+07	3.50E+00
147	RRM2B	1.30E+08	4.24E+07	3.18E+00
148	RPL18	4.76E+07	3.28E+07	3.17E+00
149	SKIV2L2	2.09E+07	5.87E+06	3.15E+00
150	RPS7	1.62E+08	5.19E+07	3.08E+00
151	SF3B4	2.26E+07	5.06E+06	3.00E+00

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
152	CSTA	7.48E+07	2.48E+09	2.87E+00
153	KPNA1	2.35E+09	1.90E+08	2.82E+00
154	RPL27	3.11E+07	7.23E+06	2.75E+00
155	PHB2	6.46E+07	3.13E+07	2.74E+00
156	MSH2	6.55E+07	1.53E+07	2.71E+00
157	DBT	4.91E+08	3.41E+08	2.60E+00
158	NXF1	4.19E+07	1.13E+07	2.50E+00
159	EIF3E	9.98E+06	8.76E+06	2.47E+00
160	LMNB1	4.82E+07	2.63E+07	2.46E+00
161	NUP93	1.16E+07	2.66E+06	2.45E+00
162	GNB2L1	1.26E+07	5.33E+07	2.43E+00
163	SNRPB;SNRPN	1.77E+07	7.18E+07	2.40E+00
164	TAGLN2	6.01E+06	3.03E+06	2.33E+00
165	UTS2	7.56E+07	9.66E+07	2.26E+00
166	HSPH1	2.62E+07	3.23E+06	2.24E+00
167	RBM39	6.12E+07	1.43E+07	2.23E+00
168	SLC16A1	1.79E+07	5.60E+06	2.19E+00
169	RAN	2.09E+07	1.91E+07	2.03E+00
170	ANXA2;ANXA2P2	1.33E+07	2.39E+09	2.00E+00
171	ANP32A	2.02E+11	2.21E+10	1.95E+00
172	DDX46	3.41E+06	2.14E+06	1.93E+00
173	GRPEL1	9.90E+07	6.72E+07	1.79E+00
174	RFC2	1.02E+07	3.31E+06	1.75E+00
175	DECR1	8.41E+06	2.49E+06	1.71E+00
176	IRS4	9.36E+06	3.15E+06	1.71E+00
177	YWHAZ	1.10E+07	1.05E+08	1.70E+00
178	RPS9	1.52E+08	8.60E+07	1.67E+00
179	KHDRBS1	7.08E+08	9.31E+08	1.67E+00
180	MRPL54	1.34E+07	5.15E+06	1.64E+00
181	PTPLAD1	8.85E+06	3.04E+06	1.64E+00
182	PRPF19	6.31E+07	3.53E+07	1.50E+00

#	Gene	Mean FC (chANP32A)	Mean FC (huANP32A)	FC (chANP32A) / FC (huANP32A)
183	RPL22	1.16E+08	7.82E+07	1.48E+00
184	EEF2	1.39E+07	2.24E+08	1.47E+00
185	KPNA2	4.21E+08	1.82E+08	1.24E+00
186	HIGD1A	5.64E+06	1.47E+07	1.20E+00
187	TOP1	3.56E+07	2.40E+07	1.17E+00
188	SLC25A3	2.93E+07	1.26E+07	1.17E+00
189	XP32	1.69E+07	1.70E+09	1.16E+00
190	CAPZA1	5.08E+06	2.36E+08	1.14E+00
191	SLTM	2.23E+08	8.80E+07	1.14E+00
192	RBM33	1.35E+07	1.10E+07	1.02E+00
193	PTBP1	1.15E+08	7.62E+07	1.01E+00
194	AIMP2	8.70E+06	7.92E+06	9.13E-01
195	SSRP1	4.04E+07	2.45E+07	7.34E-01
196	CALML5	9.86E+07	7.68E+09	5.73E-01
197	HAGH	5.63E+07	3.14E+07	4.65E-02
198	PIP	1.98E+07	1.32E+07	2.07E-02
199	TXN	2.10E+07	8.36E+08	1.59E-02
200	AZGP1	9.83E+05	2.70E+09	1.55E-02
201	ANP32B	2.62E+08	2.79E+08	9.40E-04

Appendix H Differential interactions of chicken ANP32A in HEK-239T cells after infection with influenza A/WSN/33 (H1N1) virus

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
1	PRH1	1.18E-01	4.20E+02	3.56E+03
2	YBX3	2.48E+00	8.93E+02	3.61E+02
3	RPLP2	9.95E-01	1.97E+02	1.98E+02
4	NPM1	2.87E+00	5.62E+02	1.96E+02
5	HMG1	NA	1.79E+02	1.79E+02
6	YBX1	8.18E-01	1.40E+02	1.71E+02
7	Inf_PB2	NA	1.69E+02	1.69E+02
8	CALML5	1.21E+00	1.97E+02	1.64E+02
9	EEF1B2	1.06E+00	1.62E+02	1.53E+02
10	UTS2	7.21E-01	1.09E+02	1.52E+02
11	Inf_PA	NA	1.36E+02	1.36E+02
12	HNRPD1	NA	1.31E+02	1.31E+02
13	ANP32A	1.28E+03	1.59E+05	1.25E+02
14	RPS12	1.47E+00	1.78E+02	1.21E+02
15	RPS28	1.76E+00	2.08E+02	1.18E+02
16	MSH3	2.36E+00	2.58E+02	1.09E+02
17	RRM2B	1.24E+00	1.21E+02	9.73E+01
18	GRPEL1	1.92E+00	1.81E+02	9.43E+01
19	DAP3	1.26E+00	1.04E+02	8.31E+01
20	KPNA2	1.14E+01	7.95E+02	6.98E+01
21	ENO1	1.23E+00	8.40E+01	6.82E+01
22	ACTR1A	NA	6.59E+01	6.59E+01
23	YWHAE	1.05E+00	6.32E+01	6.01E+01
24	PHB2	4.08E+00	2.35E+02	5.76E+01
25	Inf_PB1	NA	5.74E+01	5.74E+01
26	MARCKSL1	3.46E+00	1.93E+02	5.58E+01
27	SSRP1	6.63E-01	3.57E+01	5.38E+01

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
28	YWHAQ	2.69E-01	1.35E+01	5.01E+01
29	CFL1	1.57E+00	7.80E+01	4.95E+01
30	PABPN1	1.23E+00	5.78E+01	4.68E+01
31	NXF1	2.18E+00	9.96E+01	4.56E+01
32	MRPS23	1.60E+00	7.21E+01	4.50E+01
33	Inf_HA	NA	4.50E+01	4.50E+01
34	DLST	1.30E+00	5.44E+01	4.19E+01
35	CCT4	2.26E+00	9.24E+01	4.09E+01
36	FECH	2.86E+00	1.16E+02	4.05E+01
37	MSH2	3.74E+00	1.50E+02	4.01E+01
38	GNB2L1	1.13E+00	4.50E+01	3.97E+01
39	RPS9	2.01E+00	7.69E+01	3.82E+01
40	DDOST	1.30E+00	4.79E+01	3.67E+01
41	DBT	2.19E+00	7.66E+01	3.49E+01
42	RPL18	7.13E-01	2.47E+01	3.46E+01
43	RBM33	1.26E+00	3.97E+01	3.15E+01
44	KPNB1	3.28E+00	1.00E+02	3.05E+01
45	RBM39	1.90E+00	5.67E+01	2.99E+01
46	RPL23A	3.59E+00	1.06E+02	2.95E+01
47	RPL4	6.64E-01	1.93E+01	2.91E+01
48	MRPL54	9.33E-01	2.71E+01	2.90E+01
49	BCLAF1	3.17E+00	9.16E+01	2.89E+01
50	FLNA	4.44E+00	1.27E+02	2.87E+01
51	PTCD3	2.52E+00	6.87E+01	2.73E+01
52	ANXA2;ANXA2P2	1.52E+00	4.10E+01	2.69E+01
53	SNRPF	1.74E+00	4.22E+01	2.43E+01
54	PRPF19	8.30E+00	1.97E+02	2.37E+01
55	LDHA	1.40E+00	3.27E+01	2.34E+01
56	VCP	4.98E+00	1.12E+02	2.25E+01
57	SKP1	1.82E+00	3.87E+01	2.13E+01
58	SF3B1	9.69E+00	2.01E+02	2.07E+01

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
59	LMNB1	7.01E+00	1.42E+02	2.03E+01
60	SKIV2L2	3.01E+00	6.06E+01	2.01E+01
61	KPRP	3.28E+00	6.35E+01	1.93E+01
62	HMGA1	NA	1.92E+01	1.92E+01
63	DNLZ	NA	1.87E+01	1.87E+01
64	ABI1	4.54E+00	8.40E+01	1.85E+01
65	RPL6	1.14E+00	2.12E+01	1.85E+01
66	TUBA1C	1.94E+00	3.54E+01	1.82E+01
67	SSBP1	1.54E+00	2.76E+01	1.79E+01
68	Inf_NA	NA	1.77E+01	1.77E+01
69	KHDRBS1	5.24E+00	9.21E+01	1.76E+01
70	RPLP0;RPLP0P6	7.78E-01	1.24E+01	1.59E+01
71	SLC16A1	2.52E+00	3.98E+01	1.58E+01
72	PRPF40A	6.59E+00	1.03E+02	1.57E+01
73	CASP14	2.58E+00	3.98E+01	1.54E+01
74	PRPF8	1.18E+01	1.82E+02	1.54E+01
75	RPS5	2.05E+00	3.00E+01	1.47E+01
76	MRPL17	NA	1.45E+01	1.45E+01
77	EFTUD2	4.07E+00	5.77E+01	1.42E+01
78	CBX5	NA	1.37E+01	1.37E+01
79	RPS3	1.89E+00	2.59E+01	1.37E+01
80	CSTA	1.23E+00	1.66E+01	1.35E+01
81	EIF3G	4.33E+00	5.83E+01	1.34E+01
82	TAGLN2	1.49E+00	1.99E+01	1.34E+01
83	DDX21	2.55E+00	3.33E+01	1.31E+01
84	HSP90AA1	3.58E+00	4.59E+01	1.28E+01
85	RPL29	1.37E+00	1.73E+01	1.26E+01
86	DDB1	1.11E+01	1.40E+02	1.26E+01
87	SF3B3	8.59E+00	1.08E+02	1.25E+01
88	EIF3A	4.77E+00	5.86E+01	1.23E+01
89	DNAJA1	2.09E+00	2.55E+01	1.22E+01

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
90	YWHAZ	1.95E+00	2.38E+01	1.22E+01
91	UBAP2L	1.27E+00	1.54E+01	1.21E+01
92	CAPZA1	1.34E+00	1.61E+01	1.20E+01
93	SRP9	8.25E+00	9.81E+01	1.19E+01
94	RPN1	2.00E+00	2.37E+01	1.18E+01
95	ALYREF	1.06E+00	1.24E+01	1.16E+01
96	CYFIP2	1.89E+00	2.10E+01	1.11E+01
97	RPL27	6.11E+00	6.70E+01	1.10E+01
98	DECR1	NA	1.08E+01	1.08E+01
99	HNRNPA3	4.71E+00	5.09E+01	1.08E+01
100	EEF2	2.53E+00	2.66E+01	1.05E+01
101	AIFM1	1.54E+00	1.61E+01	1.05E+01
102	SLTM	3.72E+01	3.79E+02	1.02E+01
103	VIM	3.26E+00	3.29E+01	1.01E+01
104	LARS	2.75E+00	2.67E+01	9.72E+00
105	RPP30	1.42E+00	1.37E+01	9.68E+00
106	RCC1	2.62E+00	2.51E+01	9.58E+00
107	ACACA	1.81E+00	1.70E+01	9.42E+00
108	LDHB	1.92E+00	1.76E+01	9.17E+00
109	SFPQ	1.83E+00	1.67E+01	9.11E+00
110	WRN	3.13E+00	2.84E+01	9.07E+00
111	HADHA	3.93E+00	3.39E+01	8.63E+00
112	STRAP	5.07E+00	4.29E+01	8.46E+00
113	HSP90AB1	1.55E+00	1.31E+01	8.44E+00
114	API5	1.43E+00	1.15E+01	8.06E+00
115	LRPPRC	8.24E+00	6.58E+01	7.99E+00
116	HIST1H4A	1.57E+00	1.23E+01	7.85E+00
117	RFC2	2.15E+00	1.66E+01	7.69E+00
118	IRS4	2.08E+00	1.51E+01	7.26E+00
119	TMPO	2.55E+00	1.80E+01	7.08E+00
120	MRPS34	9.89E+00	6.64E+01	6.72E+00

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
121	PAICS	3.47E+00	2.18E+01	6.28E+00
122	RPL22	8.73E+00	5.29E+01	6.05E+00
123	DHX15	4.66E+00	2.62E+01	5.62E+00
124	PTBP1	2.09E+01	1.14E+02	5.46E+00
125	RPS6	3.89E+00	2.05E+01	5.26E+00
126	RARS	3.26E+00	1.71E+01	5.24E+00
127	MRPS35	2.37E+00	1.21E+01	5.10E+00
128	SF3A1	2.23E+00	1.10E+01	4.93E+00
129	IARS	7.76E+00	3.13E+01	4.03E+00
130	DDX46	2.63E+00	1.00E+01	3.80E+00
131	FLYWCH2	8.41E+00	2.40E+01	2.85E+00
132	CYFIP1	4.71E+00	1.31E+01	2.77E+00
133	NOLC1	5.43E+00	1.44E+01	2.65E+00
134	SNRPB;SNRPN	5.05E+00	1.31E+01	2.60E+00
135	PRC1	4.13E+00	1.05E+01	2.54E+00
136	EXOSC7	6.44E+00	1.55E+01	2.41E+00
137	HSPH1	9.30E+00	2.04E+01	2.19E+00
138	RFC4	1.17E+01	2.30E+01	1.97E+00
139	SF3A3	2.10E+01	3.94E+01	1.87E+00
140	ATXN2L	7.27E+00	1.11E+01	1.52E+00
141	MRPS7	4.09E+01	6.04E+01	1.48E+00
142	RPS3A	3.43E+01	5.02E+01	1.46E+00
143	XP32	1.93E+01	2.79E+01	1.45E+00
144	SF3B4	3.37E+01	4.84E+01	1.43E+00
145	DYNLL1;DYNLL2	5.42E+01	7.68E+01	1.42E+00
146	PCBP1	9.35E+00	1.32E+01	1.41E+00
147	DYNC1H1	4.36E+01	5.41E+01	1.24E+00
148	NAP1L1	6.27E+01	7.48E+01	1.19E+00
149	RAN	1.62E+01	1.92E+01	1.19E+00
150	SHMT2	1.17E+02	1.37E+02	1.17E+00
151	SF3B2	1.30E+02	1.47E+02	1.13E+00

