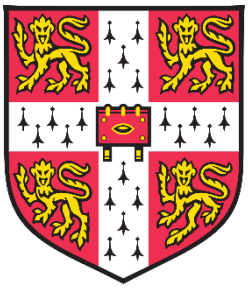
THE HCMV INTERACTOME:
A QUANTITATIVE ANALYSIS OF HUMAN CYTOMEGALOVIRUS-HOST PROTEIN INTERACTIONS



Luís Miguel Veiga Nobre

Department of Medicine

University of Cambridge

Darwin College

January 2020

This dissertation is submitted for the degree of Doctor of Philosophy

THE HCMV INTERACTOME: A QUANTITATIVE ANALYSIS OF HUMAN CYTOMEGALOVIRUS-HOST PROTEIN INTERACTIONS

Luís Miguel Veiga Nobre

Summary

Human cytomegalovirus (HCMV) is a ubiquitous pathogen that infects the majority of the adult population. Similarly to other herpesvirus, it is able to enter a state of latency, persisting within the host for life. Infection of healthy individuals is normally asymptomatic, however it is among the most common causes of allograft rejection, and can lead to several life-threatening diseases in the immunocompromised. Moreover, it is the leading viral infectious cause for congenital disease.

HCMV encodes 171 canonical genes, and a substantial number of non-canonical ORFs have been identified by ribosome profiling and proteomics. The functions of many canonical HCMV proteins remain poorly understood, and it is not yet clear how many non-canonical ORFs encode functional polypeptides.

Recent studies have provided an extensive overview on the modulation of gene expression as well as the spatio-temporal dynamics of viral and host proteins during HCMV infection. However, characterisation of specific protein-protein interactions and the exact molecular mechanisms underpinning the biological changes observed during viral infection are beyond the scope of these approaches.

Affinity-purification mass spectrometry was performed to identify binding partners for 169 canonical, and 2 non-canonical HCMV proteins in infected cells. CompPass filtering determined an extensive network of high-confidence interacting proteins, with >3,400 virus-host and >150 virus-virus protein interactions.

Domain association analysis identified protein domains co-occurring with unusual frequency, while functional enrichment analysis provided an insight into novel functions of multiple viral genes as well as how HCMV systematically modulates host environment, for example interacting with transcriptional repressive complexes or families of ubiquitin E3 ligases. Furthermore, combining interaction data with a recently published systematic analysis of HCMV-induced protein degradation identified viral interactors for 31/133 degraded host targets.

Finally, the uncharacterised, non-canonical ORFL147C protein was found to interact with elements of the mRNA splicing machinery, and a mutational study suggested its importance in viral replication. The interactome data will be important for future studies of herpesvirus infection.

Declaration

I hereby declare, that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree of qualification in this, or any other university.

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as where specified in the text and Acknowledgments.

This dissertation does not exceed the specified word limit of 60,000 words as defined by the Degree Committee, excluding figures, photographs, tables, appendices and bibliography.

Luís Miguel Veiga Nobre

January 2020

Acknowledgments

First and foremost, I would like to thank my supervisor Dr. Mike Weekes for guidance, encouragement, and continued support throughout this project. His mentorship has truly been crucial for my development as a research scientist and I am very grateful for the opportunities and teaching he has provided.

I am also grateful for the fantastic work environment provided by members of the Weekes lab, with special thanks to Dr. Katie Nightingale for helpful discussions, willingness to help with experiments and interpretation of results; and Miss Lior Soday for the R scripts that facilitated disentanglement of peptide assignment between canonical and non-canonical ORFs. I am also extremely grateful to the Wellcome Trust for funding this project.

I would like to thank Dr. Robin Antrobus, from the CIMR Proteomics Core facility, for his diligence and heedfulness in handling the multitude of mass spectrometry samples required for this project.

I am very grateful for the invaluable help provided by Dr. Edward Huttlin, from the immunoprecipitation protocol, to bioinformatics tools to filter the dataset and identify protein domain associations, as well as general advice on the project and data handling. I would like to express deep gratitude to Professor Steven Gygi and his lab, for access to the mass spectrometry data processing platform which has been fundamental to this project and many other collaborative works that I have been involved with.

I would also like to thank our several collaborators from Cardiff University: Dr. Rich Stanton, Dr. Pete Tomasec, Dr. Eddie Wang and Professor Gavin Wilkinson, for stimulating discussions and generation of viral resources; Dr. James Davies and Dr. Sepehr Seirafian for generating the recombinant adenovirus library which provided a large proportion of templates for the cloning of HCMV genes into the pHAGE-SFFV library.

My family for granting me the opportunities that have led me here, and my closest friends for the relentless support and patience they've shown in the past four years.

Finally, I would like to dedicate this work to the loving memory of Mr. Aleix Gorchs Rovira, who accompanied me throughout this journey. Aleix wrote up his PhD thesis alongside me but passed away before submitting a finalised dissertation.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

6FT-ORFs	6-Frame Translation Open Reading Frame
°C	Degrees Centigrade
µg	Microgram
µL	Microlitre
µm	Micrometre

A

aa	Amino Acid
AcN	Acetonitrile
AGC	Automatic gain control
AP-MS	Affinity-purification Mass Spectrometry

B

bp	Base Pair

C

CCR4-NOT	Carbon Catabolite Repressor 4-Negative on TATA
cm	Centimetre
CPE	Cytopathic Effect
co-IP	co-immunoprecipitation
CompPass	Comparative Proteomics Analysis Software Suite
CVC	Capsid Vertex-specific Complex

D

DAVID	Database for Annotation, Visualization and Integrated Discovery
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulphoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
DTT	Dithiothreitol
dUTPase	Deoxyuridine 5'-triphosphate nucleotidohydrolase

E

EBV	Epstein-Barr virus

F

FA	Formic Acid
FBS	Foetal Bovine Serum

G

GFP	Green Fluorescent Protein
GPCR	G protein-coupled receptor
GuHCl	Guanidine Hydrochloride

H

h	Hours

HCIP	High-confidence Interacting Proteins
HCMV	Human cytomegalovirus
HFFF	Human Foetal Foreskin Fibroblasts
HSV-1	Human simplex virus 1
HSV-2	Human simplex virus 2

I

ID	Identification
IE1	Immediate-early Protein 1
IU	International Units

K

kbp	Kilobase pair
KSHV	Kaposi's Sarcoma herpesvirus

L

LB	Luria-Bertani
LC	Liquid Chromatography
LC-MS/MS	Liquid Chromatography Tandem Mass Spectrometry
LFQ	Label-free Quantitation

M

m/z	Mass-to-charge Ratio
mg	Miligrams
min	Minutes
mL	Mililitres
mM	Milimolar
MOI	Multiplicity of infection
ms	Milisecond

N

NCoR	Nuclear receptor corepressor
NLS	Nuclear localisation signal
NuRD	Nucleosome Remodelling and Deacetylase

O

O/N	Overnight
ORF	Open Reading Frame

P

PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
ppm	Parts Per Million
PRV	Pseudorabies virus
PSM	Peptide Spectrum Matches

R

rAdv	Recombinant Adenoviral Vector
RNA	Ribonucleic Acid

RPII	RNA polymerase II
RPM	Rotations Per Minute
RT	Room Temperature

S

Sec	Seconds
SFFV	Spleen Focus-forming virus

T

TAE	Tris-acetate-EDTA
TBS	Tris Buffered Saline
TCA	Trichloroacetic acid
TE	Tris-EDTA
Th	Thompsons
TM	Transmembrane
TMT	Tandem Mass Tag

U

U	Enzyme Unit

V

V	Volt

Y

Y2H	Yeast Two-hybrid

1 | Introduction

1.1 *Herpesviridae*

The *Herpesviridae* family is comprised of viruses that can infect amniotes (mammals, birds and reptiles) [1]. This family is part of the order *Herpesvirales*, along with two other families of herpesvirus which infect mollusks (*Malacoherpesviridae*) and anamniotes (fishes and amphibians, *Alloherpesviridae*) [1]. Inclusion of viruses into the *Herpesviridae* family is based on a characteristic virion morphology, which consists of an envelope, tegument, capsid, and core containing a linear, double-stranded DNA genome [2]. Known herpesviruses share four properties: (a) their genome encodes enzymes and other factors involved in nucleic acid synthesis and metabolism, as well as proteases and protein kinases; (b) synthesis of viral DNA and capsid assembly occurs in the nucleus; (c) generation of infectious progeny is commonly followed by death of the infected cell; (d) an ability to enter a state of latency, persisting within the host for life [2, 3]. Similarities in the phylogenetic relationships among viruses and their hosts suggest that herpesviruses have co-evolved with their hosts over long periods of time and are well adapted to them [4, 5]. Thus, herpesvirus species are generally named after a taxon of the host which harbours the virus [2].

Herpesviruses have been divided into three subfamilies, the *Alpha-*, *Beta-*, and *Gamma-herpesvirinae* (see Table 1.1.1). This classification was based on biological (range of hosts, duration of the life cycle, cytopathology and latency characteristics) and genomic criteria (conservation of genes and gene clusters, arrangement of gene clusters in relation to one another, arrangement of terminal sequences involved in viral genome packaging and distribution of nucleotides subject to methylation) [2, 3].

Table 1.1.1 | *Herpesviridae* subfamily characteristics

Table removed for copyright reasons. Copyright holder is Springer.

(Adapted from Roizman, 1982) [3, 6-8]

Another characteristic of herpesvirus is the presence of direct or inverted repeat sequences greater than 100 bp in their genomes. Herpesvirus genomes can be divided according to their structure in six classes, A to F, as depicted in Figure 1.1.

Genomes of viruses comprising group A feature a large sequence from one terminus directly repeated at the other end. In group B, the terminal sequence is directly repeated numerous times at both termini, with a variable number of reiterations at both

ends. Group C genomes also have direct terminal reiterations, and additionally possess an internal set of direct repeats which is unrelated to the terminal set.

Group D genomes contain two unique regions, each flanked by inverted repeats. In these genomes, the short unique region (*S*) flanked by inverted repeats are able to invert in relation to the long unique sequence (*L*) and its flanking repeats, thus generating two equimolar genome types differing in the orientation of the short unique sequence relative to the fixed long component (see Figure 1.1, further detailed in **1.2.2**). Group E genomes are similar to group D, but contain a sequence that is repeated directly at the genome termini (*a*) and inversely repeated (*a'*) at the IR_L-IR_S junction making these genomes terminally redundant. Genomes from this group can be found in four equimolar genome types, as both the short unique region (*S*) and long unique sequence (*L*) invert their orientation. Genomes of F group lack direct or inverted repeats [2, 9].

Figure removed for copyright reasons. Copyright holder is Cambridge University Press.

Figure 1.1 | Classes of herpesvirus genome structures
Long (L) and short (S) unique sequences are shown as horizontal lines. Repeat regions (including the terminal repeat sequences TR_L/TR_S, internal repeat sequences IR_L/IR_S and terminal redundancy *a* and its inverted copy *a'*) are shown as rectangles, with arrows depicting their orientation. Adapted from Davison, 2007: 'Comparative analysis of genomes' [9].

1.1.1 Human Herpesvirus

There are currently nine herpesviruses distributed among the three subfamilies of *Herpesviridae* which are known to infect humans as their primary host (see Table 1.1.2). The two human Simplexviruses, Herpes simplex virus 1 and Herpes simplex virus 2, have approximately 70 % genomic homology and the most common clinical manifestations are mucocutaneous infections including gingivostomatitis, herpes genitalis, herpetic keratitis, and dermal whitlows. For the other human alphaherpesvirus, Varicella-zoster virus (VZV), primary infection causes chickenpox, while reactivation of latent virus causes shingles [10].

Table 1.1.2 | Human herpesvirus

Subfamily	Genus	Virus
Alphaherpesvirinae	*Simplexvirus*	(HSV-1/HHV-1)
		(HSV-2/HHV-2)
	Varicellovirus	Varicella-zoster virus (VZV/HHV-3)
Betaherpesvirinae	*Cytomegalovirus*	Human cytomegalovirus (HCMV/HHV-5)
	Roseolovirus	Human Betaherpesvirus 6A (HHV-6A)
		Human Betaherpesvirus 6B (HHV-6B)
		Human Betaherpesvirus 7 (HHV-7)
Gammaherpesvirinae	*Lymphocryptovirus*	Epstein-Barr virus (EBV/HHV-4)
	Rhadinovirus	Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8)

The two gammaherpesviruses, Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV), are able to induce oncogenesis in the host, and are predominantly associated with lymphoproliferative diseases and cancers such as lymphomas and sarcomas [10].

The betaherpesviruses of the Roseolavirus genus, Human Betaherpesvirus 6A (HHV-6A), Human Betaherpesvirus 6B (HHV-6B) and Human Betaherpesvirus 7 (HHV-7), are primarily associated with exanthem subitem (roseola) as well as rejection of

transplanted kidneys [10]. Human cytomegalovirus (HCMV) is the focus of this study and thus a more comprehensive overview is given in the following sections.

1.2 Human cytomegalovirus

HCMV is a ubiquitous pathogen that infects the majority of the population worldwide by early adulthood [11]. Although approximately 10 % of infected individuals develop mononucleosis, in healthy individuals the infection tends to be asymptomatic [10]. It is one of the most common infections that lead to serious complications following transplantation, due to transmission from the allograft, where the virus reactivates from latent infection [12]. In the immunocompromised, HCMV can cause life-threatening diseases including pneumonia, hepatitis, retinitis, encephalitis and gastrointestinal disease [13]. Congenital HCMV infection is the leading viral cause of birth defects, including hearing loss, impaired vision and learning disabilities [14-16].

1.2.1 Virion structure

The viral DNA lies within the icosahedral capsid, which is surrounded by a thick proteinaceous tegument and sheathed by a lipid envelope derived from host cell membrane, with a mature virion (see Figure 1.2.1) ranging from 150 to 200nm in diameter [17, 18].

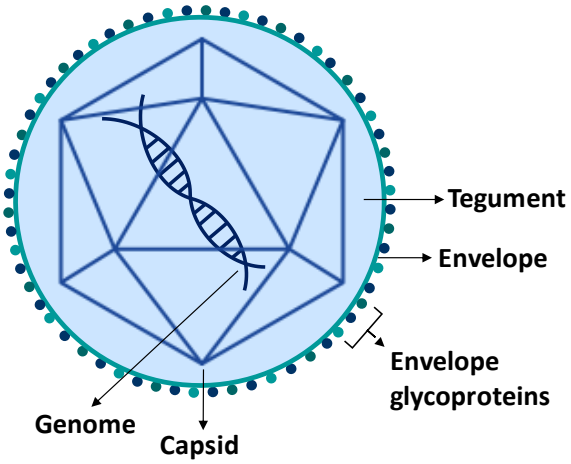


Figure 1.2.1 | The HCMV virion (not to scale)

1.2.1.1 Capsid

The capsid has an icosahedral structure, with 12 vertices and 20 faces, and is composed of 162 capsomeres and 320 triplexes. Capsomeres are termed pentons or hexons and consist of five or six major capsid protein UL86 copies respectively. Hexons form the edges and faces of the capsid, while one penton is found at each vertex. One of the vertices is occupied by a penton consisting of twelve copies of the portal protein UL104 with the remaining eleven being UL86 pentamers. The tips of each hexon are decorated with six copies of the small capsid protein UL48A [11]. Triplexes are made of one subunit of UL46 and two subunits of UL85 [19], and link the capsomeres together (see Figure 1.2.1.1).

Capsid assembly begins in the cytoplasm and is coordinated by the assembly protein precursor UL80.5 and the protease precursor UL80a. Both UL80.5 and UL80a contain a nuclear localisation signal (NLS), an amino conserved domain through which they interact with each other, and a carboxyl conserved domain which promotes binding to UL86. Nuclear translocation of UL86 is mediated by interaction with UL80.5 and UL80a, while the portal protein contains its own NLS. Triplexes also assemble in the cytoplasm and their translocation into the nucleus is enabled by the NLS of UL46. In the

nucleus, combinations of complexes containing UL86, UL80.5 and UL80a associate with the triplexes and the portal protein to form procapsids. The protease precursor is then activated resulting in elimination of both internal scaffold proteins UL80.5 and UL80a and capsid maturation.

The viral genome is then translocated into the capsid through the UL104 portal protein dodecamer via the terminase complex, a molecular motor composed of UL56, UL51 and UL89 [20]. Genome packaging generates C capsids, while capsids that lack DNA differ on the presence (B capsids) or absence (A capsids) of internal scaffold.

UL77 and UL93 then form the capsid vertex-specific complex (CVC), a heterodimer located exclusively around pentons. In addition to UL52, the CVC aid the terminase complex in the DNA packaging and cleavage process [20-25]. The CVC is also thought to label DNA-filled mature capsids as ready for nuclear egress and mediate this process by interacting with the nuclear egress complex.

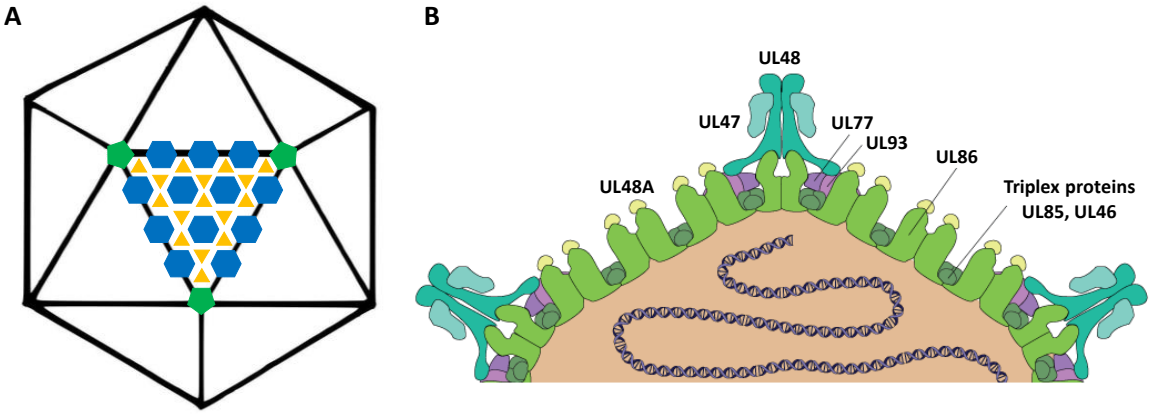


Figure 1.2.1.1 | Schematic representation of HCMV capsid organisation
(A) Surface lattice of the viral capsid, showing the positions of capsomeres and triplexes on one face. Major capsid protein UL86 pentons (green) are located in each vertex of the icosahedric structure, while UL86 hexons (blue) form the edges and faces of the capsid. Pentons and hexons are linked together by triplexes (yellow). **(B)** Cross section representation of the capsid. Small-capsid protein UL48A binds to the tips of UL86, forming a layer between the capsid and the tegument. Capsid triplex proteins UL85 and UL46 lie on the capsid floor between capsomeres. The capsid vertex specific component proteins UL77 and UL93, are located at the vertices, forming a linker between the capsid and the large tegument protein UL48. The latter plays a linker role for the association of the outer viral tegument to the capsids together with the inner tegument protein. Adapted from Viralzone (Swiss Institute of Bioinformatics) [26].

1.2.1.2 Tegument

Most of the viral proteins present in the virion can be found in the tegument, many of which are phosphorylated and carry out diverse functions, from modulating host cell signalling at the start of infection to coordinating the last stages of virion assembly [11]. Cellular proteins have also been found in the HCMV virion and are thought to be localised in the tegument. In addition, viral and cellular RNAs are also packaged within virions [27].

The tegument proteins UL50 and UL53 form the nuclear egress complex, allowing the capsid to dock onto the nuclear lamina and mediating its egress into the cytoplasm. UL50 localizes to the inner and outer leaflets of the nuclear membrane. The nuclear matrix protein UL53 interacts with the capsid through the CVC, directing it to the inner nuclear membrane by associating with UL50. Phosphorylation of the nuclear lamina is performed by the viral kinase UL97, which is recruited to the nuclear membrane by UL50 and UL53 [28]. Removal of the nuclear lamina barrier allows primary envelopment of the capsid with inner nuclear membrane, followed by translocation within the perinuclear space, de-envelopment at the outer nuclear membrane and release onto the cytoplasm. A subset of tegument proteins can be found in the nucleus, but it is thought that the majority of tegument proteins are acquired in the cytoplasm.

The most predominant proteins found in the HCMV tegument are the large tegument protein UL48, the inner tegument protein UL47, the basic phosphoprotein UL32 (BPP/pp150), the upper matrix protein UL82 (pp71), and the lower matrix protein UL83 (pp65) [29, 30]. The large tegument protein associates with the capsid by binding to the CVC. UL47 is added to the tegument by binding to UL48. The latter is thought to mediate the delivery of capsids along microtubules to the nuclear pore complex [31, 32]. UL48 has also been shown to have deubiquitinase activity [33].

The UL32 phosphoprotein (pp150) forms a filamentous net-like structure that surrounds the capsid. It associates with capsid components and promotes their stability during translocation from nucleus to cytoplasm. UL96 is added to capsids in the cytoplasm and plays a role in stabilisation of UL32-containing capsids [34].

The lower matrix protein UL83 is the most abundant tegument protein in virions, and localises to the nucleus immediately after viral entry. It has an immunomodulatory role during infection counteracting both innate and adaptive immune responses. It becomes associated with the virion during the envelopment stage.

The upper matrix protein UL82 is a sequence homolog of UL83 and similarly localises to the nucleus following entry, where it stimulates viral immediate-early (IE) transcription. The viral transactivation of UL83 is regulated by two isoforms of another tegument protein UL35 which have different temporal expressions and subcellular locations. The 75kDa isoform UL35 is expressed late during infection, is packaged into the virion and co-operates with UL82 in the activation of IE genes. Contrarily, the 22 kDa isoform UL35A is expressed early in infection, localises to the nucleus and promotes cytoplasmic UL82 accumulation, decreasing IE gene transcription.

The remaining teguments proteins are less abundant virion components and play various roles in HCMV infection including immunomodulation (UL23, UL25, UL26, UL45, US23, UL76), viral DNA replication/transcription (UL44, UL54, UL57, UL69, UL79, UL84, UL112, IRS1, TRS1, US24), modulation of host cell death pathways (UL36, UL38), latency (UL76), among others (UL24, US22, UL43, UL72, UL88).

Finally, the tegument proteins UL94, UL99, UL71 and UL103 locate to the cytoplasmic assembly complex and are involved in the last stages of virion maturation. The binding partners UL94 and UL99 are required for secondary envelopment, while

UL71 is required for final envelopment and binds to UL103, which has been implicated in virion egress from the host cell.

1.2.1.3 Envelope

The lipid envelope derives from the ER-Golgi intermediate compartment or endosomal membranes and contains approximately 23 viral glycoproteins. Five glycoproteins (gB, gH, gL, gM, gN) form complexes that contribute to attachment and entry. gB (UL55) is capable of forming homotrimers and plays a role in initial attachment by binding to heparan sulfate proteoglycans as well as membrane fusion [35, 36]. The gM:gN complex, encoded by genes UL100 (gM) and UL73 (gN), is also involved in initial attachment through binding of gM to heparan sulfate proteoglycans [37, 38]. The glycoproteins gH (UL75) and gL (UL115) form a heterodimer which can associate with additional proteins to influence attachment to different cell types. The trimeric complex gH:gL:gO facilitates entry into fibroblasts, while the pentameric gH:gL:UL128:UL130:UL131A is involved in entry into endothelial and epithelial cells [39]. UL116 has been shown to compete with gL for binding to gH, yet the function of a UL116-gH complex remains unknown [40].

Not all glycoproteins are involved in cell attachment/entry and instead have immunomodulatory roles, for example RL11, RL12, RL13 and UL119-118 can bind to IgG [41-43]. The secreted glycoprotein UL22A acts as virally-encoded decoy receptor for CC chemokines such as CCL5/RANTES (Regulated and Normal T Cell Expressed and Secreted) chemokine [44]. Other immunomodulatory glycoproteins include the viral G protein-coupled receptors (GPCR) encoded by US27, US28, UL33 and UL78. The viral GPCRs modulate host CXCR4 signalling, with US27 having an enhancing effect while the other three impair CXCR4 signalling outcomes [45]. In addition to CXCR4, UL33 and UL78

have both been shown to also heteromerize with CCR5 [46]. US27 influences viral spread, and may be required for extracellular spreading as deletion of this gene limits the virus to direct cell-to-cell spread [47]. US28 is a CC/CX3C chemokine receptor that binds and signals in response to multiple host chemokines (e.g. CX3CL1/fractalkine, CCL2/MCP-1, CCL5/RANTES, and CCL7/MCP-3) and also plays a role in latency [48-55].

Additionally, the glycoproteins RL10, UL1, UL4 and UL132 have been identified as virion envelope components [56, 57].

1.2.1.4 Other virion components

In addition to the viral genome and proteins it encodes, viral and host mRNAs, as well as cellular proteins, lipids and polyamines are also packaged into the virion [58].

Two small virus-associated RNA molecules (vRNA-1 and vRNA-2) can be found in the core of the virion, hybridized to the origin of DNA replication [59] and may act as a primer for initiation of DNA replication. Other viral RNAs encoded by the UL21A and RL13 genes, as well as the non-protein coding ORFs RL2-5 (RNA2.7), RL7 (RNA1.2) and UL106-UL109 (RNA5.0) in addition to at least 14 different small non-coding mature microRNAs (miRNAs) have also been found in virion preparations [60, 61]. However most of these species may be nonspecifically incorporated into the virion, potentially through interactions with several virion proteins [62].

Functions of host proteins incorporated into the virions include but are not limited to ATP-binding, calcium-binding (e.g. annexin I, annexin V, annexin VI, annexin A2, calreticulin), chaperones, cytoskeleton (e.g. α-Actin, β-actin, α-tubulin, β-tubulin), cellular enzymes (including several enzymatic catalysers of glycolysis), protein transport (e.g. clathrin), signal transduction (e.g. four isoforms of the signal transduction protein 14-3-3) and transcription-translation control [56].

Spermidine and spermine are two types of polyamines which have commonly been found in herpesvirus virions. Although polyamines are survival factors implicated in the regulation of programmed cell death, these positively charged molecules are thought to play a role during genome replication and packaging by neutralizing the negatively charged viral DNA, given that inhibition of polyamine biosynthesis inhibits virus growth at the level of virion assembly [63].

1.2.1.5 Non-infectious viral particles

During lytic infection, host cells also produce an excess of non-infectious by-products which can reach a 20:1 ratio to mature infectious virions [29]. These non-infectious particles can be distinguished from infectious virions using electron cryomicroscopy or separated using ultracentrifugation. They are likely to be a consequence of defective virion assembly but may also act as decoys for the immune system which facilitate survival of the infectious virions within the host [58].

Dense bodies are non-infectious viral particles which consist of a cluster of tegument proteins surrounded by the envelope lipid-bilayer but lack a viral capsid. In contrast, non-infectious enveloped particles appear to contain a B-capsid, as well as tegument and envelope. Both these types of particles are secreted from infected cells and do not contain viral DNA [64], suggesting that neither packaging of the genome nor capsid assembly are required for viral envelopment and exocytosis.

1.2.2 HCMV genome

The HCMV genome is approximately 236 kbp. It has E-type configuration consisting of two unique regions (U_L and U_S) flanked by inverted repeat segments (*a* and

a', TR_L and IR_L, TR_S and IR_S), resulting in an overall configuration *a*-TR_L-U_L-IR_L-*a'*-IR_S-U_S-TR_S-*a*. The terminal *a* sequences can recombine with the internal *a'* sequence (where the prime signal designates reverse orientation) generating four genome isomers (see Figure 1.2.2).

A partially duplicated IE gene set conserved in HCMV strains is located in the small IR_S and TR_S repeats (also designated as *c* and *c'*), thus the proteins they code (IRS1 and TRS1) have domains with nearly identical amino acid sequences. In contrast, the large TR_L and IR_L (also designated as *b* and *b'*) repeats appear to be a consequence of extensive passage in vitro [11].

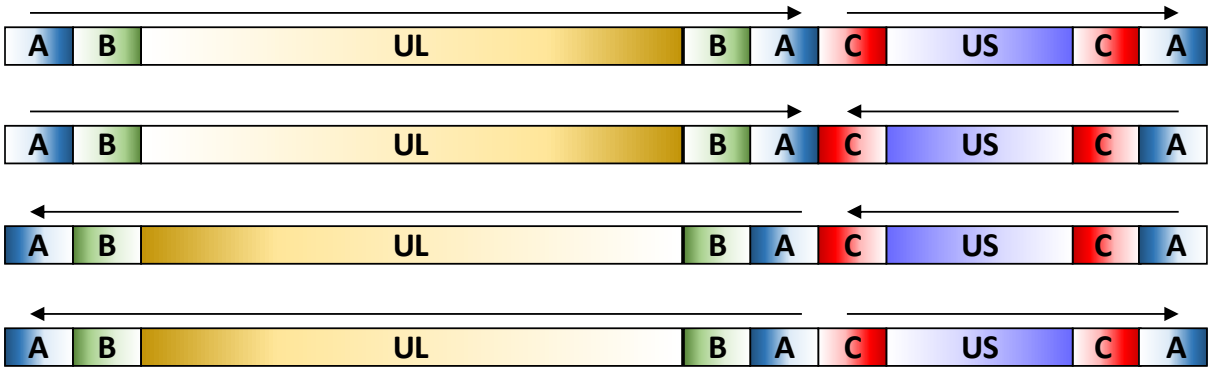


Figure 1.2.2 | Organisation and isomerisation of the HCMV genome
Representation of the four isomers of HCMV genome (not to scale) according to the orientation of U_L and U_S regions relative to each other.

1.2.2.1 Cis-*acting sequences*

The genome contains *cis*-acting sequences which are important for cleavage and packaging of newly-synthesised viral genomes (*pac*1 and *pac*2), DNA replication (*ori*Lyt) and RNA transcription.

Genome packaging is mediated through viral protein recognition of two herpesvirus-conserved sequence motifs, *pac-1* and *pac-2*, located in the *a* sequence at each terminus of the viral genome [65]. These *pac* elements are recognized by the

terminase complex to both translocate the viral DNA into a capsid and signal cleavage once a genome-length is reached. The terminase subunit UL56 binds to AT-rich regions in the *pac* sequences while the subunit UL89 cleaves the genome producing ends at the location of *pac1* and *pac2* with single overhanging 3´-nucleotides.

The origin of DNA synthesis, *ori*Lyt, is located in a position conserved in all characterized betaherpesviruses, between the genes UL57 and UL69. It is approximately 1.5 kbp and has several features, including a pyrimidine-rich sequence, reiterated elements, direct and inverted repeats, transcription factor-binding sites and RNA-DNA hybrids formed with vRNA-1 and vRNA-2.

Two other types of cis-acting sequences present in the viral genome include transcription enhancers and promoters that are activated at different times throughout infection and regulate the activity of host cell RNA polymerase II machinery. HCMV genes have individual promoters, except for a small subset of genes that are expressed by splicing from a shared 5' leader sequence (e.g. IE1/UL123 and IE2/UL122).

1.2.2.2 Gene products

HCMV encodes for 171 canonical genes that are present in all clinical isolates and are likely to code proteins (see Figure 1.2.2.2 and Table 1.2.2.2). A variety of functions has been assigned to most of these canonical proteins including: modulation of transcription, protein synthesis, immune response and host cell cycle; viral tropism, DNA replication, virion assembly/maturation/egress, latency, apoptosis inhibition and cellular trafficking [66]. Ribosome profiling studies have identified 751 translated ORFs, yet it is unclear whether these encode for functional transcripts [67].

The HCMV genome contains 40 core genes, which are shared among alpha, beta and gammaherpesvirus and are presumed to have been inherited from a common

ancestor. Similarly, it contains 7 genes which are shared among beta and gammaherpesvirus only, thus termed sub-core. Most core genes are essential for viral growth and are involved in vital processes such as DNA replication, packaging, capsid structure and assembly. Sub-core genes are required for the transcription of late transcripts. The remaining non-core canonical genes are grouped into gene families according to the likelihood of arising from gene duplication, although the proteins they code have different functions. Besides gene duplication, acquisition of genes from the cell or from other viruses has played a role in increasing the coding capacity of HCMV. Gene capture is likely to have occurred by insertion of a cDNA copy of a cellular mRNA or pre-mRNA into the viral genome, since most of the cellular homolog genes lack introns.

Additional ORFs encode non-coding RNAs including the four long non-coding RNAs (such as RNA2.7, RNA1.2, RNA4.9 and RNA5.0), at least 23 miRNAs most of which associate with the RNA-induced silencing complex, and antisense transcripts which are transcribed antisense to coding regions. The majority still has unclear functions, however cellular and viral targets for some of these RNA molecules have been identified, hinting that they may be involved in promoting changes in the infected cell to favour lytic infection [68].

Figure 1.2.2.2 | Genetic map of HCMV Merlin genome
Protein-coding regions are indicated by coloured arrows grouped in gene families as depicted by the key, with gene nomenclature below. UL72 is both a member of the deoxyuridine triphosphatase-related protein (DURP) gene family and a core gene. Introns are shown as narrow white bars. The UL and US prefixes have been omitted from UL1–UL150 and US1–US34A. Colours differentiate between genes on the basis of conservation across the α-, β- and γ-herpesvirinae (core genes) or between the β- and γ-herpesvirinae (sub-core genes). Long non-coding RNAs, the origin of lytic replication (oriLyt) and the R_L and R_S regions (which contain the '*a*' sequence as a direct repeat at the genome termini and as an inverted repeat internally) are indicated by coloured boxes. Adapted from Dolan et al (2004) and Gatherer et al (2011) [69, 70].

Table 1.2.2.2 | Canonical protein coding genes in HCMV Merlin strain

Gene	Protein name	Gene	Protein name	Gene	Protein name
RL1	IRL1, HKLF1	UL51	UL51, TRM2	UL131A	UL131A
RL5A	RL5A	UL52	UL52	UL132	UL132, gp42
RL6	RL6	UL53	UL53, NEC1	UL133	UL133
RL8A	RL8A	UL54	UL54 DPOL	UL135	UL135
RL9A	RL9A	UL55	UL55, gB, gp116, gp58	UL138	UL138
RL10	RL10, gpTRL10	UL56	UL56, TRM1	UL139	UL139
RL11	RL11, gpTRL11, gp34	UL57	UL57, SSB, p130	UL140	UL140
RL12	RL12, gpTRL12, gp95	UL69	UL69, MRE	UL141	UL141, gpUL141
RL13	gpTRL13, RL13/14	UL70	UL70, Primase	UL142	UL142
UL1	UL1, gp1	UL71	UL71, CEF1	UL144	UL144
UL2	UL2	UL72	UL72	UL145	UL145
UL4	UL4, gp48	UL73	UL73, gN	UL146	UL146, vCXCL1
UL5	UL5	UL74	UL74, gO, gp125	UL147	UL147, vCXCL2
UL6	UL6	UL74A	UL74A	UL147A	UL147A
UL7	UL7	UL75	UL75, gH, gp86	UL148	UL148
UL8	UL8	UL76	UL76	UL148A	UL148A
UL9	UL9	UL77	UL77, CVC2	UL148B	UL148B
UL10	UL10	UL78	UL78	UL148C	UL148C
UL11	UL11	UL79	UL79	UL148D	UL148D
UL13	UL13	UL80	UL80, UL80a, pPR, p38	UL150	UL150
UL14	UL14	UL80.5	UL80.5, pAP	UL150A	UL150A
UL15A	UL15A	UL82	UL82, pp71	IRS1	IRS1
UL16	UL16, gpUL16	UL83	UL83, pp65	US1	US1
UL17	UL17	UL84	UL84	US2	US2
UL18	UL18, gpUL18	UL85	UL85, TRI2, TRX2	US3	US3
UL19	UL19	UL86	UL86, MCP	US6	US6
UL20	UL20	UL87	UL87	US7	US7
UL21A	UL21	UL88	UL88	US8	US8
UL22A	UL21.5	UL89	UL89, TRM3	US9	US9
UL23	UL23	UL91	UL91	US10	US10
UL24	UL24	UL92	UL92	US11	US11
UL25	UL25, pp85	UL93	UL93, CVC1	US12	US12
UL26	UL25	UL94	UL94, CEP2	US13	US13
UL27	UL27	UL95	UL95	US14	US14
UL29	UL29/28	UL96	UL96	US15	US15
UL30	UL30	UL97	UL97, VPK	US16	US16
UL30A	UL30A	UL98	UL98, NUC	US17	US17
UL31	UL31	UL99	UL99, CEP3, pp28	US18	US18
UL32	UL32, BPP, pp150	UL100	UL100, gM	US19	US19
UL33	UL33	UL102	UL102, HEPA	US20	US20
UL34	UL34	UL103	UL103, CEP1	US21	US21
UL35	UL35	UL104	UL104, Portal protein	US22	US22
UL36	UL36, vICA	UL105	UL105, Helicase	US23	US23
UL37	UL37, vMIA, gpUL37	UL111A	UL111A, vIL-10	US24	US24
UL38	UL38	UL112	UL112-113	US26	US26
UL40	UL39	UL114	UL114	US27	US27
UL41A	UL41A, UL41.5	UL115	UL115, Gl	US28	US28
UL42	UL42	UL116	UL116,	US29	US29
UL43	UL43	UL117	UL117	US30	US30
UL44	UL44, VPAP, PPS, p52	UL119	gp68, UL119-118	US31	US31
UL45	UL45	UL120	UL120	US32	US32
UL46	UL46, TRI1, TRX1	UL121	UL121	US33A	US33A
UL47	UL47, ITP	UL122	UL122, IE2, IE-86	US34	US34
UL48	UL48, LTG	UL123	UL123, IE1, IE-72	US34A	US34A
UL48A	UL48A, SCP, UL48.5	UL124	UL124	TRS1	TRS1
UL49	UL49	UL128	UL128		
UL50	UL50, NEC2	UL130	-		

1.2.2.3 Viral DNA replication

Viral DNA replication takes place in the nucleus of infected cells, with peak production approximately around 24 h post-infection. The mechanism is not completely understood, but several viral proteins and regions in the viral genome have been found to play a role in this process.

Genomes isolated from virions are linear, yet the viral DNA circularises once it is delivered to the nucleus and replication occurs in a rolling circle mechanism, generating tandemly linked copies. As mentioned in **1.2.2.1**, these are cleaved into genome units by the viral terminase during encapsidation.

Synthesis of new genome copies initiates at the lytic origin of replication, *ori*Lyt, which contains two regions that are essential to viral DNA replication. Essential Region I, contains a bidirectional promoter and pyridine-rich sequence termed the Y-block. Essential Region II, comprises an RNA-DNA hybrid structure containing a vRNA-2 binding site and an adjacent RNA stem-loop, and is enriched in direct and inverted repeat sequences as well as transcription factor-binding sites.

Initiation of DNA synthesis is dependent on the formation of a complex between UL84 and IE2-p86 (encoded by UL122), which binds to and activates the *ori*Lyt promoter. Binding of UL84 to IE2-p86 disrupts its gene expression transactivation activity. Additionally, UL84 also binds the RNA stem-loop and acts to change its conformation [71].

The UL112-113 ORF encodes four phosphoproteins (pp34, pp43, pp50, pp84) which are produced via alternative mRNA splicing and form a complex that is thought to mediate assembly of the viral replisome. In complex with the UL112-113 proteins and IE2-p86, UL84 recruits the polymerase processivity subunit UL44, which recruits the other replication fork proteins: the DNA polymerase catalytic subunit UL54, the single-

stranded DNA-binding protein UL57, and the heterotrimeric helicase-primase complex consisting of the helicase UL105, the primase UL70 and a helicase-primase associated factor UL102. UL44 also interacts with the alkaline nuclease UL98 and the DNA glycosylase UL114, which appears to be important for efficient DNA synthesis.

Other viral proteins encoded by UL36 (vICA), UL37 (vMIA), IRS1 and TRS1 are not involved in the replication compartment but support HCMV DNA replication by modulating the host cell environment. Additionally, cellular factors such as the oriLyt-binding C-EBP and the cell-viability promoting hnRNP-K are necessary for viral DNA synthesis.

1.2.3 Lytic lifecycle

HCMV initiates infection by binding to heparin sulphate proteoglycans on the cell surface [35]. Following initial attachment, interactions between gB and cellular integrins α2β1, α6β1 and αVβ3 are thought to be required [72]. For fibroblast cells, a trimeric complex gH:gL:gO facilitates entry, in addition to gB, and the viral envelope fuses with the cellular membrane. For endothelial and epithelial cells, the pentameric complex gH:gL:UL128:UL130:UL131A mediates entry via endocytosis [39]. Once the nucleocapsid is released into the cytoplasm, tegument proteins aid its translocation towards the nucleus via microtubules. There it interacts with nuclear pores in order to release the viral genome into the nucleus. Other tegument proteins play a role in increasing efficiency of IE gene expression [73].

Viral gene expression follows a phasic pattern, with IE genes being the first to be expressed. IE gene expression is dependent on pre-formed host and viral machinery as well as the tegument protein pp71 (UL82) which functions as a transactivator. The major IE gene, transcribed from the major immediate early promotor (MIEP), encodes two

predominant nuclear phosphoproteins, IE1-p72 (from gene UL123) and IE2-p86 (from gene UL122). Together, these proteins reverse epigenetic repression, induce expression of E and L genes, self-regulate IE gene expression and contribute to the switch between latency and reactivation. The E genes are necessary for viral DNA replication, which starts at approximately 14h post-infection. Their activation requires expression of IE genes. Once DNA replication has been initiated, the E-dependent L gene expression peaks. These proteins have mostly structural functions (capsid and tegument components) and their products control capsid maturation, DNA encapsidation, virion maturation and egress [11].

Virion assembly begins in the nucleus. Firstly, the capsid is assembled (described in **1.2.1.1**) and the genome is encapsidated. Initial components of the tegument are added to the capsid, forming an inner tegument that provides stability during nuclear translocation and directs trafficking to sites of envelopment in the cytoplasm. Then the nuclear egress complex, comprised by UL50 (NEC1) and UL53 (NEC2), mediates translocation of the nucleocapsid to the cytoplasm by recruiting kinases that disrupt the nuclear lamina. Once in the cytoplasm, the secretory apparatus termed the assembly complex mediates final tegumentation and envelopment. The mature virion is then transported to the cell surface in small vesicles, exiting the cell via exocytosis.

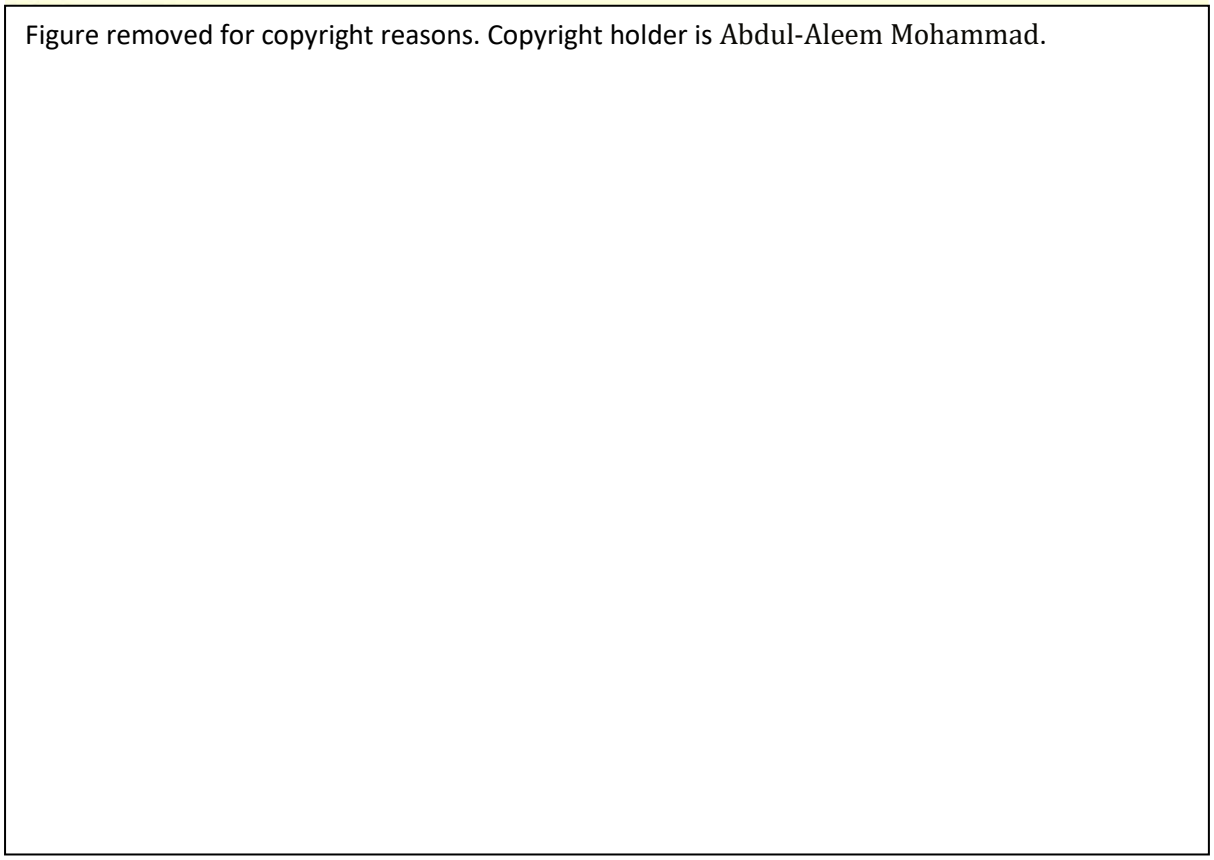


Figure 1.2.3 | Overview of the human cytomegalovirus life cycle
(A) A virion initiates infection by attaching to the cell surface, followed by entry via fusion of the viral envelope with the plasma membrane or endocytosis. Tegument proteins and the nucleocapsid are released into the cytosol. **(B)** The nucleocapsid is translocated towards the nucleus, along cytoplasmic microtubules. The viral genome is released into the nucleus where a temporal cascade of gene expression is activated: firstly immediate early (IE) genes, followed by delayed early (DE) genes, which initiate viral genome replication, and late (L) genes. **(C)** Translation of the structural virion components initiates assembly of viral capsid in the nucleus. The viral genome is then translocated into the capsid, followed by nuclear egress and acquisition of tegument proteins in the cytosol. These particles are then trafficked to the viral assembly complex (AC) which contains components of the endoplasmic reticulum (ER), Golgi apparatus, and endosomal machinery. There they acquire more tegument proteins and a viral envelope by budding into intracellular vesicles at the AC. **(D)** Mature virions are released along with non-infectious dense bodies through exocytosis. Figure and figure legends were adapted from (Jean Beltran & Cristea, 2014) and modified by Abdul-Aleem Mohammad. Reprinted from Abdul-Aleem Mohammad, 2017: "Human Cytomegalovirus: From Novel Strain, miRNAs to Interplay with Breast Cancer". [74, 75]

1.2.4 Viral latency

Like all herpesvirus, HCMV is able to persist within the host by establishing a generally asymptomatic, lifelong infection. Although viral DNA can be detected, immediate early (IE) gene expression is suppressed and no infectious virions are produced [76]. Cells of the myeloid lineage, including $CD14^+$ monocytes and their $CD34^+$ progenitor, are well established sites of latent infection, however cell differentiation triggers viral reactivation and lytic replication [77-79].

Similarly to other herpesvirus, gene expression during HCMV latency was initially thought to be restricted to a subset of viral genes and thus different from the lytic infection transcriptional programme [80]. In addition, different isoforms of viral transcripts have been shown to be expressed during latency [81, 82]. Two recent studies of the viral latent transcriptome have shown a broader pool of transcripts which largely resembles a late stage of lytic infection, although at lower expression level [83, 84].

1.2.5 Transmission and tropism

Transmission requires direct contact with infected bodily fluids (e.g. saliva, breast milk, urine and genital secretions) [85, 86]. HCMV can be vertically-transmitted from an infected mother to an embryo, fetus or baby. Horizontal transmission occurs predominantly by contact with infected children (especially in childcare centres), through sexual activity in adults [87], and via blood transfusion and cells, tissues or organ transplantation from seropositive donors [11].

The virus is able to infect a wide range of cells, including parenchyma and connective tissue of any organ, as well as several hematopoietic cell types. Infection of endothelial cells and hematopoietic cells facilitate systemic spread within host while

infection of epithelial cells in salivary glands, kidney and gastrointestinal tract likely contributes to inter-host transmission, by viral shedding into body fluids. Infection of ubiquitous cell types such as fibroblasts and smooth muscle cells enable efficient proliferation [88].

A number of immortalised cell lines (glioblastoma, monocytic) and primary cell cultures (skin and lung fibroblasts, vascular smooth muscle cells, vascular endothelial cells, retina pigment epithelial cells, placental trophoblasts, hepatocytes and kidney epithelial cells, monocyte-derived macrophages and dendritic cells, neuronal and glial brain cells) support a complete viral replication cycle [88].

Viral tropism varies among HCMV strains, however skin and lung fibroblasts are the standard cell type for isolation and propagation from patient samples and the most efficient progeny producers [11].

1.2.6 Strains of HCMV

There are two types of HCMV strains: clinical isolates that have been passaged minimally in fibroblasts and laboratory strains that have been extensively passaged and adapted to growth in fibroblasts [89].

The most commonly used lab-adapted strains are AD169 and Towne [90], which was passaged 125 times in vitro in order to develop an attenuated vaccine [91]. High-passage led to the development of mutations that affected several genes in AD169 (RL5A, RL13, UL36 and UL131A) and Towne (RL13, UL1, UL40, UL130, US1 and US9). In addition, these high-passage strains differ from wild-type HCMV as the U_L/b' region has been replaced. A 15 kbp deletion in AD169 resulted in the absence of genes between UL132-IRS1 and their replacement with a duplication of U_L/b, containing copies of TR_L1-14 termed as IR_L1-14. Similarly, in Towne, a 13 kbp deletion in U_L/b' replaced genes

between UL148-IRS1 with a duplication of U_L/b containing copies of TRL1-14 [92]. Furthermore, variants of each strain are generated as they are passaged in different laboratories accumulating diverging mutations since their initial isolation.

Another commonly used lab-strain, TB40/E, has been reported to possess the broad cell tropism of a clinical isolate and the high titre growth of a cell culture adapted strain [93, 94]. However, this strain has been shown to be genetically and phenotypically heterogeneous, with more than one virus involved in its derivation, thus complicating correlation between genotype and phenotype [69].

In 1996, Cha et al., showed that the low passage strain Toledo had a segment inversion which disrupted at least UL128 [95]. Several other low passage strains have also been shown to contain disruptions in UL128, UL130 and UL131A [96] or genes in the RL11 family, a consequence of adaptation to fibroblast cell culture.

Given that passage of virus rapidly causes adaptive mutations, no laboratory strain can be assumed to be genetically intact as any viral gene in a cultured stock could be mutated and this fact could go unrecognized, making it easy to overlook loss of gene function [90]. This led to the cloning of the genome of several HCMV strains onto bacterial artificial chromosomes (BAC) to generate clones where each genome can be maintained without acquiring additional mutations, as well as facilitating their genetic manipulation to match the sequence in the clinical isolate. Many of these BAC clones incorporated an element of the vector cassette within the U_S region replacing several viral genes with immunomodulatory function [97].

The genome sequence of HCMV strain Merlin is designated the reference HCMV sequence by the National Center for Biotechnology Information [69]. A recombinant version (RCMV 1111) of this strain was cloned into a BAC and repaired to match the viral sequence of the original clinical isolate from which it derived. Given that restoration of

RL13 and UL128 impairs growth in fibroblasts, RCMV 1111 contains point mutations in these two genes to relieve selective pressure [98].

1.2.7 Immune response and modulation

1.2.7.1 Innate immune response

With an intricate network of activating and inhibiting receptors, NK cells are crucial to controlling HCMV infection [99, 100] via direct cytotoxicity (either antibody-dependent or via perforin and granzyme) as well as the production of cytokines and chemokines such as interferon-gamma and tumour necrosis factor alpha [101].

Binding of HCMV envelope glycoproteins to cellular receptors trigger signalling cascades that result in the activation of cellular transcription factors such as NFκ-B, Sp1, and interferon regulatory factor 3 (IRF3) [102-105]. The toll-like receptor 2 (TLR2) recognises glycoproteins gB and gH and activates NFκ-B signalling resulting in the production of inflammatory cytokines such as IL-6 and IL-8 [104, 106]. The Z-DNA binding protein 1 (ZBP1) detects HCMV DNA and activates IRF3 signalling, triggering interferon type I production and expression of interferon-stimulated genes (ISGs) [107]. The PML-nuclear body constituents Daxx, Sp100 and PML proteins have been shown to repress expression of HCMV IE genes in an interferon-inducible manner [108-111].

The catalytic activity of the ISG-encoded 2'-5' oligoadenylate synthase (OAS) results in the activation of RNase L, leading to the degradation of single-stranded mRNA and rRNA [112]. Upon recognition of dsRNA, the ISG-encoded protein kinase R (PKR) halts protein synthesis by phosphorylating the eukaryotic initiation factor 2α (eIF2α) [112]. Together, OAS and PKR create an antiviral environment that prevents the virus from translating its proteins.

1.2.7.2 Adaptive immune response

An antigen specific immune response is elicited by antigen presenting cells (APCs) upon processing of viral particles. Antigens are displayed on the surface of APCs by the major histocompatibility complex (MHC) class II molecules to $CD4^+$ T cells; and in the surface of all nucleated cells by class I molecules to $CD8^+$ T cells, generating a specific subset to a range of viral epitopes. These cells restrain viral replication and are eventually maintained as long term memory T cells [113, 114]. HCMV-specific $CD4^+$ and $CD8^+$ T cells secrete cytokines (such as IFN-γ) to regulate the immune response in addition to expressing and releasing the cytolytic proteins granzyme and perforin onto the cytoplasm of infected cells upon recognition of vira antigens, triggering apoptosis [115-121].

B cells also play a role in controlling HCMV infection by producing antibodies against viral proteins such as the envelope glycoproteins (e.g. gB, gH and gL) and virion tegument proteins (e.g. UL32/pp150 and UL83/pp65) [122-127].

1.2.7.3 Immune evasion

HCMV has developed several strategies to counteract the actions of the host immune system. Modulation of MHC class I and II molecules by viral transcripts US2, US3, US6, US10, US11, UL82 (pp71), UL83 (pp65) disrupt antigen processing and presentation [128]. Furthermore, the miRNA miR-US4.1 downregulates ERAP-1, inhibiting viral peptide trimming for MHC Class I peptide presentation and consequently decreasing the susceptibility of infected cells to HCMV-specific cytotoxic T lymphocytes [129].

Multiple viral proteins are involved in the inhibition of NK cell-mediated killing [130]. The viral MHC I homolog UL18 binds the inhibitory receptor LIR1 on NK cell

surface [131]. UL142, also an MHC I homolog, promotes downregulation of MICA, a ligand for activating receptor NKG2D [132]. The glycoprotein UL40 upregulates HLA-E, which is a ligand for the inhibitory complex CD94/NKG2A on NK cells [133]. Tegument protein UL83/pp65 binds activating receptor NKp30, causing its dissociation from CD247 and repressing the NK activating ability of this complex [134]. UL16 has been shown to sequester MICB, ULBP1 and ULBP2 in the ER, preventing their binding to the NK activating receptor NKG2D [135-137]. The UL/b' region has been shown to encode for transcripts that protect from NK cell attack [138]. Besides UL142, another gene from the UL/b' region, UL141 has been shown to repress NK cell cytotoxicity by sequestering CD155 in the ER, which is a ligand for activating receptors CD226 and CD96 [138]. The miRNA miR-UL112 targets MICB mRNA suppressing cell surface expression and preventing this ligand from activating the natural killer cell receptor NKG2D [139].

To evade complement-mediated lysis, HCMV incorporates complement control proteins CD55 and CD59 into its virion, which are able to inhibit various stages of the complement cascade [140]. Additionally, two viral Fc gamma receptor homologs, encoded from genes RL11 and UL119-118 have been shown to serve as a receptor for IgG and thus may be involved in evading complement activation and antigen-dependent cell-mediated cytoxicity [141].

Modulation of the IFN signalling pathway is achieved through the 72kDa immediate early 1 protein (IE1) and 86kDa immediate early protein 2 (IE2). IE1 binds STAT1 and STAT2 preventing the transcriptional activator ISGF3 from binding to the ISG promoter [142]. IE2 represses transcription of IFN-β and CCL5 through binding to the transcription factor NFκB [143]. The miRNA miR-UL148D also targets CCL5 by inducing degradation of its mRNA, preventing its translation into a T-cell chemoattractant [129].

Viral proteins TRS1 and IRS1 can suppress block the OAS and PKR pathways, preventing activation of RNase L and protein synthesis shutoff [144].

The virally encoded interleukin-10 homolog, UL111A, mimics the function of this immunosuppressive cytokine, downregulating IFNγ and TNFα and decreasing the expression of MHC class I and II molecules [145, 146]. Additionally, UL111A can inhibit cytokine production via activation of the phosphatidylinositol 3-kinase/Akt pathway [147]. The viral GPCR homologs US27 and US28 bind and internalise cytokines, removing them from the extracellular environment [148, 149]. UL144 is able to inhibit T-cell proliferation in vitro and induce CCL22 expression, a TH2 chemoattractant, to subvert TH1 response [150, 151]. The latency-associated UL138 has been shown to induce TNFR1 hyper-responsiveness, this is thought to contribute to contribute to HCMV reactivation by re-initiation of transcriptional programmes leading to lytic infection [152, 153].

1.2.8 Antiviral treatment

Several antiviral therapies have been approved to treat HCMV infection, having been shown to eliminate or reduce viremia and viral shedding in immunocompromised adults [11]. The majority of these compounds target the DNA polymerase UL54 and are accompanied by significant toxicity, except for the recently approved Letermovir (see Table 1.2.8). Maribavir, another antiviral drug targets the viral kinase UL97 but is still being tested. However, therapeutic agents targeting viral proteins are subject to the development of resistant mutations, rendering them ineffective. Vaccines against HCMV are still in development [154].

Table 1.2.8 | Antiviral agents approved for treatment or prevention of HCMV infection in immunocompromised adults

Table removed for copyright reasons. Copyright holder is WOLTERS KLUWER HEALTH.

Adapted from Mocarski et al, 2013: 'Cytomegaloviruses'. [11, 154]

1.3 Proteomics as a tool for studying viral infection

Viruses require a living host cell for replication. Their replicative cycle relies on host-pathogen interactions which have been adapted throughout millions of years of co-evolution with their host. Although viral replication strategies may differ, host-pathogen interactions occur regularly throughout the virus lifecycle and underpin processes such as cell entry, recruitment of intracellular machinery for trafficking and replication as well as subversion of the immune system. Thus, understanding the molecular basis for the biological processes that take place throughout the viral life cycle inside the host is necessary for the development of means to prevent and treat disease caused by these pathogens.

In the last decade, genomic, transcriptomic, proteomic and metabolomics approaches have emerged as powerful and effective tools, often in a complementary manner, in basic, translational and clinical research. For example, a study by Soderholm et al (2016) compiled data from transcriptomics, metabolomics, proteomics (using iTRAQ technology, a type of isobaric tags that allows up to 8-plex analysis) and phosphoproteomics, providing a complete analysis of gene regulation during Influenza A infection [155].

Proteomics, the large-scale study of the proteome of a biological system, provides sensitive protein detection and quantification. Proteomics approaches have been developed to study modulation of protein abundance, spatial organisation of the

proteome, pathogen-induced post-translational regulation and mapping host–pathogen protein interaction networks. In addition to allowing the study of thousands of proteins in a single experiment, through the use of sample labelling methods such as stable isotope labelling of amino acids in cell culture (SILAC) [156] or tandem-mass tagging (TMT) labelling [157], proteomics allows analysis of up to 3 or 16 samples (respectively) in tandem.

1.3.1 Proteomic studies of HCMV infection

To study HCMV infection, five kinds of proteomic approaches (briefly summarised below) have been applied:

(a) Studies of the virion

(b) Temporal studies of viral and host protein changes throughout infection

(c) Studies of post-translational modifications by viral infection

(d) Studies of changes of viral and host proteins in space during infection

(e) Studies of individual viral-host protein interactions

In 2004, Varnum et al., employed mass spectrometry to identify proteins present in the particles of virions from HCMV AD169 strain [56]. This was extended by Buscher et al (2015), using proteomics to examine molar ratios of proteins in virions from different laboratory strains, suggesting that most proteins were in fact present in similar ratios across the strains studied [158].

In 2013, Weekes et al studied the regulation of the plasma membrane (PM) proteome by UL138 during latent HCMV infection and identified the regulation of the multidrug-resistance protein 1 (MRP1) [159]. This study employed PM profiling, a

technique that combined enrichment of oxidated and aminooxy-biotinylated sialylated PM proteins with SILAC-based quantitation to allow sample multiplexing [160]. In 2014, Reyda et al used label-free quantification to study protein relative abundance over three time-points of infection with two HCMV AD169-derived strains, in addition to performing proteomics on purified virions from these strains [161]. In the same year, a study by Weekes et al, analysed changes in PM and whole cell proteins during the course of HCMV Merlin strain infection. Using the multiplexing capacity of TMT tags, it provided a comprehensive quantification of over 8000 whole cell lysate proteins (WCL) and 1200 PM proteins throughout the course of HCMV infection, with 4 time points for PM and 8 time points for WCL. This study reported on the HCMV-induced modulation of cellular pathways and regulation of specific protein families, including proteins important in NK and T-cell recognition, and innate immunity. Additionally, it confirmed the identification of a subset of new viral ORFs, some of which had been previously identified by ribosomal profiling [67], as well as grouping viral proteins into five classes according to their profiles of relative abundance throughout infection. This was the first multiplexed study of viral proteomics using MS3/TMT technology and paved the way for several temporal viromic studies [162]. Another multiplexed approach by Nightingale et al (2018), used three orthogonal proteomic/transcriptomic screens to identify putative HCMV restriction factors on the basis of pathogen-induced degradation by the proteasome or lysosome early during infection. This degradation dataset was then combined with a comprehensive panel of HCMV gene block deletion mutants to predict the viral genes targeting more than 250 host proteins [163].

During infection, modulation of gene expression and function can also be regulated through post-translational modifications, such as phosphorylation and ubiquitination. The most abundant virion protein, UL83 plays a role in immune evasion

by inhibiting IFI16-mediated DNA-sensing [164]. A study by Li et al. (2013), showed that UL83 is phosphorylated in multiple sites by host kinases in order to interfere with its immune evasion function [165]. Moreover, the virus can also modulate phosphorylation of host proteins. Using SILAC, Oberstein et al. (2015), analysed UL97-induced changes on the cellular phosphoproteome [166]. In 2012, Sinigalia et al, showed that the viral DNA polymerase processivity factor UL44 has multiple sites for small ubiquitin-related modifier (SUMO) and is SUMOylated, enhancing viral replication [167].

During infection viral proteins are shuttled to the appropriate cellular compartments either to play a role in replication or for virion assembly [168]. Often, the pathogen also induces changes in the pools of specific host proteins, either through retention in a cellular compartment or shuttling to another subcellular location. In 2016, Beltran et al used a combination of density-gradient fractionation of organelles and MS2-level TMT proteomics to identify the spatial dynamics of proteins throughout the course of infection with HCMV AD169 strain [169]. This so-called spatial proteomics provided data on the subcellular localisation of individual viral proteins through the course of infection, and how HCMV may induce host proteins to translocate from one compartment to another.

Studies of individual protein-protein interactions are the most commonly used to elucidate the contribution of numerous viral protein for infection [170-174]. The most common affinity-purification (AP-MS) approaches involve heterologous expression of viral proteins tagged with an epitope for immunoprecipitation or expression of virally encoded tagged protein. However, this is normally done for individual proteins and a large-scale analysis of protein-protein interactions during HMC infection had not yet been done.

1.3.2 Affinity-purification mass spectrometry

Affinity-purification mass spectrometry (AP-MS) allows the identification of proteins interacting with a given bait. To achieve this, clarified cell lysates are incubated with antibodies conjugated to a resin (agarose, sepharose, magnetic beads) and protein complexes are isolated. Proteins are then digested into peptides, separated through liquid-chromatography and analysed by tandem mass spectrometry.

While label-free quantitation is the most common, multiplex analysis can be achieved via the use of SILAC or isobaric tag technologies (TMT, iTRAQ). Moreover, there are several approaches that employ the use of epitope tagged proteins for the study of interactions, such as cellular overexpression of tagged viral proteins or infection with virus that encodes a tagged bait.

In addition to HCMV, several studies have applied this technology to identify interactors to epitope-tagged proteins of other virus such as HSV-1, KSHV, HPV, Human respiratory syncytial virus, Influenza, Dengue virus, Sindbis virus, Ebola virus and HIV and West Nile virus [175-183]. For a small subset of virus, such as HCV, Ebola, HIV and KSHV, this approach has been used to generate a complete or near-complete map of interactions for all viral proteins [184-187].

A drawback to this approach is the identification of unspecific interactions, yet truly specific interactions may be singled out via comparison to appropriate controls which allow the establishment of a background of common contaminants. To aid this process, a repository of contaminants for AP-MS data has been compiled from negative controls of multiple AP-MS studies to increase background coverage [188].

To distinguish the interactions that are most likely to be true from the non-specific contaminants, these AP-MS datasets are usually analysed using filtering tools with algorithms that score all the putative interactors. For example, in 2015 Huttlin et al.

published a fraction of the human interactome. AP-MS analysis was used for 2,594 baits and all detected interactions were then scored using the algorithm from CompPASS. Through comparison of the AP-MS samples being analysed in tandem, CompPass identified 23,744 interactions involving 7,668 proteins [189]. These were then compiled into the BioPlex Network (https://bioplex.hms.harvard.edu/). CompPASS measured enrichment of each protein for each IP in comparison to other unrelated AP-MS data and its algorithm features several scoring metrics that altogether take into account protein abundance, detection frequency and reproducibility amongst replicates.

1.4 Project aims

The many proteomic approaches employed to study the biology of HCMV infection have provided information on how the abundance and location of cellular and viral proteins changes during infection. This may allow the prediction of potential interactions that explain patterns of gene regulation, by juxtaposing cellular and viral protein profiles. Given the number of canonical and non-canonical ORFs encoded by HCMV, identifying which viral protein is responsible for a given modulation of cellular factors may prove to be labour intensive [67, 163]. Furthermore, certain cellular proteins may be targeted by more than one viral factor [130, 190-192].

To deepen our understanding of the molecular mechanisms underlying the changes observed during HCMV infection, this project aimed to identify viral-viral and viral-human protein interactions in cells infected with HCMV.

An analysis of HCMV protein interactions can be achieved by generating stable cell lines that overexpress each of the 171 canonical proteins tagged with a V5 epitope. Additionally, two uncharacterised ORFs, ORFL147C and ORFS343C.iORF1 (referred to from here on as ORFS343C) which have been identified in previous proteomics studies

and by ribosomal profiling will be included based on the relative of these polypeptides to other HCMV proteins (see Figure 1.4) [193].

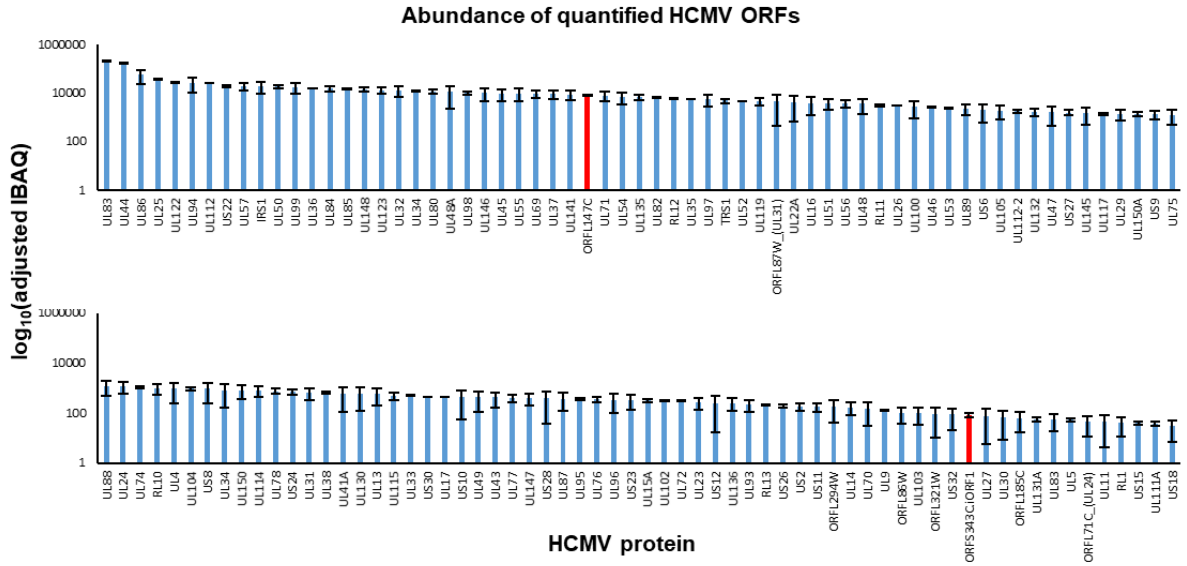


Figure 1.4 | Abundance of canonical and non-canonical HCMV ORFs quantified by proteomics from the whole cell lysate analysis
The intensity-based absolute quantification (IBAQ) method was adapted to generate normalised average IBAQ values for proteins quantified in two analysis of whole cell lysates from HCMV infection at 24, 48 and 72 h PI [162, 194]. Error bars represent the range of values for each protein. ORFL147C was the most abundant non-canonical ORF, and ORFS343C.iORF1 was one of the least abundant non-canonical ORFs (shown in red). Both were included as baits in the interactome. Adapted from Nobre et al (2019) [193].

These cell lines can subsequently be infected with low-passage Merlin strain HCMV. Lysates will then be collected at a singular time-point of infection to guarantee a similar background for the analysis, and protein complexes will be isolated through immunoprecipitation. Proteins that interact with the V5-tagged bait will be identified through mass spectrometry and through the use of the *CompPass* platform [195]. Potential interactors will be scored according to their enrichment, abundance, detection frequency and reproducibility in order to remove contaminants [189].

Thus, project aims were as follows:

- To generate the necessary resources for the assembly of a network of HCMV protein interactions in infected cells;

- To adapt and optimise the AP-MS protocol employed by Huttlin et al (2015);
- To validate a subset of novel findings.

2 | Materials and Methods

Parts of the work presented in this chapter have been published in a similar form in eLife (Nobre et al., 2019). Collaborative work is acknowledged under the header of each section, as well as in the following chapters and appendices.

2.1 Solutions

Complete medium:	Dulbecco's Modified Eagle's Medium (Sigma-Aldrich) with 10 % (v/v) Foetal bovine serum (Sigma-Aldrich), 100 IU/mL penicillin and 0.1 mg/mL streptomycin (Sigma-Aldrich).
Digitonin lysis buffer:	1 % (w/v) Digitonin (Merck Millipore) in Tris-buffered saline (Sigma) with Roche protease inhibitor cocktail (1 tablet/10 mL).
DTT:	A stock solution of 1 M was made by dissolving 1 g of DTT in 6.5 mL of HPLC grade H_2O. The stock solution was diluted to 100 mM with 200 mM HEPES pH 8.5.
Freezing medium:	90 % Foetal bovine serum (Sigma-Aldrich) and 10 % (v/v) dimethyl sulphoxide (Sigma-Aldrich).
Guanidine lysis buffer:	9 mL of 8 M Guanidine Hydrochloride were diluted with 3 mL 200 mM HEPES pH 8.5 to give final 6 M Guanidine / 50 mM HEPES pH 8.5.

HEPES buffer:	8 mL 1 M HEPES stock diluted with 32 mL HPLC-grade H_2O. pH was adjusted to 8.5 with the necessary amount of 1 M NaOH
Hydroxylamine:	A 50% stock solution (Thermo Fisher Scientific) was diluted to 5 % Hydroxylamine in 200 mM HEPES pH 8.5.
IAA:	Iodoacetamide (Sigma-Adrich) was dissolved in 200 mM HEPES pH 8.5 to a final concentration of 500 mM.
LysC:	The contents of a 10 AU vial (Wako) were dissolved in 2 mL of HPLC grade water.
MCLB buffer:	50 mM Tris pH 7.5, 300 mM NaCl, 0.5 % (v/v) NP-40, 1 mM DTT and Roche protease inhibitor cocktail (1 tablet/10 mL).
PBS:	for cell culture was purchased from Sigma-Aldrich.
PBST:	PBS with 0.2 % (v/v) Tween-20 (NBS Biologicals).
Protein Loading Dye (6X):	Tris 375 mM pH 6.8, 12 % (w/v) Sodium dodecyl sulphate (SDS), 30 % (v/v) glycerol, 0.6 M DTT, 0.06 % (w/v) bromophenol blue.
RIPA buffer (1X):	For cell lysis was purchased from CST as a 10X stock and diluted to 1X using distilled H_2O. The 1X working solution

	contains 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM Na_2EDTA, 1 mM EGTA, 1 % (v/v) NP-40, 1 % (v/v) sodium deoxycholate, 2.5 mM sodium pyrophosphate, 1 mM β-glycerophosphate, 1 mM Na_3VO_4, 1 µg/mL leupeptin. One tablet of Roche protease inhibitor cocktail was added to 10 mL of 1X RIPA buffer.
TAE buffer:	For agarose gel electrophoresis was purchased from CIMR media kitchen.
TMT reagent:	Vials containing 0.8 mg of TMT label were re-suspended in 43 µL anhydrous acetonitrile.
Trypsin:	For cell culture was purchased from Sigma-Aldrich. For proteomics, the contents of a 100 µg vial (Thermo Fisher Scientific) were dissolved in 3 mL 200 mM HEPES pH 8.5.
Semi-dry transfer buffer:	10X stock – 360 g Glycine, 75 g Tris Base, 9.25 g SDS, distilled H_2O to 2.5 L; Working solution – 0.1 L 10X stock, 0.2 L Methanol, 0.7 L ddH_2O.

2.2. Cell Culture

2.2.1 Established cell lines

Human foetal foreskin fibroblast cells immortalised with human telomerase reverse transcriptase (HFFF-TERT, male), as described in [196], were kindly provided by Dr. Peter Tomasec (School of Medicine, Cardiff University). HEK293T cells (female) were kindly supplied by Professor Paul Lehner (Department of Medicine, Cambridge University).

2.2.2 Cell culture conditions

Cell lines were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % (v/v) foetal bovine serum (FBS), and 100 IU/mL penicillin / 0.1 mg/mL streptomycin, referred to from here on as complete medium. Cell lines were generally maintained in a 175 cm^2 tissue culture flask (Falcon). All cells were incubated in a static incubator (Binder) at 37 °C and 5 % CO_2.

2.2.3 Cell line passage

Confluent layers of cells were passaged by removing the medium in the tissue culture flask, followed by washing the adherent layer with PBS and incubating cells with 6 mL Trypsin-EDTA (Sigma-Aldrich) until they detached from the surface of the culture vessel. Once the cells were in suspension, 4 mL of new complete medium was added and cells were split between 1:5 and 1:10, into 175 cm^2 tissue culture flasks containing new complete medium.

2.2.4 Cell counting

A 10 µL aliquot of the cell suspension was counted on a bright line haemocytometer (Heinz Herenz) using bright-field microscopy. The total number of cells was calculated using the following formula:

$$\text{Total number cells/ mL} = (\text{Number cells in grid}) \times 10^4$$

2.2.5 Stable cell line generation

To generate lentiviral particles, 1.5×10^5 HEK293T cells were plated in each well of a 12-well plate, 24 h prior to transfection. Cells were then transfected with 500 ng of plasmid DNA combined with 140 ng of a mixture of lentiviral packaging plasmids (VSVG, TAT1B, MGPM2, CMV-Rev1B, combined in a 1:1:1:1 ratio). The recombinant DNA mixture was diluted in 100 µL Opti-MEM Reduced Serum Medium (Thermo Fisher Scientific) and 3 µL TransIT-293 transfection reagent (Mirus) and incubated for 30 min at room temperature (RT). The mixture was then added dropwise to the cells and incubated at 37 °C, 5 % CO_2 for 24 h. After incubation, 90 % of the transfection media was replaced with complete medium. Lentiviral supernatant was harvested 48 h post-transfection and cell debris was removed with a 0.22 µm filter.

HFFF-TERT cells were plated at 1.25×10^5 cells/well in a 12-well plate, 24 h prior to transfection. These cells were then transduced with the lentiviral supernatant for 48 h and then subjected to antibiotic selection with complete medium containing 1 µg/mL Puromycin for two weeks.

2.2.6 Cryopreservation

A cell suspension from a confluent 175 cm^2 tissue culture flask was generated as described in **2.2.3** and centrifuged at 400 x g, RT for 5 min. After discarding the supernatant, the cell pellet was re-suspended in 4 mL of freezing medium (90 % FBS and 10 % (v/v) dimethyl sulphoxide (DMSO). Aliquots of 1 mL were placed in cryovials (Greiner) and stored in a Mr Frosty freezing container (Nalgene). The freezing container was kept at -70 °C for 24 h and then transferred into the liquid nitrogen storage.

Once necessary, cells were recovered from storage, thawed in a 37 °C waterbath (VWR) and transferred into a 50 mL tube containing 9 mL of complete medium. Cells were centrifuged at 400 x g, RT for 5 min. After discarding the supernatant, the pellet was re-suspended in 20 mL of complete medium and transferred into a 175 cm^2 tissue culture flask to be incubated at 37 °C, 5 % CO_2.

2.2.7 Transient transfection

HEK293T cells were plated at 7.5 x 10^5 cells/well in a 6-well plate, 24 h prior to transfection. A total of 2.5 µg plasmid DNA was diluted in 250 µL Opti-MEM Reduced Serum Medium (Thermo Fisher Scientific) and 7.5 µL TransIT-293 transfection reagent (Mirus). The mixture was incubated for 30 min at RT, then added dropwise to the cells and incubated at 37 °C, 5 % CO_2 for 48 h.

