Applying network approaches to identify genes and circular RNAs that drive Kaposi's sarcoma-associated herpesvirus's lytic replication and pathogenesis

Euan Andrew McDonnell

Submitted in accordance with the requirements for the degree of a Doctorate of Philosophy

The University of Leeds

Faculty of Biological Sciences, Faculty of Medicine

School of Molecular and Cellular Biology, Leeds Institute for Data Analytics

September 2023

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would firstly like to thank Prof David Westhead for taking over my supervision just before the COVID-19 pandemic, when things were looking very bleak. In-particular his expertise, patience and understanding; giving me the space I needed to work at my best but always present to provide support when needed. Even through some terrible thesis chapter drafts and foolish mistakes. Additionally thank you to Prof Adrian Whitehouse for conceiving the initial project.

I'm extremely grateful to Amber, Vlad and Nazia, for always being supportive colleagues. In particular Nazia for seeking me out as the only person doing a similar project in the whole University, as well as for befriending me when I was most isolated. Also Scott and the rest of the MRC DiMeN cohort. I am also thankful to members of the Whitehouse lab as well as the FBS reps for helping me to improve working conditions for (and the caffeination of) fellow PhD students.

I am indebted to Leeds 'Omics, namely Dr Elton Rosas de Vasconcelos for being a friend and endless source of advice on Bioinformatics. And to Emily Goodall and Lucy Parker for helping me seek the appropriate sources of support.

I am grateful to my parents for providing emotional, financial and culinary-based support whenever I needed, alongside of course for getting me here in the first place.

Thank you to my friends, namely Will, Jonny, Bea, Bart and Geno, Adel, Jamal, Yousef and Habibur and Freya, Peter, Layla and Max for emphasising that no matter where I am in life, I am always welcome. And thank you to my housemates Dan, Toby and Tom who helped me learn to enjoy myself away from work. In addition I am grateful to Amy and Suki for all the wisdom and coffee.

Thank you to everyone who stood on me, held me back or discounted me, for giving me the drive to overcome and the stubbornness to persevere.

Lastly thank you to Eva, Jamie and the rest of my new colleagues for being welcoming and understanding, even when thesis writing overran.

This is the most difficult thing that I have ever done and perhaps will ever do. However in completing it I know my limit is far beyond what I imagined and I am excited to keep pushing. Finally, I would like to both end and begin with a quote from Dr Viktor Frankl,"When we are no longer able to change a situation - we are challenged to change ourselves".

Abstract

Kaposi's Sarcoma-associated herpesvirus (KSHV) is an oncogenic herpesvirus that exhibits a characteristic bi-phasic life cycle, existing in a suppressed latent cycle or an actively replicating lytic cycle. One aspect of gene expression regulation crucial in controlling this is the post-transcriptional level. Circular RNAs (circRNA) are covalently closed loops of RNA whose primary function is believed to be by acting at this level, as miRNA sponges, whereby they modulate the abundance of cognate miRNAs and downstream target mRNAs. Given the extensiveness of miRNA targeting networks, circRNAs represent efficient targets of viral dysregulation as they sit at the top of hierarchical networks of such interactions, termed competing endogenous RNA (ceRNA) networks. Initial studies have highlighted the involvement of circRNA miRNA sponges in KSHV's life cycle and pathogenesis, but much remains to be elucidated and their functional characterisation is limited. To this end, the first half of this study aimed to construct and analyse a circRNA ceRNA network of differentially expressed circRNAs, miRNAs and mRNAs to characterise the involvement and purpose of dysregulation of this level of regulation. This revealed several cellular circRNAs, primarily circBAGE3, circLRCH3, circSH3PXD2A and circSMG1P1 as highly influential in the network, while the network may be targeted to promote RNA synthesis and gene expression to drive lytic replication. Such a finding represents a novel mechanism by which KSHV modulates cellular non-coding RNA-based regulatory systems to promote the progression of its life cycle.

Viral-host interplay is believed to underpin much of KSHV's pathogenesis and to contribute to Kaposi Sarcoma (KS), the cancer it is named after. However, little is known about the individual determinants for the development of KS, primarily the influence of host factors. Recent developments in transcriptomics applied to KS lesion tissue has enabled many of these determinants to begin being elucidated, but in-depth analysis is lacking. Thus, in the second half of this study, we modelled and analysed the transcriptome of KS lesions by weighted gene co-expression network analysis (WGCNA). Module partitioning revealed a positive association between latent and some lytic genes and lesion development, while hub gene analysis suggested importance of the Ras/ERK/ETS1 axis and structural maintenance of chromosomes (SMC) proteins to this process. Differential gene co-expression analysis revealed two key factors, SIMC1 and LRRK2, as possible determinants in the transformation of healthy tissue. Such analyses help to identify and characterise novel candidate determinants that drive the oncogenesis of KSHV.

Contents

Abbreviations

(HA)ART	(Highly active) anti-retroviral therapy
$<$	Less than
$>$	Greater than
$\sim$	Approximately
2D	2 dimensional
3D	3 dimensional
AAV	Adeno-associated virus
Adar	Adenosine deaminase RNA specific
AGO	Argonaute protein
AIDS	Acquired immune deficiency syndrome
AKT/PKB	Protein kinase B
ALT	Antisense-to-latency transcript / Alternative lengthening of telomeres
ANG	Angiogenin
AP-1	Activating protein-1
ARHGEF12	Rho Guanine Nucleotide Exchange Factor 12
ATF	Activating transcription factor
BAGE3	B Melanoma Antigen (BAGE) family member 3
BCBL1	Body cavity-based lymphoma
Bcl-2	B-Cell CLL/Lymphoma 2
BCR	B cell receptor
BEC	Blood endothelial cell
BH	Benjamini-Hochberg
BIR	Break-induced replication
BP	Biological process
bZIP	Basic leucine zipper protein
C/EBPa	CCAAT/enhancer-binding protein alpha
Ca2+	Calcium 2+ ions
CASR	Calcium sensing receptor
CCL	C-C motif chemokine ligand
CD	Cluster of differentiation
CDKs	Cyclin dependent kinase
CDS	Coding sequence
CENP	Centromere protein F
ceRNA	Competing endogenous RNA
cGAS	Cyclic GMP-AMP synthase
CHRNA7	Cholinergic Receptor Nicotinic Alpha 7 Subunit
circRNA	Circular RNA
CK1	Caesin kinase 1
CPM	Counts per million
CREB	cAMP response element binding protein
CTCF	CCCTC-Binding Factor
CTL	Cytotoxic T cell
CTNNB1	B-catenin
CTNNBIP1	Catenin beta interacting protein 1
DBD	DNA binding domain
DDR	DNA damage response
DDX17	DEAD Box helicase 17
DEG	Differentially expressed gene
DEGES	Differentially expressed gene elimitation strategy
DGCA	Differential gene correlation analysis
DLBCL	Diffuse large B-cell lymphoma
DLL4	Delta-like 4
DNA	Deoxyribonucleic acid
DNMT	DNA methytransferas
Dox	Doxycycline
dsDNA	Double stranded DNA
dUbase	Dubiquitinase
E genes	Early lytic genes
EBV	Epstein Barr virus

EC	Endothelial cell
ECM	Extracellular matrix
EGFR	Epithelial growth factor receptor
EHF	ETS Homologous Factor
EiciRNA	Exon-intron circular RNA
ELK4	ETS Transcription Factor ELK4
EMT	Epithelial to mesemchymal transition
EndMT	Endothelial to mesemchymal transition
ERK	Extracellular signal-related kinases
ETS1	ETS Proto-Oncogene 1, Transcription Factor
EZH2	Enhancer of zeste homolog 2
FAK	Focal adhesion kinase
FAM110A	Family With Sequence Similarity 110 Member A
FAM120A	Family With Sequence Similarity 120A
FDR	False discovery rate
FGARAT	Formyl-glycinamide-phosphoribosyl-amidotransferase
FLICE	FADD-like interleukin-1-B-converting enzyme
GAP	GTPase activating protein
gC	Global connectivity
GEF	Guanine exchange factor
GEVA	Gene expression variation analysis
gH	Glycoprotein H
gL	Glycoprotein L
gM	Glycoprotein M
gN	Glycoprotein N
GO	Gene ontology
GSEA	Gene set enrichment analysis
GSK3	Glycogen synthase kinase 3
GSVA	Gene set variation analysis
GTPase	Guanylyl transferases
HBMSCs	Human bone marrow stromal cells
HBV	Hepatitis B virus
HCC	Hepatocellular carcinoma
HCMV	Human cytomegalovirus
HCV	Hepatitis C virus
HDAC	Histone deacetylase
HELF	Human lung fibroblast
HHV	Human herpesvirus
HIF	Hypoxia-inducible factor
HIPK2	Homeodomain interacting protein kinase 2
HIPK3	Homeodomain interacting protein kinase 3
HITS-CLIP	High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation
HIV	Human immunodeficiency virus
HMGA1	High Mobility Group AT-Hook 1
HOMER2	Homer scaffold protein 2
HPV	Human polyomavirus
HSP	Heat-shock protein
HSUR	Herpesvirus saimiri U-rich RNAs
HSV	Herpes simplex virus
hTREX	Human transcription-export complex
HUNK	Hormonally Up-Regulated Neu-Associated Kinase
HVS	Herpesvirus saimiri
IC	Intramodular connectivity
IE genes	Immediate early lytic genes
IFN	Interferon
IKK	Inhibitor of kB kinase
IL	Inteleukin
IRES	Internal ribosome entry site
IRIS	Immune reconstitution inflammatory syndrome
ISG	Interferon stimulated gene
iSLK	Renal carcinoma cell line contaminant