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
152	RPS7	2.06E+02	2.27E+02	1.10E+00
153	NACA	7.34E+01	8.04E+01	1.10E+00
154	HIGD1A	1.70E+01	1.83E+01	1.07E+00
155	TPR	4.10E+01	4.39E+01	1.07E+00
156	RPN2	1.08E+01	1.08E+01	1.00E+00
157	NUP93	1.75E+01	1.70E+01	9.70E-01
158	ACTL6A	1.19E+01	1.13E+01	9.51E-01
159	MRPL37	3.30E+01	3.09E+01	9.38E-01
160	EIF3E	3.74E+01	3.49E+01	9.34E-01
161	MRPS2	1.21E+01	1.11E+01	9.20E-01
162	DDX6	7.88E+01	7.24E+01	9.19E-01
163	SRP14	2.16E+02	1.97E+02	9.12E-01
164	WASF2	3.27E+01	2.94E+01	9.00E-01
165	SLC25A3	1.70E+01	1.52E+01	8.94E-01
166	EIF3B	3.62E+01	3.23E+01	8.92E-01
167	P4HB	1.16E+01	1.03E+01	8.89E-01
168	RPS14	3.94E+01	3.37E+01	8.56E-01
169	STAT3	1.05E+02	8.86E+01	8.48E-01
170	RPL7A	2.07E+01	1.75E+01	8.43E-01
171	PHGDH	3.67E+01	2.88E+01	7.84E-01
172	TFAM	5.21E+02	4.01E+02	7.71E-01
173	PSMA6	1.94E+01	1.41E+01	7.30E-01
174	SMC4	2.55E+01	1.86E+01	7.29E-01
175	TOP1	1.57E+02	1.03E+02	6.61E-01
176	FAU	3.80E+01	2.50E+01	6.58E-01
177	KPNA6	1.55E+03	9.62E+02	6.22E-01
178	SNRNP200	1.75E+02	1.05E+02	6.00E-01
179	RPA2	1.86E+01	1.12E+01	5.99E-01
180	SMARCB1	1.53E+01	9.00E+00	5.88E-01
181	AIMP2	4.25E+01	2.48E+01	5.82E-01
182	SLIRP	4.70E+01	2.71E+01	5.77E-01

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
183	CAD	2.77E+01	1.58E+01	5.71E-01
184	CKAP4	8.71E+01	4.83E+01	5.54E-01
185	TCEB2	5.67E+01	3.05E+01	5.37E-01
186	KPNA1	4.19E+03	2.03E+03	4.84E-01
187	PHB	2.27E+02	1.04E+02	4.55E-01
188	QARS	4.29E+01	1.74E+01	4.05E-01
189	PTPLAD1	3.50E+01	1.38E+01	3.95E-01
190	RRP1B	1.59E+02	6.14E+01	3.87E-01
191	KARS	3.84E+01	1.43E+01	3.73E-01
192	MYBBP1A	7.97E+01	2.89E+01	3.63E-01
193	PRPF3	3.17E+01	9.97E+00	3.15E-01
194	LANCL1	1.12E+02	3.35E+01	2.99E-01
195	YME1L1	1.68E+01	4.51E+00	2.68E-01
196	EMD	6.66E+01	1.61E+01	2.42E-01
197	UBR4	2.85E+01	6.05E+00	2.12E-01
198	RFC1	4.28E+01	8.06E+00	1.88E-01
199	RPS19	6.20E+01	1.07E+01	1.72E-01
200	DACH1	8.72E+01	1.44E+01	1.66E-01
201	HCFC1	4.47E+01	6.87E+00	1.54E-01
202	CBX3	9.26E+01	1.37E+01	1.48E-01
203	RPL19	2.42E+01	3.32E+00	1.37E-01
204	HTATSF1	2.68E+01	3.60E+00	1.34E-01
205	HADHB	6.21E+01	7.27E+00	1.17E-01
206	SCFD1	1.00E+01	NA	9.95E-02
207	MYL12B;MYL12A	1.01E+01	NA	9.86E-02
208	ATP5O	1.05E+01	9.73E-01	9.25E-02
209	RNPS1	1.13E+01	NA	8.83E-02
210	CDV3	1.14E+01	NA	8.74E-02
211	TUBB2A	1.24E+01	NA	8.08E-02
212	ARCN1	1.27E+01	NA	7.85E-02
213	POLR1A	1.12E+01	8.53E-01	7.60E-02

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
214	MRPS28	1.23E+01	9.21E-01	7.48E-02
215	BYSL	1.10E+01	7.50E-01	6.81E-02
216	DYNC1I2	1.24E+01	8.03E-01	6.46E-02
217	RBBP7	1.55E+01	NA	6.45E-02
218	SNRPA	1.68E+01	1.05E+00	6.26E-02
219	YWHAB	1.60E+01	NA	6.25E-02
220	PSMD1	1.62E+01	NA	6.18E-02
221	PYCR1	1.33E+01	7.50E-01	5.65E-02
222	PPIB	1.84E+01	NA	5.44E-02
223	MRPL48	1.40E+01	7.50E-01	5.37E-02
224	SMARCA4;SMARCA2	1.89E+01	NA	5.30E-02
225	DDX23	1.42E+01	7.50E-01	5.28E-02
226	LSM12	2.80E+01	1.45E+00	5.20E-02
227	PSMA3	1.17E+01	5.93E-01	5.08E-02
228	MAD2L1	2.03E+01	NA	4.92E-02
229	MEPCE	2.10E+01	NA	4.76E-02
230	U2SURP	2.18E+01	NA	4.59E-02
231	EIF4A3	1.48E+01	6.56E-01	4.44E-02
232	HNRNPR	2.26E+01	NA	4.42E-02
233	MRPL53	2.34E+01	NA	4.28E-02
234	RUVBL2	2.27E+01	9.57E-01	4.22E-02
235	NHP2L1	2.39E+01	NA	4.19E-02
236	PRPF4	1.79E+01	7.50E-01	4.19E-02
237	CDKN2AIP	2.42E+01	NA	4.13E-02
238	SNRPD2	8.34E+01	3.37E+00	4.04E-02
239	PRDX6	1.26E+01	5.03E-01	3.99E-02
240	DYNC1LI1	2.47E+01	9.74E-01	3.94E-02
241	SART1	2.54E+01	NA	3.94E-02
242	GTF2I	1.47E+02	5.53E+00	3.76E-02
243	TP53	2.65E+01	9.51E-01	3.59E-02
244	HNRNPH2	2.79E+01	NA	3.58E-02

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
245	PSMD7	2.89E+01	NA	3.47E-02
246	ACP1	3.00E+01	NA	3.33E-02
247	VDAC2	1.63E+01	5.10E-01	3.12E-02
248	HNRNPH3	1.92E+01	5.87E-01	3.06E-02
249	FAM195B	3.47E+01	NA	2.88E-02
250	CHD4	3.47E+01	NA	2.88E-02
251	VAR5	3.49E+01	NA	2.86E-02
252	TCEB1	3.51E+01	NA	2.85E-02
253	MRPS18B	3.27E+01	9.31E-01	2.84E-02
254	NME1	2.74E+01	7.30E-01	2.67E-02
255	PRPF6	4.04E+01	NA	2.47E-02
256	PSMD3	4.19E+01	NA	2.39E-02
257	MRPL50	3.98E+01	9.12E-01	2.29E-02
258	MRPS14	3.28E+01	7.50E-01	2.29E-02
259	PSMC4	4.53E+01	NA	2.21E-02
260	HNRNPH1	2.45E+02	5.32E+00	2.17E-02
261	FHL1	4.64E+01	NA	2.15E-02
262	CUL5	4.59E+01	9.55E-01	2.08E-02
263	RAVER1	4.66E+01	9.54E-01	2.05E-02
264	H1FX	4.88E+01	NA	2.05E-02
265	MRPS22	3.91E+01	7.88E-01	2.02E-02
266	RPS24	3.71E+01	7.36E-01	1.99E-02
267	EIF3L	5.13E+01	NA	1.95E-02
268	PRDX4	5.36E+01	NA	1.87E-02
269	MRPL4	5.20E+01	9.35E-01	1.80E-02
270	KPNA5	5.88E+01	NA	1.70E-02
271	ATP5D	6.20E+01	NA	1.61E-02
272	PSMD2	6.35E+01	NA	1.58E-02
273	MRPS16	4.85E+01	7.50E-01	1.55E-02
274	RPS20	4.12E+01	5.73E-01	1.39E-02
275	UQCRC2	7.43E+01	NA	1.35E-02

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
276	RPS21	7.53E+01	NA	1.33E-02
277	ILF2	3.89E+01	5.13E-01	1.32E-02
278	HNRNPM	7.54E+01	9.52E-01	1.26E-02
279	SMC2	8.93E+01	NA	1.12E-02
280	CDC5L	1.01E+02	NA	9.90E-03
281	NONO	1.11E+02	9.67E-01	8.75E-03
282	HMG2;HMG3	1.65E+02	NA	6.05E-03
283	SFN	1.02E+02	5.08E-01	4.99E-03
284	HAGH	2.31E+02	8.39E-01	3.63E-03

Appendix I Differential interactions of human ANP32A in HEK-239T cells after infection with influenza A/WSN/33 (H1N1) virus

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
1	YBX3	1.24E+00	2.01E+02	1.62E+02
2	SSRP1	5.83E-01	4.86E+01	8.34E+01
3	PRH1	1.15E+00	9.58E+01	8.34E+01
4	CALML5	4.21E+00	3.45E+02	8.19E+01
5	LMNB1	1.01E+00	5.77E+01	5.69E+01
6	KHDRBS1	9.95E-01	5.51E+01	5.53E+01
7	GRPEL1	1.96E+00	1.01E+02	5.15E+01
8	TOP1	1.80E+00	8.83E+01	4.92E+01
9	PIP	NA	4.73E+01	4.73E+01
10	PTBP1	2.95E+00	1.13E+02	3.85E+01
11	ANP32A	2.18E+03	8.17E+04	3.74E+01
12	KPNA2	1.75E+01	6.40E+02	3.65E+01
13	RPS9	1.32E+00	4.60E+01	3.48E+01
14	PRPF19	3.89E+00	1.31E+02	3.38E+01
15	MRPL54	4.88E-01	1.65E+01	3.37E+01
16	RPLP2	9.33E-01	2.89E+01	3.10E+01
17	RPL23A	7.19E-01	2.18E+01	3.03E+01
18	UTS2	1.68E+00	4.84E+01	2.88E+01
19	SKIV2L2	6.95E-01	1.92E+01	2.77E+01
20	RPS12	9.74E-01	2.60E+01	2.67E+01
21	PHB2	3.26E+00	8.56E+01	2.63E+01
22	MSH3	9.32E-01	2.33E+01	2.50E+01
23	Inf_PB2	NA	2.34E+01	2.34E+01
24	FECH	1.04E+00	2.39E+01	2.29E+01
25	MSH2	2.41E+00	5.52E+01	2.29E+01
26	SLC25A3	5.86E-01	1.30E+01	2.23E+01
27	RRM2B	1.79E+00	3.79E+01	2.12E+01

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
28	RBM33	1.86E+00	3.88E+01	2.09E+01
29	STAT3	1.04E+00	2.15E+01	2.06E+01
30	TXN	1.66E+00	3.23E+01	1.95E+01
31	SLC16A1	9.30E-01	1.82E+01	1.95E+01
32	GNB2L1	1.04E+00	1.85E+01	1.79E+01
33	DBT	1.68E+00	2.95E+01	1.76E+01
34	RPS28	1.69E+00	2.94E+01	1.74E+01
35	Inf_PA	NA	1.72E+01	1.72E+01
36	NPM1	2.56E+00	4.17E+01	1.63E+01
37	Inf_PB1	NA	1.62E+01	1.62E+01
38	NXF1	2.64E+00	3.99E+01	1.51E+01
39	PRPF8	1.47E+00	1.98E+01	1.35E+01
40	FLNA	1.19E+00	1.51E+01	1.27E+01
41	DAP3	1.47E+00	1.70E+01	1.15E+01
42	ENO1	1.79E+00	2.05E+01	1.14E+01
43	RBM39	2.38E+00	2.55E+01	1.07E+01
44	KPNB1	1.15E+00	1.21E+01	1.05E+01
45	CFL1	2.32E+00	2.23E+01	9.61E+00
46	DYNLL1;DYNLL2	1.45E+00	1.24E+01	8.60E+00
47	YWHAZ	1.64E+00	1.40E+01	8.58E+00
48	CAPZA1	1.96E+00	1.41E+01	7.19E+00
49	EEF2	2.53E+00	1.81E+01	7.18E+00
50	PRPF40A	3.61E+00	2.55E+01	7.06E+00
51	SLTM	5.10E+01	3.32E+02	6.52E+00
52	RPL22	6.64E+00	3.57E+01	5.38E+00
53	SF3B1	3.90E+00	2.00E+01	5.13E+00
54	RPL27	6.09E+00	2.43E+01	4.00E+00
55	ANXA2;ANXA2P2	6.52E+00	2.04E+01	3.13E+00
56	AIMP2	9.59E+00	2.71E+01	2.83E+00
57	KPRP	6.58E+00	1.04E+01	1.59E+00
58	EIF3E	1.29E+01	1.42E+01	1.10E+00

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
59	RPS7	7.42E+01	7.38E+01	9.94E-01
60	SF3B4	1.77E+01	1.61E+01	9.12E-01
61	HIGD1A	2.43E+01	1.53E+01	6.29E-01
62	KPNA1	1.36E+03	7.20E+02	5.31E-01
63	HAGH	4.45E+01	1.80E+01	4.05E-01
64	AZGP1	9.99E+01	3.22E+01	3.22E-01
65	KPNA6	7.10E+02	2.25E+02	3.17E-01
66	XP32	9.09E+01	2.41E+01	2.65E-01
67	SF3B2	5.53E+01	1.44E+01	2.60E-01
68	PHB	7.25E+01	1.45E+01	2.00E-01
69	ANP32B	6.49E+03	1.06E+03	1.64E-01
70	IMPDH2	1.12E+01	9.94E-01	8.86E-02
71	FHL1	1.14E+01	NA	8.80E-02
72	PKP1	1.18E+01	9.93E-01	8.43E-02
73	PRDX2	1.30E+01	1.05E+00	8.06E-02
74	COPA	1.26E+01	NA	7.95E-02
75	POF1B	1.31E+01	1.00E+00	7.63E-02
76	SLC25A11	1.02E+01	7.50E-01	7.38E-02
77	SMC2	1.54E+01	NA	6.49E-02
78	HTATSF1	1.23E+01	7.50E-01	6.10E-02
79	HMG2;HMG3	1.65E+01	NA	6.08E-02
80	P4HB	1.66E+01	9.85E-01	5.94E-02
81	NACA	1.74E+01	9.19E-01	5.28E-02
82	PRSS3	1.00E+01	5.07E-01	5.06E-02
83	YWHAH	2.06E+01	NA	4.87E-02
84	PSMC4	2.50E+01	NA	4.00E-02
85	SHMT2	1.46E+01	5.74E-01	3.93E-02
86	HNRNP1	3.73E+01	1.41E+00	3.77E-02
87	SNRPD2	1.71E+01	6.23E-01	3.65E-02
88	RPS21	3.65E+01	NA	2.74E-02
89	CALML3	4.74E+01	9.99E-01	2.11E-02

#	Gene	Mean FC (mock)	Mean FC (infected)	FC (infected) / FC (mock)
90	MRPL4	2.76E+01	5.65E-01	2.05E-02
91	KPNA5	5.46E+01	NA	1.83E-02
92	S100A9	4.27E+01	7.70E-01	1.80E-02
93	NME1	4.49E+01	7.70E-01	1.71E-02
94	TFAM	3.97E+01	5.05E-01	1.27E-02
95	TGM3	4.41E+01	5.27E-01	1.20E-02
96	LGALS7	4.75E+01	5.30E-01	1.12E-02
97	DACH1	7.37E+01	5.43E-01	7.37E-03
98	SNRPA	1.59E+01	9.46E-02	5.95E-03
99	SFN	1.78E+02	9.92E-01	5.58E-03
100	LSM12	2.74E+01	1.36E-01	4.96E-03

Appendix J RNA-binding proteins in human A549 cells.