2.3 HCMV

2.3.1 Propagation of HCMV strains

All HCMV strains were propagated in HFFF-TERT cells. A list of HCMV strains used in this study can be found in Table 2.3.1. A stock of RCMV 1111 strain was provided by Dr. Pete Tomasec, while stocks of RMCV 2582 and RCMV 2697 were provided by Dr. Richard Stanton (School of Medicine, Cardiff University). Whole-genome consensus sequences of passage 2 of all recombinant viruses were derived using the Illumina platform as described in [190].

For each viral stock, cells were cultured in five to ten 175 cm^2 tissue culture flasks until they reached 80 % confluency. Cells were then infected with an aliquot of a low-passage stock of the HCMV strain to be propagated, at a multiplicity of infection (MOI) 0.0025 in 10 mL of serum-free DMEM. Infection was carried out for 2 h, at 37 °C, 5 % CO_2 in an incubator containing a rocker platform (Stuart). At the end of the incubation period, the serum-free DMEM was replaced with complete medium. Cells were monitored for cytopathic effect (CPE) and complete medium was replaced as required. Once 60 % CPE was observed, media was removed for HCMV purification (as described below in **2.3.2**). Complete media was kept and replaced every two days, until the majority of cells had been lysed.

Table 2.3.1 | List of HCMV used in this study

Laboratory designation	Description
RCMV 1111	Merlin BAC derived virus (Stanton et al, 2010)
RCMV 2582	HCMV expressing rGFP from a P2A self-cleaving peptide at the 3'-end of the UL36 coding region generated by recombineering the strain Merlin BAC (RCMV 1111) [193].
RCMV 2697	ORFL147C mutant generated by recombineering RCMV 2582. Substitutions were introduced into three in-frame ATG codons at or near the 5'-end of ORFL147C, in such a way that the coding potential of UL56, with which ORFL147C overlaps extensively in another reading frame, was unaffected [193].

2.3.2 Purification of viral stocks

Immediately after harvest, media from HCMV infections was centrifuged in sterile 250 mL Sorvall bottles (Sorvall) at 10,000 rpm, 35 °C for 2 h in an Avanti JXN-26 centrifuge (Beckman Coulter), using a pre-warmed JLA-16.250 rotor. After discarding the supernatant, the pellet was re-suspended in 1 mL complete medium pre-warmed at 37 °C, aspirated through a 21 gauge needle to disrupt pellet aggregates and stored at -80 °C. Once the infection of the propagation culture was complete, all 1 mL aliquots from harvests of the same propagation were thawed, pooled and spun at 10,000 rpm, RT for 5 min in a benchtop microcentrifuge (Eppendorf) to remove cell debris. The pooled supernatant was dispensed in 100 µL aliquots and kept at -80 °C until required. Titres of HCMV stocks were calculated by titration as described below in **2.3.3**.

2.3.3 Viral titration

HFFF-TERT cells were seeded at $1.5x10^5$ cells/well in a 12-well plate and incubated at 37 °C, 5 % CO_2. The following day, cells were infected with 4-fold serial dilutions (1:16, 1:64, 1:256, 1:1024, 1:4096, 1:16348; in duplicate) in serum-free DMEM in a total of 300 µL/well. Infections were carried out for 2 h, at 37 °C, 5 % CO_2, in an incubator containing a rocker platform. After the incubation period, serum-free DMEM was replaced by complete medium and incubated for 24 h. Cells were processed as described in **2.7** and analysed by Flow Cytometry. The values for each dilution's duplicates were averaged to give infectious titres as either IE1 or GFP-forming units/mL.

2.3.4 Infection for assays

For each assay, HFFF-TERT cells were seeded approximately at $4x10^4$ cells/ cm^2, 24 h prior to infection. The culture media was replaced with virus inocula (in serum-free DMEM), in a minimal volume of media according to the surface area of the tissue culture vessel. Cells were incubated with viral inoculums for 2 h, at 37 °C, 5 % CO_2, in an incubator containing a rocker platform. Post-incubation, the inoculum was replaced with warm complete medium and cells were cultured for the necessary amount of time according to the assay's purpose.

For a subset of experiments, cells were incubated with dexamethasone in serum-free medium 24 h prior to viral infection. This alternative method is explained in more detail in the relevant sections **2.9.1.1** and **2.10**.

2.3.5 Viral growth curve

$1x10^6$ HFFF-TERT cells were seeded into 25 cm^2 tissue culture flasks 24 h prior to infection. The following day, cells were infected with HCMV at MOI 1, as described in **2.3.4**. Starting at day 2 and then every subsequent two days, the whole supernatant from the culture was taken and replaced with new complete medium of equal volume. The supernatant was centrifuged at 400 x g, RT for 10 min to pellet any cells in suspension and frozen in 1 mL aliquots at -80 °C. Aliquots were then titred in HFFF-TERT cells, as described in **2.3.3**, to estimate the amount of virus present in the supernatant.

2.4. Molecular Biology

2.4.1 Polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify DNA fragments employing the PfuUltra II Fusion HS DNA Polymerase (Agilent) in the manufacturer's buffer (containing Magnesium to a final concentration of 2 mM), 50 mM dNTP mix and 1 µM of each primer (designed according to template and ordered from Sigma-Aldrich). Thermocycling conditions were as follows:

1. 95 °C for 2 min
2. 95 °C for 20 sec
3. Lowest primer melting temperature -5 °C for 20 sec
4. 72 °C for 15 sec/kilobase pair (kbp)
5. 72 °C for 3 min

   Steps 2 to 4 were repeated for 25 to 30 cycles depending on template type.

2.4.2 Agarose gel electrophoresis

Electrophoresis was performed to separate DNA fragments based on size using a 0.7 % (w/v) TAE agarose gel containing SYBR Safe (1:10,000). DNA samples were mixed with 6X Gel Loading Dye (NEB) prior to loading into the wells of the pre-casted agarose gel. Molecular weight-size markers (1 kbp DNA ladder or 100 bp DNA ladder, NEB) were run in parallel to estimate sample DNA fragment sizes. Gels were run at 70-90 V for 30-45 min, before visualisation with an UV transiluminator.

2.4.3 DNA purification from an agarose gel

Following agarose gel electrophoresis, selected DNA bands were excised from the gel using a scalpel and transferred to a 1.5 mL tube. DNA was then purified using the QIAquick Gel Extraction Kit (QIAGEN). Gel slices were dissolved in three times their weight of QG buffer (100 mg ~ 100 µL) at 50 °C for 10 min. One gel volume of isopropanol was added to the sample once the agarose was fully dissolved. The mixture was transferred to a QIAquick spin column placed in a collection tube, 800 µL at a time, and centrifuged for 1 min at 13,000 rpm. After discarding the flow-through, the column was washed with 0.75 mL PE buffer and centrifuged for 1 min at 13,000 rpm. The flow-through was once again discarded and the column centrifuged for an additional 1 min to remove residual buffer. A QIAquick column was placed in a new 1.5 mL tube and allowed to incubate for 1 min with 50 µL EB buffer before centrifugation for 1 min at 13,000 rpm to elute the DNA. Purified DNA was kept at -20 °C until necessary.

2.4.4 Gateway cloning

Gateway cloning (see Figure 2.4.1) allows transfer of selected DNA sequences in between vectors using lambda recombination, which is catalysed by a mixture of the integrase (Int) and excisionase (Xis) enzymes from bacteriophage λ in combination with the *E. coli* Integration host factor protein (IHF). These enzymes bind to unique sites (*att*), bring the target sites together, cleave them and covalently attach the DNA.

Figure removed for copyright reasons. Copyright holder is INVITROGEN.

Figure 2.4.1 | Gateway cloning schematic
A gene of interest is cloned into a donor vector (pDONR223) in a BP reaction generating an entry clone. This step is catalysed by the BP Clonase enzyme mix (Integrase from bacteriopaghe λ and *E. coli* Integration host factor protein). The transgene is then transferred into an expression clone in a LR reaction, catalysed by the LR Clonase enzyme mix (Integrase and Excisionase from bacteriopaghe λ in addition to the *E. coli* Integration host factor protein). Adapted from manufacturer's manual.

PCR products or double-stranded DNA fragments containing flanking attB sequences were cloned into the pDONR223 entry vector (see Figure 2.4.2), then the lentiviral destination vector pHAGE-pSFFV (see Figure 2.4.3), using Gateway cloning (Thermo Scientific).

The BP reaction was performed using 0.4 µL BP clonase II enzyme mix (containing the enzyme in its reaction buffer), 0.2 µL pDONR223 (150 ng/µL), 0.5 µL attB insert DNA (15-150 ng), 0.9 µL Tris-EDTA buffer (TE) pH 8.0 and incubated overnight (O/N) at RT.

The LR reaction was performed using 0.4 µL LR clonase, 0.2 µL pHAGE-pSFFV (150 ng/µL), 0.4 µL TE buffer pH 8.0, 1 µL entry clone (generated by the BP reaction) diluted 1:5 in TE and incubated overnight (O/N) at RT.

Following O/N incubation, both BP and LR reactions were incubated with 0.3 µL Proteinase K for 10 min at 37 °C, then incubated for 1 min on ice, before transformation into competent bacterial cells.

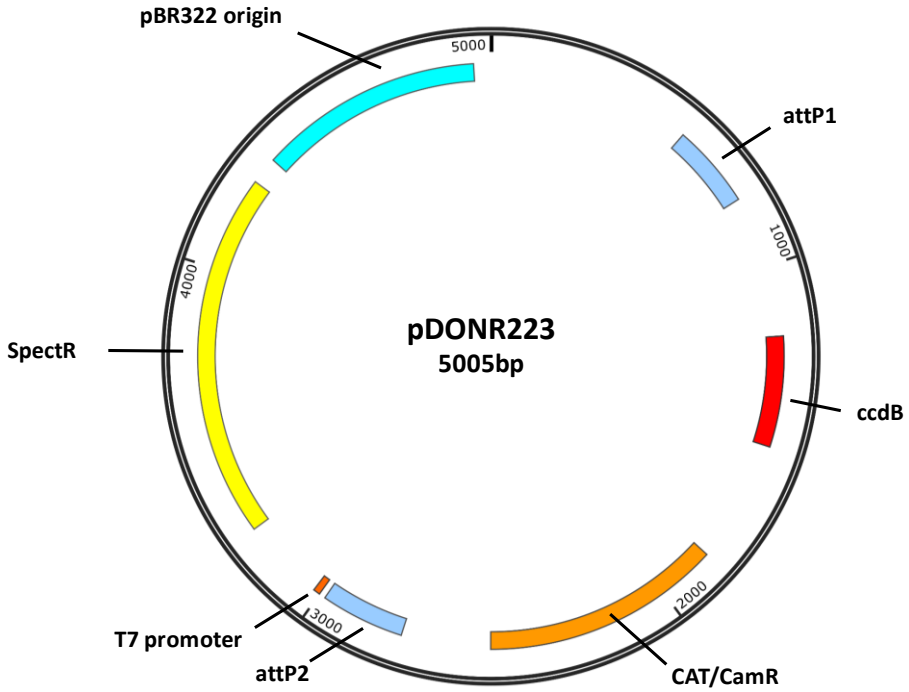


Figure 2.4.2 | pDONR223 vector map
Gateway Entry vector containing a T7 promoter, Spectinomycin resistance marker and a pBR322 origin of replication. The Gateway attP1/attP2 sites allow the replacement of the ccdB toxin coding sequence and the chloramphenicol acetyltransferase resistance marker (CAT/CamR), with an insert containing gateway attB1/attB2 flanking sequences, via a BP reaction.

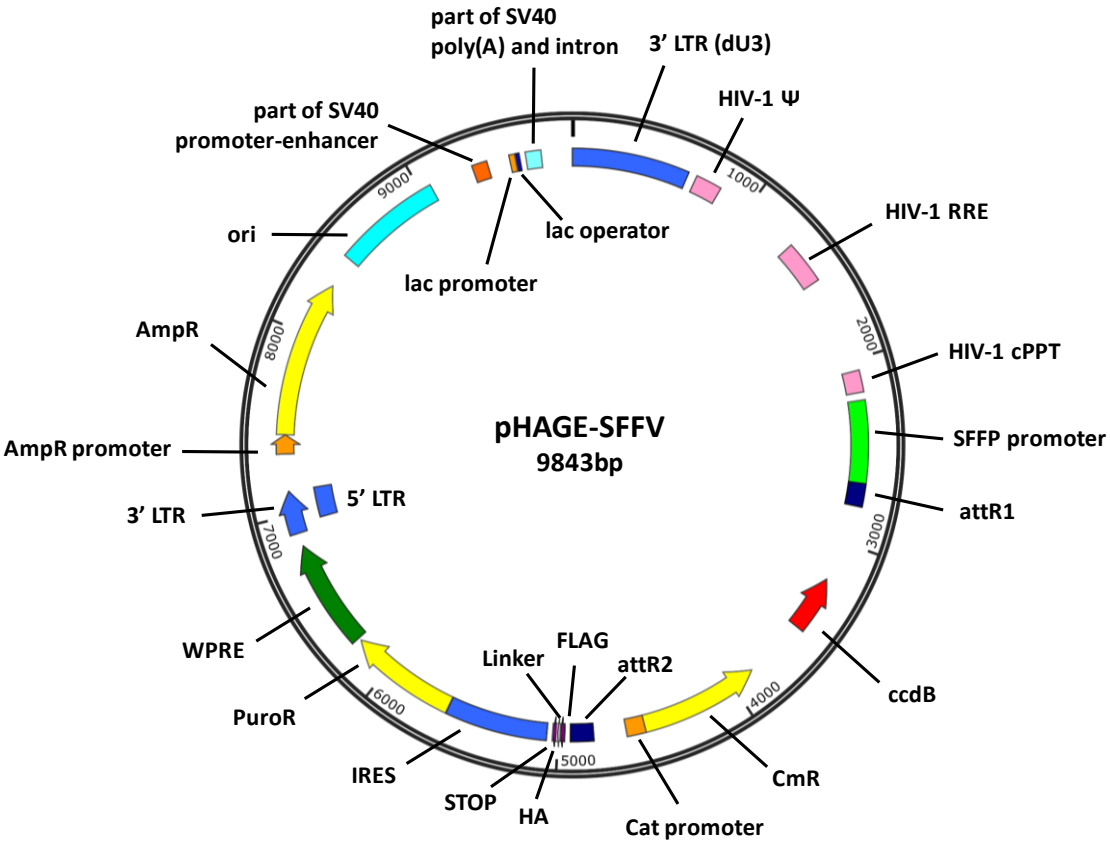


Figure 2.4.3 | pHAGE-SFFV vector map
Gateway Destination vector containing the lentiviral HIV-1 Ψ, RRE, cPPT and LTR sequences. The Gateway attR1/attR2 sites allow the replacement of the ccdB toxin coding sequence as well as the chloramphenicol acetyltransferase resistance marker (CAT/CamR) and its promoter, with an insert containing gateway attL1/attL2 flanking sequences, via an LR reaction. Expression of the insert is driven by the SFFV promoter and would contain the FLAG and HA sequences (with a short linker in between) unless a stop codon were placed at the 3' end of the insert. Other features include an internal ribosomal entry site (IRES), the Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (WPRE), an origin of replication (ORI), lac promoter and operator, SV40 promoter-enhancer sequences and ampicillin and puromycin antibiotic selection markers.

2.4.5 Bacterial transformation

Aliquots of Gold Alpha-Select Chemically Competent Cells (Bioline) were thawed on ice and 20 µL were mixed into each cloning reaction. The mixture was incubated on ice for 20 min, followed by heat-shock at 42 °C for 45 sec in a waterbath (Grant). After incubating for 1 min on ice, 200 µL Luria-Bertani (LB) medium were added and the reaction was incubated for 1 h at 37 °C, 350 rpm. The mixture was then plated on a

LB/Agar plate containing an appropriate antibiotic for selection of bacterial transformants.

2.4.6 Small-scale plasmid DNA preparation

Small-scale isolation of plasmid DNA from bacterial cultures was performed using a QIAprep Spin Miniprep kit (QIAGEN) and the proprietary reagents therein. Bacterial transformants were grown O/N at 37 °C, shaking at 350 rpm in 4 mL LB containing antibiotic for selective pressure. From this culture, 3 mL were pelleted by centrifugation at 8,000 rpm, RT for 3 min. The pellet was re-suspended in 250 µL buffer P1 and transferred to a 1.5 mL tube. The cells were lysed by adding 250 µL of buffer P2 and incubating for 5 min at RT. The reaction was neutralised with 350 µL buffer N3, mixed several times by inversion and centrifuged at 13,000 rpm, RT for 10 min. The supernatant was transferred to a QIAprep spin column placed in a collection tube and centrifuged at 13,000 rpm, RT for 1 min. After discarding the flow-through, the column was washed with 0.5 mL PB buffer and centrifuged at 13,000 rpm, RT for 1 min. The flow-through was discarded again and the column was washed with 0.75 mL PE buffer followed by centrifugation at 13,000 rpm, RT for 1 min. After discarding the flow-through, the column was centrifuged at 13,000 rpm, RT for 1 min to remove residual buffer. The column was transferred to a new 1.5 mL tube and incubated with 50 µL EB buffer for 1 min, before DNA elution by centrifugation at 13,000 rpm, RT for 1 min.

2.4.7 Large-scale plasmid DNA preparation

For transient transfection, large-scale preparation of plasmid DNA was required. This was performed using the QIAgen Plasmid MAXI kit (QIAGEN). For this procedure, a colony of bacterial transformants was used to inoculate 250 mL LB and incubated O/N at 37 °C, shaking at 350 rpm. Bacterial cells were pelleted by centrifugation at 6000 x g at 4 °C for 15 min using 250 mL Sorvall centrifuge bottles (Sorvall) and SLA1500 rotor. The pellet was re-suspended in 10 mL P1 buffer and cells were lysed by adding 10 mL P2 buffer, mixing by inversion and incubating at RT for 5 min. The reaction was neutralised with 10 mL of ice-cold buffer P3, mixed by inversion, poured into the barrel of a QIAfilter Cartridge containing a screw cap on the outlet nozzle and incubated for 10 min, RT. During this incubation, a QIAGEN-tip was equilibrated with 10 mL QBT buffer, allowing the column to empty by gravity flow. At the end of the incubation, the screw cap was removed from the QIAfilter Cartridge and a plunger was inserted, allowing the cell lysate to be filtered by gravity flow into the previously equilibrated QIAGEN-tip. The tip was washed twice with 30 mL QC buffer and DNA was eluted into a 50 mL tube with 15 mL QF buffer. DNA was precipitated with 10.5 mL RT isopropanol and the solution was mixed and centrifuged at 15,000 x g, 4 °C for 30 min. The supernatant was carefully removed and the DNA pellet washed with 5 mL RT 70 % (v/v) ethanol, followed by centrifugation at 15,000 x g for 10 min. After decanting the supernatant, the pellet was air-dried for 10 min. DNA was re-suspended in 100 µL TE buffer.

2.4.8 Nucleic acid sample quantification

Concentration of nucleic acids for DNA and RNA samples was determined by reading absorbance at 260 nm using a NanoDrop Spectrophotometer (Fisher Scientific).

Sample purity was estimated by calculating a ratio between the reading at 260, 230 and 280 nm.

2.4.9 Glycerol stock generation

Glycerol stocks for selected plasmids were made by mixing equal volumes of an O/N bacterial culture and 80 % (v/v) glycerol (sterile) in a Nunc CryoTube (Merck). Glycerol stocks were stored at -70 °C.

2.5 Construction of the expression vector library

A library of recombinant adenovirus expression vectors containing coding sequences for a large subset of HCMV proteins was generated by Dr. James Davies and Dr. Sepehr Seirafian (School of Medicine, Cardiff University). Cloning of CUL4A-HA, TRIM22-HA, MBNL1-HA, CELF1-HA as well as site-directed mutagenesis to generate the UL25-N625-V5 and NCK1-HA mutants were performed by Dr. Katie Nightingale (Department of Medicine, University of Cambridge).

2.5.1 Genes cloned from recombinant adenoviral vector library

The majority of HCMV genes and the GFP control were cloned from a library of recombinant adenoviral vectors (rAdv), which was kindly provided by Professor Gavin Wilkinson's group (School of Medicine, Cardiff University). Primers were designed to recognise the 3' end of the CMV promoter encoded within the rAdv ('rAdv fwd') and the 3' end of the V5 tag ('rAdv rev'). Both primers were flanked by four guanine residues, followed by Gateway attB1 (forward) or attB2 (reverse) sequences (highlighted in blue in the sequences of rAdv forward/reverse primers in Table 2.5.1).

Table 2.5.1 | Primer sequences for cloning and plasmid DNA sequencing

Primer	Sequence (5'-3')	PCR template
rAdv fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTGAAGACACCGGGACCGATC	rAdv library
rAdv rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTACGTAGAATCAAGACCTAGGAGC	rAdv library
No bait oligo 1	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCCAGGCGAGAACGTGTGCGTGGACAAGCGAGCAGCATACGAACCCAGCTTTCTTGTACAAAGTGGTCCCC	-
No bait oligo 2	GGGGACCACTTTGTACAAGAAAGCTGGGTTCGTATGCTGCTCGCTTGTCCACGCACACGTTCTCGCCTGGGAGCCTGCTTTTTTGTACAAACTTGTCCCC	-
US9 fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGATCCTGTGGTCCCCG	HCMV Merlin BAC
US9 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCTATCGTCTTTAGCCTCTTCTTCC	HCMV Merlin BAC
UL20 fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGCTCGGGATACGGGCTATG	HCMV Merlin BAC
UL20 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCTTGCTGGCATGCAGACCACC	HCMV Merlin BAC
UL48-H1 fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGAAAGTCACACAGGCCAGCTGCC	HCMV Merlin BAC
UL48-H1 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCTGGGCAGGGCTCCCTCGTTGG	HCMV Merlin BAC
UL48-H2 fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGAACCCACGGAAACCGCAGGC	HCMV Merlin BAC
UL48-H2 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCTCAAAAGATAGAGAAACCGCATGTGTTG	HCMV Merlin BAC
UL80.5 fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTCGCACCCTCTGAGTGC	HCMV Merlin BAC
UL80.5 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCTCTCGAGCTTATTGAGCGCAGC	HCMV Merlin BAC
UL112 fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGATCTGCCTACTACCG	cDNA of HFFFs infected with HCMV (72h)
UL112 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCTATCGTCGAAAAACGCCGCG	cDNA of HFFFs infected with HCMV (72h)
UL132 fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGCCGGCCCCGCGGGGTC	HCMV Merlin BAC
UL132 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCTGTCGTACTCGGGATCTCTGAGCGAG	HCMV Merlin BAC
RL8A fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGCCTCACGGCCATCTCC	HCMV Merlin BAC
RL8A rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCTGCTAAAAACAGCGGACAGTCC	HCMV Merlin BAC
RL9A fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTCTCTAGATGCCGCCAGC	HCMV Merlin BAC
RL9A rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCTGAGAAACAGCACGTAGGTCAGG	HCMV Merlin BAC
IRS1 fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCCCAGCGCAACGGC	HCMV Merlin BAC
IRS1 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCTATGATGAACGTGGTGAGGGGCG	HCMV Merlin BAC
M13 fwd	GTAAAACGACGGCCAG	-
M13 rev	CAGGAAACAGCTATGAC	-
SFFVp seq	CGCGCCAGTCCTCCGATTG	-

Templates were amplified by PCR using the generic primers above and the product was inserted into pDONR223, then pHAGE-pSFFV using Gateway Cloning as described in **2.4.4**. The resulting lentiviral template expresses a C-terminally V5-tagged gene under the control of a SFFV promoter, with the six base-pair linker region 'AGCGCT' between the 3' end of the gene and the tag.

For UL48 it was not possible to generate a cell line expressing the full-length construct, likely due to inefficient transduction caused by the length of the transgene (6.7 kbp). To enable cloning of UL48 in two segments without splitting protein secondary structures, the coding sequence was analysed using the YASPIN Secondary Structure Prediction tool from the Centre for Integrative Bioinformatics VU (Vrije Universiteit Amsterdam) [197] (see Appendix A). The gene was divided into two segments, one of 4.5 kbp (1-1504 aa) terminating in a stop codon (UL48-H1), and one of 2.2 kbp (1505-2241 aa), with an additional start codon (UL48-H2). Both segments were cloned as described above (primers provided in Table 2.5.1) and stably expressed in separate cell lines.

2.5.2 Genes cloned from HCMV Merlin BAC or cDNA template

For HCMV genes amplified from the HCMV Merlin BAC, primers were designed to recognize the 5' end (forward) and the 3' end (excluding the stop codon, reverse) of each gene. In addition to the gene specific sequence, the reverse primer also contained the six base-pair linker region 'AGCGCT', followed by the coding sequence for the V5 tag and a stop codon. Both primers were flanked by four guanine residues, followed by Gateway attB1 (forward) or attB2 (reverse) sequences. Cloning methods and structure of the resulting lentiviral vector were as described in **2.5.1**. Primers for gene cloning from the HCMV Merlin BAC are provided in Table 2.5.1.

HCMV genes cloned from cDNA generated from HCMV infected primary HFFF cells were amplified with primers designed similarly to primers used for cloning from the Merlin BAC. Primer sequences for these genes are provided in Table 2.5.1. A diagram of the primer design for HCMV genes cloned from the Merlin BAC or cDNA from HCMV infected primary HFFF cells is shown in Figure 2.5.2.

US9 coding sequence 3': 5′-ATGATCCTGTGGTCCCCG-3′
US9 coding sequence 5' (reverse complement): 5‘-ATCGTCTTTAGCCTCTTCTTCC-3′

attB1: 5′-ACAAGTTTGTACAAAAAAGCAGGCT-3′
Reverse complement of attB2: 5′- ACCACTTTGTACAAGAAAGCTGGGT-3′

Linker: 5′-AGCGCT-3′
V5 tag: 5′-GGTAAGCCAATCCCTAACCCGCTCCTAGGTCTTGATTCTACGTAA-3′
Stop codon: 5′-TAA-3′

Reverse complement of Linker + V5 tag + Stop codon:
5'-TTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCT-3'

US9 Primer fwd: 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGATCCTGTGGTCCCCG-3′
US9 Primer rev: 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGC GGGTTAGGGATTGGCTTACCAGCGCTATCGTCTTTAGCCTCTTCTTCC-3′

Figure 2.5.2 | Diagram of primer design for HCMV genes cloned from the Merlin BAC or cDNA from HCMV infected primary HFFF cells

2.5.3 Synthesized genes

The coding sequences of the canonical HCMV genes UL55, UL56, UL128, UL131A, UL150A and non-canonical ORFL147C and ORFS343C were synthesized as double-stranded DNA fragments (gBlocks®, Integrated DNA Technologies). Each fragment comprised the viral gene (without a stop codon), succeeded by the six base-pair linker region 'AGCGCT', the coding sequence for the V5 tag then the stop codon. The fragments had flanking Gateway attB sequences allowing cloning into pDONR223 as described in **2.4.4**.

N-terminally V5-tagged UL42 was synthesized as double-stranded DNA fragment (gBlocks®, Integrated DNA Technologies). The fragment comprised a start codon succeeded by the coding sequence for the V5 tag, followed by the six base-pair linker sequence 'AGCGCT' and the viral gene (including its start and stop codon). The fragment had flanking Gateway attB sequences allowing cloning into pDONR223 using the gateway cloning method.

2.5.4 pHAGE-pSFFV vector control

To generate an empty vector control, two complementary oligonucleotides ('No bait' oligonucleotides 1 and 2, Table 2.5.1) were designed to code for a random sequence of 14 aa, which featured neither a Methionine nor the V5 tag, with flanking Gateway attB sequences and four guanosine residues allowing cloning into pDONR223. To produce a double-stranded insert, the oligonucleotides were annealed at 95 °C for 10 min. The reaction was cooled down to 25 °C at approximately 5 °C/ min using a Thermomixer (Eppendorf). The solution was diluted 1:10 for the BP reaction, for a final concentration of 1 µM.

2.5.5 Cloning of human genes

The coding sequences for NEDD4L, CUL4A, TRIM22, NCK1, MBNL1 and CELF1 were amplified from plasmids containing the template, using primers designed to recognize the 5' end (forward) and the 3' end (excluding the stop codon, reverse) of the gene (see Table 2.5.5). The reverse primer also contained the fifteen base-pair linker region 'TCGGCCGCTGGAGGA', followed by the coding sequence for the HA tag and a stop codon. Both primers were flanked by four guanine residues, followed by Gateway attB1 (forward) or attB2 (reverse) sequences allowing cloning into pDONR223 and

subsequently pHAGE-SFFV, as described in **2.4.4.** The same strategy was used to clone TRIM22, but instead of amplifying the template from a plasmid, cDNA of HFFF-TERT cells stimulated with IFNα2a (1000 U for 24 h) was used.

The coding sequences of CNOT2 and LRFN3 were cloned from a gateway entry vector (see Table 2.5.5) into pHAGE-SFFV by performing an LR reaction. Lentiviral plasmids containing the coding sequences of CNOT7 and NEDD4 were purchased from Harvard's PlasmID or Addgene collections. Lentiviral expression vectors were then used for transient transfection of HEK293T cells.

Table 2.5.5 | Templates and primers for human gene cloning

Gene name	Forward primer (5'-3')	Reverse primer (5'-3')	Template	Source
CNOT2	–	–	pDONR223-CNOT2	Harvard PlasmID HsCD00080019
CNOT7	–	–	pHAGE-CNOT7	Harvard PlasmID HsCD00453329
NEDD4	–	–	PHAGE-P-CMVt-N-HA Nedd4 wt	Addgene 24124
NEDD4L	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCGACCGGGCTCGGGG	GGGGACCACTTTGTACAAGAAAGCTGGGTCTAGGCGTAGTCGGGCACGTCGTAGGGGTATCCTCCAGCGGCCGACACCCCTTCAAATCCTTGAGCATTTTCCACG	pCMV-SPORT6-NEDD4L	Harvard PlasmID HsCD00337956
LRFN3	–	–	pDONR221-LRFN3	Harvard PlasmID HsCD00041564
CUL4A	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCGGACGAGGCCCCG	GGGGACCACTTTGTACAAGAAAGCTGGGTTTAAGCGTAATCTGGAACATCGTATGGGTAAGCGCTGGCCACGTAGTGGTACTGATTC	pOTB7-CUL4A	Harvard PlasmID HsCD00325140
TRIM22	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGATTTCTCAGTAAAGGTAGACA	GGGGACCACTTTGTACAAGAAAGCTGGGTAGGAGCTCGGTGGGCACACAGTCATG	cDNA of HFFF-TERT cells stimulated with IFNα2a	–
MBNL1	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCTGTTAGTGTCACACC	GGGGACCACTTTGTACAAGAAAGCTGGGTTTAAGCGTAATCTGGAACATCGTATGGGTAAGCGCTCTACATCTGGGTAACATACTTG	pDONR221-MBLN1	Harvard PlasmID HsCD00079833
CELF1	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGAACGGCACCCTGGACCAC	GGGGACCACTTTGTACAAGAAAGCTGGGTTTAAGCGTAATCTGGAACATCGTATGGGTAAGCGCTTCAGTAGGGCTTGCTGTCATTC	pDONR221-CUGBP1	Harvard PlasmID HsCD00039403
NCK1	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCAGAAGAAGTGGTGGTAG	GGGGACCACTTTGTACAAGAAAGCTGGGTTTAAGCGTAATCTGGAACATCGTATGGGTAAGCGCTTGATAAATGCTTGACAAGATATAATTTTTC	pENTR223-NCK1	Harvard PlasmID HsCD00370605
NCK1 SH3-1 W38K	GGAATTTCGAACTCGCCACTTGGACTTAGAATCATCCAGA	TCTGGATGATTCTAAGTCCAAGTGGCGAGTTCGAAATTCC	pHAGE-NCK1-HA	–
NCK1 SH3-2 W143K	TTGTAGCTACCACGCCACTTCCCATCACTGCATTTCTC	GAGAAATGCAGTGATGGGAAGTGGCGTGGTAGCTACAA	pHAGE-NCK1-HA	–
NCK1 SH3-3 W229K	GATCTTCCTGCATTTCCACTTCTCTGGGTCATTTTCAGGT	ACCTGAAAATGACCCAGAGAAGTGGAAATGCAGGAAGATC	pHAGE-NCK1-HA	–

2.5.6 Site-directed mutagenesis

Site-directed mutagenesis by overlap extension PCR was used to generate point mutations in the coding sequence of NCK1. Primer sequences spanning the target region and incorporating the desired sequence changes in both forward and reverse orientations were generated (see Table 2.5.5). These, along with primers that would anneal at the 5' and 3' ends of the full-length NCK1 coding sequence ('NCK1 forward primer' and 'NCK1 reverse primer', respectively) were used to amplify two fragments of NCK1, each incorporating the point mutation, from the pHAGE-NCK1-HA template plasmid generated as described in section **2.5.5**. Fragments were purified using the QIAquick Gel Extraction Kit (QIAGEN, as described in **2.4.3**), and assembled into a full-length mutant NCK1 coding sequence by a second round of PCR using only the 'NCK1 forward primer', 'NCK1 reverse primer' and both NCK1 fragments as a template in the same reaction. The product was then purified using the QIAquick Gel Extraction Kit (QIAGEN, as described in **2.4.3**) and subcloned into the gateway system as described above.

To generate a truncation mutant of UL25, a single round of PCR was performed using a forward primer designed to recognize the 5' end of the gene, and a reverse primer designed to recognise the 3' end of the gene (bases 1856-1875, truncating the amino acid sequence at residue 625). In addition to the gene specific sequence, the reverse primer also contained the six base-pair linker region 'AGCGCT', followed by the coding sequence for the V5 tag and a stop codon. Both primers were flanked by four guanine residues, followed by Gateway attB1 (forward) or attB2 (reverse) sequences (see Figure 2.5.6).

UL25 coding sequence 3': 5′- ATGTCGTCGCGGCGTCGCAG-3′
UL25 coding sequence 5' (reverse complement): 5‘-GGATCGGCAGATTTGTTCGG-3′

attB1: 5′-ACAAGTTTGTACAAAAAAGCAGGCT-3′
Reverse complement of attB2: 5′- ACCACTTTGTACAAGAAAGCTGGGT-3′

Linker: 5′-AGCGCT-3′
V5 tag: 5′-GGTAAGCCAATCCCTAACCCGCTCCTAGGTCTTGATTCTACGTAA-3′
Stop codon: 5′-TAA-3′

Reverse complement of Linker + V5 tag + Stop codon:
5'-TTATTACGTAGAATCAAGACCTAGGAGCGGGTTAGGGATTGGCTTACCAGCGCT-3'

UL25 N625 Primer forward:
5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTATGATCCTGTGGTCCCCG-3′

UL25 N625 V5 Primer reverse:
5′-GGGGACCACTTTGTACAAGAAAGCTGGGTTTATTACGTAGAATCAAGACCTAGGAGC
GGGTTAGGGATTGGCTTACCAGCGCTGGATCGGCAGATTTGTTCGG-3′

Figure 2.5.6 | Diagram of primer design for UL25 N-terminal truncation at residue 625

2.5.7 Construct sequencing

Sequencing of all genes was conducted in the pDONR223 vector using 'M13 fwd', 'M13 rev' (see Table 2.5.1) and gene-specific primers as necessary to obtain a complete sequencing coverage. All pHAGE-pSFFV vectors underwent sequencing of the first ~700 bases from the 3' end of the SFFV promoter, using a primer that recognises the 3' end of the SFFV promoter ('SFFVp seq', Table 2.5.1), to verify that the viral construct had correctly recombined.

2.6. RT-QPCR

2.6.1 Cellular RNA Extraction

Total RNA from a subset of stable cell lines expressing viral transgenes was extracted using an RNeasy Mini Kit (QIAGEN). For each cell line, $3x10^5$ HFFF-TERT cells/well were seeded into a 6-well plate. The following day, cells were washed with PBS then disrupted with 350 µL RLT buffer/1 M DTT. The lysate was scraped, transferred into a 1.5 mL tube and homogenised by vortexing. An equal volume of 70 % (v/v) ethanol was added to the lysate, mixed by pipetting and transferred into an RNeasy spin column placed in a collection tube. The column was centrifuged at 8,000 x g, RT for 30 sec. The flow-through was discarded and 700 µl Buffer RW1 was added to the RNeasy spin column, followed by centrifugation at 8,000 x g, RT for 30 sec. After discarding the flow-through, the spin column was washed twice with 500 µl Buffer RPE, with centrifugation at 8,000 x g, RT for 30 sec after the first wash, then 2 min after the second wash. The spin column was placed in a new 1.5 mL collection tube and 50 µl RNase-free water was added directly to the spin column membrane. RNA was eluted by centrifugation at 8,000 x g, for 1 min.

2.6.2 DNase treatment

To remove residual contaminating DNA from RNA extraction samples, 1 µg total RNA was treated with 2 enzyme units (U) of Turbo DNase (Thermo Fisher Scientific) in the manufacturer's buffer for 30 min at 37 °C, for a 10 µL reaction. The reaction was quenched with 1 µL of manufacturer's DNase inactivation reagent for 5 min at RT, then centrifuged at 10.000 x g for 1.5 min and transferred to a new tube.

2.6.3 Reverse-transcription reaction

Synthesis of cDNA was performed using the GoScript Reverse Transcription system (Promega). In a 1.5 mL tube, 2 µL of the DNase-treated RNA template was incubated with 1 µL Oligo $(dT)_{15}$ primers and 2 µL Nuclease-free water for 5 min at 70 °C, then for 5 min on ice-cold water. A 15 µL aliquot of the reverse transcription reaction mix containing 1 µL GoScript Reverse Transcriptase, 2 µL 25 mM $MgCl_2$, 1 µL dNTPs (final concentration 0.5 mM each dNTP), 0.5 µL 40 U/µL RNase inhibitor, 4 µL 5X GoScript Reaction Buffer and 6.5 µL Nuclease-free water was then added to the tube containing the primers and template. The reaction was then incubated for 5 min at 25 °C to allow primer annealing to the RNA template, then 1 h at 42 °C for extension. The reverse transcriptase was inactivated by 15 min incubation at 70 °C.

2.6.4. qPCR

Quantitative PCR was used to verify viral bait expression in HFFF-TERT cell lines where these could not be detected by immunoblotting or mass spectrometry. All assays employed the Fast SYBR Green Master Mix (Applied Biosystems) and 7500 Fast & 7500 Real-Time PCR Systems (Applied Biosystems). Primers targeting HCMV genes or GAPDH (as an internal control) are shown in the Table 2.6.4. Templates prepared as described in **6.3** were diluted 1:10 with nuclease-free water and 2 µL loaded onto a MicroAmp Optical 96-well reaction plate (Applied Biosystems). An 18 µL aliquot of reaction mix containing 10 µL 2X Fast SYBR Green Master Mix, 1 µL of each primer (10 µM) and 6 µL nuclease-free water was added to each well. All samples were assayed in duplicate and RNA from each cell line as well as nuclease-free water were used as negative controls. The plate was sealed with MicroAmp Optical Adhesive film (Applied Biosystems) and then centrifuged

at 400 x g for 2 min, 10 °C. The PCR program followed the Applied Biosystems default settings, starting with activation at 95 °C for 2 min, followed by 40 cycles of denaturation at 95 °C for 5 s and annealing/extension at 60 °C for 30 s. The amplification products were then separated by agarose gel electrophoresis, purified from gel slices and sequenced to confirm viral bait expression. For UL136, UL146 and UL148D, this procedure failed to generate sequencing-quality amplicons and therefore generation of whole gene amplicons (primers listed in Table 2.6.4) by PCR (as described in **4.1**) was attempted instead.

Table 2.6.4 | Primers used for validation of transgene expression

Gene name	Short amplicon / qPCR primers	Extended amplicon / PCR primers
US13	FWD: GACCGTGTTCCATCATCAGC REV: CCTACGTAGATCAGCAGCGA	-
US18	FWD: CTGATCAACACCGGCATCAC REV: GCCAATCCCGATGACTTTCC	-
US33A	FWD: GGGTTACGAGAAACTGGGATAC REV: AACGGAAAAGTGAACGGCAA	-
UL1	FWD: CACTCACGTTGGTTGGACAG REV: TGTGTGGTTGACGTTGTTTCT	-
UL2	FWD: GAAGATGACGACGCATATCCG REV: CGTCTCGAAGCAGCTTGAG	-
UL6	FWD: GTTCCTCGAGCGGTTCAAAG REV: CAAACGCGCTGGAATTTGTC	-
UL9	FWD: TGAAGCACAACACGACACTG REV: CGCTGTATCGGTATTGTGTGA	-
UL11	FWD: CAACCACCACGAGAACAACC REV: GGCTTGGTTGTAACGGTGTT	-
UL21A	FWD: GTCGGTGAGGGAGATGAAGA REV: GAGCAAAACCGGGTACATGG	-
UL30	FWD: CCGGGGCATGATGGACTAT REV: GGCAACACAAGACAGGGAAA	-
UL33	FWD: TACTCGAGCTGCACAGTAGG REV: CATGGAAAAGATCAGCCCGG	-
UL48A	FWD: GACCGAGATCTCAGAGGCC REV: CTATCATGCGCAACAGGTCC	-
UL91	FWD: CTTTGTCGACCGCCTCTTTC REV: CAGGTGTTGCTTGTCTCCAC	-
UL136	FWD: TGCGGCTGTCATTATCCTGA REV: CCATTTCCACCGTGTCGAAG	FWD: ATGTCAGTCAAGGGCGTGGAGAT REV: TAGCGGGAGATACGGCGTTC
UL148D	FWD: GTCAAGACCAAGGAGCAGC REV: GCGTAAAGTACATCAGGGCC	FWD: ATGACGGCGCCCAAGTGTG REV: AACGGGAGCGGCAGCGGC
UL146	FWD: CGCTGCAAATGTCTTGATGG REV: GGTGATGGGGCGATAAACAT	FWD: ATGCGATTAATTTTTGGTGCGTT REV: TTCCTTCAGACCTACTAGGGTTA
RL6	FWD: CTAAAAGCGACGACTGGGAG REV: TATGACCACAGCTCGACACA	-
RL9A	FWD: ATGTCTCTAGATGCCGCCAGCC REV: TTAGAGAAACAGCACGTAGGTCA	-
RL12	FWD: CACCCCACTATGTCCCAGAT REV: GTAAGAGCCCATGTAGTGCG	-
GAPDH	FWD: AGGGCTGCTTTTAACTCTGG REV: CCCCACTTGATTTTGGAGGGA	-

2.7. Immunoblotting

2.7.1 Preparation of cell lysates

To generate lysates for immunoblotting, a confluent monolayer of HFFF-TERT cells in a 175 cm^2 tissue culture flask was washed with PBS, followed by incubation with 6mL trypsin-EDTA solution for 5 min at 37 °C, 5 % CO_2. Once the cells were in suspension, 4 mL of complete medium was added and 2.5 mL of this mixture was centrifuged at 400 x g for 5 min, RT. The cell pellet was washed with PBS and centrifuged once more at 400 x g for 5 min, RT. After discarding the supernatant, the pellet was re-suspended in 100 µL of RIPA 1X buffer (CST) containing Roche protease inhibitor cocktail (1 tablet/10 mL) and incubated on ice for 5 min. The lysate was then mixed for 30 s using a vortex homogeniser, followed by centrifugation at 16,100 x g for 10 min, 4 °C. The supernatant was then transferred to a new tube and stored at -20 °C until necessary.

2.7.2 Estimation of protein concentration

Protein quantification was performed using the Pierce BCA Protein Assay kit (Thermo Fisher Scientific). Samples were diluted both 1:5 and 1:10 in PBS and 25 µL were loaded onto a well of a flat-bottom 96-well clear plate (Greiner). Using a multi-channel pipette (Gilson), 200 µL of the BCA working reagent were added to each well. The plate was sealed with PlateMax AxySeal adhesive film (Axygen) and incubated at 37 °C for 30 min. The absorbance at 562 nm was then measured using a Spark Microplate Reader (Tecan). Serial dilutions of a 2 mg/mL Bovine Serum Albumin solution (Thermo Fisher Scientific) with PBS were assayed in parallel to generate a standard curve by linear regression analysis that allowed estimation of protein concentration.

2.7.3 Immunoblotting

Approximately 20 to 50 µg of protein for each sample was reduced with 6X Protein Loading Dye (Tris 375 mM pH 6.8, 12 % (w/v) SDS, 30 % (v/v) glycerol, 0.6 M DTT, 0.06 % (w/v/) bromophenol blue) for 5 min at 95 °C. Equal amounts of each protein sample were then separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using 4-12 % Bis-Tris Precast Protein Gels (Invitrogen) in an XCell SureLock Mini-Cell Electrophoresis System (Invitrogen), for 105 min at 130 V. Pre-stained protein molecular weight standards were run in parallel.

Proteins were then transferred to a PVDF membrane by semi-dry transfer using the Trans-Blot SD Semi-Dry Transfer Cell (Bio-rad). Gel blotting paper (Whatman) was immersed in transfer buffer (described in **2.1**), then three sheets of blotting paper were layered on the negative electrode of the transfer cell. The PVDF membrane was soaked in methanol, then immersed in transfer buffer and placed on top of the blotting paper. The polyacrylamide gel was removed from the electrophoresis system, immersed in transfer buffer and layered onto the PVDF membrane, followed by three sheets of pre-soaked blotting paper. The transfer stack was then gently pressed in a rolling motion to ensure any air bubbles in its midst were removed. The transfer cell was then assembled and protein transfer onto the membrane was performed at 20 V for 45 min.

Following transfer, the membrane was removed from the transfer cell and incubated with blocking solution (5 % (w/v) milk in PBST) for 1 h, RT. An appropriate amount of primary antibody was added to the blocking solution to achieve the dilution recommended by the suppliers and incubated O/N at 4 °C with agitation. The following day the blot was washed three times for 5 min with PBST and then incubated for 2 h with 5 % (w/v) milk in PBST containing an appropriate dilution of secondary antibody. The blot was then washed three times for 5 min with PBST, followed by digital image

acquisition using the Odyssey CLx (LI-COR). Digital images were processed using Image Studio Lite (LI-COR). Information on primary and secondary antibodies used, as well as their dilutions can be found in Table 2.7.3.

Table 2.7.3 | Antibodies

Antibody	Dilution	Isotype	Company	Clone	Application
Primary Antibodies					
α-V5	1:2000	Mouse IgG1	Thermo Fisher	E10/V4RR	IB
α-HA	1:1000	Rabbit IgG	CST	C29F4	IB
α-Calnexin	1:10000	Rabbit IgG	Lifespan BioSciences		IB
α-GAPDH	1:10000	Mouse IgG1	R&D Systems	686613	IB
α-CNOT7	1:1000	Mouse IgG	Abcam	EPR18722	IB
α-CNOT2	1:1000	Rabbit IgG	Novus		IB
α-NEDD4	1:1000	Mouse IgG_{2B}	R&D Systems	683211	IB
α-IE1	1:1000	Mouse IgG_{2A}	Merck	8B1.2	FACS
APC α-HLA-A, B, C	1:1000	Mouse IgG_{2A}	Biolegend	W6/32	FACS
Secondary Antibodies					
IRDye 680RD α-mouse	1:10000	Goat IgG	LI-COR		IB
IRDye 680RD α-rabbit	1:10000	Goat IgG	LI-COR		IB
IRDye 800CW α-mouse	1:10000	Goat IgG	LI-COR		IB
IRDye 800CW α-rabbit	1:10000	Goat IgG	LI-COR		IB
AlexaFluor 488 α-mouse	1:1000	Goat F(ab')2 Fragment	CST		FACS

2.8 Flow Cytometry

At the appropriate time-point of infection, cells were washed in PBS and brought into suspension with Trypsin-EDTA. The cell suspension was then centrifuged at 400 x g, RT for 5 min. Supernatant was discarded and cells were fixed by incubating with 200 µL of 4 % Paraformaldehyde, for 10 min at RT. The fixing solution was removed by centrifugation at 800 x g, RT for 5 min.

For cells infected with a viral strain containing a GFP-tagged gene, cells were re-suspended in 200 µL PBS and analysed by flow cytometry.

For cells infected with a viral strain lacking a GFP-tagged gene, cells were re-suspended in 200 µL ice-cold Methanol and incubated for 5 min at 4 °C, for permeabilisation of cell membranes. Cells were washed with PBS/0.2 % (v/v) FBS and pelleted by centrifugation at 800 x g, 4 °C for 5 min. After discarding the supernatant, the pellet was re-suspended in 25 µL Fc receptor blocking solution Human TruStain FcX (BioLegend) and incubated for 1 h at 4 °C. Cells were incubated with 25 µL mouse α-IE1 primary antibody in PBS (see Table 2.7.3) for 1 h at 4 °C. After washing in ice-cold PBS and centrifuging at 800 x g, 4 °C for 5 min, cells were re-suspended in 50 µL goat α-mouse IgG Alexa Fluor (see Table 2.7.3) secondary antibody in PBS for 1 h at 4 °C. The staining solution was removed by centrifugation at 800 x g, 4 °C for 5 min and cells were re-suspended in 200 µL PBS for flow cytometry analysis.

Flow cytometry was carried with a FACSCalibur Cell Analyser (BD Biosciences) with CellQuest PRO software (BD Biosciences). Cell populations were selected by forward scatter and side scatter gating before analysis with the appropriate laser excitation. Data analysis was performed with FlowJo software (BD Biosciences).

2.9. Immunoprecipitation

2.9.1 Affinity-purification mass spectrometry

2.9.1.1 Viral infection

HFFF-TERT stable cell lines expressing different viral baits were infected in batches of eight, in duplicate. For each cell line, $6x10^6$ cells were plated in complete medium in each of two 150 cm^2 dishes. The following day, media was replaced with serum-free DMEM containing 4 µg/ml Dexamethasone and incubated for 24 h, 37 °C, 5 % CO_2, as this approach has been shown to improve infection efficiency [198]. After 24 h, cells were infected in 10 mL serum-free DMEM containing requisite volume of HCMV Merlin strain stock to achieve MOI 2. Cells were gently rocked for 2 h, then the infection inoculum was replaced with complete medium and cells were incubated for a further 58 h. Infection of all cell lines was performed using aliquots of the same pool of virus.

2.9.1.2 Cell lysis

In order to preserve protein-proteins interactions, lysates were collected in one of two lysis buffers, according to the solubility of each overexpressed transgene. For soluble and single-pass transmembrane (TM) baits, cells were lysed on ice in 1.2 mL MCLB buffer (50 mM Tris pH 7.5, 300 mM NaCl, 0.5 % (v/v) NP-40, 1 M DTT and Roche protease inhibitor cocktail (1 tablet/10 mL)). Baits with more than one TM domain were solubilized on ice in 1 % (w/v) Digitonin (Merck Millipore) in TBS (Sigma) with Roche protease inhibitor cocktail (1 tablet/10 mL). Prediction of TM domains was taken from Uniprot (www.uniprot.org) for canonical HCMV proteins, and generated using TMHMM for the two uncharacterised ORFs [199]. Lysis buffers were pre-chilled before use.

Lysates were scraped on ice and transferred into a 2 mL tube. Samples were tumbled for 15 min at 4 °C, then centrifuged at 16,100 x g for 15 min at 4 °C. Lysates were clarified by filtration with a 0.7 µm filter.

For MCLB buffer, all samples were lysed using aliquots of the same batch of lysis buffer.

2.9.1.3 Anti-V5 Immunoprecipitation

Prior to incubation with samples, immobilised mouse monoclonal anti-V5 agarose resin (Sigma-Aldrich) was washed three times with 1 mL lysis buffer with intermediate centrifugation steps at 2,000 x g, 4 °C for 2 min. Each lysate was incubated for 3 h with 30 µl anti-V5 agarose resin, at 4 °C in a rotating mixer (Stuart). Duplicate samples were combined for resin washes. Washes were performed in Pierce Spin Columns (Thermo Fisher Scientific) placed in a Vac-Man Laboratory Vacuum Manifold (Promega), allowing for the flow-through waste to be directly collected into the manifold chamber. Samples lysed in NP-40-containing buffer were washed seven times with 700 µl lysis buffer, followed by seven 700 µl ice-cold PBS pH 7.4 washes. Samples lysed in Digitonin-containing buffer were washed once with 700 µl lysis buffer, twice with 700 µl 0.2 % (w/v) Digitonin in TBS and then once with 700 µl TBS. Proteins bound to the resin were incubated with 200 µl of 250 µg/ml V5 peptide (Alpha Diagnostic International) in PBS, at 37 °C for 30 min with agitation and eluted into a 1.5 mL tube by centrifugation at 4,000 rpm for 1 min. Incubation with elution buffer and centrifugation steps were repeated once more.

For MCLB buffer, all samples were washed with aliquots from the same batch of buffer. Several batches of the anti-V5 agarose resin used for immunoprecipitation were pooled and this pool was used for all samples.

2.9.1.4 Protein precipitation

Eluted proteins were precipitated with 80 µl 20 % (v/v) Trichloroacetic acid (TCA) by incubation at 4 °C for 45 min. Samples were centrifuged at 16,100 x g, 4 °C for 30 min and the majority of the supernatant was discarded. The protein pellet was washed once with 1 mL 10 % (v/v) TCA, mixed by inversion and centrifuged at 16,100 x g, 4 °C for 20 min. After removing the supernatant, samples were washed three times with 1 mL cold acetone with intermediate centrifugation steps at 16,100 x g, 4 °C for 20 min, always discarding the supernatant. Protein pellet was dried to completion using a centrifugal evaporator.

2.9.1.5 Trypsin digest

Proteins were re-suspended in 20 µL digestion buffer (50 mM Tris pH 8.5, 10 % (v/v) AcN, 1 mM DTT, 10 µg/mL Trypsin) and incubated O/N at 37 °C with agitation. The digestion reaction was quenched with 50 µL 50 % (v/v) formic acid (FA). To minimise variability in sample preparation, all protein digests were performed with aliquots from the same stock of digestion buffer.

2.9.1.6 StageTip

Peptide samples were subjected to C18 solid-phase extraction on a StageTip placed on a 2 mL collection tube. Prior to sample loading, the C18 solid-phase was washed with 50 µL Methanol and centrifuged for 1 min at 3,000 rpm. The solid-phase was washed with 50 µL 70 % (v/v) AcN/ 1 % (v/v) FA, followed by centrifugation for 1 min at 3,000 rpm. The StageTip was then equilibrated with 50 µL 1 % (v/v) FA, followed by centrifugation for 1 min at 3,000 rpm. Peptide samples were then loaded onto the tip, centrifuged for 1 min at 3,000 rpm. C18-bound peptide was washed with 100 µL 1 %

(v/v) FA, followed by centrifugation for 1 min at 3,000 rpm. The collection tube was replaced by a new 1.5 mL tube and bound peptide was eluted with 50 µL 70 % (v/v) AcN/ 1 % (v/v) FA by centrifugation for 1 min at 3,000 rpm. Samples were vacuum-centrifuged to complete dryness, followed by reconstitution in 10 µL 4 % (v/v) AcN / 5 % (v/v) FA. Samples were divided into two technical duplicates of 5 µL in glass vials, prior to LC-MS/MS on the Orbitrap Lumos.

2.9.1.7 LC-MS/MS

Peptide samples derived from each cell line were analysed in technical duplicate. The LC-MS/MS queue order was reversed between batches of technical replicates, to avoid peptide carry-over. Additionally, two washes on the liquid-chromatography (LC) system were used between each sample to further minimise carry-over. Individual batches included 16 - 22 samples. To ensure consistent performance by the mass spectrometer between batches, an identical aliquot of a control IP of uninfected cells stably expressing the viral UL123 gene with a C-terminal V5 tag was included at the beginning of each batch. The total number of peptides were very similar between each batch (**Table 5.1**, page 136).

Mass spectrometry data was acquired using an Orbitrap Lumos. An Ultimate 3000 RSLC nano UHPLC equipped with a 300 µm ID x 5 mm Acclaim PepMap µ-Precolumn (Thermo Fisher Scientific) and a 75 µm ID x 50 cm 2.1 µm particle Acclaim PepMap RSLC analytical column was used.

Loading solvent was 0.1 % (v/v) FA, while the analytical solvents were A: 0.1 % (v/v) FA and B: 80 % AcN + 0.1 % (v/v) FA. All separations were carried out at 55 °C. Samples were loaded at 5 µl/min for 5 min in loading solvent before beginning the analytical gradient. The following gradient was used: 3-7 % B over 3 min, then 7-37 % B

over 54 min followed by a 4 min wash in 95 % B and equilibration in 3 % B for 15 min. The following settings were used: MS1: 350-1500 Thompsons (Th), 120,000 resolution, $2x10^5$ automatic gain control (AGC) target, 50 ms maximum injection time. MS2: Quadrupole isolation at an isolation width of m/z 0.7, higher-energy collisional dissociation (HCD) fragmentation (normalised collision energy (NCE) 34) with fragment ions scanning in the ion trap from m/z 120, $1x10^4$ AGC target, 250 ms maximum injection time, with ions accumulated for all parallelisable time. The method excluded undetermined and very high charge states (≥25+). Dynamic exclusion was set to +/- 10 ppm for 25 s. MS2 fragmentation was trigged on precursors $5x10^3$ counts and above. Two 45 min washes were included between every affinity-purification mass spectrometry (AP-MS) analysis, to minimise carry-over between samples.

2.9.2 Immunoprecipitation for immunoblotting

HEK293T cells transiently transfected with lentiviral constructs expressing an HCMV gene and its human binding partner were plated in complete medium as described in **2.2.7**. Alternatively, $6x10^6$ HFFF-TERT cells from a stable cell line expressing a viral gene were plated in a 150 cm^2 dish. The following day, cell lysates were prepared as described in **2.9.1.2**.

Immunoprecipitation was performed as described in **2.9.1.3** with the following modifications: (a) samples were washed three times with lysis buffer, followed by two PBS pH 7.4 washes; (b) proteins bound to the anti-V5 resin were eluted once by adding 40 µl of 2.5 mg/ml V5 peptide (Alpha Diagnostic International) in PBS at 37 °C for 30 min with agitation.

The eluates were reduced with 6X Protein Loading Dye (Tris 375 mM pH 6.8, 12 % (v/v) SDS, 30 % (v/v) glycerol, 0.6 M DTT, 0.06 % (w/v) bromophenol blue) for 5 min at 95 °C. Approximately 50 µg of protein for each sample was separated by PAGE using 4-15 % TGX Precast Protein Gels (Bio-rad), then transferred to PVDF membranes using Trans-Blot Systems (Bio-rad), as described in **2.7**.

2.10 Proteomic analysis of whole-cell lysates

Proteomic analysis of whole-cell lysates was performed to compare the relative abundance of ORFL147C protein in cells infected with WT (RCMV 2582) or ΔORFL147C virus (RMCV 2697), to confirm that the deletion mutant did not express ORFL147C protein.

2.10.1 Viral infection

$1.5x10^5$ HFFF-TERT cells were seeded per well of a 12-well plate. The following day, media was replaced with serum-free DMEM containing 4 µg/ml Dexamethasone and incubated for 24 h, 37 °C, 5 % CO_2. Cells were then infected with mock (serum-free DMEM), WT (RCMV 2582) or ΔORFL147C virus (RMCV 2697) in duplicates, at MOI 2 for 48 h, as described in **2.3.4**. The following day, one set of samples (mock, WT and ΔORFL147C) were processed for flow cytometry (as described in section **2.8**) using GFP as a marker to determine percentage of infection. The other set of samples was processed for whole-cell lysate proteomics to compare ORFL147C expression.

2.10.2 Cell lysis

For whole-cell lysate proteomics cells were washed twice with PBS. After removing any residual PBS, lysates were collected from one set of duplicates using 150 µL 6 M Guanidine / 50 mM HEPES pH 8.5 lysis buffer. Lysates were homogenised by vortex for 30 sec and centrifuged at 13,000 x g for 10 min. After transferring the supernatant to a new tube, the spin and transfer steps were repeated once more.

2.10.3 Reduction and alkylation of disulphide bonds

Then, 7.5 µL of 100 mM DTT were added to the lysate, the mixture was homogenised by vortex and incubated for 20 min, RT. The samples were then alkylated by adding 4.2 µL of 500 mM Iodoacetamide. The mixture was again homogenised by vortex and incubated for 20 min, RT, in the dark. Alkylation was quenched by adding 7.5 µL of 100 mM DTT, homogenisation by vortex and incubation for 15 min, RT.

2.10.4 Protein digestion with LysC and Trypsin

The lysate was then diluted to a final concentration of 1.5 M Guanidine Hydrochloride by adding 450 µL 200 mM HEPES pH 8.5. Samples were incubated with 3 µL LysC for 3 h at RT, followed by a further dilution with 1.05 mL 200 mM HEPES pH 8.5. Lysates were then incubated at 37 °C O/N with 150 µL of 3.33 ng/µL Trypsin. The next day, samples were acidified by adding 100 µL of 50 % Formic acid (FA), homogenised by vortex, centrifuged for 10 min at 21,000 x g and transferred to a new tube.

2.10.5 Protein isolation with SepPak

Peptides were then purified using a SepPak column (Waters) placed on a vacuum manifold. The SepPack column was washed with 2 mL of 100 % Acetonitrile (AcN), followed by 1 mL of 70 % (v/v) AcN/ 1 % (v/v) FA and then 3 mL of 1 % (v/v) FA. After loading the sample, the column was washed with 3 mL 1 % (v/v) FA. Peptides were eluted with 350 µL 70 % (v/v) AcN/ 1 % (v/v) FA and dried to completion using a centrifugal evaporator.

Peptides were then resuspended in 150 µL 200 mM HEPES pH 8.5 and the concentration in each sample was estimated using a Pierce BCA Protein Assay kit (Thermo Fisher Scientific) as described in section **2.7.2**.

2.10.6 Peptide labelling with Tandem Mass Tags

Samples were then labelled with Tandem Mass Tags (TMT): mock - 126; wild-type – 127N; ΔORFL147C – 128N. Firstly, 16 µL of peptide (15 µg) were added to 6 µL AcN, homogenised by vortex and briefly spun down to collect the sample at the bottom of the tube. Approximately 93 µg of each TMT label was then added to each sample, followed by homogenisation by vortex and brief spin. The labelling reaction was incubated for 1 h at RT. TMT labelling was quenched by adding 5 µL of 5 % (v/v) Hydroxylamine and incubating for 15 min. Samples labelled with different tags were combined 1:1:1, and acidified by adding 5 µL 50 % (v/v) FA per number of samples combined. The sample was vacuum-centrifuged to near dryness and subjected to C18 SPE (Sep-Pak, Waters).

2.10.7 Offline high pH reversed phase fractionation

Six fractions were then generated using high pH reversed phase fractionation to increase the overall number of peptides quantified. TMT-labelled tryptic peptides were subjected to high pH reversed phase fractionation using an Ultimate 3000 RSLC UHPLC system (Thermo Fisher Scientific) equipped with a 2.1 mm internal diameter (ID) x 25 cm long, 1.7 µm particle Kinetix Evo C18 column (Phenomenex). Mobile phase consisted of A: 3 % acetonitrile (MeCN), B: MeCN and C: 200 mM ammonium formate pH 10. Isocratic conditions were 90 % A/10 % C, and C was maintained at 10 % throughout the gradient elution. Separations were conducted at 45 °C. Samples were loaded at 200 µl/min for 5 min. The flow rate was then increased to 400 µl/min over 5 min, after which the gradient elution proceed as follows: 0-19 % B over 10 min, 19-34 % B over 14.25 min, 34-50 % B over 8.75 min, followed by a 10 min wash at 90 % B. UV absorbance was monitored at 280 nm and 15 s fractions were collected into 96-well microplates using the integrated fraction collector. Wells were excluded prior to the start or after the cessation of elution of peptide-rich fractions, as identified from the UV trace. Fractions from adjacent columns were combined pairwise (e.g. 1+2, 3+4, 5+6) to yield 6 combined fractions, which were dried in a centrifugal evaporator and subsequently re-suspended in 10 µL of 5 % FA/ 4 % AcN prior to LC-MS3.

2.10.8 LC-MS3

Mass spectrometry data was acquired using an Orbitrap Lumos. An Ultimate 3000 RSLC nano UHPLC equipped with a 300 µm ID x 5 mm Acclaim PepMap µ-Precolumn (Thermo Fisher Scientific) and a 75 µm ID x 50 cm 2.1 µm particle Acclaim PepMap RSLC analytical column was used. Loading solvent was 0.1 % FA, analytical solvent A: 0.1 % FA

and B: 80 % AcN + 0.1 % FA. All separations were carried out at 55 °C. Samples were loaded at 5 µL/min for 5 min in loading solvent before beginning the analytical gradient. The following gradient was used: 3-7 % B over 3 min, 7-37 % B over 173 min, followed by a 4 min wash at 95 % B and equilibration at 3 % B for 15 min. Each analysis used a MultiNotch MS3-based TMT method [200]. The following settings were used: MS1: 380-1500 Th, 120,000 Resolution, $2x10^5$ AGC target, 50 ms maximum injection time. MS2: Quadrupole isolation at an isolation width of m/z 0.7, CID fragmentation (NCE 35) with ion trap scanning in turbo mode from m/z 120, $1.5x10^4$ AGC target, 120 ms maximum injection time. MS3: In Synchronous Precursor Selection mode the top 6 MS2 ions were selected for HCD fragmentation (NCE 65) and scanned in the Orbitrap at 60,000 resolution with an AGC target of $1x10^5$ and a maximum accumulation time of 150 ms. Ions were not accumulated for all parallelisable time. The entire MS/MS/MS cycle had a target time of 3 s. Dynamic exclusion was set to +/- 10 ppm for 70 s. MS2 fragmentation was trigged on precursors $5x10^3$ counts and above. Data analysis is described in section **2.11**.

2.11 Data analysis

2.11.1 Database and search parameters for protein identification

Mass spectra were processed using a Sequest-based software pipeline for quantitative proteomics, "MassPike", through a collaborative arrangement with Professor Steven Gygi's laboratory (Department of Cell Biology, Harvard Medical School). Mass spectra were converted to mzXML using an extractor built upon Thermo Fisher's RAW File Reader library (version 4.0.26). This software is a component of the MassPike software platform and is licensed by Harvard Medical School.

A combined database was constructed from (a) the human Uniprot database (26th January, 2017), (b) the HCMV strain Merlin Uniprot database, (c) all additional non-canonical human cytomegalovirus ORFs described by Stern-Ginossar et al [67], (d) a six-frame translation of HCMV strain Merlin filtered to include all potential ORFs of ≥8 residues (delimited by stop-stop rather than requiring ATG-stop) and (e) common contaminants such as porcine trypsin and endoproteinase LysC. ORFs from the six-frame translation (6FT-ORFs) were named as follows: 6FT_Frame_ORFnumber_length, where Frame is numbered 1-6, and length is the length in amino acids. The combined database was concatenated with a reverse database composed of all protein sequences in reversed order. Searches were performed using a 20 ppm precursor ion tolerance [201]. Product ion tolerance was set to 0.03 Th. Oxidation of methionine residues (15.99492Da) was set as a variable modification.

To control the fraction of erroneous protein identifications, a target-decoy strategy was employed [202, 203]. Peptide spectral matches (PSMs) were filtered to an initial peptide-level false discovery rate (FDR) of 1 % with subsequent filtering to attain a final protein-level FDR of 1 % [204, 205]. PSM filtering was performed using a linear discriminant analysis, as described [206]. This distinguishes correct from incorrect peptide IDs in a manner analogous to the widely used Percolator algorithm [207], though employing a distinct machine learning algorithm. The following parameters were considered: XCorr, ΔCn, missed cleavages, peptide length, charge state, and precursor mass accuracy.

For MS3-based TMT analysis, TMT tags on lysine residues and peptide N termini (229.162932 Da) and carbamidomethylation of cysteine residues (57.02146 Da) were included as static modifications. Proteins were quantified by summing TMT reporter ion counts across all matching peptide-spectral matches using ”MassPike”. A 0.003 Th

window around the theoretical m/z of each reporter ion (126, 127n, 128n) was scanned for ions, and the maximum intensity nearest to the theoretical m/z was used. An isolation specificity filter with a cut-off of 50 % was employed to minimise peptide co-isolation [200]. Peptide-spectral matches with poor quality MS3 spectra (more than 3 TMT channels missing and/or a combined S:N ratio of less than 100 across all TMT reporter ions) or no MS3 spectra at all were excluded from quantitation. Peptides meeting the stated criteria for reliable quantitation were then summed by parent protein, in effect weighting the contributions of individual peptides to the total protein signal based on their individual TMT reporter ion yields. Protein quantitation values were exported for further analysis in Excel. For protein quantitation, reverse and contaminant proteins were removed, then each reporter ion channel was summed across all quantified proteins and normalised assuming equal protein loading across all channels. Fractional TMT signals were used (i.e. reporting the fraction of maximal signal observed for each protein in each TMT channel, rather than the absolute normalized signal intensity). This effectively corrected for differences in the numbers of peptides observed per protein.

2.11.2 Interactor identification with CompPASS

Data from the technical replicate from each viral bait was combined to attain a summary of proteins identified in both runs. Peptides within replicates were reassembled into proteins following the principles of parsimony [206]. Where all PSMs from a given HCMV protein could be explained either by a canonical gene or non-canonical ORF, the canonical gene was picked in preference. In four cases (UL24/ORFL71C_(UL24), UL31/ORFL87W_(UL31), UL150A/ORFL321W, UL44/ORFL112C_(UL44)), PSMs assigned to a non-canonical ORF were a mixture of peptides from the canonical protein and the ORF. This occurred where the ORF was a 5'-

terminal extension of the canonical protein (thus meaning that the smallest set of proteins necessary to account for all observed peptides included the ORFs alone). In these cases, the peptides corresponding to the canonical protein were separated from those unique to the ORF, generating two separate entries.

For human and canonical HCMV proteins, Uniprot protein IDs were mapped to Entrez Gene IDs. Downstream interactor and network analysis was carried out at the level of Gene IDs to minimise confusion due to protein isoforms. CompPASS scoring was performed as described previously [189], in two analyses that were subsequently combined, one for NP40-based IPs and the other for Digitonin IPs. Data reported for each protein in every IP in the dataset included: (a) the number of peptide spectrum matches (PSMs), averaged between technical replicates; (b) a log(2)-based Shannon entropy score, comparing numbers of spectral counts observed for a protein in two technical replicates to eliminate proteins that were not detected consistently (see Figure 2.11.2.1). (c) a z-score, calculated in comparison to the average and standard deviation of PSMs observed across all IPs. (d) an NWD score, calculated as described in [208] using the fraction of runs in which a protein was observed, the observed number of PSMs, the average and standard deviation of PSMs observed for that protein across all IPs, and the number of replicates (1 or 2) containing the protein of interest. NWD scores were normalized so that the top 2 % earned scores were equal to or greater than 1.0. For NP40-based IPs, the top 2 % of z-scores were greater than 6.676, and for Digitonin-based IPs greater than 4.329. Equations for NWD and Z scores are provided in Figure 2.11.2.2.

As the set of Digitonin-based IPs was necessarily smaller than NP40-based IPs (18 compared to 153 viral genes examined respectively), additional control IPs were included. Biological duplicates of cells transduced with empty vector controls ('No bait'), and biological duplicates of cells transduced with the pHAGE-SFFV vector encoding GFP

were included in the Digitonin set. A single replicate of the 'No bait' control was included in the NP40 set. These controls had the effect of increasing the number of IPs that identified non-specific interacting proteins, thus decreasing NWD and z-scores for these proteins. Mass spectrometry RAW files from control UL123 IPs included to ensure batch-to-batch consistency were not included in the final data analysis to avoid modification of NWD and z-scores for the infected UL123-expressing sample.

Following CompPASS analysis, a series of filters were applied to remove inconsistent and low-confidence protein identifications across all IPs and minimize both false protein identifications and associations. These included: (a) a minimum PSM score of 1.5 (i.e. a minimum of 3 peptides per protein across both replicates); (b) a minimum entropy score of 0.75; (c) a top 2 % NWD or z-score. Previous studies have estimated a 5% false discovery rate when employing a similar strategy with a top 2 % NWD score [195]. As found in prior human interactome investigations, certain known interactions fell just below the stringent top 2 % NWD or z-score cut-offs. Proteins were therefore also included with top 5 % NWD or z-scores (>0.434 and >3.688 respectively), if they were reported to interact with the bait in a prior study [209]. For protein UL133 (2 TM regions), an initial Digitonin-based AP-MS analysis failed to generate any interactors after filtering. This IP was repeated using the NP40-based lysis buffer. Protein-protein interaction visualisation and diagram generation was performed using Cytoscape ver 3.7.1 [210].

Figure 2.11.2.1 | Equations for the CompPass parameter Entropy (adapted from Huttlin et al, 2015). This parameter is particularly useful for detection of LC-carry-over and inconsistent proteins, requiring technical replicates for each IP, and inverse ordering of the second replicates, as shown in the table for prey Z. Rows correspond to different AP-MS experiments (Baits I to X); PSMs quantified in each replicate for baits I to X are listed in middle columns for prey Z. Although prey Z was mostly observed in bait V IP, decreasing levels were seen in subsequent runs. Because run order was reversed for replicate B, carry-over can be distinguished based on substantial PSM differences between replicates, and can be quantified with an entropy score [211]. To enable calculation when 0 PSMs are observed, a pseudo-count of 0.5 is added to each replicate. In Huttlin et al, (2015) an analysis of the distribution of entropy values identified that a cut-off of 0.75 removes irreproducible proteins (red) while leaving consistently-detected proteins (green). However, in the present study, two washes were included between each sample run (as opposed to the Huttlin study where, due to the much larger interactome size, no washes could be included). The degree of carry-over would therefore be expected to be considerably lower. Adapted from Huttlin et al, 2015.

Figure 2.11.2.2 | Equations for the CompPass parameters Z and NWD
CompPass generates an interaction matrix in which the rows are the proteins identified from all AP-MS experiments and the columns represent each bait used for these experiments. Each cell in the matrix contains the Total Spectral Count (TSC) for a specific interacting protein from a particular bait's AP-MS experiment. The TSC for a protein provides a good estimation of that protein's abundance in the AP-MS experiment. Scores are then calculated for each interactor, for each bait. After identifying peptides and proteins, given that each bait was run twice, the duplicates are combined into a single "merged" run for that bait. When merging, the TSC for an interactor is the average TSC observed from the duplicate runs. The Z-score determines the number of standard deviations away from the mean (Eq. 1) at which a measurement lies (Eq. 2). The NWD score is calculated as shown in Eq. 5 and incorporates the frequency of the observed interactor, its abundance, the mean and standard deviation of TSCs observed for that protein across all IPs, and the reproducibility of that interaction. In Eq. 4, f is a term which is 0 or 1 depending on whether or not the interacting protein was immunoprecipitated by a given bait. Placed in the summation across all baits, it is a counting term and therefore, '$k/\sum f$' represents the frequency (λ) of an interactor across all baits. The smaller f, the larger the frequency becomes, up-weighting interactors that are rare. The weight factor, ωi (Eq. 3), is added as a multiplicative factor to the frequency term in order to offset this low value for interactors that are found frequently across baits but will only be >1 if the conditions in Eq. 5 are met. In this way, only if a frequent interactor displays the observed characteristics of a true interactor will its score increase due to the weight factor. The power term r takes into account the reproducibility of the interaction and allows for discrimination between a true 'one hit wonder' (a protein found with 1 peptide in a single run, not in the duplicate) which is likely a false positive versus a true interactor with low TSC (even 1) that is found in both duplicate runs. The term $X_{b,i}$ is the TSC for interactor i from bait b and therefore multiplying by this value scales the score with increasing interactor TSC. This provides a higher score to interactors having high TSC and are therefore more abundant and less likely to be stochastically sampled. Adapted from Sowa et al, 2009.

2.11.3 Interaction database comparisons

Lists of physical interactions between viral proteins and human proteins were downloaded in October 2018 from: BioGRID [212], IntAct [213], Uniprot (www.uniprot.org), MINT [214], and Virus Mentha [215]. Entries were then compiled into a single file with duplicate entries removed. Search terms for each database were as follows:

- BioGRID: “Human Herpesvirus 5”
- IntAct: “Human cytomegalovirus (strain AD169) (HHV-5)”
- Uniprot: “Human cytomegalovirus (HHV-5) (Human herpesvirus 5) [10359]”
- MINT: “Human cytomegalovirus” & “Human cytomegalovirus (strain AD169) (HHV-5)”
- Virus Mentha: “Human herpesvirus 5” & “Human herpesvirus 5 strain AD169” & “Human herpesvirus 5 strain Merlin” & “Human herpesvirus 5 strain Towne”

2.11.4 Functional enrichment analysis

The Database for Annotation, Visualisation and Integrated Discovery (DAVID) version 6.8 was used to determine functional term enrichment [216]. All human HCIPs for all viral baits were searched against a background of all human proteins, using default settings provided by this bioinformatics tool.

2.11.5 Domain association analysis

Domain associations were generated by mapping Pfam domains [217] provided by Uniprot onto all proteins in the dataset. Then the total number of interactions for each domain as well as the number of interactions involving pairs of domains from interacting

proteins were counted, and the significance of the association between domains was calculated using Fisher's Exact Test [189]. p-values were corrected for multiple hypothesis testing [218]. Domains were considered significantly associated if their adjusted p-value was less than 0.01.

2.11.6 ORFL147C DNA and amino acid sequence alignment analysis

Basic local alignment search tool (BLAST, National Center for Biotechnology Information) [219] was used to search for homologues of ORFL147C in other human herpesvirus. The nucleotide and amino acid sequences of ORFL147C were searched against a database of 'Nucleotide collection (nr/nt)' or 'Non-redundant protein sequences (nr)', specifying the following organisms HSV-1 (taxid: 10298), HSV-2 (taxid: 10310), VZV (taxid: 10335), EBV (taxid: 10376), HHV-6 (taxid: 10368), HHV-7 (taxid: 10372) and KSHV (taxid: 37296).

To identify a degree of conservation for the amino acid sequence of ORFL147C between HCMV strains, the DNA sequences for this ORF in strains Merlin (NC_006273), Towne (FJ616285), Toledo (GU937742) and AD169 (FJ527563) were translated using the ExPASy Translate tool (Swiss Institute of Bioinformatics) [220] and aligned using the Clustal Omega (EMBL-EBI) [221] with default settings.