ITCH	Itchy E3 Ubiquitin Protein Ligase
JAK	Janus kinase
JNK	C-Jun N-terminal Kinase
KAP1	Krüppel-associated box domain-associated protein 1
KapA/B/C	Kaposin A/B/C
KCNJ6	Potassium inwardly rectifying channel subfamily J member 6
KCNN3	Potassium Calcium-Activated Channel Subfamily N Member 3
KCNQ5	Potassium Voltage-Gated Channel Subfamily Q Member 5
KCP	Viral complement control protein
KEGG	Kyoto encyclopedia of genes and genomes
KICS	KSHV-induced inflammatory cytokine syndrome
KIF9	Kinesin family 9
KS	Kaposi's sarcoma
KSHV	Kaposi's sarcoma-associated herpesvirus
L genes	Late lytic genes
LAC	Lung adrenocarcinoma
LANA	Latency associated nuclear antigen
LEC	Lymphatic endothelial cell
lncRNA	Long non-coding RNA
log2AE	Log2 average expression
log2FC/LFC	Log2 fold-change
LRCH3	Leucine Rich Repeats And Calponin Homology Domain Containing 3
LRRK2	Leucine rich-repeat kinase 2
LRT	Likelihood ratio test
m6A	N6-methyladenosine
m7G cap	5' methyl-7-guanosine cap
MAP3K4	Mitogen-Activated Protein Kinase Kinase Kinase 4
MAPK	Mitogen-activated protein kinase
MAVS	Mitochondrial antiviral signalling protein
MBS	miRNA binding site
MCD	Multicentric castleman's disease
MCP	Minor capsid protein
MCV	Merkel cell polyomavirus
MDM2	Mouse double minute 2
MDV	Marek's disease virus
ME	Module eigengene
MendT	Mesemchymal to endothelial transition
MGAT5	Alpha-1,6-Mannosylglycoprotein 6-Beta-N-Acetylglucosaminyltransferase
MHV68	Murine herpesvirus 68
MIP	Macrophage inflammatory protein
miRDE	miRNA decay element
miRNA	microRNA
MKK/MAPKK	MAPK kinase
MKKK/MAPKKK	MAPK kinase kinase
MM	Module membership
MMP	Matrix metalloproteinase
Mod	Module
MOR	Median-of-rations
MRGPRF	MAS Related GPR Family Member
mRNA	Messenger RNA
MTA	mRNA transcript accumulation
mtDNA	Mitochondrial DNA
MTOC	Microtubule organising center
MTOR	Mechanistic target of rapamycin
NaBu	Sodium butyrate
NB	Nuclear body
ncRNA	Non-coding RNA
ND10	Nuclear domains 10
NF-kB	Nuclear factor kappa B
NFAT	Nuclear factor of activated T cells
NGS	Next generation sequencing

NHEJ	Non homologous end joining
NLRP	Nucleotide-binding oligomerization domain
NO	Nitric oxide
NOS	Nitric oxide synthase
NPC	Nasopharyngeal carcinoma
NPM	Nucleophosmin
NRF	Nuclear respiratory factor
NSCLC	Non-small cell lung cancer
OA	Osteoarthritis
OIS	Oncogene-induced senescence
OLAP	Origin of lytic replication associated protein
OMA1	OMA1 Zinc Metallopeptidase
ORA	Over-representation analysis
ORF	Open reading frame
P-body	Processing body
PAA	Phosphonoacetic acid
PABPC1	Poly(A) Binding Protein Cytoplasmic 1
PAN	Polyadenylated nuclear transcript
PARE-Seq	Parallel analysis of read ends sequencing
PARP	Poly(ADP-ribose) polymerase 1
PARS	Protein abundance regulatory signal
PCA	Principal components analysis
PCALM	Phosphatidylinositol Binding Clathrin Assembly Protein
PCR	Polymerase chain reaction
PD	Parkinson's disease
PDGF	Platelet-derived growth factor
PDZRN3	PDZ Domain Containing Ring Finger 3
PEL	Primary effusion lymphoma
PI3K	phosphoinositide 3-kinase
PKA	Protein kinase A
PKC	Protein kinase C
PKR	Protein kinase R
PLC	Phospholipase C
PML	promyelocytic leukemia
pRb	Retinoblastoma protein
PRC	Polycomb repressor complex
PROX1	Prospero homeobox 1
PRR	Pathogen recognition receptor
PS	Phosphatidylserine
PTEN	Phosphatase and tension homologue
PWM	position weight matrix
qCLASH	quick Cross-Linking, Ligation, and Sequencing of Hybrids
QKI	Quaking
QRT-PCR	Quantitative reverse-transcriptase polymerase chain reaction
RBP	RNA binding protein
RBPJ	Recombination signal binding protein for immunoglobulin kappa J region
RIG-I	Retinoic acid-inducible gene I
RISC	RNA induced silencing complex
rLCV	rhesus lymphocryptovirus
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
RNAi	RNA interference
RNAPII	RNA polymerase II
ROCK	Rho Associated Coiled-Coil Containing Protein Kinase
ROS	Reactive oxygen species
RREs	RTA response elements
rRNA	Ribosomal RNA
RSK/RPS6KA3	Ribosomal S6 protein kinase
RTA	Replication and transcription activator
RV1/2	Rhadinovirus ½
SCP	Smallest capsid protein

ScRNA-Seq	Single-cell RNA-Seq
SH3PXD2A	SH3 And PX Domains 2A
SIM	SUMO interacting motif
SIMC1	SUMO Interacting Motifs Containing 1
SLF2	SMC5-SMC6 Complex Localization Factor 2
SMC	Structural maintenance of chromosomes
SMG1P1	SMG1 pseudogene 1
snoRNA	Small nucleolar RNA
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
SNX5	Sorting Nexin 5
SOX	Host shut-off factor
ssDNA	Single-stranded DNA
STAT	Signal transducer and activator of transcripton
STING	Stimulator of interferon response cGAMP
SUMO	Small ubiquitin-like modifier
TAD	Topologically associated domain
TCDD	2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD)
TCF	T-cell factor
TDMD	Target dependent miRNA degradation
TF	Transcription factor
TGF	Transforming growth factor
TIME	Telomerase-immortalized human microvascular endothelium cell line
TLR	Toll-like receptor
TMM	Trimmed-mean of M
TNBC	Triple negative breast cancer
TNF	Tumour necrosis factor
TNK1/KOS1	Tyrosine non-receptor kinase 1, Thirty-eight non-receptor kinase 1
TOM	Topological overlap matrix
TPA	12-O-tetradecanoyl-phorbol-13-acetate
TPM	Transcripts per million
TRAF6	TNF receptor associated factor 6
TRI1	Capsid trimer protein 1
TRI2	Capsid trimer protein 2
TRIB3	Tribbles 3 kinase
TRIO	Trio Rho Guanine Nucleotide Exchange Factor
tRNA	Transfer RNA
TWEAK	TNF superfamily member 12
TWEAKER	TNF Receptor Superfamily Member 12A
Ub	Ubiquitin
UBAP2	Ubiquitin Associated Protein 1
Ubase	Ubiquitin ligase
UCDS	Unique coding sequence
UTR	Untranslated region
vCyclin	Viral cyclin D1 homologue
VEGF	Vascular endothelial growth factor
VEGFR	Vascular endothelial growth factor receptor
vFLIP	Viral FLICE inhibitory protein
vGPCR	Viral G protein-coupled receptors
vIL6	Viral interleukin 6 homologue
vIRF	Viral interferon regulatory factor
VMP1	Vacuole membrane protein 1
vOX2	Viral CD200 homologue
vPK	Viral protein kinase
VZV	Varicella-Zoster virus
wED	Weighted euclidean distance
WGCNA	Weighted gene co-expression network analysis
WT	Wild-type
wTO	Weighted topological overlap
XKR4	XK related 4
ZEB1	Zinc Finger E-Box Binding Homeobox 1

ZFHX4	Zinc Finger Homeobox 4
ZKSCAN1	Zinc Finger With KRAB And SCAN Domains 1
ZSWIM5	Zinc Finger SWIM-Type Containing 5

1 Introduction

1.1 Herpesviruses

Herpesviruses are a broad collection of viruses that share a common phylogenetic origin as well as similarities in their virion and genome architecture, life cycle and infectious dynamics. They are all enveloped, large ($>$100kb) linear double-stranded DNA (dsDNA) viruses that exhibit bi-phasic life-cycles, whereby they undergo either a dormant latency or an actively replicating lytic replication cycle. The prior is associated with minimal gene expression and genome replication, while the later involves the expression of most genes, with the eventual production of infectious progeny virions.

Herpesviruses belong to *Herpesviridae*, a diverse family within the broader order *Herpesvirales* (Fig 1.1). Members of *Herpesviridae* can infect all forms of complex life as well as microorganisms and even prokaryotes. Within *Herpesviridae* there are 9 major herpesviruses that infect humans (often termed "human herpesviruses", HHVs) and generally these are the most well-studied members. They include Herpes simplex virus-1 and 2 (HSV-1/2), varicella-zoster virus (VZV), cytomegalovirus (HCMV), human herpesvirus 6 variants A/B (HHV6A/B) human herpesvirus 7 (HHV7), Epstein-Barr virus (EBV) and Kaposi's Sarcoma-associated herpesvirus (KSHV) [1]. These can be grouped into 3 based on phylogeny, with HSV-1, HSV-2 and VZV belonging to the α-herpesviruses, HCMV, HHV-6 and -7 belonging to the β-herpesviruses and KSHV and EBV belonging to the γ-herpesviruses [2]. While sharing common phylogenetic origins such herpesviruses vary not just in their genome sequence and structure but also in the cell types that they are able to infect, their life cycle and the pathologies that they are associated with [3]. Overall they represent a diverse and clinically important family of viruses for which much remains uncharacterised.

1.1.1 Alpha-Herpesviruses

As well as differing at a genetic level, human α-herpesviruses vary from β- and γ-herpesviruses in features of their life cycle and pathology. Specifically, unlike the latter two they have relatively short replication cycles and their infection progresses through to their rapid dissemination in cultured cells [4]. *In vivo*, they primarily infect and undergo such lytic replication in epithelial cells, but tend to latently infect terminally differentiated and non-replicative neuronal cells [2]. Because of this, unlike β- and γ-herpesviruses, they don't tether their genome to the host chromosome [4].

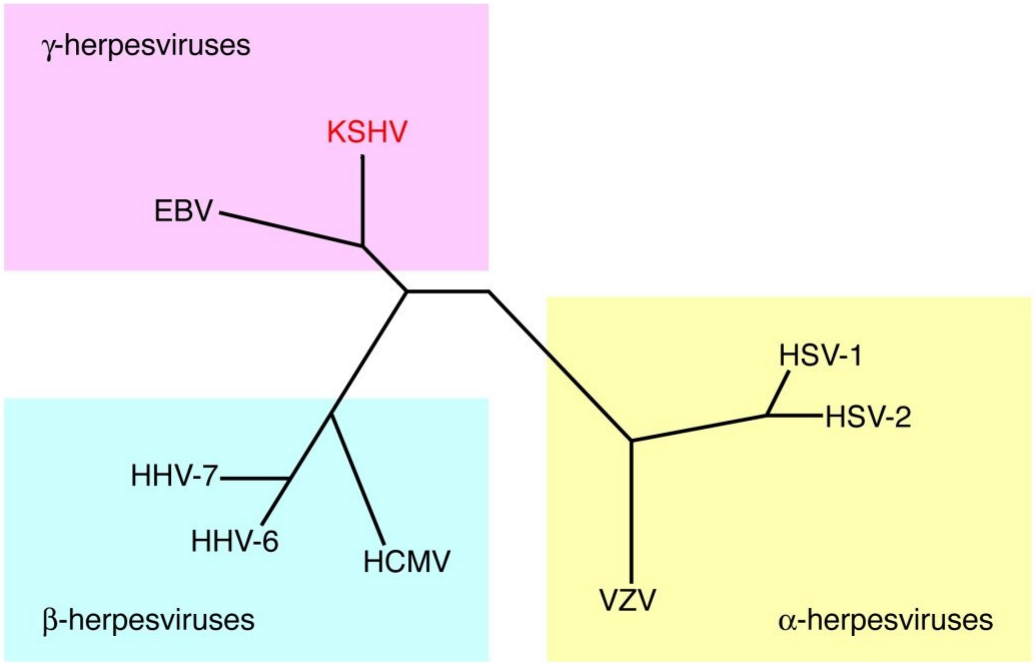


Figure 1.1: Phylogeny of human herpesviruses. Figure adapted from [5].