#	ENSG	Symbol	Adjusted <i>p</i> -value	Quantification	RBP
1	ENSG00000188846	RPL14	6.93E-10	quantified	yes
2	ENSG00000161016	RPL8	8.65E-10	quantified	yes
3	ENSG00000181163	NPM1	1.30E-09	quantified	yes
4	ENSG00000162244	RPL29	3.99E-09	quantified	yes
5	ENSG00000131469	RPL27	5.05E-09	quantified	yes
6	ENSG00000065978	YBX1	2.81E-07	quantified	yes
7	ENSG00000135316	SYNCRIP	6.90E-07	quantified	yes
8	ENSG00000166441	RPL27A	1.57E-06	quantified	yes
9	ENSG00000011304	PTBP1	2.91E-06	quantified	yes
10	ENSG00000099622	CIRBP	3.40E-06	quantified	yes
11	ENSG00000106344	RBM28	7.56E-06	quantified	yes
12	ENSG00000099783	HNRNPM	1.03E-05	quantified	yes
13	ENSG00000142534	RPS11	2.23E-05	quantified	yes
14	ENSG00000130305	NSUN5	2.63E-05	quantified	yes
15	ENSG00000066044	ELAVL1	4.90E-05	quantified	yes
16	ENSG00000186468	RPS23	1.30E-04	quantified	yes
17	ENSG00000105640	RPL18A	2.61E-04	quantified	yes
18	ENSG00000140988	RPS2	2.70E-04	quantified	yes
19	ENSG00000169045	HNRNPH1	3.85E-04	quantified	yes
20	ENSG00000052749	RRP12	4.33E-04	quantified	yes
21	ENSG00000142937	RPS8	4.51E-04	quantified	yes
22	ENSG00000121390	PSPC1	6.47E-04	quantified	yes
23	ENSG00000125944	HNRNPR	8.20E-04	quantified	yes
24	ENSG00000134697	GNL2	9.72E-04	quantified	yes
25	ENSG00000115368	WDR75	1.95E-03	quantified	yes
26	ENSG00000115816	CEBPZ	2.24E-03	quantified	yes
27	ENSG00000116560	SFPQ	3.25E-03	quantified	yes
28	ENSG00000148773	MKI67	3.78E-03	quantified	yes
29	ENSG00000186416	NKRF	4.05E-03	quantified	yes

30	ENSG00000107937	GTPBP4	4.83E-03	quantified	yes
31	ENSG00000173145	NOC3L	5.21E-03	quantified	yes
32	ENSG00000152795	HNRNPDL	5.88E-03	quantified	yes
33	ENSG00000090621	PABPC4	6.90E-03	quantified	yes
34	ENSG00000126945	HNRNPH2	1.02E-02	quantified	yes
35	ENSG00000261236	BOP1	1.04E-02	quantified	yes
36	ENSG00000144713	RPL32	1.40E-02	quantified	yes
37	ENSG00000135486	HNRNPA1	1.65E-02	quantified	yes
38	ENSG00000162613	FUBP1	1.76E-02	quantified	yes
39	ENSG00000147140	NONO	1.92E-02	quantified	yes
40	ENSG00000130713	EXOSC2	2.22E-02	quantified	yes
41	ENSG00000138385	SSB	2.25E-02	quantified	yes
42	ENSG00000170889	RPS9	2.25E-02	quantified	yes
43	ENSG00000088247	KHSRP	2.28E-02	quantified	yes
44	ENSG00000130935	NOL11	3.63E-02	quantified	yes
45	ENSG00000142207	URB1	3.65E-02	quantified	yes
46	ENSG00000111364	DDX55	4.99E-02	quantified	yes
47	ENSG00000185298	CCDC137	9.40E-02	semi_quantified	yes
48	ENSG00000089220	PEBP1	1.34E-01	semi_quantified	yes
49	ENSG00000161011	SQSTM1	1.64E-01	semi_quantified	yes
50	ENSG00000171824	EXOSC10	1.74E-01	semi_quantified	yes
51	ENSG00000011260	UTP18	1.78E-01	semi_quantified	yes
52	ENSG00000137547	MRPL15	2.10E-01	semi_quantified	yes
53	ENSG00000155438	NIFK	2.20E-01	semi_quantified	yes
54	ENSG00000182718	ANXA2	2.20E-01	semi_quantified	yes
55	ENSG00000083845	RPS5	2.41E-01	semi_quantified	yes
56	ENSG00000178209	PLEC	2.49E-01	semi_quantified	yes
57	ENSG00000134644	PUM1	4.67E-01	semi_quantified	yes
58	ENSG00000167526	RPL13	4.73E-01	semi_quantified	yes
59	ENSG00000103194	USP10	5.19E-01	semi_quantified	yes
60	ENSG00000067334	DNTTIP2	6.27E-01	semi_quantified	yes
61	ENSG00000120800	UTP20	7.18E-01	semi_quantified	yes
62	ENSG00000107164	FUBP3	7.60E-01	semi_quantified	yes

63	ENSG00000162408	NOL9	7.68E-01	semi_quantified	yes
64	ENSG00000001497	LAS1L		semi_quantified	yes
65	ENSG00000005007	UPF1		semi_quantified	yes
66	ENSG00000005100	DHX33		semi_quantified	yes
67	ENSG00000008988	RPS20		semi_quantified	yes
68	ENSG00000009307	CSDE1		semi_quantified	yes
69	ENSG00000037474	NSUN2		semi_quantified	yes
70	ENSG00000039123	SKIV2L2		semi_quantified	yes
71	ENSG00000040341	STAU2		semi_quantified	yes
72	ENSG00000047849	MAP4		semi_quantified	yes
73	ENSG00000055044	NOP58		semi_quantified	yes
74	ENSG00000055483	USP36		semi_quantified	yes
75	ENSG00000055917	PUM2		semi_quantified	yes
76	ENSG00000056097	ZFR		semi_quantified	yes
77	ENSG00000060138	YBX3		semi_quantified	yes
78	ENSG00000063046	EIF4B		semi_quantified	yes
79	ENSG00000063244	U2AF2		semi_quantified	yes
80	ENSG00000065183	WDR3		semi_quantified	yes
81	ENSG00000065526	SPEN		semi_quantified	yes
82	ENSG00000067225	PKM		semi_quantified	yes
83	ENSG00000068784	SRBD1		semi_quantified	yes
84	ENSG00000070756	PABPC1		semi_quantified	yes
85	ENSG00000071082	RPL31		semi_quantified	yes
86	ENSG00000071626	DAZAP1		semi_quantified	yes
87	ENSG00000074800	ENO1		semi_quantified	yes
88	ENSG00000076067	RBMS2		semi_quantified	yes
89	ENSG00000079246	XRCC5		semi_quantified	yes
90	ENSG00000079785	DDX1		semi_quantified	yes
91	ENSG00000080608	PUM3		semi_quantified	yes
92	ENSG00000083223	ZCCHC6		semi_quantified	yes

93	ENSG00000083520	DIS3	semi_quantified	yes
94	ENSG00000087269	NOP14	semi_quantified	yes
95	ENSG00000088930	XRN2	semi_quantified	yes
96	ENSG00000089048	ESF1	semi_quantified	yes
97	ENSG00000089280	FUS	semi_quantified	yes
98	ENSG00000089737	DDX24	semi_quantified	yes
99	ENSG00000092199	HNRNPC	semi_quantified	yes
100	ENSG00000092201	SUPT16H	semi_quantified	yes
101	ENSG00000096080	MRPS18A	semi_quantified	yes
102	ENSG00000096746	HNRNPH3	semi_quantified	yes
103	ENSG00000099821	POLRMT	semi_quantified	yes
104	ENSG00000100029	PES1	semi_quantified	yes
105	ENSG00000100138	SNU13	semi_quantified	yes
106	ENSG00000100201	DDX17	semi_quantified	yes
107	ENSG00000100353	EIF3D	semi_quantified	yes
108	ENSG00000100403	ZC3H7B	semi_quantified	yes
109	ENSG00000100650	SRSF5	semi_quantified	yes
110	ENSG00000100823	APEX1	semi_quantified	yes
111	ENSG00000101361	NOP56	semi_quantified	yes
112	ENSG00000101811	CSTF2	semi_quantified	yes
113	ENSG00000102317	RBM3	semi_quantified	yes
114	ENSG00000103550	KNOP1	semi_quantified	yes
115	ENSG00000103994	ZNF106	semi_quantified	yes
116	ENSG00000104177	MYEF2	semi_quantified	yes
117	ENSG00000104824	HNRNPL	semi_quantified	yes
118	ENSG00000105202	FBL	semi_quantified	yes
119	ENSG00000105364	MRPL4	semi_quantified	yes
120	ENSG00000105447	GRWD1	semi_quantified	yes
121	ENSG00000105671	DDX49	semi_quantified	yes
122	ENSG00000105939	ZC3HAV1	semi_quantified	yes

123	ENSG00000107223	EDF1	semi_quantified	yes
124	ENSG00000107625	DDX50	semi_quantified	yes
125	ENSG00000108256	NUFIP2	semi_quantified	yes
126	ENSG00000108312	UBTF	semi_quantified	yes
127	ENSG00000108518	PFN1	semi_quantified	yes
128	ENSG00000108654	DDX5	semi_quantified	yes
129	ENSG00000108829	LRRC59	semi_quantified	yes
130	ENSG00000109475	RPL34	semi_quantified	yes
131	ENSG00000109534	GAR1	semi_quantified	yes
132	ENSG00000109606	DHX15	semi_quantified	yes
133	ENSG00000109971	HSPA8	semi_quantified	yes
134	ENSG00000110367	DDX6	semi_quantified	yes
135	ENSG00000110700	RPS13	semi_quantified	yes
136	ENSG00000111142	METAP2	semi_quantified	yes
137	ENSG00000111615	KRR1	semi_quantified	yes
138	ENSG00000111786	SRSF9	semi_quantified	yes
139	ENSG00000112081	SRSF3	semi_quantified	yes
140	ENSG00000112306	RPS12	semi_quantified	yes
141	ENSG00000112578	BYSL	semi_quantified	yes
142	ENSG00000113387	SUB1	semi_quantified	yes
143	ENSG00000113460	BRIX1	semi_quantified	yes
144	ENSG00000114391	RPL24	semi_quantified	yes
145	ENSG00000114416	FXR1	semi_quantified	yes
146	ENSG00000114686	MRPL3	semi_quantified	yes
147	ENSG00000114867	EIF4G1	semi_quantified	yes
148	ENSG00000115268	RPS15	semi_quantified	yes
149	ENSG00000115524	SF3B1	semi_quantified	yes
150	ENSG00000115677	HDLBP	semi_quantified	yes
151	ENSG00000115761	NOL10	semi_quantified	yes
152	ENSG00000115875	SRSF7	semi_quantified	yes

153	ENSG00000116120	FARSB	semi_quantified	yes
154	ENSG00000116251	RPL22	semi_quantified	yes
155	ENSG00000117133	RPF1	semi_quantified	yes
156	ENSG00000117395	EBNA1BP2	semi_quantified	yes
157	ENSG00000118246	FASTKD2	semi_quantified	yes
158	ENSG00000119285	HEATR1	semi_quantified	yes
159	ENSG00000119314	PTBP3	semi_quantified	yes
160	ENSG00000119616	FCF1	semi_quantified	yes
161	ENSG00000119705	SLIRP	semi_quantified	yes
162	ENSG00000119812	FAM98A	semi_quantified	yes
163	ENSG00000120948	TARDBP	semi_quantified	yes
164	ENSG00000121060	TRIM25	semi_quantified	yes
165	ENSG00000121274	PAPD5	semi_quantified	yes
166	ENSG00000121486	TRMT1L	semi_quantified	yes
167	ENSG00000121774	KHDRBS1	semi_quantified	yes
168	ENSG00000122026	RPL21	semi_quantified	yes
169	ENSG00000122299	ZC3H7A	semi_quantified	yes
170	ENSG00000122406	RPL5	semi_quantified	yes
171	ENSG00000122566	HNRNPA2B1	semi_quantified	yes
172	ENSG00000124193	SRSF6	semi_quantified	yes
173	ENSG00000124214	STAU1	semi_quantified	yes
174	ENSG00000124541	RRP36	semi_quantified	yes
175	ENSG00000124571	XPO5	semi_quantified	yes
176	ENSG00000124942	AHNAK	semi_quantified	yes
177	ENSG00000125485	DDX31	semi_quantified	yes
178	ENSG00000125691	RPL23	semi_quantified	yes
179	ENSG00000125970	RALY	semi_quantified	yes
180	ENSG00000127804	METTL16	semi_quantified	yes
181	ENSG00000128534	LSM8	semi_quantified	yes
182	ENSG00000129245	FXR2	semi_quantified	yes