Alignment of the amino acid sequence of ORFL147C with cy89 protein (Uniprot: G8H195) from *Cynomolgus macaque* cytomegalovirus (CyCMV) strain Ottawa, Cy89 protein (Uniprot: A0A0K1H0A6) from CyCMV strain Mauritius and Rh91.1 (Uniprot: Q2FAM5) from Rhesus cytomegalovirus (RhCMV) was also performed using the Clustal Omega (EMBL-EBI) multiple sequence alignment tool with default settings.

3 | Generating resources for AP-MS

Normalised IBAQ quantification was performed by Dr. Michael Weekes (Department of Medicine, University of Cambridge). Codon optimisation of UL74, US14 and US17 was performed by Dr. James Davies and Dr. Sepehr Seirafian (School of Medicine, Cardiff University). Parts of the work presented in this chapter have been published in a similar form in eLife (Nobre et al., 2019).

Affinity-purification mass-spectometry allows identification of isolated protein complexes bound to a specific protein bait. Systematic analysis of these datasets in tandem generates a network map of protein interactions, designated as an interactome, providing valuable insights into gene function.

To enable the identification of interactors for each HCMV protein (HCMV interactome), three types of resources were necessary: (a) expression vectors with the coding sequences of each viral protein followed by a tag for affinity-purification; (b) cell lines that would constitutively express the tagged viral proteins individually; (c) an homogeneous stock of HCMV, sufficient to infect all cell lines to a similar MOI.

3.1 Generation of the expression construct library

Expression vectors for each viral bait were generated by cloning the coding sequences for 171 canonical protein-coding genes from HCMV strain Merlin and 2 uncharacterised viral ORFs (see Figure 3.1) into the pHAGE-SFFV lentiviral vector. The expression vector contained a spleen focus-forming virus (SFFV) promoter replacing the

CMV promoter in the parental pHAGE-pCMV, to prevent promoter inactivation during HCMV infection by the UL122 gene product IE86 [222, 223].

Coding sequence for ORFL147C

ATGTCTCTGGCCGGCGCCAGACCGGACGACAGCGTCTCGTACGTGAGCGAGTCGAGTCATGGAGATGAATTTGTT
ACAGAAACTATGCGTAGTGTGTTCGAAATGCAACGAATACGCCATGGAGCTGGAGTGTCTAAAGTACTGCGATCC
GAACGTGTTACTGGCGGAGTCCACGCCGTTCAAGAGAAACGCGGCGGCTATAGTGTATCTGTACCGGAAGATCTA
CCCGGAGGTGGTGGCGCAGAATCGTACGCAGAGTTCGCTGCTGACTCTCTATCTGGAGATGCTGCTGAAGGCGCT
GCACGAGGATACGGCTTTGCTGGATCGGGCGCTGATGGCCTACTCGCGCCAGCCGGACCGGGCGGCCTTCTACCG
TACCGTCCTCCGTTTGGATCGCTGCGATCGCCATCACACCGTGGAGCTCCAGTTTACGGACAACGTCCGTTTCAG
CGTCAGTCTGGCCACACTCAACGACATCGAGCGCTTCCTGTGCAAAATGAACTACGTGTACGGGATCCTGGCGCC
GGAGGCCGGCCTGGAGGTCTGCGCGCAGCTGCTGGAGCTCCTCCGTCGCCTATGCGGCATCTCGCCGGTGGCGCG
TCAGGAAGTGTACGTCGAAGGGACGACATGCGCCCAATGCTACGAGGAGCTGACCATCATCCCGAATCAGGGCCG
CTCGCTGAACAAGCGGCTGCAGGGCTTGCTGTGCAACCATATAGCGGTCCACCGTCCGTCAAGCCAGTCCGATGT
GAATATCCAGACGGTGGAGCAGGACCTGCTGGACCTGACAACGCGCATCCCCCACTTGGCTGGAGTCCTTTCGGC
CCTCAAAAGCCTATTCTCTTCTTCATCGGCCTACCACAGCTACATCCAGGAGGCGGAGGAGGCGCTGAGGGAGTA
CAACCTGTTTACGGATATACCGGAACGAATATATTCCTTGTCGGATTTTACCTACTGGTCCCGTACCTCGGAGGT
TATCGTCAAGCGGGTGGGCATCACCATCCAGCAGCTAAATGTGTATCACCAGCTGTGCCGGGCGCTCATGAACGG
CATCAGTCGCCATCTGTACGGGGAGGACGTGGAGGACATCTTCGTGCTCGGGGAAAAGGCGTTGGACGGGGAGGA
GCGCATGTTCGTGGGGTCGGTCTTTGCCGCCCCCAACAGGATCATCGACCTCATCACATCCCTCAGCATTCAAGC
TTTCGAGGACAACCCGGTGTTCAACAAGCTCCACGAAAGCAACGAGATGTACACCAAAATCAAGCATATTCTCGA
GGAGATTCGACGTCCGCTGCCCGATGGCACGGGGGGCGACGGCCCCGAGGGCGAGGTTATTCACCTGCGTGGACG
GGAGGCGATGTCGGGGACGGGTACGACTTTGATGACGGCCAGCAACAGCAGCAACAGCAGTACTCACAGTCAGAG
GAA

Amino-acid sequence for ORFL147C

MSLAGARPDDSVSYVSESSHGDEFVTETMRSVFEMQRIRHGAGVSKVLRSERVTGGVHAVQEKRGGYSVSVPEDL
PGGGGAESYAEFAADSLSGDAAEGAARGYGFAGSGADGLLAPAGPGGLLPYRPPFGSLRSPSHRGAPVYGQRPFQ
RQSGHTQRHRALPVQNELRVRDPGAGGRPGGLRAAAGAPPSPMRHLAGGASGSVRRRDDMRPMLRGADHHPESGP
LAEQAAAGLAVQPYSGPPSVKPVRCEYPDGGAGPAGPDNAHPPLGWSPFGPQKPILFFIGLPQLHPGGGGGAEGV
QPVYGYTGTNIFLVGFYLLVPYLGGYRQAGGHHHPAAKCVSPAVPGAHERHQSPSVRGGRGGHLRARGKGVGRGG
AHVRGVGLCRPQQDHRPHHIPQHSSFRGQPGVQQAPRKQRDVHQNQAYSRGDSTSAARWHGGRRPRGRGYSPAWT
GGDVGDGYDFDDGQQQQQQQYSQSEE

Coding sequence for ORFS343C

ATGAAGCGGCCGTGGTGGCCGTGCTGGGCTGGGTGCATCGCTTCGAGGTTGTCGTCCGAATCGCCGGCTTGCTCC
TCTTCCAAATCTCCACGGCGGTGGCCGTCTTGGGTAGCTTCTCTCTGGTCTTCCCTACAGCGACTCTCAAGTCGC
GTCCGGGCTTTCCTTGTCACGTTGTCTGGGCGCCCGAGGTGCTCCTCCTGGTGCCTGTGGCCTCCGCGCTCTTCG
TGTATTTCCGCTATGAGCGACCGGTTCTCGCTCAGCGAAACCGGCACCCGCGCTGCCGTCGTCCGTTCCGACAGC
TGGTGCTATTACTCGCCGGCCTCCTGGCGCACATCCCGGCGC

Amino-acid sequence for ORFS343C

MKRPWWPCWAGCIASRLSSESPACSSSKSPRRWPSWVASLWSSLQRLSSRVRAFLVTLSGRPRCSSWCLWPPRSS
CISAMSDRFSLSETGTRAAVVRSDSWCYYSPASWRTSRR

Figure 3.1 | Coding and amino-acid sequences for the non-canonical uncharacterised HCMV ORFs, ORFL147C and ORFS343C

Bait sequences for five proteins were modified from canonical annotation in the GenBank entry AY446894.2. Adaptation to fibroblasts leads to mutations in RL13 and UL128 genes, as a consequence, the coding sequence for UL128 in the AY446894.2 entry contains a single nucleotide substitution that causes premature truncation by introducing an in-frame termination codon into the third exon [69]. The mutation was absent from the clinical sample (designated '742') from which Merlin was derived and thus Uniprot entry V9LLX6 does not contain the premature truncation. To clone UL128, the coding sequence from HCMV strain 6397 (EBI AFR54607.1) was used as its translation matches the protein sequence for Merlin UL128 annotated in Uniprot. Attempts to generate a stable cell line expressing the large tegument protein deneddylase UL48 failed, likely due to the length of the transgene exceeding the packaging limit of the vector. UL48 contains a predicted α-helix from residues 540-1500, but no predicted secondary structure between residues 1501-1509 (see Appendix A). The gene was thus divided into two segments, one of 4.5 kbp (1-1504 aa) terminating in a stop codon (UL48-H1), and one of 2.2 kbp (1505-2241 aa), with an additional start codon (UL48-H2), with both segments transduced in separate cell lines. Protein sequences for UL74, US14 and US17 were codon optimised by Dr. James Davies and Dr. Sepehr Seirafian (School of Medicine, Cardiff University), as the expression of annotated canonical sequences could not be detected (see Appendix B).

Reagents for detection of individual HCMV proteins are limited to a small subset of genes. To enable isolation of protein complexes bound to viral baits using AP-MS, a two amino acid linker sequence followed by a V5 epitope tag (GKPIPNPLLGLDST) was added to the C-terminus of each viral coding sequence. This 14 aa epitope was chosen as it should have minimal impact on transgene protein folding and function [224]. By using the same tag for all baits, variability among the datasets is decreased, given that the pool

of non-specific interactors that bind the epitope should be the same and thus more easily identifiable and removed from analysis.

In addition to the viral constructs, two controls were included in the interactome to promote the identification of non-specific interactors. One control was generated using the same cloning strategy for the viral baits, containing a two amino acid linker sequence followed by the V5 tag, but coding for the *A. victoria* GFP instead of a viral gene. The second control, referred to as 'no-bait' from here on, contained neither the linker sequence nor the V5 tag, consisting only of a 42 bp sequence devoid of start codons in the place of the gateway recombination cassette.

Coding sequences of all expression constructs were sequenced and compared to the Merlin strain genome, to ensure no mutations had been acquired during cloning.

3.2 Detecting the expression of recombinant constructs

Lentiviral expression constructs were stably transduced into HFFF-TERT cells and the expression of all baits was validated either by immunoblotting, mass spectrometry or PCR-based methods. Initially, lysates for all cell lines were collected as described in **2.7.1**, and analysed by immunoblotting using an anti-V5 antibody. Blots were also incubated with an anti-Calnexin antibody to control for equal protein loading among samples. Bands were detected for 130 viral baits, validating transgene expression in 75 % of cell lines (see Figure 3.2.1-3.2.2). For a subset of baits, multiple bands or a band with a molecular size differing from theoretical (see Table 3.2.1) were observed.

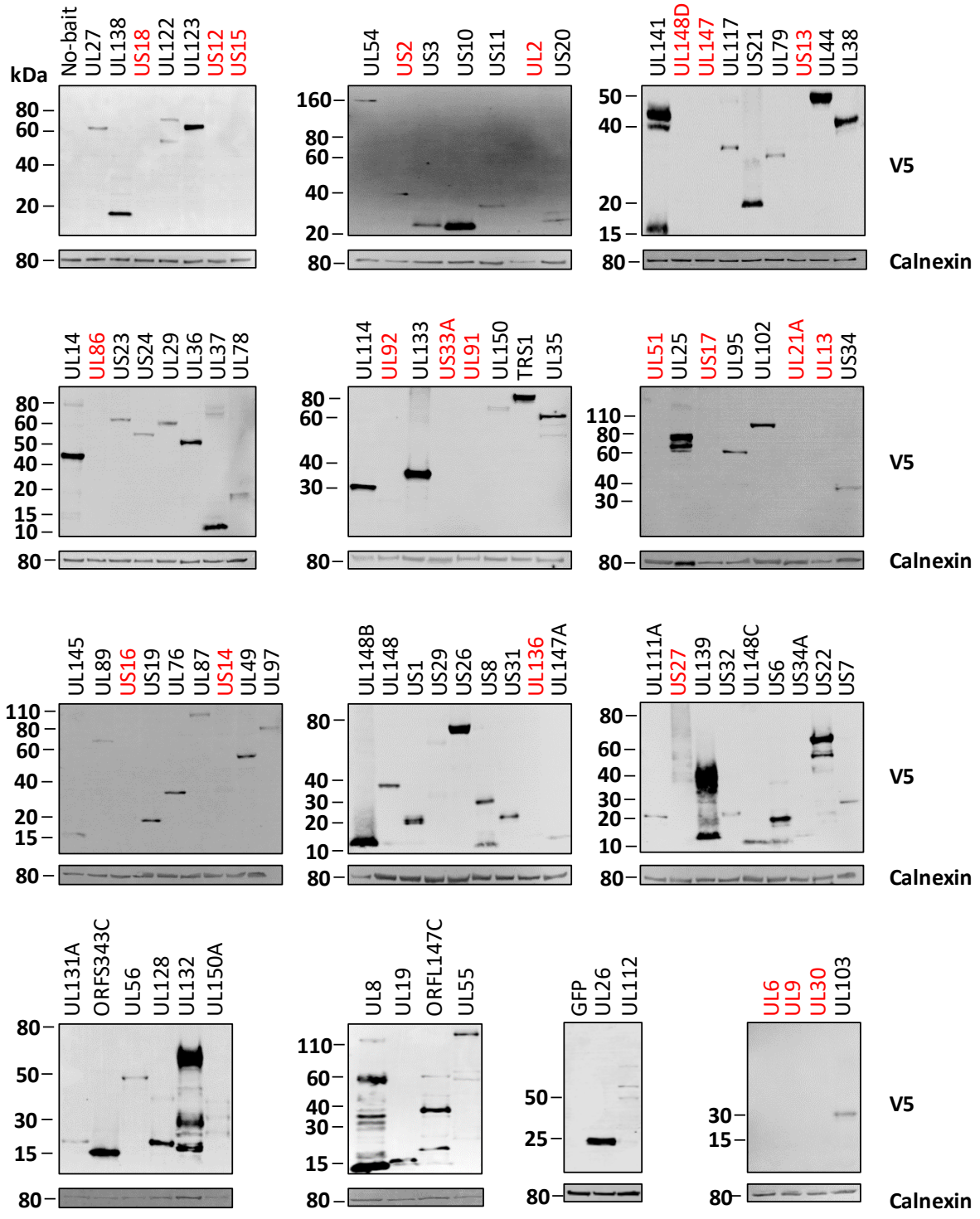


Figure 3.2.1 | Expression of V5-tagged HCMV proteins in stable cell lines
PVDF membranes were incubated with anti-V5 antibody to detect expression of viral transgenes and anti-calnexin as a control for equal protein loading among samples. Expression of proteins highlighted in red could not be confirmed by immunoblot and thus expression of these constructs was assessed using mass spectrometry.

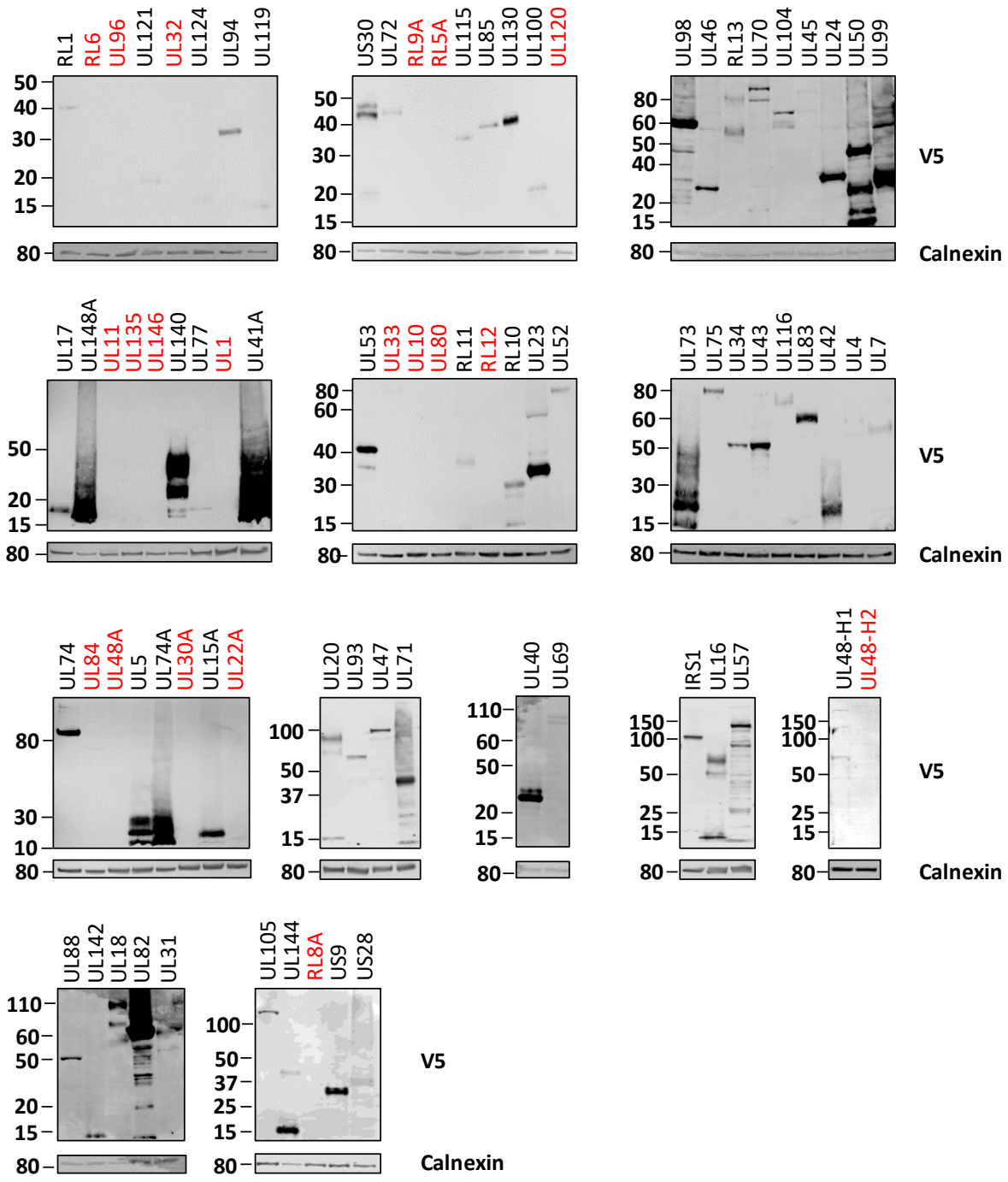


Figure 3.2.2 | Expression of V5-tagged HCMV proteins in stable cell lines
PVDF membranes were incubated with anti-V5 antibody to detect expression of viral transgenes and anti-calnexin as a control for equal protein loading among samples. Expression of proteins highlighted in red could not be confirmed by immunoblot and thus expression of these constructs was assessed using mass spectrometry.

Table 3.2.1 | Predicted molecular sizes for HCMV proteins

Gene	Size	Gene	Size	Gene	Size	Gene	Size	Gene	Size
US1	18	**UL5**	19	**UL42**	14	**UL85**	35	**UL133**	28
US2	23	**UL6**	31	**UL43**	48	**UL86**	154	**UL135**	33
US3	22	**UL7**	24	**UL44**	46	**UL87**	105	**UL136**	27
US6	21	**UL8**	36	**UL45**	102	**UL88**	48	**UL138**	19
US7	26	**UL9**	27	**UL46**	33	**UL89**	77	**UL139**	14
US8	27	**UL10**	29	**UL47**	110	**UL91**	12	**UL140**	21
US9	28	**UL11**	31	**UL48-H1**	170	**UL92**	23	**UL141**	39
US10	21	**UL13**	54	**UL48-H2**	84	**UL93**	69	**UL142**	35
US11	25	**UL14**	37	**UL48A**	8	**UL94**	38	**UL144**	20
US12	32	**UL15A**	11	**UL49**	64	**UL95**	57	**UL145**	15
US13	29	**UL16**	26	**UL50**	43	**UL96**	14	**UL146**	14
US14	34	**UL17**	13	**UL51**	17	**UL97**	78	**UL147**	19
US15	29	**UL18**	42	**UL52**	74	**UL98**	65	**UL147A**	8
US16	35	**UL19**	11	**UL53**	42	**UL99**	21	**UL148**	37
US17	32	**UL20**	38	**UL54**	137	**UL100**	43	**UL148A**	9
US18	30	**UL21A**	14	**UL55**	102	**UL102**	94	**UL148B**	9
US19	26	**UL22A**	11	**UL56**	96	**UL103**	29	**UL148C**	9
US20	29	**UL23**	33	**UL57**	134	**UL104**	79	**UL148D**	7
US21	27	**UL24**	40	**UL69**	82	**UL105**	107	**UL150**	70
US22	65	**UL25**	74	**UL70**	108	**UL111A**	20	**UL150A**	29
US23	69	**UL26**	25	**UL71**	40	**UL112**	70	**TRS1**	84
US24	58	**UL27**	69	**UL72**	43	**UL114**	28	**RL1**	35
US26	70	**UL29**	79	**UL73**	14	**UL115**	31	**RL5A**	11
US27	42	**UL30**	14	**UL74**	55	**UL116**	34	**RL6**	12
US28	41	**UL30A**	9	**UL74A**	8	**UL117**	46	**RL8A**	10
US29	51	**UL31**	66	**UL75**	84	**UL119**	39	**RL9A**	5
US30	39	**UL32**	113	**UL76**	36	**UL120**	23	**RL10**	19
US31	19	**UL33**	46	**UL77**	71	**UL121**	20	**RL11**	27
US32	22	**UL34**	45	**UL78**	47	**UL122**	63	**RL12**	47
US33A	7	**UL35**	73	**UL79**	34	**UL123**	55	**RL13**	33
US34	17	**UL36**	55	**UL80**	74	**UL124**	16	**IRS1**	91
US34A	8	**UL37**	56	**UL80.5**	38	**UL128**	20	**ORFL147C**	50
UL1	25	**UL38**	37	**UL82**	62	**UL130**	25	**ORFS343C**	13
UL2	7	**UL40**	24	**UL83**	63	**UL131A**	15		
UL4	17	**UL41A**	9	**UL84**	65	**UL132**	30		

Predicted molecular sizes (in kDa) for UL48-H1, UL48-H2, ORFL147C and ORFS343C were calculated with the ProtParam tool (Swiss Institute of Bioinformatics) [225]. Molecular sizes for all other HCMV proteins were obtained from their Uniprot entries for HCMV Merlin.

Affinity-purification mass-spectrometry was performed on 43 stable cell lines for which the detection of V5 tagged bait was unsuccessful by immunoblotting, enabling the validation of another 24 lines (see Table 3.2.2). iBAQ quantification values from previous proteomics datasets in HCMV infected cells [162, 194], show that over half of the baits which were not validated using AP-MS were also not quantified in those datasets, suggesting that even in the context of infection, the abundance of these proteins may be particularly low (see Table 3.2.2). The iBAQ quantification for these datasets incorporates

maximum MS1 precursor intensity for each peptide quantified for these proteins across 24, 48 and 72 hr or 24, 48, 72 and 96 hr of HCMV infection, while the cell lines used in the AP-MS detection method expressed the bait protein constitutively. Thus, it is unlikely that these proteins were detected due to the analysis of an inappropriate time point of expression. A possible explanation for the inability to detect expression of these baits in the stably-overexpressing cell lines is that these viral proteins may require co-expression of viral binding partners for stabilisation. Alternatively, constitutive expression outside the context of infection may lead to degradation of these baits by host factors that could be antagonised by other viral proteins in a time-dependent manner.

Expression of the 19 remaining baits was tested by PCR-based methods. For this approach, total mRNA was extracted from each cell line and reverse-transcribed into cDNA, as described in section **2.6**. Quantitative PCR was then performed on the template cDNA using custom-made primers, as a generic programme to generate small amplicons ranging from 150-200 bp. Products from the PCR reaction were then run on an agarose gel, isolated from gel bands and sequenced to verify homology to the bait's coding sequence. Alignment of sequenced PCR products and amplicon sequence are shown in Table 3.2.3, where sequence identity is highlighted in blue. Short amplicons for UL136, UL146 and UL148D yielded low-quality sequencing that could not be aligned to the coding sequence, thus primers were designed to amplify full-length PCR products. While it was possible to obtain sequences for the PCR products with a high percentage of identity relative to the coding sequence of UL146 and UL148D, this was unsuccessful for UL136, and thus this bait was removed from any further analysis. Altogether, PCR-based validation verified the expression of viral transgenes in 18 cell lines.

Table 3.2.2 | Mass spectrometry quantification of viral protein baits

Bait	Bait peptides	Normalised IBAQ experiment WCL3 (Weekes et al, Cell 2014)	Normalised IBAQ experiment 3 (Fielding et al, eLife 2017)	Average normalised IBAQ
US2	15	2.49E-04	1.23E-04	1.86E-04
US12	45	1.77E-05	4.94E-04	2.56E-04
US13	NQ	-	-	-
US14	15	-	1.93E-04	1.93E-04
US15	21	3.43E-05	4.56E-05	4.00E-05
US16	1	-	-	-
US17	5	-	-	-
US18	NQ	5.30E-05	7.06E-06	3.00E-05
US27	1	1.99E-03	1.21E-03	1.60E-03
US33A	NQ	-	-	-
UL1	NQ	-	-	-
UL2	NQ	-	-	-
UL6	NQ	-	1.09E-04	1.09E-04
UL9	NQ	1.34E-04	1.21E-04	1.27E-04
UL10	71	-	-	-
UL11	NQ	8.52E-05	4.18E-06	4.47E-05
UL13	16	2.04E-04	9.33E-04	5.69E-04
UL21A	NQ	-	-	-
UL22A	10	7.43E-03	6.74E-04	4.05E-03
UL30	NQ	1.28E-04	8.86E-06	6.84E-05
UL30A	11	-	-	-
UL32	292	7.14E-03	1.96E-02	1.34E-02
UL33	NQ	5.15E-04	4.71E-04	4.93E-04
UL48-H2	6	1.37E-03	5.97E-03	3.67E-03
UL48A	NQ	-	-	-
UL51	11	5.89E-03	1.99E-03	3.94E-03
UL80	68	1.39E-02	9.43E-03	1.17E-02
UL84	139	1.20E-02	1.96E-02	1.58E-02
UL86	61	2.48E-02	8.84E-02	5.66E-02
UL91	NQ	-	-	-
UL92	3	-	9.75E-06	9.75E-06
UL96	23	5.81E-04	1.06E-04	3.44E-04
UL120	1	-	-	-
UL135	21	5.35E-03	8.54E-03	6.95E-03
UL136	NQ	3.83E-04	1.22E-04	2.52E-04
UL146	NQ	1.52E-02	4.91E-03	1.01E-02
UL147	9	6.06E-04	2.09E-04	4.08E-04
UL148D	NQ	-	-	-
RL5A	1	-	-	-
RL6	NQ	-	-	-
RL8A	12	-	-	-
RL9A	NQ	-	-	-
RL12	NQ	6.00E-03	6.30E-03	6.15E-03

Note: NQ – Not quantified

Table 3.2.3 | Homology between sequenced PCR products and target amplicon

Bait	Forward primer sequencing (5' > 3')	Reverse primer sequencing (5' > 3')
US13	GACCGTGTTCCATCATCAGCGTATTCTACTC TACGGCTACGGCGCCATCGTCTTCCTTATGA TGACCGTGACTTTCTACGGCACGCGTTACAT CCGCGACGAACTGCCGGCTGCTCAGACGTTA CGCGGCTCGCTGCTGATCTACGTAGG	GACCGTGTTCCATCATCAGCGTATTCTACTC TACGGCTACGGCGCCATCGTCTTCCTTATGA TGACCGTGACTTTCTACGGCACGCGTTACAT CCGCGACGAACTGCCGGCTGCTCAGACGTTA CGCGGCTCGCTGCTGATCTACGTAGG
US18	CTGATCAACACCGGCATCACCGTGTGCACGG GCTTTTGCGGAGAAAGGCGCGTCATCGGTCT GTCGTTTGCCCTGGTGATGGTCTTTTTCGTT CTCTGCAGCGGTCTCACCTACCTGGCCGGCA ACAATCCCACGCGCTGGAAAGTCATCGGGAT TGGC	CTGATCAACACCGGCATCACCGTGTGCACGG GCTTTTGCGGAGAAAGGCGCGTCATCGGTCT GTCGTTTGCCCTGGTGATGGTCTTTTTCGTT CTCTGCAGCGGTCTCACCTACCTGGCCGGCA ACAATCCCACGCGCTGGAAAGTCATCGGGAT TGGC
US33A	Low-quality sequencing	GGGTTACGAGAAACTGGGATACCGCCCGCAT GCCAAACGCGTGTGGGTGCATGACCCGTTGG GATTGACGCGGTTTATCATGAGGCAACTCAT GATGTACCCGCTGGTGTTGCCGTTCACTTTT CCGTT
UL1	CACTCACGTTGGTTGGACAGCCACCGTGGTG ATAATTATCTGCGTTTTAACTTACGTTAACG TTACAACAACCCTGAAGCACAGACTACGAAC TAGAAACAACGTCAACCACACA	CACTCACGTTGGTTGGACAGCCACCGTGGTG ATAATTATCTGCGTTTTAACTTACGTTAACG TTACAACAACCCTGAAGCACAGACTACGAAC TAGAAACAACGTCAACCACACA
UL2	GAAGATGACGACGCATATCCGTCGTTCGGCA GCCTACCCGCCTCGCACGCTCAGTACGGCTT TCGACTGCTACGCAGCATATTTTTGATCATG CTTGTCATTTGGACCGCAGTGTGGCTCAAGC TGCTTCGAGACG	GAAGATGACGACGCATATCCGTCGTTCGGCA GCCTACCCGCCTCGCACGCTCAGTACGGCTT TCGACTGCTACGCAGCATATTTTTGATCATG CTTGTCATTTGGACCGCAGTGTGGCTCAAGC TGCTTCGAGACG
UL6	Low-quality sequencing	GTTCCTCGAGCGGTTCAAAGCACAACCACCG TAATGACACCCACGCTGGTTACAAACTCCAC ATTCAGTGTGTCACTTGTTGCGTTGAGACTG ACGACAAATTCCAGCGCGTTTG
UL9	TGAAGCACAACACGACACTGCCACTTCACAT ACAATGTGGATCATACCCCTAGTTATCGTTA TAACAATCATCGTTTTAATTTGTTTCAAGTT CCCCCAAAAAGCTTGGAATAAATTCACACAA TACCGATACAGCG	TGAAGCACAACACGACACTGCCACTTCACAT ACAATGTGGATCATACCCCTAGTTATCGTTA TAACAATCATCGTTTTAATTTGTTTCAAGTT CCCCCAAAAAGCTTGGAATAAATTCACACAA TACCGATACAGCG
UL11	CAACCACCACGAGAACAACCACCACCGCCAA GAAGACGACGATAAGCACTACCCATCATAAA CACCCCAGTCCCAAAAAATCCACCACCCCTA ACAGTCACGTAGAACATCACGTTGGTTTTGA AGCCACAGCAGCGGAAACACCGTTACAACCA AGCC	CAACCACCACGAGAACAACCACCACCGCCAA GAAGACGACGATAAGCACTACCCATCATAAA CACCCCAGTCCCAAAAAATCCACCACCCCTA ACAGTCACGTAGAACATCACGTTGGTTTTGA AGCCACAGCAGCGGAAACACCGTTACAACCA AGCC
UL21A	GTCGGTGAGGGAGATGAAGAGATGTTGCCGG ATCTGCCGATGGAGATCGACATCGTCATCGA CCGACCTCCGCAGCAACCCCTACCCAATCCG CTGGTGCTACTGCTGGACGATGTTCCCCCCC ATGTACCCGGTTTTGCTC	GTCGGTGAGGGAGATGAAGAGATGTTGCCGG ATCTGCCGATGGAGATCGACATCGTCATCGA CCGACCTCCGCAGCAACCCCTACCCAATCCG CTGGTGCTACTGCTGGACGATGTTCCCCCCC ATGTACCCGGTTTTGCTC
UL30	CCGGGGCATGATGGACTATCACGACGGGCTC TCGCGCCGTCAACAGCGTGCCTTTTGCCGCG CGGGTCGCGTGTTGACGGACCCGGAGCCCAT CCAGAGCGAGACGGAGGGGGAGAATAAACAG TTTACGGAGCACACACACAAAGTAGTCTCGT TTTTTATTAAAAGTGTCTTTGTATTTCCCTG TCTTGTGTTGCC	CCGGGGCATGATGGACTATCACGACGGGCTC TCGCGCCGTCAACAGCGTGCCTTTTGCCGCG CGGGTCGCGTGTTGACGGACCCGGAGCCCAT CCAGAGCGAGACGGAGGGGGAGAATAAACAG TTTACGGAGCACACACACAAAGTAGTCTCGT TTTTTATTAAAAGTGTCTTTGTATTTCCCTG TCTTGTGTTGCC
UL33	TACTCGAGCTGCACAGTAGGCTTTGCCACCG TAGCCCTGATCGCCGCCGACCGATACCGCGT TCTTCATAAGCGTACCTACGCGCGGCAGTCG TACCGCTCCACCTATATAATTTTGCTATTGA CCTGGTTTGCCGGGCTGATCTTTTCCATG	Low-quality sequencing

UL48A	GACCGAGATCTCAGAGGCCACCCACCCGGTG CTGGCCACCATGCTGAGCAAGTATACGCGCA TGTCCAGTCTGTTTAACGACAAGTGCGCCTT TAAGCTGGACCTGTTGCGCATGATAG	GACCGAGATCTCAGAGGCCACCCACCCGGTG CTGGCCACCATGCTGAGCAAGTATACGCGCA TGTCCAGTCTGTTTAACGACAAGTGCGCCTT TAAGCTGGACCTGTTGCGCATGATAG
UL91	CTTTGTCGACCGCCTCTTTCAACACTTTTCC TTCCTTTTCCAGGCCGAGGAGTCAGGCCCGC GCCGCTTGGAACTGGTCGCGTCCGTGTTCGA GCACCTGACGGTGGAGTGCGTCAACGACATC CTGGACGCCTGCAGCCACCCGGACGTGAACG TCGTGGAGACAAGCAACACCTG	CTTTGTCGACCGCCTCTTTCAACACTTTTCC TTCCTTTTCCAGGCCGAGGAGTCAGGCCCGC GCCGCTTGGAACTGGTCGCGTCCGTGTTCGA GCACCTGACGGTGGAGTGCGTCAACGACATC CTGGACGCCTGCAGCCACCCGGACGTGAACG TCGTGGAGACAAGCAACACCTG
UL136	Low-quality sequencing	Low-quality sequencing
UL146	ATGCGATTAATTTTTGGTGCGTTGATTATTT CTTTAACGTATATGTATTATTATGAAGTGCA TGGAACGGAATTACGCTGCAAATGTCTTGAT GGTAAAAAACTGCCGCCCAAAACAATTATGT TGGGTAATTTTTGGTTTCATCGCGAATCTGG TGGTCCCAGATGCAATAACAATGAATATTTC TTGTATCTAGGCGGAGGAAAAAAACATGGAC CTGGAGTATGTTTATCGCCCCATCACCCTTT TTCAAAATGGCTAGACAAACGCAACGATAAC AGGTGGTATAATGTTAATGTAACAAGACAAC CGGAACGAGGGCCGGGAAAAATAACTGTAAC CCTAGTAGGTCTGAAGGAA	ATGCGATTAATTTTTGGTGCGTTGATTATTT CTTTAACGTATATGTATTATTATGAAGTGCA TGGAACGGAATTACGCTGCAAATGTCTTGAT GGTAAAAAACTGCCGCCCAAAACAATTATGT TGGGTAATTTTTGGTTTCATCGCGAATCTGG TGGTCCCAGATGCAATAACAATGAATATTTC TTGTATCTAGGCGGAGGAAAAAAACATGGAC CTGGAGTATGTTTATCGCCCCATCACCCTTT TTCAAAATGGCTAGACAAACGCAACGATAAC AGGTGGTATAATGTTAATGTAACAAGACAAC CGGAACGAGGGCCGGGAAAAATAACTGTAAC CCTAGTAGGTCTGAAGGAA
UL148D	ATGACGGCGCCCAAGTGTGTCACCACCACGA CCTATCTGGTCAAGACCAAGGAGCAGCCCTG GTGGCCCGACAACGCCATCAGGAGATGGTGG ATCAGCGTTGCCATCGTCATCTTCATCGGAG TCTGTCTGGTGGCCCTGATGTACTTTACGCA GCAGCAGGCACGCAACGGGAGCGGCAGCGGC	ATGACGGCGCCCAAGTGTGTCACCACCACGA CCTATCTGGTCAAGACCAAGGAGCAGCCCTG GTGGCCCGACAACGCCATCAGGAGATGGTGG ATCAGCGTTGCCATCGTCATCTTCATCGGAG TCTGTCTGGTGGCCCTGATGTACTTTACGCA GCAGCAGGCACGCAACGGGAGCGGCAGCGGC
RL6	CTAAAAGCGACGACTGGGAGTAATTTTACCA TTACGCATAGGAAAGATCCGTTGACAACTAA GTGGAAAACCGTTTTTGGTAACAATGGTGAT CAGTGGTTGTGCAACGTTACGGGTATAGGTA ATGCTACTGTGAATGGTAACGCAACTATTTG TGTGTCGAGCTGTGGTCATA	CTAAAAGCGACGACTGGGAGTAATTTTACCA TTACGCATAGGAAAGATCCGTTGACAACTAA GTGGAAAACCGTTTTTGGTAACAATGGTGAT CAGTGGTTGTGCAACGTTACGGGTATAGGTA ATGCTACTGTGAATGGTAACGCAACTATTTG TGTGTCGAGCTGTGGTCATA
RL9A	ATGTCTCTAGATGCCGCCAGCCACCAACCGG CGGCACGGCGGCTCTTGGATTCGGCATTGGT GCGCCGCGTCTTGGCCTGCATGATCATCGTC ATCATGATCGCCATTAGCATCTGGATCCTGA CCTACGTGCTGTTTCTCTAA	ATGTCTCTAGATGCCGCCAGCCACCAACCGG CGGCACGGCGGCTCTTGGATTCGGCATTGGT GCGCCGCGTCTTGGCCTGCATGATCATCGTC ATCATGATCGCCATTAGCATCTGGATCCTGA CCTACGTGCTGTTTCTCTAA
RL12	CACCCCACTATGTCCCAGATACGTAGGAACA CAATCAGAAGAAGACGAAGACGACGATTATA CACTAAGCACTATCACAAATAATAACATGCG CAAAACAAGTCACCGTGACATCTCACATGGC ACGCGCACTACATGGGCTCTTAC	CACCCCACTATGTCCCAGATACGTAGGAACA CAATCAGAAGAAGACGAAGACGACGATTATA CACTAAGCACTATCACAAATAATAACATGCG CAAAACAAGTCACCGTGACATCTCACATGGC ACGCGCACTACATGGGCTCTTAC

Note: homology between sequenced PCR product and amplicon was determined by Basic local alignment search tool [219] and is displayed in blue.

3.3 Generation of HCMV stock

In order to minimise variability, for the interactome all stable cell lines were infected with the same stock of HCMV strain Merlin. To generate the required amount of virus, aliquots of 17 supernatant harvests from 7 independent viral propagation experiments were pooled immediately prior to infection. To titrate the combined virus pool, IE1 (UL123) expression of cells infected for 24 h with serial dilutions of the stock was estimated by flow cytometry (see Figure 3.3.1). For the 1:64 dilution, a high degree of cell death was observed compared to other dilutions and thus this data was omitted from the calculation of the stock titre, to avoid an underestimation of the titre. The percentage of infected cells in the 1:16384 dilution (single sample) was below 1 %, and thus too small to reflect an accurate titre. Using an average of the three remaining serial dilutions (in duplicate), the combined stock was estimated to be $4.8x10^7$ IE1-forming units/ mL.

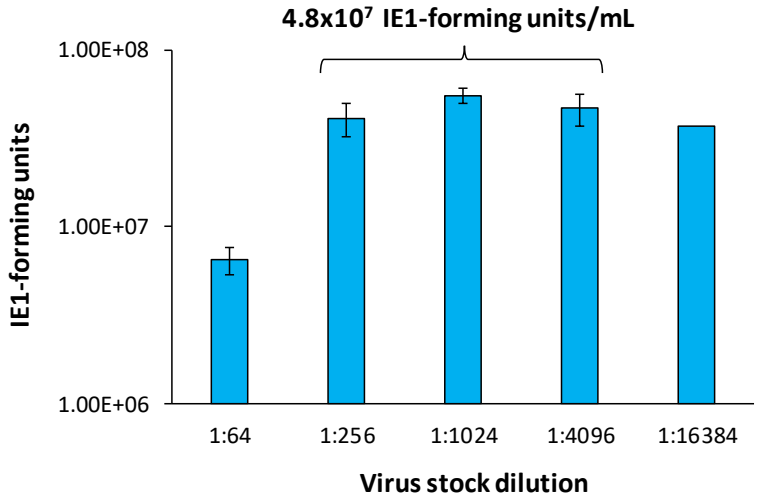


Figure 3.3.1 | Titration of the interactome virus stock using IE1 expression
HFFF-TERT cells were infected with serial dilutions (in duplicate) of the combined HCMV virus stock and IE1 expression was analysed by flow cytometry at 24 h post-infection. An average titre for the stock was estimated using values from the 1:256, 1:1024 and 1:4096 dilutions.

3.4 Infection conditions

Single AP-MS experiments can yield lists of hundreds to thousands of putative interactors, making it difficult to distinguish between contaminants and true interactors. Overlap of individual lists of putative interactors enables the identification of common contaminants, which can be enhanced by increasing sample size. Using similar experimental conditions, such as a single time-point post-infection to harvest all samples, inherently makes the interactor lists more comparable.

Expression of HCMV proteins occurs with different temporal profiles, according to the roles they play in the viral life cycle. Proteins from all temporal classes are expressed to some extent at 60 h post-infection (see Figure 3.4.1), thus harvest of lysates from infected cells was performed at this time-point.

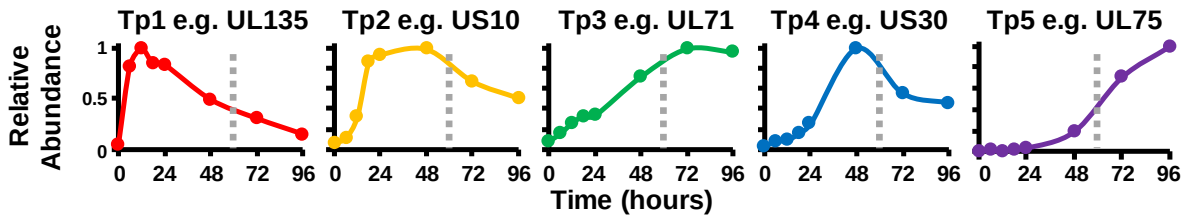


Figure 3.4.1 | Temporal classes of HCMV protein expression (Adapted from Weekes *et al,* 2014). Example profiles of protein abundance throughout the course of HCMV infection for temporal classes Tp1 to Tp5. At 60 h post-infection (grey dotted line), peptides from 139/139 quantified canonical HCMV proteins and 14/14 quantified non-canonical ORFs were detected.

Dexamethasone is a glucocorticoid that has been shown to enhance HCMV replication, increasing synthesis of immediate early proteins and HCMV DNA [198]. Additionally, serum starvation by pre-incubation with serum-free media similarly increases the percentage of cells infected by inducing cell cycle synchronisation.

In order to maximise the percentage of infection that could be obtained with our viral stock, HFFF-TERT cells were pre-treated with 4 µg/mL dexamethasone in serum-free media for 24 h, prior to infection with HCMV for 60 h. For MOI 0.5, and according to

the Poisson distribution that is applied to calculate the probability that a cell will absorb a number of virus particles when inoculated at a specific MOI, it would be expected that approximately 39 % of cells would be infected. However, in this experiment, 66 % of cells were IE1-positive. This observation may be explained by the effect of serum-starvation in enhancing viral infection, or alternatively by an underestimation of the titre of the viral stock.

Incubation with Dexamethasone resulted in a considerable increase in percentage infection, from 66 to 92 % (see Figure 3.4.2), thus dexamethasone pre-treatment was performed prior to infection of the interactome stable cell lines.

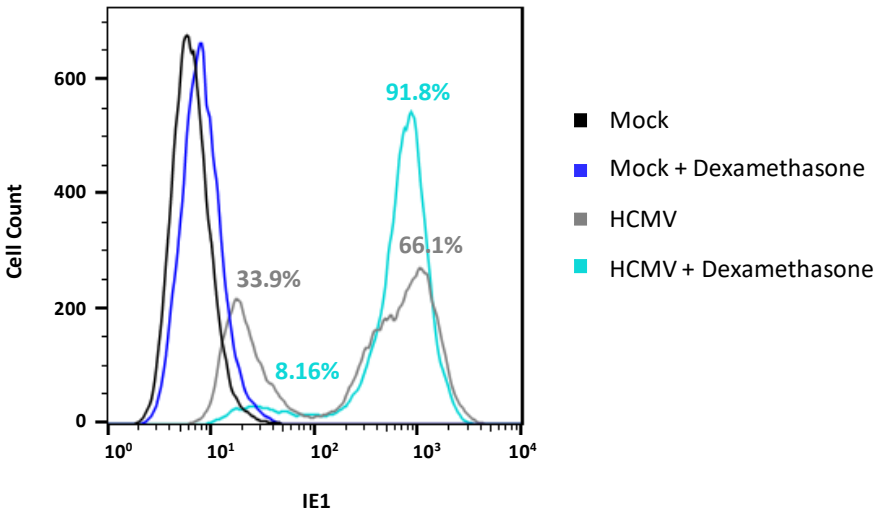


Figure 3.4.2 | Dexamethasone treatment augments viral infection in HFFF-TERT cells. HFFF-TERT cells were incubated O/N in serum-free DMEM containing 4 µg/ml Dexamethasone, 24 h after seeding. Cells were then mock or HCMV infected (MOI 0.5) for 24 h and collected for IE1 expression analysis by flow-cytometry.

Finally, an estimation of the percentage infection obtained using the exact volume of viral stock for each cell line was determined by flow cytometry. Using cell-surface downregulation of MHC Class I molecules as well as IE1 expression in cells pre-treated with dexamethasone, followed by infection with the combined virus stock for 60 h,

showed that approximately 76 % cells were infected in these conditions (see Figure 3.4.3).

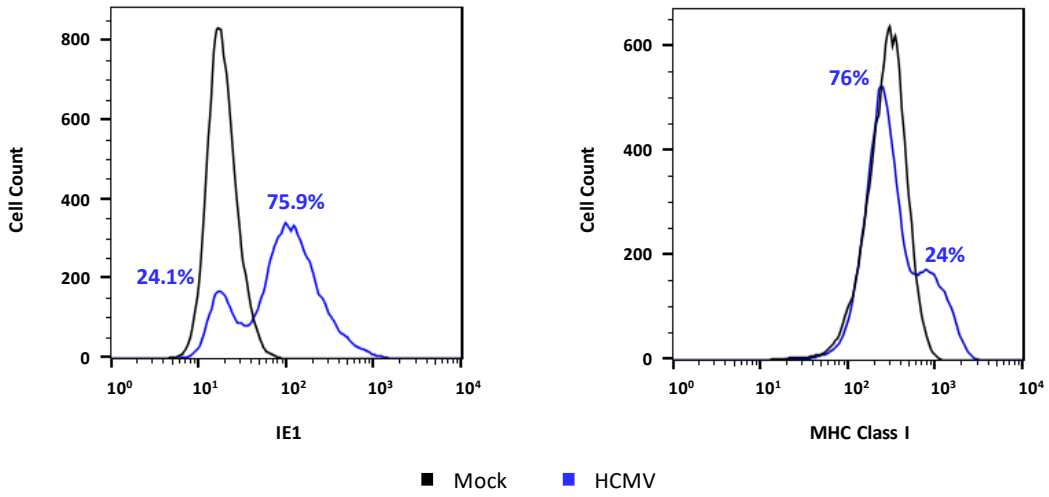


Figure 3.4.3 | IE1 expression and cell surface downregulation of MHC-1 as indicators of percentage infection in a cell population. HFFF-TERT cells were incubated O/N in serum-free DMEM containing 4 µg/ml Dexamethasone, 24 h after seeding. Cells were then mock or HCMV infected for 60 h and collected for analysis of expression of IE1 and MHC Class I molecules by flow-cytometry.

3.5 Discussion

To generate an HCMV interactome in the context of viral infection, two approaches could be employed: (a) infecting a multitude of cell lines individually expressing each tagged viral protein with one strain of HCMV; (b) infecting one cell line with multiple HCMV recombinants separately, with each recombinant virus containing a tag at the extremity of a different viral gene. The latter strategy may add variability to each AP-MS dataset, as HCMV canonical protein-coding genes, in addition to uncharacterised ORFs, overlap in their coding sequences. Thus, the addition of tags to the viral genome could disrupt different genes in each dataset, resulting in changes in viral protein expression, and a possible inability to identify a subset of viral protein-protein interactions. Combining tandem AP-MS datasets enabled the identification of common contaminants by comparing and scoring putative interactors among each dataset, while also increasing

confidence in the identification of true binding partners. This scoring is less effective if non-specific interactors vary in an inconsistent manner throughout the datasets, resulting in lower confidence/lower scores for true-positive interactions.

Even though expression of all viral baits was driven under the same promoter, the level of each transgene expression was quite variable (see Figure 3.2.1-3.2.2). Part of this variability could be attributed to differential protein turnover, the balance between protein synthesis and degradation. Difficulty in detecting bait expression in uninfected lysates may be due to a requirement for other viral factors for their stabilisation.

Detection of most members of the US12-21 family was only possible using mass spectrometry, suggesting that these highly hydrophobic multi-transmembrane spanning proteins may have been expressed at low level, or poorly solubilised.

The molecular weight of many constructs differed from the theoretical value. These differences can be due to post-translational editing. While modifications such as glycosylation (for example UL7 has an expected molecular weight of 24 kDa based on its amino acid sequence, yet it has been reported to be highly glycosylated and was observed approximately 55 kDa) can explain a higher molecular weight, protein cleavage into multiple forms can explain the observation of several bands of a lower weight.

Expression of 23 V5-tagged transgenes was detected by MS but not by immunoblot. Detection by immunoblot is solely dependent on the binding of an antibody to an epitope, while quantification by MS is achieved orthogonally by identification of any peptide derived from tryptic digest, which may have greater overall sensitivity. A particular advantage of the use of MS includes sequence-based validation of the protein detected, since all baits that were detected only by MS were quantified by unique peptides or peptides that were redundant among viral ORFs and thus could not have derived from a cellular protein.

Detection of 18 V5-tagged transgenes was only possible by PCR-based methods. Previously published proteomics data show that over half of these baits weren't quantified by MS in HCMV infected HFFF-TERT cells either, which could be due to a very low abundance, or small size which would mean that few tryptic peptides were available for analysis. It has been found that certain baits are co-degraded with host proteins they target (for example, UL138 [159] and UL145 [163]), which would further limit bait expression.

UL136 is expressed as five protein isoforms, yet transgene expression was not validated by any of the methods employed. iBAQ quantification in previous proteomics datasets (see Table 3.2.2) showed that although these gene products were not highly abundant, they were within the limits of detection. Codon optimisation of the canonical sequence might increase the translation efficiency of this gene. Nevertheless, the inability to validate UL136-V5 expression by any of the methods employed led to its exclusion from the interactome.

4 | Optimising the immunoprecipitation protocol

Optimization of the liquid chromatography gradient and injection of samples in the Orbitrap Fusion Lumos Tribrid Mass spectrometer was performed by Dr. Robin Antrobus (Cambridge Institute for Medical Research, University of Cambridge). The starting protocol for immunoprecipitation used for optimisation throughout this chapter was developed by Professor Steven Gygi's laboratory (Department of Cell Biology, Harvard Medical School).

The immunoprecipitation protocol used by Huttlin *et al* (2015) for the human interactome project was adapted for a smaller scale throughput [189]. The adapted protocol was then optimised to ensure a robust peptide quantification.

Cell lines expressing viral proteins were chosen for use in optimisation of the protocol based on (a) transgene solubility (absence of transmembrane domains), (b) length and abundance (>15 kDa to avoid excessively small proteins, and proteins easily detected by immunoblot), and (c) having interactors reported in the literature, which could be used as positive controls. UL27, UL54 and UL123 fitted these criteria. Multiple steps of the protocol (see Figure 4) were then optimised, and optimal conditions were determined based on the number of bait, positive control and overall peptides as well as the number of protein identifications (IDs) quantified in each sample. None of the experiments discussed in this chapter were performed in the context of HCMV infection to simplify the experimental workflow.

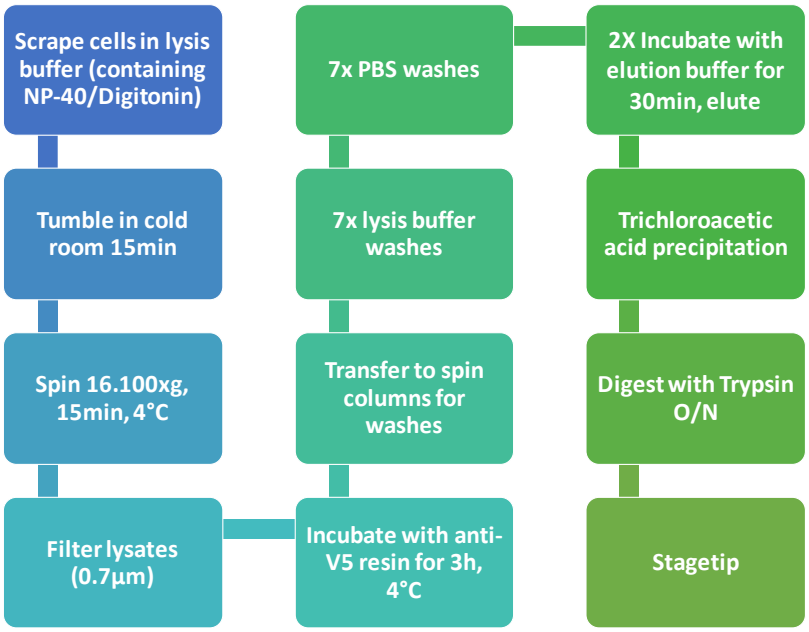


Figure 4 | Overview of sample processing for affinity-purification mass-spectrometry

4.1 Anti-V5 agarose beads

In the original protocol from Huttlin et al (2015), 60 µL of a 50 % slurry of immobilized mouse monoclonal anti-HA agarose resin (Sigma-Aldrich), that was pre-washed in lysis buffer (without protease inhibitor), was added to each lysate (comprised of five confluent 10 cm^2 dishes) [189]. Given that the majority of baits for the HCMV interactome were cloned from recombinant adenoviral vectors containing the viral coding sequences followed by a V5 tag, this similarly small epitope was chosen for this interactome instead of the HA tag. Thus, it was necessary to select an anti-V5 agarose resin for affinity purification.

Agarose beads conjugated to anti-V5 antibody from two different suppliers, Sigma-Aldrich and Abcam, were tested for optimal isolation of the V5-tagged bait. Resin from the Sigma-Aldrich product was conjugated to approximately 2 mg of mouse monoclonal (clone V5-10) per mL of bead volume. The Abcam product contained a goat polyclonal anti-V5 antibody attached to the agarose at a ratio of 500 µg of antibody per

mL of bead. According to the manufacturer's instructions, both antibodies were coupled to the agarose using cyanogen bromide and supplied as a 50 % slurry.

In this optimisation experiment the most important measurement was the number of bait peptides as this provided a measurement of affinity between the anti-V5 agarose beads and the V5-tagged bait, reflecting an enrichment in the protein of interest. The UL54-expressing cell line was chosen for this purpose given that the catalytic subunit of the DNA polymerase is a large protein, capable of yielding relatively abundant peptides. Lysates from two confluent 15 cm^2 dishes were incubated with 60 µL of anti-V5 agarose from either supplier and processed for affinity purification as described in section **2.9.1**. Two confluent 15 cm^2 dishes were used instead of five 10 cm^2 dishes as this provided an equivalent surface area and cell density and facilitate sample handling.

Immunoprecipitation with anti-V5 beads from Sigma-Aldrich yielded more bait peptides, suggesting that the overall sensitivity to detect specific interactions would also increase (see Figure 4.1). The total number of peptides and proteins was considered at best a secondary measure of performance, since this also included known and possible contaminants. The difference in overall sensitivity may be partly due to the ratio of anti-V5 antibody conjugated to the resin. Yet, given the difference in volume of slurry provided and cost, Sigma-Aldrich's anti-V5 agarose proved to be more cost-effective for this purpose and hence was used for all consecutive experiments.

To reduce variability in sample preparation, ten vials of the same batch of anti-V5 resin were combined and the same pool (a total of 20 mL) was used for all interactome samples.

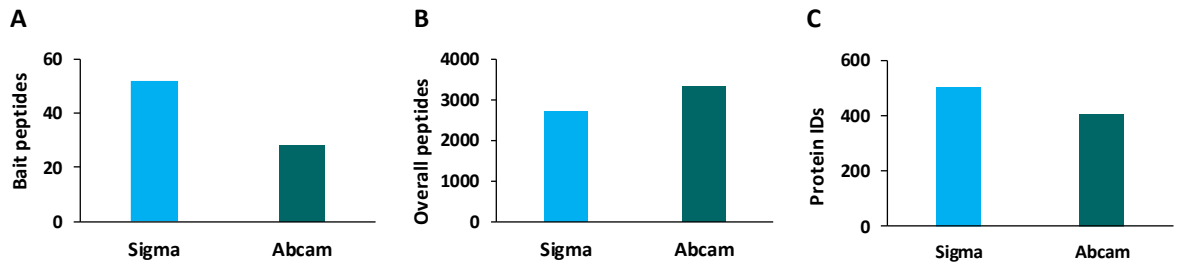


Figure 4.1 | Comparing Sigma-Aldrich and Abcam's anti-V5 agarose binding capacity
Numbers of UL54-V5 peptides (**A**), overall quantified peptides (**B**) and protein IDs (**C**) obtained from immunoprecipitation with anti-V5 resins from either Sigma-Aldrich (Sigma) or Abcam. This figure is representative of n= 1 experiment.

4.2 Bead volume per sample

Upon selection of the Sigma-Aldrich anti-V5 resin for affinity purification, a titration of the agarose slurry was performed to determine the optimal bead volume per IP. In the original protocol 60 µL of immobilized agarose resin were added to each lysate. For this titration, three different volumes of anti-V5 conjugated agarose used per sample (30, 60 or 120 µL) were compared. As in **4.1**, each sample comprised lysates from two confluent 15cm^2 dishes. Samples were processed for affinity purification as described in section **2.9.1**.

The cell line expressing UL27 was chosen for this experiment, as the human interactors that have been described in the literature for this protein could provide an additional measure of sensitivity in addition to the number of bait peptides. UL27 promotes cell cycle arrest in G_0/G_1 by targeting the host histone acetyltransferase Tip60/KAT5 to the proteasome for degradation. Previous AP-MS data in primary human foreskin fibroblasts infected with a recombinant HCMV encoding UL27-FLAG, has suggested that UL27 interacts with members of the Tip60 acetyltransferase complex, 26S proteasome subunits, proteins involved in ubiquitin E3 ligase complexes, among others [226]. Table 4.2 shows 27 suggested putative interactors of UL27 from this AP-MS dataset

that were quantified by at least 2 peptides, and Tip60 which was identified as an interactor using co-IP with an antibody to the endogenous protein in cells infected with the recombinant HCMV (AD169 background) encoding UL27-FLAG. Similarly, this study also validated the interactions with TRRAP, EP400 and PSME3 using the same co-IP experimental setup. However, these interactions have not been reported in any further studies.

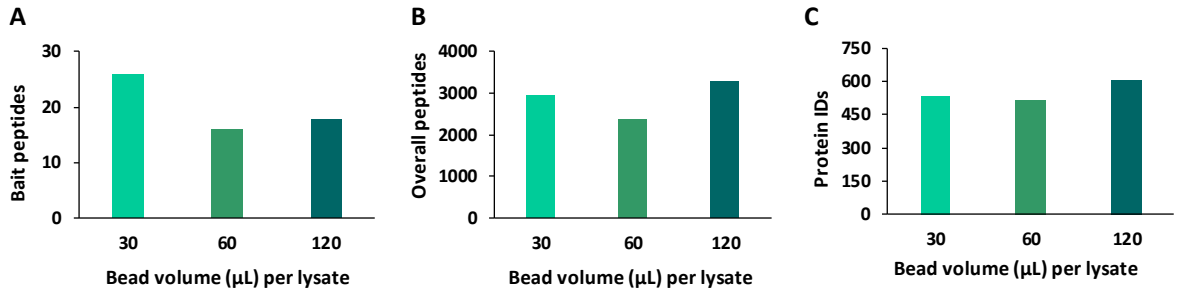
Table 4.2 | UL27-interacting proteins identified in Reitsma et al (2011) [226]

UL27 Interactors						
TRRAP	ACTL6A	PSMC4	PSMD3	UBR5	RAN	NUDT21
EP400	STAT3	PSMD14	PSMA3	DDB1	HNRNPH3	TMEM43
RUVBL1	Tip60/KAT5	PSMD2	PSMB6	WDR26	CBR1	RAB1A
RUVBL2	PSME3	PSMC6	PSMB4	ACLY	FKBP10	PDIA4

Increasing the volume of resin did not lead to an increase in peptide numbers or protein IDs, suggesting that 30 µl of beads provided an excess of V5 binding sites (see Figure 4.2). Furthermore, the total number of peptides from known UL27 interactors also did not increase.

The number of peptides per known interactor of UL27 was quite variable, although this pattern was evened out by assessing the total number of peptides from all known interactors. In fact, only 10/27 known interactors of UL27 were detected in this experiment, with just PSMD2, RAN, EP400 and PSME3 being co-purified in all samples. This may reflect differences in the experimental conditions such as cell type and affinity-purification in the context infection instead of transgene bait overexpression.

Given that there was no substantial gain in using more than 30 µl per sample, all immunoprecipitation experiments onwards were performed using this volume of resin.



D

UL27 Interactor	30µL	60µL	120µL
RAB1A	0	1	0
KAT5	0	0	1
DDB1	0	1	0
TRRAP	0	0	1
ACLY	1	8	3
PSMD3	1	2	0
PSMD2	1	2	1
RAN	3	1	1
EP400	5	1	13
PSME3	15	7	11
Interactor peptide total	26	23	31
Positive controls identified	6	8	7

Figure 4.2 | Titration of Anti-V5 resin
Numbers of UL27-V5 peptides (**A**), overall quantified peptides (**B**) and protein IDs (**C**) obtained from immunoprecipitation with three different amounts (30, 60 or 120 µl per lysate) of anti-V5 resin. **D**) Numbers of peptides from known UL27 interactors quantified for each condition. This figure is representative of n= 1 experiment.

4.3 Input material per immunoprecipitation reaction

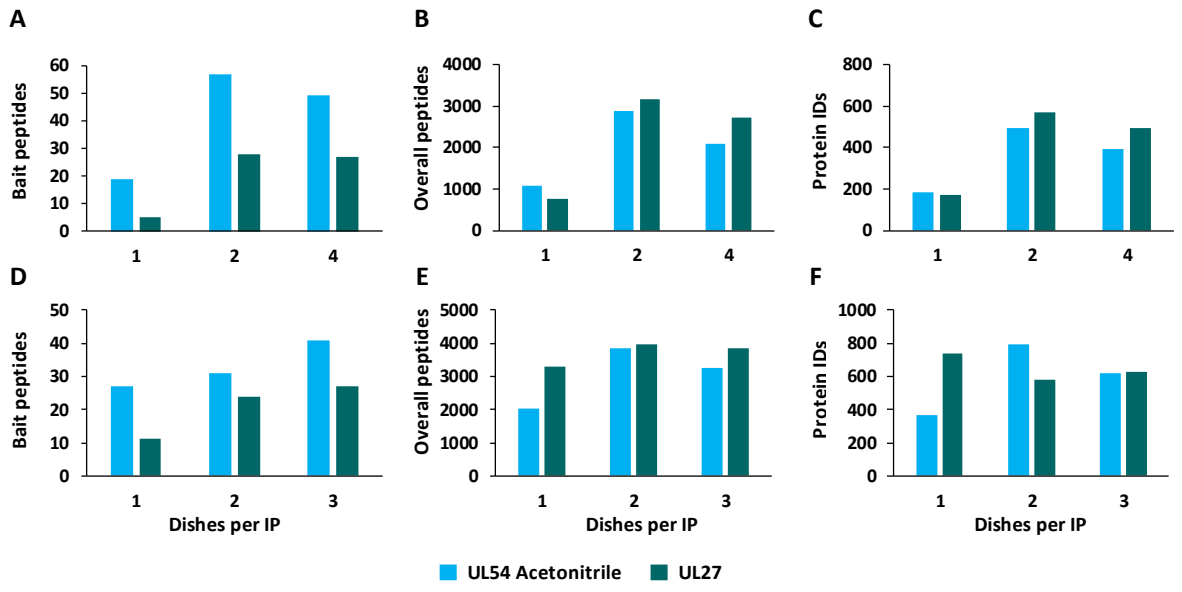
In the previous optimisation steps, two confluent 15 cm^2 dishes were used per immunoprecipitation reaction to provide an equivalent surface area and cell density to the original protocol.

The next optimisation step aimed to determine the optimal amount of protein added to each immunoprecipitation reaction. For this purpose, lysates from one, two or four confluent 15 cm^2 tissue culture dishes of UL27 or UL54-V5 expressing stable cell lines were compared. As determined in **4.2**, 30 µl of anti-V5 agarose was used per lysate. Consequently, 30 µl, 60 µl or 120 µl were used for the lysates of one, two or four confluent 15 cm^2 tissue culture dishes, respectively. Samples were processed for affinity purification as described in section **2.9.1**.

The number of bait peptides, overall peptides and protein IDs increased two-fold between lysates from one and two dishes. However, this direct proportionality was not observed when using lysates from four dishes, (see Figure 4.3 A-C). A second experiment, comparing the protein contents of one, two and three confluent 15 cm^2 tissue culture dishes yielded similar results (see Figure 4.3.D-F), which were supported by the quantification of known interactors of UL27 (see Figure 4.3.G-H).

Similarly to the results in **4.2**, the number of peptides per known interactor of UL27 was quite variable, with a substantial number of these interactors being detected inconsistently across the AP-MS samples. In this experiment, PSMD2, RAN and EP400 were detected in 4/5 IP reactions and PSME3 was the only known interactor of UL27 being co-purified in all samples.

Increasing the amount of input material beyond two 15 cm^2 dishes did not result in an increase of bait peptide or positive interactors suggesting that maximum resin binding capacity might have been achieved. Using both a higher volume of beads and more lysate could improve quantification, however this would hinder the throughput of the project as it would have increased the number of dishes necessary per cell line and reduce the number of cell lines that could be infected in parallel. Most importantly, it would have substantially increased the amount of virus required to infect these cells. Therefore, lysates from two confluent 15 cm^2 tissue culture dishes were used in all immunoprecipitations from here onwards.



G

UL27 Interactor	1 dish	2 dishes	4 dishes
PSMC4	0	1	0
FKBP10	0	4	4
PSMA3	0	1	1
HNRNPH3	0	6	7
EP400	0	4	9
ACLY	0	5	9
RAN	0	4	4
PSMD2	0	2	2
PSMB6	0	1	0
ACTL6A	0	1	4
DDB1	0	1	1
PSME3	2	7	8
Interactor peptide total	2	37	49
Positive controls identified	1	12	10

H

UL27 Interactor	1 dish	2 dishes	3 dishes
PSMC4	0	2	0
DDB1	0	3	1
RAN	0	0	1
PSMB6	0	1	0
CBR1	1	0	1
PSMC6	1	2	0
PSMA3	1	0	1
RAB1A	1	2	0
PSMD2	1	2	1
ACTL6A	1	3	5
RUVBL1	2	0	0
RUVBL2	2	0	0
NUDT21	2	0	3
PSMB4	2	0	0
PDIA4	2	1	0
FKBP10	4	7	5
HNRNPH3	6	10	8
EP400	7	9	17
ACLY	8	7	4
PSME3	9	19	20
Interactor peptide total	50	68	67
Positive controls identified	16	13	12

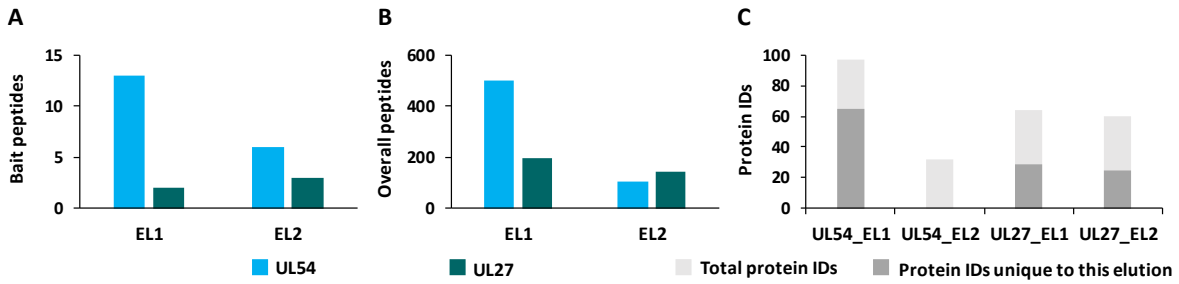
Figure 4.3 | Assessing immunoprecipitation input material
Two stable cell lines overexpressing UL27-V5 and UL54-V5 were used in two independent experiments with different amounts of input material (1, 2, 3, 4x confluent 15 cm^2 dishes per reaction). Numbers of peptides (**A, D**), overall quantified peptides (**B, E**), protein ID's (**C, F**) and UL27 known interactors (**G, H**) obtained for these conditions are shown. This figure is representative of n= 2 experiments.

4.4 Peptide and protein yields of V5 elutions

The original protocol from Huttlin et al (2015) employed two elutions with 200 µL of 250 µg/mL HA peptide in PBS at 37 °C with agitation for 30 min, to elute protein complexes [189]. Both elutions were then combined before protein precipitation with Trichloroacetic acid. For the HCMV interactome, the concentration of eluting peptide (in this case V5) and elution conditions were kept the same.

In order to test whether both elutions with V5 peptide provided comparable peptide and protein yields, lysates from UL27 and UL54-V5 expressing cell lines were used to compare the first (EL1) and second (EL2) elution of protein complexes from the anti-V5 resin. For each lysate, instead of combining both elutions as per the initial protocol, these were processed separately (as described in **2.9.1**) and analysed on the mass spectrometer as two separate samples.

This experiment showed that a variable amount of protein complexes were eluted with EL2 across the two baits, ranging from approximately ¼ to ½ of the number of peptides compared to the total sum of both elutions (see Figure 4.4A, B and D). Furthermore, the number of proteins IDs that were unique to each elution also varied between the two baits. In the case of UL54-V5, there were no unique protein IDs in the second elution. For UL27-V5, approximately half of the protein IDs in elution 1 and 2 were unique, while the other half comprised proteins that were present in both elutions (see Figure 4.4C and D). Thus, similarly to the initial protocol by Huttlin et al, two elutions were used and combined for all interactome samples.



D

	UL54		UL27	
	EL1	EL2	EL1	EL2
% Total peptides	76%	24%	47%	53%
% Bait peptides	68%	32%	40%	60%
% Unique protein IDs	67%	0%	45%	39%

Figure 4.4 | Assessing efficiency of V5 elutions
Numbers of UL27-V5 and UL54-V5 bait peptides (**A**) and overall quantified peptides (**B**). Total protein IDs and the fraction of unique protein IDs for each elution are shown in (**C**). Percentages of total and bait peptides as well as unique protein IDs obtained with the first (EL1) and second (EL2) V5 elutions for each viral bait (**D**). This figure is representative of n= 1 experiment.

4.5 Peptide solubilising agents

After elution, protein samples were precipitated with TCA to remove the PBS salt that derived from the elution buffer, as high monovalent salt concentrations may interfere with trypsin activity. Proteins were then re-suspended directly into the digestion buffer which contained 10% acetonitrile for the purposes of solubilising dried-down protein.

The last optimisation step aimed to determine whether replacing 10 % acetonitrile by 10 % guanidine hydrochloride (GuHCl) would increase protein solubility, which can be measured by an increase in the number of quantified. Using two sets of V5 immunoprecipitation lysates from both UL54 and UL123-V5 expressing cell lines, samples were processed as described in **2.9.1**, but for protein digest samples were resuspended either in digestion buffer containing 10 % acetonitrile or 10 % GuHCl. As depicted in Figure 4.5, acetonitrile-containing buffer led to the greatest number of bait and overall peptides, and was therefore used.

To reduce variability in sample preparation for the interactome samples, a batch of digestion buffer containing acetonitrile was prepared and aliquoted, so that all samples could be processed as similarly as possible. After digestion peptide samples were loaded onto a StageTip. This is a pipette tip containing a fixed C18 silica-based medium, used in proteomics for single-step desalting, enrichment and purification of protein/peptide samples. It can be seen as a buffer-exchange step that placed samples in injection buffer for liquid chromatography tandem mass spectrometry.

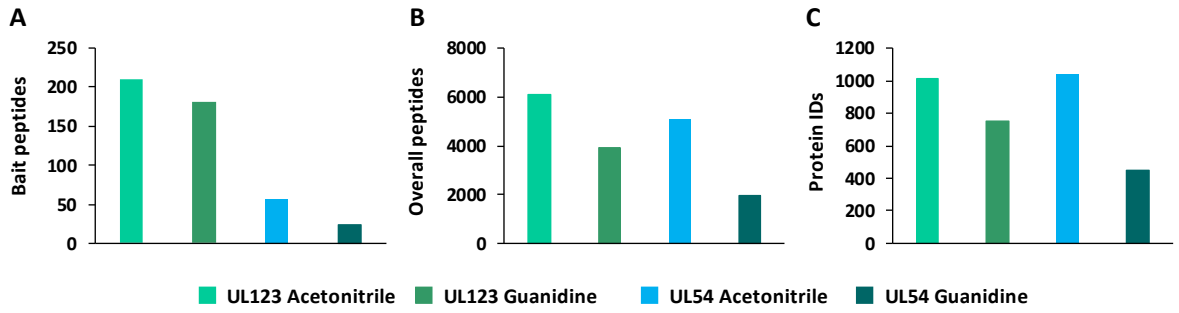


Figure 4.5 | Efficacy of solubilising agents in trypsin digestion buffer
Numbers of bait peptides (**A**), overall quantified peptides (**B**) and protein IDs (**C**) obtained by solubilising proteins for trypsin digestion using guanidine hydrochloride or acetonitrile. This figure is representative of n= 1 experiment.

4.6 Discussion

Optimisation of the immunoprecipitation protocol was necessary to yield a robust quantification of interacting proteins while guaranteeing an efficient use of resources. The number of bait peptides and quantified positive control interactors were the most specific measurements used to assess the optimisation procedure, as the number of total peptides and protein IDs include common AP-MS contaminants, and non-specific interacting proteins.

Identification of UL27-interacting proteins was inconsistent among experiments, with only one of the 28 known interactors, PSME3, being quantified in all datasets. Almost

a third of UL27-interacting proteins were not quantified in any of the optimisation experiments. This could be due to differences in experimental conditions between different studies. Potential reasons could include differences in most parts of the immunoprecipitation protocol. For example, different tags were used for affinity purification, presence of virus and time-point of harvest post-infection, different lysis buffer composition including distinct detergents, as well as cell type were not the same as described in Reitsma *et al* (2011) [226]. Furthermore, a subset of the interactors seemed to vary with the amount of input material, alluding to their relative abundance. In fact, despite its relative low molecular weight (29.5 kDa), PSME3 was quantified by the greatest number of peptides in all experiments, which may indicate that it is particularly likely to be a specific UL27 interactor. Other possible reasons for this observation may include relative protein abundance or UL27 binding affinity.

Both the organic solvent acetonitrile and the chaotropic agent GuHCl solubilise proteins through denaturation [227]. In a direct comparison with either 10 % acetonitrile or 10 % GuHCl in the digestion buffer, the acetonitrile-containing buffer yielded a greater number of quantified bait and total peptides. GuHCl has been shown to increase identification of hydrophobic peptides [228] whereas hydrophobic proteins have generally low solubility in acetonitrile/aqueous solutions [229]. The observed effect may however not solely reflect an effect on solubility, with another possible explanation of this data including improved protein digestion in the presence of acetonitrile.

One step that was not optimised was a comparison between elution of interacting proteins using excess V5 peptide followed by in-solution digest and elution by boiling the resin with SDS sample buffer, running the sample on a gel followed by in-gel band digest. However, the latter approach has a few caveats: (a) elution by boiling includes nonspecific interactions of proteins that bind to the resin but not the bait; (b)

hydrophobic peptides may prove difficult to isolate from gel bands; (c) each gel lane is typically separated into multiple slices that require individual MS analysis.

In order to increase detection of weaker interactions, two other common strategies that could have been trialled as modifications to the starting immunoprecipitation protocol from Huttlin et al, are the use of cross-linking agents and cryogenic cell lysis. Cross-linking agents stabilise transient and weak interactions, while cryogenic lysis helps preserving protein complexes by immediately freezing and then mechanically grinding the sample. However, both of these may also increase detection of non-specific interactions [230].

Mass spectrometry is a powerful technique that has increasingly become the preferred method for analysis of complex protein samples. Viral interactome studies commonly employ yeast two-hybrid (Y2H) screens or AP-MS. Y2H allows identification of direct interactions between a pair of proteins, but has a relatively high false-positive rate resulting from non-physiological expression of baits in a different organism. Y2H may also give false negative results, missing relevant interactions since pathogen proteins are expressed outside the context of infection, and may not fold properly in yeast [230]. AP-MS can identify members of stable protein complexes in the context of infection, however does not provide information on whether all the isolated proteins directly interact with each other.