1.1.2 Beta-Herpesviruses

β-herpesviruses have similarities to both α- and γ-herpesviruses, but are phylogenetically closest to γ-herpesviruses (Fig 1.1). Like β-herpesvirus, initial infection tends to default to lytic replication, however this process is much slower, more like that observed for γ-herpesviruses [4, 4]. Moreover, similarly to γ-herpesviruses, lytic infection is associated with differentiation of infected cells, which tend to be lymphoreticular cells of the lymphoid organs [6].

1.1.3 Gamma-Herpesviruses

γ-herpesvirus's infection cycles vary from most α- and β-herpesviruses and this contributes to their pathogenicity. Infection by γ-herpesviruses generally progresses to latency after an initial burst of lytic gene expression [4]. However *in vivo*, establishment of latency varies by cell tropisms as, while γ-herpesviruses can infect a range of cell types including lymphocytes, epithelial and endothelial cells, they generally establish latency in B lymphocytes [4]. Nonetheless, the establishment of latency is a major reason for their propensity to persist in infected hosts, alongside their capacity to cause chronic illnesses. Accordingly, γ-herpesviruses are those most associated with cancers and this aspect of their interaction with hosts has garnered much research and makes them of paramount clinical relevance [4].

Further γ-herpesviruses related to KSHV (and to a lesser extend EBV) are often used as

models of human γ-herpesviruses [7, 8]. These include Murine Herpesvirus 68 (MHV-68) and Herpesvirus Saimiri (HVS) that, like KSHV, are grouped into the Rhadinovirus genus and respectively infect mice and primates [7, 8]. Rhadinovirus 1 and 2 (RV1 and RV2) are less well-studied *Rhadinoviruses* that are even more closely related to KSHV and infect old world primates [9]. Despite infecting different organisms, many insights gained from these herpesviruses have relevance to all γ-herpesviruses and herpesviruses in general.

1.2 Kaposi's Sarcoma-associated Herpesvirus

Since Kaposi's Sarcoma-associated Herpesvirus's (KSHV's) discovery its association with chronic and complex pathologies was immediately evident. This is in-part due to its discovery via isolation from Kaposi's Sarcoma (KS), a highly vascular AIDS-associated lesion, from which it derives its name and was subsequently found to be the primary causative agent of [10]. It is now known to drive at least 2 further cancer/cancer-like diseases including primary effusion lytmphoma (PEL), multicentric Castlemann's disease (MCD), alongside several further illnesses including Kaposi sarcoma inflammatory cytokine syndrome (KICS) and immune reconstitution inflammatory syndrome (IRIS), all detailed in a later section (Section 1.3.6) [11]. While KSHV is capable of causing illness in immunocompetent individuals, its maladies are generally associated with immunocompromised patients, particularly with AIDS and as such its epidemeology generally follows that of HIV, with high prevalence in Sub-Saharan Africa [12, 13]. It is now known to infect humans and is the only known member of the *Rhadinovirus* genus to do so.

Phylogenetically, it has features common and variable when compared to other γ-herpesviruses. As a *Rhadinovirus*, it is most-closely related to RV1, followed by RV2, for which it shares many genes unique to only these viruses (Fig 1.2). However it also has many similarities with the more well-studied γ-herpesvirus EBV, which also infects humans and is oncogenic, driving various lymphomas as well as certain nasopharyngeal and gastric carcinomas [14]. However KSHV varies in the exact nature of its life cycle, the pathologies it causes as well as the gene repertoire that it encodes (Fig 1.2). While the relationship between KSHV and human disease is well-established, the underlying mechanisms that underpin this association remain unclear.

1.2.1 Genome organisation

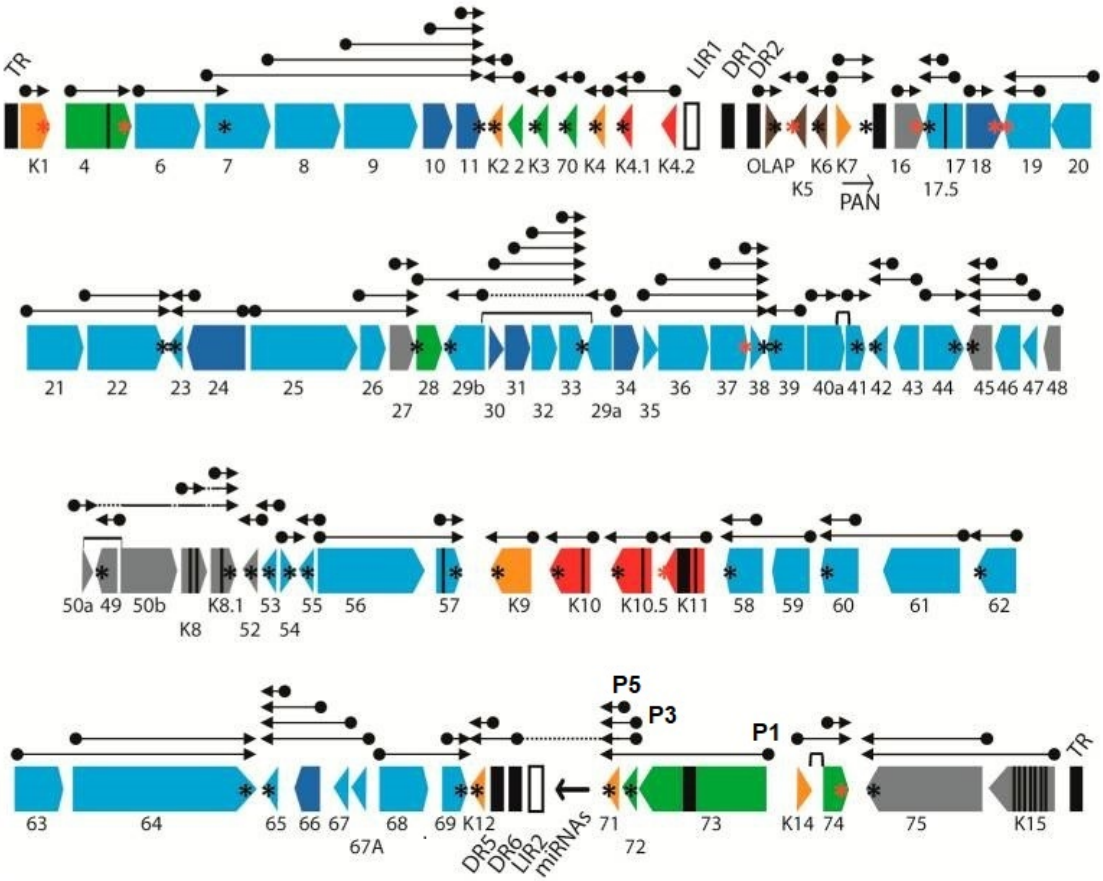


Figure 1.2: KSHV genes and genome organisation. Arrows correspond to differing transcripts produced by each locus, with solid lines corresponding mature spliced transcripts and dashed lines corresponding to introns. Known polyadenylation sites are marked with an asterix. ORFs are coloured according to their conservation, with genes conserved across all herpesviruses in light blue, genes conserved in β- and γ-herpesviruses in dark blue, γ-herpesviruses in grey, γ-2-herpesviruses (including EBV) in green, HVS/RV2 Rhadinoviruses in orange, HVS rhadinovirus in red and KSHV-only in brown. Where numbers refer to ORFs, (ORF4, ORF7, ORF8 etc..), genes starting with "K" are γ-2-herpesvirus-specific genes. TR = terminal repeat, LIR = long inverted repeats, DR = direct repeats, OLAP = origin of lytic replication associated protein and PAN = polyadenylated nuclear transcript. Figure is adapted from [15].

The genome of KSHV comprises $\sim$165 kb of linear dsDNA, sub-divided into two main components. The first comprises the $\sim$ 135kbp central "long unique region" (LUR) that contains all the coding and non-coding potential of the virus. This LUR region is flanked on either side by GC-rich 801bp long terminal repeat (TR) sequences that encompass the remaining $\sim$30 kb [13]. These TR regions mediate circularisation of the viral genome during lytic and latent replication, as well as facilitating its attachment to host chromosomes [13].

The genome contains over 95 genes, many of which contain open-reading frames (ORFs)

that express mRNAs which are translated by cellular machinery to produce $>$78 proteins, which facilitate many functions during viral life-cycle (Fig 1.2 & Table 1.1,1.2 & 1.3). These include receptor binding glycoproteins and proteins involved in viral genome replication and gene expression, virion assembly and egress, virion structure (capsid), the tegument as well as those that target host processes such as gene expression, immunomodulation, cellular replication, proliferation and the cell cycle (Table 2.3) [16–18]. Moreover many of these have described and proposed roles in cellular transformation and tumourigenesis [19]. These are broadly split into the latency locus (Kaposin-ORF73) at the 3' end of the genome downstream of ORF69, while all genes upstream of this locus encode lytic gene products (Fig 1.2 & 1.4). ORF74/5 and K14/15 are unusual lytic genes in that they are 3' to the latency locus (Fig 1.2).

As well as containing protein coding genes, KSHV encodes several non-coding RNAs that perform a range of cellular functions. These include the polyadenylated nuclear transcript (PAN), antisense-to-latency (ALT) transcript transcribed antisense to the latency region, T3.0 and T1.2 which overlaps with the ORF50 gene and K7.3 that overlaps antisense to PAN [17, 20]. Moreover like most herpesviruses, KSHV encodes a cohort of 12 pre microRNAs (miRNAs) [15, 21]. Recent studies have identified several further more esoteric gene products such as circular RNAs (discussed in a later section) [22–27].