183	ENSG00000129347	KRI1	semi_quantified	yes
184	ENSG00000129460	NGDN	semi_quantified	yes
185	ENSG00000130119	GNL3L	semi_quantified	yes
186	ENSG00000130255	RPL36	semi_quantified	yes
187	ENSG00000130312	MRPL34	semi_quantified	yes
188	ENSG00000130811	EIF3G	semi_quantified	yes
189	ENSG00000130826	DKC1	semi_quantified	yes
190	ENSG00000131051	RBM39	semi_quantified	yes
191	ENSG00000132275	RRP8	semi_quantified	yes
192	ENSG00000132300	PTCD3	semi_quantified	yes
193	ENSG00000132382	MYBBP1A	semi_quantified	yes
194	ENSG00000132463	GRSF1	semi_quantified	yes
195	ENSG00000132467	UTP3	semi_quantified	yes
196	ENSG00000132507	EIF5A	semi_quantified	yes
197	ENSG00000132591	ERAL1	semi_quantified	yes
198	ENSG00000132603	NIP7	semi_quantified	yes
199	ENSG00000134987	WDR36	semi_quantified	yes
200	ENSG00000135046	ANXA1	semi_quantified	yes
201	ENSG00000135372	NAT10	semi_quantified	yes
202	ENSG00000135763	URB2	semi_quantified	yes
203	ENSG00000136231	IGF2BP3	semi_quantified	yes
204	ENSG00000136270	TBRG4	semi_quantified	yes
205	ENSG00000136450	SRSF1	semi_quantified	yes
206	ENSG00000136463	TACO1	semi_quantified	yes
207	ENSG00000136527	TRA2B	semi_quantified	yes
208	ENSG00000137073	UBAP2	semi_quantified	yes
209	ENSG00000138326	RPS24	semi_quantified	yes
210	ENSG00000138668	HNRNPD	semi_quantified	yes
211	ENSG00000138757	G3BP2	semi_quantified	yes
212	ENSG00000139131	YARS2	semi_quantified	yes

213	ENSG00000139168	ZCRB1	semi_quantified	yes
214	ENSG00000139233	LLPH	semi_quantified	yes
215	ENSG00000141076	CIRH1A	semi_quantified	yes
216	ENSG00000142864	SERBP1	semi_quantified	yes
217	ENSG00000143319	ISG20L2	semi_quantified	yes
218	ENSG00000143368	SF3B4	semi_quantified	yes
219	ENSG00000143436	MRPL9	semi_quantified	yes
220	ENSG00000143569	UBAP2L	semi_quantified	yes
221	ENSG00000143621	ILF2	semi_quantified	yes
222	ENSG00000143889	HNRNPLL	semi_quantified	yes
223	ENSG00000143947	RPS27A	semi_quantified	yes
224	ENSG00000144161	ZC3H8	semi_quantified	yes
225	ENSG00000145425	RPS3A	semi_quantified	yes
226	ENSG00000145741	BTF3	semi_quantified	yes
227	ENSG00000145907	G3BP1	semi_quantified	yes
228	ENSG00000145912	NHP2	semi_quantified	yes
229	ENSG00000146223	RPL7L1	semi_quantified	yes
230	ENSG00000146909	NOM1	semi_quantified	yes
231	ENSG00000147274	RBMX	semi_quantified	yes
232	ENSG00000147403	RPL10	semi_quantified	yes
233	ENSG00000147679	UTP23	semi_quantified	yes
234	ENSG00000147905	ZCCHC7	semi_quantified	yes
235	ENSG00000148296	SURF6	semi_quantified	yes
236	ENSG00000148300	REXO4	semi_quantified	yes
237	ENSG00000149187	CELF1	semi_quantified	yes
238	ENSG00000150990	DHX37	semi_quantified	yes
239	ENSG00000151304	SRFBP1	semi_quantified	yes
240	ENSG00000151846	PABPC3	semi_quantified	yes
241	ENSG00000154719	MRPL39	semi_quantified	yes
242	ENSG00000155363	MOV10	semi_quantified	yes

243	ENSG00000156482	RPL30	semi_quantified	yes
244	ENSG00000158417	EIF5B	semi_quantified	yes
245	ENSG00000159217	IGF2BP1	semi_quantified	yes
246	ENSG00000160208	RRP1B	semi_quantified	yes
247	ENSG00000160214	RRP1	semi_quantified	yes
248	ENSG00000160710	ADAR	semi_quantified	yes
249	ENSG00000160818	GPATCH4	semi_quantified	yes
250	ENSG00000161547	SRSF2	semi_quantified	yes
251	ENSG00000161813	LARP4	semi_quantified	yes
252	ENSG00000161960	EIF4A1	semi_quantified	yes
253	ENSG00000162231	NXF1	semi_quantified	yes
254	ENSG00000162664	ZNF326	semi_quantified	yes
255	ENSG00000162775	RBM15	semi_quantified	yes
256	ENSG00000163214	DHX57	semi_quantified	yes
257	ENSG00000163694	RBM47	semi_quantified	yes
258	ENSG00000163811	WDR43	semi_quantified	yes
259	ENSG00000163931	TKT	semi_quantified	yes
260	ENSG00000163938	GNL3	semi_quantified	yes
261	ENSG00000164338	UTP15	semi_quantified	yes
262	ENSG00000164346	NSA2	semi_quantified	yes
263	ENSG00000164587	RPS14	semi_quantified	yes
264	ENSG00000165119	HNRNPK	semi_quantified	yes
265	ENSG00000165271	NOL6	semi_quantified	yes
266	ENSG00000166197	NOLC1	semi_quantified	yes
267	ENSG00000166889	PATL1	semi_quantified	yes
268	ENSG00000167112	TRUB2	semi_quantified	yes
269	ENSG00000167658	EEF2	semi_quantified	yes
270	ENSG00000169564	PCBP1	semi_quantified	yes
271	ENSG00000169714	CNBP	semi_quantified	yes
272	ENSG00000169813	HNRNPF	semi_quantified	yes

273	ENSG00000170144	HNRNPA3	semi_quantified	yes
274	ENSG00000171863	RPS7	semi_quantified	yes
275	ENSG00000172172	MRPL13	semi_quantified	yes
276	ENSG00000173141	MRPL57	semi_quantified	yes
277	ENSG00000173545	ZNF622	semi_quantified	yes
278	ENSG00000173914	RBM4B	semi_quantified	yes
279	ENSG00000173933	RBM4	semi_quantified	yes
280	ENSG00000174720	LARP7	semi_quantified	yes
281	ENSG00000174953	DHX36	semi_quantified	yes
282	ENSG00000177971	IMP3	semi_quantified	yes
283	ENSG00000178105	DDX10	semi_quantified	yes
284	ENSG00000179041	RRS1	semi_quantified	yes
285	ENSG00000179119	SPTY2D1	semi_quantified	yes
286	ENSG00000182154	MRPL41	semi_quantified	yes
287	ENSG00000182944	EWSR1	semi_quantified	yes
288	ENSG00000183311	TUBB	semi_quantified	yes
289	ENSG00000183520	UTP11L	semi_quantified	yes
290	ENSG00000183570	PCBP3	semi_quantified	yes
291	ENSG00000183684	ALYREF	semi_quantified	yes
292	ENSG00000183808	RBM12B	semi_quantified	yes
293	ENSG00000184967	NOC4L	semi_quantified	yes
294	ENSG00000187051	RPS19BP1	semi_quantified	yes
295	ENSG00000187772	LIN28B	semi_quantified	yes
296	ENSG00000188529	SRSF10	semi_quantified	yes
297	ENSG00000188573	FBLL1	semi_quantified	yes
298	ENSG00000188739	RBM34	semi_quantified	yes
299	ENSG00000188976	NOC2L	semi_quantified	yes
300	ENSG00000196141	SPATS2L	semi_quantified	yes
301	ENSG00000196419	XRCC6	semi_quantified	yes
302	ENSG00000196924	FLNA	semi_quantified	yes

303	ENSG00000196943	NOP9		semi_quantified	yes
304	ENSG00000197111	PCBP2		semi_quantified	yes
305	ENSG00000197157	SND1		semi_quantified	yes
306	ENSG00000197451	HNRNPAB		semi_quantified	yes
307	ENSG00000197498	RPF2		semi_quantified	yes
308	ENSG00000197728	RPS26		semi_quantified	yes
309	ENSG00000198000	NOL8		semi_quantified	yes
310	ENSG00000198015	MRPL42		semi_quantified	yes
311	ENSG00000198034	RPS4X		semi_quantified	yes
312	ENSG00000198301	SDAD1		semi_quantified	yes
313	ENSG00000198755	RPL10A		semi_quantified	yes
314	ENSG00000198917	C9orf114		semi_quantified	yes
315	ENSG00000204221	WDR46		semi_quantified	yes
316	ENSG00000204469	PRRC2A		semi_quantified	yes
317	ENSG00000204574	ABCF1		semi_quantified	yes
318	ENSG00000204628	GNB2L1		semi_quantified	yes
319	ENSG00000205560	CPT1B		semi_quantified	yes
320	ENSG00000213782	DDX47		semi_quantified	yes
321	ENSG00000215301	DDX3X		semi_quantified	yes
322	ENSG00000241945	PWP2		semi_quantified	yes
323	ENSG00000247315	ZCCHC3		semi_quantified	yes
324	ENSG00000253729	PRKDC		semi_quantified	yes
325	ENSG00000254772	EEF1G		semi_quantified	yes
326	ENSG00000257103	LSM14A		semi_quantified	yes
327	ENSG00000259956	RBM15B		semi_quantified	yes
328	ENSG00000265681	RPL17		semi_quantified	yes
329	ENSG00000270647	TAF15		semi_quantified	yes
330	ENSG00000275700	AATF		semi_quantified	yes
331	ENSG00000277594	DDX52		semi_quantified	yes
332	ENSG00000115053	NCL	2.43E-22	quantified	yes
333	ENSG00000171490	RSL1D1	1.83E-13	quantified	yes

334	ENSG00000136942	RPL35	4.13E-13	quantified	yes
335	ENSG00000165733	BMS1	1.45E-11	quantified	yes
336	ENSG00000147604	RPL7	1.67E-10	quantified	yes
337	ENSG00000063177	RPL18	3.99E-09	quantified	yes
338	ENSG00000165732	DDX21	3.99E-09	quantified	yes
339	ENSG00000123064	DDX54	4.83E-09	quantified	yes
340	ENSG00000135829	DHX9	5.11E-09	quantified	yes
341	ENSG00000135404	CD63	6.66E-09	quantified	yes
342	ENSG00000136271	DDX56	1.03E-08	quantified	yes
343	ENSG00000108592	FTSJ3	1.94E-08	quantified	yes
344	ENSG00000089009	RPL6	6.04E-08	quantified	yes
345	ENSG00000239306	RBM14	6.04E-08	quantified	yes
346	ENSG00000143977	SNRPG	6.55E-08	quantified	yes
347	ENSG00000153187	HNRNPU	1.42E-07	quantified	yes
348	ENSG00000048162	NOP16	1.59E-07	quantified	yes
349	ENSG00000138095	LRPPRC	2.29E-07	quantified	yes
350	ENSG00000174444	RPL4	1.84E-06	quantified	yes
351	ENSG00000124228	DDX27	3.07E-06	quantified	yes
352	ENSG00000198242	RPL23A	4.50E-06	quantified	yes
353	ENSG00000122965	RBM19	1.23E-05	quantified	yes
354	ENSG00000133773	CCDC59	1.23E-05	quantified	yes
355	ENSG00000129351	ILF3	1.55E-05	quantified	yes
356	ENSG00000148303	RPL7A	1.60E-05	quantified	yes
357	ENSG00000048828	FAM120A	2.63E-05	quantified	yes
358	ENSG00000106682	EIF4H	2.63E-05	quantified	yes
359	ENSG00000174748	RPL15	3.72E-05	quantified	yes
360	ENSG00000151923	TIAL1	4.90E-05	quantified	yes
361	ENSG00000100836	PABPN1	1.75E-04	quantified	yes
362	ENSG00000170515	PA2G4	2.82E-04	quantified	yes
363	ENSG00000135387	CAPRIN1	2.97E-04	quantified	yes
364	ENSG00000116001	TIA1	3.08E-04	quantified	yes
365	ENSG00000142541	RPL13A	3.47E-04	quantified	yes
366	ENSG00000177733	HNRNPA0	5.04E-04	quantified	yes

367	ENSG00000156697	UTP14A	9.08E-04	quantified	yes
368	ENSG00000108298	RPL19	1.09E-03	quantified	yes
369	ENSG00000149273	RPS3	1.12E-03	quantified	yes
370	ENSG00000117523	PRRC2C	1.95E-03	quantified	yes
371	ENSG00000104356	POP1	3.26E-03	quantified	yes
372	ENSG00000108107	RPL28	3.32E-03	quantified	yes
373	ENSG00000132153	DHX30	3.47E-03	quantified	yes
374	ENSG00000124383	MPHOSPH10	3.57E-03	quantified	yes
375	ENSG00000088205	DDX18	3.87E-03	quantified	yes
376	ENSG00000111641	NOP2	3.87E-03	quantified	yes
377	ENSG00000138768	USO1	4.03E-03	quantified	yes
378	ENSG00000172867	KRT2	4.05E-03	quantified	yes
379	ENSG00000151834	GABRA2	4.72E-03	quantified	yes
380	ENSG00000158406	HIST1H4H	5.66E-03	quantified	yes
381	ENSG00000096150	RPS18	8.90E-03	quantified	yes
382	ENSG00000185163	DDX51	1.39E-02	quantified	yes
383	ENSG00000105373	GLTSCR2	1.54E-02	quantified	yes
384	ENSG00000015479	MATR3	1.67E-02	quantified	yes
385	ENSG00000086189	DIMT1	1.68E-02	quantified	yes
386	ENSG00000155506	LARP1	1.68E-02	quantified	yes
387	ENSG00000139343	SNRPF	1.92E-02	quantified	yes
388	ENSG00000156508	EEF1A1	2.10E-02	quantified	yes
389	ENSG00000107581	EIF3A	2.25E-02	quantified	yes
390	ENSG00000065150	IPO5	2.95E-02	quantified	yes
391	ENSG00000137154	RPS6	3.26E-02	quantified	yes
392	ENSG00000182774	RPS17	3.26E-02	quantified	yes
393	ENSG00000148843	PDCD11	3.40E-02	quantified	yes
394	ENSG00000105323	HNRNPUL1	3.44E-02	quantified	yes
395	ENSG00000104915	STX10	3.94E-02	quantified	yes
396	ENSG00000100316	RPL3	4.55E-02	quantified	yes
397	ENSG00000164934	DCAF13	5.19E-02	not significant	no
398	ENSG00000106028	SSBP1	5.67E-02	not significant	no
399	ENSG00000110172	CHORDC1	6.86E-02	not significant	no