Sample labelling with tandem-mass tags (TMT) or SILAC has been used in the generation of small sample-size viral interactomes, as for example HSV-1 UL37 and the NS1 and NS2 proteins of Human respiratory syncytial virus [177, 181]. However, label-free quantitation (LFQ) remains the most common method for relative quantitative analysis. SILAC quantitation has a maximum sample number of three (in tandem), with an additional requirement that cells be cultured in labelling medium for two to three

weeks prior to analysis [231]. A recent study comparing TMT and LFQ quantitative coverage, has shown that the former is more sensitive due to fewer missing values [232]. However, for the interactome dataset, there were two caveats of using TMT labelling: (a) the current number of samples that can be analysed in tandem is eleven; (b) given that substantially different levels of bait expression were observed, each IP would yield disparate amounts of protein resulting in very unequal protein content between channels. If mixing resultant samples 1:1:1:1:1:1:1:1:1:1:1, quantitation of proteins in channels with low-abundant baits would suffer, as the majority of ions would derive from other channels. The benefits of LFQ include unlimited sample number, no additional labelling steps, and most importantly, slight differences in bait abundance between samples are tolerated [231].

5 | HCMV-host protein interactions

CompPass filtering was performed in collaboration with Dr. Edward Huttlin (Department of Cell Biology, Harvard Medical School). DAVID functional enrichment analysis was performed by Dr. Michael Weekes (Department of Medicine, University of Cambridge). PFAM domain association analysis was performed by Dr. Edward Huttlin and Dr. Michael Weekes. Validation of the interactions between RL1-CUL4A, UL71-TRIM22 and Ul25-NCK1 was performed by Dr. Katie Nightingale (Department of Medicine, University of Cambridge). Parts of the work presented in this chapter have been published in a similar form in eLife (Nobre et al., 2019).

To generate the HCMV interactome, stable cell lines were made from HFFF-TERT cells, each expressing a C-terminally V5-tagged HCMV ORF to enable affinity-purification. The immunoprecipitation baits included 170 canonical ORFs and two non-canonical ORFs, ORFL147C and ORFS343C. These were included as they ranked close to both ends of HCMV protein relative abundance detected in previous proteomics studies (see Figure 1.4) [162, 194]. Expression of all tagged 'baits' but UL136 was validated by IB, MS or RT-qPCR (see Figure 5.0.1) before AP-MS.

To enable detection of virus-virus interactions, the affinity-purification was performed on lysates of the stable cell lines, which were harvested after 60 h of infection with HCMV Merlin at an MOI of 2. As detailed in **1.2.6**, this strain was cloned into a BAC to minimise genetic instability and prevent loss of gene function. It contains a full-length genome which expresses all HCMV genes apart from UL128 and RL13. Lysates were collected at 60 h post-infection as all detectable proteins have been shown to be

expressed at this point with this strain [162] (see Figure 3.4.1). HCMV ORFs were divided into two sets of baits according to their solubility: proteins with zero or one TM region (n=153), were solubilized using an NP40-based lysis buffer; proteins with 2-8 TM regions, were solubilized using a digitonin-based buffer. Digitonin was chosen to solubilize multi-pass membrane baits as mild detergent extraction has been shown to improve identifications of interacting proteins ('prey') with hydrophobic membrane proteins [233, 234]. A schematic of the AP-MS strategy is shown in Figure 5.0.2.

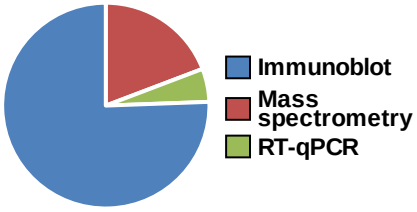


Figure 5.0.1 | Percentage of HCMV interactome baits validate by IB, MS or RT-qPCR
Prior to AP-MS in the context of infection, uninfected lysates were used to validate bait expression by IB, MS or RT-qPCR (as detailed in **3.2**). Expression of 78 % of baits was further validated in the interactome itself, including 9/18 baits which had only been detected by RT-qPCR. In this figure, these nine baits were included in the 'validated by Mass spectrometry' class. From the remaining baits which were only detected by RT-qPCR, four were small proteins of 47-111 aa, with 1-4 theoretically observable peptides, and none were detected in two previous proteomic analyses of HCMV infection [162, 163]. Another 4/9 viral proteins in the 'validated by RT-qPCR' class were only detected by a median of 0-2 peptides in these two previous proteomic analysis, with two containing multiple transmembrane domains. This suggested that detection may have been limited by protein abundance or hydrophobicity. Known interactions for these nine viral baits were detected in the AP-MS data (for example the interaction between UL48A and UL86 [235], and US18 and the natural killer cell cytotoxicity receptor 3 ligand 1 (NCR3LG1) [194]) and thus data for baits in the 'validated by RT-qPCR' class was included in this interactome. UL136 was excluded from further analysis as its expression could not be validated by any method.

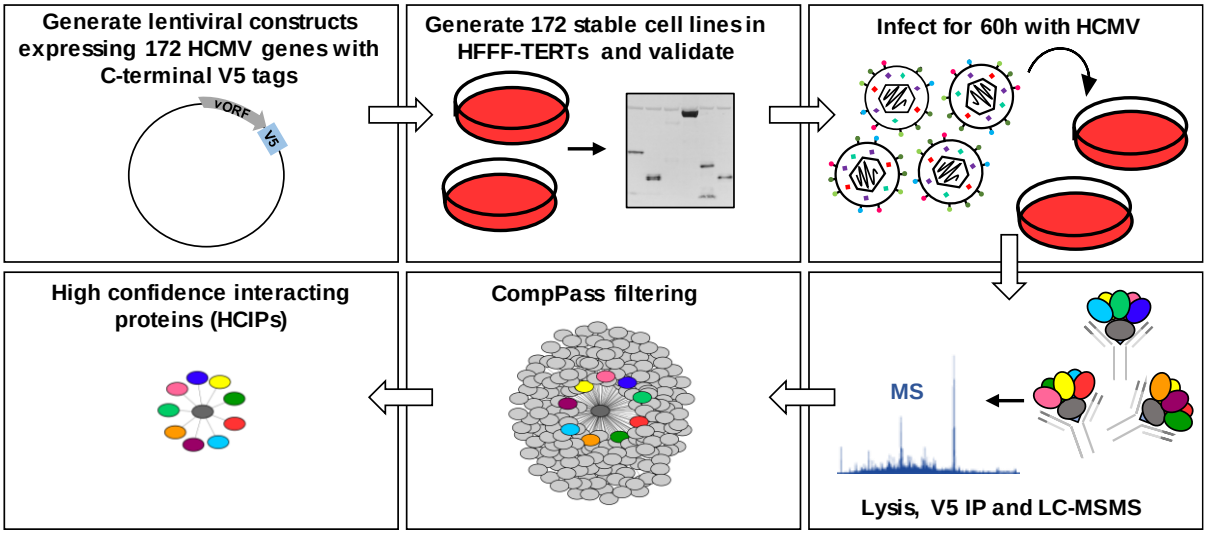


Figure 5.0.2 | Schematic of the AP-MS strategy.
Lentiviral constructs were generated, each containing the coding sequence for one of 172 HCMV ORFs, followed by a C-terminal V5 tag to facilitate immunoprecipitation (IP). Stable cell lines were then generated from HFFF-TERT cells, and the expression of each construct was validated by either immunoblot, mass spectrometry or RT-qPCR. Cells were infected with Merlin strain HCMV at an MOI of 2 for 60 h. IP samples were generated and analysed in technical duplicate, using the method originally described in Huttlin et al (2015) and detailed in **2.9.1** [189, 236]. The CompPASS algorithm was used to assign scores for all quantified interactors for each bait. Then, stringent filters were applied to remove inconsistent and low-confidence protein identifications across all IPs [189, 195]. Interactions passing these criteria were designated 'high confidence interacting proteins' (HCIPs), and were used in the analyses shown in **Chapters 5** and **6**.

AP-MS experiments can yield lists of hundreds to thousands of interactors. Data filtering thus becomes a crucial tool for the distinction between true interactors and common background between experiments. This chapter will focus on the strategies (controls, combined database for protein identification, filtering criteria) used for the identification of 'high-confidence interacting proteins' (HCIPs) and the subsequent bioinformatic analysis of the filtered dataset, while showcasing experimental validation for a subset of interactions.

5.1 Controls and correlation between replicate samples

In this interactome, the biological replicates were pooled and samples were analysed in technical duplicate, as described in Huttlin et al., 2015. This approach was chosen in order to address potential carry-over of peptides between consecutive samples with different baits. Furthermore, two 'wash' injections were queued in between AP-MS samples, with the overall run order of replicate batches being reversed. This guaranteed that carry-over for each sample in either batch was different (i.e. Batch A1: Sample 1, wash x2, Sample 2, wash x2, Sample 3; Batch A2 (replicate of A1): Sample 3, wash x2, Sample 2, wash x2, Sample 3). Then, a score incorporated within CompPass was used to filter out carry-over contaminants. This score, designated 'entropy', compared the number of peptide-spectrum matches (PSM) between technical replicate injections and eliminated prey that were not detected consistently (detailed further in **2.11.2,** Figure 2.11.2.1 and **5.3**).

It was therefore important that replicate injection material was as similar as possible to ensure this filter was effective.

Using technical replicates also aimed to address sufficiency of the amount of injected material, ensuring that enough material was present for MS analysis after all sample preparation steps had been completed. Using biological replicates would have required the use of additional technical replicates for each for compatibility with the entropy score. This would have doubled the required MS instrument time and reagents such as the amount of virus to infect interactome samples.

In order to assess the impact of biological variability in HCIP identification, six AP-MS experiments were re-run with biological replicates instead of technical replicates. In this independent analysis, depicted in Figure 5.1.1, a very good correlation between the

numbers of PSM from each identified HCIP between biological replicates was observed, adding confidence that HCIP identified by CompPass are likely to be genuine interactors.

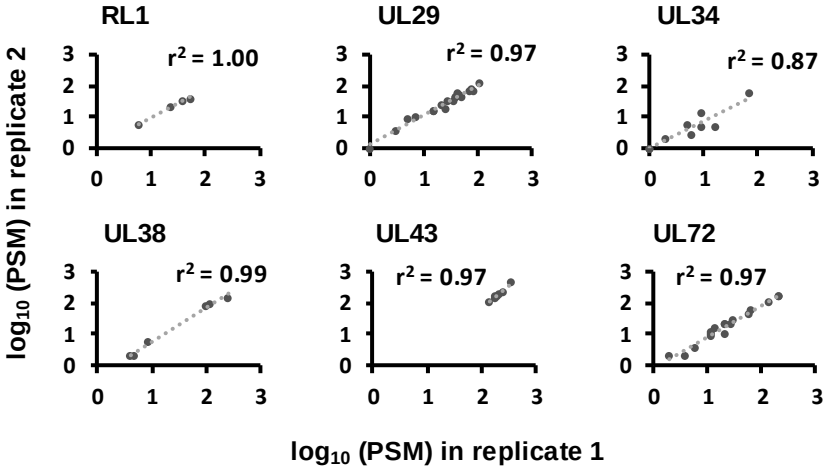


Figure 5.1.1 | Reproducibility of biological replicate AP-MS samples
For the HCMV interactome samples were analysed as technical replicates. To assess the reproducibility between biological replicate samples, six AP-MS experiments were repeated with independent analysis of each biological replicate. PSM are shown for HCIPs predicted by the interactome for each of the baits.

A control to assess instrument performance was included at the start of the queue of samples to run on the mass spectrometer, in order to prevent samples being analysed in suboptimal conditions. The control sample was generated by harvesting 30 confluent 15 cm^2 dishes of uninfected UL123-V5 expressing HFFF-TERT cells, and processing those lysates in the same way as the samples that constituted the HCMV interactome. Peptides from these lysates were pooled and aliquoted to generate 30 identical control samples (technical replicates). Three uninfected UL123 controls were run using the same settings as the interactome samples, and averages of summary search statistics (generated by the software pipeline for quantitative proteomics) were calculated to produce a standard for comparison (see 'Benchmark', Table 5.1**)**. Summary statistics for all subsequent uninfected UL123 control samples were compared against the standard to assess instrument performance (see Table 5.1). Data from these controls was not included in the

final data analysis, to avoid modification of NWD and z-scores for the infected UL123-expressing sample.

Table 5.1 | Summary statistics of UL123 control samples

Control	Sensitivity	Success rate	xCorr	Total peptides	Unique peptides	Total peptides ratio to benchmark	Unique peptides ratio to benchmark
Benchmark	94.5	30.7	2.8	11228	7264	1	1
UL123 A1	91.7	38.89	2.80	18662	10918	1.7	1.5
UL123 A2	95.3	29.67	2.77	13575	7932	1.2	1.1
UL123 A3	93.4	35.16	2.86	12435	7222	1.1	1.0
UL123 A4	96.8	31.63	2.88	10990	7036	1.0	1.0
UL123 A5	94.4	31.28	2.85	10083	6446	0.9	0.9
UL123 A6	93.2	28.40	2.76	10674	5940	1.0	0.8
UL123 A7	93.8	26.07	2.77	8338	5465	0.7	0.8
UL123 A8	97.7	27.64	2.66	6570	4651	0.6	0.6
UL123 A9	94.1	28.85	2.70	11892	7962	1.1	1.1
UL123 A10	94.9	30.64	2.62	11064	7572	1.0	1.0
UL123 A11	96.6	29.57	2.62	10751	7255	1.0	1.0
UL123 A12	95.8	30.75	2.63	10090	6672	0.9	0.9
UL123 A13	94.7	30.06	2.58	10417	7185	0.9	1.0
UL123 A14	93.0	33.72	2.65	11849	8660	1.1	1.2
UL123 A15	92.9	33.87	2.48	10133	7499	0.9	1.0
UL123 A16	93.1	28.53	2.63	9418	6447	0.8	0.9
UL123 A17	95.2	32.62	2.63	9902	7017	0.9	1.0
UL123 A18	91.5	28.15	2.65	10739	6656	1.0	0.9
UL123 A19	92.8	30.76	2.59	11511	7902	1.0	1.1
UL123 A20	93.9	33.81	2.87	11434	7307	1.0	1.0
UL123 A21	93.2	32.17	2.82	11552	8283	1.0	1.1
UL123 A22	92.9	35.05	2.82	11373	7428	1.0	1.0
UL123 A23	92.9	32.05	2.84	11390	7387	1.0	1.0

Each batch of samples was run in a specific order to prevent baits with similar function or virion location from running consecutively. As an additional quality control, replicate samples were compared in terms of summary search statistics and LC chromatogram, to ensure instrument performance had not decreased throughout the batch. A high level of correlation in peptide quantification between technical replicates was observed (see Figure 5.1.2).

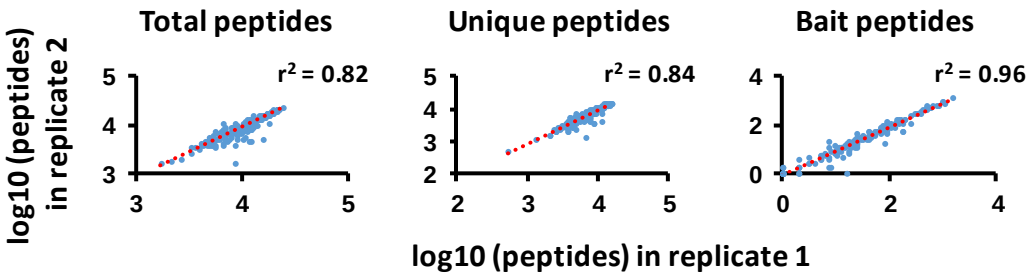


Figure 5.1.2 | Correlation of the number of total, unique and bait peptides for proteins identified in technical replicates 1 and 2 of the HCMV interactome. 'Total peptides' refers to the sum of all peptides identified in each sample. 'Unique peptides' refers to peptides that can only be matched to the tryptic profile of the sequence of one protein from proteomes in the combined database (in this case, human and HCMV Merlin). 'Bait peptides' refers to peptides derived from the bait protein for each IP (i.e. from US1 in the US1 IP). This metric is independent of the V5 epitope sequence, as the sequences of the baits featured in the database do not feature the linker region nor the tag. Additionally, contrary to the 'total peptides' and 'unique peptides', 'bait peptides' is not provided by the summary search statistics function within 'MassPike'. All data for this figure are also shown in **(**Appendix C, C1**)**.

5.2 Data filtering with CompPass

The raw data was searched against a combined database containing human and HCMV strain Merlin Uniprot entries (detailed in section **2.11.1**), to assign mass spectra to peptide sequences which were then assembled into proteins. The full list of interacting proteins quantified in all AP-MS experiments was then processed using CompPass, which calculated scoring metrics for each interactor (as detailed in **2.11.2**). Each of the two buffers used for lysis and IP yielded a distinct background of non-specific interactors. Thus for CompPass filtering, samples were segregated into two datasets according to lysis buffer, and the datasets were scored independently to better account for detergent-specific variation in the AP-MS background [189, 195].

For each prey protein in every IP, CompPass calculated: (a) an average of the number of peptide spectral matches (PSMs) between the two replicates; (b) an entropy score, aiming to eliminate proteins inconsistently detected proteins by comparing the number of PSMs between replicates of the same IP; (c) a z-score, calculated by comparing

the average and standard deviation of PSMs observed across all IPs; and (d) a normalized WD (NWD) score, which incorporated each protein's frequency of detection across all IPs, as well as their reproducibility among replicates. The NWD score was calculated as described in **2.11.2** (see Figure 2.11.2.2) using an equation that incorporated the fraction of runs in which a protein was detected, the number of PSMs as well as their average and standard deviation from the mean of PSMs observed for that protein across all samples, and the number of replicates (1 or 2) where it was identified [208]. NWD scores were normalized so that the top 2 % earned scores of ≥1.0.

HCIPs were identified using the following criteria: (a) a PSM score ≥1.5 (i.e. a minimum of 3 peptides per protein across both replicates); (b) an entropy score ≥0.75; (c) a top 2 % NWD or Z-score. A similar filtering strategy has been applied by a previous study, estimating a 5 % false discovery rate within the top 2 % NWD score [195].

Setting the thresholds for HCIP identification affects the interactor lists for all baits simultaneously. Applying the same cut-offs across the whole dataset disregards factors such as the total number of interactors quantified in the AP-MS data for each bait. This may result in the identification of a high number of HCIPs for a subset of baits (for example through binding to protein complexes with several subunits or co-immunoprecipitation of secondary interactors to directly-interacting proteins) while no HCIPs are identified for another subset. Thus, threshold setting is a critical and non-trivial task that aims to balance identification of known interaction partners while avoiding loss of stringency and consequently, a compromise had to be made.

As reported in prior interactome studies, certain known interactions scored below these stringent NWD or Z-score criteria. For example, the interaction between the DNA polymerase subunits UL54 and UL44 had an NWD score below 1 with either protein as a bait, but met all other filtering criteria. Thus, proteins scoring within the top 5 % NWD or

Z-scores (>0.434 and >3.688 respectively) were also included, if they were reported to interact with the bait in a prior study [209]. In order to use an unbiased collection of reported interactions, HCMV protein interaction data from a combination of online databases was used. This was comprised by entries from BioGRID, IntAct, Uniprot, MINT and Virus Mentha (search terms provided in section **2.11.3**).

After CompPass filtering, 132 baits were themselves identified as self-interacting proteins. There are examples of known oligomeric or multimeric viral proteins (e.g. UL104 assembles as a dodecamer, UL86 forms homomultimeric pentons and hexons, among others), however from this interactome, it was not possible to make a distinction between peptides who derive from the V5-tagged transgene or the virally-expressed protein. Thus, these HCIPs resulting from bait-enrichment were removed from the filtered dataset.

5.3 High-confidence interacting proteins

An initial Digitonin-based AP-MS analysis failed to generate any interactors for protein UL133 (2 TM regions) after filtering. This IP was repeated using the NP40-based lysis buffer which identified 13 HCIPs.

No interacting proteins passed the stringent filters employed for UL120 and UL142. For seven further proteins (US2, US11, UL93, UL96, UL119, UL146 and RL13), only the bait itself passed filtering, leaving 162 viral baits with ≥1 HCIP (see Figure 5.3.1). A total of 3440 HCIPs (excluding self-HCIP identification) were quantified across all 162 baits (see Figure 5.3.2 – 5.3.19), with a range of 1-174 and median of 9 interactions per bait, similar to what had been previously observed in the Bioplex 2.0 human interactome [189].

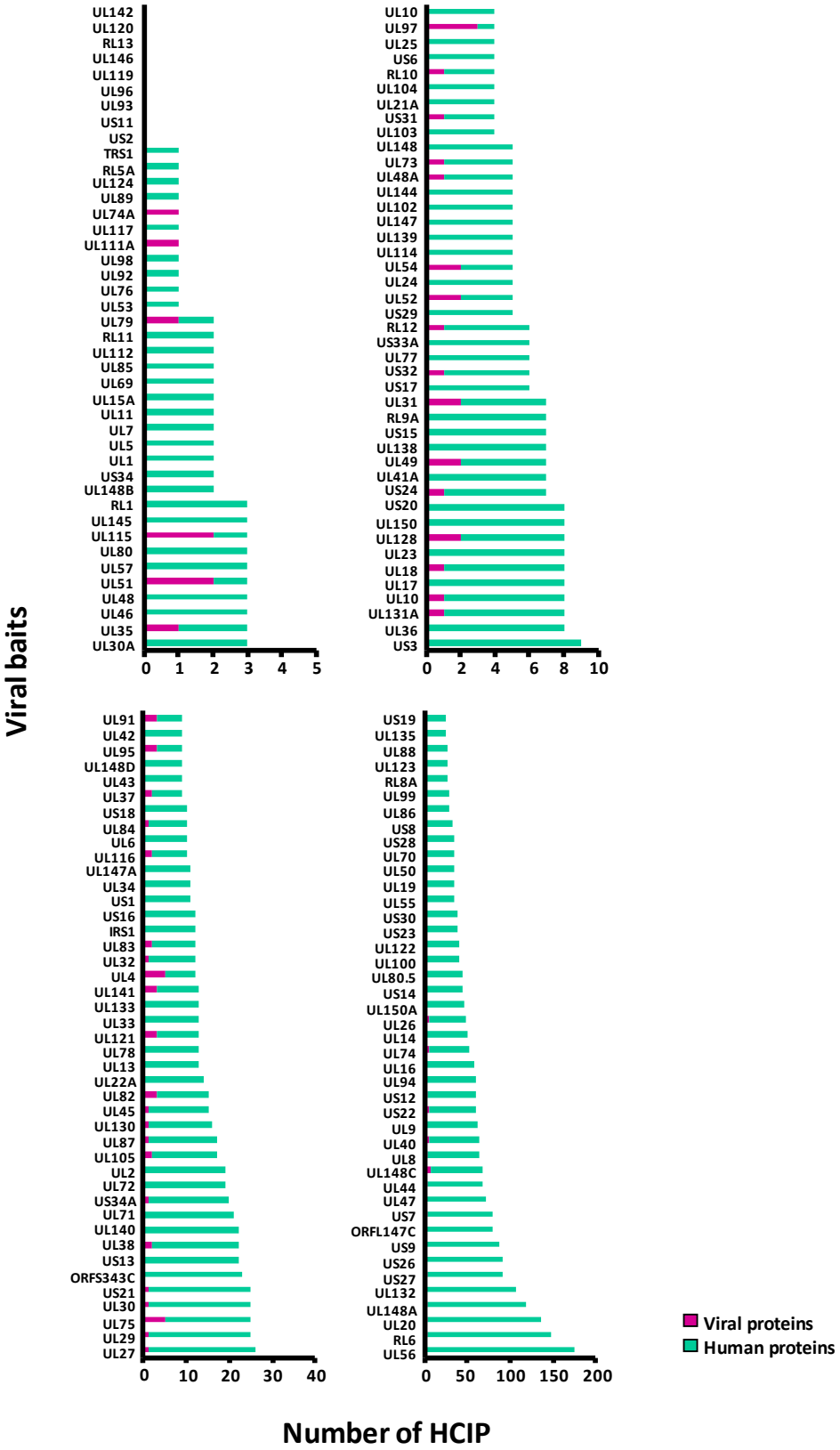


Figure 5.3.1 | High-confidence interacting proteins quantified in the HCMV interactome
Numbers of human (green) and viral (purple) HCIPs per bait, excluding bait-self interactions.

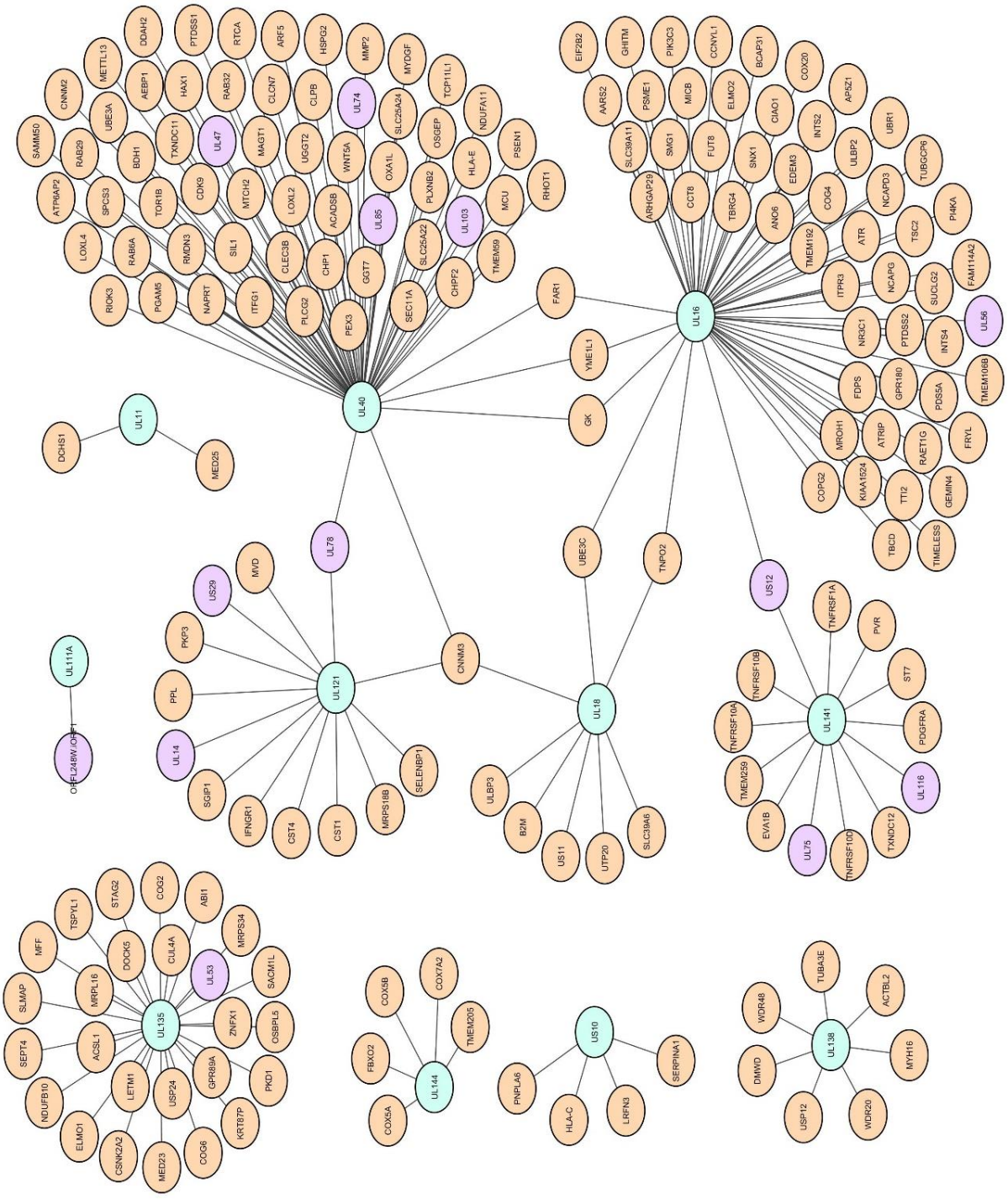


Figure 5.3.2 | Interaction diagrams for US10, UL11, UL16, UL18, UL40, UL111A, UL121, UL135, UL138, UL140 and UL144. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

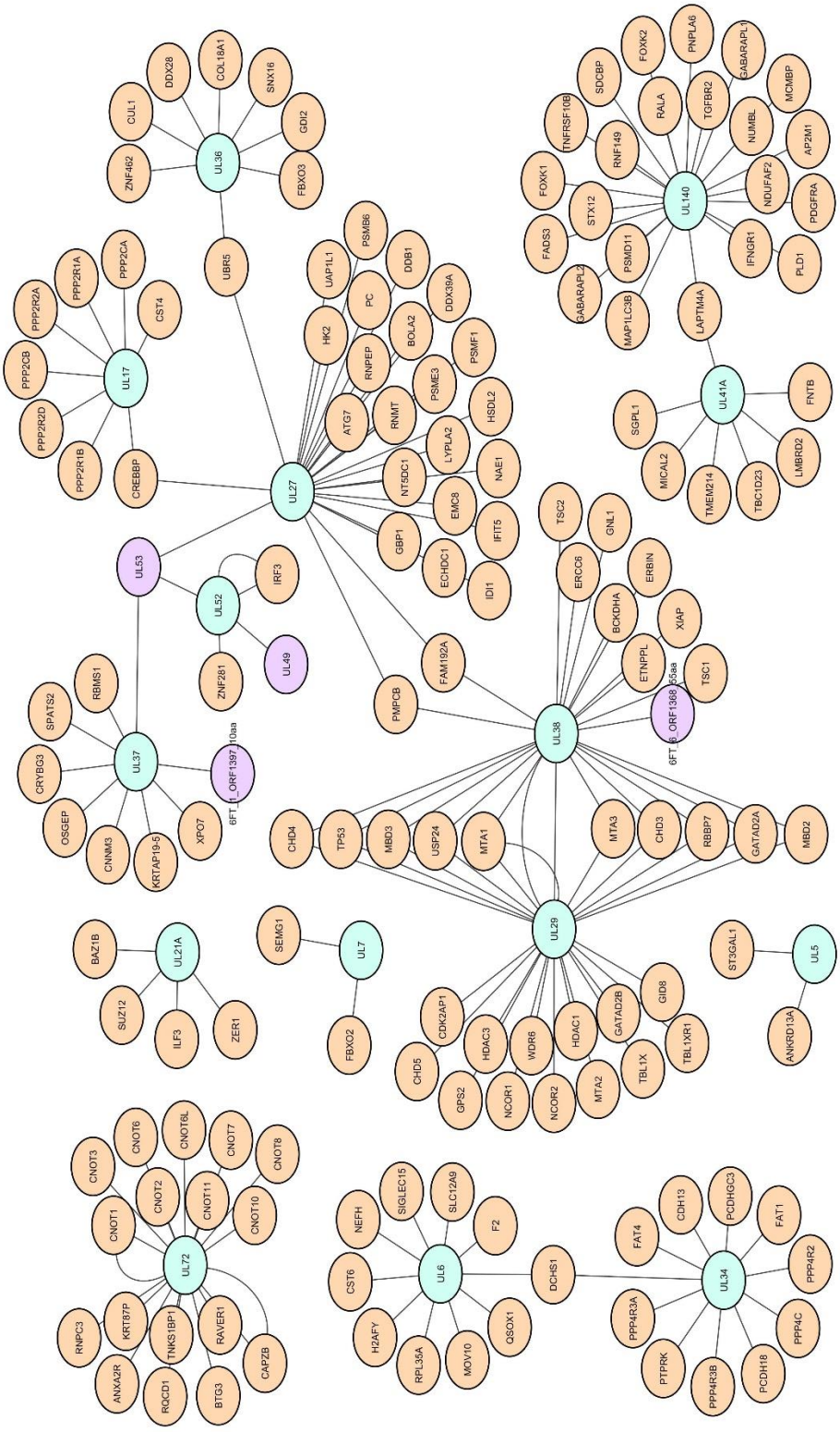


Figure 5.3.3 | Interaction diagrams for UL5, UL6, UL7, UL17, UL21A, UL27, UL29, UL34, UL36, UL37, UL38, UL41A, UL52, UL72 and UL140. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

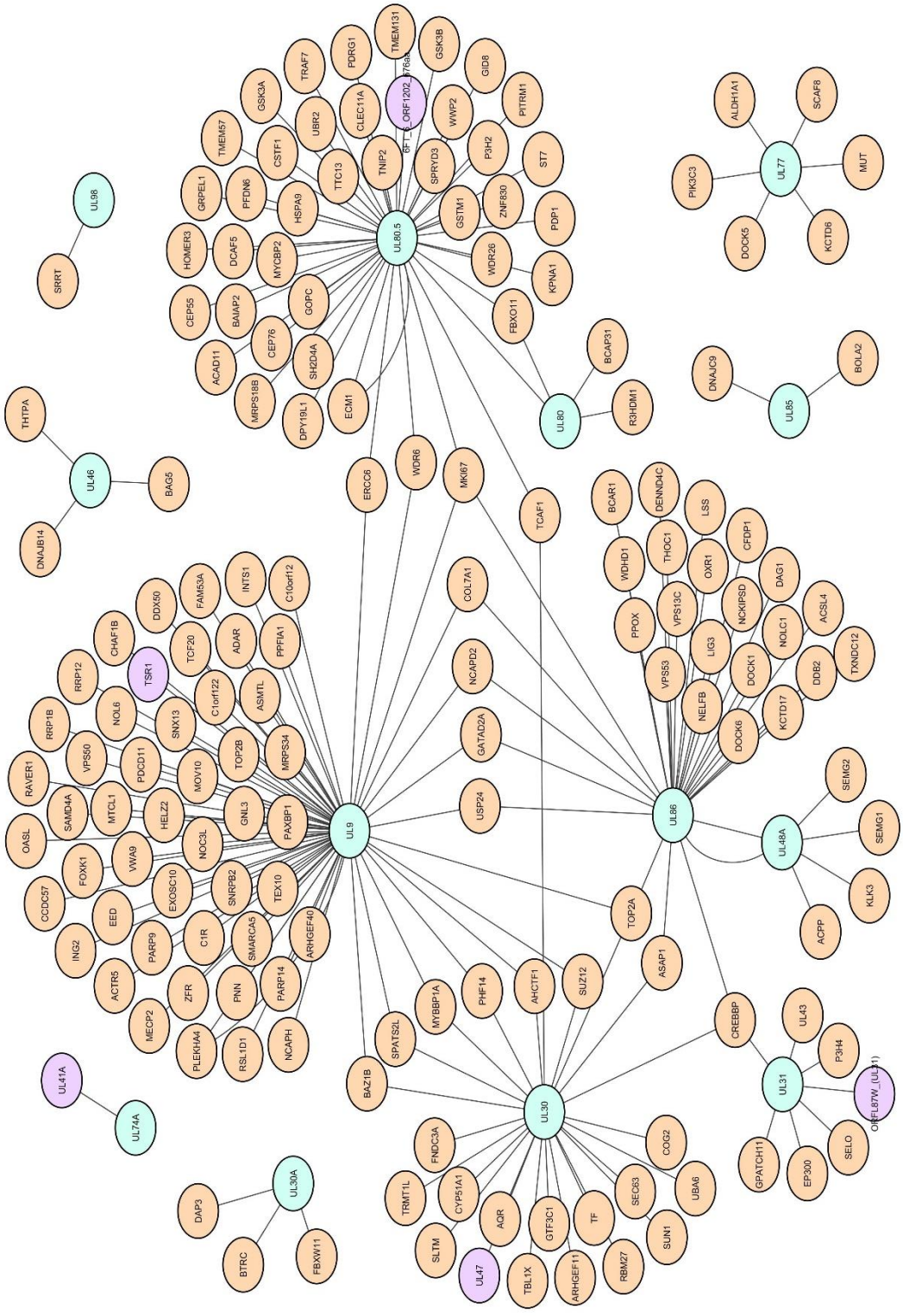


Figure 5.3.4 | Interaction diagrams for UL9, UL30, UL30A, UL31, UL46, UL48A, UL74A, UL77, UL80, UL80.5, UL85, UL86 and UL98. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

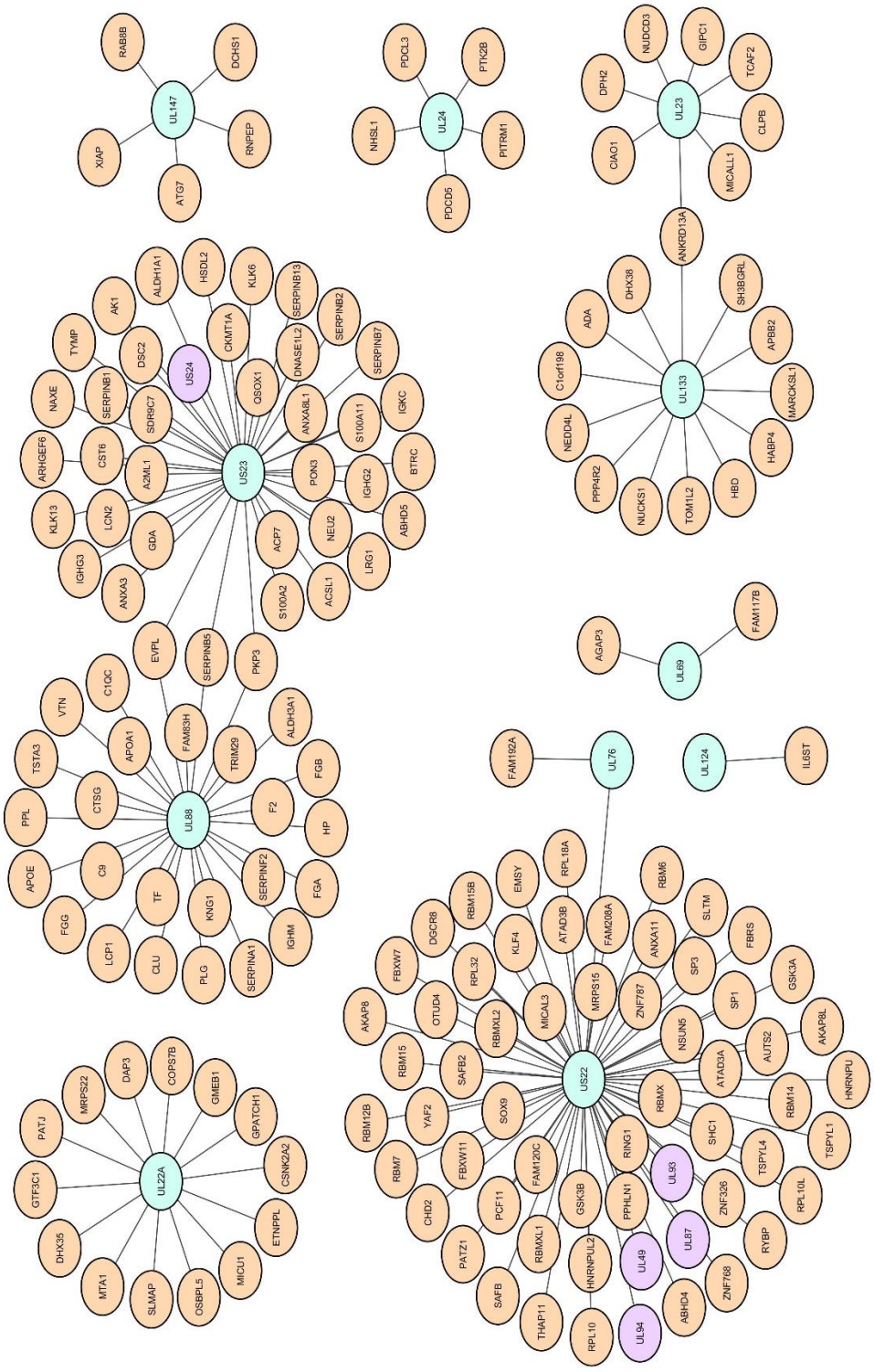


Figure 5.3.5 | Interaction diagrams for US22, US23, UL22A, UL23, UL24, UL69, UL76, UL88, UL124, UL133 and UL147. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

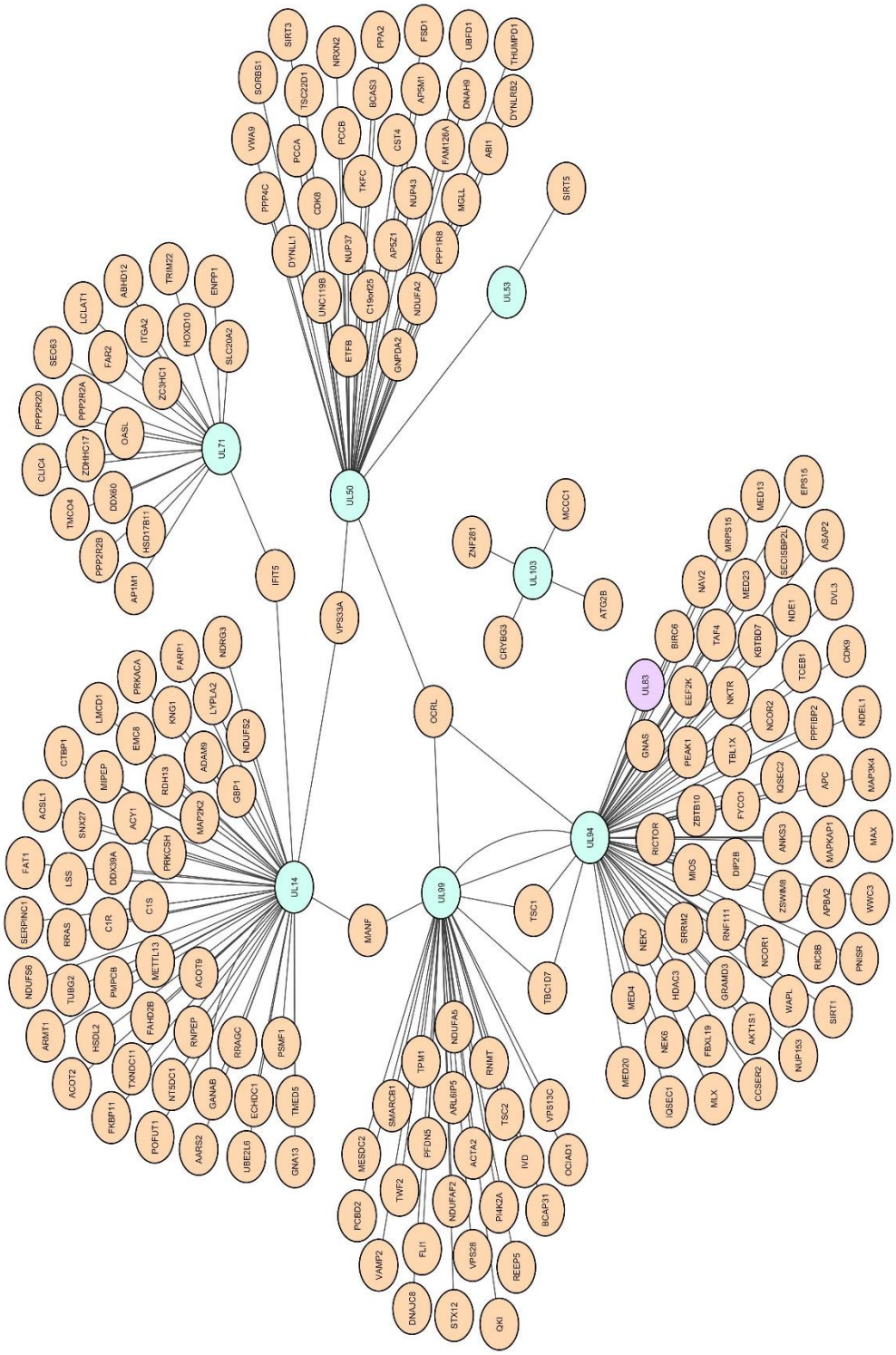


Figure 5.3.6 | Interaction diagrams for UL14, UL50, UL53, UL71, UL94, UL99, UL71, UL94, UL99 and UL103. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

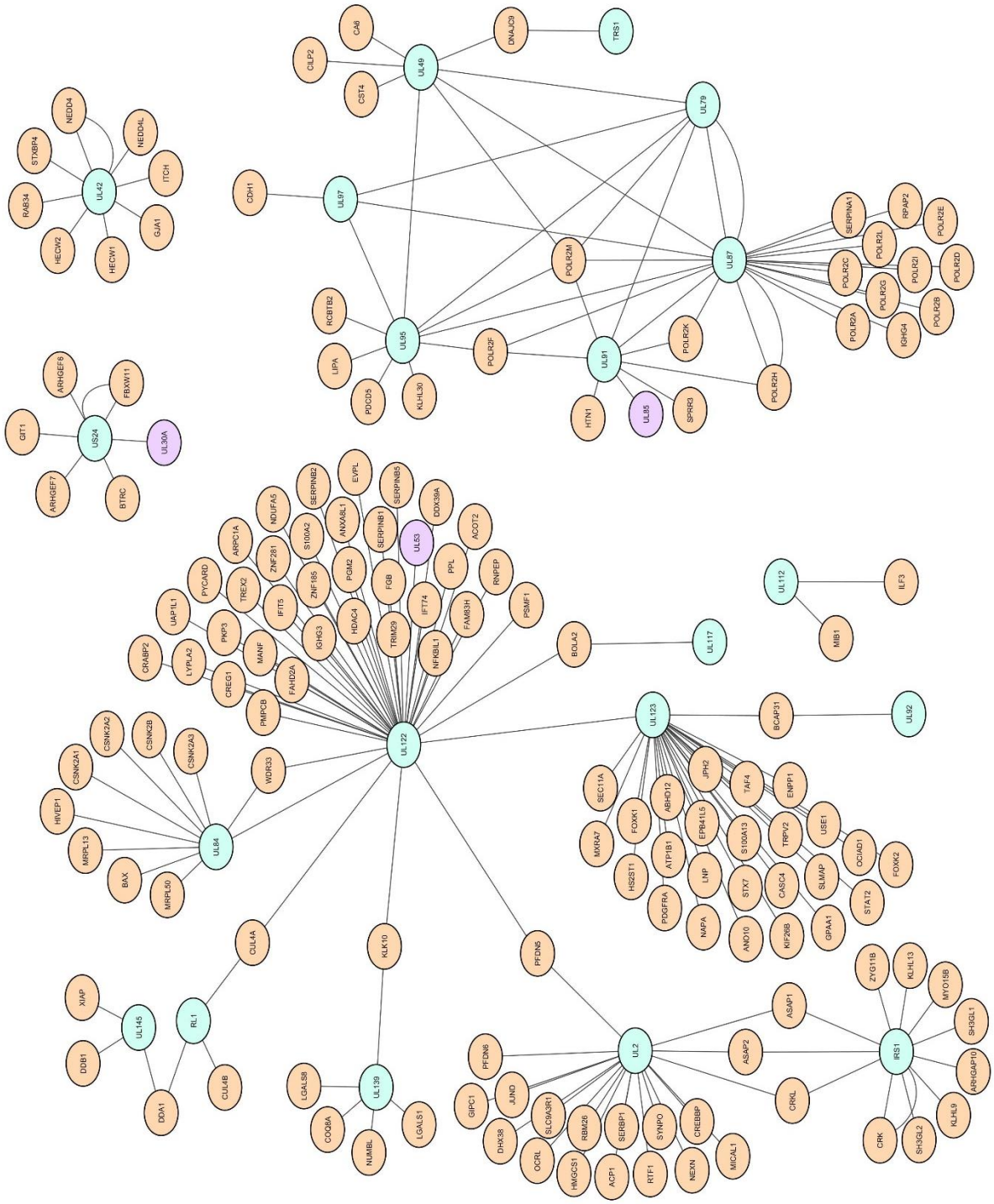


Figure 5.3.7 | Interaction diagrams for IRS1, TRS1, RL1, US24, UL2, UL42, UL49, UL79, UL84, UL87, UL91, UL92, UL95, UL97, UL112, UL117, UL122, UL123 and UL139. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

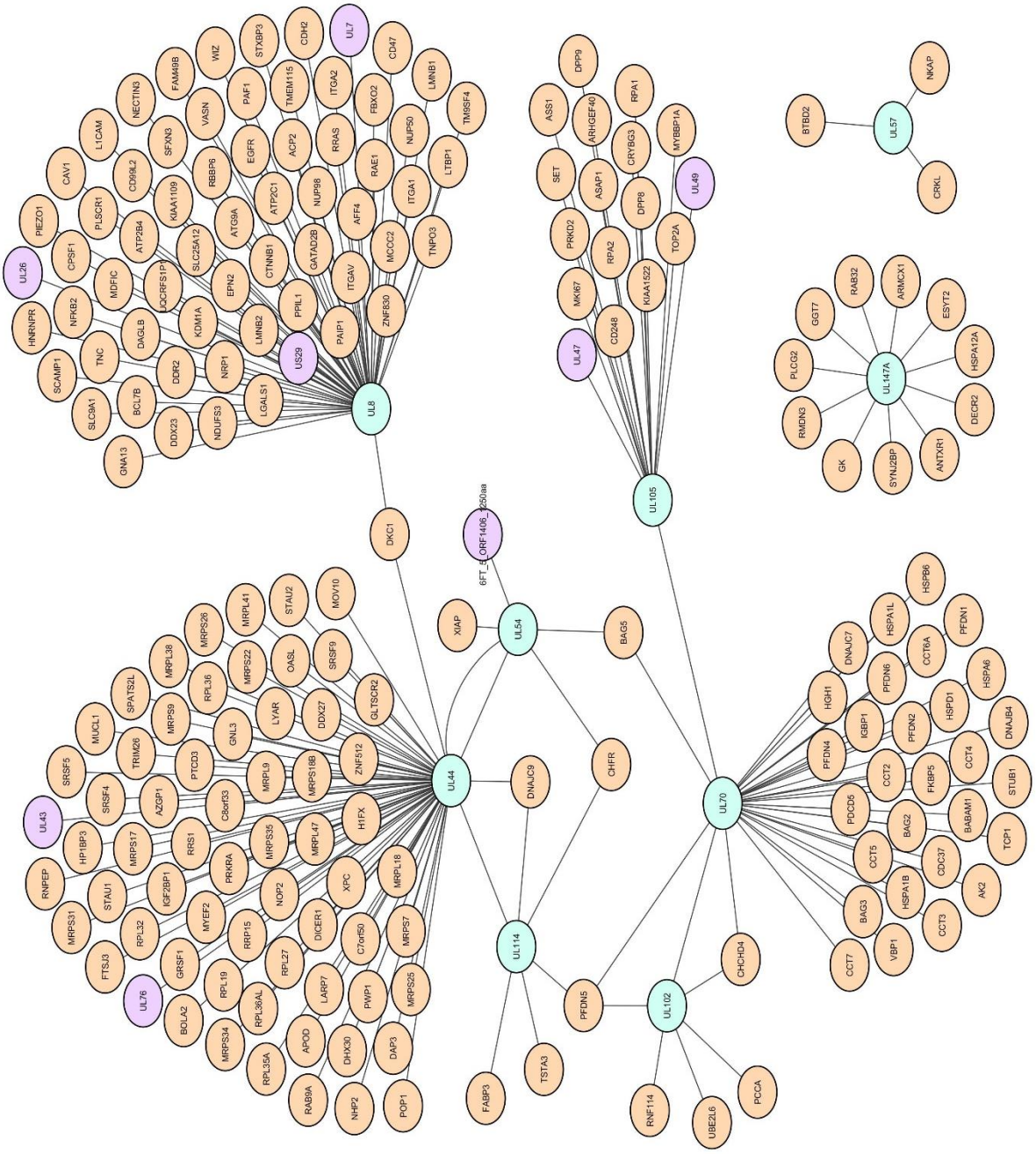


Figure 5.3.8 | Interaction diagrams for UL8, UL44, UL54, UL57, UL70, UL102, UL105, UL114 and UL147A. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

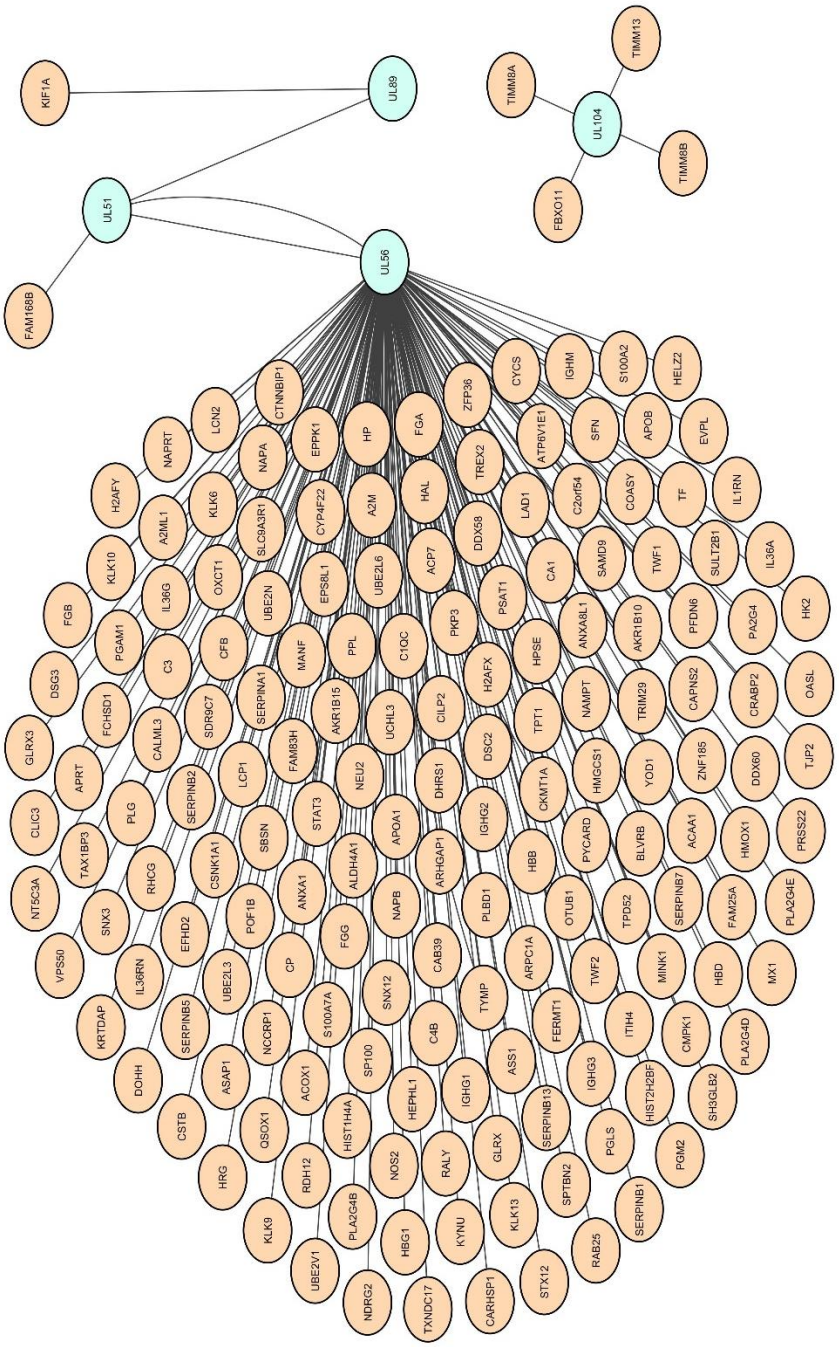


Figure 5.3.9 | Interaction diagrams for UL51, UL56, UL89 and UL104. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

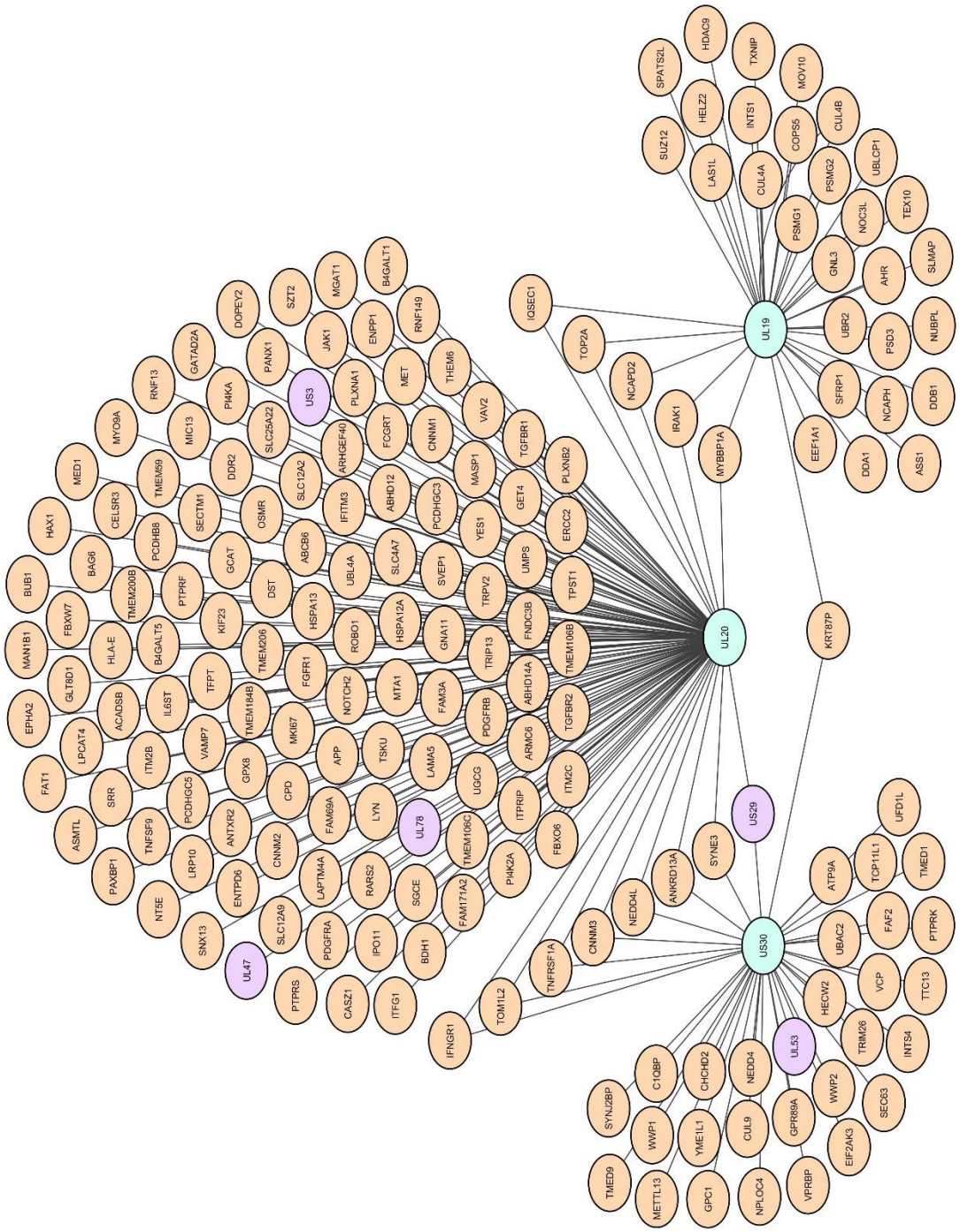


Figure 5.3.10 | Interaction diagrams for UL19, UL20 and US30. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

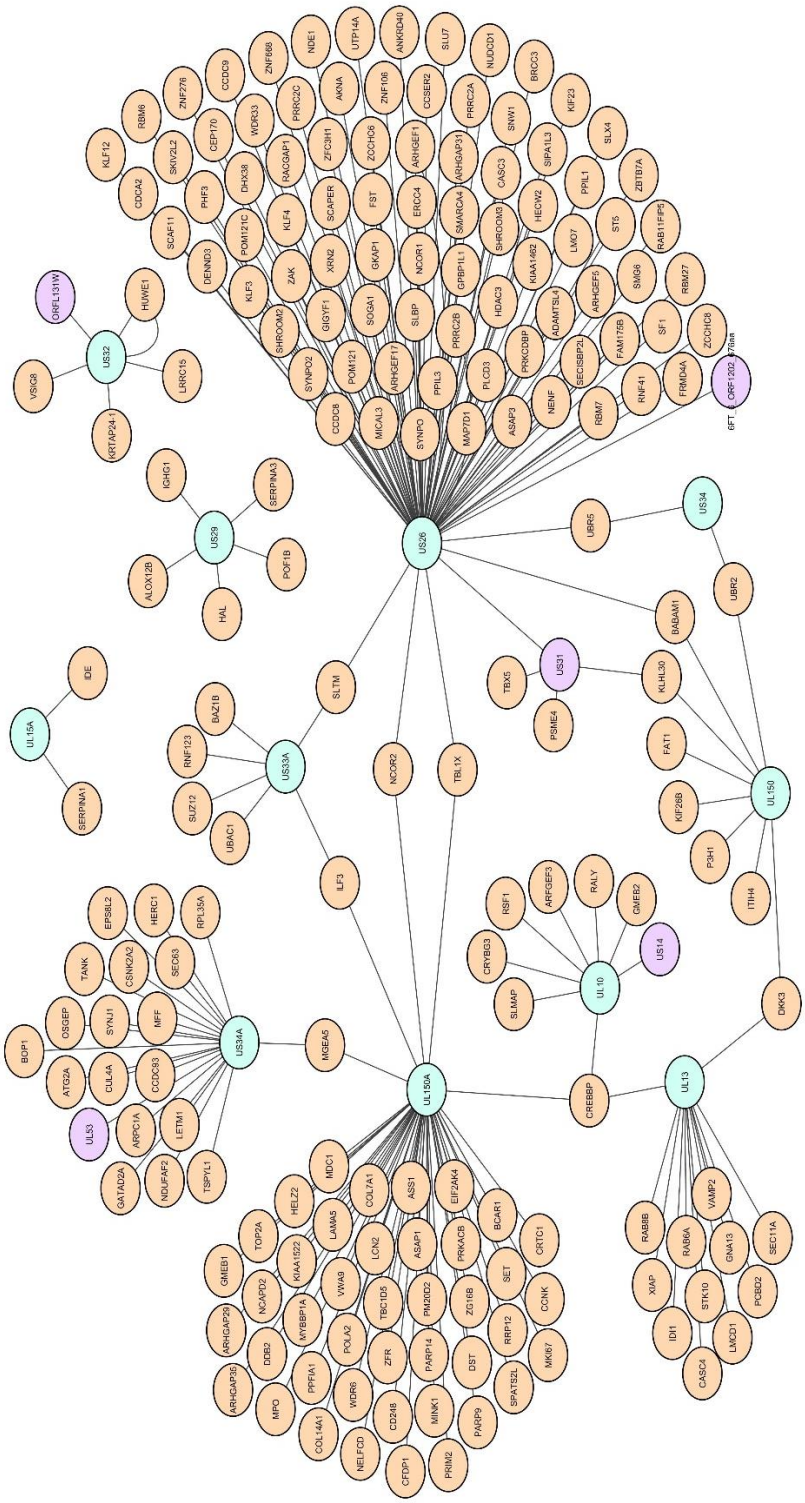


Figure 5.3.11 | Interaction diagrams for US26, US29, US32, US33A, US34, US34A, UL10, UL13, UL15A, UL150 and UL150A. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

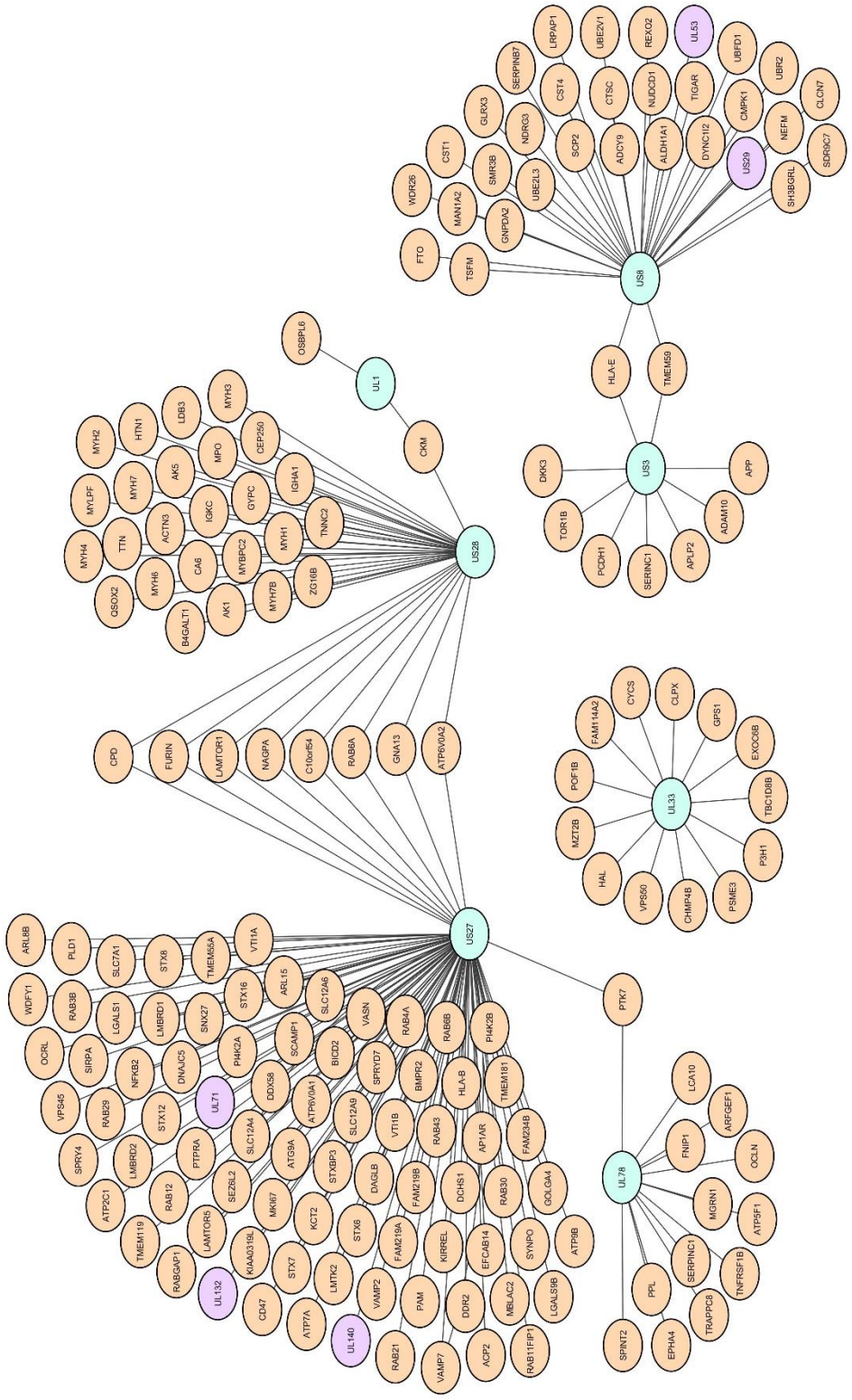


Figure 5.3.12 | Interaction diagrams for US3, US8, US27, US28, UL1, UL33 and UL78. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

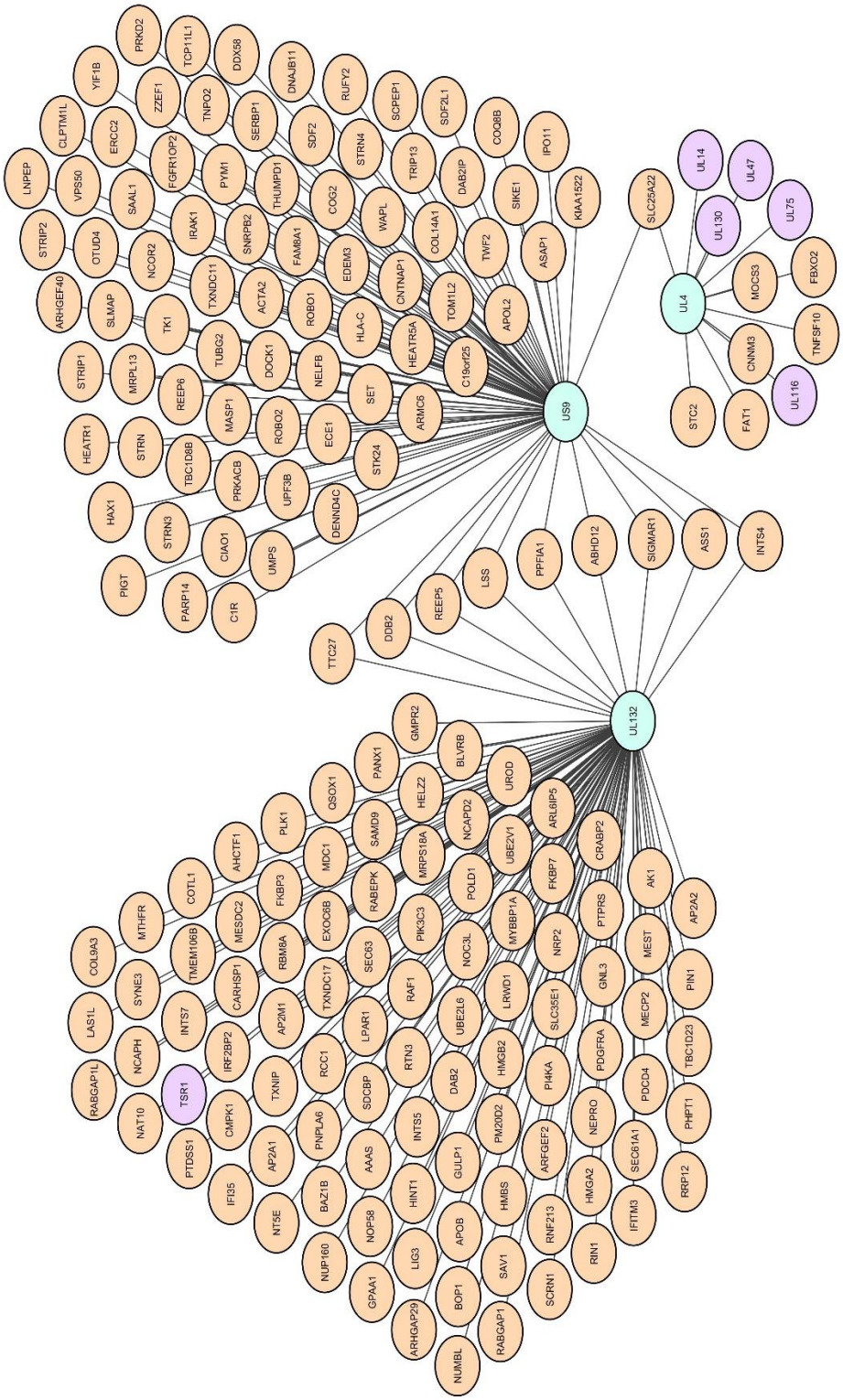


Figure 5.3.13 | Interaction diagrams for US9, UL4 and UL132. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

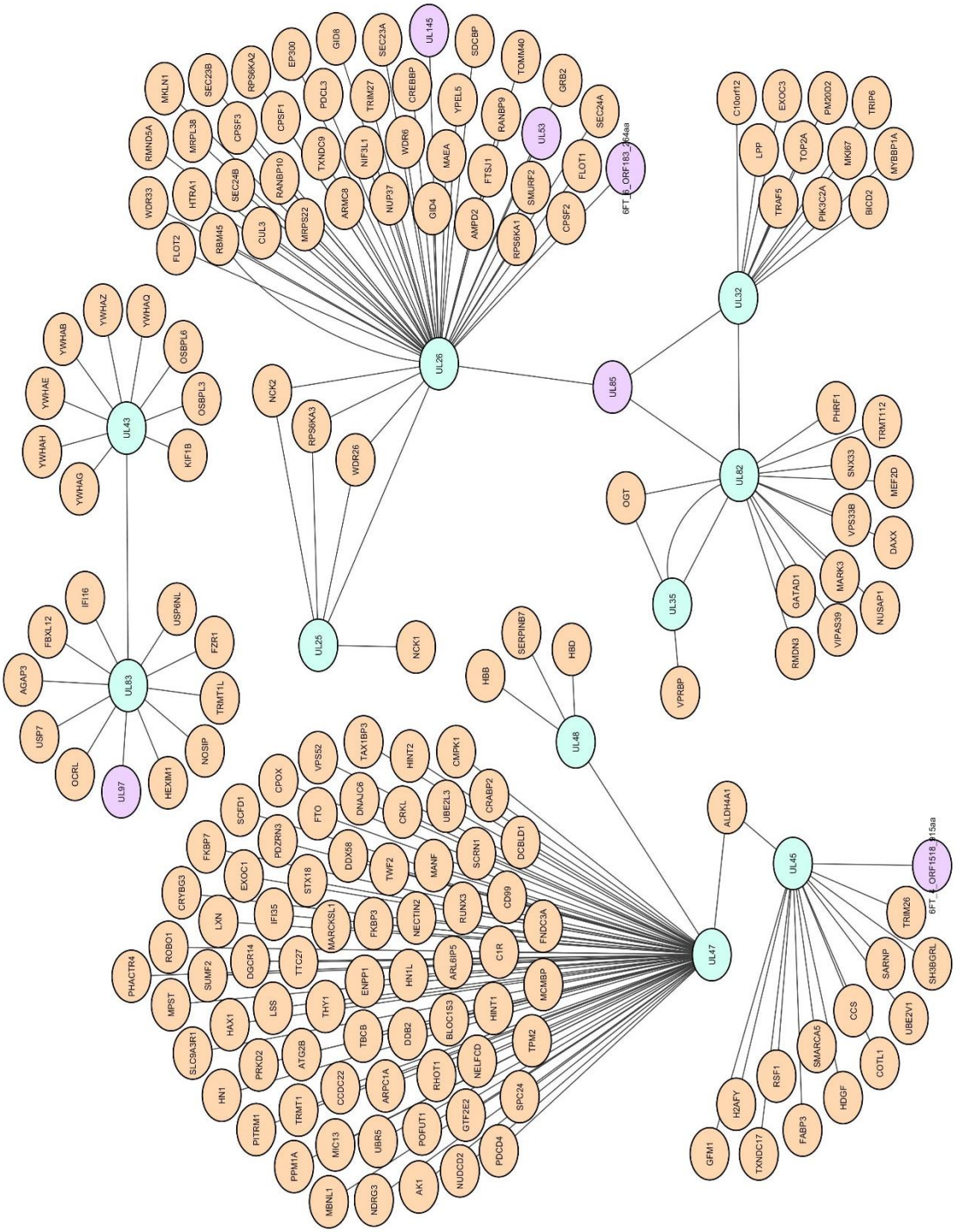


Figure 5.3.14 | Interaction diagrams for UL25, UL26, UL32, UL35, UL43, UL45, UL47, UL48, UL82 and UL83. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

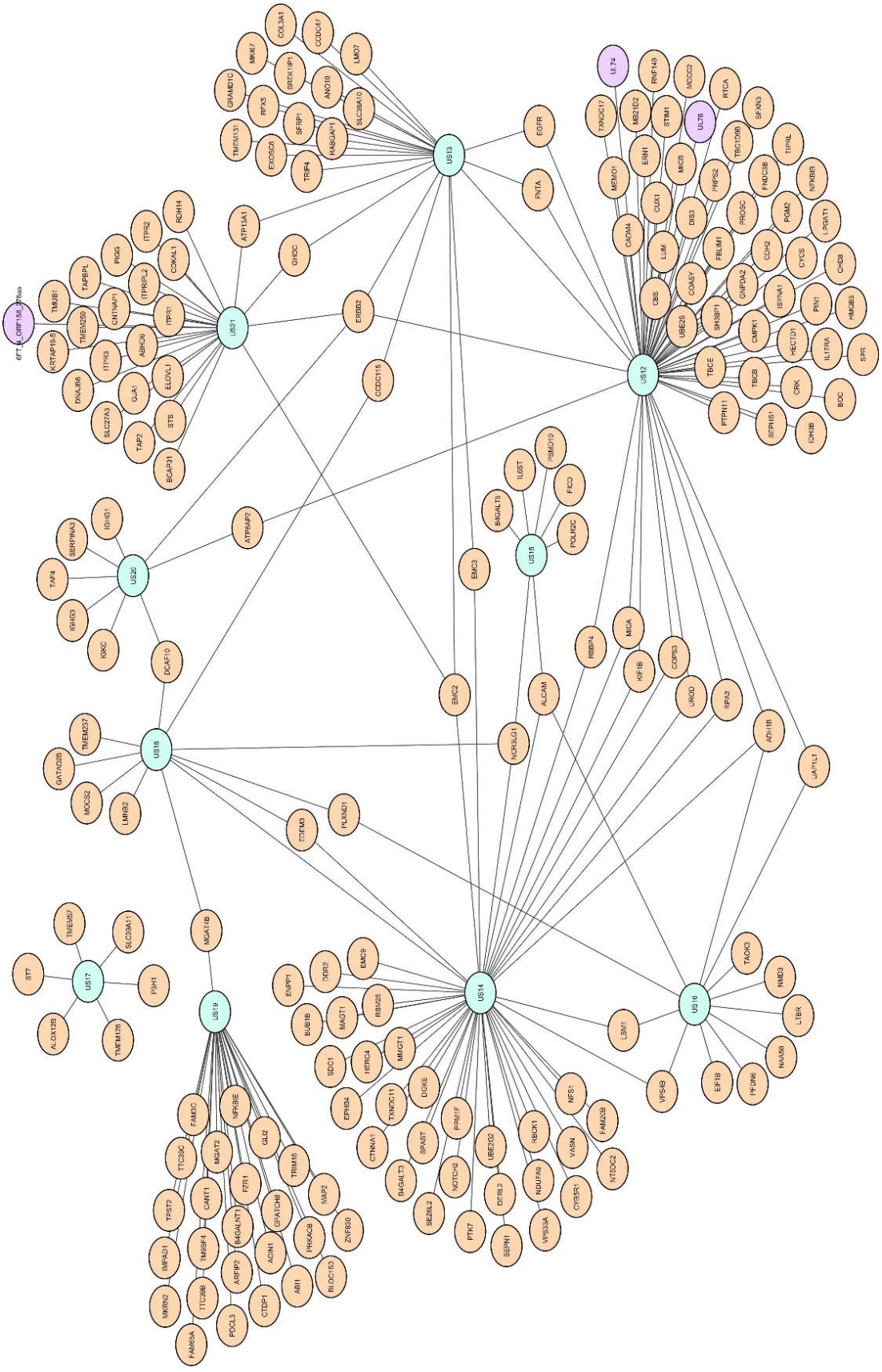


Figure 5.3.15 | Interaction diagrams for US12, US13, US14, US15, US16, US17, US18, US19, US20 and US21. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