While most ORFs (ORF4-ORF75) show some conservation between γ-herpesviruses and HHVs, KSHV and other related *Rhadinoviruses* encode some genes unique to their genus. These include at least 15 unique protein-coding genes that are denoted by "K" and a number (1-15) (Fig 1.2) [13]. Additionally many KSHV genes, including protein coding but also miRNAs, are orthologous to host genes [28–30]. In summary the structure of KSHV's genome relates to its life-cycle, while the pattern and expression characteristics of the genes that it encodes reflect its bi-phasic replication.

ORF	Alias	Grouping	Full Name/Function
K12A	KapA	Classical Latent	Kaposin A
K12B	KapB	Classical Latent	Kaposin B
K12C	KapC	Classical Latent	Kaposin C
mIRregion	miR-Ks/K-miRs	Classical Latent	viral miRNAs
ORF71	vFLIP	Classical Latent	Viral FADD-like interleukin-1-B-converting enzyme [FLICE/ caspase 8]-inhibitory protein
ORF72	vCycl	Classical Latent	Viral cyclin D2 homologue
ORF73	LANA	Classical Latent	Latency associated nuclear antigen
K1	-	Relaxed_Latent	B cell receptor homologue
K14	vOX2	Relaxed_Latent	Viral CD200 homologue
K15a	LMP1/2	Relaxed_Latent	EBV LMP1/2 orthologue
K2	vIL-6	Relaxed_Latent	Viral interleukin 6 homologue
K5	-	Relaxed_Latent	E3 Ubiquitin ligase
ORF74	vGPCR	Relaxed_Latent	Viral G-protein coupled receptor
ORF75	-	Relaxed_Latent	FGARAT protein
vIRF-3	-	Relaxed_Latent	viral interferon regulatory factor 3
K4.2A	-	Immediate Early	-
K8	bZIP	Immediate Early	Basic leucine zipper protein
ORF45	-	Immediate Early	RSK activator
ORF50	RTA	Immediate Early	Replication and transcription activator
ORF57	MTA	Immediate Early	mRNA export/splicing

Table 1.1: Details on KSHV's classical and relaxed latent and immediate early lytic genes. Table adapted from [31].

ORF	Alias	Grouping	Full Name/Function
K3	-	Early Lytic	E3 Ubiquitin ligase
K4	vCCL2, vMIP-1a	Early Lytic	Viral MIP-1B homologue
K4.1	vCCL3	Early Lytic	viral chemokine (C-C) ligand 2 homologue
K4.2	-	Early Lytic	-
K6	vCCL2, vMIP-1a	Early Lytic	Viral MIP-1a homologue
K7	-	Early Lytic	Mitochondrial membrane protein
OLAP	-	Early Lytic	Origi-l of lytic replication associated protein
ORF10	-	Early Lytic	Regulator of Interferon Function
ORF11	-	Early Lytic	Promotes specialised ribosome biogenesis
ORF16	vBcl-2	Early Lytic	vBcl-2 homologue
ORF17	-	Early Lytic	Protease
ORF17.5	-	Early Lytic	Assembly protein
ORF18	-	Early Lytic	Late gene regulation
ORF2	-	Early Lytic	Dihydrofolate reductase
ORF21	vTK	Early Lytic	Thymidine kinase
ORF24	-	Early Lytic	Viral pre-initiation complex protein
ORF29A	-	Early Lytic	packaging protein
ORF29B	-	Early Lytic	packaging protein
ORF30	-	Early Lytic	Late gene regulation
ORF31	-	Early Lytic	Nuclear and cytoplasmic
ORF34	-	Early Lytic	-
ORF35	-	Early Lytic	-
ORF36	vPK	Early Lytic	Serine protein kinase
ORF37	SOX	Early Lytic	Host shut-off factor
ORF38	-	Early Lytic	Myristylated protein
ORF4	KCP	Early Lytic	Complement binding protein
ORF40	-	Early Lytic	Helicase
ORF40A	-	Early Lytic	-
ORF41	-	Early Lytic	Primase
ORF44	-	Early Lytic	Helicase
ORF46	-	Early Lytic	Uracil deglycosylase
ORF48	-	Early Lytic	-
ORF49	-	Early Lytic	Activates JNK/p38
ORF4A	-	Early Lytic	-
ORF54	-	Early Lytic	dUTPase/Immunmodulator
ORF56	-	Early Lytic	DNA replication
ORF59	-	Early Lytic	Processivity factor
ORF6	-	Early Lytic	ssDNA Binding protein
ORF60	-	Early Lytic	Ribonucleoprotein reductase
ORF61		Early Lytic	Ribonucleoprotein reductase
ORF62	TRI1	Early Lytic	Viral capsid triplex
ORF63	-	Early Lytic	NLR homologue
ORF64	-	Early Lytic	Deubiquiti-se
ORF66	-	Early Lytic	capsid
ORF67	-	Early Lytic	Nuclear egress complex
ORF67A	-	Early Lytic	-
ORF69	-	Early Lytic	BRLF2 Nuclear egress
ORF7	-	Early Lytic	virion protein
ORF9	-	Early Lytic	DNA Polymerase
PAN	Nut1	Early Lytic	Late gene expression
vIRF-1	-	Early Lytic	viral interferon regulatory factor 1
vIRF-2	-	Early Lytic	viral interferon regulatory factor 2
vIRF-4	-	Early Lytic	viral interferon regulatory factor 4

Table 1.2: Details on KSHV's early lytic genes. Table adapted from [31].

ORF	Alias	Grouping	Full Name/Function
K8.1	-	Late Lytic	Glycoprotein
K8a	-	Late Lytic	-
ORF19	-	Late Lytic	Genome packaging protein
ORF20	-	Late Lytic	ORF59-associated protein
ORF22	gH	Late Lytic	Glycoprotein H
ORF23	-	Late Lytic	Glycoprotein associated with late gene trancription
ORF25	MCP	Late Lytic	Minor capsid protein
ORF26	TRI2	Late Lytic	Viral capsid triplex
ORF27	-	Late Lytic	Glycoprotein
ORF28	-	Late Lytic	EBV BDLF3 homologue
ORF32	-	Late Lytic	Tegument protein
ORF33	-	Late Lytic	Tegument protein
ORF39	gM	Late Lytic	Glycoprotein M
ORF42	-	Late Lytic	Tegument protein
ORF43	-	Late Lytic	Capsid portal protein
ORF47	gL	Late Lytic	glycoprotein L
ORF4B	-	Late Lytic	-
ORF52	-	Late Lytic	Tegument protein
ORF53	gN	Late Lytic	Glycoprotein N
ORF55	-	Late Lytic	Tegument protein
ORF58	-	Late Lytic	EBV BMRF2 homologue
ORF65	SCP	Late Lytic	Smallest capsid protein
ORF68	-	Late Lytic	Glycoprotein
ORF70	-	Late Lytic	thymidylate synthasee
ORF8	gB	Late Lytic	Glycoprotein B

Table 1.3: Details on KSHV's late lytic genes. Table adapted from [31].

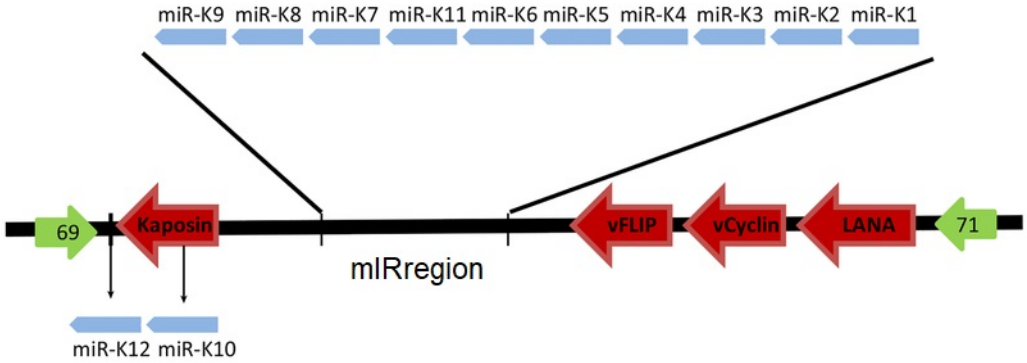


Figure 1.3: Genomic arrangement of KSHV miRNAs. Figure adapted from [32].

1.2.2 Unique coding sequence annotation

Many KSHV and herpesviral transcripts show extensive overlap and complex regulatory schemas, with differential promoter usage, splicing and antisense transcription [15, 21]. For example ORF57, vIRF-2, -3 and -4 and ORF50 are spliced, while anti-sense transcription occurs from a locus overlapping with K1-11, K4.1, PAN, ORF3, 50, 58, 59, 71-3 [15]. Previous annotations used for RNA-Sequencing (RNA-Seq) didn't account for this, however a pair of recent studies have aimed to further investigate this and provide an annotation and associated quantification schema in order to mitigate some of these issues and provide high-confidence viral gene expression estimates [15, 17, 21, 33]. Key to this is the delineation of "unique coding sequence" (UCDS) features that shrink the canonical ORF annotations to be greater than the RNA-Seq library read-length apart, to eliminate ambiguous overlap between annotated features [15, 17]. Such feature annotations have

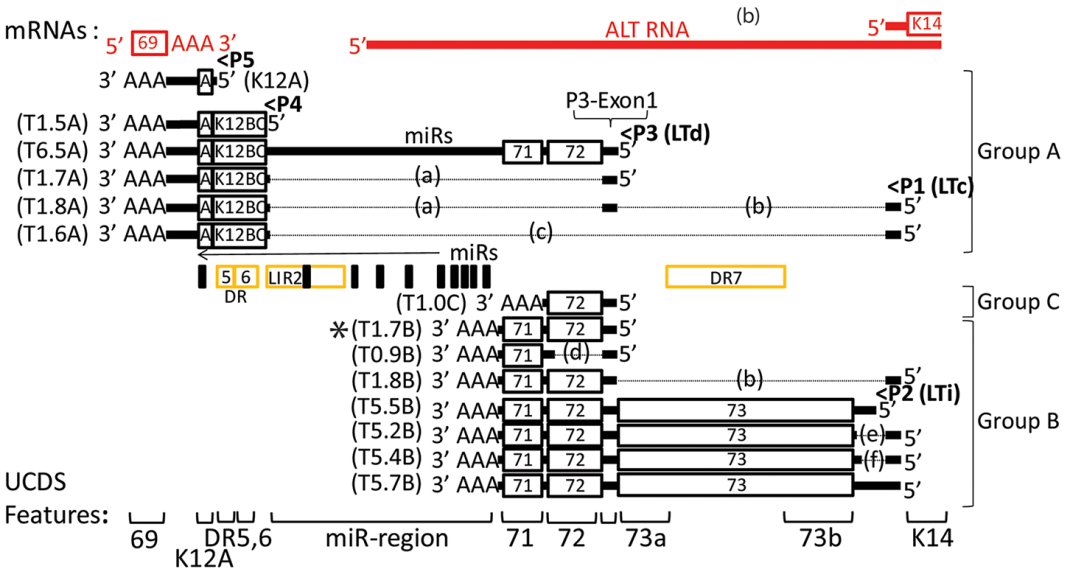


Figure 1.4: Genomic and transcript structure of the latency locus and associated UCDS feature annotations [15]. Figure adapted from [15].

been developed for KSHV and applied to interrogate viral gene expression in a panel of cell lines as well as KS lesion tissue, while similar annotations have been developed for EBV [15, 18, 33].