400	ENSG00000087086	FTL	7.79E-02	not significant	no
401	ENSG00000179820	MYADM	8.52E-02	not significant	no
402	ENSG00000270885	RASL10B	9.98E-02	not significant	no
403	ENSG00000161634	DCD	1.08E-01	not significant	no
404	ENSG00000116030	SUMO1	1.12E-01	not significant	no
405	ENSG00000134760	DSG1	1.12E-01	not significant	no
406	ENSG00000125844	RRBP1	1.45E-01	not significant	no
407	ENSG00000168028	RPSA	1.56E-01	not significant	no
408	ENSG00000147687	TATDN1	1.58E-01	not significant	no
409	ENSG00000100028	SNRPD3	1.63E-01	not significant	no
410	ENSG00000131747	TOP2A	1.97E-01	not significant	no
411	ENSG00000125743	SNRPD2	2.10E-01	not significant	no
412	ENSG00000173801	JUP	2.90E-01	not significant	no
413	ENSG00000136891	TEX10	2.90E-01	not significant	no
414	ENSG00000121552	CSTA	4.34E-01	not significant	no
415	ENSG00000096696	DSP	4.67E-01	not significant	no
416	ENSG00000136026	CKAP4	4.69E-01	not significant	no
417	ENSG00000116221	MRPL37	4.73E-01	not significant	no
418	ENSG00000137491	SLCO2B1	4.73E-01	not significant	no
419	ENSG00000146676	PURB	4.73E-01	not significant	no
420	ENSG00000182004	SNRPE	4.73E-01	not significant	no
421	ENSG00000133983	COX16	5.21E-01	not significant	no
422	ENSG00000134765	DSC1	5.22E-01	not significant	no
423	ENSG00000146109	ABT1	5.22E-01	not significant	no
424	ENSG00000114850	SSR3	5.57E-01	not significant	no
425	ENSG00000182872	RBM10	6.82E-01	not significant	no
426	ENSG00000112651	MRPL2	7.68E-01	not significant	no
427	ENSG00000142700	DMRTA2	7.76E-01	not significant	no
428	ENSG00000105193	RPS16	8.24E-01	not significant	no
429	ENSG00000132341	RAN	8.24E-01	not significant	no
430	ENSG00000172349	IL16	8.34E-01	not significant	no
431	ENSG00000055950	MRPL43	8.40E-01	not significant	no
432	ENSG00000167088	SNRPD1	8.63E-01	not significant	no

433	ENSG00000203786	KPRP	8.63E-01	not significant	no
434	ENSG00000179950	PUF60	9.04E-01	not significant	no
435	ENSG00000168280	KIF5C	9.24E-01	not significant	no
436	ENSG00000174231	PRPF8	9.31E-01	not significant	no
437	ENSG00000172850	LSM2	9.82E-01	not significant	no
438	ENSG00000114650	SCAP	9.91E-01	not significant	no
439	ENSG00000011243	AKAP8L		not significant	no
440	ENSG00000015171	ZMYND11		not significant	no
441	ENSG00000023608	SNAPC1		not significant	no
442	ENSG00000025796	SEC63		not significant	no
443	ENSG00000031691	CENPQ		not significant	no
444	ENSG00000033050	ABCF2		not significant	no
445	ENSG00000050748	MAPK9		not significant	no
446	ENSG00000054118	THRAP3		not significant	no
447	ENSG00000063438	AHRR		not significant	no
448	ENSG00000065548	ZC3H15		not significant	no
449	ENSG00000066654	THUMPDI		not significant	no
450	ENSG00000067221	STOML1		not significant	no
451	ENSG00000067248	DHX29		not significant	no
452	ENSG00000072210	ALDH3A2		not significant	no
453	ENSG00000073792	IGF2BP2		not significant	no
454	ENSG00000075420	FNDC3B		not significant	no
455	ENSG00000075624	ACTB		not significant	no
456	ENSG00000077097	TOP2B		not significant	no
457	ENSG00000080824	HSP90AA1		not significant	no
458	ENSG00000081277	PKP1		not significant	no
459	ENSG00000085662	AKR1B1		not significant	no
460	ENSG00000086589	RBM22		not significant	no
461	ENSG00000086827	ZW10		not significant	no
462	ENSG00000090905	TNRC6A		not significant	no
463	ENSG00000096384	HSP90AB1		not significant	no

464	ENSG00000099139	PCSK5	not significant	no
465	ENSG00000100038	TOP3B	not significant	no
466	ENSG00000100567	PSMA3	not significant	no
467	ENSG00000100605	ITPK1	not significant	no
468	ENSG00000102271	KLHL4	not significant	no
469	ENSG00000104131	EIF3J	not significant	no
470	ENSG00000106355	LSM5	not significant	no
471	ENSG00000108100	CCNY	not significant	no
472	ENSG00000108255	CRYBA1	not significant	no
473	ENSG00000108344	PSMD3	not significant	no
474	ENSG00000108551	RASD1	not significant	no
475	ENSG00000108578	BLMH	not significant	no
476	ENSG00000110321	EIF4G2	not significant	no
477	ENSG00000111057	KRT18	not significant	no
478	ENSG00000111640	GAPDH	not significant	no
479	ENSG00000111907	TPD52L1	not significant	no
480	ENSG00000114767	RRP9	not significant	no
481	ENSG00000115946	PNO1	not significant	no
482	ENSG00000117114	ADGRL2	not significant	no
483	ENSG00000117115	PADI2	not significant	no
484	ENSG00000117477	CCDC181	not significant	no
485	ENSG00000117713	ARID1A	not significant	no
486	ENSG00000117859	OSBPL9	not significant	no
487	ENSG00000120802	TMPO	not significant	no
488	ENSG00000120860	CCDC53	not significant	no
489	ENSG00000121897	LIAS	not significant	no
490	ENSG00000123737	EXOSC9	not significant	no
491	ENSG00000124092	CTCF	not significant	no
492	ENSG00000124721	DNAH8	not significant	no
493	ENSG00000125482	TTF1	not significant	no

494	ENSG00000126254	RBM42	not significant	no
495	ENSG00000130035	GALNT8	not significant	no
496	ENSG00000130764	LRRC47	not significant	no
497	ENSG00000130816	DNMT1	not significant	no
498	ENSG00000131095	GFAP	not significant	no
499	ENSG00000132361	CLUH	not significant	no
500	ENSG00000132466	ANKRD17	not significant	no
501	ENSG00000133313	CNDP2	not significant	no
502	ENSG00000133315	MACROD1	not significant	no
503	ENSG00000134504	KCTD1	not significant	no
504	ENSG00000135269	TES	not significant	no
505	ENSG00000135972	MRPS9	not significant	no
506	ENSG00000136535	TBR1	not significant	no
507	ENSG00000138061	CYP1B1	not significant	no
508	ENSG00000140416	TPM1	not significant	no
509	ENSG00000140694	PARN	not significant	no
510	ENSG00000142230	SAE1	not significant	no
511	ENSG00000142676	RPL11	not significant	no
512	ENSG00000143748	NVL	not significant	no
513	ENSG00000143933	CALM2	not significant	no
514	ENSG00000145050	MANF	not significant	no
515	ENSG00000146416	AIG1	not significant	no
516	ENSG00000147649	MTDH	not significant	no
517	ENSG00000148377	IDI2	not significant	no
518	ENSG00000148688	RPP30	not significant	no
519	ENSG00000149136	SSRP1	not significant	no
520	ENSG00000151623	NR3C2	not significant	no
521	ENSG00000153404	PLEKHG4B	not significant	no
522	ENSG00000153487	ING1	not significant	no
523	ENSG00000154258	ABCA9	not significant	no

524	ENSG00000154864	PIEZO2	not significant	no
525	ENSG00000154957	ZNF18	not significant	no
526	ENSG00000155926	SLA	not significant	no
527	ENSG00000156587	UBE2L6	not significant	no
528	ENSG00000158710	TAGLN2	not significant	no
529	ENSG00000160213	CSTB	not significant	no
530	ENSG00000160679	CHTOP	not significant	no
531	ENSG00000161653	NAGS	not significant	no
532	ENSG00000161847	RAVER1	not significant	no
533	ENSG00000163584	RPL22L1	not significant	no
534	ENSG00000163825	RTP3	not significant	no
535	ENSG00000164068	RNF123	not significant	no
536	ENSG00000164134	NAA15	not significant	no
537	ENSG00000164548	TRA2A	not significant	no
538	ENSG00000164687	FABP5	not significant	no
539	ENSG00000165092	ALDH1A1	not significant	no
540	ENSG00000165140	FBP1	not significant	no
541	ENSG00000166152	C16orf78	not significant	no
542	ENSG00000166226	CCT2	not significant	no
543	ENSG00000166902	MRPL16	not significant	no
544	ENSG00000167491	GATAD2A	not significant	no
545	ENSG00000168066	SF1	not significant	no
546	ENSG00000168071	CCDC88B	not significant	no
547	ENSG00000168488	ATXN2L	not significant	no
548	ENSG00000169239	CA5B	not significant	no
549	ENSG00000169871	TRIM56	not significant	no
550	ENSG00000171262	FAM98B	not significant	no
551	ENSG00000171365	CLCN5	not significant	no
552	ENSG00000171848	RRM2	not significant	no
553	ENSG00000171861	RNMTL1	not significant	no

554	ENSG00000172757	CFL1	not significant	no
555	ENSG00000174332	GLIS1	not significant	no
556	ENSG00000174780	SRP72	not significant	no
557	ENSG00000177409	SAMD9L	not significant	no
558	ENSG00000179526	SHARPIN	not significant	no
559	ENSG00000179912	R3HDM2	not significant	no
560	ENSG00000183431	SF3A3	not significant	no
561	ENSG00000184083	FAM120C	not significant	no
562	ENSG00000184640	SEPT9	not significant	no
563	ENSG00000185721	DRG1	not significant	no
564	ENSG00000188089	PLA2G4E	not significant	no
565	ENSG00000189306	RRP7A	not significant	no
566	ENSG00000189403	HMGB1	not significant	no
567	ENSG00000196139	AKR1C3	not significant	no
568	ENSG00000197312	DDI2	not significant	no
569	ENSG00000198363	ASPH	not significant	no
570	ENSG00000198900	TOP1	not significant	no
571	ENSG00000198959	TGM2	not significant	no
572	ENSG00000204120	GIGYF2	not significant	no
573	ENSG00000204634	TBC1D8	not significant	no
574	ENSG00000214753	HNRNPUL2	not significant	no
575	ENSG00000218823	PAPOLB	not significant	no
576	ENSG00000225921	NOL7	not significant	no
577	ENSG00000233927	RPS28	not significant	no
578	ENSG00000262874	C19orf84	not significant	no
579	ENSG00000271793	RP11-321N4.5	not significant	no
580	ENSG00000274209	ANTXRL	not significant	no
581	ENSG00000274523	WBSCR16	not significant	no
582	ENSG00000277936	MRPL45	not significant	no

Appendix K RBPome of A549 cells 9 h.p.i. with influenza A/WSN/33 (H1N1) virus

#	ENSG	Symbol	Known RBP	FC (log2)	Adjusted <i>p</i> -value
1	Inf_P15682_NCAP	Inf_NP	no	6.068	3.33E-71
2	Inf_Q20MH3_NS1	Inf_NS1	yes	4.228	1.94E-37
3	ENSG00000028203	VEZT	no	7.217	1.34E-36
4	ENSG00000173545	ZNF622	yes	-3.185	3.54E-07
5	ENSG00000130119	GNL3L	yes	1.765	1.16E-06
6	ENSG00000143368	SF3B4	yes	-1.660	6.67E-06
7	ENSG00000177728	TMEM94	no	-2.005	8.57E-06
8	ENSG00000106211	HSPB1	yes	-2.684	3.37E-05
9	ENSG00000005007	UPF1	yes	-1.534	3.86E-05
10	ENSG00000100722	ZC3H14	yes	-1.579	1.53E-03
12	ENSG00000070756	PABPC1	yes	1.229	2.93E-03
11	ENSG00000039123	SKIV2L2	yes	-1.234	2.93E-03
13	ENSG00000090621	PABPC4	yes	1.216	3.26E-03
14	ENSG00000149187	CELF1	yes	-1.205	3.49E-03
15	ENSG00000111057	KRT18	yes	-1.159	5.95E-03
17	ENSG00000169251	NMD3	yes	1.995	5.95E-03
16	ENSG00000048991	R3HDM1	yes	-1.150	5.95E-03
18	ENSG00000183684	ALYREF	yes	-1.113	9.01E-03
19	ENSG00000183570	PCBP3	yes	-1.328	1.21E-02
20	ENSG00000185721	DRG1	no	-1.071	1.35E-02
21	ENSG00000163811	WDR43	yes	1.066	1.37E-02
22	ENSG00000160710	ADAR	yes	1.052	1.56E-02
23	ENSG00000143742	SRP9	no	-1.266	1.77E-02
24	ENSG00000151923	TIAL1	yes	-1.036	1.77E-02
25	ENSG00000152601	MBNL1	yes	-0.990	2.83E-02
26	ENSG00000040341	STAU2	yes	-0.984	2.89E-02
28	ENSG00000066044	ELAVL1	yes	-0.947	4.06E-02
27	ENSG00000179912	R3HDM2	yes	-0.949	4.06E-02

29	ENSG00000168066	SF1	yes	1.607	4.81E-02
30	ENSG00000101811	CSTF2	yes	-1.556	6.37E-02
31	ENSG00000160679	CHTOP	yes	-0.895	6.41E-02
33	ENSG00000138646	HERC5	yes	-1.073	7.08E-02
32	ENSG00000173110	HSPA6	no	-1.516	7.08E-02
34	ENSG00000134644	PUM1	yes	-0.878	7.08E-02
35	ENSG00000116001	TIA1	yes	-0.873	7.08E-02
36	ENSG00000107164	FUBP3	yes	-0.865	7.38E-02
37	ENSG00000242732	RGAG4	no	-0.864	7.38E-02
38	ENSG00000163694	RBM47	yes	-0.853	8.03E-02
39	ENSG00000119314	PTBP3	yes	-0.837	9.18E-02
40	ENSG00000047849	MAP4	yes	-0.826	9.93E-02
41	ENSG00000137449	CPEB2	yes	-0.822	1.01E-01
42	ENSG00000157110	RBPMS	yes	-0.817	1.04E-01
43	ENSG00000147065	MSN	no	-0.805	1.13E-01
44	ENSG00000204842	ATXN2	yes	0.799	1.17E-01
45	ENSG00000182872	RBM10	yes	-0.788	1.28E-01
47	ENSG00000182481	KPNA2	yes	1.342	1.40E-01
46	ENSG00000163931	TKT	no	-0.773	1.40E-01
48	ENSG00000254772	EEF1G	no	-0.770	1.41E-01
49	ENSG00000132361	CLUH	no	-1.327	1.42E-01
50	ENSG00000143486	EIF2D	yes	-0.763	1.42E-01
51	ENSG00000188529	SRSF10	yes	-0.762	1.42E-01
52	ENSG00000155438	NIFK	yes	0.752	1.53E-01
54	ENSG00000071626	DAZAP1	yes	-0.740	1.61E-01
55	ENSG00000103550	KNOP1	yes	0.744	1.61E-01
53	ENSG00000142864	SERBP1	yes	-0.742	1.61E-01
56	ENSG00000130935	NOL11	yes	0.735	1.66E-01
57	ENSG00000107223	EDF1	yes	-0.727	1.75E-01
58	ENSG00000048828	FAM120A	yes	-0.721	1.79E-01
59	ENSG00000139343	SNRPF	no	0.721	1.79E-01
61	ENSG00000152795	HNRNPDL	yes	-0.718	1.80E-01
62	ENSG00000197111	PCBP2	yes	-0.714	1.80E-01