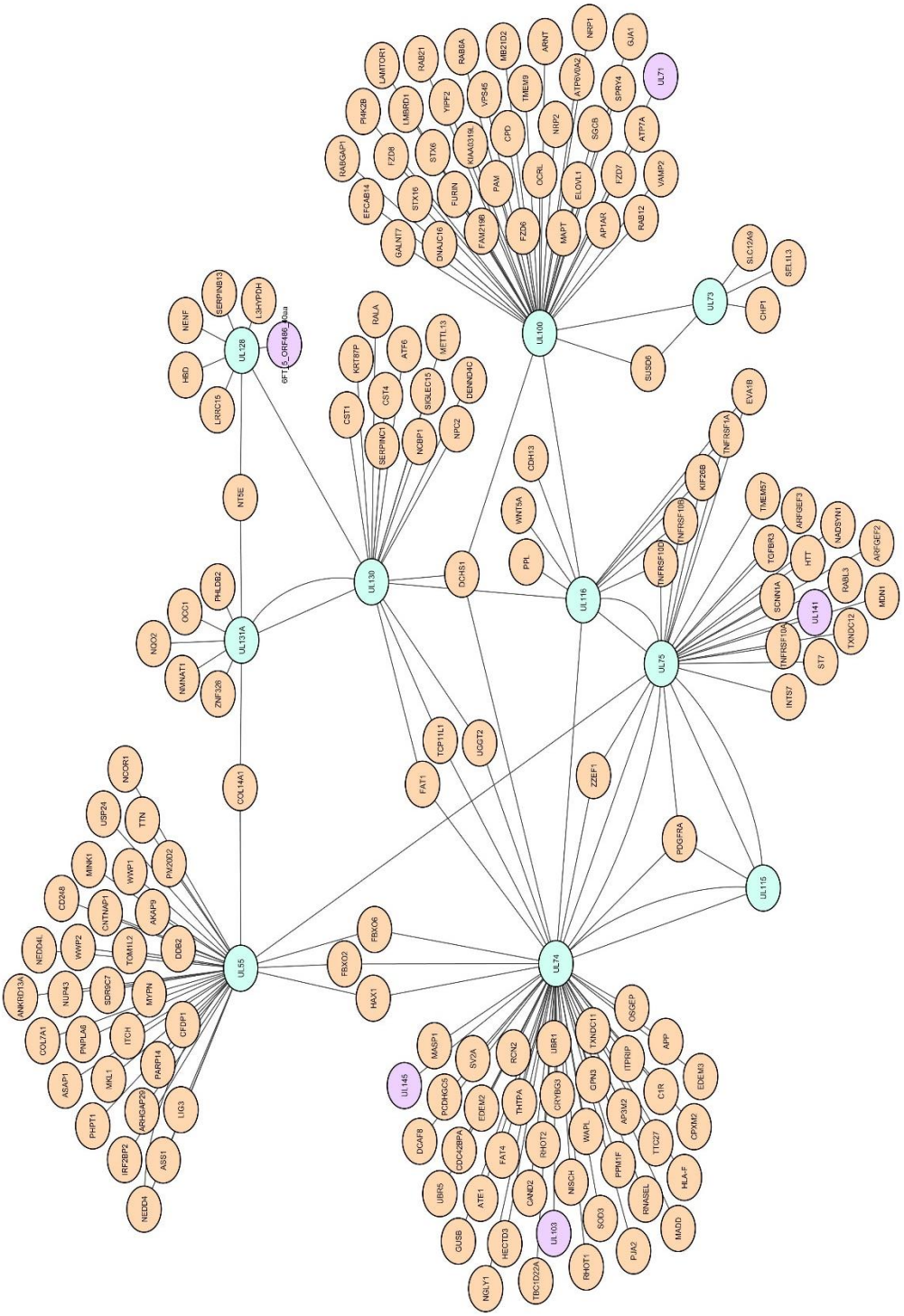


Figure 5.3.16 | Interaction diagrams for UL55, UL73, UL74, UL75, UL100, UL115, UL116, UL128, UL130 and UL131A. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

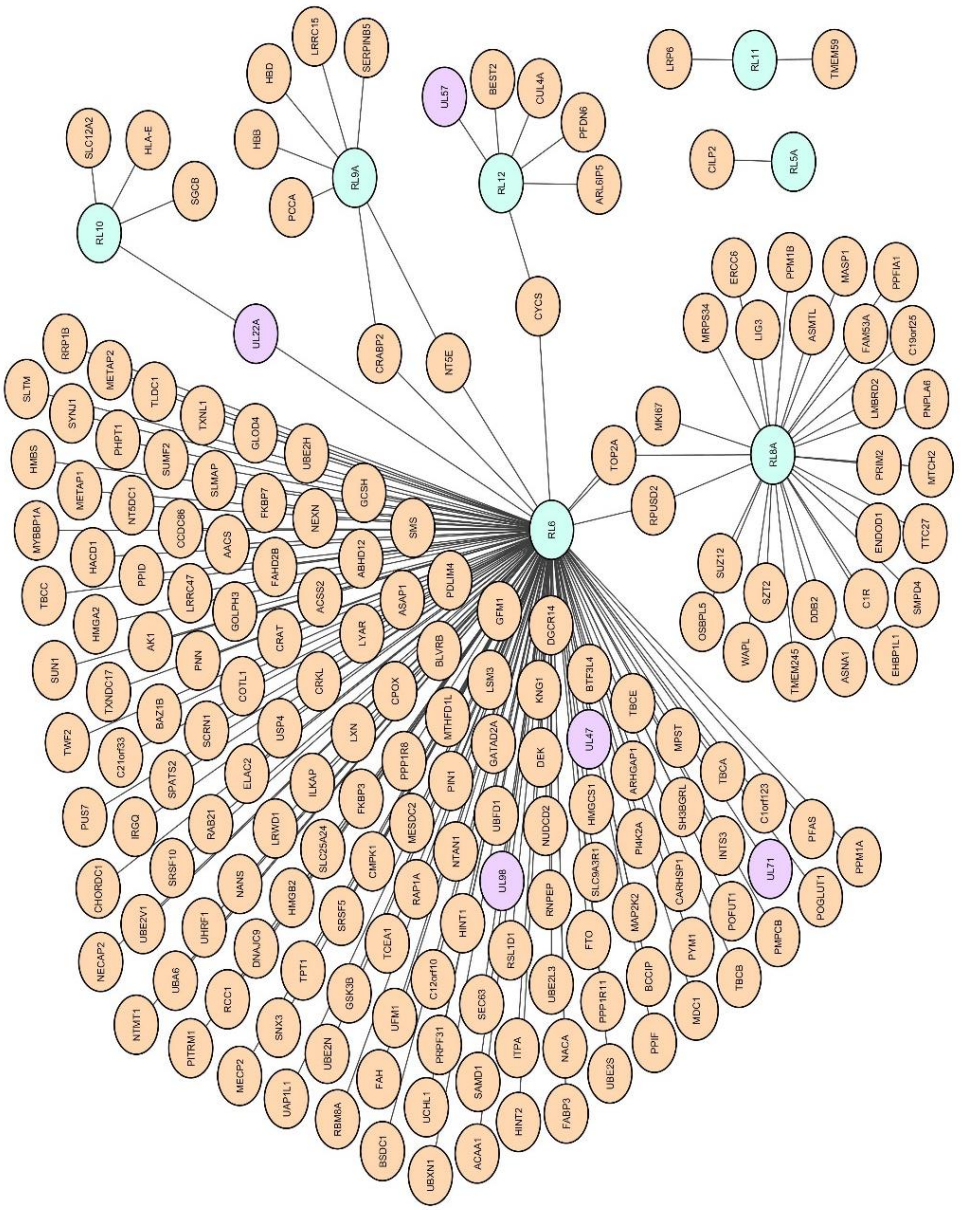


Figure 5.3.17 | Interaction diagrams for RL5A, RL6, RL8A, RL9A, RL10, RL11 and RL12. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

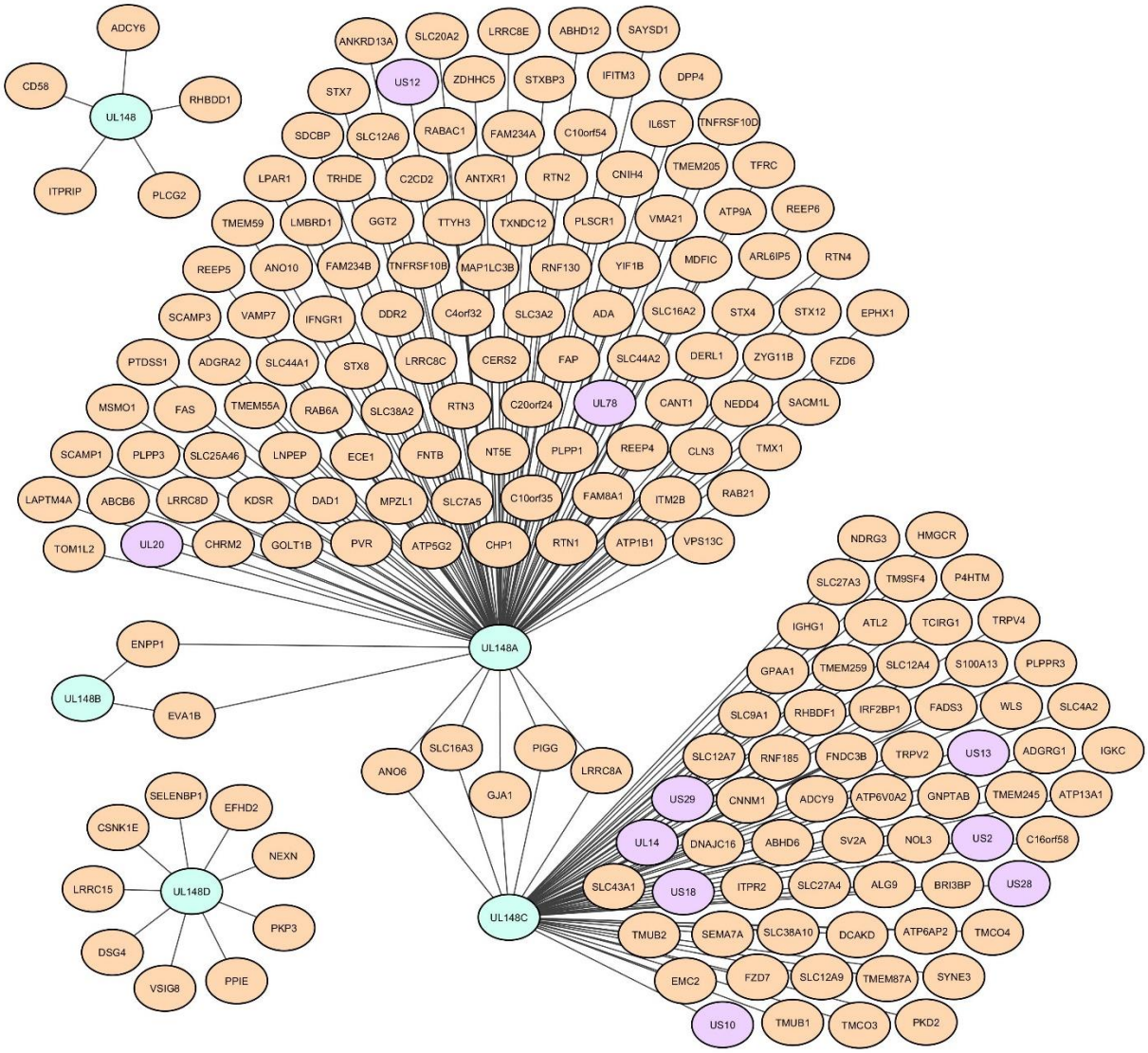


Figure 5.3.18 | Interaction diagrams for UL148, UL148A, UL148B, UL148C and UL148D. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

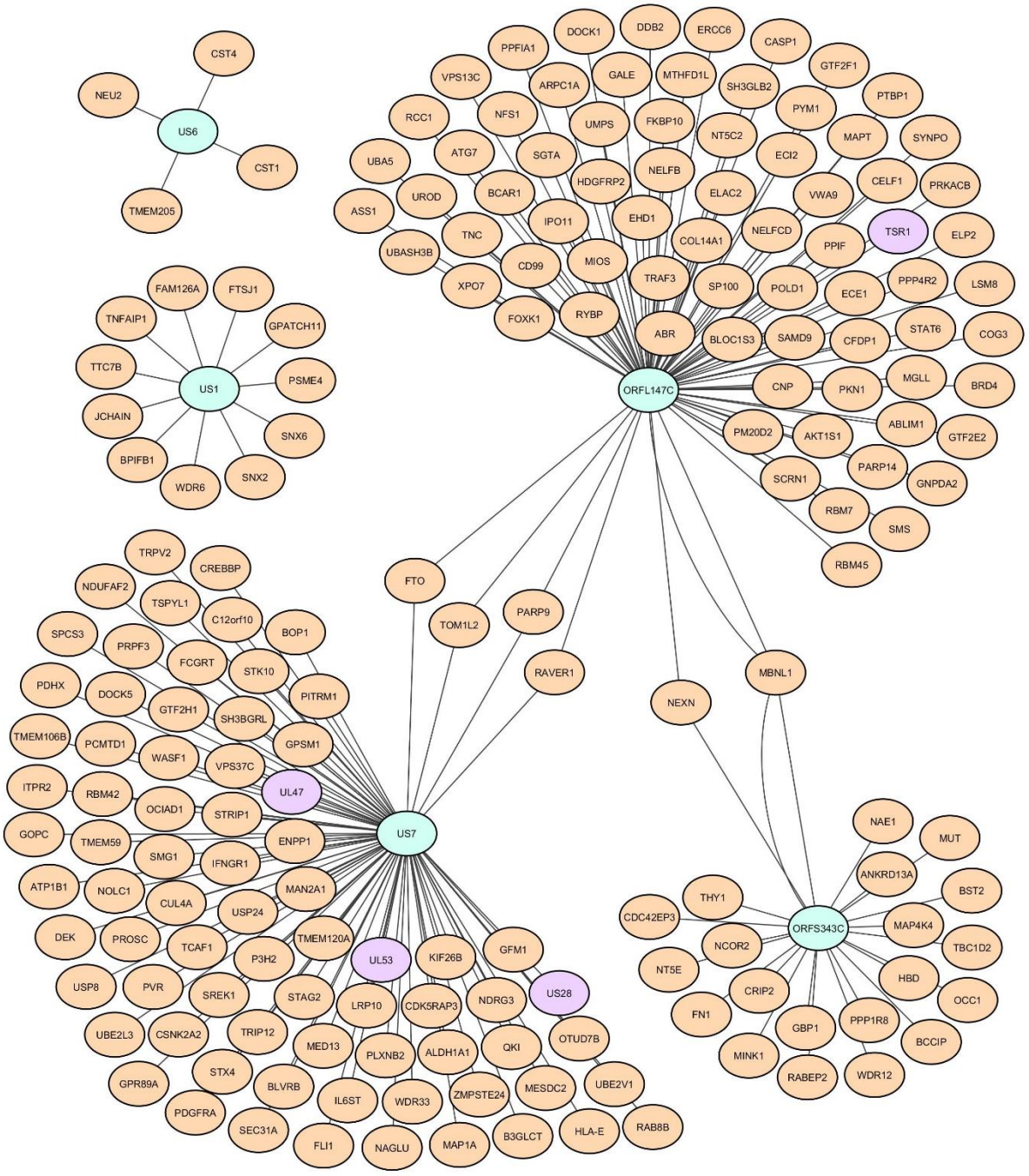


Figure 5.3.19 | Interaction diagrams for US1, US6, US7, ORFL147C and ORFS343C. Straight lines connect the viral baits (light blue) to their human interactors (light pink) and viral interactors (light purple).

A comparison of the list of HCIPs with the previously reported interactions from the combined database validated 59 HCIPs identified in this interactome (see Figure 5.3.20A).

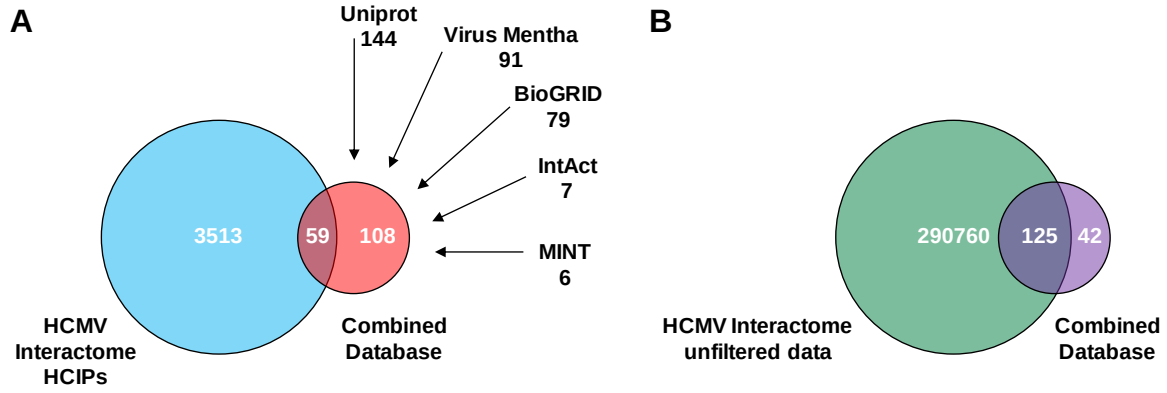


Figure 5.3.20 | Comparing HCMV interactome HCIPs with curated protein interaction data **(A)** HCMV protein interaction data annotated on Uniprot, Virus Mentha, BioGRID, IntAct and MINT was compiled into a combined database (167 interactions, red circle). Of these, 127 were identified in the unfiltered dataset from the HCMV interactome, with 59 (blue circle) meeting the stringent filtering criteria. **(B)** Overlap between the combined database and the unfiltered interactome data.

In the overlap with the filtered data, there were 26 interactions between viral proteins (see Figure 5.3.21), with 16 being reciprocal (i.e. A was identified as an HCIP of bait B, and B was identified as an HCIP of bait A). These included structurally relevant interactions for the virion such as binding between the major capsid (UL86) and small capsomere-interacting protein (UL48A) [235], large (UL48) and inner (UL47) tegument proteins [237], in addition to the envelope glycoproteins gM (UL100) and gN (UL73) [38]; gB (UL55), gH (UL75), gL (UL115) and gO (UL74) [39, 238, 239]. Viral-viral protein interactions relevant to the viral lifecycle and identified by this dataset included the binding between the tripartite terminase subunits TRM1 (UL56), TRM2 (UL51) and TRM3 (UL89) [20, 240], the nuclear egress proteins NEC1 (UL53) and NEC2 (UL50) [28], as well as the cytoplasmic envelopment proteins CEP2 (UL94) and CEP3 (UL99) [241].

The remaining viral-viral interaction positive controls were part of the DNA replication machinery and included binding between the lytic DNA synthesis trans-acting factors UL84 and IE2 (UL122) [242, 243]; the complex formed by the viral helicase (UL105), primase (UL70) and the helicase-primase associated factor (UL102) [244]; the viral DNA polymerase catalytic subunit (UL54), its processivity factor (UL44) and the uracyl-DNA glycosylase (UL114) [245-247].

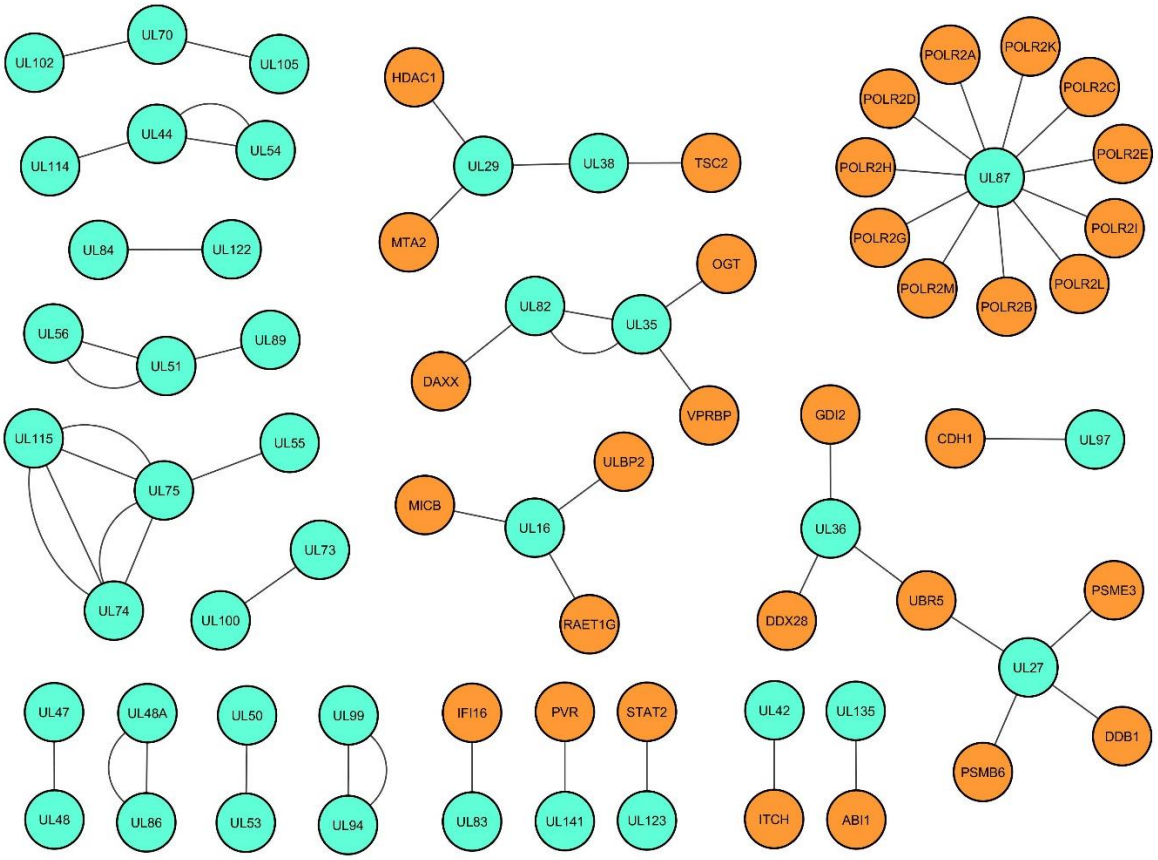


Figure 5.3.21 | Positive controls highlighted from overlap of HCMV interactome with curated databases. Representation of known interactions between HCMV (light blue) and host (orange) proteins. Reciprocal interactions are shown as two lines (one straight, one curved) connecting the protein names.

Interactions between viral and human proteins provided additional positive controls, many relating to subversion of the host immune system. For example, binding

between pp65 (UL83) and the innate immune viral DNA sensor IFI16 or pp71 (UL82) and hDaxx, a component of the viral genome silencing complex Promyelocytic leukemia protein nuclear bodies (PML-NBs) [165, 248]. Binding of IE1 to STAT2 disrupts interferon signalling, while interaction between UL135 and the WAVE2 complex member ABI1 triggers remodelling of the actin cytoskeleton in order to reduce efficiency of immune synapse formation [249, 250]. Other synergistic interactions that result in override of host defence mechanisms are the interaction of UL141 with PVR, causing retention in the ER for this ligand of activating NK-cell receptors, as well as the interaction of UL16 with the NK-cell activating KLRK1/NKG2D receptor ligands ULBP2, ULBP5 and MICB [251-254].

Other viral-human protein interactions related to the hijacking of cellular machinery in order to facilitate viral replication, as for example the interaction of several subunits of RNA polymerase II with UL87, a component of a viral protein complex required for the transcription of true late genes [187].

Of the remaining 108 protein interactions featured in the combined database list, 42 were not detected in the interactome, as the prey was not quantified, and 66 did not pass the stringent scoring thresholds employed. A larger degree of overlap (approximately 75%) was observed when comparing the combined database with the unfiltered interactome data (see Figure 5.3.20B). Full data regarding the overlap with the combined database is shown in Appendix D.

5.4 Functional enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) [216] is a bioinformatics platform that provides a comprehensive set of functional annotation tools to enable understanding of biological meaning behind large lists of genes

by identifying enriched biological themes such as 'Gene Ontology' terms, protein domains and cellular pathways, among others. Thus, to gain an overview on the pathways targeted by all proteins during infection, DAVID functional annotation clustering was used to determine which cellular and molecular terms were enriched amongst the 3416 human proteins that interact with viral baits.

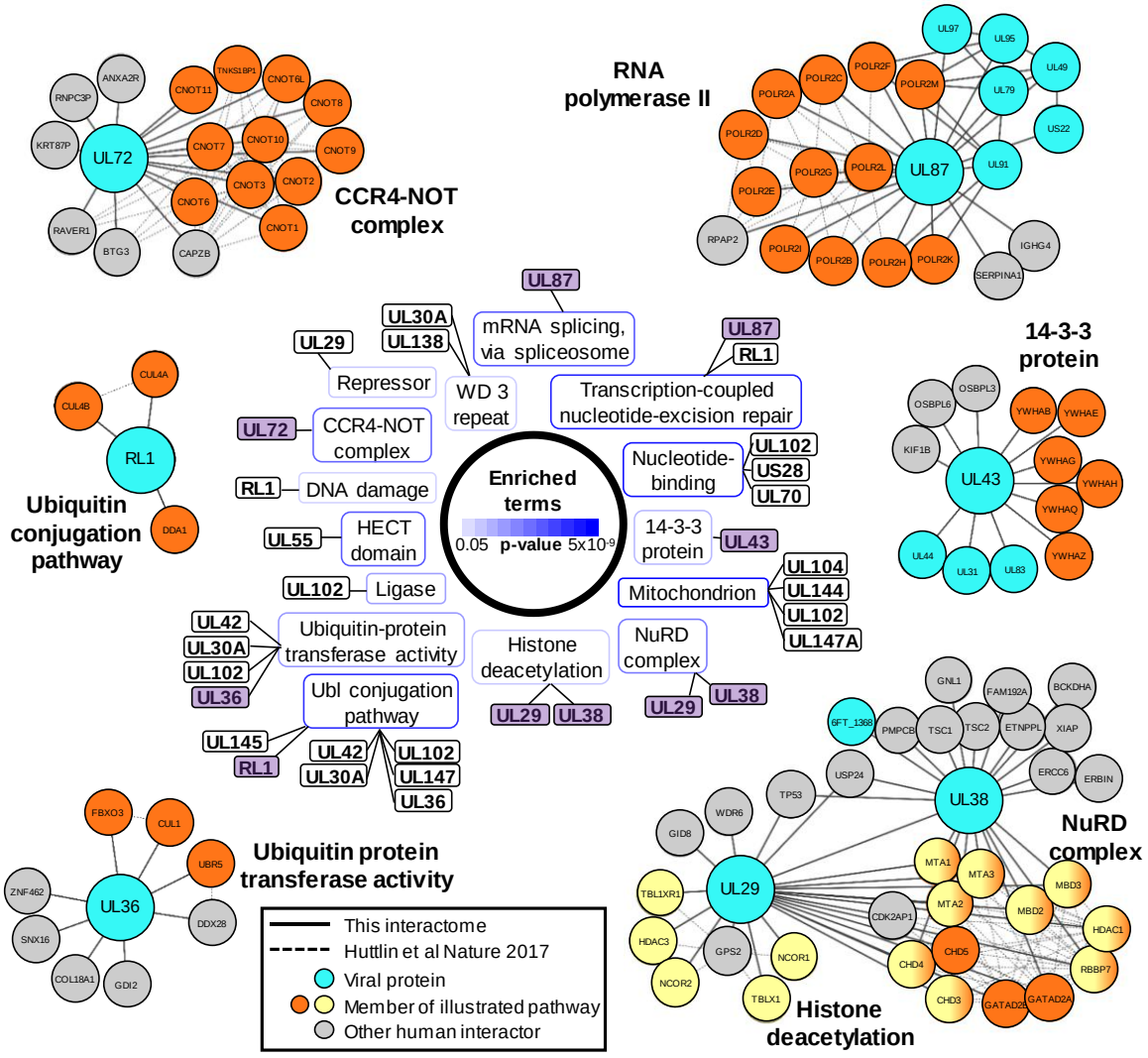


Figure 5.4.1 | Functional enrichment analysis of HCMV interactome data
DAVID software with default settings [216] was applied to determine which pathways were enriched amongst all interactome HCIPs, using the human proteome as a 'background'. Benjamini-Hochberg adjusted p-values are shown as blue surrounds to each pathway enriched at p<0.05. Viral baits are linked to enriched pathways where >33 % of human interacting proteins belonged to a given pathway, and examples are shown around the outside of the figure. For example, 6/9 (67 %) human HCIPs for UL43 were part of the 14-3-3 protein. These examples are indicated in the central part of the figure by purple shading. Viral baits are shown as large turquoise circles, and interacting viral proteins as smaller turquoise circles. Members of enriched pathways are shown in orange or yellow (for NuRD complex and histone deacetylation, protein membership of both pathways is indicated by half-orange, half-yellow circles). Solid lines indicate interactions identified in the HCMV interactome, and dashed lines indicated interactions drawn

from the human interactome (Bioplex 2.0) and subsequent unpublished data (http://bioplex.hms.harvard.edu/downloadInteractions.php).

Components of the Nucleosome Remodelling and Deacetylase (NuRD) complex were significantly enriched among HCMV-interacting proteins. The NuRD complex (see Figure 5.4.2A) is one of the major chromatin remodeling complexes in mammalian cells, and is known to be co-opted by HCMV UL29 and UL38 to enhance expression of immediate-early genes [174, 255]. The interactome confirmed that UL29 and UL38 interact in a complex with all components of the NuRD complex, in addition to p53 [255]. UL29 was also found to interact with several human proteins that function in histone deacetylation. Five of these proteins (NCOR1, NCOR2, TBL1X, TBL1XR1 and HDAC3) interacted with UL29, but not UL38, and are in fact components of the Nuclear receptor corepressor (NCoR) complex (see Figure 5.4.2B) [256-258]. Another component of the NCoR complex, GPS2 was also identified in the interactome as an HCIP of UL29. These interactions had not been previously reported.

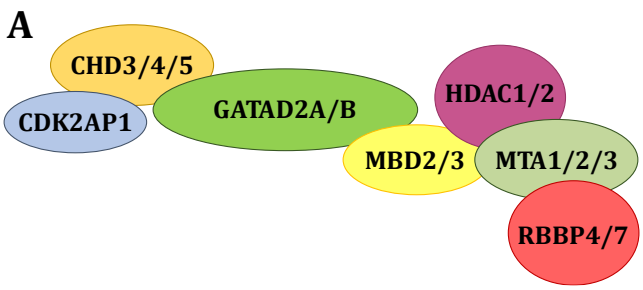


Figure B removed for copyright reasons. Copyright holder is *Springer Nature*

Figure 5.4.2 | Diagram of the NuRD and NCoR complexes
(A) The **Nu**cleosome **R**emodeling and **D**eacetylase (NuRD) complex is comprised of proteins with chromatin remodelling and histone deacetylation activity. It consists of two subcomplexes containing a total of seven different proteins: the nucleosome remodelling subcomplex contains one chromodomain-helicase-DNA-binding protein (CHD3/CHD4/CHD5), one cyclin-dependent kinase 2-associated protein 1 (CDK2AP1), and one transcriptional repressor (GATAD2A\GATAD2B); one methyl-CpG-binding domain protein MBD2/MBD3 protein bridges the remodelling subcomplex to the histone deacetylase subcomplex, which consists of histone deacetylase core proteins HDAC1/HDAC2 proteins, two metastasis-associated proteins MTA1/MTA2/MTA3 proteins, and four histone-binding proteins RBBP4/7 proteins (adapted from Hoffmann and Spengler, 2019) [259]. **(B)** The **N**uclear receptor **Co**-**R**epressor complex binds to ligand-free nuclear receptors and represses transcription, partly by deacetylating histones. The

complex consists of the F-box-like/WD repeat-containing proteins TBL1X and TBL1XR1, G protein pathway suppressor 2 (GPS2), the nuclear receptor co-repressor 1 or 2 (NCOR1/NCOR2) and the histone deacetylase HDAC3 (adapted from Emmett and Lazar, 2019) [260].

HCIP analysis of the tegument protein UL43 showed an enrichment for 14-3-3 proteins, a family of regulatory molecules that can bind a wide range of functionally diverse signalling proteins, including kinases, phosphatases, and transmembrane receptors [261]. Members of this protein family have been identified in virions from HCMV, HSV-1, KSHV and PRV (pseudorabies virus/suid herpesvirus 1 – does not infect humans), but not EBV (see Table 5.4.1). For HSV-1, 14-3-3 beta has been shown to interact with its tegument protein UL46 [262], while a small interfering RNA functional screen found that 14-3-3 zeta/delta supported viral proliferation in cell culture [263]. It remains to be determined whether 14-3-3 proteins can influence HCMV infection.

Table 5.4.1 | 14-3-3 Proteins quantified in virions from the *Herpesviridae* family

14-3-3 Family Members	HCMV [56]	HSV-1 [264]	KSHV [265]	EBV [266]	PRV [267]
14-3-3 beta/alpha (YWHAB)	X	X	-	-	X
14-3-3 epsilon (YWHAE)	X	X	-	-	X
14-3-3 gamma (YWHAG)	-	X	-	-	-
14-3-3 eta (YWHAH)	-	-	-	-	-
14-3-3 theta (YWHAQ)	X	-	-	-	X
14-3-3 sigma (YWHAS)	-	-	-	-	X
14-3-3 zeta/delta (YWHAZ)	X	X	X	-	-

The HCMV genes UL87, UL49, UL79, UL88, UL91, UL92 and UL95 are conserved between beta- and gammaherpesvirus, but not alphaherpesvirus. All these genes except for UL88 have been shown to be necessary for transcriptional activation of viral genes expressed with 'true late' kinetics, and it has been suggested that these proteins may form one or more complexes that modulate the activity of RNA polymerase II [268-271]. Interactome data confirmed that UL87 interacted with UL79, UL49, UL91 and UL95 but

did not detect a high-confidence interaction with UL92. Several interactions between UL92 and UL91, UL87, UL95 and UL79 were detected but did not meet the filtering criteria. The lack of identification of any high-confidence viral-viral UL92 interactions may be due to UL92 being one of the two least abundantly expressed viral proteins during HCMV infection (see Figure 1.4). UL87 also interacted with all 12 components of the RNA polymerase II (RPII) complex and the associated protein RPII Associated Protein 2 (RPAP2). The UL87-RPII interaction was anticipated by analogy to the orthologous RPII-interacting EBV protein BcRF1, but had not previously been demonstrated. Interaction of UL87, UL95 and UL79 with the UL97 protein kinase was also novel. UL97 has been shown to phosphorylate the carboxyl-terminal domain of RPII located in the POLR2A subunit [272], however this interaction was not detected in the interactome potentially due to its transient nature.

UL72 is regarded as the evolutionary counterpart of the deoxyuridine 5'-triphosphate nucleotidohydrolase (dUTPase) in other herpesviruses, but lacks dUTPase activity [273]. UL72 interacted with all 11 components of the CCR4-NOT (carbon catabolite repressor 4-negative on TATA) complex (see Figure 5.4.3), which is a key regulator of gene expression from production of mRNAs in the nucleus to their degradation in the cytoplasm [274]. Additionally, UL72 also interacted with BTG3 which has been previously isolated by co-IP with subunits of this complex [275]. The interaction between UL72 and CNOT2/CNOT7 was confirmed by co-IP in transiently transfected HEK293T cells, and in HFFF-TERT cells stably overexpressing V5-tagged UL72 (see Figure 5.4.4).

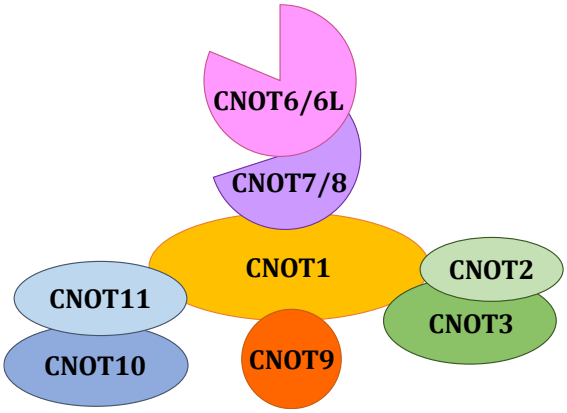


Figure 5.4.3 | Diagram of the CCR4-NOT complex
The carbon catabolite repression 4 (CCR4)–negative on TATA-less (NOT) complex plays a fundamental role in eukaryotic mRNA metabolism and has a multitude of different roles that impact eukaryotic gene expression. CNOT1 acts as a scaffold subunit around which the other members of the complex assemble, including the deadenylase subunits CNOT6/6L and CNOT7/8 (adapted from Shirai et al, 2014) [276].

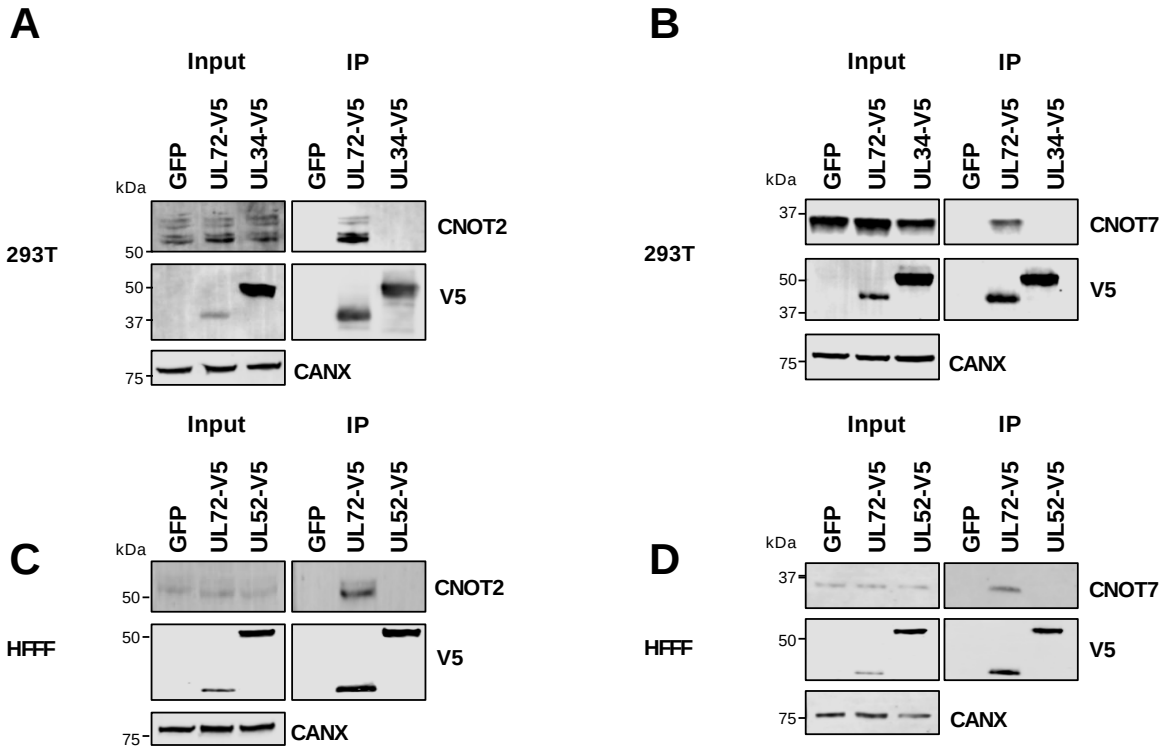


Figure 5.4.4 | Interaction of UL72 with members of CCR4-NOT complex CNOT2 and CNOT7
Co-IPs validating that UL72 interacts with CNOT2 and CNOT7, conducted in HEK293T cells (**A** and **B**) or HFFFs (**C** and **D**). For all experiments in this figure, left panels show an IB of 1-2 % of input sample, and right panels show an anti-V5 co-IP. HEK293T cells were transiently transfected with two plasmids, one expressing the C-terminally V5-tagged viral protein and the other expressing the C-terminally HA-tagged cellular prey. Bait proteins were detected with anti-V5, and prey with antibodies against CNOT7 or CNOT2 protein. Controls included GFP and the viral proteins UL34 or UL52. CANX – calnexin loading control. This figure is representative of n= 1 experiment (A); n= 2 experiments (B); n= 2 experiments (C); n= 1 experiment (D). Expected sizes: CNOT7: 33 kDa; CNOT2: 52 kDa; CANX: 72 kDa; UL72: 44 kDa; UL34: 45 kDa.

It is possible that UL72 directly binds just one subunit of the CCR4-NOT complex and co-IPs the whole complex via interaction with that protein. To determine if a single CCR4-NOT complex subunit was preferentially enriched in the UL72 AP-MS experiments, a ratio between the average number of PSMs (in both replicates) and the length of each protein was calculated (see Table 5.4.2). It is necessary to account for the effect of protein length, as large proteins tend to yield more peptides than small ones. This ratio is equivalent to a normalised spectral abundance factor, but without the normalisation by dividing by the sum of all PSMs/protein length for all proteins in the experiment. This analysis showed a wide spread of ratios, and whereas it appears that CNOT6 was relatively less enriched than CNOT2, there was no clear result. Thus, preferential affinity for just one subunit cannot be identified from this analysis. Other complementary methods would be required to address this question, for example yeast two-hybrid could be employed to test interactions from UL72 with each subunit.

Table 5.4.2 | Details of CCR4-NOT complex subunits in UL72-V5 AP-MS experiments

CCR4-NOT complex subunit	Uniprot	Average PSMs	Protein length (aa)	Average PSMs/ Protein length ratio
CNOT1 (isoform 1)	A5YKK6	296.5	2,376	12.5 %
CNOT1 (isoform 2)	A5YKK6-2	4	2,371	0.2 %
CNOT2	Q9NZN8	101	540	18.7 %
CNOT3	O75175	51.5	753	6.8 %
CNOT6	Q9ULM6	11.5	557	2.1 %
CNOT6L	Q96LI5	24	555	4.3 %
CNOT7	Q9UIV1	29	285	10.2 %
CNOT8	Q9UFF9	10	292	3.4 %
CNOT9	Q92600-2	11	331	3.3 %
CNOT10	Q9H9A5-6	30	804	3.7 %
CNOT11	Q9UKZ1	21	510	4.1 %

The CCR4-NOT complex is targeted for degradation during adenovirus infection in order to promote expression of early viral proteins and increase concentration of viral DNA [277]. Degradation of CCR4-NOT complex members does not seem to be induced by

HCMV infection [162], thus the biological function of UL72 binding to this complex remains to be determined.

UL145 has recently been shown to recruit the Cullin 4 E3 ligase scaffold and associated adaptor proteins in order to degrade the helicase-like transcription factor HLTF [163]. Interactome data suggested that all human proteins interacting with UL145 and the paralogous RL1 were part of the ubiquitin conjugation pathway, and furthermore that RL1 interacted with Cullin 4. Using co-immunoprecipitation in transiently transfected 293Ts with RL1-V5 and CUL4A-HA, Dr Katie Nightingale validated the RL1-CUL4A interaction (see Figure 5.4.5A). Proteins that are degraded after binding RL1/CUL4 still require identification; it is possible that their abundance after degradation may have been insufficient to enable identification in this study, as for example the interaction between UL145 and its degraded target HLTF was not detected in the interactome. Multiple other HCMV proteins additionally interacted with elements of the ubiquitin transfer or conjugation pathways, including the DNA helicase/primase associated factor UL102, which interacted with the E3 ligase RNF114 and E2 conjugating enzyme UBE2L6. Similarly, the inhibitor of apoptosis UL36 which bound the Cullin 1 scaffold, E3 ligase UBR5, and F-box component FBOX3.

UL71 has been shown to play a role in the secondary envelopment. However, this viral protein is expressed with Tp3 kinetics, suggesting that this tegument protein may also play a role earlier during infection [162, 278, 279]. The interactome identified interactions of UL71 with multiple interferon-stimulated proteins (see Figure 5.4.5B), including TRIM22, a restriction factor for HIV-1, influenza A and hepatitis B and C viruses [280]. Using co-immunoprecipitation in transiently transfected 293Ts with UL71-V5 and TRIM22-HA, Dr. Katie Nightingale validated the UL71-TRIM22 interaction, suggesting

that investigation of a putative innate immune role for UL71 will be important (see Figure 5.4.5C).

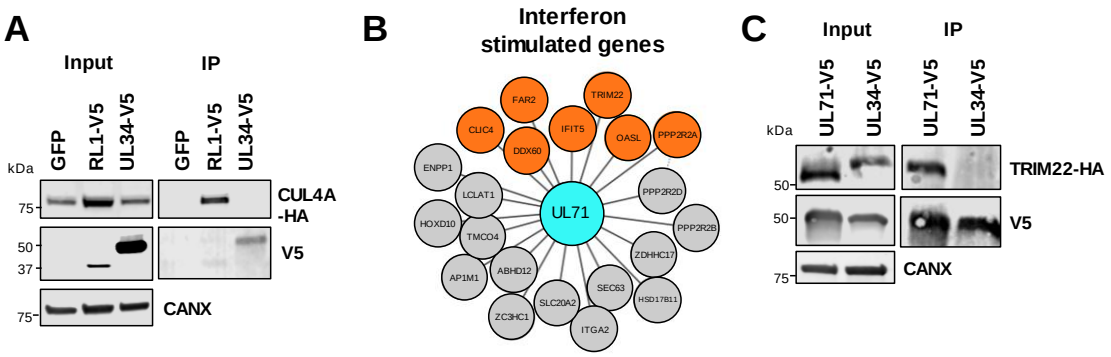


Figure 5.4.5 | Validation of the interaction between RL1 and CUL4A, UL71 and TRIM22. **(A)** Co-IP validating the interaction between RL1 and CUL4A, conducted in HEK293T cells as described in Figure 5.4.4, but with detection of CUL4A using anti-HA. This figure is representative of n= 4 experiments. Expected sizes: CUL4A: 77 kDa; RL1: 35 kDa; UL34: 45 kDa; CANX: 72 kDa. **(B)** HCMV UL71 interacted with multiple interferon-stimulated proteins, including TRIM22. **(C)** Co-IP validating the interaction between UL71 and TRIM22, conducted as described in Figure 5.4.4. This figure is representative of n= 3 experiments. Expected sizes: TRIM22: 56 kDa; UL71: 40 kDa; UL34: 45 kDa; CANX: 72 kDa.

Using functional enrichment analysis, an alternative approach identified enriched terms whose members interacted predominantly with single baits. Figure 5.4.6 shows enriched terms with $p<0.05$ (after Benjamini-Hochberg adjustment) and for which >33 % of the identified members of the enriched term interacted with a given viral bait. For example, US28 interacted with all quantified members of thick filament/muscle myosin complexes (myosin heavy and light chain components, a myosin binding protein and titin), suggesting a putative role for US28 in processes such as cytoskeletal remodeling [281]. Another viral GPCR, US27 interacted with multiple components of the SNARE complex, which mediates vesicle fusion [282]. The envelope glycoprotein UL132 interacted with the AP-2 adaptor complex, which functions in clathrin-mediated endocytosis (see Figure 5.4.6) [283]. Altogether, this data suggested a putative role for these viral proteins in modulation of vesicular transport.

Functional enrichment analysis was then used to gain insights into the temporal regulation of protein-protein interactions. For this, viral baits were segregated according to their temporal class and lists of their human HCIPs were analysed, using a list of all human HCIP as a background [162]. This analysis yielded an enrichment in functions required at different stages of the viral life-cycle (see Figure 5.4.7A). For example, HCIPs of Tp1 and Tp2 proteins were enriched in NuRD complex members, proteins involved in histone deacetylation and proteins with SANT domains (which function in chromatin remodelling). HCIPs of Tp3 proteins were enriched in functions required for viral DNA replication and immune evasion. HCIPs of Tp5 proteins were enriched in terms related to intracellular trafficking and secretion (see Figure 5.4.7A).

A similar analysis was performed on viral HCIPs to assess the temporal regulation of viral-viral protein interactions. Two general patterns were observed from this analysis, viral proteins either interacted mainly with others from the same class or adjacent classes, or with proteins from the largest class Tp5 (see Figure 5.4.7B). For example, Tp1 and Tp2 class proteins UL29 and UL38 interacted, as previously reported (see Figure 5.4.1). Tp1-class tegument proteins US23 and US24 interacted. The majority of Tp5 interactions were with other Tp5 proteins, 15/37 of which were tegument-tegument, capsid-capsid or tegument-capsid protein interactions (see Figure 5.4.7B). Interactions between proteins in different temporal classes have long been reported and include the interaction between the DNA polymerase subunits UL54 (Tp2) and UL44 (Tp5). This interactome has now identified novel interactions between other distinctly expressed proteins, such as the interaction between the functionally unknown membrane protein UL14 (Tp2) and two Tp5-class proteins, membrane protein UL121 and envelope glycoprotein UL4.

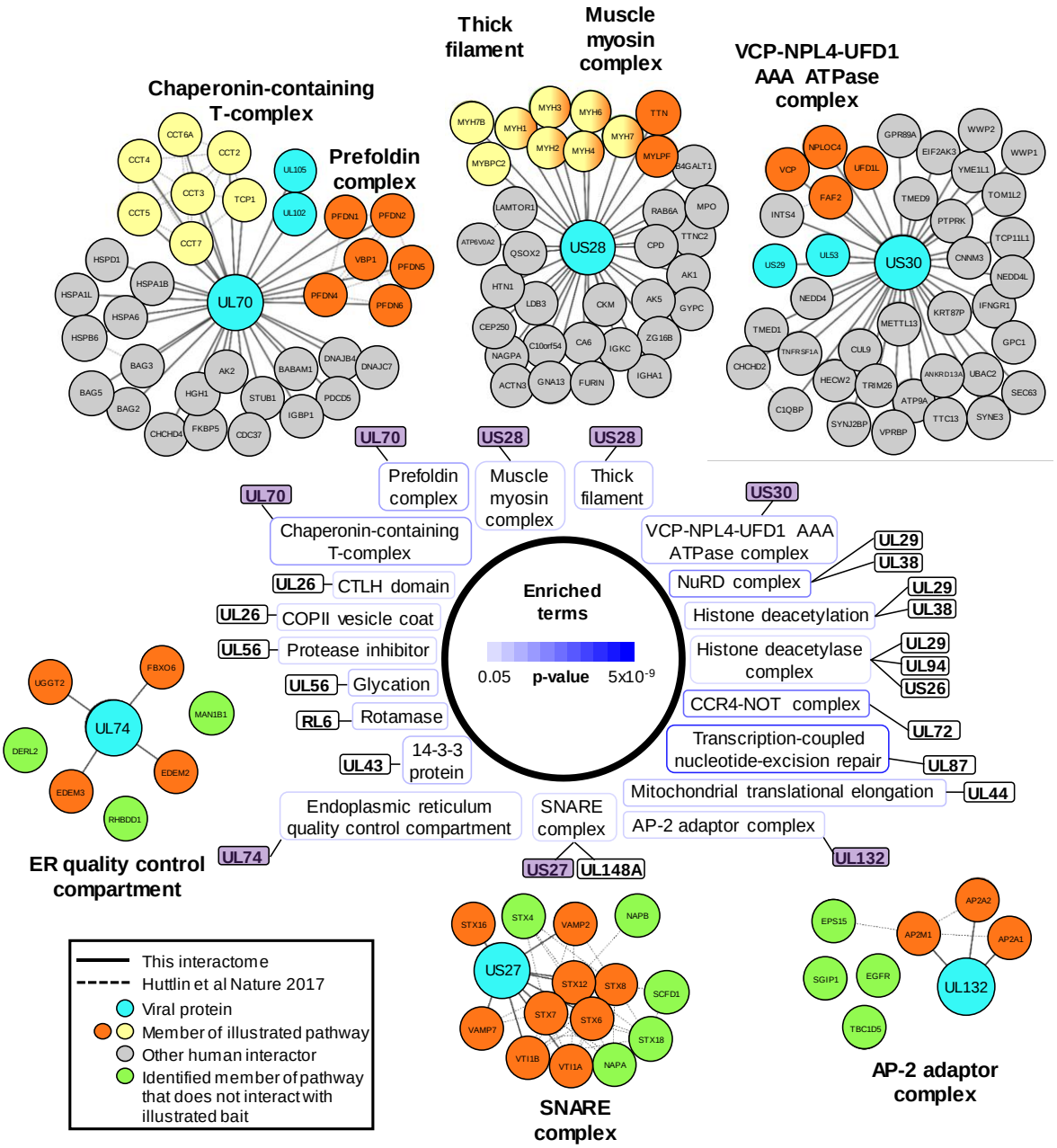


Figure 5.4.6 | Pathways enriched with p<0.05 (after Benjamini-Hochberg adjustment) and for which >33 % of the identified components interacted with a given viral bait. Examples are indicated in the central part of the figure by purple shading. All members of the thick filament/muscle myosin complex detected in this interactome interacted with US28 (100 %). For the bottom three complexes (UL74, US27 and UL132), each viral bait interacted with a total of 52-107 proteins. For simplicity, only members of the illustrated pathway identified in this interactome are displayed. For example, 14 members of the SNARE complex were enriched in the interactome, of which 9 interacted with US27 (64 %). Green circles show members of a pathway that were detected in the interactome but did not interact with the bait. Viral baits are shown as large turquoise circles, and interacting viral proteins as smaller turquoise circles. Members of enriched pathways are shown in orange or yellow (membership of two pathways is indicated by half-orange, half-yellow circles). Solid lines indicate interactions identified by this interactome, and dashed lines indicated interactions derived from the human interactome (Bioplex 2.0) and subsequent unpublished data (http://bioplex.hms.harvard.edu/downloadInteractions.php).

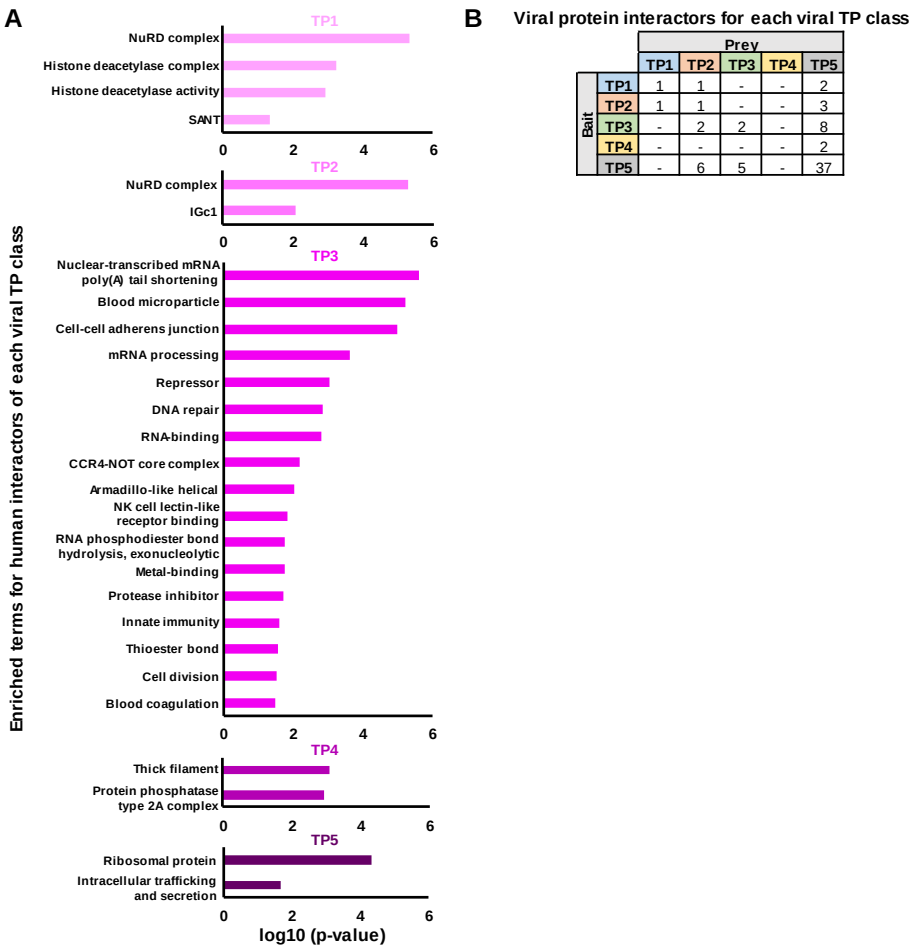


Figure 5.4.7 | Further details of interactions according to viral protein temporal class
(A) Functional enrichment of human HCIPs according to the temporal class of their viral bait. DAVID software with default settings [216] was applied using all human HCIP as a 'background'. Benjamini-Hochberg adjusted p-values are shown on the x-axis. **(B)** Temporal analysis of viral-viral protein interactions. This analysis was only performed on interactions for which both bait and prey had a defined Tp class according to Weekes et al (2014) [162].

DAVID analysis can also be applied to individual lists of genes with a considerable number of interactors. US22 is a tegument protein and an uncharacterised member of the betaherpesvirus-specific US22 family. DAVID analysis of the 61 HCIPs of US22 revealed an enrichment for proteins containing RNA recognition motifs (see Figure 5.4.8). Among the 61 US22 HCIPs, 24 were associated with the term 'transcription', while 14 were related to the term 'repressor' including TASOR and Periphilin-1. These two proteins, in addition to MPP8, form the epigenetic repressor Human Silencing Hub (HUSH) complex

[284]. Validation of these interactions is necessary, however DAVID analysis may provide an unbiased approach to determine individual gene function.

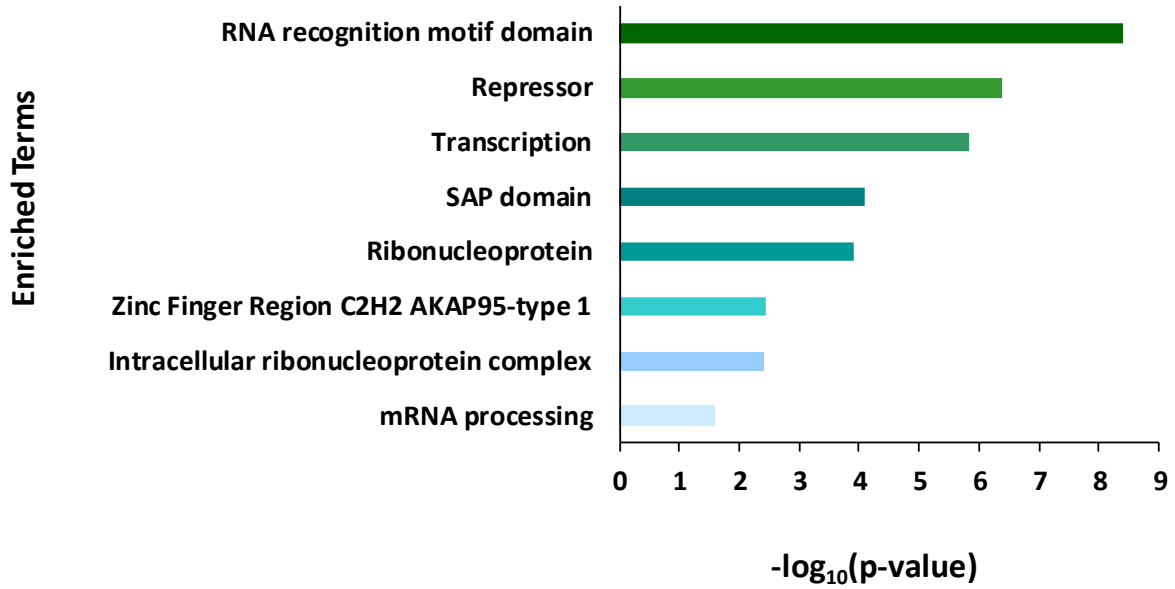


Figure 5.4.8 | Functional enrichment analysis of US22 HCIPs
DAVID software with default settings [216] was applied to determine which pathways were enriched amongst all US22 HCIPs, in comparison to all human proteins as background. Negative log base 10 for Benjamini-Hochberg adjusted p-values is shown as a measurement of enrichment.

5.5 Viral proteins that degrade cellular prey

A multiplexed approach has recently been employed to discover proteins that have innate immune function on the basis of their active degradation by the proteasome or lysosome during the early phase of HCMV infection. In Nightingale et al, 2018, three orthogonal proteomic/transcriptomic screens were employed to quantify protein degradation. The combination of the three screens included a shortlist of 133 proteins degraded by the proteasome or lysosome during early phase infection [163]. A final screen employed a panel of HCMV gene-block deletion mutants, to aid mapping of viral gene functions. However, this screen was unable to confidently identify the genetic region containing the viral effector which targeted 121/133 degraded proteins. Additionally, depending on the number of genes within each block, identification of which individual

viral genes degraded cellular targets could prove arduous. For example, identification of the viral factor within the UL133-UL150 block, UL145, which targeted HLTF to the proteasome required testing of 19 single viral gene deletion mutants [163].

In order to extend the findings of the study by Nightingale et al (2018), HCIP data from this interactome was combined electronically to identify viral factors interacting with the shortlist of 133 degraded host factors. This overlap identified viral interactors for 31 of these degraded prey (see Table 5.5). A subset of these interactions has already been described in the literature and are shown on Figure 5.3.3, for example the binding and sequestering of Signal transducer and activator of transcription 2 (STAT2) by IE1-72 [285]. Additionally, the ubiquitin E3 ligase ITCH (Itchy E3 Ubiquitin Protein Ligase) is also known to be targeted for degradation by viral UL42 [286].

Isoform A of the nuclear autoantigen Sp-100 is known to be targeted for degradation via proteasome and to interact directly with IE1 through the N-terminal dimerization domain [129]. Overlap between the interactome data and the shortlist of 133 host degraded targets from Nightingale et al (2018) identified UL56 and the non-canonical ORFL147C as HCIPs for the degraded isoform C of Sp-100, indicating that the seven isoforms of this nuclear autoantigen may be targeted by different viral proteins.

Despite the inclusion of 14-3-3 epsilon, theta, zeta/delta and beta/alpha in HCMV virions, the latter is the only protein from this family in the shortlist of host factors degraded in all three screens.

Table 5.5 | Viral interactors for degraded host proteins identified in Nightingale et al, 2018 [163]

Uniprot	Gene Symbol	Description	Interactome
P52630	STAT2	Signal transducer and activator of transcription 2	UL123
Q96J02	ITCH	E3 ubiquitin-protein ligase Itchy homolog	UL42, UL55
Q96PU5	NEDD4L	E3 ubiquitin-protein ligase NEDD4-like	US30, UL20, UL42, UL55, UL133
P78357	CNTNAP1	Contactin-associated protein 1	US9, US21, UL55
P22694-2	PRKACB	Isoform 2 of cAMP-dependent protein kinase catalytic subunit beta	US9, US19, UL150A, ORFL147C
P55899	FCGRT	IgG receptor FcRn large subunit p51	US7, UL20
P11908-2	PRPS2	Isoform 2 of Ribose-phosphate pyrophosphokinase 2	US12
O75663	TIPRL	TIP41-like protein	US12
P61244	MAX	MYC associated factor X	UL94
Q15003	NCAPH	Condensin complex subunit 2	UL9, UL19, UL132
P56945-6	BCAR1	Isoform 6 of Breast cancer anti-estrogen resistance protein 1	UL86, UL150A, ORFL147C
Q96JK2	DCAF5	DDB1- and CUL4-associated factor 5	UL80.5
P23497-4	SP100	Isoform Sp100-C of Nuclear autoantigen Sp-100	UL56, ORFL147C
Q99685	MGLL	Monoglyceride lipase	UL50, ORFL147C
P31946	YWHAB	14-3-3 protein beta/alpha	UL43
P17302	GJA1	Gap junction alpha-1 protein	UL42, UL100, UL148A, UL148C, US21
Q86Y39	NDUFA11	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 11	UL40
O14730	RIOK3	Serine/threonine-protein kinase RIO3	UL40
P16333	NCK1	Non-catalytic region of protein tyrosine kinase 1	UL25
Q8IVD9	NUDCD3	NudC domain-containing protein 3	UL23
P14324	FDPS	Farnesyl pyrophosphate synthase	UL16
Q9NRY4	ARHGAP35	Rho GTPase-activating protein 35	UL150A
P01008	SERPINC1	Antithrombin-III	UL14, UL78, UL130
P17612	PRKACA	cAMP-dependent protein kinase catalytic subunit alpha	UL14
Q9Y6D5	ARFGEF2	Brefeldin A-inhibited guanine nucleotide-exchange protein 2	UL132, UL75
P04049-2	RAF1	Isoform 2 of RAF proto-oncogene serine/threonine-protein kinase	UL132
Q13671	RIN1	Ras and Rab interactor 1	UL132
Q13907-2	IDI1	Isoform 2 of Isopentenyl-diphosphate Delta-isomerase 1	UL13, UL27
P61088	UBE2N	Ubiquitin-conjugating enzyme E2 N	RL6, UL56
Q9UKI2	CDC42EP3	Cdc42 effector protein 3	ORFS343C
A1A4S6	ARHGAP10	Rho GTPase-activating protein 10	IRS1

In addition to ITCH, UL42 interacted with Neural Precursor Cell Expressed, Developmentally Down-Regulated 4 (NEDD4)-family E3 ligases NEDD4 (both isoforms 1 and 4) and NEDD4-like (NEDD4L), as well as the HECT, C2 and WW Domain Containing E3 Ubiquitin Protein Ligases HECW1 and 2 (see Figure 5.5.1A). Both NEDD4 and NEDD4L have been shown to be degraded during early HCMV infection (see Figure 5.5.1B) [163].

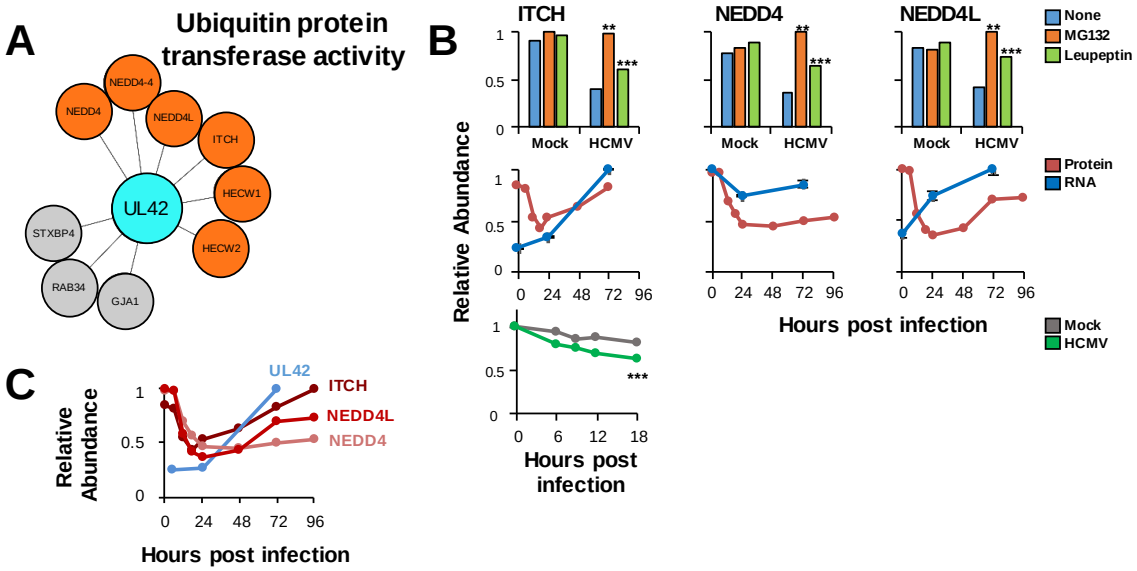


Figure 5.5.1 | UL42 as a hub of E3 destruction
(A) High-confidence cellular interactors of UL42. Interactors in orange circles exhibited ubiquitin protein transferase activity. **(B)** ITCH, NEDD4 and NEDD4L are degraded during early HCMV infection (data from Nightingale et al, 2018). Protein degradation was measured using three orthogonal tandem mass tag (TMT)-based proteomic screens. The first measured protein abundance throughout early infection in the presence or absence of inhibitors of the proteasome or lysosome. The second compared transcript and protein abundance over time to distinguish between degraded and transcriptionally regulated proteins. The third employed an unbiased global pulse-chase to compare the rates of protein degradation during HCMV infection against mock infection (NEDD4 and NEDD4L were not quantified in this latter screen). Benjamini-Hochberg adjusted Significance A values were used to estimate p-values in the top panels; **p<0.005, ***p<0.0005. Mean and SEM are shown for transcript quantitation (n= 3) in the middle panels. A p-value for the difference between rates of degradation is shown in the bottom panel; ***p<0.0005. **(C)** UL42 transcript is expressed contemporaneously with NEDD4 and NEDD4L degradation. Protein profiles from Figure 5.5.1B (red colour, data from Nightingale et al, 2018) are overlaid with a UL42 transcript profile (blue colour, data from Stern-Ginossar et al, 2012).

UL42 protein was not detected in the previous proteomic studies by Weekes et al (2014), Fielding et al (2017) or Nightingale et al (2018) [162, 163, 194]. UL42 transcript was quantified by Stern-Ginossar et al (2012) [67], with a peak of expression 72 h post-infection, yet still detectable at earlier time points. This suggested that UL42 protein may be expressed coincidentally with degradation of NEDD4 and NEDD4L (see Figure 5.5.1C).

The interactions between UL42 and NEDD4/NEDD4L were validated by co-IP in transiently transfected HEK293T cells using C- and N-terminally tagged UL42 (see Figures 5.5.2A and 5.5.3). Furthermore, stable overexpression of UL42 in HFFF-TERT cells was shown to be sufficient for degradation of NEDD4 (see Figure 5.5.2B).

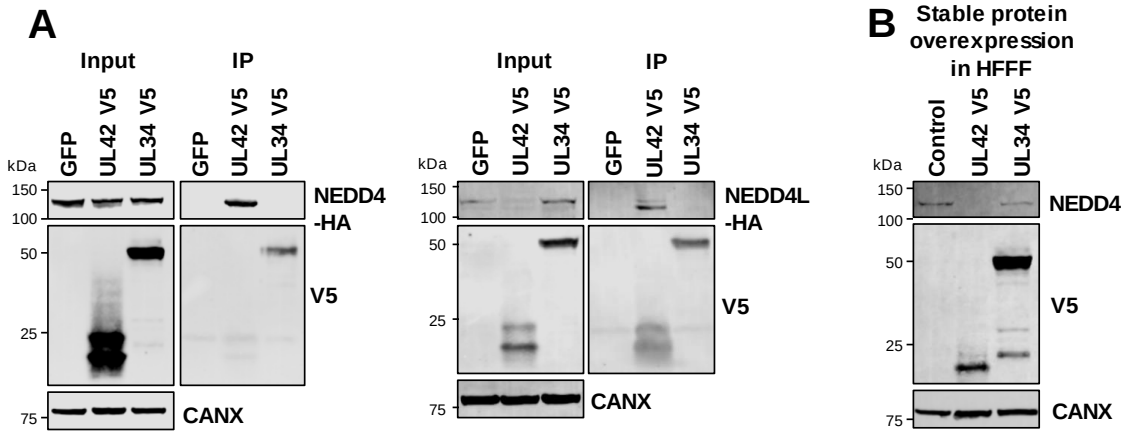


Figure 5.5.2 | Validation of the interaction between UL42 and NEDD4/NEDD4L
(A) Validation of interaction between UL42 and NEDD4/NEDD4L by co-IP. HEK293T cells were transiently transfected with the indicated plasmids, one expressing the C-terminally V5-tagged viral protein and the other expressing C-terminally HA-tagged NEDD4 or NEDD4L. These proteins were detected with anti-V5 and anti-HA. This figure is representative of n= 2 experiments (NEDD4); n= 1 experiment (NEDD4L). Expected sizes: NEDD4: 104-149 kDa; NEDD4L: 96-111 kDa; UL42: 14 kDa; UL34: 45 kDa; CANX: 72 kDa. **(B)** UL42 was sufficient to degrade NEDD4. HFFF-TERTs expressing UL42 or controls were lysed and immunoblotted as indicated. Anti-NEDD4 was used to detect endogenous NEDD4. This figure is representative of n= 1 experiment. Expected sizes: NEDD4: 104-149 kDa; UL42: 14 kDa; UL34: 45 kDa; CANX: 72 kDa.

The route of degradation of these interactors of UL42 requires further characterization. As depicted in Figure 5.5.1B, MG132 and leupeptin both inhibited degradation of ITCH, NEDD4 and NEDD4L, although this may be explained by the known effects of MG132 on lysosomal cathepsins in addition to the proteasome [287], or conversely by effects of leupeptin on certain proteasomal proteases in addition to lysosomal proteases.

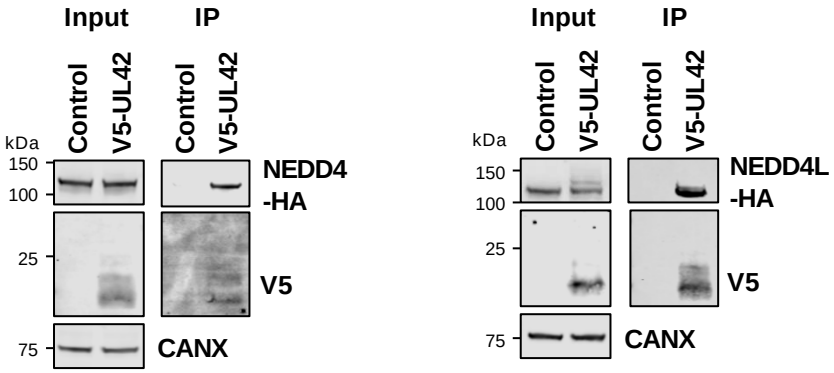


Figure 5.5.3 | Further validation of the interaction between UL42 and NEDD4/NEDD4L Validation of interaction between UL42 and NEDD4 (left panel) and NEDD4L (right panel) by co-IP, conducted as described in Figure 5.5.1. HEK293T cells were transiently transfected with the indicated plasmids, one expressing N-terminally V5-tagged UL42 and the other expressing C-terminally HA-tagged NEDD4 or NEDD4L. These proteins were detected with anti-V5 and anti-HA. This figure is representative of n= 1 experiments. Expected sizes: NEDD4: 104-149 kDa; NEDD4L: 96-111 kDa; UL42: 14 kDa; CANX: 72 kDa.

The sensitivity of the HCMV interactome for detecting interactions with weakly-expressed prey was assessed by examining the interaction between the cell surface adhesion molecule Leucine Rich Repeat And Fibronectin Type III Domain Containing 3 (LRFN3) and US10.

LRFN3 was previously quantified by 1 peptide in samples enriched for PM proteins only [162, 163], being rapidly downregulated from the PM and accompanied by upregulation of transcript over the same period. This suggested that LRFN3 is either degraded or retained within the infected cell (see Figure 5.5.4 A). The ER-resident transmembrane glycoprotein US10 was the only bait interacting with LRFN3. Using the

same methodology as for interactions validated in this chapter, the interaction between US10 and LRFN3 was validated by co-IP (see Figures 5.5.4 B-C). This suggested that US10 may regulate LRFN3 in a manner similarly to the reported degradation for HLA-G [288].

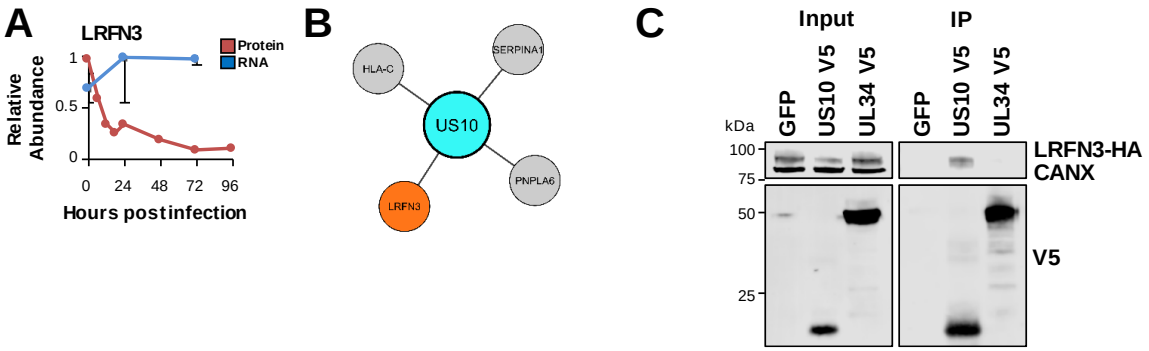


Figure 5.5.4 | US10 interacts with HCMV degradation target LRFN3
(A) LRFN3 was rapidly downregulated from the PM during HCMV infection, in the presence of upregulated transcript (mean and SEM are shown for transcript quantitation (n= 3); data are from (Nightingale et al., 2018)) **(B)** HCIPs of US10, including LRFN3. **(C)** Validation of the interaction between US10 and LRFN3 by co-IP, conducted as described in Figure 5.4.2. Prey were detected using anti-HA. This figure is representative of n= 2 experiments. Expected sizes: LRFN3: 66 kDa; US10: 21 kDa; UL34: 45 kDa; CANX: 72 kDa.

5.6 Protein domain associations inferred from interaction data

A protein domain can perform similar functions within different proteins, often via interactions with complementary protein domains or molecular structures. The function and interaction(s) of these domains can be predicted by analysing interactions between their parent proteins [189, 289]. Domains that are necessary for an interaction between two proteins tend to co-occur at higher frequency, as depicted in Figure 5.6.1. Certain protein domains co-occur frequently but do not necessarily interact directly, however these associations can still provide insights into their function.

The Pfam database is a repository where proteins have been grouped into families according to characteristic domains using multiple sequence alignments and hidden Markov models. This database has 168 entries for HCMV strain Merlin, comprised of 96 protein domains annotated for 115 viral proteins. By mapping Pfam domains to every

bait and prey protein in the interactome, it was possible to identify domain pairs that interact with unusual frequency (see Figure 5.6.2).

Figure B removed for copyright reasons. Copyright holder is *Elsevier Inc.*

Figure 5.6.1 | Schematic representation of domain association analysis
The association of protein domains 'a', 'b' and 'c' with domains '1', '2', '3', '4' and '5' is mapped onto an interaction matrix to determine unusual frequency of interaction between two domains (i.e. domain 'a' and domain '1' interact more frequently with each other than with other domains).

Significant associations for domains shared by at least two viral proteins were found for seven viral protein domains (see Table 5.6). A full list of significant domain associations for each viral protein, including prey proteins and interacting domains, is provided in Appendix E.