Many of KSHV's transcribed mRNAs are polycistronic and such transcripts can be of differing lengths, thus comprising differing numbers of ORFs and encoded features. This includes the latency locus where differential promoter usage of the P1, P3 and P5 promoter drivers differential inclusion of ORF71, ORF71 and ORF73 (Fig 1.4) [15]. This can result in ambiguity in the attribution of mapped RNA-Seq reads to ORFs in polycistronic transcripts, particularly as the same loci can be transcribed with different combinations of ORFs [15]. Importantly, Bruce *et al.,* 2017 observed a step-wise gradient of aligned reads between successive ORFs in polycistronic transcripts, implying redundancy in the expression measurements of downstream ORFs relative to those upstream [15, 17, 34]. The UCDS schema accounts for this as the measured counts for each downstream ORF encoded were subtracted from those upstream [15]. Moreover, in the case of the P1, P3 and P5 latency promoters, the abundance of exons transcribed immediately downstream are used in the UCDS schema as proxies for their activities [18].

Spliced or overlapped transcripts are also modified in order to account for more complex structures and to limit ambiguity. For example, the Kaposin locus can be spliced to form 3 main isoforms, K12A, B and C [15]. This involves the inclusion of the DR5 and DR6 repeat sequences as proxies for K12B and K12C respectively. However there remains some ambiguity in the quantification of reads for these two genes when measuring their expression by the abundance of the DR5/6 repeat sequences (Fig 1.4) [15].

Interestingly, the structure and location of most of the virally encoded miRNAs makes a measurement of their abundances quantifiable from long-read RNA-Seq data. This is because the miRNAs are transcribed from 2 genomic loci: one intronic encoding 10 premiRNAs (20 miRNAs) and one within an exon of the Kaposin gene, encoding 2 pre-miRNAs (5 miRNAs) (Fig 1.3) [32]. While the latter 2 miRNAs are difficult to disambiguate due to their confounding with the Kaposin locus, Bruce *et al.,* 2017 provided an annotation that corresponds to the intron the remaining 10 viral miRNAs (miR-K1 to K9 and miR-K-11) are encoded by, enabling a proxy measurement for their collective abundance [15]. This annotation as well as those previously described allow for an expanded, as well as more accurate and reliable, estimation of the expression of viral genes from short read RNA-Seq data.

1.2.3 Virion Structure

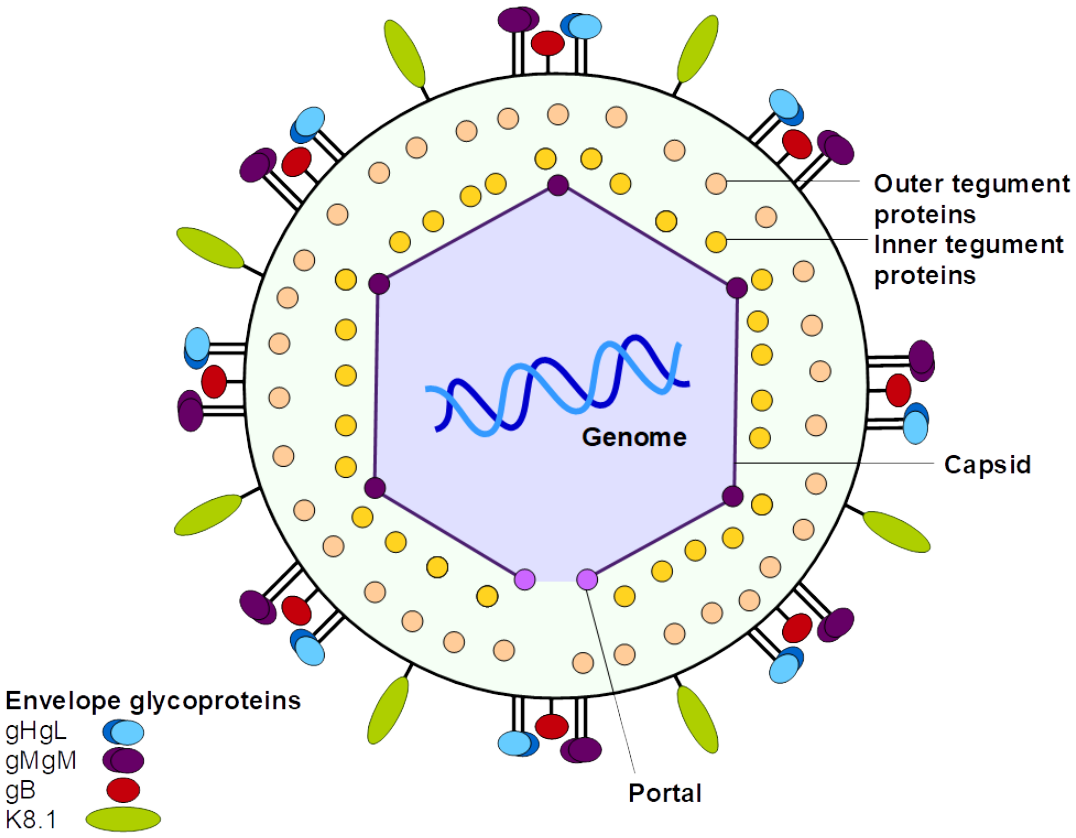


Figure 1.5: KSHV virion capsid structure.

The structure of KSHV's virion is reminiscent of most herpesviruses. This includes a four-layered structure comprising the genomic core, encased in a icosapentahedral protein capsid, surrounded by a semi-amorphous protein tegument and a final lipid bilayer studded with glycoproteins (Fig 1.5) [2].

Notably the capsid of KSHV comprises the same overall structure as HSV-1 and HCMV despite differing sequences of the proteins that constitute its capsomers [35]. 5 ORFs (ORF25, ORF26, ORF62, ORF65 and ORF43) encode for 4 heteromutlimeric capsomer proteins that comprise the viral capsid [13]. These capsomers comprise hexamers and pentamers of the major capsid protein (MCP, ORF25) which are held together with the aid of the smallest cap protein (SCP, ORF65), alongside heterotrimeric triplexes comprising one Tri2 (minor capsid protein, ORF26) and two Tri1 (ORF62) proteins [36]. MCP Pentamers are present at the capsid vertices and MCP hexamers are present at the faces of the capsid, however one pentamer in every capsid is replaced with a portal complex that comprises dodecameric ORF43 protein [36]. This capsid acts to contain and protect the viral dsDNA genome. Additional studies have indicated that various RNA species are present in the viral capsid as well, including mRNAs as well as viral and host miRNAs, "unusual small" RNAs (usRNAs) and circular RNAs (circRNAs) have been identified [26, 37].

The viral capsid is surrounded by a tegument layer, which contains at least 13 confirmed viral proteins, including a pentameric capsid associated tegument complex (CATC) that comprises ORF19, 32 and 65, alongside ORF7, 11, 21 33, 45, 50, 52, 55, 63 and ORF75 in the general tegument (Fig 1.5) [13]. While some of these proteins have historically had poorly characterised functions, work in recent years has indicated some of these as being involved in a range of key processes. This includes ORF11 with recently described roles in modulating specialised ribosomes, ORF75 which dissipates anti-viral ND10 bodies and ORF45 which has proposed roles in suppressing TLR signalling, promoting viral envelopment and activating ribosomal protein S6 kinase [38, 39]. Various host proteins have been observed to be incoorporated into the KSHV virions, including HSP70/90, which are important for the formation of viral replication and transcription compartments [40]. This tegument layer was canonically considered amorphous, however more recent research has suggested at least two semi-structured layers; the inner "capsid-associated" and outer "membrance-associated" layers (Fig 1.5).

The virus is shrouded in a lipid bilayer derived from the cellular nuclear membrane that forms an envelope around the tegument. Like most viruses, this layer is embedded with glycoproteins that mediate interactions with host cells. These membrane proteins are composed of ORF8, K8.1A/B, ORF22, ORF47, ORF39, ORF53, ORF28, ORF66 [13] (Fig 1.5). Further proteins found in KSHV virions via proteomics include ORF9, 23, 35, 48, 58, 72, K3, K9 (vIRF-1), K10 (vIRF-4) and K10.5 (vIRF-3) [41]. These may be bound to the genome and capsid or associated proteins or present in the viral tegument. In general,

KSHV's virion structure is reminiscent of many dsDNA viruses, including herpesviruses and is relatively well understood.

1.2.4 Cell tropisms

KSHV can infect a range of cell types, including endothelial and epithelial cells, keratinocytes, B cells and monocytes. B cells in particular are a major target for KSHV and are considered an important reservoir of infection and transformation, that directly or indirectly contirbute to KSHV's pathologies.

In vitro infection does not necessarily imply biologically relevant infection of cells *in vivo* and much cell line work doesn't immediately translate to *in vivo* settings. For example, while not historically considered major targets, KSHV has been shown to enter both CD4+ and CD8+ primary human tonsillar T cells, resulting in a form of abortive lytic replication and no infectious virion production [42]. Conversely, KSHV's DNA has been isolated from rare T cell lymphomas, implying pathogenically-relevant evidence for KSHV's association with T cells [42]. Moreover, KSHV can infect fibroblasts and keratinocytes *in vitro*, however limited detection of KSHV antigens in these cell types have been observed in KS lesions [42].

1.2.5 Primary infection and establishment of latency

KSHV entry into cells is a complex and highly coordinated process. KSHV first binds to heparin sulfate via several of its envelope glycoproteins (gB, gHgL, K8.1 and ORF4), followed by binding to integrin ($\alpha 3\beta 1$, $\alpha V\beta 3$ and $\alpha\delta\beta 5$) and xCT (Fig 1.6) [37, 43]. Engagement of these host receptors stimulates the recruitment of various host factors that modulate a suite of cellular signalling pathways to facilitate successful entry and establishment of infection, including PI3K, Rho GTPases, Src, ERK, FAK, PKCδ, NF-kB, c-Cbl, GEF-C3G, CIB1, Crk, p130Cas, ROS and Dia-2 [37]. These induce several transcription factors, namely ERK1's downstream effectors AP-1, STAT1, MEF2, c-Myc and ATF2, alongside NF-kB and NRF2 [44]. Once efficient binding has been initiated, KSHV enters cells via endocytosis, specifically clathrin-mediated or via macropinocytosis and membrane blebbing. In order to properly enter the cell, KSHV must be released from these vesicles via a poorly defined process of fusion of it and the vesicle's membranes [45].