60	ENSG00000124201	ZNFX1	yes	-1.240	1.80E-01
63	ENSG00000105321	CCDC9	yes	-0.872	1.80E-01
64	ENSG00000116030	SUMO1	yes	-0.709	1.81E-01
65	ENSG00000100403	ZC3H7B	yes	-0.708	1.81E-01
66	ENSG00000241945	PWP2	yes	0.697	1.96E-01
67	ENSG00000067248	DHX29	yes	1.203	1.97E-01
68	ENSG00000056097	ZFR	yes	0.693	1.97E-01
70	ENSG00000088247	KHSRP	yes	-0.682	2.07E-01
71	ENSG00000156697	UTP14A	yes	0.685	2.07E-01
69	ENSG00000060138	YBX3	yes	-0.684	2.07E-01
73	ENSG00000125844	RRBP1	yes	-0.676	2.09E-01
72	ENSG00000108312	UBTF	yes	-1.172	2.09E-01
74	ENSG00000183520	UTP11L	yes	0.679	2.09E-01
75	ENSG00000111364	DDX55	yes	0.824	2.12E-01
76	ENSG00000182810	DDX28	yes	0.818	2.14E-01
77	ENSG00000160818	GPATCH4	yes	0.819	2.14E-01
78	ENSG00000196943	NOP9	yes	0.665	2.17E-01
79	ENSG00000135486	HNRNPA1	yes	-0.659	2.19E-01
81	ENSG00000129250	KIF1C	yes	-0.658	2.19E-01
80	ENSG00000188976	NOC2L	yes	-0.658	2.19E-01
82	ENSG00000125482	TTF1	no	1.145	2.19E-01
84	ENSG00000117523	PRRC2C	yes	-0.651	2.25E-01
83	ENSG00000121060	TRIM25	yes	-0.653	2.25E-01
85	ENSG00000197451	HNRNPAB	yes	-0.645	2.36E-01
86	ENSG00000065978	YBX1	yes	-0.639	2.44E-01
87	ENSG00000154277	UCHL1	no	-0.629	2.60E-01
89	ENSG00000164024	METAP1	no	-0.618	2.81E-01
88	ENSG00000181019	NQO1	yes	-0.754	2.81E-01
90	ENSG00000174231	PRPF8	yes	-0.615	2.81E-01
91	ENSG00000257103	LSM14A	yes	-0.609	2.91E-01
92	ENSG00000244509	APOBEC3C	yes	-0.744	2.91E-01
93	ENSG00000083223	ZCCHC6	yes	-1.039	3.05E-01
95	ENSG00000149925	ALDOA	yes	-0.714	3.27E-01

99	ENSG00000104131	EIF3J	yes	0.584	3.27E-01
98	ENSG00000189403	HMGB1	yes	-0.582	3.27E-01
97	ENSG00000128534	LSM8	yes	-0.583	3.27E-01
96	ENSG00000148773	MKI67	yes	-0.584	3.27E-01
94	ENSG00000167004	PDIA3	yes	-1.006	3.27E-01
100	ENSG00000174748	RPL15	yes	0.586	3.27E-01
102	ENSG00000096746	HNRNPH3	yes	-0.576	3.34E-01
101	ENSG00000124383	MPHOSPH10	yes	-0.576	3.34E-01
103	ENSG00000147905	ZCCHC7	yes	-0.572	3.40E-01
104	ENSG00000147140	NONO	yes	-0.569	3.44E-01
105	ENSG00000115368	WDR75	yes	0.564	3.53E-01
106	ENSG00000130713	EXOSC2	yes	-0.685	3.61E-01
107	ENSG00000139726	DENR	no	-0.966	3.63E-01
111	ENSG00000185567	AHNAK2	no	-0.551	3.68E-01
108	ENSG00000110367	DDX6	yes	-0.553	3.68E-01
109	ENSG00000198000	NOL8	yes	-0.553	3.68E-01
110	ENSG00000116560	SFPQ	yes	-0.551	3.68E-01
112	ENSG00000114416	FXR1	yes	-0.548	3.72E-01
113	ENSG00000138757	G3BP2	yes	-0.546	3.74E-01
114	ENSG00000125944	HNRNPR	yes	-0.545	3.74E-01
115	ENSG00000196262	PPIA	yes	-0.542	3.78E-01
116	ENSG00000086189	DIMT1	yes	-0.923	3.96E-01
117	ENSG00000179041	RRS1	yes	0.533	3.96E-01
119	ENSG00000169564	PCBP1	yes	-0.527	4.06E-01
118	ENSG00000161547	SRSF2	yes	-0.528	4.06E-01
120	ENSG00000160211	G6PD	no	-0.518	4.27E-01
121	ENSG00000186468	RPS23	yes	-0.517	4.29E-01
123	ENSG00000185163	DDX51	yes	0.619	4.56E-01
122	ENSG00000135372	NAT10	yes	0.506	4.56E-01
124	ENSG00000134987	WDR36	yes	0.616	4.59E-01
126	ENSG00000113742	CPEB4	yes	-0.498	4.64E-01
125	ENSG00000165119	HNRNPK	yes	-0.500	4.64E-01
127	ENSG00000142534	RPS11	yes	-0.498	4.64E-01

128	ENSG00000144895	EIF2A	no	-0.495	4.68E-01
131	ENSG00000184083	FAM120C	yes	-0.489	4.73E-01
130	ENSG00000165525	NEMF	no	-0.601	4.73E-01
132	ENSG00000055917	PUM2	yes	-0.488	4.73E-01
129	ENSG00000083312	TNPO1	yes	-0.603	4.73E-01
133	ENSG00000055483	USP36	yes	0.489	4.73E-01
134	ENSG00000198034	RPS4X	yes	-0.486	4.76E-01
138	ENSG00000132153	DHX30	yes	-0.481	4.77E-01
137	ENSG00000259956	RBM15B	no	-0.482	4.77E-01
136	ENSG00000120948	TARDBP	yes	-0.482	4.77E-01
135	ENSG00000149658	YTHDF1	yes	-0.838	4.77E-01
139	ENSG00000162775	RBM15	yes	-0.477	4.86E-01
140	ENSG00000139131	YARS2	yes	-0.824	4.87E-01
141	ENSG00000122870	BICC1	yes	-0.580	4.91E-01
147	ENSG00000145741	BTF3	yes	-0.463	4.94E-01
150	ENSG00000184110	EIF3C	yes	0.466	4.94E-01
144	ENSG00000127804	METTL16	yes	-0.467	4.94E-01
151	ENSG00000186416	NKRF	yes	0.471	4.94E-01
145	ENSG00000011304	PTBP1	yes	-0.466	4.94E-01
142	ENSG00000185088	RPS27L	yes	-0.573	4.94E-01
148	ENSG00000115524	SF3B1	yes	-0.463	4.94E-01
143	ENSG00000136450	SRSF1	yes	-0.468	4.94E-01
146	ENSG00000270647	TAF15	no	-0.464	4.94E-01
149	ENSG00000131747	TOP2A	yes	-0.462	4.94E-01
153	ENSG00000092199	HNRNPC	yes	-0.459	4.97E-01
152	ENSG00000086589	RBM22	yes	-0.459	4.97E-01
157	ENSG00000182718	ANXA2	yes	-0.450	5.05E-01
159	ENSG00000089048	ESF1	yes	0.453	5.05E-01
158	ENSG00000075420	FNDC3B	yes	-0.450	5.05E-01
156	ENSG00000178209	PLEC	yes	-0.452	5.05E-01
154	ENSG00000003756	RBM5	yes	-0.555	5.05E-01
155	ENSG00000122406	RPL5	yes	-0.454	5.05E-01
160	ENSG00000099622	CIRBP	yes	-0.443	5.21E-01

161	ENSG00000162613	FUBP1	yes	-0.442	5.21E-01
162	ENSG00000080824	HSP90AA1	yes	-0.442	5.21E-01
164	ENSG00000126777	KTN1	yes	-0.438	5.28E-01
163	ENSG00000197345	MRPL21	yes	-0.438	5.28E-01
165	ENSG00000120800	UTP20	yes	-0.437	5.29E-01
166	ENSG00000102317	RBM3	yes	-0.434	5.34E-01
168	ENSG00000105821	DNAJC2	yes	-0.745	5.35E-01
170	ENSG00000111640	GAPDH	no	-0.429	5.35E-01
169	ENSG00000153187	HNRNPU	yes	-0.430	5.35E-01
171	ENSG00000109971	HSPA8	yes	-0.429	5.35E-01
167	ENSG00000109814	UGDH	no	-0.746	5.35E-01
173	ENSG00000033050	ABCF2	no	-0.426	5.40E-01
176	ENSG00000112578	BYSL	yes	-0.423	5.40E-01
174	ENSG00000119616	FCF1	yes	-0.424	5.40E-01
175	ENSG00000107929	LARP4B	yes	-0.424	5.40E-01
172	ENSG00000089195	TRMT6	yes	-0.518	5.40E-01
178	ENSG00000116120	FARSB	yes	-0.514	5.45E-01
177	ENSG00000044574	HSPA5	yes	-0.726	5.45E-01
179	ENSG00000117450	PRDX1	yes	-0.418	5.47E-01
180	ENSG00000122566	HNRNPA2B1	yes	-0.412	5.61E-01
181	ENSG00000179950	PUF60	yes	-0.411	5.63E-01
182	ENSG00000125977	EIF2S2	yes	0.406	5.72E-01
183	ENSG00000160208	RRP1B	yes	0.407	5.72E-01
184	ENSG00000162664	ZNF326	yes	-0.405	5.73E-01
185	ENSG00000117133	RPF1	yes	-0.404	5.73E-01
186	ENSG00000104852	SNRNP70	yes	-0.402	5.76E-01
187	ENSG00000138668	HNRNPD	yes	-0.401	5.76E-01
188	ENSG00000055950	MRPL43	yes	-0.398	5.80E-01
189	ENSG00000135316	SYNCRIP	yes	-0.398	5.80E-01
190	ENSG00000079246	XRCC5	yes	-0.397	5.80E-01
191	ENSG00000141367	CLTC	yes	-0.394	5.84E-01
192	ENSG00000100029	PES1	yes	0.395	5.84E-01
196	ENSG00000275700	AATF	no	-0.388	5.92E-01

194	ENSG00000130813	C19orf66	no	-0.389	5.92E-01
195	ENSG00000177733	HNRNPA0	yes	-0.388	5.92E-01
193	ENSG00000119705	SLIRP	yes	-0.389	5.92E-01
197	ENSG00000070814	TCOF1	yes	0.391	5.92E-01
198	ENSG00000112651	MRPL2	yes	-0.385	5.98E-01
204	ENSG00000164934	DCAF13	yes	0.381	5.99E-01
200	ENSG00000150990	DHX37	no	-0.659	5.99E-01
205	ENSG00000165271	NOL6	yes	0.382	5.99E-01
199	ENSG00000170473	PYM1	yes	-0.663	5.99E-01
203	ENSG00000105193	RPS16	yes	0.380	5.99E-01
202	ENSG00000123352	SPATS2	yes	-0.378	5.99E-01
201	ENSG00000137073	UBAP2	yes	-0.379	5.99E-01
220	ENSG00000261236	BOP1	no	0.371	6.05E-01
217	ENSG00000133773	CCDC59	yes	0.362	6.05E-01
213	ENSG00000215301	DDX3X	yes	-0.368	6.05E-01
212	ENSG00000169813	HNRNPF	yes	-0.368	6.05E-01
211	ENSG00000143889	HNRNPLL	yes	-0.369	6.05E-01
207	ENSG00000130520	LSM4	yes	-0.446	6.05E-01
218	ENSG00000048162	NOP16	yes	0.362	6.05E-01
206	ENSG00000059378	PARP12	yes	-0.647	6.05E-01
216	ENSG00000121390	PSPC1	yes	-0.363	6.05E-01
208	ENSG00000125970	RALY	yes	-0.373	6.05E-01
219	ENSG00000197498	RPF2	yes	0.364	6.05E-01
221	ENSG00000187051	RPS19BP1	yes	0.374	6.05E-01
210	ENSG00000106803	SEC61B	yes	-0.370	6.05E-01
222	ENSG00000113387	SUB1	yes	0.376	6.05E-01
209	ENSG00000169871	TRIM56	yes	-0.371	6.05E-01
215	ENSG00000063244	U2AF2	yes	-0.363	6.05E-01
214	ENSG00000142207	URB1	yes	-0.364	6.05E-01
223	ENSG00000161847	RAVER1	yes	-0.442	6.07E-01
225	ENSG00000130589	HELZ2	yes	0.441	6.07E-01
224	ENSG00000121774	KHDRBS1	yes	-0.359	6.07E-01
226	ENSG00000198918	RPL39	no	-0.356	6.12E-01

227	ENSG00000164548	TRA2A	yes	0.437	6.12E-01
228	ENSG00000167526	RPL13	yes	0.355	6.12E-01
229	ENSG00000123908	AGO2	yes	-0.354	6.13E-01
230	ENSG00000104824	HNRNPL	yes	-0.353	6.15E-01
231	ENSG00000130305	NSUN5	yes	0.350	6.22E-01
232	ENSG00000074800	ENO1	yes	-0.603	6.24E-01
233	ENSG00000135046	ANXA1	no	-0.598	6.27E-01
235	ENSG00000132463	GRSF1	yes	-0.345	6.27E-01
234	ENSG00000015479	MATR3	yes	-0.346	6.27E-01
236	ENSG00000129351	ILF3	yes	0.344	6.27E-01
237	ENSG00000153250	RBMS1	yes	-0.342	6.31E-01
238	ENSG00000198917	C9orf114	yes	0.591	6.32E-01
240	ENSG00000196419	XRCC6	yes	-0.339	6.35E-01
239	ENSG00000058673	ZC3H11A	yes	-0.587	6.35E-01
241	ENSG00000111615	KRR1	yes	0.335	6.44E-01
242	ENSG00000170144	HNRNPA3	yes	-0.333	6.50E-01
244	ENSG00000130956	HABP4	yes	0.328	6.54E-01
246	ENSG00000111142	METAP2	yes	0.330	6.54E-01
243	ENSG00000067225	PKM	yes	-0.330	6.54E-01
245	ENSG00000146223	RPL7L1	yes	0.329	6.54E-01
247	ENSG00000198074	AKR1B10	no	-0.325	6.65E-01
252	ENSG00000067048	DDX3Y	no	-0.316	6.68E-01
253	ENSG00000204120	GIGYF2	yes	-0.315	6.68E-01
255	ENSG00000134697	GNL2	yes	0.317	6.68E-01
254	ENSG00000107937	GTPBP4	yes	0.316	6.68E-01
249	ENSG00000170860	LSM3	yes	-0.320	6.68E-01
250	ENSG00000104177	MYEF2	yes	-0.319	6.68E-01
257	ENSG00000173145	NOC3L	yes	0.321	6.68E-01
248	ENSG00000123131	PRDX4	no	-0.386	6.68E-01
256	ENSG00000156482	RPL30	yes	0.319	6.68E-01
258	ENSG00000124486	USP9X	no	0.546	6.68E-01
251	ENSG00000144161	ZC3H8	yes	-0.318	6.68E-01
259	ENSG00000076067	RBMS2	yes	-0.314	6.69E-01