Table 5.6 | Annotated PFAM domains in canonical HCMV proteins

Domain name	Number of proteins	HCMV protein members	PFAM accession
US22	13	IRS1, TRS1, US22, US23, US24, US26, UL23, UL24, UL26, UL29, UL36, UL38, UL43	PF02393
Bax1-I	5	US12, US15, US17, US20, US21	PF01027
CMV_US	4	US7, US8, US9, US11	PF08001
7tm_1	3	US27, US28, UL33	PF00001
RL11D	2	RL5A, RL6	PF11088
Herpes_IE2_3	2	UL117, UL122	PF03361
UL141	2	UL141, UL14	PF16758
Herpes_pp85	2	UL25, UL35	PF04637
Herpes_UL82_83	2	UL82, UL83	PF05784
Cytomega_US3	2	US2, US3	PF05963

As depicted in Figure 5.6.2, this analysis correctly predicted that HCMV glycoprotein UL141 interacts with TNFR cysteine-rich domains (TNFR c6), which has been demonstrated for TNFRSF10B and predicted for TNFRSF10A [251]. UL141 also interacted with TNFRSF10D as reported in Smith et al (2013) and was found to interact with TNFRSF1A suggesting that these interactions may also occur via the TNFR c6 domain [290].

Domain analysis also predicted an interaction between Herpes pp85 proteins and SH3 domains (see Figure 5.6.2). HCIP data suggested that the phosphoprotein pp85 (UL25) interacted with SH3 domain-containing proteins NCK1 (Non-catalytic region of protein tyrosine kinase 1) and NCK2. UL25 also interacted with two other human proteins, WDR26 and RPS6KA3, in addition to the viral tegument protein UL26. The tegument protein UL26 had more diverse targets, including proteins with E3 ligase activity, ribosomal S6 kinases, COPII vesicle coat proteins, members of the CPSF and CTLH complexes, and NCK2 but not NCK1 (see Figure 5.6.3).

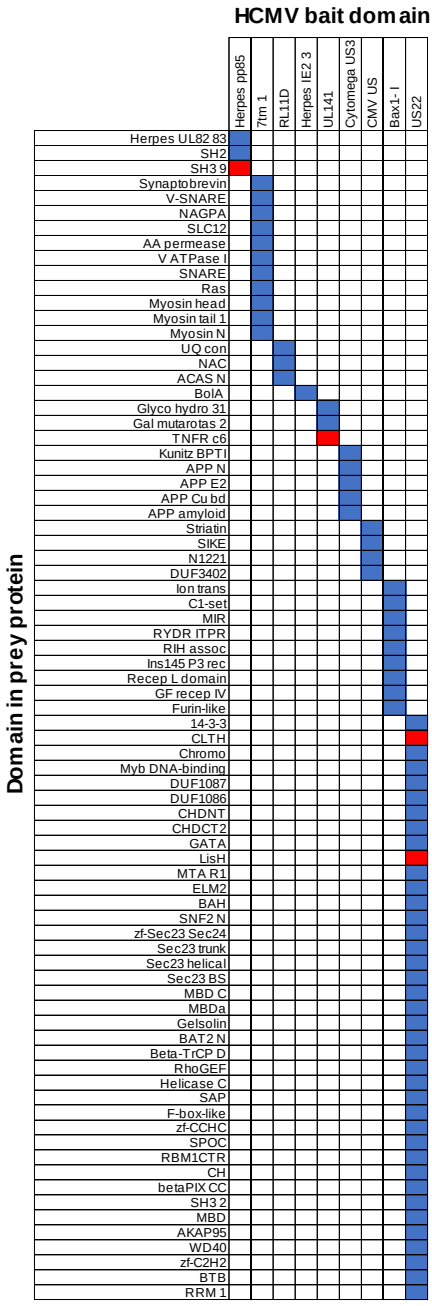


Figure 5.6.2 | PFAM protein domain association analysis

Heatmap depicting significant associations between domains present in HCMV baits (top) and human or viral prey (side). Pfam domains were mapped onto every bait and prey protein in the interactome [289]. The numbers of interactions emanating from proteins containing each domain were tallied individually, along with the numbers of interactions linking each observed domain pair. Contingency tables were then populated to relate domain associations. For each pair, Fisher's exact test determined the likelihood of a non-random association. p-values were adjusted for multiple hypothesis testing [218]. Coloured boxes identify domain pairs that associate at a 1 % false discovery rate. Red boxes indicate domain pairs from this analysis discussed in the text. Domain associations are only shown for domains occurring in at least two viral proteins. Appendix E shows the full underlying data.

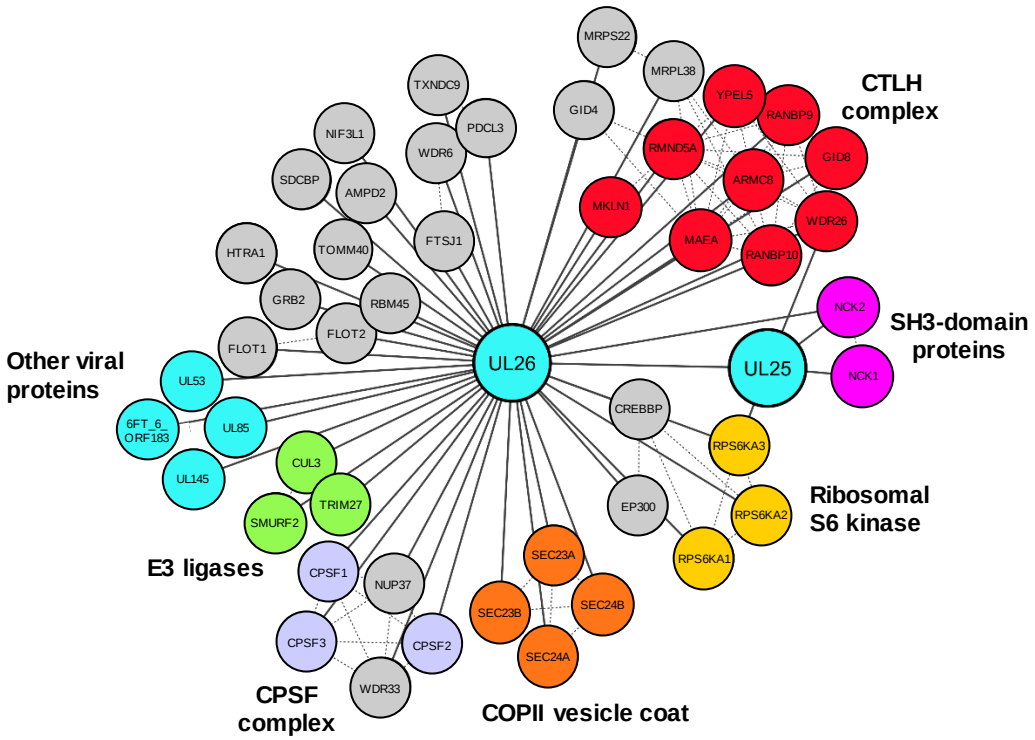


Figure 5.6.3 | HCIPs of UL25 and UL26
DAVID analysis identified that members of the C-terminal to LisH (CTLH) complex and COPII vesicle coat proteins were enriched among UL26 HCIPs (see Figure 5.4.6). Domain association analysis suggested that interaction of UL26 with CTLH components may occur via interaction of the viral US22 domain with either cellular CLTH or LisH domains. Dashed lines represent human-human interactions derived either from Bioplex 2.0 as described in Figure 5.4.1 or from curated or experimental data in the STRING database. CPSF - Cleavage and polyadenylation specificity factor.

SH3 domains have been shown to interact with proline-rich regions [291]. Given that UL25 has a proline-rich C-terminus, and NCK1 has three N-terminal SH3 regions, Dr. Katie Nightingale designed a series of mutations on each SH3 domain of NCK1 and cloned a C-terminally truncated form of UL25 without the proline-rich region (see Figure 5.6.4A). These mutations were based on previous functional studies of NCK1 function [292, 293]. Using transiently transfected HEK293T cells with plasmids expressing the NCK1-SH3 mutants, in combination with full-length or truncated UL25, Dr. Katie Nightingale performed a series of co-IPs which suggested that the UL25 C-terminus interacts with the first NCK1 SH3 domain alone, validating and extending the prediction from domain association analysis (see Figure 5.6.4B).

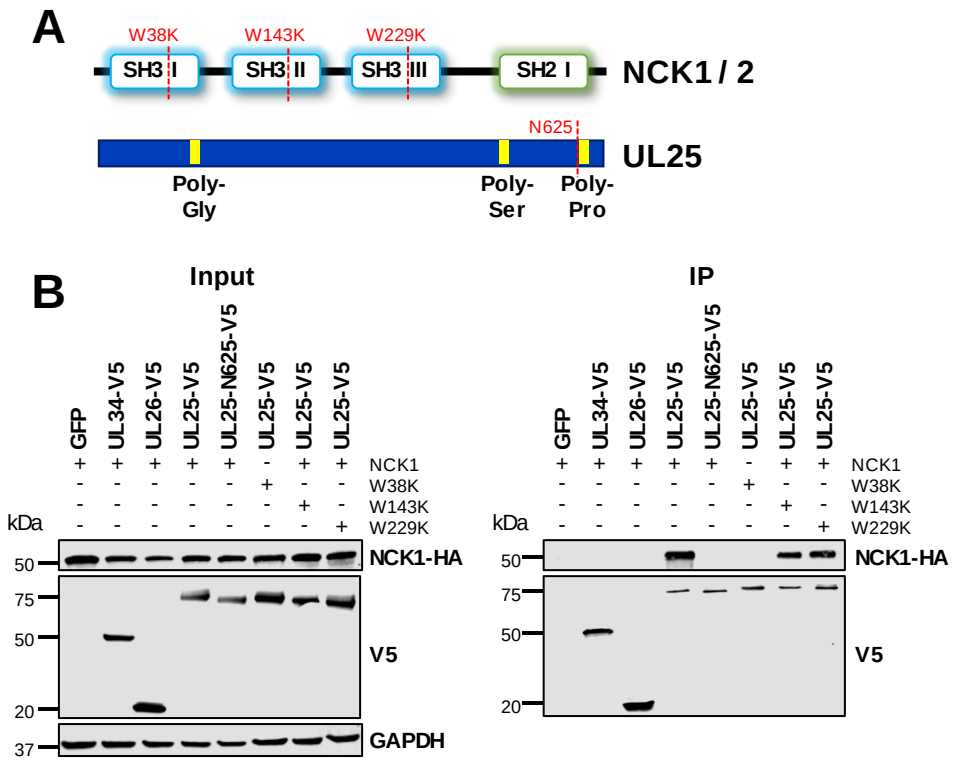


Figure 5.6.4 | Validation of the interaction between UL25 and NCK1
(A) Schematic of NCK1 and UL25 protein structures, indicating the position of point mutations or truncation used in (B). NCK1 point mutation design was based on previous studies on the function of this adaptor protein [292, 293]. **(B)** Co-IP demonstrating that the UL25 proline-rich C-terminal domain associates with the first NCK1 SH3 domain, conducted as described in Figure 3. HEK293T cells were transiently transfected with the indicated plasmids, one expressing the C-terminally V5-tagged viral protein and the other expressing C-terminally HA-tagged NCK1. These proteins were detected with anti-V5 and anti-HA. Mutations or truncations of each gene are indicated in the figure. GAPDH – loading control. This figure is representative of n= 3 experiments. Expected sizes: NCK1: 43 kDa; UL25: 74 kDa; UL26: 21 kDa; GAPDH: 36 kDa.

5.7 Discussion

Instrument performance controls as well as monitoring sample summary statistics were necessary to avoid samples being run in suboptimal conditions which would impair overall protein quantification. Additionally, disparities in quantification between replicates would undermine CompPass filtering, given that parameters (for example 'entropy') looked for inconsistently detected prey.

The ‘no bait’ control had no V5 tag but was still able to co-purify over a thousand proteins. A single replicate of this control was included in the NP-40 set. Technical replicates of biological duplicates of cells transduced with the empty vector control, as well as biological duplicates of cells transduced with a vector encoding GFP were included in the digitonin set in order to increase the sample number for CompPass filtering. This was necessary to help distinguish genuine interactors from common contaminants, given that the latter were different between lysates containing NP-40 and digitonin, and thus could not be processed together. In both datasets, these controls had the effect of increasing the number of IPs that identified non-specific interacting proteins, thus decreasing NWD and z-scores for these proteins.

Given the extensive list of proteins identified in each AP-MS dataset, a stringent filtering strategy was necessary to remove false positives. CompPass has been previously employed in large-scale protein-protein interaction studies including the human interactome project where similar filtering thresholds for each parameter were used [189, 195, 208, 209, 236].

The PSM filter employed to identify HCIPs for this interactome guaranteed that at least 3 peptides had to be identified among the two replicates in order for a protein to be considered a genuine interactor.

The entropy score used probability to measure variability of PSMs from each protein among the replicates, attributing low scores for proteins which had very disparate numbers of PSMs when comparing replicate ‘A’ to ‘B’ or vice-versa. This was particularly useful in the case of carry-over of unusually abundant proteins which would result in one replicate having a high number of PSMs, whereas the other would have a very low number (close to or indeed zero PSMs).

Another CompPass parameter, the Z-score, was calculated for each prey for each bait, and thus the same prey had a different score depending on the bait. This parameter can identify bait-prey pairs where the number of PSMs differs significantly from the mean and was most useful when analyzing proteins that were present in multiple complexes but were found at much higher levels in a subset of these (thus having high standard deviation in these cases). However, it does not differentiate an interactor with 1 PSM from another with 30 PSM, resulting in a tendency to upweight unique proteins regardless of their abundance.

The NWD score is a version of the z-score with higher discriminating power, as it takes into account the frequency of an interactor across all baits, up-weighting interactors that are rare. Additionally, previous studies have shown that by taking only the top 2 % of interactors according to NWD and Z parameters, false protein identifications should be limited to 5 %. In this case, it is possible that out of the 3440 HCIPs (excluding self-HCIP identification), approximately 172 interactors may be false positives. However, these false positives are likely to predominate amongst prey with low NWD or Z-scores; particularly high scores reflect particularly confident bait-prey interactions [189].

Interactomes of this type include false discoveries, however simultaneous analysis of AP-MS experiments decreases false discovery rate in comparison to isolated AP-MS experiments, given that non-specific interacting proteins can be identified and excluded due to their frequent identification in several IPs [195].

Interactomes may also fail to detect genuine interactions. Assessing false negative is less straightforward, and literature reporting previously identified interactions also suffers from false discoveries. A potential cause for missed identifications is the abundance of the prey protein. The HCMV interactome clearly had the ability to identify prey with low cellular abundance such as LRFN3, which was below the limit of detection

in two unbiased quantitative proteomic studies from whole cell lysates of HFFFs [193]. While 36 % of previously described interactions that were not identified in this interactome were also unquantified in whole cell lysates [193], degradation of human factors during HCMV infection may also decrease their abundance below the limit of detection by MS. Future studies aiming to validate interactions from this dataset with a prey protein with low cellular abundance may be more successful using co-IP of overexpressed prey rather than endogenous protein.

Presence and overexpression of each viral bait prior to and throughout the course of infection may have induced temporal dysregulation of the expression of other viral proteins. Thus changing the regulation of interactions that would usually commence earlier or later than 60 h of infection. Yet, given that 153/153 quantified viral ORFs were expressed at 60 h [162], the interactions observed in this interactome should occur at this phase of infection, regardless of whether either bait or prey protein (or both) were not maximally expressed. Furthermore, the abundance of certain stably expressed proteins may actually have been below the level of expression usually observed during HCMV infection. Additionally, previous human interactome studies found no correlation between bait protein expression and the number of HCIPs [195].

Alternative approaches to generate an interactome would suffer from other confounding issues. For example, introduction of a tag in the viral genome may generate a construct for AP-MS with similar abundance and temporal profile as during infection with unmodified virus. However, the transcription of overlapping viral ORFs during infection may disrupt expression of neighbouring genes [67].

The interactome did not identify HCIPs for nine HCMV canonical proteins. The only interactors identified for US2, US11 and UL93 were the baits themselves. Nevertheless, US2, US11 and UL93 were detected after filtering as prey for UL148C, UL18

and US22 respectively. The viral immunoevasins US2 and US11 are known to interact with the SEC61 translocon to redirect MHC I heavy chains from the endoplasmic reticulum to the cytosol where they are degraded by the proteasome [287, 294, 295]. Both these viral baits contain a single-pass transmembrane domain and were quantified in their own IPs. For US2, an interaction with HLA-A, HLA-C and SEC61A2 was detected in the unfiltered dataset but these were all quantified by 1 PSM and had low NWD and Z-scores. Similarly, US11 immunoprecipitated HLA-A (1 PSM), HLA-B (1 PSM), HLA-C (2 PSM) and SEC61A2 (1 PSM) but these did not meet the stringent filtering criteria. Given that certain US2 and US11 interactions occur via the transmembrane domain (e.g. interaction between US2 and TRC8, [296]; interaction between US11 and Derlin-1, [297, 298]), and are thus inherently weak, it may not be surprising that we did not detect any non-bait HCIP for these proteins, particularly given that NP40 was employed as opposed to Digitonin. A future avenue would be to determine if additional interactors were detected with Digitonin as a detergent (likewise for other single-pass TM proteins). Additionally, US2 and US11 targets are rapidly degraded (for example MHC Class I molecules are degraded with a half-time of less than one minute in cells expressing US11) [287]. Thus, repeating these IPs in cells pre-treated with inhibitors of proteasome degradation may also enable detection of host targets for US2 and US11. UL93 is a positional homolog for HSV-1 capsid vertex component CVC1 and is known to interact with the HCMV ortholog for CVC2, UL77. Both these proteins play a role in viral genome cleavage and packaging [22]. Although the soluble UL93 was enriched in its own IP, an interaction with the UL77 was not detected. Additionally, UL93 was not quantified as a prey for UL77. Interaction with other capsid proteins as well as proteins that have been reported to be present in the virion were detected but had an overall low score for most parameters.

UL96 has been reported to interact with the UL32 gene product pp150 [299]. Both these bait and prey were quantified in the UL96 AP-MS samples but the interaction scored low for NWD and Z-scores and was quantified by a single PSM between the two replicates. In contrast, UL96 was not detected in the UL32 IPs.

UL119 encodes for a viral Fc-gamma receptor-like protein with a single-transmembrane domain. This bait was enriched in its own IP and even though an interaction with IGHG1 (Immunoglobulin heavy constant gamma 1) was detected, this prey was quantified by 1 PSM between both replicates and scored low for NWD and Z-scores, and therefore did not pass the final stringent filtering criteria employed. This example illustrates that the unfiltered data may nevertheless be useful in testing hypotheses – since it is likely to include false negative interactions. Identification of such interactions could potentially be made by lowering the PSM criteria, or adjusting this criteria according to the overall estimated cellular abundance of each host protein, which has been estimated in Nobre et al (2019) [193].

UL146 encodes for a viral CXCL1 chemokine. As UL146 is a secreted protein, this may explain why this bait was poorly detected in its own IP. UL146 has been reported to interact with CXCR1 and CXCR2 receptors [300, 301], but these baits were not detected in the dataset in any IP.

The membrane protein RL13 was also enriched in its own IP. It has been shown to modulate viral DNA replication by interacting with Nudix Hydrolase 14 (NUDT14) [302], however this prey was not detected in the interactome. Additionally, an interaction with Fc region of Immunoglobulin G has been reported [43], and even though IGHG1 was detected as prey for RL13, it scored low for entropy, NWD and Z-score.

Despite previous validation of transgene expression, neither UL120 nor UL142 baits were quantified in their respective IPs. UL120 was also not quantified in the

proteomics dataset from Weekes et al (2014), indicating that this protein may have low abundance. In contrast, UL142 has previously been quantified using PM profiling, although this method relies on enrichment of sialylated PM glycoproteins and employs a lysis buffer with different composition (e.g contains Triton instead of NP-40) [162]. There are no studies showing binding partners for the single-pass membrane protein UL120, but ULBP3 [303] and SNAPIN [304] have previously been identified as interactors for the glycoprotein UL142. For both UL120 and UL142 the only prey to meet the entropy, NWD and Z-score filtering criteria were quantified by 1 PSM only.

Taken altogether, the stringency of the filtering criteria may partly explain the lack of identification of HCIP for these nine baits. Protein abundance and interaction strength are additional factors. Weakening the filtering criteria would need to be undertaken with caution due to the increase in false positive identifications, however it could be undertaken with specific testable hypotheses in mind.

Overlap with published data annotated in interaction databases was only 35 %. For 25 % of the annotated interactions in the combined positive control database (167 listed interactions, see Appendix D), the prey was not detected in the AP-MS data for the expected bait. There are a number of possible reasons for these observations. These include: differences in experimental conditions employed; differences in cell type; whether or not HCMV infection or simply single-gene overexpression alone was employed; differences in the protein tag and whether this was attached to the N- or C-terminus. Even though a 5 % cut-off for NWD and Z-score was applied to include previously published interactions that didn't meet the more stringent threshold of 2 %, approximately 40 % of the interactions in the combined dataset still failed to meet the filtering criteria. This was partly caused by inconsistent detection in the interactome (quantification by 1 PSM or entropy score of 0) but most predominantly due to low Z and

NWD scores, a consequence of a high frequency of detection of these interactors across all baits, a low standard deviation of PSMs or a combination of these two. Nevertheless, lack of detection of a given 'positive control' does not mean that the interactome has failed to identify a correct interaction. There is no 'gold standard' for correct interactions; and many of the published interactions for HCMV are based on less-well controlled AP-MS studies, yeast two-hydrid, in vitro binding assays with purified protein or immunofluorescence. For example, from the list of known interactors of UL27 (see Table 4.2 or Appendix D), only 4/27 were identified as HCIP for this bait in this interactome. Three of these published interactors were not quantified in the dataset, which could reflect a difference in protein abundance in the cell types used in both experiments. For the remaining 20/27 published interactors of UL27, the entropy scores were close to 1, yet the NWD scores were below 0.2, suggesting that although these proteins may be consistently detected in AP-MS experiments they are detected to a similar degree in several other IPs and thus may not be specific interactors of UL27. Furthermore, a multitude of published interactions has not yet been curated into a database, limiting the extent of the evaluation of the HCMV interactome data.

DAVID analysis of all HCIPs revealed enrichment in biological terms which had been reported previously in the literature, such as 14-3-3 proteins which have been identified in HCMV virions; the association of UL29 and UL38 with the NuRD complex; as well as the interaction of several RNAII subunits with a complex of viral proteins required for the transcription of late genes. This analysis has also highlighted biological terms that have not been studied in the context of HCMV infection such as the CNOT and NCoR complex. As seen for UL72, UL43 and US22, which have an enrichment for interacting proteins with related functions, DAVID analysis can provide an unbiased approach to generate hypothesis for determining individual gene function.

Combination of HCIP data with the screens of protein degradation during early HCMV infection from Nightingale et al (2018) showed that UL42 targets multiple ubiquitin E3 ligases for degradation, and predicted novel interactions between viral baits and 29 other degraded host factors. In fact, the interactome suggested that multiple viral proteins interact with members of the ubiquitin conjugation pathway, with 51 viral proteins interacting with one at least one E3 ligase (defined in Medvar et al., 2016) [305]. These interactions may identify previously uncharacterised viral effectors of cellular protein degradation. For example, UL25 interacted with the adaptor protein WD Repeat Domain 26 (WDR26), which can recruit substrates to the Cullin-4 RING ubiquitin ligase family [306]. UL25 also interacted with UL26, which itself had HCIPS with E3 ligase activity, such as the members of CTLH complex [307, 308]. Additionally, UL26 also interacted with other ligases and scaffolds, such as Cullin 3 and SMAD Specific E3 Ubiquitin Protein Ligase 2 (SMURF2). Future work may identify whether UL25 or UL26 prey are degraded and via which of these degradation mechanisms.

The HCMV interactome also highlighted UL20 as a hub of degradation. This viral protein was previously reported to be rapidly degraded, with the suggestion that it may target unidentified cellular proteins to lysosomes [309]. Several HCIPs of UL20 have previously been shown to be rescued from degradation by application of the lysosomal protease inhibitor Leupeptin, such as Interleukin 6 Signal Transducer (IL6ST), the neonatal Fc receptor (FCGRT), Ephrin A2 (EPHA2), and Interferon Gamma Receptor 1 (IFNGR1) [163]. These four HCIPs were shown to be rescued upon deletion of members of the viral US12-US21 family [194], suggesting that there may be cooperativity between these proteins and UL20, potentially with UL20 acting as a common final mediator of degradation.

DAVID analysis also highlighted several viral baits that interact with proteins involved in ubiquitin mediated degradation. While for UL42 and US10 it was possible to detect interactions with degraded prey, the interactome did not identify host targets for baits such as RL1 or UL145. For these baits, an interaction with SLFN11 and HLTF was expected, respectively [163, 310]. However, for HLTF at least, prey overexpression was required to observe this interaction, which might have been due to low prey abundance, weak bait-prey interaction or both. It will therefore be important to repeat this interactome with cells pre-treated with lysosomal or proteasomal inhibitors prior to AP-MS in order to prevent the degradation and increase the abundance of such targets, which could enable their identification. In fact, unpublished work by Fletcher-Etherington et al (in preparation), has employed this approach to identify an interaction between UL36 (for which DAVID analysis determined an enrichment for Ubiquitin protein transferase activity) and the necroptosis mediator MLKL [311].

Higher frequency of co-occurrence between two domains does not necessarily mean that these interact or that two proteins containing each of those domains bind to each other. For example, the interaction between UL35 and UL82 may not be due to a binding between the respective PFAM domains 'Herpes pp65' and 'Herpes UL82 UL83', given that no interaction is observed between UL35 and UL83.

In cases where an interaction is indeed observed between proteins containing domains that co-occur with high frequency, PFAM domain association analysis can provide insight into which domains may be necessary for this interaction. The exact residues required for binding can be determined using single-residue mutants or truncations for each binding partner. For example, Dr Katie Nightingale validated the interaction between the viral tegument pp85 phosphoprotein UL25 and the SH3 domain-containing protein NCK1, and using the domain association predictions determined that

both the C-terminus of UL25 and the first SH3 domain of NCK1 are necessary for this interaction. Given the multifunctional roles of NCK1 in signal transduction, including receptor tyrosine kinases, cytoplasmic remodelling via regulation of actin polymerization, apoptosis and the DNA damage response, the interaction with UL25 may fulfil a variety of functions [292, 293, 312]. One hypothesis includes inhibition of the immune synapse formation via regulation of actin polymerisation, similarly to UL135 which has been reported to dispel association between F-actin filaments in target cells and the immune synapse [250].

6 | Characterisation of ORFL147C

Functional enrichment analysis of ORFL147C HCIPs was performed by Dr. Michael Weekes (Department of Medicine, University of Cambridge). Generation of HCMV Merlin recombinants used for viral growth experiments was performed by Dr. Richard Stanton (School of Medicine, Cardiff University). Validation of the interactions between ORFL147C and MBNL1/CELF1 was performed by Dr. Katie Nightingale (Department of Medicine, University of Cambridge). Parts of the work presented in this chapter have been published in a similar form in eLife (Nobre et al., 2019).

The 604 HCMV ORFs identified by ribosome profiling remain uncharacterised and thus it is unclear whether they play a functional role in HCMV infection [67]. As depicted in Figure 1.4, the two uncharacterised ORFs included in the HCMV interactome showed a relative abundance within the range of canonical HCMV proteins. ORFL147C was the most abundant non-canonical ORF and was quantified at approximately 25x lower copy number than the most abundant viral protein UL83. ORFS343C was observed in the lower end of the abundance range, yet it was still quantified at approximately 3x higher copy number than the least abundant viral protein US18.

HCIPs have now been identified for the polypeptides encoded by these two non-canonical ORFs. This chapter will focus on a preliminary characterisation of the protein encoded by the ORFL147C ORF.

6.1 Sequence analysis and conservation

The coding sequence of ORFL147C initiates upstream to the 5' end (in a different reading frame) of UL56, which is a canonical gene encoding a subunit of the viral terminase, and approximately 0.5 kbp from the 3' end of the canonical gene encoding the single-stranded DNA binding protein UL57 (see Figure 6.1).

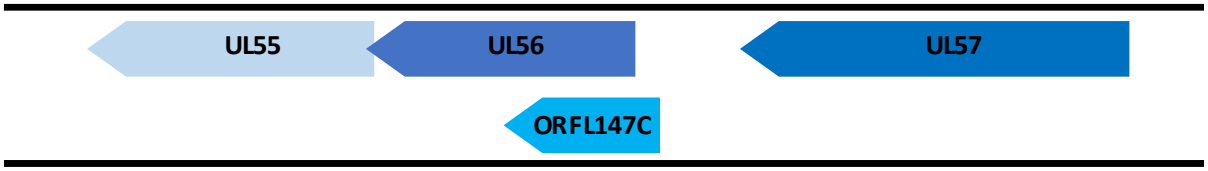


Figure 6.1.1 | ORFL147C coding sequence and relation to neighbouring viral genes
The coding sequence for the uncharacterised ORFL147C starts 58bp upstream of the start codon for UL56. Coding sequences for both these proteins are located on the same strand but not in frame, thus ORFL147C is not an N-terminal extension of UL56.

This region of the HCMV genome contains genes conserved across Alpha, Beta- and Gammaherpesvirinae (thus designated core genes), including UL56 and UL57 [69]. A basic local alignment search of the nucleotide and protein sequences for the uncharacterised ORFL147C against HSV-1, HSV-2, VZV, EBV, HHV-6, HHV-7 and KSHV entries found no significant similarity, thus this ORF is not conserved among other genus of human herpesvirus.

In contrast, alignment of ORFL147C protein sequences for HCMV strains Merlin, Toledo, AD169 and Towne showed that not only do these four strains contain the coding sequence for this ORF, they also share 98 % residue identity, with the high-passage strains being slightly more dissimilar (see Figure 6.1.2).

A basic local alignment search of the protein sequence against the 'Non-redundant protein sequences (nr)' database generated only three significant alignments with 65.4 – 66.6% identity (see Figure 6.1.3): the cy89 protein from *Cynomolgus macaque* cytomegalovirus (CyCMV) strain Ottawa, Cy89 protein from CyCMV strain Mauritius and

Rh91.1 from Rhesus cytomegalovirus (RhCMV). These three homologs of ORFL147C are similarly positioned between the homologs of UL56 (cyUL56 or Rh91/RhUL56) and UL57 (cyUL57 or Rh92/RhUL57) [313, 314], but no function has been assigned to either of the homologs of this ORF. ORFL147C thus appears to be conserved in human cytomegalovirus strains and have homologs in other primate CMVs, but not in other human herpesvirus genus.

```
Towne       MSLAGARPDDSVSYVSESSHGDEFVTETMRSVFEMQRIRHGAGVSKVLRSERVTGGVHAV    60
AD169       MSLAGARPDDSVSYVSESSHGDEFVTETMRSVFEMQRIRHGAGVSKVLRSERVTGGVHAV    60
Merlin      MSLAGARPDDSVSYVSESSHGDEFVTETMRSVFEMQRIRHGAGVSKVLRSERVTGGVHAV    60
Toledo      MSLAGARPDDSVSYVSESSHGDEFVTETMRSVFEMQRIRHGAGVSKVLRSERVTGGVHAV    60
            ************************************************************

Towne       QEKRGGYSVSVPEDLPGGGGAESYAEFAADALSGDAAEGAARGYGFAGPGADGLLAPAGP    120
AD169       QEKRGGYSVSVPEDLPGGGGAESYAEFVADALPGDAAEGAARGYGFAGSGADGLLAPAGP    120
Merlin      QEKRGGYSVSVPEDLPGGGGAESYAEFAADSLSGDAAEGAARGYGFAGSGADGLLAPAGP    120
Toledo      QEKRGGYSVSVPEDLPGGGGAESYAEFAADALSGDAAEGAARGYGFAGSGADGLLAPAGP    120
            ****************************.**:* *************** ***********

Towne       GGLLPYRPPFGSLRSPSHRGAPVYGQRPFQRQSGHTQRHRALPVQDELRVRDPGAGGRPG    180
AD169       GGLLPYRPPFGSLRSPSHRGAPVYGQRPFQRQSGHTQRHRALPVQNELRVRDPGAGGRPG    180
Merlin      GGLLPYRPPFGSLRSPSHRGAPVYGQRPFQRQSGHTQRHRALPVQNELRVRDPGAGGRPG    180
Toledo      GGLLPYRPPFGSLRSPSHRGAPVYGQRPFQRQSGHTQRHRALPVQNELRVRDPGAGGRPG    180
            ********************************************:***************

Towne       GLRAAAGASPSPMRHLAGGASGSVRRRDDMRPMLRGADHHPESGPLVEQAAAGLAVQPYS    240
AD169       GLRAVAGAPPSPMRHLAGGASGSVRRRDDMRPMLRGADHHPESGPLAEQAAAGLAVQPYS    240
Merlin      GLRAAAGAPPSPMRHLAGGASGSVRRRDDMRPMLRGADHHPESGPLAEQAAAGLAVQPYS    240
Toledo      GLRAAAGAPPSPMRHLAGGASGSVRRRDDMRPMLRGADHHPESGPLAEQAAAGLAVQPYS    240
            ****.*** ************************************.*************

Towne       GPPSVKPVRCEYPDGGAGPAGPDNAHPPLGWSPFGPQKPILFFIGLPQLHPGGGGGAEGV    300
AD169       GPPSVKPVRCEYPDGGAGPAGPDNAHPPLGWSPFGPQKPILFFIGLPQLHPGGGGGAEGV    300
Merlin      GPPSVKPVRCEYPDGGAGPAGPDNAHPPLGWSPFGPQKPILFFIGLPQLHPGGGGGAEGV    300
Toledo      GPPSVKPVRCEYPDGGAGPAGPDNAHPPLGWSPFGPQKPILFFIGLPQLHPGGGGGAEGV    300
            ************************************************************

Towne       QPVYGYTGTNIFLVGFYLLVPYLGGYRQAGGHHHPAAKCVSPAVPGAHERHQSPSVRGGR    360
AD169       QPVYGYTGTNIFLVGFYLLVPYLGGYRQAGGHHHPAAKCVSPAVPGAHERHQSPSVRGGR    360
Merlin      QPVYGYTGTNIFLVGFYLLVPYLGGYRQAGGHHHPAAKCVSPAVPGAHERHQSPSVRGGR    360
Toledo      QPVYGYTGTNIFLVGFYLLVPYLGGYRQAGGHHHPAAKCVSPAVPGAHERHQSPSVRGGR    360
            ************************************************************

Towne       GGHLRARGKGVGRGGAHVRGVGLCRPQQDHRPHHIPQHSSFRGQPGVQQAPRKQRDVHQN    420
AD169       GGHLRARGKGVGRGGAHVRGVGLCRPQQDHRPHHIPQHSSFRGQPGVQQAPRKQRDVHQN    420
Merlin      GGHLRARGKGVGRGGAHVRGVGLCRPQQDHRPHHIPQHSSFRGQPGVQQAPRKQRDVHQN    420
Toledo      GGHLRARGKGVGRGGAHVRGVGLCRPQQDHRPHHIPQHSSFRGQPGVQQAPRKQRDVHQN    420
            ************************************************************

Towne       QAYSRGDSTSAARWHGGRRPRGRGYSPAWTGGDVGDGYDFDDGQQQQQQQYSQSEE        476
AD169       QAYSRGDSTSAARWHGGRRPRGRGYSPAWTGGDVGDGYDFDDGQQQQQQQYSQSEE        476
Merlin      QAYSRGDSTSAARWHGGRRPRGRGYSPAWTGGDVGDGYDFDDGQQQQQQQYSQSEE        476
Toledo      QAYSRGDSTSAARWHGGRRPRGRGYSPAWTGGDVGDGYDFDDGQQQQQQQYSQSEE        476
            ********************************************************
```

Figure 6.1.2 | Alignment of ORFL147C amino acid sequences from different HCMV strains. Multiple sequence alignment was generated using Clustal Omega (EMBL-EBI) with default settings for protein sequences. Consensus symbols key: '*' denotes positions that have a single and fully conserved residue; ':' denotes conservation between groups of strongly similar properties; '.' denotes conservation between groups of weakly similar properties.

```
ORFL147C      MSLAGARPDDSVSYVSESSHGDEFVTETMRSVFEMQRIRHGAGVSKVLRSERVTGGVHAV  60
Rh91.1        MSLSEQITEDAVCYADEQGAGDEFVAETVRSVFQMQRIRHGTGVSKVLRSERVTGGVKSV  60
cy89          MSLSEQITEDAVCYADEQGAGDEFVAETVRSVFQMQRIRHGTGVSKVLRSERVTGGVKSV  60
Cy89          MSLSEQITEDAVCYADEQGAGDEFVAETVRSVFQMQRIRHGTGVSKVLRSERVTGGVKSV  60
              ***:    :*:*.*..*.. *****:**:****:*******:****************::*

ORFL147C      QEKRGGYSVSVPEDLPGGGGAESYAEFAADSLSGDAAEGAARGYGFAGSGADGLLAPAGP  120
Rh91.1        QEERTGYRLSVPQDLPGAGAPEPYTDFAVEPVHGDDFEGSVRGYGAARPRSQGVLQEAGP  120
cy89          QEERTGYRLSVPQDLPGAGAPEPYTDFAVESVHGDDFEGSVRGYGAARPRSQGVLQEAGP  120
Cy89          QEERTGYRLSVPQDLPGAGAPEPYTDFAVESVHGDDFEGSVRGYGAARPRSQGVLQEAGP  120
              **:* ** :***:****.*. * *::**.: : **  **:.**** *   ::*:*  ***

ORFL147C      GGLLPYRPPFGSLRSPSHRGAPVYGQRPFQRQSGHTQRHRALPVQNELRVRDPGAGGRPG  180
Rh91.1        GGVLPYSPSLRSVRSAPHGGAYVHGQRKVQRHASNAQRHRALPVQDELRIRDPVPGERAG  180
cy89          GGVLPYSPSFRSVRSAPHGGAYVHGQRKVQRHASNAQRHRALPVQDELRIRDPVPGERTG  180
Cy89          GGVLPYSPSFRSVRSAPHGGAYVHGQRKVQRHASNAQRHRALPVQDELRIRDPVPGERAG  180
              **:*** * : *:**  * ** *:*** .**::.::*********:***:***  * * *

ORFL147C      GLRAAAGAPPSPMRHLAGGASGSVRRRDDMRPMLRGADHHPESGPLAEQAAAGLAVQPYS  240
Rh91.1        GLLATPPAAASPVRSVTGSLPRGVRGGNNVRPMLRGTDHHPQPGALAEQAPAGLALQPYC  240
cy89          GLLATPPAAASPVRSVAGSLPRGVRGGNNVRPMLRGADHHPQPGALAEQAPAGLALQPYC  240
Cy89          GLLATPPAAASPVRSVAGSLPRGVRGGNNVRPMLRGADHHPQPGALAEQAPAGLALQPYC  240
              ** *:  *  **:* ::*.   .**  :::******:****: * ***** ****:***.

ORFL147C      GPPSVKPVRCEYPDGGAGPAGPDNAHPPLGWSPFGPQKPILFFIGLPQLHPGGGGGAEGV  300
Rh91.1        GPPAVEPVRCEYTNGGAGSHGADAAHPQPFWGPHGPQKPVLFFFGLPQLYPGGGGGAEGV  300
cy89          GPPAVEPVRCEYPNGGAGPHGADAAHPQPFWGPHGPQKPVLFFFGLPQLYPGGGGGAEGV  300
Cy89          GPPAVEPVRCEYPNGGAGPHGADAAHPQPFWGPHGPQKPVLFFFGLPQLYPGGGGGAEGV  300
              ***:*:****** :****  * * ***   *.*.*****:***:*****:**********

ORFL147C      QPVYGYTGTNIFLVGFYLLVPYLGGYRQAGGHHHPAAKCVSPAVPGAHERHQSPSVRGGR  360
Rh91.1        QPVYGYTGTNIFLVGFYLLVPYLGGYSQAGGDKHAAAKCLSPPVQGPHERPQQVPVRGGR  360
cy89          QPVYGYTGTNIFLVGFYLLVPYLGGYSQAGGDKHAAAKCLSSPVQGPHERPQQVPVRGGR  360
Cy89          QPVYGYTGTNIFLVGFYLLVPYLGGYSQAGGDKHAATKCLSSPVQSPHERPQQVPVRGGR  360
              ************************** ****.:* *:**:*  * . *** *.  *****

ORFL147C      GGHLRARGKGVGRGGAHVRGVGLCRPQQDHRPHHIPQHSSFRGQPGVQQAPRKQRDVHQN  420
Rh91.1        GGHFRARGKAAFSRAASFRGVGLCRPQQDHRPNYIPQHSSFRGQPGFQQTPRKQRNVYQN  420
cy89          GGHFRARGKAAFSRAASFRGVGLCRPQQDHRPNYIPQHSSFRGQPGFQQAPRKQRNVYQN  420
Cy89          GGHFRARGKAAFPRAASFRGVGLCRPQQDHRPNYIPQHSSFRGQPGFQQAPRKQRNVYQN  420
              ***:*****..   .* .**************::************.**:*****:*:**

ORFL147C      QAYSRGDSTSAARWHGGRRPRGRGYSPAWTGGDVGDGYDFDDGQQQQQQQYSQSEE       476
Rh91.1        QMPPGRDQASRARWGGG-G-RGGGRGHSWAGPPEYVQ--------------------       455
cy89          QMPPGRDQAPRARWGGGGG-RGGGRGHSWAGPPEYVQ--------------------       456
Cy89          QMPPGRDQAPRARWGGG-G-RGGGRGHSWAGPPEYVQ--------------------       455
              *     *.:  *** **   ** * . :*:*
```

Figure 6.1.3 | ORFL147C amino acid sequence alignment with homologs from other primate CMV species. Multiple sequence alignment for ORFL147C (HCMV), Rh91.1 (Macacine betaherpesvirus 3/Rhesus cytomegalovirus), cy89 (Cynomolgus macaque cytomegalovirus strain Ottawa) and Cy89 (Cynomolgus macaque cytomegalovirus strain Mauritius) generated using Clustal Omega (EMBL-EBI) with default settings for protein sequences. Consensus symbols key: '*' denotes positions that have a single and fully conserved residue; ':' denotes conservation between groups of strongly similar properties; '.' denotes conservation between groups of weakly similar properties.

6.2 ORFL147C interactors

CompPass filtering identified 80 HCIP for this bait, with no viral interactors. Additionally, ORFL147C was not identified as an HCIP for any of the other viral baits. Functional enrichment analysis of ORFL147C HCIPs predicted functions in RNA binding, mRNA splicing or transcription (see Figures 6.2.1 and 6.2.2).

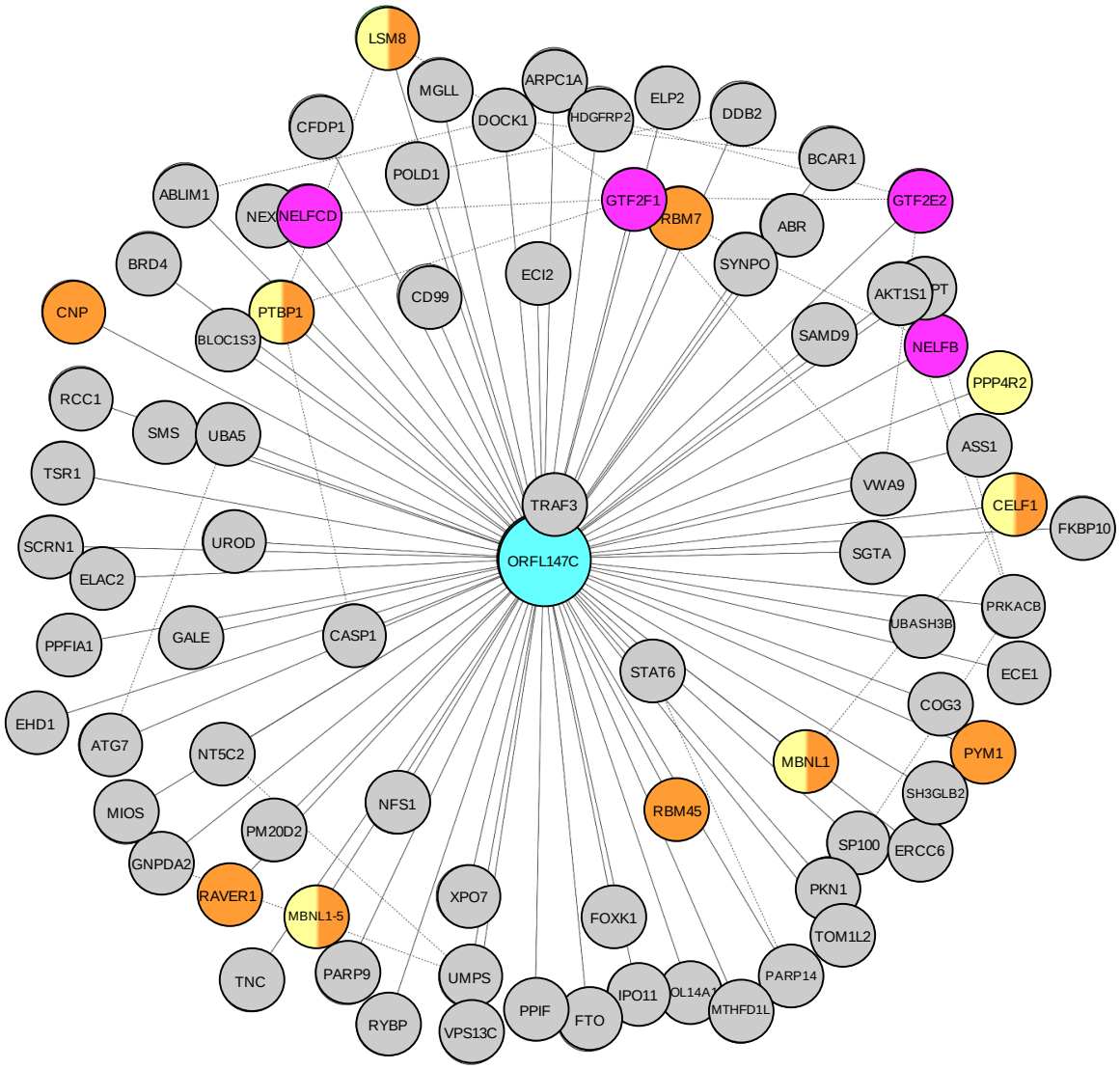


Figure 6.2.1 | ORFL147C high-confidence interacting proteins
Interactors of ORFL147C are shown in small circles, coloured according to functional enrichment (pink: RNA Pol II pre-transcription events; orange: RNA binding; yellow: mRNA splicing; dual membership of pathways is indicated by half-coloured circles). Solid lines indicate interactions identified in the HCMV interactome while dashed lines indicate interactions drawn from the human interactome (Bioplex 2.0) and subsequent unpublished data (Huttlin et al, 2017 and http://bioplex.hms.harvard.edu/downloadInteractions.php).

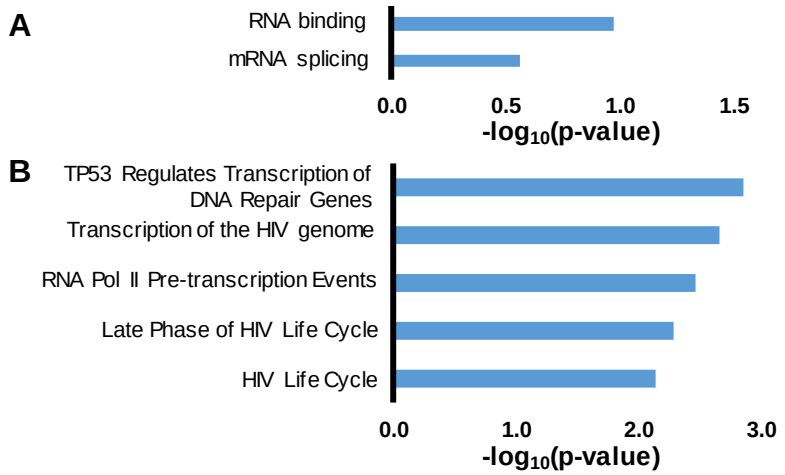


Figure 6.2.2 | Enrichment analysis of HCIPs of ORFL147C
(A) DAVID functional enrichment analysis was performed using all human proteins as a 'background'. Benjamini-Hochberg adjusted p-values are shown. **(B)** Reactome database analysis [315] showing results with a minimum of 4 entities per enriched pathway.

Furthermore, the interaction between ORFL147C and the Muscleblind-Like Splicing Regulator 1 (MBNL1) and CUG Triplet Repeat RNA-Binding Protein 1 (CELF1), two proteins with roles in mRNA splicing and RNA binding, was validated by Dr. Katie Nightingale, using co-immunoprecipitation of transiently transfected ORFL147C-V5 and MBNL1-HA or CELF1-HA [193].

Similarly to HSV-1 ICP27 and HCMV UL69, ORFL147C possesses RGG box RNA-binding motifs. HSV-1 ICP27 and its HCMV homolog UL69 bind to intronless viral RNAs and shuttle between the nucleus and cytoplasm, working as viral mRNA export factors [316, 317]. In contrast, interactome data suggested that ORFL147C does not bind any of the components of the mRNA nuclear export machinery employed by ICP27 and UL69, although it co-precipitated Exportin-7 (XPO7), a broad-spectrum bidirectional nuclear transporter [318].

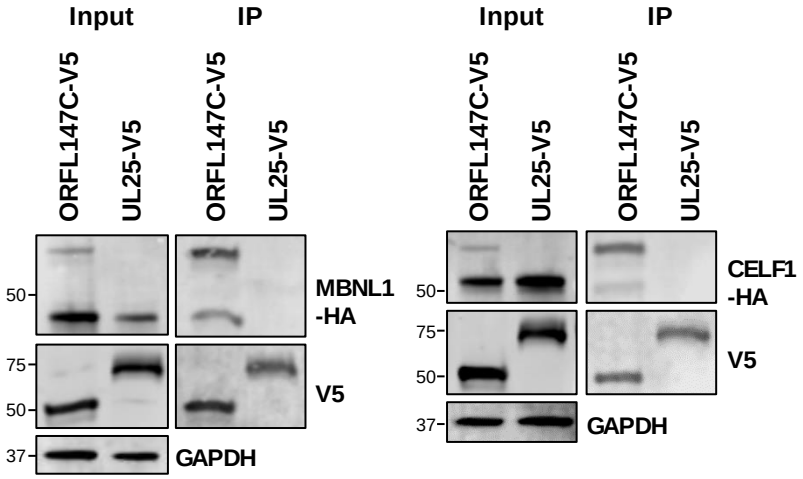


Figure 6.2.3 | Validation of interaction between ORFL147C, MBNL1 and CELF1
HEK293T cells were transiently transfected with the indicated plasmids, one expressing the C-terminally V5-tagged viral protein and the other expressing C-terminally HA-tagged MBNL1 or CELF1. These proteins were detected with anti-V5 and anti-HA. GAPDH – calnexin loading control. This figure is representative of n= 1 experiment. Expected sizes: MBNL1: 33-42 kDa; CELF1: 50-55 kDa; ORFL147C: 50 kDa; UL25: 74 kDa; GAPDH: 36 kDa.

6.3 Deletion of ORFL147C impairs viral growth

To assess whether ORFL147C plays an important role in viral replication, Dr. Rich Stanton mutated the three most N-terminal methionine residues in ORFL147C without modifying the coding sequence of UL56, generating an HCMV recombinant with ORFL147C deletion (see Figure 6.3.1).

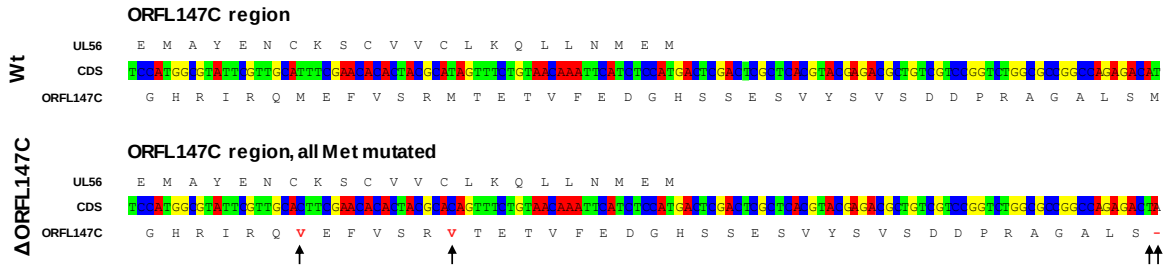


Figure 6.3.1 | Construction of a viral ORFL147C deletion mutant
The N-terminal protein sequences and coding sequences of ORFL147C and UL56 are shown, with point mutations signalled by black arrows, and residue mutations shown in red for the recombinant. The three most N-terminal methionines in ORF147C were mutated (M to Stop, or V) without affecting the amino acid sequences coded by UL56.

Absence of ORFL147C expression was verified by tandem-mass tag-base quantitative proteomics of HFFF-TERT cells which were mock-infected or infected with WT Merlin HCMV or the ORFL147C deletion mutant (see Figure 6.3.2A). Due to difficulties in growing and tittering the ORFL147C deletion virus, replicates to the samples analysed by proteomics were collected for flow cytometry and analysed for GFP expression as a measure of percentage of infection (see Figure 6.3.2B). The percentage of cells infected with WT and the ORFL147C deletion mutant was very similar (82.7 % and 82.1 % respectively), suggesting that differences in the abundance of ORFL147C protein observed in Figure 6.3.2A were not a consequence of disparities in infection.

Growth of the ORFL147C deletion virus in HFFF-TERT cells was significantly impaired compared to WT Merlin, suggesting that ORFL147C plays a role in viral replication (see Figure 6.3.2C).

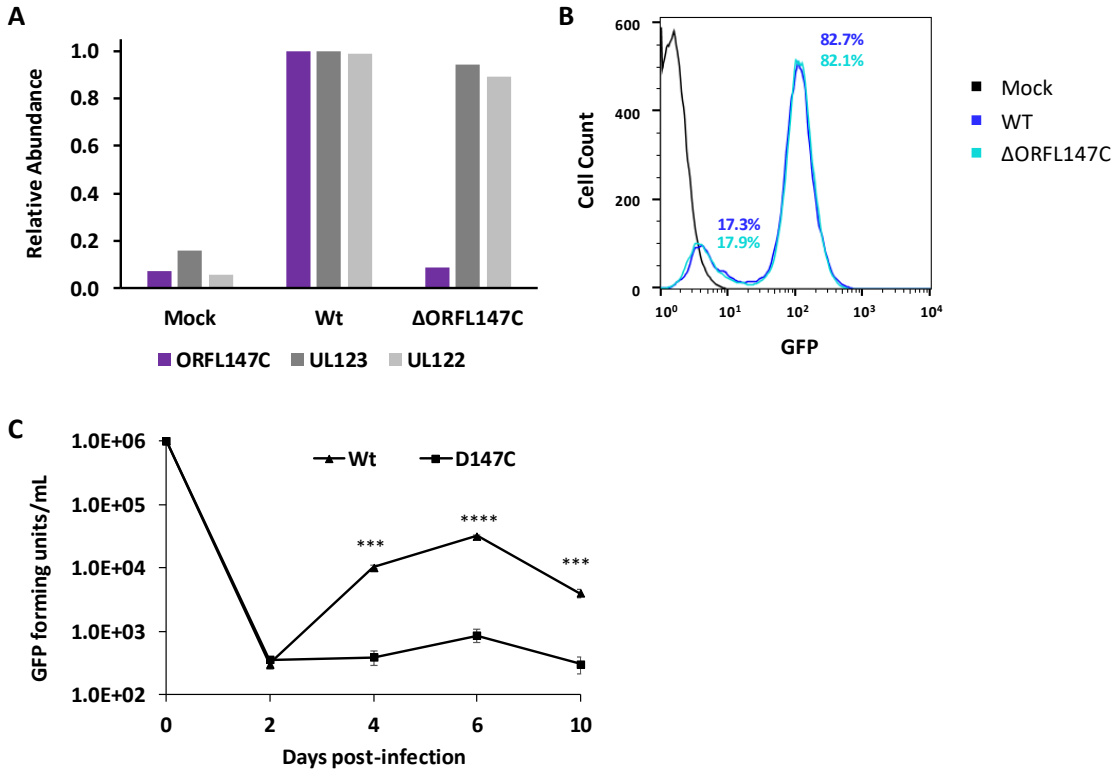


Figure 6.3.2 | Growth analysis of an ORFL147C-deficient recombinant
(A) HFFF-TERT cells were incubated with 4 µg/ml Dexamethasone in serum-free DMEM O/N, followed by mock-infected, infected with wild-type or ORFL147C-deficient recombinant (MOI=2) for 48 h. Lysates were then collected and processed for tandem mass tag-based proteomics. Both recombinant viruses contained an immediate-early gene UL36 3'-terminally fused with enhanced GFP (eGFP) gene and a self-cleaving P2A peptide. This peptide released the eGFP marker following synthesis. Additionally, insertion of the marker did not impede UL36 function in such recombinants [163]. ORFL147C relative protein abundance was similar in mock and ORFL147C-deficient recombinant infected cells. **(B)** Biological replicates of the samples analysed in A were collected for flow-cytometry analysis of GFP expression. Samples had simultaneously been infected for 48 h at MOI=2. **(C)** Cells were infected in biological duplicates at an MOI of 1, and supernatants harvested every two days. Viral supernatants were then titred in technical duplicates. Mean values are shown, and error bars represent SD. p-values for a difference between wild-type and ORFL147C-deficient virus or wild-type and ORFL147C-V5 overexpressing HFFF-TERT cells were estimated using a two-tailed Student's t-test. *p<0.05, ***p<0.001, ****p<0.0001.

6.4 Discussion

Despite previous detection by ribosome profiling or mass spectrometry, the functions of non-canonical ORFs remain uncharacterised. Three ORFs previously identified by ribosomal profiling and two 6FT-ORFs were identified as interactors of canonical HCMV proteins. Additionally, seven interacting 6FT-ORFs had not previously been identified. This data suggests that at least a subset of uncharacterised ORFs may play a role in infection, warranting further characterisation. Initial predictions of these functions can be achieved by interaction analysis.

ORFL147C contains RGG RNA-binding motifs similarly to other herpesvirus proteins that interact with host factors involved in mRNA processing. Additionally, interactions with proteins involved in mRNA splicing, MBNL1 and CELF1, have been validated [193]. Another interactor of ORFL147C, Ribonucleotide PTB-binding 1 (RAVER1) modulates alternative splicing events. In fact, up to 100 splice junctions have been identified in HCMV [67, 70, 319], with several spliced transcripts identified at all times post infection [320].

Deletion of ORFL147C slowed viral replication. However, its large HCIP network suggests that various putative mechanisms (for example splicing or transcriptional effects) which could underlie this observation and thus need to be examined.

The ORFL147C protein is the most abundant non-canonical ORF [193], yet the absence of homologues in other genera of human herpesvirus may explain why its role in viral infection has not been studied to date.

7 | Conclusion and future work

Parts of the work presented in this chapter have been published in a similar form in eLife (Nobre et al., 2019).

This dataset identifies virus-virus and virus-host interactions for 160/171 canonical HCMV proteins, comprising the largest host-pathogen protein interaction network to date and the first interactome for a DNA virus in infected cells. Thus, this dataset may prove valuable in future studies of HCMV and other herpesviruses.

Systematic analysis of this dataset has predicted functions and domain associations for multiple uncharacterized or partly characterized viral proteins, in addition to providing evidence that the non-canonical HCMV protein ORFL147C may play a role in viral infection.

The combination of interactome data generated in the present study with previous screens of protein degradation during early HCMV infection identified viral interactors for degraded host factors, revealing potential viral mechanisms of cellular protein degradation. Functional enrichment analysis of the interactome data suggested that HCMV devotes multiple proteins to interactions with the ubiquitin conjugation pathway, with a subset of viral proteins as hubs of degradation for multiple ubiquitin E3 ligases.

In addition to the vast protein-protein interaction dataset, the HCMV interactome has generated other resources which have already proven useful to other researchers in the field, such as the library of lentiviral expression vectors for each HCMV canonical gene, a collection of stable cell lines overexpressing each canonical HCMV protein as well as lysates from these, which allow screening of single-gene effects.

There are several lines of research arising from this work which could be pursued. One line of research would be to understand how changes in the methodology could better results. CompPass filtering was unable to identify HCIPs for nine canonical HCMV genes, thus it would be important to understand which changes to the methodology could address this. For example, cryogenic cell lysis or a different detergent could be trialled to improve identification of weak interactions. The use of an inducible expression system where abundance of the bait protein can be tuned, may also improve results for baits with low constitutive expression.

Further research could aim to understand the biological implications of the fixed time-point of 60 h on the HCMV interactome. For example, the viral proteins could be divided in five batches according to their relative abundance temporal profile and immunoprecipitated at the time point for which their expression peaks. This could potentially reveal novel interactions which are time-point dependent but biologically relevant. Additionally, for HCMV proteins expressed during latency, future research could aim to determine how this part of the viral life cycle affects the list of binding partners.

Several lines of research stem from chapter 5. One example would be to apply the methodology in this study to other ORFs identified by ribosome profiling and proteomics to determine whether they are functional polypeptides. Novel interactions reported in the HCMV interactome dataset require further validation followed by studies into their biological significance in viral infection. The role of UL72 binding to CCR4-NOT complex or UL43 binding to the 14-3-3 proteins which are involved in diverse signalling functions and present in the virion, are only two examples of several that could be examined.

The dataset did not reveal every viral bait that interacts with degraded human prey. A possible explanation is that such proteins were degraded to below the limit of detection by MS. Further research could aim to repeat this interactome in the presence of

lysosomal and proteasomal inhibition to identify such targets. Regarding the overlap of this dataset with previously published degradation data, future experiments may focus on showing sufficiency of the identified viral genes for degradation of host factors.

Protein domain association analysis requires experimental validation to determine whether the co-occurring domains are necessary for the binding between proteins which contain these domains. Further research may employ single-residue mutants and truncated protein forms to determine how changes in these domains affect binding to interactors.

Another line of research, which follows from chapter 6, would be to elucidate the function of ORFL147C protein in HCMV infection. Preliminary analysis on this ORF raises the hypothesis that it could play a role in mRNA export using different cellular machinery from UL69/ICP27. Firstly, it could be tested whether the presence of RGG box RNA-binding domains allow ORFL147C protein the ability to bind mRNA molecules. Secondly, the subcellular location of this protein should be examined to determine whether it is able to shuttle between nucleus and cytoplasm. Validation of interaction with its HCIP XPO7, which mediates nuclear export, could then be performed in a similar way to validation of other interactions reported in this study. Finally, it could be tested whether an ORFL147C mutant that is unable to bind nuclear export machinery or mRNA, results in a similar viral growth phenotype as observed for the ORFL147C deletion virus.

Apart from the newly-approved Letermovir, all drugs used to treat HCMV infection induce significant side effects and are subject to the development of mutations which can confer resistance. Innovative therapeutic strategies are thus required and may arise from the identification of key interactions in virus-virus or virus-host protein complexes. For example, small molecule inhibitors may be developed to selectively target and disrupt these interactions, potentially restoring innate antiviral restriction through the inhibition

of degradation of host factors [321-323]. Factors targeted by a single viral protein provide a more straightforward opportunity for therapeutic interruption by small molecule inhibitors, such as for example the interaction of UL72 with members of the CCR4-NOT complex. If this interaction proves important for HCMV infection, the function of the CCR4-NOT complex may be disrupted, for example by employing inhibitors of the CNOT7 deadenylase, providing another target for antiviral therapy [324]. Furthermore, a combination of therapies simultaneously targeting several pathways may also be employed to inhibit viral replication. For example, cytotoxins may be developed based on ligands for the viral GPCRs, enabling targeting and killing of cells infected with HCMV [53].

More broadly, future research may compare this dataset with other published viral interactomes in order to identify common pathways targeted by viruses as well as diverging strategies of host factor modulation.

8 | Publications

8.1 Journal articles published during PhD

Nobre L*, Nightingale K, Ravenhill BJ, Antrobus R, Soday L, Nichols J, Seirafian S, Davies J, Wang ECY, Davison AJ, Wilkinson GWG, Stanton RJ, Huttlin EL, Weekes MP. "Global analysis of the human cytomegalovirus interactome identifies degradation hubs, domain associations and viral protein functions."
eLife (2019)

Ravenhill BJ, Kanjee U, Ahouidi A, **Nobre L**, Williamson J, Goldberg JM, Antrobus R, Dieye T, Duraisingh MT, Weekes MP. "Quantitative comparative analysis of erythrocyte cell surface proteins between individuals from two genetically distinct populations".
Commun Biol (2019)

Wang LW, Wang Z, Ersing I, **Nobre L**, Guo, Jiang S, Trudeau S, Zhao B, Weekes MP, Gewurz BE. "Epstein-Barr Virus Subverts Mevalonate and Fatty Acid Pathways to Promote Infected B-cell Proliferation and Survival".
PLoS Pathog (2019), 15(9):e1008030

Wang LW, Shen H, **Nobre L***, Ersing I, Paulo JA, Trudeau S, Wang Z, Smith NA, Ma Y, Reinstadler B, Nomburg J, Sommermann T, Cahir-McFarland E, Gygi SP, Mootha VK, Weekes MP, Gewurz BE. "Epstein-Barr-Virus-Induced One-Carbon Metabolism Drives B Cell Transformation."
Cell Metab (2019), pii: S1550-4131(19)30306-7

Nightingale K, Lin KM, Ravenhill BJ, Davies C, **Nobre L**, Fielding CA, Ruckova E, Fletcher-Etherington A, Soday L, Nichols H, Sugrue D, Wang ECY, Moreno P, Umrania Y, Huttlin EL, Antrobus R, Davison AJ, Wilkinson GWG, Stanton RJ, Tomasec P, Weekes MP. "High-Definition Analysis of Host Protein Stability during Human Cytomegalovirus Infection Reveals Antiviral Factors and Viral Evasion Mechanisms."
Cell Host Microbe (2018), 24 (3): 447-460

Wang ECY, Pjechova M, Nightingale K, Vlahava VM, Patel M, Ruckova E, Forbes SK, **Nobre L**, Antrobus R, Roberts D, Fielding CA, Seirafian S, Davies J, Murrell I, Lau B, Wilkie GS, Suárez NM, Stanton RJ, Vojtesek B, Davison A, Lehner PJ, Weekes MP, Wilkinson GWG, Tomasec P. "Suppression of costimulation by human cytomegalovirus promotes evasion of cellular immune defences."
Proc Natl Acad Sci U S A (2018), 115 (19): 4998-5003

Dankwa S, Chaand M, Kanjee U, Jiang RHY, **Nobre LV**, Goldberg JM, Bei AK, Moechtar MA, Grüring C, Ahouidi AD, Ndiaye D, Dieye TN, Mboup S, Weekes MP, Duraisingh MT. "Genetic Evidence for Erythrocyte Receptor Glycophorin B Expression Levels Defining a Dominant Plasmodium falciparum Invasion Pathway into Human Erythrocytes."
Infect Immun (2017), 85(10). pii: e00074-17

Ersing I, **Nobre L***, Wang LW, Soday L, Ma Y, Paulo JA, Narita Y, Ashbaugh CW, Jiang C, Grayson NE, Kieff E, Gygi SP, Weekes MP, Gewurz BE. "A Temporal Proteomic Map of Epstein-Barr Virus Lytic Replication in B Cells."
Cell Rep (2017), 19(7):1479-1493

Fielding CA, Weekes MP, **Nobre LV**, Ruckova E, Wilkie GS, Paulo JA, Chang C, Suárez NM, Davies JA, Antrobus R, Stanton RJ, Aicheler RJ, Nichols H, Vojtesek B, Trowsdale J, Davison AJ, Gygi SP, Tomasec P, Lehner PJ, Wilkinson GW. "Control of immune ligands by members of a cytomegalovirus gene expansion suppresses natural killer cell activation."
eLife (2017), 6. pii: e22206

8.2 Manuscripts referenced in dissertation and currently in preparation for publication:

Fletcher-Etherington A, **Nobre L**, Nightingale K, Antrobus R, Nichols J, Davison AJ, Stanton RJ, Weekes MP. "Human cytomegalovirus protein pUL36: a dual cell death pathway inhibitor."

Nightingale K, Fielding C, **Nobre L**, Davies C, Wang E. C. Y, Antrobus R, Davison AJ, Tomasec P, Stanton RJ, Weekes MP. "Human Cytomegalovirus Protein RL1 degrades the antiviral factor SLFN11 Through Recruitment of the CRL4 E3 Ubiquitin Ligase Complex."

ORCID iD: 0000-0003-0467-8989

9 | References

1. Davison, A.J., et al., *The order Herpesvirales.* Archives of virology, 2009. **154**(1): p. 171-177.
2. Roizmann, B., et al., *The family Herpesviridae: an update. The Herpesvirus Study Group of the International Committee on Taxonomy of Viruses.* Arch Virol, 1992. **123**(3-4): p. 425-49.
3. Roizman, B., *The Family Herpesviridae: General Description, Taxonomy, and Classification*, in *The Herpesviruses*, B. Roizman, Editor. 1982. p. 1-17.
4. Davison, A.J., *Evolution of the herpesviruses.* Vet Microbiol, 2002. **86**(1-2): p. 69-88.
5. Davison, A., *Comments on the phylogenetics and evolution of herpesviruses and other large DNA viruses.* Virus Res, 2002. **82**(1-2): p. 127-32.
6. Cohrs, R.J. and D.H. Gilden, *Human herpesvirus latency.* Brain Pathol, 2001. **11**(4): p. 465-74.
7. Pantry, S.N. and P.G. Medveczky, *Latency, Integration, and Reactivation of Human Herpesvirus-6.* Viruses, 2017. **9**(7): p. 194.
8. Speck, S.H. and D. Ganem, *Viral latency and its regulation: lessons from the gamma-herpesviruses.* Cell host & microbe, 2010. **8**(1): p. 100-115.
9. Davison, A.J., *Comparative analysis of the genomes*, in *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, A. Arvin, et al., Editors. 2007, ambridge: Cambridge University Press: Cambridge.
10. Whitley, R.J., *Herpesvirus*, in *Medical Microbiology*, S. Baron, Editor. 1996, University of Texas Medical Branch at Galveston: Galveston (TX).
11. Mocarski, E.S., et al., eds. *Cytomegaloviruses*. 6 ed. Fields Virology, ed. D.M. Knipe and P.M. Howley. Vol. 2. 2013, Lipincott Williams and Wilkins, Philadelphia. 1960-2014.
12. Azevedo, L.S., et al., *Cytomegalovirus infection in transplant recipients.* Clinics (Sao Paulo), 2015. **70**(7): p. 515-23.
13. Streblow, D.N., S.L. Orloff, and J.A. Nelson, *Acceleration of allograft failure by cytomegalovirus.* Curr Opin Immunol, 2007. **19**(5): p. 577-82.
14. Malm, G. and M.L. Engman, *Congenital cytomegalovirus infections.* Semin Fetal Neonatal Med, 2007. **12**(3): p. 154-9.
15. Dollard, S.C., S.D. Grosse, and D.S. Ross, *New estimates of the prevalence of neurological and sensory sequelae and mortality associated with congenital cytomegalovirus infection.* Rev Med Virol, 2007. **17**(5): p. 355-63.
16. Grosse, S.D., D.S. Ross, and S.C. Dollard, *Congenital cytomegalovirus (CMV) infection as a cause of permanent bilateral hearing loss: a quantitative assessment.* J Clin Virol, 2008. **41**(2): p. 57-62.
17. Gibson, W., *Structural and nonstructural proteins of strain Colburn cytomegalovi.* Virology, 1981. **111**(2): p. 516-537.
18. Baldick, C.J., Jr. and T. Shenk, *Proteins associated with purified human cytomegalovirus particles.* Journal of Virology, 1996. **70**(9): p. 6097-105.
19. Britt, W.J. and S. Boppana, *Human cytomegalovirus virion proteins.* Hum Immunol, 2004. **65**(5): p. 395-402.
20. Borst, E.M., et al., *The human cytomegalovirus UL51 protein is essential for viral genome cleavage-packaging and interacts with the terminase subunits pUL56 and pUL89.* J Virol, 2013. **87**(3): p. 1720-32.
21. Scheffczik, H., et al., *The terminase subunits pUL56 and pUL89 of human cytomegalovirus are DNA-metabolizing proteins with toroidal structure.* Nucleic Acids Res, 2002. **30**(7): p. 1695-703.
22. Borst, E.M., et al., *The Essential Human Cytomegalovirus Proteins pUL77 and pUL93 Are Structural Components Necessary for Viral Genome Encapsidation.* J Virol, 2016. **90**(13): p. 5860-5875.

23. Koppen-Rung, P., A. Dittmer, and E. Bogner, *Intracellular Distribution of Capsid-Associated pUL77 of Human Cytomegalovirus and Interactions with Packaging Proteins and pUL93.* J Virol, 2016. **90**(13): p. 5876-5885.
24. DeRussy, B.M. and R. Tandon, *Human Cytomegalovirus pUL93 Is Required for Viral Genome Cleavage and Packaging.* J Virol, 2015. **89**(23): p. 12221-5.
25. DeRussy, B.M., M.T. Boland, and R. Tandon, *Human Cytomegalovirus pUL93 Links Nucleocapsid Maturation and Nuclear Egress.* J Virol, 2016. **90**(16): p. 7109-7117.
26. Hulo, C., et al., *ViralZone: a knowledge resource to understand virus diversity.* Nucleic Acids Res, 2011. **39**(Database issue): p. D576-82.
27. Kalejta, R.F., *Tegument proteins of human cytomegalovirus.* Microbiology and Molecular Biology Reviews, 2008. **72**(2): p. 249-65, table of contents.
28. Sharma, M., et al., *Human cytomegalovirus UL50 and UL53 recruit viral protein kinase UL97, not protein kinase C, for disruption of nuclear lamina and nuclear egress in infected cells.* J Virol, 2014. **88**(1): p. 249-62.
29. Gibson, W., et al., *Human cytomegalovirus (HCMV) smallest capsid protein identified as product of short open reading frame located between HCMV UL48 and UL49.* J Virol, 1996. **70**(8): p. 5680-3.
30. Mocarski, E.S. and C.T. Courcelle, *Cytomegalovirus and their replication*, in *Fields Virology*, B.N. Fields, et al., Editors. 2001, Lippincott Williams & Wilkins: Philadelphia.
31. Brock, I., et al., *Nuclear targeting of human cytomegalovirus large tegument protein pUL48 is essential for viral growth.* Journal of virology, 2013. **87**(10): p. 6005-6019.
32. Kim, Y.E., et al., *Involvement of the N-Terminal Deubiquitinating Protease Domain of Human Cytomegalovirus UL48 Tegument Protein in Autoubiquitination, Virion Stability, and Virus Entry.* J Virol, 2016. **90**(6): p. 3229-42.
33. Kim, E.T., et al., *Cleavage specificity of the UL48 deubiquitinating protease activity of human cytomegalovirus and the growth of an active-site mutant virus in cultured cells.* J Virol, 2009. **83**(23): p. 12046-56.
34. Tandon, R. and E.S. Mocarski, *Cytomegalovirus pUL96 is critical for the stability of pp150-associated nucleocapsids.* Journal of virology, 2011. **85**(14): p. 7129-7141.
35. Compton, T., D.M. Nowlin, and N.R. Cooper, *Initiation of human cytomegalovirus infection requires initial interaction with cell surface heparan sulfate.* Virology, 1993. **193**(2): p. 834-41.
36. Wille, P.T., et al., *Human cytomegalovirus (HCMV) glycoprotein gB promotes virus entry in trans acting as the viral fusion protein rather than as a receptor-binding protein.* mBio, 2013. **4**(3): p. e00332-13.
37. Kari, B. and R. Gehrz, *A human cytomegalovirus glycoprotein complex designated gC-II is a major heparin-binding component of the envelope.* J Virol, 1992. **66**(3): p. 1761-4.
38. Mach, M., et al., *Complex formation by human cytomegalovirus glycoproteins M (gpUL100) and N (gpUL73).* Journal of Virology, 2000. **74**(24): p. 11881-92.
39. Ryckman, B.J., et al., *Characterization of the human cytomegalovirus gH/gL/UL128-131 complex that mediates entry into epithelial and endothelial cells.* J Virol, 2008. **82**(1): p. 60-70.
40. Calo, S., et al., *The Human Cytomegalovirus UL116 Gene Encodes an Envelope Glycoprotein Forming a Complex with gH Independently from gL.* J Virol, 2016. **90**(10): p. 4926-38.
41. Lilley, B.N., H.L. Ploegh, and R.S. Tirabassi, *Human cytomegalovirus open reading frame TRL11/IRL11 encodes an immunoglobulin G Fc-binding protein.* Journal of Virology, 2001. **75**(22): p. 11218-21.
42. Sprague, E.R., et al., *The human cytomegalovirus Fc receptor gp68 binds the Fc CH2-CH3 interface of immunoglobulin G.* J Virol, 2008. **82**(7): p. 3490-9.
43. Cortese, M., et al., *Recombinant human cytomegalovirus (HCMV) RL13 binds human immunoglobulin G Fc.* PLoS One, 2012. **7**(11): p. e50166.