Once the virus has entered the cytosol, several steps proceed that results in the virus's deposition into the nucleus, the site of both latent and lytic replication. Firstly, tegument proteins and transcripts are released, which include ORF50, ORF73 and ORF59, which

have been hypothesised to play roles during these initial stages of *de novo* infection by modulatin ghost and viral processes [13]. Next it is transported to the nucleus along microtubules. To facilitate entry into the nucleus, the viral genome undergoes extrusion through the capsid portal complex via the nuclear pore in a process predominantly mediated by ORF19 [46].

Like cellular genomes, epigenetics plays a key role in regulating KSHV's gene expression and life cycle. However KSHV's genomes are epigenetically naive upon entry and must undergo rapid chromatinisation, with wave of active H3K4me3 and H3K27ac mark depositions which enables an abortive gene expression program that resembles early lytic replication [44]. After 24-72hrs, KSHV's epigenetic profile shifts towards a repressive state with enrichments of repressive H3K27me3 and H2AK119 ubiquitinylation marks, resulting in latency [47]. One contributor to this is the recruitment of the polycombe repressive complex (PRC) in particular EZH2, by the latency associated nuclear antigen (LANA), encoded by ORF73, among other factors that facilitate the silencing of most lytic genes via addition of repressive histone acetylation and methylation marks [48, 49]. Cohesion complexes (formed from SMC1/3 heterodimers and accessory proteins) and CTCF play key roles in recruiting PRC components and overal transcriptional silencing of the nascent viral genome [50]. Importantly while most of the viral's lytic genes gain repressive modifications, the latency locus remains as euchromatin with an enrichment of active H3K4me3 and H3K27ac marks [48]. The eventual end result of this process is the repression of lytic gene expression, which facilitates the eventual shift towards latency exhibited $\sim$72hrs post-infection.

Genome circularisation is known to be essential for episome maintenance during latency, as well rolling circle replication during lytic replication [51]. How this process is accomplished is not fully understood however EBV is know to make use of the activation of DNA damage response (DDR) signalling [52]. An additional key step in establishing latency is the tethering of the viral episome to host chromosomal centromere complexes by LANA [53]. To facilitate this, LANA's N-terminal domain binds H2A and H2B histones, while its DNA binding domain (DBD) binds to the TR regions of the viral genome [15]. Additional chromatin- and chromosome-associated proteins are known to be important, including BRD2 and BRD4, the centromere protein CENPF, cohesin and CTCF [50]. Recent work has indicated that episome binding is localised to ChAHP (comprising CDH4, ADNP and HP1γ)-rich regions of host chromosomes via LANA-CHD4 binding to the proximal terminal repeat regions as well as the PAN RNA promoter [53]. Genome circularisation is particularly key for proper tethering of the viral episome to host chromosomes, which fa-

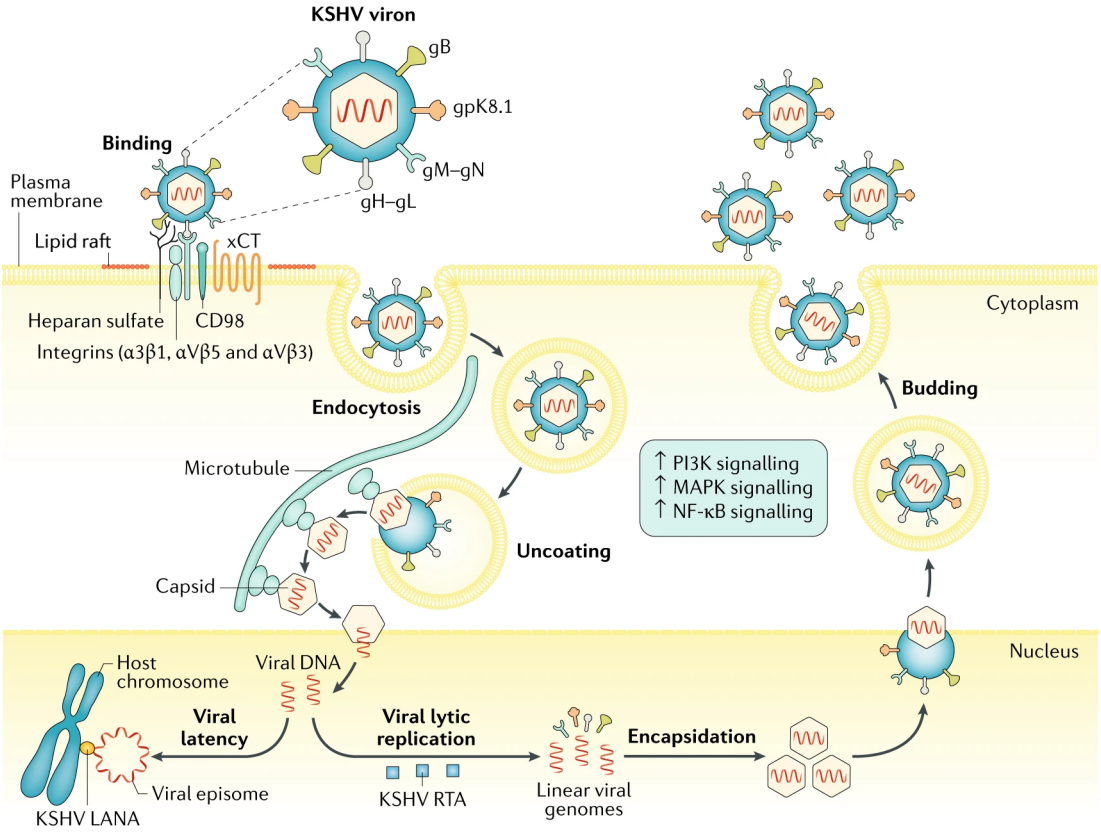


Figure 1.6: Life cycle of KSHV. Figure adapted from [11].

cilitates replication and segregation of the episome with host cellular genomic replication and daughter cell segregation, respectively.

1.2.6 Latency

It is largely because of latency that KSHV is able to persist, both in cell culture and in infected complex organisms as it facilitates the duplication and segragation of the viral episome alongside the host's genome, alongside the evasion of the host's immune system (Fig 1.6) [47]. Genes believed to be expressed during latency can be split into the "classical" latent genes (ORF71, ORF72, ORF73, kaposin (K12A, B & C) and 25 miRNAs) which are encoded by the latency locus and have been strongly associated with latency. However further canonically lytic genes have been observed to be co-expressed with the classical latent genes during latency [15]. Many of these lytic genes are encoded near the latency locus and include ORF74, ORF75, K14, K15, K1, K2, K5 and vIRF-3, depending on the exact cell line and study (Fig 1.2) [15, 18]. These genes (hereby termed "relaxed latent" genes) have been suggested to be expressed via a "leaky" or "relaxed" expression of some lytic genes during latency or during bouts of spontaneous lytic abortive replication within predominantly latently infected cell populations [14]. The status of these genes as truly expressed during latency remains unconfirmed, however.

Of the latency genes ORF73 (LANA) has pleiotropic roles in latently infected cells, as well as roles in proper viral DNA replication and persistence. These include targeting signalling pathways and host TFs, such as NF-kB, JAK-STAT and various aspects of epigenetic machinery (Fig 1.7) [54]. Importantly, it directly inhibits the expression and activity of the key instigator of lytic replication, the replication and transcription activator (RTA) protein (ORF50), thus promoting latency by inhibiting the induction of lytic replication (Fig 1.7) [14]. It also represses lytic gene expression via DNA methyltransferase (DMNTs) or lysine demethylases (KDM)-mediated repression of lytic gene expression and lytic replication via MDM2-mediated degradation of p53, a key factor for proper lytic viral genomic replication (Fig 1.6) [47]. It also indirectly represses lytic gene expression via recruitment of the SUMO-2 histone SUMOylation complex thus regulating chromatin conformation and recruiting Krüppel-associated box domain-associated protein 1 (KAP1) to lytic promoters, facilitating their transcriptional repression [55]. Additionally, LANA directly competes with RTA to bind RBP-Jk, a key co-factor for RTA's transactivation activity [47].

Other latent factors contribute to latency. For example vFLIP (encoded by ORF71) activates the inhibitor of kB kinase (IKK) which subsequently phosphorylates the inhibitor of kB (IkB) complex to facilitate its degradation, promoting constitutive activation of NF-kB to induce the expression of anti-apoptotic genes (Fig 1.7) [56]. One consequence of this NF-kB activation is to promote the expression of the key PRC component H3K27 methyltransferase EZH2, thus promoting latency [47]. vFLIP also inhibits autophagy in latently infected lymphoctes and endothelial cells, which limits cellular senescence [56]. Additionally, the viral cyclin D2 homologue, vCyclin (encoded by ORF72) helps maintain (and establish) latency as it phosphorylates nucleophosmin (NPM) in latently and *de novo* infected cells, which facilitates recruitment of the histone deacetylase HDAC1 to promote transcriptional repression (Fig 1.6) [56]. The viral miRNAs also have known roles in suppressing lytic replication; by directly targeting viral transcripts such as RTA (miR-K7/K9), or by indirectly targeting pro-lytic host factors [32]. The role of the latent genes in oncogenesis are discussed in later sections (Section 1.2.10).