260	ENSG00000164338	UTP15	yes	0.313	6.71E-01
261	ENSG00000159217	IGF2BP1	yes	-0.310	6.75E-01
262	ENSG00000166889	PATL1	yes	0.310	6.75E-01
263	ENSG00000147649	MTDH	yes	-0.308	6.79E-01
265	ENSG00000124942	AHNAK	yes	-0.305	6.81E-01
266	ENSG00000171824	EXOSC10	yes	0.306	6.81E-01
264	ENSG00000140988	RPS2	yes	-0.305	6.81E-01
267	ENSG00000108592	FTSJ3	yes	0.303	6.85E-01
268	ENSG00000116221	MRPL37	yes	-0.302	6.87E-01
269	ENSG00000131711	MAP1B	no	-0.368	6.90E-01
271	ENSG00000165733	BMS1	yes	-0.297	6.95E-01
270	ENSG00000138326	RPS24	yes	-0.298	6.95E-01
273	ENSG00000111641	NOP2	yes	0.295	6.96E-01
272	ENSG00000197958	RPL12	yes	-0.296	6.96E-01
278	ENSG00000182944	EWSR1	yes	-0.290	6.98E-01
274	ENSG00000196924	FLNA	yes	-0.293	6.98E-01
279	ENSG00000089154	GCN1	yes	0.507	6.98E-01
276	ENSG00000198015	MRPL42	yes	-0.291	6.98E-01
277	ENSG00000117569	PTBP2	yes	-0.290	6.98E-01
275	ENSG00000196141	SPATS2L	yes	-0.292	6.98E-01
280	ENSG00000135972	MRPS9	yes	-0.284	7.16E-01
281	ENSG00000142937	RPS8	yes	0.282	7.22E-01
285	ENSG00000156508	EEF1A1	yes	-0.279	7.22E-01
287	ENSG00000105373	GLTSCR2	yes	0.279	7.22E-01
284	ENSG00000073792	IGF2BP2	yes	-0.279	7.22E-01
282	ENSG00000111716	LDHB	no	-0.340	7.22E-01
283	ENSG00000099821	POLRMT	yes	-0.280	7.22E-01
286	ENSG00000114391	RPL24	yes	-0.278	7.22E-01
288	ENSG00000159176	CSRP1	yes	-0.334	7.33E-01
289	ENSG00000132275	RRP8	yes	0.273	7.33E-01
292	ENSG00000136231	IGF2BP3	yes	-0.268	7.36E-01
291	ENSG00000169288	MRPL1	yes	-0.269	7.36E-01
294	ENSG00000204469	PRRC2A	yes	-0.267	7.36E-01

290	ENSG00000131018	SYNE1	yes	-0.270	7.36E-01
295	ENSG00000198113	TOR4A	no	0.328	7.36E-01
293	ENSG00000144481	TRPM8	no	-0.268	7.36E-01
296	ENSG00000065548	ZC3H15	yes	-0.267	7.37E-01
301	ENSG00000114867	EIF4G1	yes	0.261	7.48E-01
303	ENSG00000125691	RPL23	yes	-0.258	7.48E-01
297	ENSG00000174444	RPL4	yes	-0.261	7.48E-01
304	ENSG00000149273	RPS3	yes	-0.258	7.48E-01
298	ENSG00000100650	SRSF5	yes	-0.260	7.48E-01
299	ENSG00000124193	SRSF6	yes	-0.260	7.48E-01
300	ENSG00000132478	UNK	yes	-0.259	7.48E-01
302	ENSG00000198492	YTHDF2	yes	0.262	7.48E-01
305	ENSG00000129245	FXR2	yes	0.255	7.55E-01
306	ENSG00000171863	RPS7	yes	-0.252	7.62E-01
308	ENSG00000126945	HNRNPH2	yes	-0.249	7.67E-01
307	ENSG00000107951	MTPAP	yes	-0.305	7.67E-01
309	ENSG00000099783	HNRNPM	yes	-0.248	7.69E-01
312	ENSG00000187772	LIN28B	yes	0.245	7.71E-01
311	ENSG00000122965	RBM19	yes	0.245	7.71E-01
310	ENSG00000131732	ZCCHC9	yes	-0.245	7.71E-01
315	ENSG00000007080	CCDC124	yes	-0.289	7.72E-01
321	ENSG00000009307	CSDE1	yes	-0.236	7.72E-01
342	ENSG00000123064	DDX54	yes	0.239	7.72E-01
319	ENSG00000055332	EIF2AK2	yes	-0.238	7.72E-01
343	ENSG00000119285	HEATR1	yes	0.291	7.72E-01
333	ENSG00000158406	HIST1H4H	yes	-0.225	7.72E-01
318	ENSG00000169045	HNRNPH1	yes	-0.241	7.72E-01
317	ENSG00000177971	IMP3	yes	-0.243	7.72E-01
313	ENSG00000002834	LASP1	no	-0.404	7.72E-01
330	ENSG00000172172	MRPL13	yes	-0.227	7.72E-01
338	ENSG00000146909	NOM1	yes	0.223	7.72E-01
329	ENSG00000037474	NSUN2	yes	-0.227	7.72E-01
334	ENSG00000108256	NUFIP2	yes	-0.224	7.72E-01

324	ENSG00000100836	PABPN1	yes	-0.230	7.72E-01
331	ENSG00000253729	PRKDC	yes	-0.225	7.72E-01
320	ENSG00000131051	RBM39	yes	-0.236	7.72E-01
335	ENSG00000173933	RBM4	yes	-0.224	7.72E-01
314	ENSG00000126254	RBM42	yes	-0.291	7.72E-01
341	ENSG00000148300	REXO4	yes	0.232	7.72E-01
339	ENSG00000122026	RPL21	yes	0.224	7.72E-01
328	ENSG00000089009	RPL6	yes	-0.228	7.72E-01
340	ENSG00000096150	RPS18	yes	0.227	7.72E-01
332	ENSG00000197728	RPS26	yes	-0.225	7.72E-01
326	ENSG00000168028	RPSA	yes	-0.229	7.72E-01
336	ENSG00000160214	RRP1	yes	-0.224	7.72E-01
337	ENSG00000167088	SNRPD1	yes	-0.224	7.72E-01
327	ENSG00000115875	SRSF7	yes	-0.229	7.72E-01
322	ENSG00000124214	STAU1	yes	-0.232	7.72E-01
325	ENSG00000136463	TACO1	yes	-0.229	7.72E-01
316	ENSG00000106638	TBL2	yes	-0.282	7.72E-01
323	ENSG00000136810	TXN	yes	-0.232	7.72E-01
344	ENSG00000182004	SNRPE	no	-0.222	7.73E-01
345	ENSG00000108829	LRRC59	yes	-0.221	7.73E-01
346	ENSG00000100201	DDX17	yes	-0.217	7.86E-01
348	ENSG00000164164	OTUD4	yes	0.264	7.86E-01
347	ENSG00000185728	YTHDF3	yes	-0.216	7.86E-01
349	ENSG00000134333	LDHA	yes	-0.215	7.86E-01
350	ENSG00000183311	TUBB	no	-0.259	7.94E-01
351	ENSG00000011260	UTP18	yes	-0.211	7.94E-01
352	ENSG00000110321	EIF4G2	yes	0.208	8.04E-01
355	ENSG00000146540	C7orf50	yes	0.206	8.06E-01
354	ENSG00000135829	DHX9	yes	0.205	8.06E-01
353	ENSG00000115268	RPS15	yes	-0.205	8.06E-01
356	ENSG00000130811	EIF3G	yes	0.204	8.06E-01
357	ENSG00000247315	ZCCHC3	no	-0.203	8.07E-01
359	ENSG00000158417	EIF5B	yes	0.201	8.12E-01

358	ENSG00000189306	RRP7A	yes	-0.200	8.12E-01
360	ENSG00000204574	ABCF1	yes	-0.198	8.16E-01
361	ENSG00000117395	EBNA1BP2	yes	0.198	8.16E-01
364	ENSG00000115816	CEBPZ	yes	0.195	8.17E-01
362	ENSG00000124177	CHD6	no	-0.241	8.17E-01
363	ENSG00000161016	RPL8	yes	-0.195	8.17E-01
365	ENSG00000161960	EIF4A1	yes	-0.193	8.21E-01
366	ENSG00000147403	RPL10	yes	0.192	8.21E-01
367	ENSG00000110700	RPS13	yes	0.193	8.21E-01
370	ENSG00000213782	DDX47	yes	0.233	8.21E-01
368	ENSG00000173141	MRPL57	no	-0.190	8.21E-01
369	ENSG00000105939	ZC3HAV1	yes	-0.190	8.21E-01
371	ENSG00000242265	PEG10	yes	-0.188	8.27E-01
372	ENSG00000111669	TPI1	no	-0.320	8.34E-01
373	ENSG00000107581	EIF3A	yes	0.183	8.35E-01
374	ENSG00000165209	STRBP	yes	0.225	8.35E-01
376	ENSG00000161813	LARP4	yes	-0.180	8.42E-01
377	ENSG00000179820	MYADM	no	0.311	8.42E-01
375	ENSG00000087365	SF3B2	yes	-0.311	8.42E-01
378	ENSG00000145912	NHP2	yes	0.178	8.43E-01
382	ENSG00000079785	DDX1	yes	-0.172	8.48E-01
383	ENSG00000089737	DDX24	yes	0.172	8.48E-01
384	ENSG00000277936	MRPL45	no	0.173	8.48E-01
381	ENSG00000101361	NOP56	yes	-0.172	8.48E-01
379	ENSG00000132819	RBM38	yes	-0.302	8.48E-01
380	ENSG00000100316	RPL3	yes	-0.173	8.48E-01
385	ENSG00000083544	TDRD3	yes	0.212	8.48E-01
405	ENSG00000146109	ABT1	yes	0.167	8.54E-01
397	ENSG00000088205	DDX18	yes	0.156	8.54E-01
389	ENSG00000124228	DDX27	yes	-0.164	8.54E-01
386	ENSG00000125868	DSTN	no	-0.271	8.54E-01
403	ENSG00000167658	EEF2	yes	0.163	8.54E-01
399	ENSG00000100353	EIF3D	yes	0.158	8.54E-01

394	ENSG00000204628	GNB2L1	yes	-0.158	8.54E-01
391	ENSG00000114686	MRPL3	yes	-0.161	8.54E-01
395	ENSG00000153944	MSI2	yes	-0.157	8.54E-01
400	ENSG00000055044	NOP58	yes	0.160	8.54E-01
396	ENSG00000164346	NSA2	yes	-0.157	8.54E-01
398	ENSG00000188739	RBM34	yes	0.158	8.54E-01
401	ENSG00000198755	RPL10A	yes	0.160	8.54E-01
390	ENSG00000162244	RPL29	yes	-0.163	8.54E-01
392	ENSG00000124614	RPS10	yes	-0.160	8.54E-01
388	ENSG00000008988	RPS20	yes	-0.165	8.54E-01
393	ENSG00000083845	RPS5	yes	-0.158	8.54E-01
404	ENSG00000171490	RSL1D1	yes	0.164	8.54E-01
387	ENSG00000111786	SRSF9	yes	-0.168	8.54E-01
402	ENSG00000100038	TOP3B	yes	0.161	8.54E-01
406	ENSG00000026025	VIM	no	0.288	8.54E-01
416	ENSG00000128394	APOBEC3F	yes	0.147	8.58E-01
418	ENSG00000113460	BRIX1	yes	0.150	8.58E-01
420	ENSG00000117877	CD3EAP	yes	0.263	8.58E-01
410	ENSG00000128886	ELL3	no	-0.151	8.58E-01
414	ENSG00000163938	GNL3	yes	0.146	8.58E-01
412	ENSG00000104356	POP1	yes	-0.146	8.58E-01
407	ENSG00000132300	PTCD3	yes	-0.153	8.58E-01
419	ENSG00000239306	RBM14	yes	0.150	8.58E-01
411	ENSG00000147274	RBMX	yes	-0.149	8.58E-01
408	ENSG00000131469	RPL27	yes	-0.152	8.58E-01
417	ENSG00000148303	RPL7A	yes	0.149	8.58E-01
409	ENSG00000112081	SRSF3	yes	-0.152	8.58E-01
415	ENSG00000148296	SURF6	yes	0.147	8.58E-01
413	ENSG00000054118	THRAP3	yes	-0.146	8.58E-01
421	ENSG00000138095	LRPPRC	yes	-0.142	8.67E-01
425	ENSG00000106682	EIF4H	yes	-0.138	8.69E-01
424	ENSG00000171861	RNMTL1	yes	-0.138	8.69E-01
426	ENSG00000063177	RPL18	yes	0.138	8.69E-01

423	ENSG00000160633	SAFB	yes	-0.140	8.69E-01
422	ENSG00000136527	TRA2B	yes	-0.140	8.69E-01
427	ENSG00000204221	WDR46	no	0.139	8.69E-01
429	ENSG00000136448	NMT1	yes	0.137	8.69E-01
428	ENSG00000067167	TRAM1	no	-0.237	8.69E-01
431	ENSG00000182473	EXOC7	no	-0.130	8.84E-01
430	ENSG00000116251	RPL22	yes	-0.131	8.84E-01
432	ENSG00000100028	SNRPD3	yes	-0.128	8.89E-01
433	ENSG00000274523	WBSCR16	no	0.157	8.89E-01
435	ENSG00000277594	DDX52	no	0.120	9.04E-01
436	ENSG00000136271	DDX56	yes	0.120	9.04E-01
434	ENSG00000155506	LARP1	yes	-0.121	9.04E-01
437	ENSG00000119596	YLPM1	yes	0.121	9.04E-01
438	ENSG00000141076	CIRH1A	yes	0.119	9.04E-01
441	ENSG00000145907	G3BP1	yes	-0.117	9.05E-01
440	ENSG00000163584	RPL22L1	no	-0.118	9.05E-01
439	ENSG00000125835	SNRPB	yes	-0.118	9.05E-01
442	ENSG00000179115	FARSA	yes	-0.197	9.13E-01
443	ENSG00000130255	RPL36	yes	0.113	9.14E-01
452	ENSG00000198363	ASPH	no	0.108	9.15E-01
450	ENSG00000153113	CAST	yes	-0.105	9.15E-01
451	ENSG00000109534	GAR1	yes	-0.105	9.15E-01
455	ENSG00000146834	MEPCE	yes	0.136	9.15E-01
454	ENSG00000146676	PURB	yes	0.110	9.15E-01
453	ENSG00000142541	RPL13A	yes	0.109	9.15E-01
449	ENSG00000164587	RPS14	yes	-0.105	9.15E-01
445	ENSG00000137154	RPS6	yes	-0.109	9.15E-01
447	ENSG00000124541	RRP36	yes	-0.107	9.15E-01
446	ENSG00000198301	SDAD1	yes	-0.109	9.15E-01
444	ENSG00000099995	SF3A1	yes	-0.131	9.15E-01
448	ENSG00000197157	SND1	yes	-0.107	9.15E-01
458	ENSG00000105127	AKAP8	yes	-0.112	9.15E-01
475	ENSG00000196139	AKR1C3	no	0.164	9.15E-01