44. Wang, D., W. Bresnahan, and T. Shenk, *Human cytomegalovirus encodes a highly specific RANTES decoy receptor.* Proc Natl Acad Sci U S A, 2004. **101**(47): p. 16642-7.
45. Tu, C.C., et al., *Human Cytomegalovirus UL111A and US27 Gene Products Enhance the CXCL12/CXCR4 Signaling Axis via Distinct Mechanisms.* J Virol, 2018. **92**(5).
46. Tadagaki, K., et al., *Human cytomegalovirus-encoded UL33 and UL78 heteromerize with host CCR5 and CXCR4 impairing their HIV coreceptor activity.* Blood, 2012. **119**(21): p. 4908-18.
47. O'Connor, C.M. and T. Shenk, *Human cytomegalovirus pUS27 G protein-coupled receptor homologue is required for efficient spread by the extracellular route but not for direct cell-to-cell spread.* J Virol, 2011. **85**(8): p. 3700-7.
48. Neote, K., et al., *Molecular cloning, functional expression, and signaling characteristics of a C-C chemokine receptor.* Cell, 1993. **72**(3): p. 415-25.
49. Gao, J.L. and P.M. Murphy, *Human cytomegalovirus open reading frame US28 encodes a functional beta chemokine receptor.* Journal of Biological Chemistry, 1994. **269**(46): p. 28539-42.
50. Kuhn, D.E., C.J. Beall, and P.E. Kolattukudy, *The Cytomegalovirus US28 Protein Binds Multiple CC Chemokines with High Affinity.* Biochemical and Biophysical Research Communications, 1995. **211**(1): p. 325-330.
51. Kledal, T.N., M.M. Rosenkilde, and T.W. Schwartz, *Selective recognition of the membrane-bound CX3C chemokine, fractalkine, by the human cytomegalovirus-encoded broad-spectrum receptor US28.* FEBS Lett, 1998. **441**(2): p. 209-14.
52. Humby, M.S. and C.M. Connor, *Human Cytomegalovirus US28 Is Important for Latent Infection of Hematopoietic Progenitor Cells.* Journal of Virology, 2016. **90**(6): p. 2959.
53. Krishna, B.A., et al., *Targeting the latent cytomegalovirus reservoir with an antiviral fusion toxin protein.* Nat Commun, 2017. **8**: p. 14321.
54. Krishna, B.A., et al., *Latency-Associated Expression of Human Cytomegalovirus US28 Attenuates Cell Signaling Pathways To Maintain Latent Infection.* mBio, 2017. **8**(6): p. e01754-17.
55. Krishna, B.A., W.E. Miller, and C.M. O'Connor, *US28: HCMV's Swiss Army Knife.* Viruses, 2018. **10**(8).
56. Varnum, S.M., et al., *Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome.* Journal of Virology, 2004. **78**(20): p. 10960-6.
57. Shikhagaie, M., et al., *The human cytomegalovirus-specific UL1 gene encodes a late-phase glycoprotein incorporated in the virion envelope.* Journal of Virology, 2012. **86**(8): p. 4091-101.
58. Liu, F. and Z.H. Zhou, *Comparative virion structures of human herpesviruses*, in *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, A. Arvin, et al., Editors. 2007, Cambridge University Press: Cambridge.
59. Prichard, M.N., et al., *Identification of persistent RNA-DNA hybrid structures within the origin of replication of human cytomegalovirus.* J Virol, 1998. **72**(9): p. 6997-7004.
60. Bresnahan, W.A. and T. Shenk, *A subset of viral transcripts packaged within human cytomegalovirus particles.* Science, 2000. **288**(5475): p. 2373-6.
61. Mohammad, A.A., et al., *Human cytomegalovirus microRNAs are carried by virions and dense bodies and are delivered to target cells.* J Gen Virol, 2017. **98**(5): p. 1058-1072.
62. Terhune, S.S., J. Schroer, and T. Shenk, *RNAs are packaged into human cytomegalovirus virions in proportion to their intracellular concentration.* J Virol, 2004. **78**(19): p. 10390-8.
63. Gibson, W., et al., *D,L-alpha-difluoromethylornithine inhibits human cytomegalovirus replication.* Journal of virology, 1984. **50**(1): p. 145-154.
64. Irmiere, A. and W. Gibson, *Isolation of human cytomegalovirus intranuclear capsids, characterization of their protein constituents, and demonstration that the B-capsid assembly protein is also abundant in noninfectious enveloped particles.* J Virol, 1985. **56**(1): p. 277-83.
65. Mocarski, E.S., Jr., *Cytomegaloviruses and their replication.* Fields Virology, 2000.

66. Van Damme, E. and M. Van Loock, *Functional annotation of human cytomegalovirus gene products: an update.* Front Microbiol, 2014. **5**: p. 218.
67. Stern-Ginossar, N., et al., *Decoding human cytomegalovirus.* Science, 2012. **338**(6110): p. 1088-93.
68. Poole, E., M. Wills, and J. Sinclair, *Human Cytomegalovirus Latency: Targeting Differences in the Latently Infected Cell with a View to Clearing Latent Infection.* New Journal of Science, 2014. **2014**: p. 10.
69. Dolan, A., et al., *Genetic content of wild-type human cytomegalovirus.* Journal of General Virology, 2004. **85**(Pt 5): p. 1301-12.
70. Gatherer, D., et al., *High-resolution human cytomegalovirus transcriptome.* Proc Natl Acad Sci U S A, 2011. **108**(49): p. 19755-60.
71. Colletti, K.S., et al., *Human cytomegalovirus UL84 interacts with an RNA stem-loop sequence found within the RNA/DNA hybrid region of oriLyt.* J Virol, 2007. **81**(13): p. 7077-85.
72. Feire, A.L., H. Koss, and T. Compton, *Cellular integrins function as entry receptors for human cytomegalovirus via a highly conserved disintegrin-like domain.* Proc Natl Acad Sci U S A, 2004. **101**(43): p. 15470-5.
73. Smith, R.M., S. Kosuri, and J.A. Kerry, *Role of human cytomegalovirus tegument proteins in virion assembly.* Viruses, 2014. **6**(2): p. 582-605.
74. Jean Beltran, P.M. and I.M. Cristea, *The life cycle and pathogenesis of human cytomegalovirus infection: lessons from proteomics.* Expert review of proteomics, 2014. **11**(6): p. 697-711.
75. Mohammad, A.-A., *Human Cytomegalovirus: From Novel Strain, miRNAs to Interplay with Breast Cancer*. 2017.
76. Elder, E. and J. Sinclair, *HCMV latency: what regulates the regulators?* Medical microbiology and immunology, 2019. **208**(3-4): p. 431-438.
77. Taylor-Wiedeman, J., et al., *Monocytes are a major site of persistence of human cytomegalovirus in peripheral blood mononuclear cells.* Journal of General Virology, 1991. **72**: p. 2059-64.
78. Mendelson, M., et al., *Detection of endogenous human cytomegalovirus in CD34+ bone marrow progenitors.* J Gen Virol, 1996. **77**: p. 3099-102.
79. Reeves, M. and J. Sinclair, *Aspects of Human Cytomegalovirus Latency and Reactivation*, in *Human Cytomegalovirus*, T.E. Shenk and M.F. Stinski, Editors. 2008, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 297-313.
80. Schwartz, M. and N. Stern-Ginossar, *The Transcriptome of Latent Human Cytomegalovirus.* J Virol, 2019.
81. Jenkins, C., A. Abendroth, and B. Slobedman, *A novel viral transcript with homology to human interleukin-10 is expressed during latent human cytomegalovirus infection.* J Virol, 2004. **78**(3): p. 1440-7.
82. Tarrant-Elorza, M., Cyprian C. Rossetto, and Gregory S. Pari, *Maintenance and Replication of the Human Cytomegalovirus Genome during Latency.* Cell Host & Microbe, 2014. **16**(1): p. 43-54.
83. Shnayder, M., et al., *Defining the Transcriptional Landscape during Cytomegalovirus Latency with Single-Cell RNA Sequencing.* mBio, 2018. **9**(2): p. e00013-18.
84. Cheng, S., et al., *Transcriptome-wide characterization of human cytomegalovirus in natural infection and experimental latency.* Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(49): p. E10586-E10595.
85. Lang, D.J. and J.F. Kummer, *Cytomegalovirus in semen: observations in selected populations.* J Infect Dis, 1975. **132**(4): p. 472-3.
86. Stagno, S., et al., *Breast milk and the risk of cytomegalovirus infection.* N Engl J Med, 1980. **302**(19): p. 1073-6.

87. Britt, W., *Manifestations of human cytomegalovirus infection: proposed mechanisms of acute and chronic disease.* Curr Top Microbiol Immunol, 2008. **325**: p. 417-70.
88. Sinzger, C., M. Digel, and G. Jahn, *Cytomegalovirus cell tropism.* Curr Top Microbiol Immunol, 2008. **325**: p. 63-83.
89. Murphy, E. and T. Shenk, *Human cytomegalovirus genome.* Curr Top Microbiol Immunol, 2008. **325**: p. 1-19.
90. Wilkinson, G.W., et al., *Human cytomegalovirus: taking the strain.* Med Microbiol Immunol, 2015. **204**(3): p. 273-84.
91. Plotkin, S.A., et al., *Candidate cytomegalovirus strain for human vaccination.* Infect Immun, 1975. **12**(3): p. 521-7.
92. Bradley, A.J., et al., *High-throughput sequence analysis of variants of human cytomegalovirus strains Towne and AD169.* J Gen Virol, 2009. **90**(Pt 10): p. 2375-80.
93. Sinzger, C., et al., *Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E.* J Gen Virol, 2008. **89**(Pt 2): p. 359-68.
94. Murrell, I., et al., *Impact of sequence variation in the UL128 locus on production of human cytomegalovirus in fibroblast and epithelial cells.* J Virol, 2013. **87**(19): p. 10489-500.
95. Cha, T.A., et al., *Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains.* J Virol, 1996. **70**(1): p. 78-83.
96. Akter, P., et al., *Two novel spliced genes in human cytomegalovirus.* J Gen Virol, 2003. **84**(Pt 5): p. 1117-22.
97. Murphy, E., et al., *Reevaluation of human cytomegalovirus coding potential.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13585-90.
98. Stanton, R.J., et al., *Reconstruction of the complete human cytomegalovirus genome in a BAC reveals RL13 to be a potent inhibitor of replication.* Journal of Clinical Investigation, 2010. **120**(9): p. 3191-208.
99. Biron, C.A., K.S. Byron, and J.L. Sullivan, *Severe herpesvirus infections in an adolescent without natural killer cells.* N Engl J Med, 1989. **320**(26): p. 1731-5.
100. Gazit, R., et al., *Expression of KIR2DL1 on the entire NK cell population: a possible novel immunodeficiency syndrome.* Blood, 2004. **103**(5): p. 1965-6.
101. Vivier, E., et al., *Functions of natural killer cells.* Nat Immunol, 2008. **9**(5): p. 503-10.
102. Yurochko, A.D., et al., *The human cytomegalovirus UL55 (gB) and UL75 (gH) glycoprotein ligands initiate the rapid activation of Sp1 and NF-kappaB during infection.* J Virol, 1997. **71**(7): p. 5051-9.
103. Navarro, L., et al., *Cytomegalovirus activates interferon immediate-early response gene expression and an interferon regulatory factor 3-containing interferon-stimulated response element-binding complex.* Mol Cell Biol, 1998. **18**(7): p. 3796-802.
104. Boehme, K.W., M. Guerrero, and T. Compton, *Human cytomegalovirus envelope glycoproteins B and H are necessary for TLR2 activation in permissive cells.* J Immunol, 2006. **177**(10): p. 7094-102.
105. Boehme, K.W., et al., *Human cytomegalovirus elicits a coordinated cellular antiviral response via envelope glycoprotein B.* J Virol, 2004. **78**(3): p. 1202-11.
106. Compton, T., et al., *Human cytomegalovirus activates inflammatory cytokine responses via CD14 and Toll-like receptor 2.* Journal of Virology, 2003. **77**(8): p. 4588-96.
107. DeFilippis, V.R., et al., *Human cytomegalovirus induces the interferon response via the DNA sensor ZBP1.* Journal of Virology, 2010. **84**(1): p. 585-98.
108. Woodhall, D.L., et al., *Human Daxx-mediated repression of human cytomegalovirus gene expression correlates with a repressive chromatin structure around the major immediate early promoter.* Journal of Biological Chemistry, 2006. **281**(49): p. 37652-60.
109. Saffert, R.T. and R.F. Kalejta, *Inactivating a cellular intrinsic immune defense mediated by Daxx is the mechanism through which the human cytomegalovirus pp71 protein stimulates viral immediate-early gene expression.* J Virol, 2006. **80**(8): p. 3863-71.

110. Tavalai, N., et al., *Nuclear domain 10 components promyelocytic leukemia protein and hDaxx independently contribute to an intrinsic antiviral defense against human cytomegalovirus infection.* J Virol, 2008. **82**(1): p. 126-37.
111. Tavalai, N., et al., *Evidence for a role of the cellular ND10 protein PML in mediating intrinsic immunity against human cytomegalovirus infections.* J Virol, 2006. **80**(16): p. 8006-18.
112. Marshall, E.E., et al., *Essential role for either TRS1 or IRS1 in human cytomegalovirus replication.* Journal of Virology, 2009. **83**(9): p. 4112-20.
113. Klenerman, P. and A. Oxenius, *T cell responses to cytomegalovirus.* Nat Rev Immunol, 2016. **16**(6): p. 367-77.
114. Sylwester, A.W., et al., *Broadly targeted human cytomegalovirus-specific CD4+ and CD8+ T cells dominate the memory compartments of exposed subjects.* The Journal of experimental medicine, 2005. **202**(5): p. 673-685.
115. Appay, V., et al., *Characterization of CD4(+) CTLs ex vivo.* J Immunol, 2002. **168**(11): p. 5954-8.
116. van Leeuwen, E.M., et al., *Emergence of a CD4+CD28- granzyme B+, cytomegalovirus-specific T cell subset after recovery of primary cytomegalovirus infection.* J Immunol, 2004. **173**(3): p. 1834-41.
117. Casazza, J.P., et al., *Acquisition of direct antiviral effector functions by CMV-specific CD4+ T lymphocytes with cellular maturation.* The Journal of experimental medicine, 2006. **203**(13): p. 2865-2877.
118. van den Berg, S.P.H., et al., *The hallmarks of CMV-specific CD8 T-cell differentiation.* Med Microbiol Immunol, 2019. **208**(3-4): p. 365-373.
119. Jackson, S.E., G.M. Mason, and M.R. Wills, *Human cytomegalovirus immunity and immune evasion.* Virus Res, 2011. **157**(2): p. 151-60.
120. Jackson, S.E., et al., *Human Cytomegalovirus (HCMV)-Specific CD4(+) T Cells Are Polyfunctional and Can Respond to HCMV-Infected Dendritic Cells In Vitro.* J Virol, 2017. **91**(6).
121. van Leeuwen, E.M., et al., *Strong selection of virus-specific cytotoxic CD4+ T-cell clones during primary human cytomegalovirus infection.* Blood, 2006. **108**(9): p. 3121-7.
122. Britt, W.J., et al., *Cell surface expression of human cytomegalovirus (HCMV) gp55-116 (gB): use of HCMV-recombinant vaccinia virus-infected cells in analysis of the human neutralizing antibody response.* Journal of Virology, 1990. **64**(3): p. 1079-85.
123. Boppana, S.B. and W.J. Britt, *Recognition of human cytomegalovirus gene products by HCMV-specific cytotoxic T cells.* Virology, 1996. **222**(1): p. 293-6.
124. Landini, M.P., E. Rossier, and H. Schmitz, *Antibodies to human cytomegalovirus structural polypeptides during primary infection.* J Virol Methods, 1988. **22**(2-3): p. 309-17.
125. Rasmussen, L., et al., *Antibody response to human cytomegalovirus glycoproteins gB and gH after natural infection in humans.* J Infect Dis, 1991. **164**(5): p. 835-42.
126. Macagno, A., et al., *Isolation of human monoclonal antibodies that potently neutralize human cytomegalovirus infection by targeting different epitopes on the gH/gL/UL128-131A complex.* J Virol, 2010. **84**(2): p. 1005-13.
127. Urban, M., et al., *Glycoprotein H of human cytomegalovirus is a major antigen for the neutralizing humoral immune response.* J Gen Virol, 1996. **77 (Pt 7)**: p. 1537-47.
128. Noriega, V., et al., *Diverse immune evasion strategies by human cytomegalovirus.* Immunol Res, 2012. **54**(1-3): p. 140-51.
129. Kim, Y.E., et al., *Human cytomegalovirus infection causes degradation of Sp100 proteins that suppress viral gene expression.* Journal of Virology, 2011. **85**(22): p. 11928-37.
130. Wilkinson, G.W., et al., *Modulation of natural killer cells by human cytomegalovirus.* Journal of Clinical Virology, 2008. **41**(3): p. 206-12.
131. Cosman, D., et al., *A novel immunoglobulin superfamily receptor for cellular and viral MHC class I molecules.* Immunity, 1997. **7**(2): p. 273-82.

132. Chalupny, N.J., et al., *Down-regulation of the NKG2D ligand MICA by the human cytomegalovirus glycoprotein UL142.* Biochemical and Biophysical Research Communications, 2006. **346**(1): p. 175-81.
133. Tomasec, P., et al., *Surface expression of HLA-E, an inhibitor of natural killer cells, enhanced by human cytomegalovirus gpUL40.* Science, 2000. **287**(5455): p. 1031.
134. Arnon, T.I., et al., *Inhibition of the NKp30 activating receptor by pp65 of human cytomegalovirus.* Nat Immunol, 2005. **6**(5): p. 515-23.
135. Kubin, M., et al., *ULBP1, 2, 3: novel MHC class I-related molecules that bind to human cytomegalovirus glycoprotein UL16, activate NK cells.* Eur J Immunol, 2001. **31**(5): p. 1428-37.
136. Welte, S.A., et al., *Selective intracellular retention of virally induced NKG2D ligands by the human cytomegalovirus UL16 glycoprotein.* European Journal of Immunology, 2003. **33**(1): p. 194-203.
137. Spreu, J., T. Stehle, and A. Steinle, *Human cytomegalovirus-encoded UL16 discriminates MIC molecules by their alpha2 domains.* J Immunol, 2006. **177**(5): p. 3143-9.
138. Tomasec, P., et al., *Downregulation of natural killer cell-activating ligand CD155 by human cytomegalovirus UL141.* Nat Immunol, 2005. **6**(2): p. 181-8.
139. Stern-Ginossar, N., et al., *Host immune system gene targeting by a viral miRNA.* Science, 2007. **317**(5836): p. 376-81.
140. Miller-Kittrell, M. and T.E. Sparer, *Feeling manipulated: cytomegalovirus immune manipulation.* Virol J, 2009. **6**: p. 4.
141. Pandey, J.P., et al., *The decoy Fcgamma receptor encoded by the cytomegalovirus UL119-UL118 gene has differential affinity to IgG proteins expressing different GM allotypes.* Hum Immunol, 2015. **76**(8): p. 591-4.
142. Paulus, C. and M. Nevels, *The human cytomegalovirus major immediate-early proteins as antagonists of intrinsic and innate antiviral host responses.* Viruses, 2009. **1**(3): p. 760-779.
143. Taylor, R.T. and W.A. Bresnahan, *Human cytomegalovirus immediate-early 2 gene expression blocks virus-induced beta interferon production.* J Virol, 2005. **79**(6): p. 3873-7.
144. Child, S.J., et al., *Evasion of cellular antiviral responses by human cytomegalovirus TRS1 and IRS1.* Journal of Virology, 2004. **78**(1): p. 197-205.
145. Spencer, J.V., et al., *Potent immunosuppressive activities of cytomegalovirus-encoded interleukin-10.* J Virol, 2002. **76**(3): p. 1285-92.
146. Jenkins, C., et al., *Immunomodulatory properties of a viral homolog of human interleukin-10 expressed by human cytomegalovirus during the latent phase of infection.* J Virol, 2008. **82**(7): p. 3736-50.
147. Spencer, J.V., *The cytomegalovirus homolog of interleukin-10 requires phosphatidylinositol 3-kinase activity for inhibition of cytokine synthesis in monocytes.* J Virol, 2007. **81**(4): p. 2083-6.
148. Bodaghi, B., et al., *Chemokine sequestration by viral chemoreceptors as a novel viral escape strategy: withdrawal of chemokines from the environment of cytomegalovirus-infected cells.* Journal of Experimental Medicine, 1998. **188**(5): p. 855-66.
149. Vomaske, J., J.A. Nelson, and D.N. Streblow, *Human Cytomegalovirus US28: a functionally selective chemokine binding receptor.* Infect Disord Drug Targets, 2009. **9**(5): p. 548-56.
150. Cheung, T.C., et al., *Evolutionarily divergent herpesviruses modulate T cell activation by targeting the herpesvirus entry mediator cosignaling pathway.* Proc Natl Acad Sci U S A, 2005. **102**(37): p. 13218-23.
151. Poole, E., et al., *The UL144 gene product of human cytomegalovirus activates NFkappaB via a TRAF6-dependent mechanism.* The EMBO journal, 2006. **25**(18): p. 4390-4399.
152. Le, V.T., M. Trilling, and H. Hengel, *The cytomegaloviral protein pUL138 acts as potentiator of tumor necrosis factor (TNF) receptor 1 surface density to enhance ULb'-encoded modulation of TNF-alpha signaling.* J Virol, 2011. **85**(24): p. 13260-70.

153. Montag, C., et al., *The latency-associated UL138 gene product of human cytomegalovirus sensitizes cells to tumor necrosis factor alpha (TNF-alpha) signaling by upregulating TNF-alpha receptor 1 cell surface expression.* J Virol, 2011. **85**(21): p. 11409-21.
154. Krishna, B.A., M.R. Wills, and J.H. Sinclair, *Advances in the treatment of cytomegalovirus.* Br Med Bull, 2019. **131**(1): p. 5-17.
155. Soderholm, S., et al., *Multi-Omics Studies towards Novel Modulators of Influenza A Virus-Host Interaction.* Viruses, 2016. **8**(10).
156. Oda, Y., et al., *Accurate quantitation of protein expression and site-specific phosphorylation.* Proc Natl Acad Sci U S A, 1999. **96**(12): p. 6591-6.
157. McAlister, G.C., et al., *Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses.* Analytical Chemistry, 2012. **84**(17): p. 7469-78.
158. Buscher, N., et al., *The proteome of human cytomegalovirus virions and dense bodies is conserved across different strains.* Med Microbiol Immunol, 2015. **204**(3): p. 285-93.
159. Weekes, M.P., et al., *Latency-associated degradation of the MRP1 drug transporter during latent Human Cytomegalovirus infection.* Science, 2013. **340**: p. 199-202.
160. Weekes, M.P., et al., *Proteomic plasma membrane profiling reveals an essential role for gp96 in the cell surface expression of LDLR family members, including the LDL receptor and LRP6.* J.Proteome Res., 2012. **11**(3): p. 1475-1484.
161. Reyda, S., et al., *Proteomic analyses of human cytomegalovirus strain AD169 derivatives reveal highly conserved patterns of viral and cellular proteins in infected fibroblasts.* Viruses, 2014. **6**(1): p. 172-188.
162. Weekes, M.P., et al., *Quantitative temporal viromics: an approach to investigate host-pathogen interaction.* Cell, 2014. **157**: p. 1460-1472.
163. Nightingale, K., et al., *High-Definition Analysis of Host Protein Stability during Human Cytomegalovirus Infection Reveals Antiviral Factors and Viral Evasion Mechanisms.* Cell Host Microbe, 2018.
164. Cristea, I.M., et al., *Human cytomegalovirus pUL83 stimulates activity of the viral immediate-early promoter through its interaction with the cellular IFI16 protein.* Journal of virology, 2010. **84**(15): p. 7803-7814.
165. Li, T., J. Chen, and I.M. Cristea, *Human cytomegalovirus tegument protein pUL83 inhibits IFI16-mediated DNA sensing for immune evasion.* Cell Host Microbe, 2013. **14**(5): p. 591-9.
166. Oberstein, A., et al., *Human cytomegalovirus pUL97 kinase induces global changes in the infected cell phosphoproteome.* Proteomics, 2015. **15**(12): p. 2006-22.
167. Sinigalia, E., et al., *The human cytomegalovirus DNA polymerase processivity factor UL44 is modified by SUMO in a DNA-dependent manner.* PLoS One, 2012. **7**(11): p. e49630.
168. Tandon, R. and E.S. Mocarski, *Viral and host control of cytomegalovirus maturation.* Trends Microbiol, 2012. **20**(8): p. 392-401.
169. Jean Beltran, P.M., R.A. Mathias, and I.M. Cristea, *A Portrait of the Human Organelle Proteome In Space and Time during Cytomegalovirus Infection.* Cell Syst, 2016. **3**(4): p. 361-373.e6.
170. Moorman, N.J., et al., *A targeted spatial-temporal proteomics approach implicates multiple cellular trafficking pathways in human cytomegalovirus virion maturation.* Mol Cell Proteomics, 2010. **9**(5): p. 851-60.
171. Gao, Y., K. Colletti, and G.S. Pari, *Identification of human cytomegalovirus UL84 virus- and cell-encoded binding partners by using proteomics analysis.* J Virol, 2008. **82**(1): p. 96-104.
172. Du, G. and M.F. Stinski, *Interaction network of proteins associated with human cytomegalovirus IE2-p86 protein during infection: a proteomic analysis.* PLoS One, 2013. **8**(12): p. e81583.
173. Milbradt, J., et al., *Proteomic analysis of the multimeric nuclear egress complex of human cytomegalovirus.* Mol Cell Proteomics, 2014. **13**(8): p. 2132-46.

174. Terhune, S.S., et al., *Human cytomegalovirus UL29/28 protein interacts with components of the NuRD complex which promote accumulation of immediate-early RNA.* PLoS Pathog, 2010. **6**(6): p. e1000965.
175. White, E.A., et al., *Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity.* J Virol, 2012. **86**(24): p. 13174-86.
176. Cristea, I.M., et al., *Tracking and elucidating alphavirus-host protein interactions.* J Biol Chem, 2006. **281**(40): p. 30269-78.
177. Wu, W., et al., *The interactome of the human respiratory syncytial virus NS1 protein highlights multiple effects on host cell biology.* J Virol, 2012. **86**(15): p. 7777-89.
178. Luo, Y., et al., *HIV-host interactome revealed directly from infected cells.* Nat Microbiol, 2016. **1**(7): p. 16068.
179. Yamayoshi, S., et al., *Ebola virus matrix protein VP40 uses the COPII transport system for its intracellular transport.* Cell Host Microbe, 2008. **3**(3): p. 168-77.
180. Garcia-Dorival, I., et al., *Elucidation of the Cellular Interactome of Ebola Virus Nucleoprotein and Identification of Therapeutic Targets.* J Proteome Res, 2016. **15**(12): p. 4290-4303.
181. Hernandez Duran, A., et al., *Protein interactions and consensus clustering analysis uncover insights into herpesvirus virion structure and function relationships.* PLoS Biol, 2019. **17**(6): p. e3000316.
182. Mayer, D., et al., *Identification of cellular interaction partners of the influenza virus ribonucleoprotein complex and polymerase complex using proteomic-based approaches.* J Proteome Res, 2007. **6**(2): p. 672-82.
183. Youn, S., et al., *Evidence for a genetic and physical interaction between nonstructural proteins NS1 and NS4B that modulates replication of West Nile virus.* J Virol, 2012. **86**(13): p. 7360-71.
184. Batra, J., et al., *Protein Interaction Mapping Identifies RBBP6 as a Negative Regulator of Ebola Virus Replication.* Cell, 2018. **175**(7): p. 1917-1930.e13.
185. Germain, M.A., et al., *Elucidating novel hepatitis C virus-host interactions using combined mass spectrometry and functional genomics approaches.* Mol Cell Proteomics, 2014. **13**(1): p. 184-203.
186. Jager, S., et al., *Global landscape of HIV-human protein complexes.* Nature, 2011. **481**(7381): p. 365-70.
187. Davis, Z.H., et al., *Global mapping of herpesvirus-host protein complexes reveals a transcription strategy for late genes.* Mol Cell, 2015. **57**(2): p. 349-60.
188. Mellacheruvu, D., et al., *The CRAPome: a contaminant repository for affinity purification-mass spectrometry data.* Nat Methods, 2013. **10**(8): p. 730-6.
189. Huttlin, E.L., et al., *The BioPlex Network: A Systematic Exploration of the Human Interactome.* Cell, 2015. **162**(2): p. 425-440.
190. Fielding, C.A., et al., *Two novel human cytomegalovirus NK cell evasion functions target MICA for lysosomal degradation.* PLoS Pathog, 2014. **10**(5): p. e1004058.
191. Hsu, J.L., et al., *Plasma membrane profiling defines an expanded class of cell surface proteins selectively targeted for degradation by HCMV US2 in cooperation with UL141.* PLoS Pathog, 2015. **11**(4): p. e1004811.
192. van der Wal, F.J., M. Kikkert, and E. Wiertz, *The HCMV gene products US2 and US11 target MHC class I molecules for degradation in the cytosol.* Current Topics in Microbiology and Immunology, 2002. **269**: p. 37-55.
193. Nobre, L.V., et al., *Human cytomegalovirus interactome analysis identifies degradation hubs, domain associations and viral protein functions.* Elife, 2019. **8**.
194. Fielding, C.A., et al., *Control of immune ligands by members of a cytomegalovirus gene expansion suppresses natural killer cell activation.* Elife, 2017. **6**.

195. Sowa, M.E., et al., *Defining the human deubiquitinating enzyme interaction landscape.* Cell, 2009. **138**(2): p. 389-403.
196. McSharry, B.P., et al., *Human telomerase reverse transcriptase-immortalized MRC-5 and HCA2 human fibroblasts are fully permissive for human cytomegalovirus.* J Gen Virol, 2001. **82**(Pt 4): p. 855-63.
197. Lin, K., et al., *A simple and fast secondary structure prediction method using hidden neural networks.* Bioinformatics, 2004. **21**(2): p. 152-159.
198. Tanaka, J., et al., *Enhanced replication of human cytomegalovirus in human fibroblasts treated with dexamethasone.* J Gen Virol, 1984. **65 (Pt 10)**: p. 1759-67.
199. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* J Mol Biol, 2001. **305**(3): p. 567-80.
200. McAlister, G.C., et al., *MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes.* Anal Chem, 2014. **86**(14): p. 7150-8.
201. Haas, W., et al., *Optimization and use of peptide mass measurement accuracy in shotgun proteomics.* Mol Cell Proteomics, 2006. **5**(7): p. 1326-37.
202. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nat Methods, 2007. **4**(3): p. 207-14.
203. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for mass spectrometry-based proteomics.* Methods in Molecular Biology, 2010. **604**: p. 55-71.
204. Wu, R., et al., *Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes.* Mol Cell Proteomics, 2011. **10**(8): p. M111 009654.
205. Kim, W., et al., *Systematic and quantitative assessment of the ubiquitin-modified proteome.* Mol Cell, 2011. **44**(2): p. 325-40.
206. Huttlin, E.L., et al., *A tissue-specific atlas of mouse protein phosphorylation and expression.* Cell, 2010. **143**(7): p. 1174-89.
207. Kall, L., et al., *Semi-supervised learning for peptide identification from shotgun proteomics datasets.* Nat Methods, 2007. **4**(11): p. 923-5.
208. Behrends, C., et al., *Network organization of the human autophagy system.* Nature, 2010. **466**(7302): p. 68-76.
209. Gallegos, L.L., et al., *A protein interaction map for cell-cell adhesion regulators identifies DUSP23 as a novel phosphatase for beta-catenin.* Sci Rep, 2016. **6**: p. 27114.
210. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
211. Shannon, C.E., *A Mathematical Theory of Communication.* Bell System Technical Journal, 1948. **27**(3): p. 379-423.
212. Chatr-Aryamontri, A., et al., *The BioGRID interaction database: 2013 update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D816-23.
213. Orchard, S., et al., *The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases.* Nucleic Acids Res, 2014. **42**(Database issue): p. D358-63.
214. Licata, L., et al., *MINT, the molecular interaction database: 2012 update.* Nucleic Acids Res, 2012. **40**(Database issue): p. D857-61.
215. Calderone, A., L. Licata, and G. Cesareni, *VirusMentha: a new resource for virus-host protein interactions.* Nucleic Acids Res, 2015. **43**(Database issue): p. D588-92.
216. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
217. El-Gebali, S., et al., *The Pfam protein families database in 2019.* Nucleic Acids Res, 2019. **47**(D1): p. D427-D432.
218. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate - a practical and powerful approach to multiple testing.* J. R. Stat. Soc. Ser. B-Methodol., 1995. **57**: p. 289-300.

219. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
220. Gasteiger, E., et al., *ExPASy: The proteomics server for in-depth protein knowledge and analysis.* Nucleic Acids Res, 2003. **31**(13): p. 3784-8.
221. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Mol Syst Biol, 2011. **7**: p. 539.
222. Liu, B., T.W. Hermiston, and M.F. Stinski, *A cis-acting element in the major immediate-early (IE) promoter of human cytomegalovirus is required for negative regulation by IE2.* Journal of virology, 1991. **65**(2): p. 897-903.
223. Cherrington, J.M., E.L. Khoury, and E.S. Mocarski, *Human cytomegalovirus ie2 negatively regulates alpha gene expression via a short target sequence near the transcription start site.* J Virol, 1991. **65**(2): p. 887-96.
224. Kimple, M.E., A.L. Brill, and R.L. Pasker, *Overview of affinity tags for protein purification.* Current protocols in protein science, 2013. **73**: p. 9.9.1-9.9.23.
225. Wilkins, M.R., et al., *Protein identification and analysis tools in the ExPASy server.* Methods Mol Biol, 1999. **112**: p. 531-52.
226. Reitsma, J.M., et al., *Antiviral inhibition targeting the HCMV kinase pUL97 requires pUL27-dependent degradation of Tip60 acetyltransferase and cell-cycle arrest.* Cell Host Microbe, 2011. **9**(2): p. 103-14.
227. Proc, J.L., et al., *A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin.* J Proteome Res, 2010. **9**(10): p. 5422-37.
228. Waas, M., et al., *Combine and conquer: surfactants, solvents, and chaotropes for robust mass spectrometry based analyses of membrane proteins.* Anal Chem, 2014. **86**(3): p. 1551-9.
229. Schindler, P.A., A. Van Dorsselaer, and A.M. Falick, *Analysis of hydrophobic proteins and peptides by electrospray ionization mass spectrometry.* Anal Biochem, 1993. **213**(2): p. 256-63.
230. Jean Beltran, P.M., et al., *Proteomics and integrative omic approaches for understanding host-pathogen interactions and infectious diseases.* Molecular systems biology, 2017. **13**(3): p. 922-922.
231. Gerold, G., J. Bruening, and T. Pietschmann, *Decoding protein networks during virus entry by quantitative proteomics.* Virus research, 2016. **218**: p. 25-39.
232. O'Connell, J.D., et al., *Proteome-Wide Evaluation of Two Common Protein Quantification Methods.* J Proteome Res, 2018. **17**(5): p. 1934-1942.
233. Babu, M., et al., *Interaction landscape of membrane-protein complexes in Saccharomyces cerevisiae.* Nature, 2012. **489**(7417): p. 585-9.
234. Malty, R.H., et al., *A Map of Human Mitochondrial Protein Interactions Linked to Neurodegeneration Reveals New Mechanisms of Redox Homeostasis and NF-kappaB Signaling.* Cell Syst, 2017. **5**(6): p. 564-577.e12.
235. Lai, L. and W.J. Britt, *The interaction between the major capsid protein and the smallest capsid protein of human cytomegalovirus is dependent on two linear sequences in the smallest capsid protein.* J Virol, 2003. **77**(4): p. 2730-5.
236. Huttlin, E.L., et al., *Architecture of the human interactome defines protein communities and disease networks.* Nature, 2017. **545**(7655): p. 505-509.
237. Tullman, J.A., et al., *Recovery of an HMWP/hmwBP (pUL48/pUL47) complex from virions of human cytomegalovirus: subunit interactions, oligomer composition, and deubiquitylase activity.* J Virol, 2014. **88**(15): p. 8256-67.
238. Wu, Y., et al., *Human cytomegalovirus glycoprotein complex gH/gL/gO uses PDGFR-alpha as a key for entry.* PLoS Pathog, 2017. **13**(4): p. e1006281.
239. Vanarsdall, A.L., et al., *Human Cytomegalovirus gH/gL Forms a Stable Complex with the Fusion Protein gB in Virions.* PLoS Pathog, 2016. **12**(4): p. e1005564.

240. Hwang, J.S. and E. Bogner, *ATPase activity of the terminase subunit pUL56 of human cytomegalovirus.* J Biol Chem, 2002. **277**(9): p. 6943-8.
241. Liu, Y., et al., *The tegument protein UL94 of human cytomegalovirus as a binding partner for tegument protein pp28 identified by intracellular imaging.* Virology, 2009. **388**(1): p. 68-77.
242. Samaniego, L.A., M.J. Tevethia, and D.J. Spector, *The human cytomegalovirus 86-kilodalton immediate-early 2 protein: synthesis as a precursor polypeptide and interaction with a 75-kilodalton protein of probable viral origin.* J Virol, 1994. **68**(2): p. 720-9.
243. Spector, D.J. and M.J. Tevethia, *Protein-protein interactions between human cytomegalovirus IE2-580aa and pUL84 in lytically infected cells.* J Virol, 1994. **68**(11): p. 7549-53.
244. McMahon, T.P. and D.G. Anders, *Interactions between human cytomegalovirus helicase-primase proteins.* Virus Res, 2002. **86**(1-2): p. 39-52.
245. Prichard, M.N., et al., *Human cytomegalovirus uracil DNA glycosylase associates with ppUL44 and accelerates the accumulation of viral DNA.* Virol J, 2005. **2**: p. 55.
246. Ranneberg-Nilsen, T., et al., *Characterization of human cytomegalovirus uracil DNA glycosylase (UL114) and its interaction with polymerase processivity factor (UL44).* J Mol Biol, 2008. **381**(2): p. 276-88.
247. Loregian, A., et al., *Residues of human cytomegalovirus DNA polymerase catalytic subunit UL54 that are necessary and sufficient for interaction with the accessory protein UL44.* J Virol, 2004. **78**(1): p. 158-67.
248. Hofmann, H., H. Sindre, and T. Stamminger, *Functional interaction between the pp71 protein of human cytomegalovirus and the PML-interacting protein human Daxx.* J Virol, 2002. **76**(11): p. 5769-83.
249. Scherer, M., et al., *Characterization of Recombinant Human Cytomegaloviruses Encoding IE1 Mutants L174P and 1-382 Reveals that Viral Targeting of PML Bodies Perturbs both Intrinsic and Innate Immune Responses.* J Virol, 2016. **90**(3): p. 1190-205.
250. Stanton, R.J., et al., *HCMV pUL135 remodels the actin cytoskeleton to impair immune recognition of infected cells.* Cell Host Microbe, 2014. **16**(2): p. 201-14.
251. Nemcovicova, I., C.A. Benedict, and D.M. Zajonc, *Structure of human cytomegalovirus UL141 binding to TRAIL-R2 reveals novel, non-canonical death receptor interactions.* PLoS Pathog, 2013. **9**(3): p. e1003224.
252. Cosman, D., et al., *ULBPs, novel MHC class I-related molecules, bind to CMV glycoprotein UL16 and stimulate NK cytotoxicity through the NKG2D receptor.* Immunity, 2001. **14**(2): p. 123-33.
253. Wittenbrink, M., J. Spreu, and A. Steinle, *Differential NKG2D binding to highly related human NKG2D ligands ULBP2 and RAET1G is determined by a single amino acid in the alpha2 domain.* Eur J Immunol, 2009. **39**(6): p. 1642-51.
254. Eagle, R.A., et al., *ULBP6/RAET1L is an additional human NKG2D ligand.* Eur J Immunol, 2009. **39**(11): p. 3207-16.
255. Savaryn, J.P., et al., *Human cytomegalovirus pUL29/28 and pUL38 repression of p53-regulated p21CIP1 and caspase 1 promoters during infection.* Journal of Virology, 2013. **87**(5): p. 2463-74.
256. Guenther, M.G., et al., *A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-repeat protein linked to deafness.* Genes Dev, 2000. **14**(9): p. 1048-57.
257. Li, J., et al., *Both corepressor proteins SMRT and N-CoR exist in large protein complexes containing HDAC3.* Embo j, 2000. **19**(16): p. 4342-50.
258. Yoon, H.G., et al., *N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso.* Mol Cell, 2003. **12**(3): p. 723-34.
259. Hoffmann, A. and D. Spengler, *Chromatin Remodeling Complex NuRD in Neurodevelopment and Neurodevelopmental Disorders.* Front Genet, 2019. **10**: p. 682.

260. Emmett, M.J. and M.A. Lazar, *Integrative regulation of physiology by histone deacetylase 3.* Nat Rev Mol Cell Biol, 2019. **20**(2): p. 102-115.
261. Fu, H., R.R. Subramanian, and S.C. Masters, *14-3-3 proteins: structure, function, and regulation.* Annu Rev Pharmacol Toxicol, 2000. **40**: p. 617-47.
262. Lin, A.E., et al., *A proteomic perspective of inbuilt viral protein regulation: pUL46 tegument protein is targeted for degradation by ICP0 during herpes simplex virus type 1 infection.* Mol Cell Proteomics, 2013. **12**(11): p. 3237-52.
263. Stegen, C., et al., *Analysis of virion-incorporated host proteins required for herpes simplex virus type 1 infection through a RNA interference screen.* PLoS One, 2013. **8**(1): p. e53276.
264. Loret, S., G. Guay, and R. Lippe, *Comprehensive characterization of extracellular herpes simplex virus type 1 virions.* J Virol, 2008. **82**(17): p. 8605-18.
265. Zhu, F.X., et al., *Virion proteins of Kaposi's sarcoma-associated herpesvirus.* J Virol, 2005. **79**(2): p. 800-11.
266. Johannsen, E., et al., *Proteins of purified Epstein-Barr virus.* Proc Natl Acad Sci U S A, 2004. **101**(46): p. 16286-91.
267. Kramer, T., et al., *Proteomic characterization of pseudorabies virus extracellular virions.* J Virol, 2011. **85**(13): p. 6427-41.
268. Omoto, S. and E.S. Mocarski, *Cytomegalovirus UL91 is essential for transcription of viral true late (gamma2) genes.* J Virol, 2013. **87**(15): p. 8651-64.
269. Omoto, S. and E.S. Mocarski, *Transcription of True Late (gamma2) Cytomegalovirus Genes Requires UL92 Function That Is Conserved among Beta- and Gammaherpesviruses.* Journal of Virology, 2014. **88**(1): p. 120-30.
270. Isomura, H., et al., *The human cytomegalovirus gene products essential for late viral gene expression assemble into prereplication complexes before viral DNA replication.* J Virol, 2011. **85**(13): p. 6629-44.
271. Aubry, V., et al., *Epstein-Barr virus late gene transcription depends on the assembly of a virus-specific preinitiation complex.* Journal of virology, 2014. **88**(21): p. 12825-12838.
272. Baek, M.C., et al., *Phosphorylation of the RNA polymerase II carboxyl-terminal domain in human cytomegalovirus-infected cells and in vitro by the viral UL97 protein kinase.* Virology, 2004. **324**(1): p. 184-93.
273. Caposio, P., et al., *Evidence that the human cytomegalovirus 46-kDa UL72 protein is not an active dUTPase but a late protein dispensable for replication in fibroblasts.* Virology, 2004. **325**(2): p. 264-76.
274. Yi, H., et al., *PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay.* Mol Cell, 2018. **70**(6): p. 1081-1088 e5.
275. Lau, N.C., et al., *Human Ccr4-Not complexes contain variable deadenylase subunits.* Biochem J, 2009. **422**(3): p. 443-53.
276. Shirai, Y.T., et al., *Multifunctional roles of the mammalian CCR4-NOT complex in physiological phenomena.* Front Genet, 2014. **5**: p. 286.
277. Chalabi Hagkarim, N., et al., *Degradation of a Novel DNA Damage Response Protein, Tankyrase 1 Binding Protein 1, following Adenovirus Infection.* J Virol, 2018. **92**(12).
278. Dietz, A.N., et al., *A Tyrosine-Based Trafficking Motif of the Tegument Protein pUL71 Is Crucial for Human Cytomegalovirus Secondary Envelopment.* J Virol, 2018. **92**(1).
279. Meissner, C.S., et al., *A leucine zipper motif of a tegument protein triggers final envelopment of human cytomegalovirus.* J Virol, 2012. **86**(6): p. 3370-82.
280. Lian, Q. and B. Sun, *Interferons command Trim22 to fight against viruses.* Cell Mol Immunol, 2017. **14**(9): p. 794-796.
281. Wang, L., et al., *Thick Filament Protein Network, Functions, and Disease Association.* Compr Physiol, 2018. **8**(2): p. 631-709.
282. Han, J., K. Pluhackova, and R.A. Bockmann, *The Multifaceted Role of SNARE Proteins in Membrane Fusion.* Front Physiol, 2017. **8**: p. 5.

283. Collins, B.M., et al., *Molecular architecture and functional model of the endocytic AP2 complex.* Cell, 2002. **109**(4): p. 523-35.
284. Tchasovnikarova, I.A., et al., *GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells.* Science, 2015. **348**(6242): p. 1481-1485.
285. Paulus, C., S. Krauss, and M. Nevels, *A human cytomegalovirus antagonist of type I IFN-dependent signal transducer and activator of transcription signaling.* Proc Natl Acad Sci U S A, 2006. **103**(10): p. 3840-5.
286. Koshizuka, T., K. Tanaka, and T. Suzutani, *Degradation of host ubiquitin E3 ligase Itch by human cytomegalovirus UL42.* J Gen Virol, 2016. **97**(1): p. 196-208.
287. Wiertz, E.J., et al., *The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol.* Cell, 1996. **84**(5): p. 769-79.
288. Park, B., et al., *The HCMV membrane glycoprotein US10 selectively targets HLA-G for degradation.* Journal of Experimental Medicine, 2010. **207**(9): p. 2033-41.
289. Finn, R.D., et al., *Pfam: the protein families database.* Nucleic Acids Res, 2014. **42**(Database issue): p. D222-30.
290. Smith, W., et al., *Human cytomegalovirus glycoprotein UL141 targets the TRAIL death receptors to thwart host innate antiviral defenses.* Cell Host Microbe, 2013. **13**(3): p. 324-35.
291. Kurochkina, N. and U. Guha, *SH3 domains: modules of protein-protein interactions.* Biophys Rev, 2013. **5**(1): p. 29-39.
292. Keyvani Chahi, A., C.E. Martin, and N. Jones, *Nephrin Suppresses Hippo Signaling through the Adaptor Proteins Nck and WTIP.* J Biol Chem, 2016. **291**(24): p. 12799-808.
293. Ngoenkam, J., et al., *Non-overlapping functions of Nck1 and Nck2 adaptor proteins in T cell activation.* Cell Commun Signal, 2014. **12**: p. 21.
294. Wiertz, E.J., et al., *Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction.* Nature, 1996. **384**(6608): p. 432-8.
295. Jones, T.R. and L. Sun, *Human cytomegalovirus US2 destabilizes major histocompatibility complex class I heavy chains.* J Virol, 1997. **71**(4): p. 2970-9.
296. Stagg, H.R., et al., *The TRC8 E3 ligase ubiquitinates MHC class I molecules before dislocation from the ER.* J Cell Biol, 2009. **186**(5): p. 685-92.
297. Lilley, B.N., D. Tortorella, and H.L. Ploegh, *Dislocation of a type I membrane protein requires interactions between membrane-spanning segments within the lipid bilayer.* Mol Biol Cell, 2003. **14**(9): p. 3690-8.
298. Lilley, B.N. and H.L. Ploegh, *A membrane protein required for dislocation of misfolded proteins from the ER.* Nature, 2004. **429**(6994): p. 834-40.
299. Brechtel, T.M., E.S. Mocarski, and R. Tandon, *Highly acidic C-terminal region of cytomegalovirus pUL96 determines its functions during virus maturation independently of a direct pp150 interaction.* J Virol, 2014. **88**(8): p. 4493-503.
300. Penfold, M.E., et al., *Cytomegalovirus encodes a potent alpha chemokine.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(17): p. 9839-44.
301. Luttichau, H.R., *The cytomegalovirus UL146 gene product vCXCL1 targets both CXCR1 and CXCR2 as an agonist.* J Biol Chem, 2010. **285**(12): p. 9137-46.
302. Wang, G., et al., *Human cytomegalovirus RL13 protein interacts with host NUDT14 protein affecting viral DNA replication.* Mol Med Rep, 2016. **13**(3): p. 2167-74.
303. Bennett, N.J., et al., *Intracellular sequestration of the NKG2D ligand ULBP3 by human cytomegalovirus.* Journal of Immunology, 2010. **185**(2): p. 1093-102.
304. Liu, C., et al., *Interaction between the human cytomegalovirusencoded UL142 and cellular Snapin proteins.* Mol Med Rep, 2015. **11**(2): p. 1069-72.
305. Medvar, B., et al., *Comprehensive database of human E3 ubiquitin ligases: application to aquaporin-2 regulation.* Physiol Genomics, 2016. **48**(7): p. 502-12.
306. Higa, L.A., et al., *CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation.* Nat Cell Biol, 2006. **8**(11): p. 1277-83.

307. Francis, O., F. Han, and J.C. Adams, *Molecular phylogeny of a RING E3 ubiquitin ligase, conserved in eukaryotic cells and dominated by homologous components, the muskelin/RanBPM/CTLH complex.* PLoS One, 2013. **8**(10): p. e75217.
308. Salemi, L.M., et al., *Cell signalling pathway regulation by RanBPM: molecular insights and disease implications.* Open Biol, 2017. **7**(6).
309. Jelcic, I., et al., *The polymorphic HCMV glycoprotein UL20 is targeted for lysosomal degradation by multiple cytoplasmic dileucine motifs.* Traffic, 2011. **12**(10): p. 1444-56.
310. Nightingale, K., et al., *Human Cytomegalovirus Protein RL1 degrades the antiviral factor SLFN11 Through Recruitment of the CRL4 E3 Ubiquitin Ligase Complex.* In preparation.
311. Fletcher-Etherington, A., et al., *Human cytomegalovirus protein pUL36: a dual cell death pathway inhibitor.* PNAS, In preparation.
312. Buvall, L., et al., *Proteasomal degradation of Nck1 but not Nck2 regulates RhoA activation and actin dynamics.* Nat Commun, 2013. **4**: p. 2863.
313. Hansen, S.G., et al., *Complete sequence and genomic analysis of rhesus cytomegalovirus.* J Virol, 2003. **77**(12): p. 6620-36.
314. Marsh, A.K., et al., *Genomic sequencing and characterization of cynomolgus macaque cytomegalovirus.* J Virol, 2011. **85**(24): p. 12995-3009.
315. Fabregat, A., et al., *The Reactome Pathway Knowledgebase.* Nucleic Acids Res, 2018. **46**(D1): p. D649-D655.
316. Sandri-Goldin, R.M., *ICP27 mediates HSV RNA export by shuttling through a leucine-rich nuclear export signal and binding viral intronless RNAs through an RGG motif.* Genes Dev, 1998. **12**(6): p. 868-79.
317. Lischka, P., et al., *The UL69 transactivator protein of human cytomegalovirus interacts with DEXD/H-Box RNA helicase UAP56 to promote cytoplasmic accumulation of unspliced RNA.* Mol Cell Biol, 2006. **26**(5): p. 1631-43.
318. Aksu, M., et al., *Xpo7 is a broad-spectrum exportin and a nuclear import receptor.* J Cell Biol, 2018. **217**(7): p. 2329-2340.
319. Balazs, Z., et al., *Long-Read Sequencing of Human Cytomegalovirus Transcriptome Reveals RNA Isoforms Carrying Distinct Coding Potentials.* Sci Rep, 2017. **7**(1): p. 15989.
320. Rawlinson, W.D. and B.G. Barrell, *Spliced transcripts of human cytomegalovirus.* J Virol, 1993. **67**(9): p. 5502-13.
321. Cen, S., et al., *Small molecular compounds inhibit HIV-1 replication through specifically stabilizing APOBEC3G.* Journal of Biological Chemistry, 2010. **285**(22): p. 16546-52.
322. Nathans, R., et al., *Small-molecule inhibition of HIV-1 Vif.* Nature Biotechnology, 2008. **26**(10): p. 1187-92.
323. Pery, E., et al., *Identification of a novel HIV-1 inhibitor targeting Vif-dependent degradation of human APOBEC3G protein.* J Biol Chem, 2015. **290**(16): p. 10504-17.
324. Maryati, M., et al., *A fluorescence-based assay suitable for quantitative analysis of deadenylase enzyme activity.* Nucleic Acids Res, 2014. **42**(5): p. e30.
325. Madeira, F., et al., *The EMBL-EBI search and sequence analysis tools APIs in 2019.* Nucleic acids research, 2019. **47**(W1): p. W636-W641.

Appendix A | UL48 secondary structure prediction

Secondary structure prediction was generated using the YASPIN Secondary Structure Prediction tool from the Centre for Integrative Bioinformatics VU (Vrije Universiteit Amsterdam) [197].

```
Ruler :      .........10........20........30........40........50........60
Sequence :   MKVTQASCHQGDIARFGARAGNQCVCNGIMFLHALHLGGTSAVLQTEALDAIMEEGARLD
Prediction: -EEEEEE--------------EEEEEHHHHHHHHHHH--------HHHHHHHHHHHHHHH
Ruler :      .........70........80........90........100.......110.......120
Sequence :   ARLERELQKKLPAGGRLPVYRLGDEVPRRLESRFGRTVHALSRPFNGTTETCDLDGYMCP
Prediction: HHHHH--------------EEE-----EEEEE------------------------EEEE
Ruler :      .......130.......140.......150.......160.......170.......180
Sequence :   GIFDFLRYAHAKPRPTYVLVTVNSLARAVVFTEDHMLVFDPHSSAECHNAAVYHCEGLHQ
Prediction: -HHHHHH--------EEEEEEE----EEEEEE--EEEEE----------EEEEEEE-HHH
Ruler :      .......190.......200.......210.......220.......230.......240
Sequence :   VLMVLTGFGVQLSPAFYYEALFLYMLDVATVPEAEIAARLVSTYRDRDIDLTGVVRESAD
Prediction: HHHHH-----------EEEEEEEEEEE-----HHHHHHHHHHH-----------------
Ruler :      .......250.......260.......270.......280.......290.......300
Sequence :   TAATTTTAAPSLPPLPDPIVDPGCPPGVAPSIPVYDPSSSPKKTPEKRRKDLSGSKHGGK
Prediction: ------------------------------------------------------------
Ruler :      .......310.......320.......330.......340.......350.......360
Sequence :   KKPPSTTSKTLATASSSPSAIAAASSSSAVPPSYSCGEGALPALGRYQQLVDEVEQELKA
Prediction: --------------------HHH--------------HHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .......370.......380.......390.......400.......410.......420
Sequence :   LTLPPLPANTSAWTLHAAGTESGANAATATAPSFDEAFLTDRLQQLIIHAVNQRSCLRRP
Prediction: -------------EEEE-----------------HHHHHHHHHHHHHHHHH-EEEE---
Ruler :      .......430.......440.......450.......460.......470.......480
Sequence :   CGPQSAAQQAVRAYLGLSKKLDAFLLNWLHHGLDLQRMHDYLSHKTTKGTYSTLDRALLE
Prediction: ---HHHHHHHHHHHHHHHHHHHHHHHHHHH-HHHHHHHHHH------HHHHHHHHHHHHH
Ruler :      .......490.......500.......510.......520.......530.......540
Sequence :   KMQVVFDPYGRQHGPALIAWVEEMLRYVESKPTNELSQRLQRFVTKRPMPVSDSFVCLRP
Prediction: HHHHHH--HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-----HHHHHHH-H
Ruler :      .......550.......560.......570.......580.......590.......600
Sequence :   VDFQRLTQVIEQRRRVLQRQREEYHGVYEHLAGLITSIDIHDLDASDLNRREILKALQPL
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .......610.......620.......630.......640.......650.......660
Sequence :   DDNAKQELFRLGNAKMLELQMDLDRLSTQLLTRVHNHILNGFLPVEDLKQMERVVEQVLR
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .......670.......680.......690.......700.......710.......720
Sequence :   LFYDLRDLKLCDGSYEEGFVVIREQLSYLMTGTVRDNVPLLQEILQLRHAYQQATQQNEG
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .......730.......740.......750.......760.......770.......780
Sequence :   RLTQIHDLLHVIETLVRDPGSRGSALTLALVQEQLAQLEALGGLQLPEVQQRLQNAQLAL
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .......790.......800.......810.......820.......830.......840
Sequence :   SRLYEEEEETQRFLDGLSYDDPPTEQTIKRHPQLREMLRRDEQTRLRLINAVLSMFHTLV
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .......850.......860.......870.......880.......890.......900
Sequence :   MRLARDESPRPTFFDAVSLLLQQLPPDSHEREDLRAANATYAQMVKKLEQIEKAGTGASE
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
```

```
Ruler :      .......910.......920.......930.......940.......950.......960
Sequence :   KRFQALRELVYFFRNHEYFFQHMVGRLGVGPQVTELYERYQHEMEEQHLERLEREWQEEA
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .......970.......980.......990.......1000.........1010.........1020
Sequence :   GKLTVTSVEDVQRVLARAPSHRVMHQMQQTLTTKMQDFLDKEKRKQEEQQRQLLDGYQKK
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .........1030.........1040.........1050.........1060.........1070.........1080
Sequence :   VQQDLQRVVDAVKGEMLSTIPHQPLEATLELLLGLDQRAQPLLDKFNQDLLSALQQLSKK
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .........1090.........1100.........1110.........1120.........1130.........1140
Sequence :   LDGRINECLHGVLTGDVERRCHPHREAAMQTQASLNHLDQILGPQLLIHETQQALQHAVH
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .........1150.........1160.........1170.........1180.........1190.........1200
Sequence :   QAQFIEKCQQGDPTTAITGSEFESDFARYRSSQQKMEGQLQETRQQMTETSERLDRSLRQ
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .........1210.........1220.........1230.........1240.........1250.........1260
Sequence :   DPGSSSVTRVPEKPFKGQELAGRITPPPADFQRPVFKTLLDQQADAARKALSDEADLLNQ
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .........1270.........1280.........1290.........1300.........1310.........1320
Sequence :   KVQTQLRQRDEQLSTAQNLWTDLVTRHKMSGGLDVTTPDAKALMEKPLETLRELLGKATQ
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .........1330.........1340.........1350.........1360.........1370.........1380
Sequence :   QLPYLSAERTVRWMLAFLEEALAQITADPTHPHHGSRTHYRNLQQQAVESAVTLAHQIEQ
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .........1390.........1400.........1410.........1420.........1430.........1440
Sequence :   NAACENFIAQHQEATANGASTPRVDMVQAVEAVWQRLEPGRVAGGAARHQKVQELLQRLG
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .........1450.........1460.........1470.........1480.........1490.........1500
Sequence :   QTLGDLELQETLATEYFALLHGIQTFSYGLDFRSQLEKIRDLRTRFAELAKRCGTRLSNE
Prediction: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Ruler :      .........1510.........1520.........1530.........1540.........1550.........1560
Sequence :   GALPNPRKPQATTSLGAFTRGLNALERHVQLGHQYLLNKLNGSSLVYRLEDIPSVLPPTH
Prediction: ---H-----HHHHHHHHHHHHHHHHHHHHHHHHHHHH-----EEE------------
Ruler :      .........1570.........1580.........1590.........1600.........1610.........1620
Sequence :   ETDPALIMRDRLRRLCFARHHDTFLEVVDVFGMRQIVTQAGEPIYLVTDYGNVAFKYLAL
Prediction: ----HHHHHHHHHHHH------EEEE---------------EEEEE----EEEEEEE-
Ruler :      .........1630.........1640.........1650.........1660.........1670.........1680
Sequence :   RDDGRPLAWRRRCSGGGLKNVVTTRYKAITVAVAVCQTLRTFWPQISQYDLRPYLTQHQS
Prediction: --------------------EEEEEEEEEEEEHHHHHHHHHHHHHHHH---HHHHH----
Ruler :      .........1690.........1700.........1710.........1720.........1730.........1740
Sequence :   HTHPAETHTLHNLKLFCYLVSTAWHQRIDTQQELTAADRVGSGEGGDVGEQRPGRGTVLR
Prediction: ---------EEEEEEEEEEEHHHHHHHH----------------------------EEE
Ruler :      .........1750.........1760.........1770.........1780.........1790.........1800
Sequence :   LSLQEFCVLIAALYPEYIYTVLKYPVQMSLPSLTAHLHQDVIHAVVNNTHKMPPDHLPEQ
Prediction: E-HHHHHHHHHH--HHHHHHHH---HHHHHHHHHHHHH---HHHHH----------HHH
Ruler :      .........1810.........1820.........1830.........1840.........1850.........1860
Sequence :   VKAFCITPTQWPAMQLNKLFWENKLVQQLCQVGPQKSTPSLGKLWLYAMATLVFPQDMLQ
Prediction: HHHHH------------HHHH--HHHHHHHHH-----HHHHHHHHHHHHH-EEE-HHHHH
Ruler :      .........1870.........1880.........1890.........1900.........1910.........1920
Sequence :   CLWLELKPQYAETYASVSELVQTLFQIFTQQCEMVTEGYTQPQLPTGEPVLQMIRVRRQD
Prediction: HHHHHH--------HHHHHHHHHHHHHH-----EEE-----------EEEEEEE-----
Ruler :      .........1930.........1940.........1950.........1960.........1970.........1980
Sequence :   TTTTDTNTTTEPGLLDVFIQTETALDYALGSWLFGIPVCLGVHVADLLKGQRVLVARHLE
Prediction: --------------HHHHH--HHHHHHHH--EEEEEEEEEEEEE------EEEEEEHHH
```

(Continued)

```
Ruler :      .........1990.........2000.........2010.........2020.........2030.........2040
Sequence :   YTSRDRDFLRIQRSRDLNLSQLLQDTWTETPLEHCWLQAQIRRLRDYLRFPTRLEFIPLV
Prediction: -------HHHHHH----------------HHHHHHHHHHHHHHHH----------EEE
Ruler :      .........2050.........2060.........2070.........2080.........2090.........2100
Sequence :   IYNAQDHTVVRVLRPPSTFEQDHSRLVLDEAFPTFPLYDQDDNTSADNVAASGAAPTPPV
Prediction: EEE----EEEEEEE----------EEEEE-------EEE---------------------
Ruler :      .........2110.........2120.........2130.........2140.........2150.........2160
Sequence :   PFNRVPVNIQFLRENPPPIARVQQPPRRHRHRAAAAADDDGQIDHAQDDTSRTADSALVS
Prediction: ------------------------------------------------HHHH----H
Ruler :      .........2170.........2180.........2190.........2200.........2210.........2220
Sequence :   TAFGGSVFQENRLGETPLCRDELVAVAPGAASTSFASPPITVLTQNVLSALEILRLVRLD
Prediction: HHH-------------------EEEEE-------------HHHHHHHHHHHHHHHHHHH
Ruler :      .........2230.........2240.
Sequence :   LRQLAQSVQDTIQHMRFLYLL
Prediction: HHHHHHHHHHHHHHHHHHH--
```

Key: H=helix; E=strand; -=coil; highlighted in yellow=amino acid 1504, where the sequence was divided into two halves (1-1504 and 1505-2241).

Appendix B | Codon Optimised Sequences

Codon optimisation of UL74, US14 and US17 was performed by Dr. James Davies and Dr. Sepehr Seirafian (School of Medicine, Cardiff University). Sequence alignments were generated using the EMBOSS Stretcher bioinformatics tool [325].

B.1 Sequence alignment for canonical and codon optimised UL74

```
UL74_Canon          1 ATGGGGAAAAAAGAGATGATAATGGTGAAAGGCATTCCTAAAATTATGCT     50
                      |||||.||.||||||.|||||.|||||.||.|||||.||.||.||.|||||
UL74_OPTIM          1 ATGGGCAAGAAAGAAATGATCATGGTCAAGGGCATCCCCAAGATCATGCT     50

UL74_Canon         51 CCTGATCTCTATAACGTTCTTGCTCCTTTCCCTCATAAATTGTAATGTAT    100
                      .|||||....||.||.||..||||.||.|||||.||.||.||.||.||..
UL74_OPTIM         51 GCTGATTAGCATCACCTTTCTGCTGCTGTCCCTGATCAACTGCAACGTGC    100

UL74_Canon        101 TGGTAAACTCCAGAGGAACAAGACGTTCCTGGCCGTATACCGTGCTATCT    150
                      ||||.|||..|.|.||.||.|||.|.||||||||.||.||||||||.||.
UL74_OPTIM        101 TGGTCAACAGCCGGGGCACCAGAAGATCCTGGCCCTACACCGTGCTGTCC    150

UL74_Canon        151 TATCGAGGTAAAGAGATTCTGAAGAAACAGAAGGAAGATATCTTAAAACG    200
                      ||.||.||.||||||||.||||||||.|||||.||.||.|||.|.||.||
UL74_OPTIM        151 TACCGGGGCAAAGAGATCCTGAAGAAGCAGAAAGAGGACATCCTGAAGCG    200

UL74_Canon        201 ATTGATGTCTAC-ATCATCTGACGGATACCGGTTTTTAATGTACCCCAGT    249
                      ..||||||...|| |.||.| |||||.||||||||..|.||||||||||.
UL74_OPTIM        201 GCTGATGAGCACCAGCAGC-GACGGCTACCGGTTCCTGATGTACCCCAGC    249

UL74_Canon        250 CAGCAAAAATTTCATGCCATCGTTATTAGCATGGATAAATTTCCTCAAGA    299
                      |||||.|||||.||.||||||||.||.||||||||.||.||.||.||.||
UL74_OPTIM        250 CAGCAGAAATTCCACGCCATCGTGATCAGCATGGACAAGTTCCCCCAGGA    299

UL74_Canon        300 CTACATTTTAGCGGGTCCCATTAGAAATGATAGCATTACCCATATGTGGT    349
                      ||||||..|.||.||.||||||..|.||.||.||||||.|||||.|||||||
UL74_OPTIM        300 CTACATCCTGGCCGGACCCATCCGGAACGACAGCATCACCCACATGTGGT    349

UL74_Canon        350 TTGACTTTTACAGTACTCAACTCCGAAAACCAGCCAAGTACGTATATTCC    399
                      |.|||||.|||||.||.||.||.||.||.||.||||||.|||||.||...|
UL74_OPTIM        350 TCGACTTCTACAGCACCCAGCTGCGGAAGCCCGCCAAATACGTGTACAGC    399

UL74_Canon        400 GAATATAATCACACGGCCCACAAAATAACGTTACGACCCCCACCTTGCGG    449
                      ||.||.||.||||||.||||||||.||.|||..|..|.||.||.||||||.||
UL74_OPTIM        400 GAGTACAACCACACCGCCCACAAGATCACCCTGAGGCCTCCCCCTTGTGG    449

UL74_Canon        450 CACAGTGCCTTCTATGAACTGCCTATCCGAAATGTTAAATGTTTCCAAAC    499
                      |||.|||||....||||||||||...|||.|||.|.||.||.||||||.|
UL74_OPTIM        450 CACCGTGCCCAGCATGAACTGCCTGAGCGAGATGCTGAACGTGTCCAAGC    499

UL74_Canon        500 GCAATGATACCGGCGAAAAAGGTTGCGGTAATTTCACCACGTTTAATCCT    549
                      |.||.||.||||||||.||.||.|||||.||.||||||||.||.||.||.
UL74_OPTIM        500 GGAACGACACCGGCGAGAAGGGCTGCGGCAACTTCACCACCTTCAACCCC    549

UL74_Canon        550 ATGTTTTTCAACGTACCACGTTGGAACACAAAACTGTACATAGGTTCCAA    599
                      |||||.||||||||.||.||.||||||||.||.||||||||.||...|||
UL74_OPTIM        550 ATGTTCTTCAACGTGCCCCGGTGGAACACCAAGCTGTACATCGGCAGCAA    599

UL74_Canon        600 CAAAGTCAACGTGGATAGTCAGACAATCTACTTTTTGGGCCTAACCGCCC    649
                      ||||||.||||||||.||.|||||.|||||||||.|||||||.|||||||
UL74_OPTIM        600 CAAAGTGAACGTGGACAGCCAGACCATCTACTTTCTGGGCCTGACCGCCC    649

UL74_Canon        650 TACTTTTACGATACGCGCAACGTAACTGCACTCGCAGTTTCTACCTGGTT    699
                      |.||..|..||||||||.||.||.||||||||.||....||||||||||.
UL74_OPTIM        650 TGCTGCTGAGATACGCCCAGCGGAACTGCACCCGGTCCTTCTACCTGGTC    699
```

```
UL74_Canon       700 AACGCCATGAGCCGAAATTTATTCCGCGTTCCCAAGTATATTAACGGCAC    749
                     ||||||||||||||.||..|.|||||.||.||||||||.||.||||||||
UL74_OPTIM       700 AACGCCATGAGCCGGAACCTGTTCCGGGTGCCCAAGTACATCAACGGCAC    749

UL74_Canon       750 CAAGTTGAAAAACACTATGCGAAAACTCAAACGTAAACAAGCGCTTGTCA    799
                     ||||.||||.|||||.|||||.||.||.||.||.||.||.||.||.||||
UL74_OPTIM       750 CAAGCTGAAGAACACCATGCGGAAGCTGAAGCGGAAGCAGGCCCTGGTCA    799

UL74_Canon       800 AAGAACAACCACAAAAAAAGAATAAGAAATCTCAAAGTACTACTACGCCA    849
                     ||||.||.||.||.||.|||||.|||||.||.||.||.||.||.||.||.
UL74_OPTIM       800 AAGAGCAGCCCCAGAAGAAGAACAAGAAGTCCCAGAGCACCACCACCCCC    849

UL74_Canon       850 TATCTTTCCTATACAACGTCTACCGCTTTCAACGTCACCACTAACGTGAC    899
                     ||.||...|||.||.||.||.|||||.||||||||.|||||.||||||||
UL74_OPTIM       850 TACCTGAGCTACACCACCTCCACCGCCTTCAACGTGACCACCAACGTGAC    899

UL74_Canon       900 TTATAGTGCTACCGCTGCTGTAACGCGGGTTGCCACATCTACGACAGGTT    949
                     .||.||.||.||.||.||.||.||..|.||.|||||||...||.||.||.|
UL74_OPTIM       900 CTACAGCGCCACAGCCGCCGTGACCAGAGTGGCCACAAGCACCACCGGCT    949

UL74_Canon       950 ATCGTCCTGATAGTAACTTTATGAAATCCATTATGGCCACGCAGTTAAGA    999
                     |.||.||.||.||.||||||||||.|||||.||||||||.|||.|.|||
UL74_OPTIM       950 ACCGGCCCGACAGCAACTTTATGAAGTCCATCATGGCCACCCAGCTGAGA    999

UL74_Canon      1000 GATCTCGCAACATGGGTATATACTACTCTGCGGTATCGGAATGAACCCTT   1049
                     |||||.||.||.|||||.||.||.||.||||||||..|.||.||.|||||
UL74_OPTIM      1000 GATCTGGCCACCTGGGTGTACACCACCCTGCGGTACAGAAACGAGCCCTT   1049

UL74_Canon      1050 TTGTAAACCAGACCGTAACCGTACCGCCGTGTCAGAATTTATGAAAAACA   1099
                     .||.||.||.|||||.|||.|.|||||||||...||.||.|||||.||.|
UL74_OPTIM      1050 CTGCAAGCCCGACCGGAACAGAACCGCCGTGAGCGAGTTCATGAAGAATA   1099

UL74_Canon      1100 CGCACGTACTGATTCGTAACGAAACGCCGTACACTATTTATGGCACTCTT   1149
                     |.|||||.||||||.|.|||||.||.||.|||||.||.||.|||||.||.
UL74_OPTIM      1100 CCCACGTGCTGATCAGAAACGAGACACCCTACACCATCTACGGCACCCTG   1149

UL74_Canon      1150 GACATGAGCTCCTTATATTACAACGAAACCATGTCCGTGGAAAACGAAAC   1199
                     |||||||||..|.|.||.||||||||.||.||||..||||||.||||||||
UL74_OPTIM      1150 GACATGAGCAGCCTGTACTACAACGAGACAATGAGCGTGGAGAACGAGAC   1199

UL74_Canon      1200 GGCTTCCGATAATAACGAAACTACACCTACGTCACCATCGACGAGGTTTC   1249
                     .||...|||.||.||||||||.||.||.||.||.||....||..|||||.|
UL74_OPTIM      1200 AGCCAGCGACAACAACGAAACCACCCCCACCTCCCCCAGCACCCGGTTCC   1249

UL74_Canon      1250 AGAGAACGTTCATAGATCCCCTATGGGACTATCTAGACTCGCTGCTGTTT   1299
                     ||.|.||.|||||.||.|||||.||||||||.||.|||...|||||||||.
UL74_OPTIM      1250 AGCGGACCTTCATCGACCCCCTGTGGGACTACCTGGACAGCCTGCTGTTC   1299

UL74_Canon      1300 CTAGATAAAATCCGTAACTTTAGCCTCCAGTTACCCGCGTATGGAAATCT   1349
                     ||.||.||.|||||.|||||.|||||.|||.|.|||||.||.||.|||||
UL74_OPTIM      1300 CTGGACAAGATCCGGAACTTCAGCCTGCAGCTGCCCGCCTACGGCAATCT   1349

UL74_Canon      1350 TACCCCGCCGGAACACCGCCGGGCTGCAAATCTATCCACCCTCAATAGCC   1399
                     .|||||.||.||.|||.|..||||.||.||.||...||||||.||.||||
UL74_OPTIM      1350 GACCCCCCCTGAGCACAGAAGGGCCGCCAACCTGAGCACCCTGAACAGCC   1399

UL74_Canon      1400 TTTGGTGGTGGTCGCAG   1416
                     |.||||||||||...|||
UL74_OPTIM      1400 TGTGGTGGTGGAGCCAG   1416
```

B.2 Sequence alignment for canonical and codon optimised US14

```
US14_Canon          1 ATGGAGACAGTTTCCACGCAGCGGGAAACCGCGTCGTCGGAAACGGAGCG     50
                      ||||||||.||...|||.|||.|.||||||||.||.||.||.||.||..|
US14_OPTIM          1 ATGGAGACCGTCAGCACCCAGAGAGAAACCGCCTCATCCGAGACAGAAAG     50

US14_Canon         51 TACCAGGGAAGCCGCGTCCGCGGAAACGACGGACGCTACTTTCCGATCCT    100
                      .||.||.||||||||.||.||.||.||.||.|||||.||.||||||||..
US14_OPTIM         51 GACAAGAGAAGCCGCCTCAGCCGAGACTACCGACGCCACATTCCGATCAC    100

US14_Canon        101 TGGAAGAGGGAAGCACGATCTCCTCCCGGTATTCGGAAACAGCGTCCACA    150
                      ||||.||.||.|||||.|||..|||||||||.||.||.||.||....||.
US14_OPTIM        101 TGGAGGAAGGCAGCACCATCAGCTCCCGGTACTCTGAGACTGCAAGTACC    150

US14_Canon        151 GCTTCCGAAGATGCGGTGTGTTGGCTACGTCGGACCGCGATCGTCATGCG    200
                      ||.||.||||||||.|||||.|||||.||..|.||.||.||||||||||
US14_OPTIM        151 GCCTCAGAAGATGCCGTGTGCTGGCTGCGGAGAACAGCAATCGTCATGCG    200

US14_Canon        201 TGTCTACGGATTGCTGACGCTGGAAACCGCTTTCAGCGTACTCATTAGCG    250
                      .||.||.||..||||||||.|||||.||.||.||||||||.||.||.||||
US14_OPTIM        201 GGTGTATGGGCTGCTGACACTGGAGACTGCCTTCAGCGTGCTGATCAGCG    250

US14_Canon        251 CGCTGGTTTGGCTGGGTTACCCCTCACTGGGCTACGAGTGCCGCGACGAT    300
                      |.|||||.||||||||.|||||||...|||||.||.||.|||.|.||||||
US14_OPTIM        251 CCCTGGTGTGGCTGGGCTACCCCAGTCTGGGGTATGAATGCAGGGACGAT    300

US14_Canon        301 CCCTCGCCGCTATTGCTTAGCTGCACACCGGTGTTGGTTCTGGGCGCGCT    350
                      |||||.||.||..||||....||.||.||.|||.||||.|||||.||.||
US14_OPTIM        301 CCCTCCCCTCTGCTGCTGTCTTGTACTCCAGTGCTGGTCCTGGGGGCCCT    350

US14_Canon        351 GGAGCTCACCGACCACAGACACCCCAGCAACGGCCTGGTGTTTGCACTGT    400
                      ||||||.|||||||||.|.||.||||||||||.|||||.||.||.||||
US14_OPTIM        351 GGAGCTGACCGACCACCGGCATCCCAGCAACGGACTGGTCTTCGCTCTGT    400

US14_Canon        401 ATGTGGCGCTTCTCTCGTTCACCACCGCCGGGCTCAACCTGTGCGCCACA    450
                      |.|||||.||.||.||.||.|||||.||.||.||.||.||||||||||.
US14_OPTIM        401 ACGTGGCACTGCTGTCCTTTACCACAGCTGGACTGAATCTGTGCGCCACC    450

US14_Canon        451 GCGCCCATCGGCGTTTCCAGCCTCATCCTAACGTGGACGTTGTTCGTGGC    500
                      ||.||.||||||||.||.||.||.||.||.||.|||||..|||||||.||
US14_OPTIM        451 GCTCCTATCGGCGTGTCTAGTCTGATTCTGACATGGACTCTGTTCGTCGC    500

US14_Canon        501 CTGCAACGGCGTGGCTTGGGAACACCGCCTCAGCTCTGTGTGGCGTGACG    550
                      |||.|||||.|||||||||||||||.|.||.......||||||||.||||
US14_OPTIM        501 CTGTAACGGAGTGGCTTGGGAACACAGGCTGTCAAGCGTGTGGCGCGACG    550

US14_Canon        551 CGCTTTTCACCTCCACACTTTTGACGGTGATGGTCAGCGTCCTGGCTTCT    600
                      |.||.||.||||..|||||..||||.|||||||||...||.||||||...|
US14_OPTIM        551 CTCTGTTTACCAGCACACTGCTGACCGTGATGGTCTCTGTGCTGGCAAGT    600

US14_Canon        601 ACTTACACCTGGTTGCATAAGACTCTGCTGTGTCTCTACACCGTGTTCGT    650
                      ||||||||||||.||||||||||.||||||||.||.||.||.||.|||||
US14_OPTIM        601 ACTTACACCTGGCTGCATAAGACCCTGCTGTGCCTGTATACAGTCTTCGT    650

US14_Canon        651 GGGCTGCATCCTGGCCGTCCTTTTTCCAAGACGTGCGCTACATCGCCACCA    700
                      ||||||.|||||||||||||||.||.||.||.|||.|.|||||.||.||.|
US14_OPTIM        651 GGGCTGTATCCTGGCCGTCCTGTTTCAGGATGTGAGATACATTGCTACAA    700

US14_Canon        701 AAATGCCCGTGAGCCACATCATCCGCAGCTCGCTGGTACTTTACGCCACC    750
                      ||||||||||||||||||||||..|...|||.|||||.||.||||||||.
US14_OPTIM        701 AAATGCCCGTGAGCCACATCATTAGGTCCTCTCTGGTGCTGTACGCCACA    750

US14_Canon        751 GAGACACTCATCTATCACACCACACTCCTGATGCTCACGCCCGTGGTGTG    800
                      |||||.||.||||||.||.||.||.||.||||||||.||.||.||||||.||
US14_OPTIM        751 GAGACTCTGATCTACCATACTACCCTGCTGATGCTGACACCAGTGGTCTG    800
```

```
US14_Canon       801 GTCGGCACGCTGGGACCAAATGTTCAGCTACCTGGCCAAGCTGGGCACCT    850
                     |...|||||||||||||||.||||||...||.|||||||||||||||.||.|
US14_OPTIM       801 GAGTGCACGCTGGGACCAGATGTTCTCATATCTGGCCAAGCTGGGGACTT    850

US14_Canon       851 ATCATCATTACCGGGTTGACAACGGCACGCTCAGCGTCATCCTCAACAGC    900
                     |.||.|||||.||.||.||.||||||||.||....|||||.||.|||..|
US14_OPTIM       851 ACCACCATTATCGAGTGGATAACGGCACTCTGTCAGTCATTCTGAACTCC    900

US14_Canon       901 ACCACCGCGACGTTCCAGAGCAGGGTTGCT    930
                     ||.|||||.||.||||||....||||.||.
US14_OPTIM       901 ACAACCGCCACTTTCCAGTCACGGGTCGCA    930
```