1.2.7 Lytic Replication

While infection eventually defaults to latency maintained by LANA, a range of triggers result in increased KSHV gene expression that can result in the induction of lytic replication. For example, environmental changes can drive this, including hypoxia and oxidative stress, co-infection and inflammatory and growth factors [57]. Additional changes to the

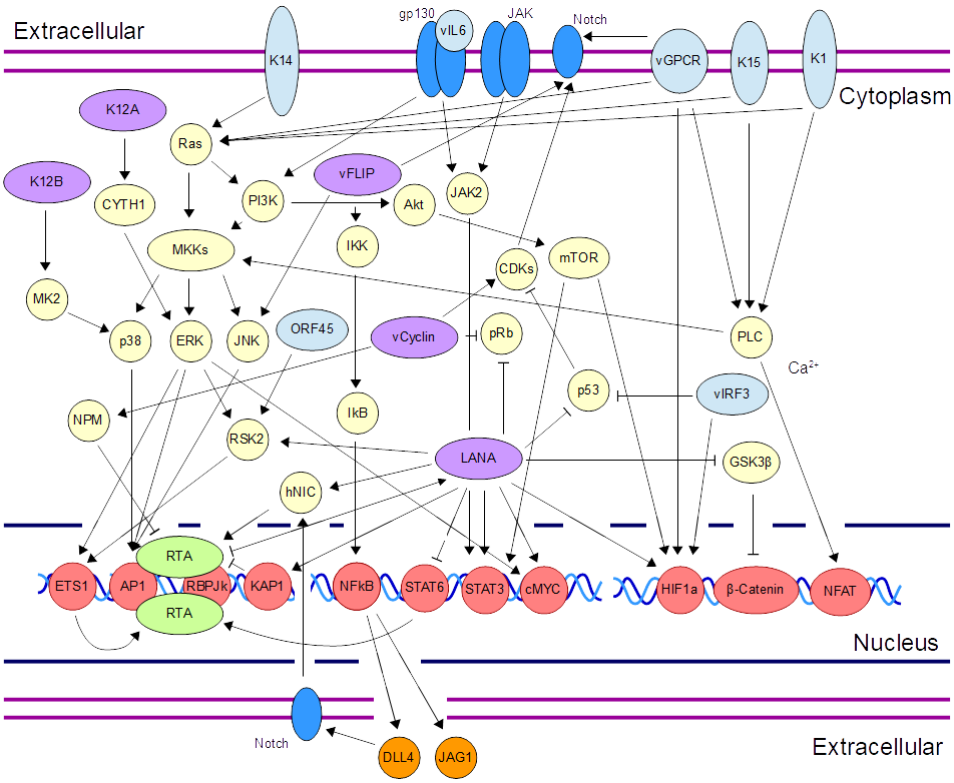


Figure 1.7: Signalling pathways targeted by KSHV.

cellular signalling and contexts can activate lytic replication such as calcium ion influx, MAPK, PKA and PI3K pathway activation and adrenergic signalling, alongside epigenetic modifications that promote de-repression of the viral episome [53, 58, 59]. Co-infection can also trigger lytic replication, including the presence of bacterial metabolites, HIV-1 proteins such as Tat and HSV-1/2, HCMV, HHV-6/7 and HPV co-infection [51]. Overall, these triggers shift the balance between LANA and RTA activity in favour of RTA, triggering the lytic gene expression cascade [14, 47, 51].

As LANA is to latency, RTA is the key orchestrator of lytic replication and usually considered the first transcript to be expressed during it and as such is its major driver. Accordingly its expression in latently infected cells is sufficient for induction and progression of a productive lytic replication cycle [14, 51]. To facilitate this it can activate the expression of many lytic viral genes (including itself and many host genes) via three main mechanisms (Fig 1.8) [51, 60]. The first is direct binding to and activation of transcription via its N-terminal DNA binding domain binding to RTA response elements (RREs) in promoters, while its C-terminal host factor recruitment transactivation domain, which recruits key proteins that facilitate transcription, including RNAPII (Fig 1.8) [60]. Secondly, it can complex with a range of co-factors including the downstream effectors of Notch and MAPK signalling, RBP-Jk and AP-1, respectively, which enables it to bind and induce the transcription of genes whose promoters don't contain RREs (Fig 1.8) [61, 62]. Third,

RTA also has E3 ubiquitin ligase activity that facilitates its modulation and degradation of many transcriptional repressors including the host-encoded Zinc finger protein of the cerebellum 2 (ZIC2), Hey1 and ID as well as LANA (Fig 1.8) [63, 64]. This multi-pronged regulatory schema emphasises the potency of RTA in driving lytic gene expression and replication.

Upon sufficient expression and activity of RTA the lytic cycle is induced. Classically this proceeds in a triphasic cascade of the overlapping expression of sets of genes defined as the "lytic cascade", comprising the immmediate early (IE), early (E) and late (L) lytic genes (Fig 1.9). IE genes were originally defined by their transcription not being cyclohexamide-resistant and thus independent of *de novo* translation [65, 66] (Fig 1.9). RTA is the archetypal IE lytic gene while other genes such as K8 (bZIP), PAN, ORF45 and ORF57 are sometimes considered IE genes, however this is debated and context-dependent [15]. Otherwise these latter genes are usually considered E lytic (or sometimes delayed early (DE) lytic) genes and prime the virus and cell to undergo viral transcription and DNA replication, as well as immunomodulation (Fig 1.9) [14, 18, 67]. Late genes tend to mediate viral assembly and thus encode capsid and other structural proteins as well as those involved in viral egress and release from the cell (Fig 1.9) [18]. Note that such classifications (E, IE and L genes) generally refer to when expression of these viral genes is detectable, not when they are maximally expressed. However there is overlap as IE, E and L lytic genes tend to be maximally expressed 0-8hrs, 8-24hrs and 48-72hrs post lytic induction, respectively, but this varies by cell type and broader environmental context [18].

ORF57 (MTA) is a key lytic gene that is required for proper lytic progression and infectious virion production [68, 69]. It has key roles in proper splicing and nuclear export of intronless viral transcripts and through this, evasion of nonsense-mediated decay (NMD) and subsequent immune responses or apoptosis [68, 70]. Interestingly in more recent years, ORF57 has been proposed to promote the biogenesis via back-splicing of host and viral circular RNAs (circRNAs) [27, 71] (discussed further in Section 1.4.6).

PAN is one of the earliest and the most highly expressed lytic transcripts in cells undergoing lytic replication. Several debated roles have been proposed for its function. These include sequestering RNAPII in proximity to the viral genome or acting as a scaffold to recruit UTX/JMJD3 to facilitate the removal of repressinve H3K27me3 marks [72, 73]. Alternatively it has been suggested to directly promote lytic gene expression by suppressing LANA activity either by directly binding and sequestering it and, most recently, competing with LANA to bind CHD4 which promotes dissociation of the viral episome

from host chromosomes to facilitate lytic gene expression [13, 53]. An additional viral lncRNA, T3.0, encodes small ORFs that can be translated to generate the micropeptide vSP-1 that binds RTA and prevents its ubiquitinylation and subsequent degradation [74].

As previously stated, early lytic genes encode proteins that facilitate lytic genome replication. For example, those that result in the generation of ribonucleotide and nucleotide substrates for transcription and genomic replication, respectively (Table 1.2) [15, 21, 66]. Once enough of these genes have been expressed into proteins they assemble to replication complexes that begin the process of lytic genome replication [51]. Like many DNA viruses this proceeds via a rolling-circle mechanism whereby nascent genomes are synthesised as periodic concatamers that must be cleaved prior to encapsidation. This replication process initiates at one of two *ori-Lyt* sites [41]. Interestingly like LANA, RTA also facilitates the initiation of replication as it is recruited to *ori-Lyt* by an RRE, alongside bZIP, to promote lytic replication [75, 76].

One notable early gene is ORF37, which encodes SOX, termed the "host shut-off factor". It is an alkaline endonuclease that cleaves cellular transcripts via a loose degenerate RNA sequence motif [77]. In fact, SOX's primary function is to degrade a majority (75%) of host transcripts in order to promoter immune evasion as well as favour expression of viral genes [78–80]. However SOX also has DNase activity is also believed to cleave viral genomes from newly synthesised concatemers [81].

Early genes also encode a complex termed the late gene transcription complex that faciliates the expression of late lytic genes [82]. Accordingly late lytic genes were originally defined as dependent on viral genome replication as evidenced by the sensitivity of their expression to gangcyclovir, which inhibits viral genome replication, and phosphonoacetic acid (PAA), which inhibits viral DNA polymerase activity (Fig 1.9) [83]. Therefore the functions that they encode, virion assembly and egress, capsid etc, are generally believed to not occur until newly synthesised genomes are present to be packaged.

The later stages of lytic replication, namely capsid assembly, virion egress and release are poorly defined. In terms of assembly, KSHV capsomers are believed to self-assemble in a similar manner to other herpesviruses (Fig 1.6) [51]. Virion egress is assumed to follow several steps common to herpesviruses, whereby it exits the nucleus, undergoes tegumentation (formation of the tegument) in the cytoplasm, gains an envelope and exit cells (Fig 1.6) [51]. The end result of effective lytic replication is the formation of new infectious virion particles which exit the cell to infect further naive cells. The role of lytic genes in oncogenesis are described in a later section (Section 1.2.12).

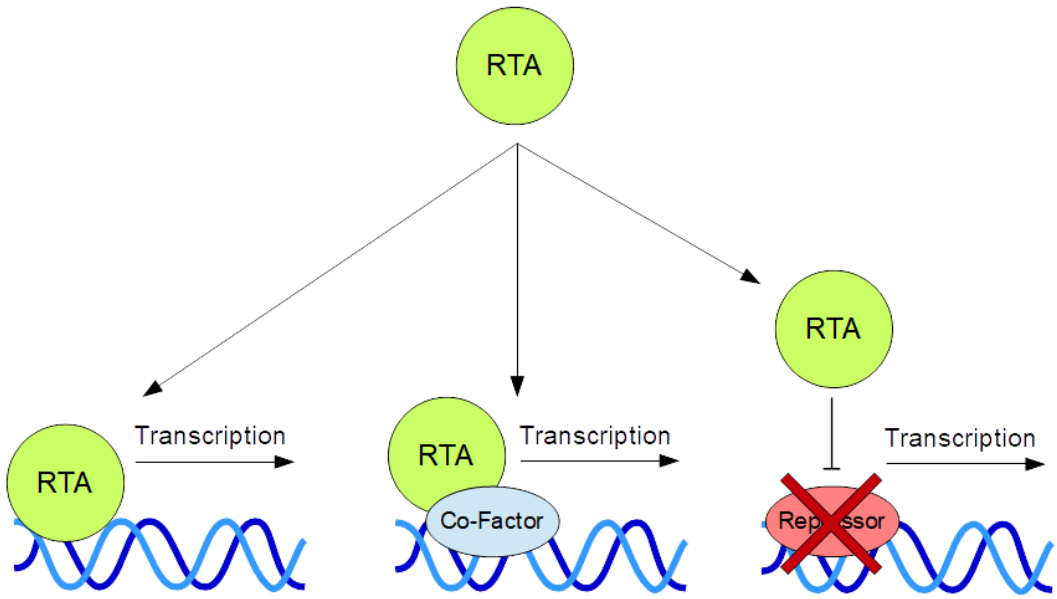


Figure 1.8: Mechanisms of KSHV RTA's regulation of gene expression.