470	ENSG00000029363	BCLAF1	yes	-0.093	9.15E-01
462	ENSG00000110104	CCDC86	yes	-0.098	9.15E-01
466	ENSG00000108654	DDX5	yes	-0.094	9.15E-01
463	ENSG00000119812	FAM98A	yes	-0.098	9.15E-01
456	ENSG00000181610	MRPS23	yes	-0.166	9.15E-01
460	ENSG00000144029	MRPS5	yes	-0.101	9.15E-01
457	ENSG00000086102	NFX1	yes	-0.164	9.15E-01
472	ENSG00000087269	NOP14	yes	0.092	9.15E-01
469	ENSG00000121274	PAPD5	yes	-0.093	9.15E-01
471	ENSG00000089220	PEBP1	yes	-0.091	9.15E-01
467	ENSG00000161057	PSMC2	no	-0.094	9.15E-01
468	ENSG00000185129	PURA	yes	-0.093	9.15E-01
473	ENSG00000106344	RBM28	yes	0.095	9.15E-01
474	ENSG00000166441	RPL27A	yes	0.095	9.15E-01
459	ENSG00000109475	RPL34	no	-0.102	9.15E-01
461	ENSG00000233927	RPS28	yes	-0.100	9.15E-01
465	ENSG00000065526	SPEN	yes	-0.098	9.15E-01
476	ENSG00000158710	TAGLN2	yes	0.165	9.15E-01
464	ENSG00000075035	WSCD2	no	-0.098	9.15E-01
477	ENSG00000136942	RPL35	yes	0.089	9.21E-01
480	ENSG00000105202	FBL	yes	0.086	9.23E-01
479	ENSG00000167005	NUDT21	yes	-0.086	9.23E-01
478	ENSG00000134308	YWHAQ	no	-0.149	9.23E-01
482	ENSG00000165732	DDX21	yes	0.083	9.27E-01
481	ENSG00000174173	TRMT10C	yes	-0.144	9.27E-01
486	ENSG00000136026	CKAP4	yes	-0.080	9.27E-01
484	ENSG00000102081	FMR1	yes	-0.097	9.27E-01
485	ENSG00000182154	MRPL41	yes	-0.082	9.27E-01
483	ENSG00000181991	MRPS11	yes	-0.136	9.27E-01
487	ENSG00000071082	RPL31	yes	-0.080	9.27E-01
488	ENSG00000182774	RPS17	yes	-0.079	9.27E-01
489	ENSG00000177954	RPS27	yes	0.079	9.27E-01
490	ENSG00000185298	CCDC137	yes	-0.078	9.29E-01

491	ENSG00000063046	EIF4B	yes	0.076	9.31E-01
493	ENSG00000174953	DHX36	yes	-0.074	9.33E-01
494	ENSG00000067334	DNTTIP2	yes	-0.072	9.33E-01
492	ENSG00000166197	NOLC1	yes	-0.075	9.33E-01
496	ENSG00000130669	PAK4	no	-0.072	9.33E-01
498	ENSG00000180228	PRKRA	yes	0.071	9.33E-01
497	ENSG00000188846	RPL14	yes	-0.071	9.33E-01
495	ENSG00000143569	UBAP2L	yes	-0.072	9.33E-01
500	ENSG00000147604	RPL7	yes	-0.068	9.37E-01
499	ENSG00000132467	UTP3	yes	-0.069	9.37E-01
501	ENSG00000139233	LLPH	yes	0.067	9.38E-01
502	ENSG00000196235	SUPT5H	yes	0.115	9.38E-01
503	ENSG00000142657	PGD	no	-0.112	9.40E-01
504	ENSG00000174780	SRP72	yes	-0.065	9.40E-01
505	ENSG00000178105	DDX10	yes	-0.064	9.40E-01
509	ENSG00000118246	FASTKD2	yes	-0.060	9.42E-01
506	ENSG00000108826	MRPL27	yes	-0.102	9.42E-01
508	ENSG00000156471	PTDSS1	yes	-0.061	9.42E-01
510	ENSG00000170889	RPS9	yes	0.059	9.42E-01
512	ENSG00000156990	RPUSD3	yes	0.108	9.42E-01
507	ENSG00000137776	SLTM	yes	-0.073	9.42E-01
511	ENSG00000125743	SNRPD2	yes	0.062	9.42E-01
516	ENSG00000129347	KRI1	yes	0.069	9.42E-01
515	ENSG00000115761	NOL10	yes	0.056	9.42E-01
513	ENSG00000132341	RAN	yes	-0.057	9.42E-01
514	ENSG00000198242	RPL23A	yes	-0.056	9.42E-01
527	ENSG00000165092	ALDH1A1	no	0.052	9.48E-01
518	ENSG00000188573	FBLL1	no	-0.050	9.48E-01
529	ENSG00000143321	HDGF	yes	0.079	9.48E-01
520	ENSG00000155363	MOV10	yes	-0.048	9.48E-01
523	ENSG00000132382	MYBBP1A	yes	0.047	9.48E-01
517	ENSG00000129460	NGDN	yes	-0.065	9.48E-01
524	ENSG00000148843	PDCD11	yes	0.048	9.48E-01

528	ENSG00000173914	RBM4B	yes	0.052	9.48E-01
522	ENSG00000142676	RPL11	yes	0.046	9.48E-01
519	ENSG00000108107	RPL28	yes	-0.050	9.48E-01
525	ENSG00000138385	SSB	yes	0.050	9.48E-01
526	ENSG00000121486	TRMT1L	yes	0.050	9.48E-01
521	ENSG00000122299	ZC3H7A	yes	-0.046	9.48E-01
530	ENSG00000143947	RPS27A	yes	0.044	9.50E-01
531	ENSG00000174720	LARP7	yes	0.043	9.52E-01
532	ENSG00000085662	AKR1B1	no	-0.051	9.53E-01
534	ENSG00000168488	ATXN2L	yes	-0.039	9.56E-01
533	ENSG00000125485	DDX31	yes	-0.039	9.56E-01
536	ENSG00000107625	DDX50	yes	0.038	9.58E-01
535	ENSG00000143977	SNRPG	yes	-0.038	9.58E-01
537	ENSG00000105323	HNRNPUL1	yes	-0.036	9.61E-01
538	ENSG00000136270	TBRG4	yes	-0.035	9.61E-01
539	ENSG00000169067	ACTBL2	no	-0.038	9.64E-01
542	ENSG00000135387	CAPRIN1	yes	0.031	9.64E-01
541	ENSG00000158941	CCAR2	yes	0.031	9.64E-01
543	ENSG00000105364	MRPL4	yes	0.032	9.64E-01
540	ENSG00000103194	USP10	yes	-0.032	9.64E-01
544	ENSG00000144713	RPL32	yes	-0.028	9.69E-01
545	ENSG00000163214	DHX57	yes	-0.025	9.74E-01
547	ENSG00000096080	MRPS18A	no	-0.024	9.74E-01
546	ENSG00000130723	PRRC2B	yes	-0.024	9.74E-01
548	ENSG00000108298	RPL19	yes	0.025	9.74E-01
549	ENSG00000138231	DBR1	yes	-0.038	9.78E-01
550	ENSG00000265681	RPL17	yes	0.020	9.81E-01
555	ENSG00000109606	DHX15	yes	0.017	9.81E-01
552	ENSG00000089280	FUS	yes	-0.017	9.81E-01
554	ENSG00000143621	ILF2	yes	0.017	9.81E-01
551	ENSG00000065183	WDR3	yes	-0.018	9.81E-01
553	ENSG00000088930	XRN2	yes	-0.017	9.81E-01
557	ENSG00000130826	DKC1	yes	-0.010	9.87E-01

563	ENSG00000115677	HDLBP	yes	0.013	9.87E-01
560	ENSG00000115053	NCL	yes	-0.009	9.87E-01
556	ENSG00000181163	NPM1	yes	-0.010	9.87E-01
559	ENSG00000080608	PUM3	yes	-0.010	9.87E-01
562	ENSG00000112183	RBM24	yes	0.011	9.87E-01
558	ENSG00000145425	RPS3A	yes	-0.010	9.87E-01
561	ENSG00000106028	SSBP1	yes	0.010	9.87E-01
565	ENSG00000169714	CNBP	yes	-0.007	9.88E-01
564	ENSG00000114767	RRP9	yes	-0.010	9.88E-01
566	ENSG00000052749	RRP12	yes	0.006	9.89E-01
567	ENSG00000105640	RPL18A	yes	0.005	9.90E-01
568	ENSG00000120289	MAGEB4	no	-0.001	9.99E-01
569	ENSG00000170515	PA2G4	yes	-0.000	9.99E-01
616	ENSG00000162688	AGL	no	NA	NA
592	ENSG00000121057	AKAP1	yes	NA	NA
583	ENSG00000101974	ATP11C	no	NA	NA
613	ENSG00000160862	AZGP1	no	NA	NA
589	ENSG00000108578	BLMH	no	NA	NA
620	ENSG00000167644	C19orf33	no	NA	NA
577	ENSG00000088543	C3orf18	no	NA	NA
629	ENSG00000178363	CALML3	no	NA	NA
630	ENSG00000178372	CALML5	no	NA	NA
586	ENSG00000105141	CASP14	no	NA	NA
594	ENSG00000121691	CAT	no	NA	NA
604	ENSG00000137197	CDSN	no	NA	NA
585	ENSG00000103187	COTL1	no	NA	NA
593	ENSG00000121552	CSTA	no	NA	NA
615	ENSG00000161634	DCD	yes	NA	NA
639	ENSG00000198947	DMD	no	NA	NA
602	ENSG00000134765	DSC1	no	NA	NA
601	ENSG00000134760	DSG1	no	NA	NA
580	ENSG00000096696	DSP	yes	NA	NA
608	ENSG00000143924	EML4	no	NA	NA

618	ENSG00000164687	FABP5	no	NA	NA
624	ENSG00000171262	FAM98B	yes	NA	NA
603	ENSG00000136068	FLNB	yes	NA	NA
591	ENSG00000117228	GBP1	no	NA	NA
576	ENSG00000084207	GSTP1	yes	NA	NA
597	ENSG00000125812	GZF1	no	NA	NA
642	ENSG00000206053	HN1L	no	NA	NA
579	ENSG00000096384	HSP90AB1	yes	NA	NA
599	ENSG00000131724	IL13RA1	no	NA	NA
646	Inf_P03454_HEMA	Inf_HA	no	NA	NA
647	Inf_Q1K9H1_NA	Inf_NA	no	NA	NA
648	Inf_Q20MG9_PB2	Inf_PB2	no	NA	NA
649	Inf_Q20MH1_PB1	Inf_PB1	no	NA	NA
650	Inf_Q20MH2_PA	Inf_PA	no	NA	NA
651	Inf_Q67183_M1	Inf_M1	no	NA	NA
584	ENSG00000102313	ITIH6	no	NA	NA
626	ENSG00000173801	JUP	no	NA	NA
635	ENSG00000187961	KLHL17	no	NA	NA
640	ENSG00000203786	KPRP	no	NA	NA
625	ENSG00000172867	KRT2	no	NA	NA
631	ENSG00000178934	LGALS7B	no	NA	NA
578	ENSG00000090382	LYZ	no	NA	NA
638	ENSG00000196531	NACA	no	NA	NA
643	ENSG00000263956	NBPF11	no	NA	NA
633	ENSG00000185624	P4HB	yes	NA	NA
607	ENSG00000143799	PARP1	yes	NA	NA
570	ENSG00000040633	PHF23	no	NA	NA
575	ENSG00000081277	PKP1	no	NA	NA
595	ENSG00000124429	POF1B	no	NA	NA
588	ENSG00000105618	PRPF31	yes	NA	NA
645	ENSG00000275903	PSMB3	no	NA	NA
628	ENSG00000177548	RABEP2	no	NA	NA
582	ENSG00000100320	RBFOX2	yes	NA	NA

571	ENSG00000056586	RC3H2	yes	NA	NA
641	ENSG00000205937	RNPS1	yes	NA	NA
644	ENSG00000271793	RP11-321N4.5	no	NA	NA
605	ENSG00000137818	RPLP1	no	NA	NA
600	ENSG00000134419	RPS15A	yes	NA	NA
587	ENSG00000105372	RPS19	yes	NA	NA
581	ENSG00000100220	RTCB	yes	NA	NA
636	ENSG00000188011	RTP5	no	NA	NA
637	ENSG00000189334	S100A14	no	NA	NA
606	ENSG00000143546	S100A8	no	NA	NA
617	ENSG00000163220	S100A9	no	NA	NA
598	ENSG00000126524	SBDS	yes	NA	NA
627	ENSG00000175793	SFN	no	NA	NA
611	ENSG00000157593	SLC35B2	no	NA	NA
623	ENSG00000169474	SPRR1A	no	NA	NA
622	ENSG00000169469	SPRR1B	no	NA	NA
614	ENSG00000161011	SQSTM1	no	NA	NA
610	ENSG00000151304	SRFBP1	yes	NA	NA
621	ENSG00000168610	STAT3	no	NA	NA
572	ENSG00000064607	SUGP2	yes	NA	NA
590	ENSG00000109111	SUPT6H	yes	NA	NA
573	ENSG00000065491	TBC1D22B	no	NA	NA
634	ENSG00000187735	TCEA1	no	NA	NA
596	ENSG00000125780	TGM3	no	NA	NA
574	ENSG00000077097	TOP2B	no	NA	NA
632	ENSG00000180098	TRNAU1AP	yes	NA	NA
609	ENSG00000147679	UTP23	yes	NA	NA
619	ENSG00000164924	YWHAZ	yes	NA	NA
612	ENSG00000158545	ZC3H18	yes	NA	NA