B.3 Sequence alignment for canonical and codon optimised US17

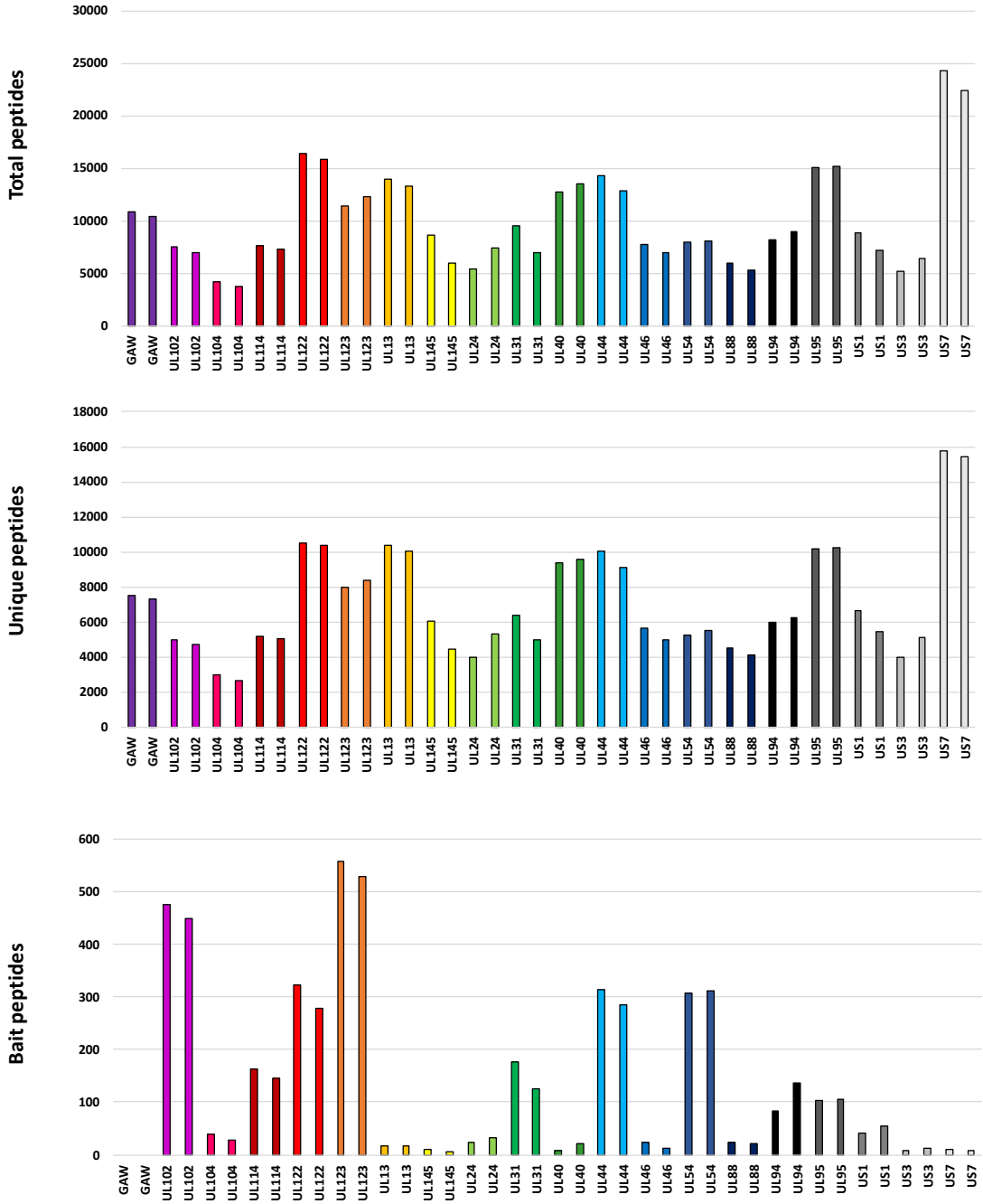
```
US17_Canon         1 ATGTCTCCGAACTCAGAGGCCACCGGGACGGCCTGGGCGCCCCCACCCCC     50
                     |||...||.||.|||||.||.|||||.||.||||||||.||.||.|||||
US17_OPTIM         1 ATGAGCCCCAATTCAGAAGCAACCGGCACCGCCTGGGCACCACCCCCCCC     50

US17_Canon        51 GCGACCCTCGCGCGGAGTCATTATGATTTCCTCCGTCTCGACGAACGACG    100
                     .|||||.....|.|||||.|||||||||..|||||||.||.||.|||||||
US17_OPTIM        51 CCGACCTAGCAGAGGAGTGATTATGATTAGCTCCGTGTCAACAAACGACG    100

US17_Canon       101 TACGTCGCTTTTTTACTTTGTATGCGGGTCTACAGCACCGTGGCCGTGCAG    150
                     |.||.||.||..|.||.||.||||||||.||||||||.|||||.||.|||
US17_OPTIM       101 TGCGGCGGTTCCTGCTGTGCATGCGGGTGTACAGCACTGTGGCAGTCCAG    150

US17_Canon       151 GGCACCTGCACCTTCTTGCTCTGTCTGGGCCTGGTGCTGGCTTTTCCGCA    200
                     ||.||||||||.|||.||||.||.||||||||||||||||||||||.||
US17_OPTIM       151 GGAACCTGCACATTCCTGCTGTGCCTGGGCCTGGTGCTGGCTTTTCCCCA    200

US17_Canon       201 TCTTAAAGGCACCGTCTTTCTCTGTTGCACCGGCTTTATGCCGCCCTTGA    250
                     .||.||.||||||||||.||.||.||.||.||||||||.||..|||
US17_OPTIM       201 CCTGAAGGGCACCGTCTTCCTGTGCTGTACAGGGTTTATGCCCCCTCTGA    250

US17_Canon       251 GTTTGATGGTGCCCACCATCTGTTTGGCCCTGCTGCACGGCAAACGCGAT    300
                     |..||||||||||.|||||||||.||||||||||||||.||.|||||.|||
US17_OPTIM       251 GCCTGATGGTGCCTACCATCTGTCTGGCCCTGCTGCATGGAAAACGGGAT    300

US17_Canon       301 GAAGGATCGTTCACGTCGCCGCCGAGCCCGGGCCTGCTCACCATTTATAG    350
                     ||.||.||.||.||....||.||....||.||.|||||.|||||.|||||
US17_OPTIM       301 GAGGGCTCTTTTACTAGTCCACCCTCACCCGGGCTGCTGACCATCTATAG    350

US17_Canon       351 CGTGCTCACGACGCTTTCGGTGATCGTGGCCAGCGCCTGCTCCTCCTCTA    400
                     .|||||.||.||.||.||.|||||.||.||.||.||||||||....||.|
US17_OPTIM       351 TGTGCTGACCACACTGTCTGTGATTGTCGCAAGTGCCTGCTCTAGTTCAA    400

US17_Canon       401 CGCTGGTGACCTTCTCGGGCCTCTTGGCTTGCGTGCTTTTTAGCCTCTGC    450
                     |.||||||||||||...||.||..||||.||||||||.||.|||||.|||
US17_OPTIM       401 CTCTGGTGACCTTCAGCGGACTGCTGGCCTGCGTGCTGTTCAGCCTGTGC    450

US17_Canon       451 AGCTGCGTCACGGGTCTAGCCGGCCATAATCACCGTCGATGGCAGGTCAT    500
                     ..||||||.||.||.||.||.||.||||||..|.||.||||||||||||
US17_OPTIM       451 TCCTGCGTGACCGGACTGGCTGGGCACAATCATAGGCGCTGGCAGGTCAT    500

US17_Canon       501 CGTCACGCTGTTTGTGATCGGCGTTATCGCGTTCTTGATCGCACTTTACC    550
                     ||||||.|||||.||||||||||.||.||.||..||||.||.||.||||
US17_OPTIM       501 CGTCACACTGTTCGTGATCGGCGTCATTGCATTTCTGATTGCCCTGTACC    550

US17_Canon       551 TGCAGCCCGTACCGTTGGGGCACAAACTTTTTTTGGGCTATTACGCCATG    600
                     |||||||.||.||..||||||||||.||.||..||||.||.||.||.|||
US17_OPTIM       551 TGCAGCCTGTGCCACTGGGGCACAAGCTGTTCCTGGGATACTATGCTATG    600

US17_Canon       601 GCGCTCAGCTTCATGCTGGTCGTCACGGTCTTTGACACCACGCGCCTGTT    650
                     ||.||...|||.||||||||.|||||.||.||.|||||.||.||.|||||
US17_OPTIM       601 GCACTGTCCTTTATGCTGGTGGTCACCGTGTTCGACACTACCCGACTGTT    650
```

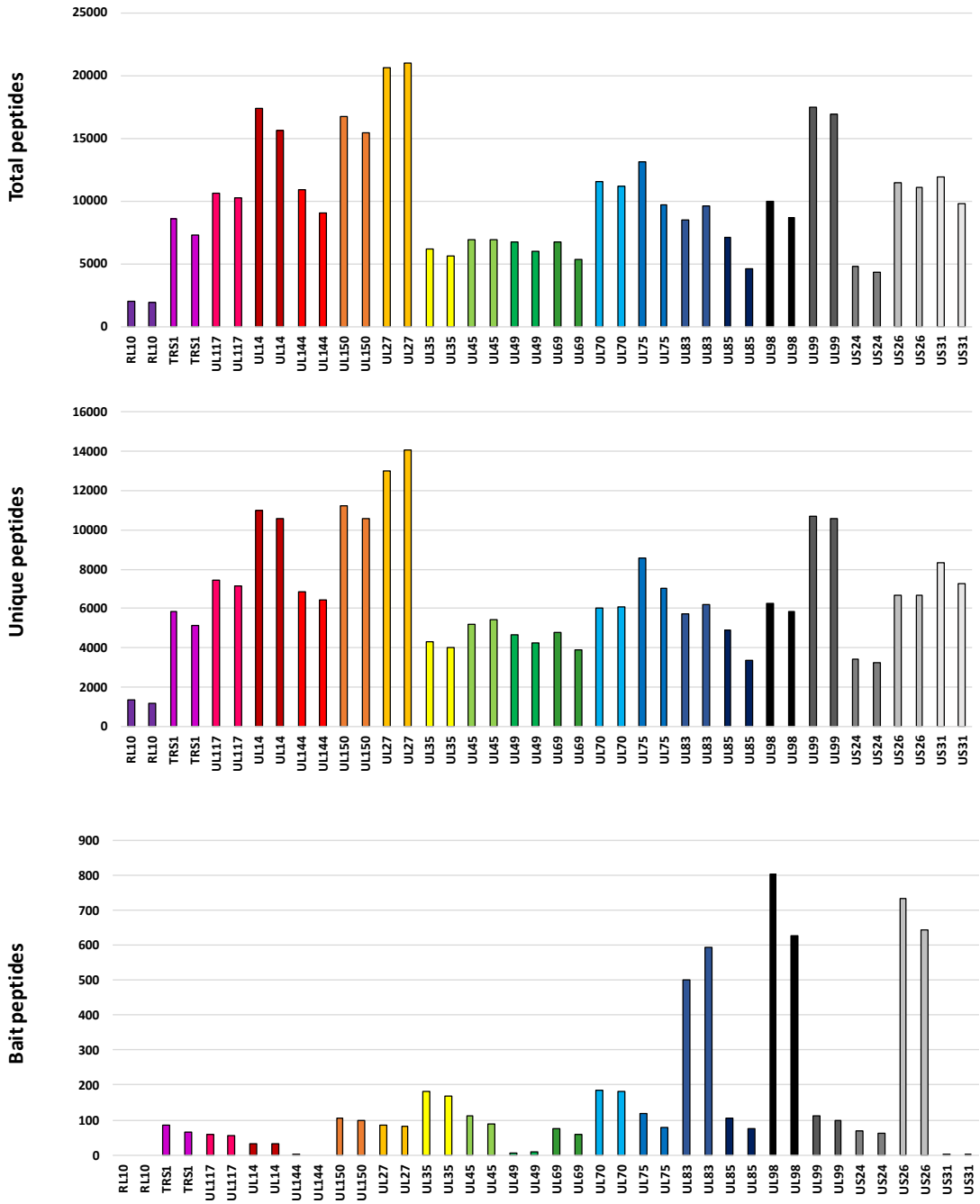
```
US17_Canon       651 TGAGATCGCGTGGTCCGAGGCCGACCTGCTCACCTTGTGTCTCTATGAGA    700
                     |||||||||.|||||.||.||||||||||.||..||||.||.|||||||
US17_OPTIM       651 TGAGATCGCTTGGTCTGAAGCCGACCTGCTGACACTGTGCCTGTATGAGA    700

US17_Canon       701 ACCTGGTGTACCTGTACCTGCTCATTCTCATCCTTTTCACCACCGAGGAC    750
                     |||||||||||||||.|||||.||.||.||.||.|||||.||.||.|||
US17_OPTIM       701 ACCTGGTGTACCTGTATCTGCTGATCCTGATTCTGTTCACAACTGAAGAC    750

US17_Canon       751 TCATTAGACAAACTCATCGCTTGGATGACCTGGTTATCGTCACGCGCCAC    800
                     ||..|.||.|||||.||.||.||||||||||||.|....||..|.|||||
US17_OPTIM       751 TCCCTGGATAAACTGATTGCCTGGATGACCTGGCTGAGCTCCAGAGCCAC    800

US17_Canon       801 CGGGGCCACCAACGCGGCCTCCATTTCGGGCTGTGACCTTTTGCGGGAGG    850
                     .|||||.||.||.||.||...|||.||.|||||.|||||..||||.||.|
US17_OPTIM       801 AGGGGCTACTAATGCCGCTAGCATCTCCGGCTGCGACCTGCTGCGCGAAG    850

US17_Canon       851 TACAGAGAAACCTCACGCGAACCATGGCG    879
                     |.|||||.|||||.||.|||||.|||||.
US17_OPTIM       851 TCCAGAGGAACCTGACACGAACTATGGCA    879
```

Appendix C | Correlation of peptide quantification between replicate samples and controls

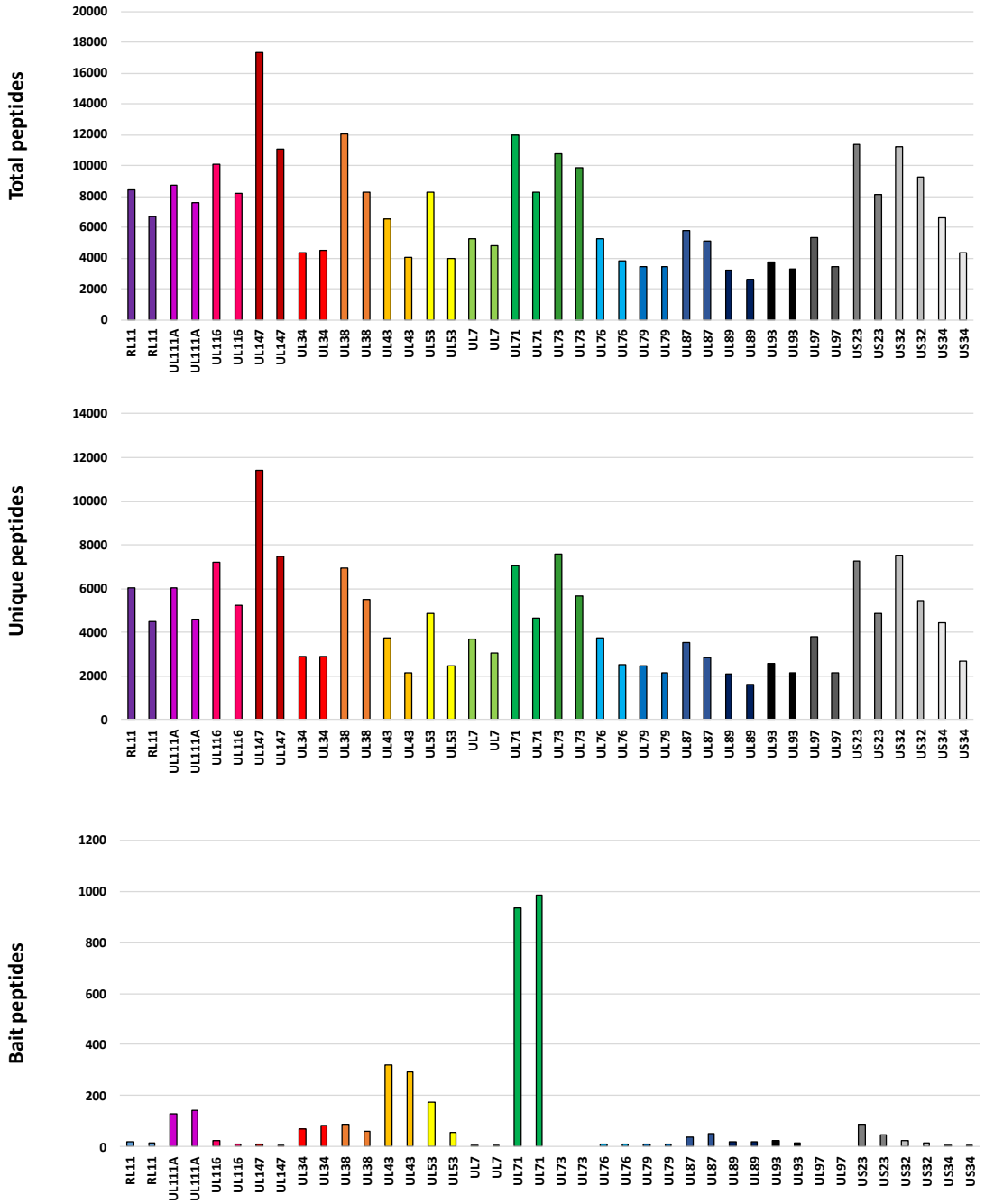
*Data underlying the graphs in Figure 5.1.2, depicting the correlation of the number of total, unique and bait peptides from each protein identified in the technical replicates. Peptide numbers were obtained from the 'Summary Plots' function provided within "MassPike", the Sequest-based software pipeline for quantitative proteomics (described in **2.11.1**)*

C.1 Replicate samples for viral baits

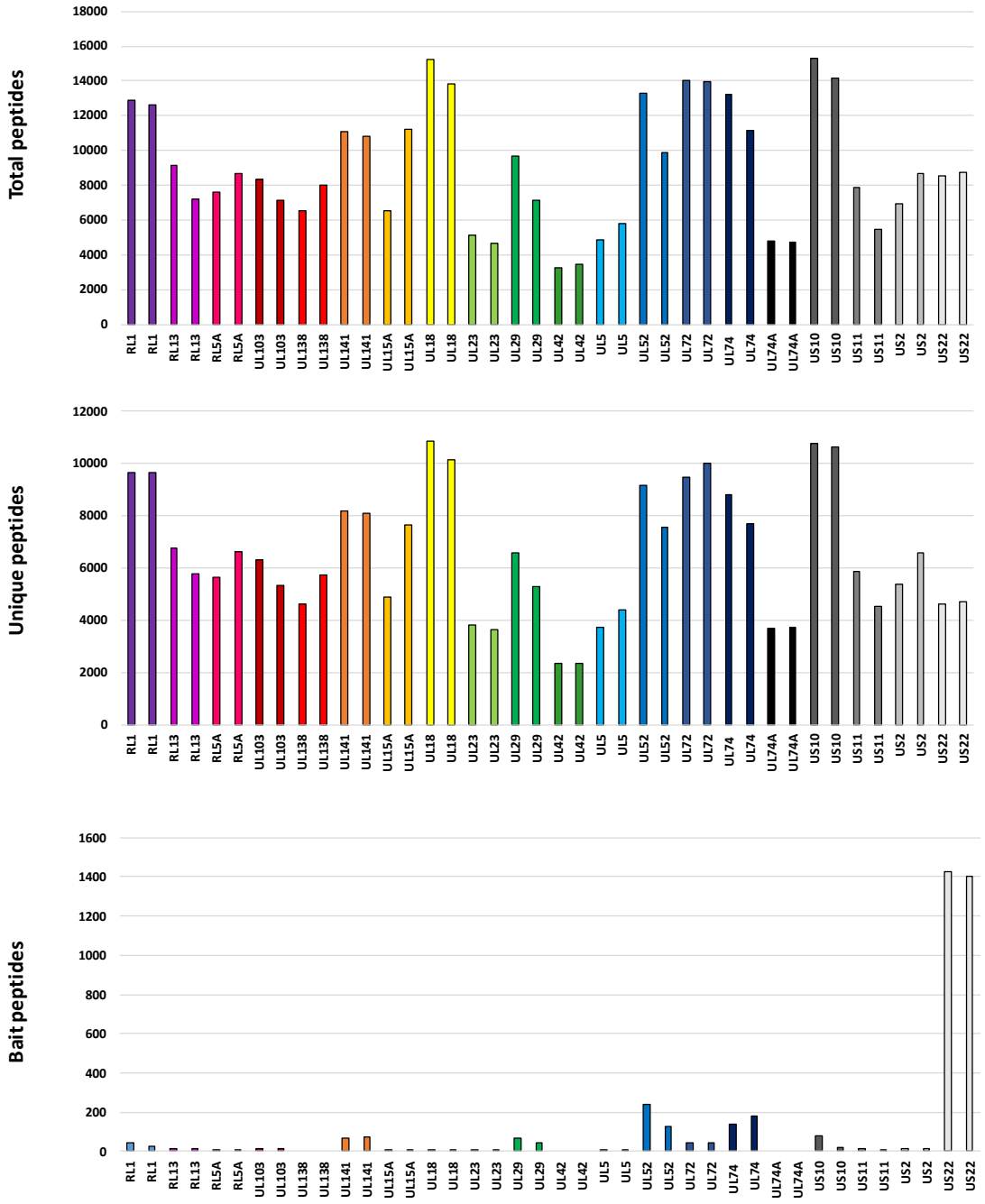



Total peptides
25000
20000
15000
10000
5000
0
Unique peptides
16000
14000
12000
10000
8000
6000
4000
2000
0
Bait peptides
900
800
700
600
500
400
300
200
100
0
RL10 RL10 TRS1 TRS1 UL117 UL117 UL14 UL14 UL144 UL144 UL150 UL150 UL27 UL27 UL35 UL35 UL45 UL45 UL49 UL49 UL69 UL69 UL70 UL70 UL75 UL75 UL83 UL83 UL85 UL85 UL98 UL98 UL99 UL99 US24 US24 US26 US26 US31 US31

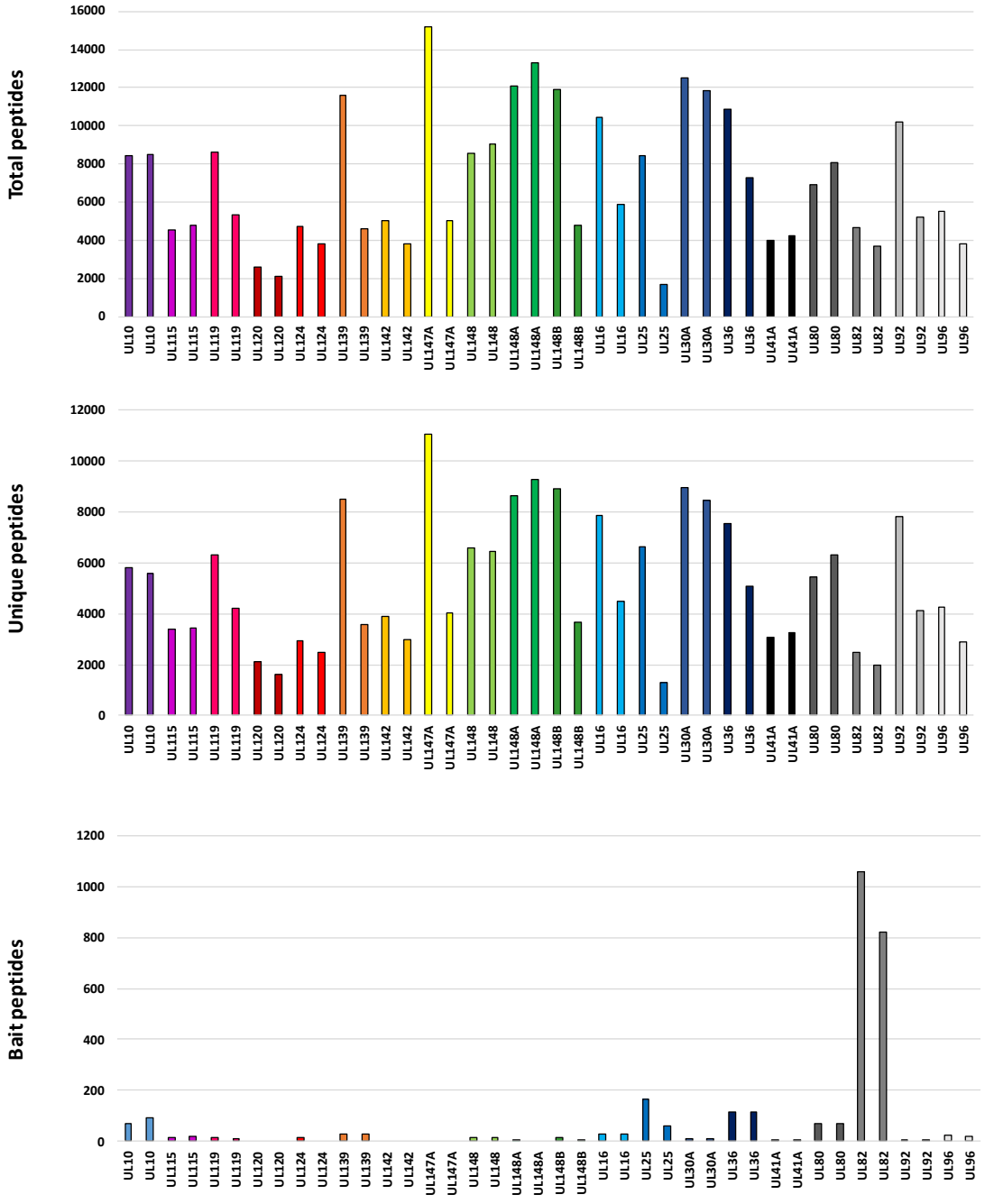
(Continued)


Total peptides
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Unique peptides
14000
12000
10000
8000
6000
4000
2000
0
Bait peptides
1200
1000
800
600
400
200
0
RL11 RL11 UL111A UL111A UL116 UL116 UL147 UL147 UL34 UL34 UL38 UL38 UL43 UL43 UL53 UL53 UL7 UL7 UL71 UL71 UL73 UL73 UL76 UL76 UL79 UL79 UL87 UL87 UL89 UL89 UL93 UL93 UL97 UL97 US23 US23 US32 US32 US34 US34

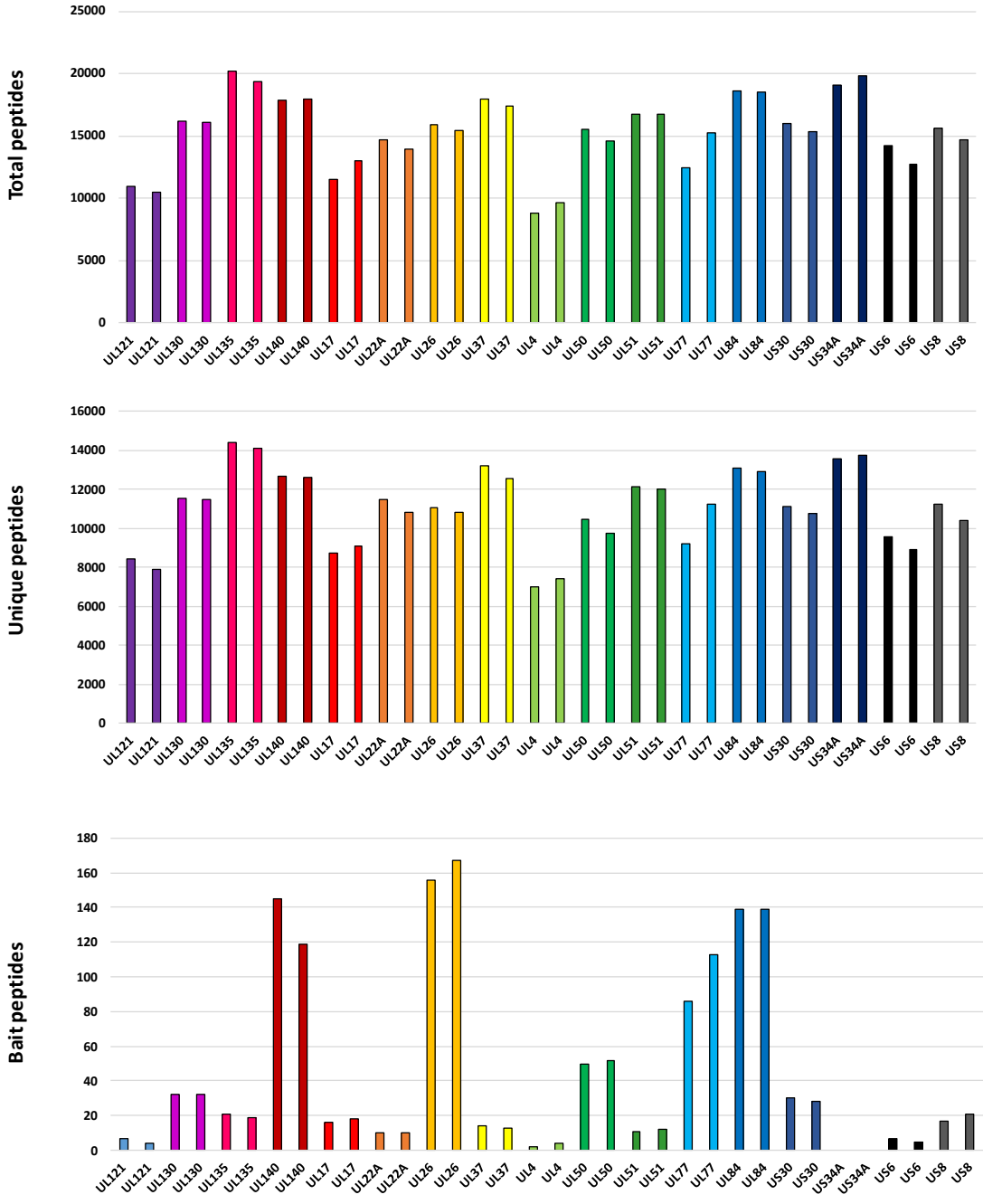
(Continued)


Total peptides
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Unique peptides
12000
10000
8000
6000
4000
2000
0
Bait peptides
1600
1400
1200
1000
800
600
400
200
0
RL1 RL1 RL13 RL13 RL5A RL5A UL103 UL103 UL138 UL138 UL141 UL141 UL15A UL15A UL18 UL18 UL23 UL23 UL29 UL29 UL42 UL42 UL5 UL5 UL52 UL52 UL72 UL72 UL74 UL74 UL74A UL74A US10 US10 US11 US11 US2 US2 US22 US22

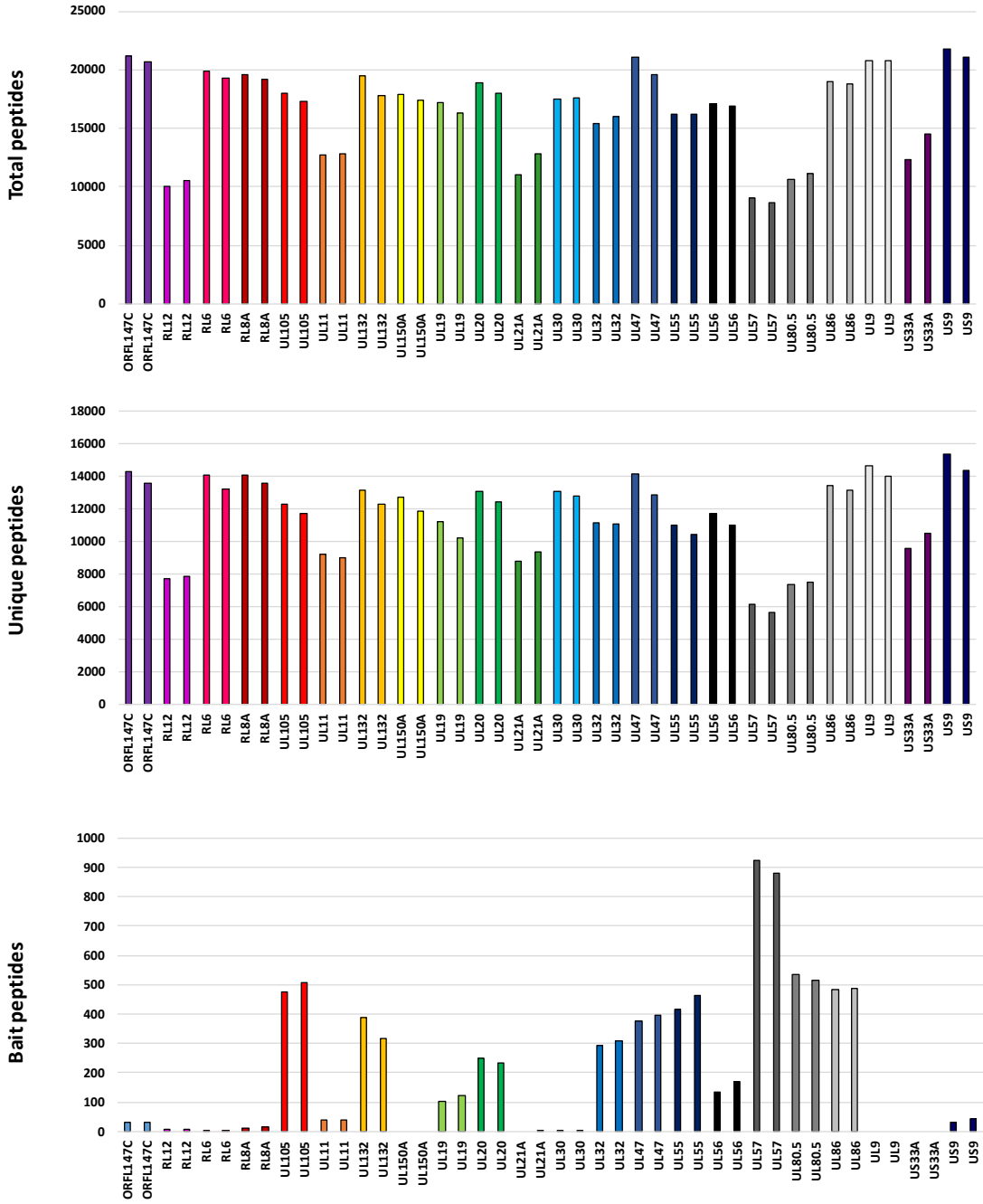
(Continued)


Total peptides
16000
14000
12000
10000
8000
6000
4000
2000
0
Unique peptides
12000
10000
8000
6000
4000
2000
0
Bait peptides
1200
1000
800
600
400
200
0
UL10 UL10 UL115 UL115 UL119 UL119 UL120 UL120 UL124 UL124 UL139 UL139 UL142 UL142 UL147A UL147A UL148 UL148 UL148A UL148A UL148B UL148B UL16 UL16 UL25 UL25 UL30A UL30A UL36 UL36 UL41A UL41A UL80 UL80 UL82 UL82 UL92 UL92 UL96 UL96

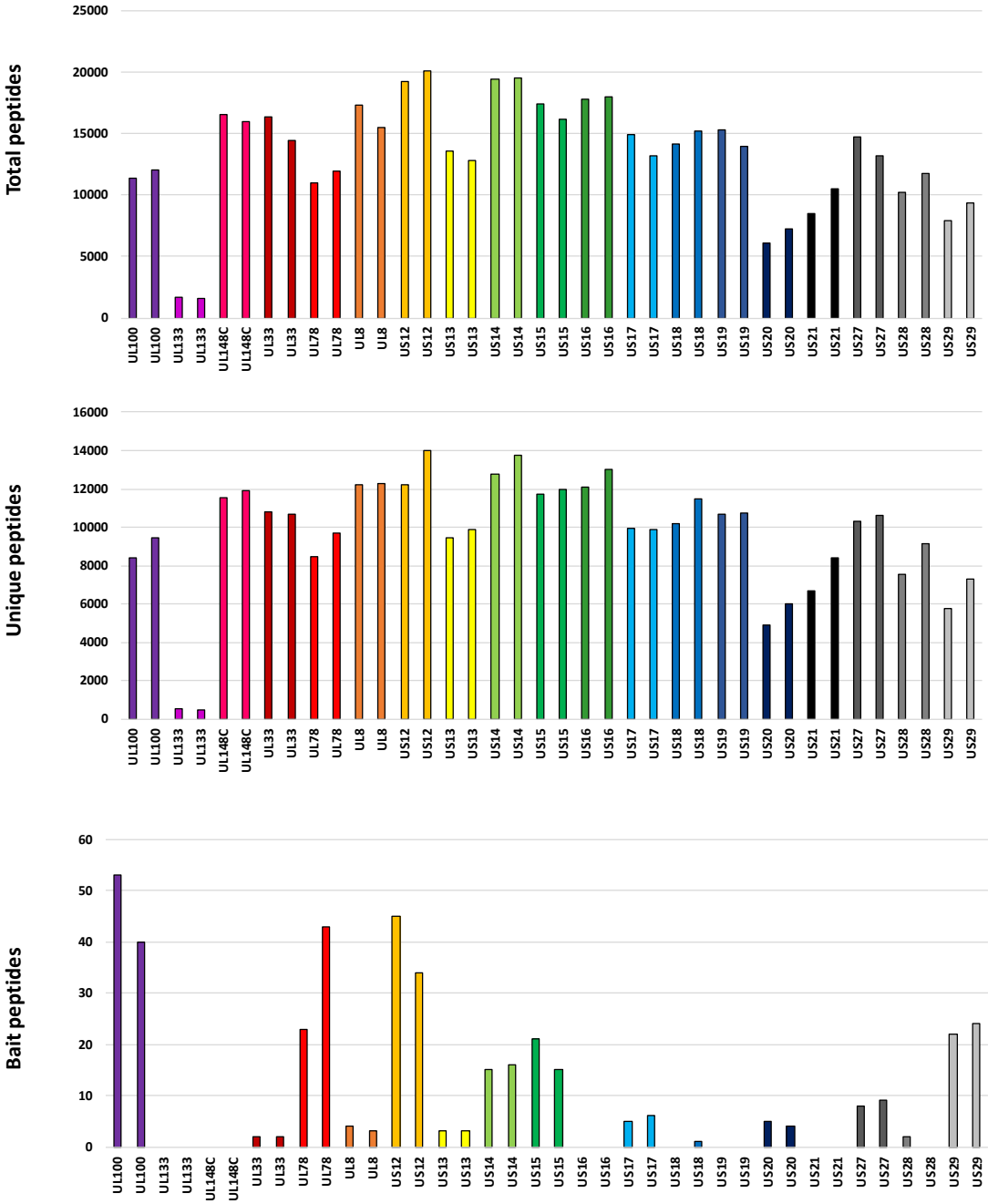
(Continued)


Total peptides
25000
20000
15000
10000
5000
0
Unique peptides
16000
14000
12000
10000
8000
6000
4000
2000
0
Bait peptides
180
160
140
120
100
80
60
40
20
0
UL121 UL121 UL130 UL130 UL135 UL135 UL140 UL140 UL17 UL17 UL22A UL22A UL26 UL26 UL37 UL37 UL4 UL4 UL50 UL50 UL51 UL51 UL77 UL77 UL84 UL84 US30 US30 US34A US34A US6 US6 US8 US8

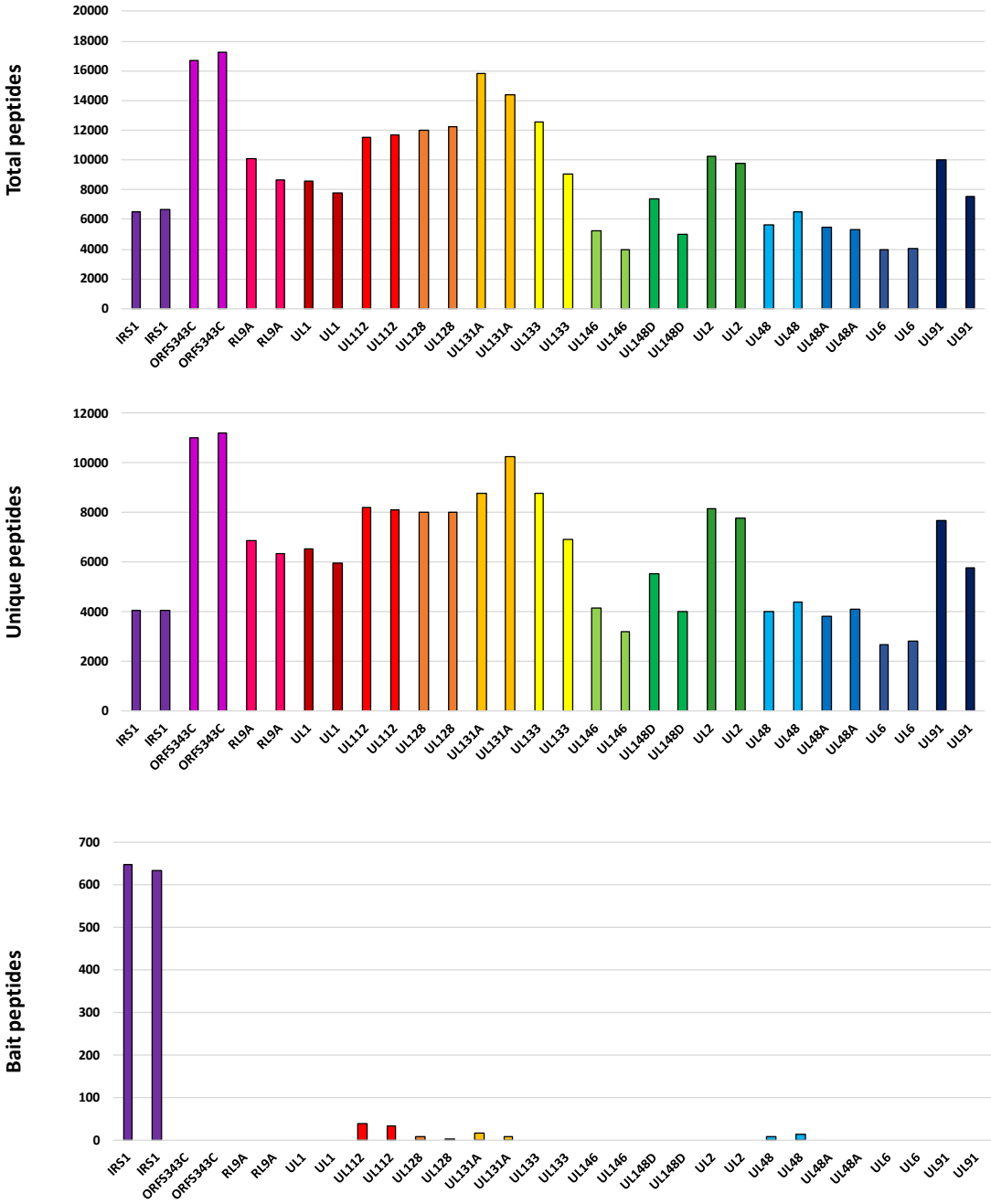
(Continued)


Total peptides
25000
20000
15000
10000
5000
0
Unique peptides
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Bait peptides
1000
900
800
700
600
500
400
300
200
100
0
ORFL147C
ORFL147C
RL12
RL12
RL6
RL6
RL8A
RL8A
UL105
UL105
UL11
UL11
UL132
UL132
UL150A
UL150A
UL19
UL19
UL20
UL20
UL21A
UL21A
UL30
UL30
UL32
UL32
UL47
UL47
UL55
UL55
UL56
UL56
UL57
UL57
UL80.5
UL80.5
UL86
UL86
UL9
UL9
US33A
US33A
US9
US9

(Continued)


Total peptides
25000
20000
15000
10000
5000
0
Unique peptides
16000
14000
12000
10000
8000
6000
4000
2000
0
Bait peptides
60
50
40
30
20
10
0
UL100 UL100 UL133 UL133 UL148C UL148C UL33 UL33 UL78 UL78 UL8 UL8 US12 US12 US13 US13 US14 US14 US15 US15 US16 US16 US17 US17 US18 US18 US19 US19 US20 US20 US21 US21 US27 US27 US28 US28 US29 US29

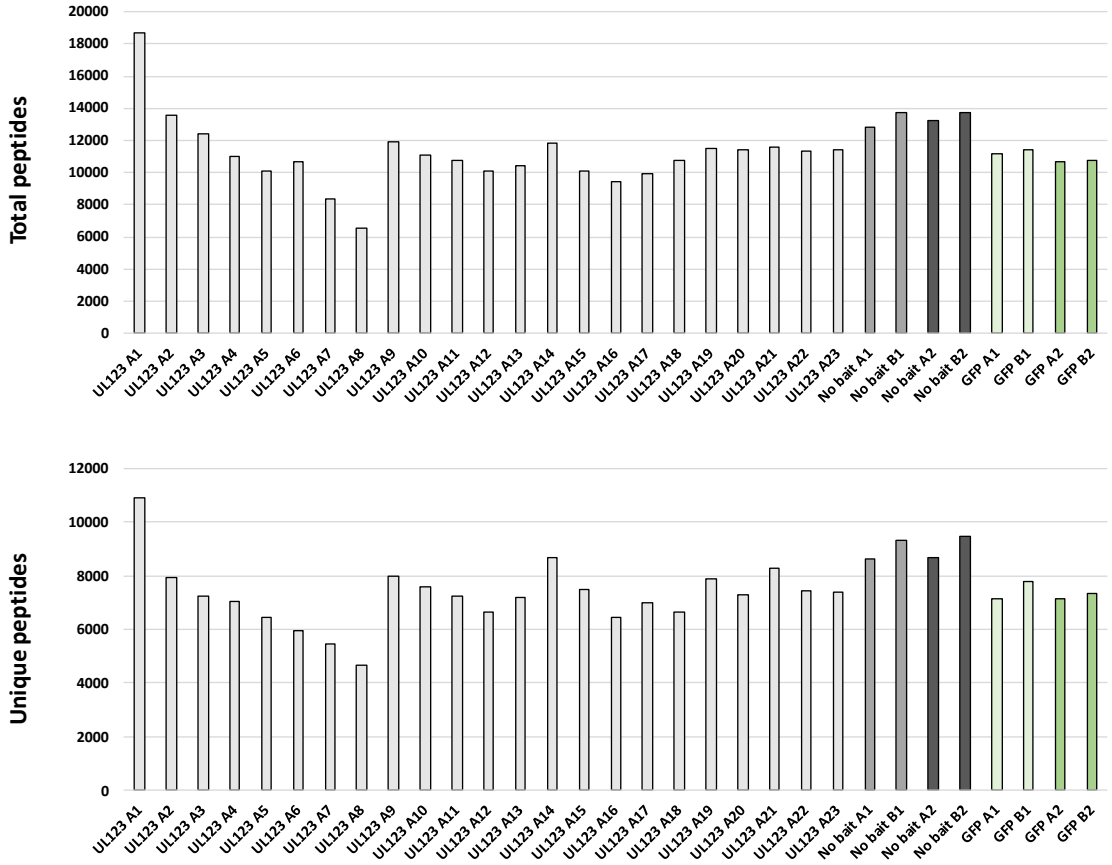
(Continued)


Total peptides
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
IRS1 IRS1 ORF5343C ORF5343C RL9A RL9A UL1 UL1 UL112 UL112 UL128 UL128 UL131A UL131A UL133 UL133 UL146 UL146 UL148D UL148D UL2 UL2 UL48 UL48 UL48A UL48A UL6 UL6 UL91 UL91
Unique peptides
12000
10000
8000
6000
4000
2000
0
IRS1 IRS1 ORF5343C ORF5343C RL9A RL9A UL1 UL1 UL112 UL112 UL128 UL128 UL131A UL131A UL133 UL133 UL146 UL146 UL148D UL148D UL2 UL2 UL48 UL48 UL48A UL48A UL6 UL6 UL91 UL91
Bait peptides
700
600
500
400
300
200
100
0
IRS1 IRS1 ORF5343C ORF5343C RL9A RL9A UL1 UL1 UL112 UL112 UL128 UL128 UL131A UL131A UL133 UL133 UL146 UL146 UL148D UL148D UL2 UL2 UL48 UL48 UL48A UL48A UL6 UL6 UL91 UL91

(Continued)

C.2 Control samples

Correlation of the number of total and unique peptides from each protein identified in the uninfected UL123 control samples (A1-A23) used to assess instrument performance (detailed in Table 5.1), and the technical replicates (A and B) and biological duplicates (A1, A2) for 'No bait' and GFP controls.



Appendix D | Annotated interactions in curated databases

Data underlying the analysis on the overlap between HCMV interactome HCIPs and annotated interactions on Uniprot, Virus Mentha, BioGRID, IntAct and MINT, depicted in Figure 5.3.20.

Bait	Prey	Databases	Detection Method	PubMed ID	Identified in cells infected with HCMV	Cell type	Validated by interactome
UL115	UL74	UNIPROT	IP	28403202	No	HFF	Yes
UL74	UL115	UNIPROT	IP	28403202	No	HFF	Yes
UL73	UL100	UNIPROT	IP	11090188	No	293T	Yes
UL48A	UL86	UNIPROT	IP	12552013	No	SF9	Yes
UL86	UL48A	UNIPROT	IP	12552013	No	SF9	Yes
UL70	UL102	UNIPROT, MINT, INTACT, VM	IP	12076828	No	BHK-21	Yes
UL75	UL115	UNIPROT	IP	17942555	Yes	HDF	Yes
UL115	UL75	UNIPROT	IP	17942555	Yes	HDF	Yes
UL99	UL94	UNIPROT	Y2H	19345970	No	AH109	Yes
UL94	UL99	UNIPROT	Y2H	19345970	No	AH109	Yes
UL51	UL56	UNIPROT	IP	23175377	Yes	HFF	Yes
UL56	UL51	UNIPROT	IP	23175377	Yes	HFF	Yes
UL70	UL105	UNIPROT, MINT, INTACT, VM	IP	12076828	No	BHK-21	Yes
UL47	UL48	UNIPROT	IP	24829352	Yes	HFF	Yes
UL50	UL53	UNIPROT	IP	24155370	Yes	HFF	Yes
UL29	UL38	UNIPROT	IP	23236067	No	U2OS	Yes
UL44	UL114	UNIPROT	IP	16022730, 18599070	Yes	Primary HFF, HE	Yes
UL75	UL55	UNIPROT	IP	27082872	Yes	HFF	Yes
UL35	UL82	UNIPROT	IP	15308743	Yes	HFF	Yes
UL82	UL35	UNIPROT	IP	15308743	Yes	HFF	Yes
UL84	UL122	UNIPROT	IP	8289376, 7933141	Yes	HEL	Yes
UL54	UL44	UNIPROT	IP	14671097	No	BL21	Yes
UL44	UL54	UNIPROT, VM	IP	14671097	No	BL21	Yes
UL74	UL75	UNIPROT	IP	28403202	No	HFF	Yes
UL75	UL74	UNIPROT	IP	28403202	No	HFF	Yes
UL51	UL89	UNIPROT	IP	23175377	Yes	HFF	Yes
UL35	OGT	UNIPROT, BIOGRID, VM	AP-MS	22072767	No	293A	Yes
UL27	PSME3	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	Yes
UL29	MTA2	UNIPROT	IP	20585571	Yes	HFF	Yes

UL87	POLR2M	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	Yes
UL87	POLR2G	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	Yes
UL141	PVR	UNIPROT	SPR	23555243	No	-	Yes
UL38	TSC2	UNIPROT	IP	18407068	Yes	HFF	Yes
UL83	IFI16	UNIPROT, INTACT	IP	24237704	No	293T	Yes
UL36	UBR5	UNIPROT, VM	AP-MS	22810585	No	293	Yes
UL35	VPRBP	UNIPROT, BIOGRID, VM	AP-MS	22072767	No	293A	Yes
UL87	POLR2D	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	Yes
UL87	POLR2H	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	Yes
UL36	DDX28	VM	AP-MS	22810585	No	293	Yes
UL16	MICB	UNIPROT	PPBA	11239445	No	CV-1/EBNA	Yes
UL82	DAXX	UNIPROT, BIOGRID, VM	IP	11992005	No	293	Yes
UL16	RAET1G	VM	PPBA	19658097	No	CV1	Yes
UL16	ULBP2	UNIPROT, VM	PPBA	19424970	No	Cos-7	Yes
UL87	POLR2K	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	Yes
UL29	HDAC1	UNIPROT	IP	20585571	Yes	HFF	Yes
UL123	STAT2	UNIPROT, BIOGRID	IF	26559840	No	HFF	Yes
UL42	ITCH	UNIPROT, BIOGRID, VM	IP	26555021	No	293T	Yes
UL27	PSMB6	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	Yes
UL36	GDI2	VM	AP-MS	22810585	No	293	Yes
UL87	POLR2A	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	Yes
UL87	POLR2B	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	Yes
UL87	POLR2C	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	Yes
UL87	POLR2E	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	Yes
UL87	POLR2I	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	Yes
UL87	POLR2L	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	Yes
UL97	CDH1	UNIPROT, BIOGRID, VM	PPBA	20686030	No	-	Yes
UL135	ABI1	UNIPROT	IP	25121749	Yes	HFF	Yes
UL27	DDB1	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	Yes
UL27	UBR5	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	Yes
UL102	UL70	UNIPROT, MINT, INTACT	IP	12076828	No	BHK-21	No
UL105	UL70	UNIPROT	IP	12076828	No	BHK-21	No
UL53	UL50	UNIPROT	IP	24155370	Yes	HFF	No
UL97	UL83	UNIPROT	IP	17634236	Yes	HFF	No
UL80	UL86	UNIPROT	Y2H	8985337	No	PCY2	No
UL86	UL80	UNIPROT	Y2H	8985337	No	PCY2	No
UL105	UL102	UNIPROT	IP	12076828	No	BHK-21	No

(Continued)

UL102	UL105	UNIPROT, MINT, INTACT, VM	IP	12076828	No	BHK-21	No
UL46	UL86	UNIPROT	Y2H	8892863	No	PCY2	No
UL84	UL44	UNIPROT	IP	17959680	Yes	HFF	No
UL44	UL84	UNIPROT	IP	17959680	Yes	HFF	No
TRS1	UL44	UNIPROT	IP	20444996	Yes	HFF	No
IRS1	UL44	UNIPROT	IP	20444996	Yes	HFF	No
UL104	UL89	UNIPROT	IP	16282466	No	BL21	No
UL44	UL112	UNIPROT	IP	20538862	Yes	HFF	No
UL112	UL44	UNIPROT	IP	20538862	Yes	HFF	No
UL89	UL56	UNIPROT	IP	11744697	Yes	-	No
UL56	UL89	UNIPROT	IP	11744697	Yes	-	No
UL122	RB1	UNIPROT, BIOGRID, VM	Y2H	9671498	No	Y190	No
US32	DLG1	MINT	PPPD	24550280	No	-	No
UL27	RAB1A	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	RAN	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL122	RPS17	UNIPROT, BIOGRID, VM	AP-MS	22810585	No	293	No
UL76	PSMD4	UNIPROT, BIOGRID, VM	IP	23966401	Yes	Hel 299	No
UL76	UBC	UNIPROT, BIOGRID, VM	FRET	23966401	No	293	No
UL35	SART3	UNIPROT, BIOGRID, VM	AP-MS	22072767	No	293A	No
UL122	MCM3	UNIPROT	IP	20545442	No	U373MG	No
UL36	RAB10	VM	AP-MS	22810585	No	293	No
UL27	CBR1	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL123	TRIM5	UNIPROT, BIOGRID, VM	IP	25412268	No	293	No
UL27	PDIA4	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL123	PML	UNIPROT, BIOGRID, VM	Y2H	9671498	No	Y190	No
UL27	PSMD3	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	PSMD2	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	STAT3	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	EP400	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	ACLY	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL135	ABI2	UNIPROT	IP	25121749	Yes	HFF	No
UL135	TLN1	UNIPROT	IP	25121749	Yes	HFF	No
UL55	EGFR	UNIPROT	IP	12879076	Yes	HEL	No
UL35	DDB1	UNIPROT, BIOGRID, VM	AP-MS	22072767	No	293A	No
UL36	MYO9A	VM	AP-MS	22810585	No	293	No
UL114	SMARCB1	UNIPROT, BIOGRID, VM	Y2H, IP	22479537	Yes	HFF	No
UL122	CSNK2A1	UNIPROT, BIOGRID, VM	AP-MS	22810585	No	293	No

(Continued)

UL122	DYNLL1	UNIPROT, BIOGRID, VM	AP-MS	22810585	No	293	No
UL122	DYNLL2	UNIPROT, BIOGRID, VM	AP-MS	22810585	No	293	No
UL122	HSPA5	UNIPROT, BIOGRID, VM	AP-MS	22810585	No	293	No
UL122	KPNA3	UNIPROT, BIOGRID, VM	AP-MS	22810585	No	293	No
UL27	ACTL6A	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	FKBP10	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	HNRNPH3	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	NUDT21	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	PSMA3	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	PSMB4	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	PSMC4	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	PSMC6	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	RUVBL1	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	RUVBL2	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL27	TMEM43	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	No
UL36	ATAD3A	VM	AP-MS	22810585	No	293	No
UL36	HNRNPH2	VM	AP-MS	22810585	No	293	No
UL36	HSPA5	VM	AP-MS	22810585	No	293	No
UL36	NEFH	VM	AP-MS	22810585	No	293	No
UL36	RPS27	VM	AP-MS	22810585	No	293	No
UL36	TUBB	VM	AP-MS	22810585	No	293	No
UL44	SMARCB1	UNIPROT, BIOGRID, VM	IP	22479537	Yes	HFF	No
UL100	UL73	UNIPROT	IP	11090188	No	293T	ND
UL48	UL47	UNIPROT	IP	24829352	Yes	HFF	ND
UL55	UL75	UNIPROT	IP	27082872	Yes	HFF	ND
UL86	UL46	UNIPROT	Y2H	8892863	No	PCY2	ND
UL89	UL51	UNIPROT	IP	23175377	Yes	HFF	ND
UL89	UL104	UNIPROT	IP	16282466	No	BL21	ND
UL27	TRRAP	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	ND
UL111A	IL10RA	VM, UNIPROT	XRC	12093920, 15837194	No	-	ND
UL122	CSNK2B	UNIPROT, BIOGRID, VM	AP-MS	22810585	No	293	ND
UL122	ZMYND11	UNIPROT, BIOGRID, VM	AP-MS	22810585	No	293	ND
UL123	DAXX	UNIPROT, BIOGRID, VM	IP	20444888	Yes	U373	ND
UL123	HES1	UNIPROT, BIOGRID	IP	28750047	Yes	NPC	ND
UL123	SP100	UNIPROT, BIOGRID	IP	28750047	No	293T	ND
UL123	UBE2D1	UNIPROT, BIOGRID	IP	28750047	Yes	NPC	ND
UL144	TRAF6	UNIPROT, BIOGRID, VM, MINT	IP	19176615	No	HFF	ND

(Continued)

UL144	TRIM23	UNIPROT, BIOGRID, VM	IP	19176615	No	HFF	ND
UL16	RAET1L	VM	PPBA	19658097	No	CV1	ND
UL21A	ANAPC7	UNIPROT, BIOGRID, VM	AP-MS	22792066	Yes	MRC-5	ND
UL21A	CDC23	UNIPROT, BIOGRID, VM	AP-MS, IP	22792066	Yes	MRC-5	ND
UL21A	CDC27	UNIPROT, BIOGRID, VM	AP-MS, IP	22792066	Yes	MRC-5	ND
UL27	KAT5	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	ND
UL27	WDR26	UNIPROT, BIOGRID, VM	AP-MS	21320693	Yes	Primary HFFs	ND
UL35	DDA1	UNIPROT, BIOGRID, VM	AP-MS	22072767	No	293A	ND
UL35	IPO4	UNIPROT, BIOGRID, VM	AP-MS	22072767	No	293A	ND
UL35	USP7	UNIPROT, BIOGRID, VM	AP-MS, IP	22072767	No	293A	ND
UL36	ACTA2	VM	AP-MS	22810585	No	293	ND
UL36	ACTG2	VM	AP-MS	22810585	No	293	ND
UL36	CASP8	UNIPROT, VM	AP-MS	22810585	No	293	ND
UL36	IRS4	VM	AP-MS	22810585	No	293	ND
UL36	RPS27L	VM	AP-MS	22810585	No	293	ND
UL36	TUBA1A	VM	AP-MS	22810585	No	293	ND
UL36	TUBA1C	VM	AP-MS	22810585	No	293	ND
UL36	USP54	VM	AP-MS	22810585	No	293	ND
UL82	ATRX	UNIPROT, BIOGRID, VM	IP	23135716	No	HepaRG	ND
UL83	MNDA	INTACT	IP	24237704	No	293T	ND
UL83	PYHIN1	INTACT	IP	24237704	No	293T	ND
UL87	POLR2J	UNIPROT, BIOGRID	AP-MS	25544563	No	293T	ND
US11	DERL1	VM	IP	15215856	No	Astrocytoma	ND
US11	TRAM1	UNIPROT, BIOGRID, VM	IP	19121997	No	U373	ND
US2	TRAM1	UNIPROT, BIOGRID, VM	IP	19121997	No	U373	ND
US3	PDIA2	UNIPROT, BIOGRID, VM	IP	17055437	No	HeLa	ND
US3	TAPBP	UNIPROT, BIOGRID, VM	IP	17055437	No	HeLa	ND

VM: VIRUSMENTHA; AP-MS: Affinity-purification mass spectrometry; IP: Immunoprecipitation; Y2H: Yeast two-hybrid; SPR: Surface plasmon resonance; FRET: Fluorescent Resonance Energy Transfer; IF: Immunofluorescence; XRC: X-ray crystallography; PPBA: Purified protein binding assay; PPPD: Proteomic peptide-phage display.

Appendix E | PFAM domain association

Data underlying the protein domain-domain association predictions shown in Figure 5.6.2.

Bait	PFAM Domain	Prey Uniprot-l	Prey Gene Name	Interacting PFAM domains
UL25	Herpes pp85	O43639	NCK2	SH2, SH3 9
UL25	Herpes pp85	P16333	NCK1	SH2, SH3 9
UL35	Herpes pp85	F5HBC6	UL82	Herpes UL82 83
US27	7tm 1	P30825	SLC7A1	AA permease
US27	7tm 1	Q9UK23	NAGPA	NAGPA
US27	7tm 1	P20338	RAB4A	Ras
US27	7tm 1	Q86YS6	RAB43	Ras
US27	7tm 1	Q9NRW1	RAB6B	Ras
US27	7tm 1	Q15771	RAB30	Ras
US27	7tm 1	P20340-2	RAB6A	Ras
US27	7tm 1	Q6IQ22	RAB12	Ras
US27	7tm 1	P20337	RAB3B	Ras
US27	7tm 1	Q9UL25	RAB21	Ras
US27	7tm 1	O14966	RAB29	Ras
US27	7tm 1	Q9UP95	SLC12A4	SLC12, AA permease
US27	7tm 1	Q9UHW9	SLC12A6	SLC12, AA permease
US27	7tm 1	Q9BXP2	SLC12A9	SLC12, AA permease
US27	7tm 1	O43752	STX6	SNARE
US27	7tm 1	O14662	STX16	SNARE
US27	7tm 1	Q86Y82	STX12	SNARE
US27	7tm 1	O15400	STX7	SNARE
US27	7tm 1	Q9UNK0	STX8	SNARE
US27	7tm 1	P51809-2	VAMP7	Synaptobrevin
US27	7tm 1	P51809	VAMP7	Synaptobrevin
US27	7tm 1	P63027	VAMP2	Synaptobrevin
US27	7tm 1	Q9Y487	ATP6V0A2	V ATPase I
US27	7tm 1	Q9UEU0	VTI1B	V-SNARE, SNARE
US27	7tm 1	Q96AJ9	VTI1A	V-SNARE, SNARE
US28	7tm 1	Q9UKX2	MYH2	Myosin head, Myosin tail 1, Myosin N
US28	7tm 1	P13533	MYH6	Myosin head, Myosin tail 1, Myosin N
US28	7tm 1	A7E2Y1	MYH7B	Myosin head, Myosin tail 1, Myosin N
US28	7tm 1	Q9Y623	MYH4	Myosin head, Myosin tail 1, Myosin N
US28	7tm 1	P11055	MYH3	Myosin head, Myosin tail 1, Myosin N
US28	7tm 1	P12882	MYH1	Myosin head, Myosin tail 1, Myosin N
US28	7tm 1	P12883	MYH7	Myosin head, Myosin tail 1, Myosin N
US28	7tm 1	Q9UK23	NAGPA	NAGPA
US28	7tm 1	P20340-2	RAB6A	Ras
US28	7tm 1	Q9Y487	ATP6V0A2	V ATPase I

RL6	RL11D	Q86V21	AACS	ACAS N
RL6	RL11D	Q96K17	BTF3L4	NAC
RL6	RL11D	E9PAV3	NACA	NAC
RL6	RL11D	P62256	UBE2H	UQ con
RL6	RL11D	Q16763	UBE2S	UQ con
RL6	RL11D	P61088	UBE2N	UQ con
RL6	RL11D	Q13404-1	UBE2V1	UQ con
UL117	Herpes IE2 3	Q9H3K6	BOLA2	BolA
UL122	Herpes IE2 3	Q9H3K6	BOLA2	BolA
UL14	UL141	Q14697-2	GANAB	Glyco hydro 31, Gal mutarotas 2
UL14	UL141	Q14697	GANAB	Glyco hydro 31, Gal mutarotas 2
UL141	UL141	O00220	TNFRSF10A	TNFR c6
UL141	UL141	Q9UBN6	TNFRSF10D	TNFR c6
UL141	UL141	O14763	TNFRSF10B	TNFR c6
UL141	UL141	P19438	TNFRSF1A	TNFR c6
US3	Cytomega US3	Q06481	APLP2	Kunitz BPTI, APP N, APP E2, APP Cu bd, APP amyloid
US3	Cytomega US3	P05067	APP	Kunitz BPTI, APP N, APP E2, APP Cu bd, APP amyloid
US7	CMV US	Q5VSL9	STRIP1	N1221, DUF3402
US9	CMV US	Q9ULQ0	STRIP2	N1221, DUF3402
US9	CMV US	Q5VSL9	STRIP1	N1221, DUF3402
US9	CMV US	Q9NVK5	FGFR1OP2	SIKE
US9	CMV US	Q9NVK5-2	FGFR1OP2	SIKE
US9	CMV US	Q9BRV8-2	SIKE1	SIKE
US9	CMV US	O43815	STRN	Striatin
US9	CMV US	Q13033	STRN3	Striatin
US12	Bax1-I	Q29980	MICB	C1-set
US12	Bax1-I	Q29983	MICA	C1-set
US12	Bax1-I	P00533	EGFR	Recep L domain, GF recep IV, Furin-like
US12	Bax1-I	P04626	ERBB2	Recep L domain, GF recep IV, Furin-like
US15	Bax1-I	Q68D85	NCR3LG1	C1-set
US20	Bax1-I	P01860	IGHG3	C1-set
US20	Bax1-I	P01834	IGKC	C1-set
US20	Bax1-I	P01857	IGHG1	C1-set
US20	Bax1-I	P04626	ERBB2	Recep L domain, GF recep IV, Furin-like
US21	Bax1-I	Q14643	ITPR1	Ion trans,MIR, RYDR ITPR, RIH assoc, Ins145 P3 rec
US21	Bax1-I	Q14573	ITPR3	Ion trans, MIR, RYDR ITPR, RIH assoc, Ins145 P3 rec
US21	Bax1-I	Q14571	ITPR2	Ion trans, MIR, RYDR ITPR, RIH assoc, Ins145 P3 rec
US21	Bax1-I	P04626	ERBB2	Recep L domain, GF recep IV, Furin-like
IRS1	US22	Q9P2N7-5	KLHL13	CH, BTB
IRS1	US22	Q9P2J3	KLHL9	CH, BTB
IRS1	US22	P46109	CRKL	SH3 2

(Continued)

IRS1	US22	P46108-2	CRK	SH3 2
IRS1	US22	P46108	CRK	SH3 2
IRS1	US22	Q99962	SH3GL2	SH3 2
IRS1	US22	Q96JP2	MYO15B	SH3 2
UL23	US22	Q8N3F8	MICALL1	CH
UL23	US22	Q8IZ07	ANKRD13A	CH
UL23	US22	O76071	CIAO1	WD40
UL26	US22	Q9UL63	MKLN1	CH
UL26	US22	Q9H871	RMND5A	CLTH
UL26	US22	Q7L5Y9	MAEA	CLTH
UL26	US22	Q6VN20	RANBP10	CLTH, LisH
UL26	US22	Q96S59	RANBP9	CLTH, LisH
UL26	US22	Q9NWU2	GID8	CLTH, LisH
UL26	US22	Q8IUH3-3	RBM45	RRM 1
UL26	US22	Q8IUH3	RBM45	RRM 1
UL26	US22	Q9NNW5	WDR6	WD40
UL26	US22	Q9H7D7	WDR26	WD40
UL26	US22	Q9C0J8	WDR33	WD40
UL26	US22	Q8NFH4	NUP37	WD40
UL26	US22	O95486	SEC24A	zf-Sec23 Sec24, Sec23 trunk, Sec23 helical, Sec23 BS, Gelsolin
UL26	US22	Q15437	SEC23B	zf-Sec23 Sec24, Sec23 trunk, Sec23 helical, Sec23 BS, Gelsolin
UL26	US22	Q15436	SEC23A	zf-Sec23 Sec24, Sec23 trunk, Sec23 helical, Sec23 BS, Gelsolin
UL29	US22	Q9UPU5	USP24	CH
UL29	US22	Q14839-2	CHD4	Chromo, DUF1087, DUF1086, CHDNT, CHDCT2, SNF2 N, Helicase C, CH
UL29	US22	Q8TDI0	CHD5	Chromo, DUF1087, DUF1086, CHDNT, CHDCT2, SNF2 N, Helicase C, CH
UL29	US22	Q9NWU2	GID8	CLTH, LisH
UL29	US22	Q8WXI9	GATAD2B	GATA
UL29	US22	O94776	MTA2	GATA, MTA R1, ELM2, BAH
UL29	US22	O60907	TBL1X	LisH, WD40
UL29	US22	Q9BZK7	TBL1XR1	LisH, WD40
UL29	US22	Q9UBB5	MBD2	MBD C, MBDa, MBD
UL29	US22	O95983	MBD3	MBD C, MBDa, MBD
UL29	US22	O75376	NCOR1	Myb DNA-binding
UL29	US22	Q9Y618	NCOR2	Myb DNA-binding
UL29	US22	Q9BTC8	MTA3	Myb DNA-binding, GATA, MTA R1, ELM2, BAH
UL29	US22	Q13330	MTA1	Myb DNA-binding, GATA, MTA R1, ELM2, BAH
UL29	US22	Q9NNW5	WDR6	WD40
UL36	US22	Q9UK99	FBXO3	F-box-like
UL36	US22	Q9NUL7	DDX28	Helicase C
UL38	US22	Q9UPU5	USP24	CH

(Continued)

UL38	US22	Q14839-2	CHD4	Chromo, DUF1087, DUF1086, CHDNT, CHDCT2, SNF2 N, Helicase C, CH
UL38	US22	Q9UBB5	MBD2	MBD C, MBDa, MBD
UL38	US22	O95983	MBD3	MBD C, MBDa, MBD
UL38	US22	Q9BTC8	MTA3	Myb DNA-binding, GATA, MTA R1, ELM2, BAH
UL38	US22	Q13330	MTA1	Myb DNA-binding, GATA, MTA R1, ELM2, BAH
UL38	US22	Q03468	ERCC6	SNF2 N, Helicase C
UL38	US22	Q16576	RBBP7	WD40
UL43	US22	P62258	YWHAE	14-3-3
UL43	US22	Q04917	YWHAH	14-3-3
UL43	US22	P61981	YWHAG	14-3-3
UL43	US22	P63104	YWHAZ	14-3-3
UL43	US22	P31946	YWHAB	14-3-3
UL43	US22	P27348	YWHAQ	14-3-3
US22	US22	Q5BKZ1	ZNF326	AKAP95
US22	US22	Q9ULX6	AKAP8L	AKAP95
US22	US22	O43823	AKAP8	AKAP95
US22	US22	Q9UKB1	FBXW11	Beta-TrCP D, F-box-like, WD40
US22	US22	Q7RTP6	MICAL3	CH
US22	US22	P78332	RBM6	CH
US22	US22	O14647	CHD2	Chromo, SNF2 N, Helicase C, CH
US22	US22	Q969H0	FBXW7	F-box-like, WD40
US22	US22	Q96E39	RBMXL1	RBM1CTR, RRM 1
US22	US22	O75526	RBMXL2	RBM1CTR, RRM 1
US22	US22	P38159	RBMX	RBM1CTR, RRM 1
US22	US22	Q8IXT5	RBM12B	RRM 1
US22	US22	Q9Y580	RBM7	RRM 1
US22	US22	Q96PK6	RBM14	RRM 1
US22	US22	Q1KMD3	HNRNPUL2	SAP
US22	US22	Q00839	HNRNPU	SAP
US22	US22	Q9NWH9	SLTM	SAP, RRM 1
US22	US22	Q14151	SAFB2	SAP, RRM 1
US22	US22	Q8NDT2	RBM15B	SPOC, RRM 1
US22	US22	Q96T37	RBM15	SPOC, RRM 1
US22	US22	Q9H5H4	ZNF768	zf-C2H2
US22	US22	Q02447	SP3	zf-C2H2
US22	US22	Q6DD87	ZNF787	zf-C2H2
US22	US22	O43474	KLF4	zf-C2H2
US22	US22	P08047	SP1	zf-C2H2
US22	US22	Q9HBE1	PATZ1	zf-C2H2, BTB
US23	US22	Q9Y297	BTRC	Beta-TrCP D, F-box-like, WD40
US23	US22	Q15052	ARHGEF6	RhoGEF, CH,betaPIX CC
US24	US22	Q9UKB1-3	FBXW11	Beta-TrCP D, F-box-like, WD40

(Continued)

US24	US22	Q9Y297	BTRC	Beta-TrCP D, F-box-like, WD40
US24	US22	Q9UKB1	FBXW11	Beta-TrCP D, F-box-like, WD40
US24	US22	Q15052	ARHGEF6	RhoGEF, CH, betaPIX CC
US24	US22	Q14155	ARHGEF7	RhoGEF, CH, SH3 2
US26	US22	Q5JSZ5-4	PRRC2B	BAT2 N
US26	US22	Q9Y520-7	PRRC2C	BAT2 N
US26	US22	Q5JSZ5	PRRC2B	BAT2 N
US26	US22	P48634	PRRC2A	BAT2 N
US26	US22	Q8IY92	SLX4	BTB
US26	US22	Q7RTP6	MICAL3	CH
US26	US22	Q8WWI1	LMO7	CH
US26	US22	P78332	RBM6	CH
US26	US22	Q9P2N5	RBM27	CH
US26	US22	Q92620	DHX38	Helicase C
US26	US22	P42285	SKIV2L2	Helicase C, CH
US26	US22	O60907	TBL1X	LisH, WD40
US26	US22	O75376	NCOR1	Myb DNA-binding
US26	US22	Q9Y618	NCOR2	Myb DNA-binding
US26	US22	Q12774	ARHGEF5	RhoGEF
US26	US22	Q96PE2	ARHGEF17	RhoGEF
US26	US22	Q92888-3	ARHGEF1	RhoGEF
US26	US22	Q9Y580	RBM7	RRM 1
US26	US22	Q9NWH9	SLTM	SAP, RRM 1
US26	US22	P51532	SMARCA4	SNF2 N, Helicase C
US26	US22	Q92576	PHF3	SPOC
US26	US22	Q9H2Y7	ZNF106	WD40
US26	US22	Q9C0J8	WDR33	WD40
US26	US22	Q9Y4X4	KLF12	zf-C2H2
US26	US22	P57682	KLF3	zf-C2H2
US26	US22	Q8N554	ZNF276	zf-C2H2
US26	US22	O43474	KLF4	zf-C2H2
US26	US22	O95365	ZBTB7A	zf-C2H2, BTB
US26	US22	Q5VYS8	ZCCHC6	zf-CCHC, CH
US26	US22	Q6NZY4	ZCCHC8	zf-CCHC, CH
US26	US22	Q15637-5	SF1	zf-CCHC, CH