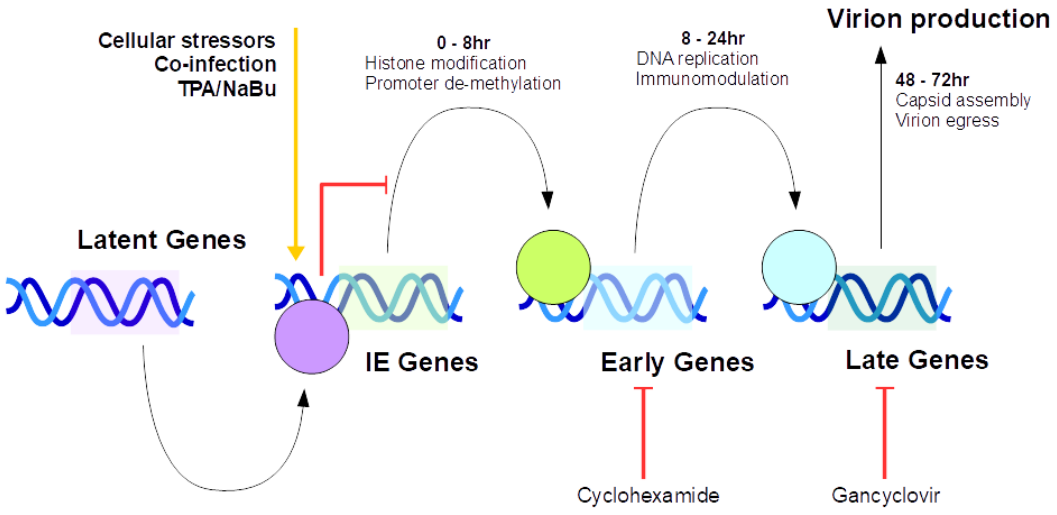


Figure 1.9: KSHV's lytic gene cascade.

1.2.8 Study of KSHV *in vitro*

The study of KSHV's life cycle and general molecular biology has predominantly occurred in transformed cell lines that stably maintain the viral episome. *In vitro* various methods have been developed to induce lytic replication in an experimental setting, including the chemical inducers 12-O-tetradecanoyl-phorbol-13-acetate (TPA) and sodium butyrate (NaBu) [51]. A more developed system is the TREx-RTA model, which comprises transformed cell lines containing KSHV's latent episome alongside an ORF50 (RTA) gene cassette under the control of a Doxycycline (Dox)-inducible promoter. Using this sys-

tem, introduction of Dox to the cell culture triggers ORF50 expression, thus facilitating the reactivation of latent KSHV into lytic replication [84]. These methods and variations on them have allowed for the reproducible induction of lytic replication within controlled laboratory settings, allowing many facets of KSHV's life cycle and pathogenesis to isolated and elucidated.

1.2.9 Infection state heterogeneity

While latent and lytic gene expression programs are often considered mutually exclusive, there is considerable overlap in terms of the expression of canonically latent and lytic genes, both within-cells and at the cell population level. Namely that most latent genes are still expressed during the lytic cycle, while LANA in particular has a long half life of several days and is transcribed from a second, lytic-specific promoter [55]. Moreover in populations of cells, both in cell culture and *in vivo* tissues, different cells can undergo latency or lytic replication, with the latter believed to be important for the persistence of KSHV in the cell population as a whole, by enabling *de novo* infection of further cells [51].

Recent single cell RNA-Seq (scRNA-Seq) studies on PEL and epithelial cell lines, primary cells and a 3D organoid model have emphasised such heterogeneity, with cells clustering into latent, IE, E and L lytic gene expression, as well as various apparent hybrid latent/lytic gene expression programs [85**?** –89]. Working in KSHV-infected iSLK.219 cells, one study by Tabtieng *et al.,* 2022 identified a sub-population of IFN-1β-expressing early lytic cells, with particularly high K5 expression and an enrichment for NF-kB transcription family members [89]. Interestingly this reflected a previous study on Herpes simplex virus (HSV) in which a sub-population of IFN-producing abortive lytically replicating cells were observed [6]. The 3D structure of tissue has also been shown to be important as by using a 3D air-liquid interface (ALI) organoid model of oral epithelial cells, Jung *et al.,* 2022 greater expression of lytic genes in the superficial layers while the basal layers exhibited greater latent gene expression, reminiscent of HPV infection of the epidermis [85].

While no scRNA-Seq studies have been performed from *in vivo* biopsies, variable infection programs have been observed from biochemical and bulk RNA-Seq studies into KSHV-induced pathologies, including KS, with lesions showing predominantly latent, lytic or mixed expression profiles [67]. This did show some dependency on the patient and type of infection as, at least in Lidenge *et al.,* 2020, nearly all endemic samples exhibited predominant lytic replication while most epidemic samples showed lytic or mixed. This likely relates to the absence of HIV-1 as, interestingly, HIV-1 co-infection is known to stimulate lytic replication and HIV-1 and KSHV gene products are known to synergis-

tically contribute to transformation and associated processes [90–92]. Generally «5% of infected cells in KS undergo spontaneous lytic viral reactivation [13]. Such studies and the ones previously discussed emphasise the additional complexity of studying KSHV in an *in vivo* context and the gulf between such contexts and 2D cell line culture models.

1.2.10 Latent oncogenes

KSHV's canonical latent genes are the most strongly associated with oncogenesis and as such they are believed to be the predominant drivers of KS, among the other virally-induced cancers. Several contextual clues for this come from the detection of their expression in and association with KSHV-driven pathologies, including KS lesions and PEL cells [67, 93, 94]. Moreover the ectopic expression of at least ORF71, ORF72 and ORF73 has been shown to induce tumours in animal models [95–97]. However how they facilitate this varies, dependent on their interactions and below details how each classical latent gene product has been suggested to contribute to oncogenesis.

LANA As the major promoter of latent persistence as well as having extensive capacity to modulate pro-growth and mitogenic pathways, LANA is assumed to be a great contributor to the oncogenic potential of KSHV. One aspect of this is likely its modulation of aspects of the DNA damage response (DDR) by inhibiting the major tumour suppressors p53 and pRb (Fig 1.7) [52]. Interestingly, the activity of LANA and HRas have been shown to be mutually enhancing, acting to promote transformation of primary rat cells [98]. LANA can additionally induce the modification of proteins, including itself [47]. This in particular enables several of its interactions with factors involved in mitotic process and DNA binding [47, 54]. For example LANA binds and inhibits GSK3β leading to accumulation of β-catenin in the cytoplasm that facilitates increased LEF-mediated expression of c-Myc and cyclin D, promoting cellular growth (Fig 1.7) [52, 99]. GSK3β and pRB are G1/S checkpoint proteins and as such LANA's inhibition of these proteins modulates transition through this boundary [98]. Moreover LANA itself is regulated by by a range of cellular kinases. For example, LANA's interaction with chromatin and repression of lytic gene expression is promoted by phosphorylation by glycogen synthase 3 (GSK3), casein kinase 1 (CK1) and ribosomal S6 kinase (RSK) (Fig 1.7) [47]. As such it effectively sits at a nexus between host cellular signalling pathways and viral processes. Overall, LANA is perhaps the most well-characterised viral oncogene and its role in KSHV-driven cancers is paramount.

vCyclin KSHV vCyclin (ORF72) is a cellular cyclin D2 homologue that can complex with cyclin-dependent kinases (CDKs) to form cell cycle-modulatory and thus pro-oncogenic complexes, as well as modulating various onocgenic processes (Fig 1.7) [99]. Like LANA, it can inactivate pRb by phosphorylation in complex with CDK6 and possibly p53 in complex with CDK9 (Fig 1.7) [52]. Moreover, the vCyclin-CDK6 complex is able to further target proteins involved in autophagy and apoptosis (Bcl-2), chromatin remodeling (histone H1), genome stability and replication (NPM, CDC6, ORC1) [100]. The vCyclin/CDK6 complex can activate Notch to promote cell cycle progression and lymphamogenesis (Fig 1.7) [101]. Thus ORF72 is a knonw oncogene with potent transforming potential.

vFLIP ORF71 (vFLIP) is suggested to contribute to oncogenesis directly via modulation of cellular processes and indirectly via its immunomodulation of cells and whole-organisms. For example it contains death domains that can interact with caspase 8 as well as FADD to prevent caspase 8's cleavage, thus inhibiting apoptosis [52]. Similarly its persistent activation of NF-kB further contributes to apoptotic evasion as it prevents TNFα-induced apoptosis (Fig 1.7) [52]. vFLIP is also proposed to contribute to excessive endothelial-to-mesemchymal transition (EndMT) of infected endothelial cells via activation of Notch signalling, which may contribute to the development of KS (Fig 1.7) [87, 99]. Moreover it inhibits autophagy by prevening ATG3 processing of LC3 to mitigate oncogene-induced senescence (OIS) in response to vCyclin's activity [102]. Therefore ORF71's proposed pro-oncogenic roles is believed to be mediated by its immunomodulation as well as inhibition of apoptotic and autophagic processes.

Kaposins K12A and B are some of the most abundantly detected genes in KS lesions and PEL cells and have long been observed to have pro-oncogenic roles [18, 52, 67, 93]. K12A has been suggested to activate the ERK/MAPK pathway via interaction with the ARF guanine nucleotide exchange factor cytohesin-1 as well as modulating adhesion junctions and the cytoskeleton (Fig 1.7) [103]. However there is ambiguity as miR-K10a is embedded within K12A's coding sequence and found to target p27 and p120 to dysregulate the cell cycle and adherens junctions, respectively, thus these miRNAs and not Kaposin A proteins may responsible for the observed transformting potential of the K12A (Fig 1.3) [104].

K12B has been shown to modulate the host cell by impacting mRNA stability. It impedes the degradation of cytokine mRNAs by directly binding and activating MK2, a target of p38 MAPK, with immunomodulatory consequences (Fig 1.7) [105, 106]. Moreover it stabilises PROX1, the master regulator of lymphatic endothelial cell differentiaiton and has been

suggested to be critical in driving the shift towards a lymphatic phenotype characteristic of KSHV-infected endothelial cells, both *in vitro* and in KS tissue [107]. Consequently K12B is sufficient but not necessary to induce KS spindle cell-like morphology including tubule and actin stress fibre formation in cultured endothelial cells [108]. Therefore K12B as been strongly linked to the transformed phenotype observed in KS lesion tissue.

Viral miRNAs KSHV encodes a repertoire of 25 mature viral miRNAs that are some of the most highly abundant miRNAs types in PEL cells and are suggested to be expressed at even higher levels *in vivo* [32, 109, 110]. As previously stated viral miRNAs target a plethora of host and viral transcripts and all together can modulate aspects of the cell cycle, cell growth and proliferation, apoptosis, the immune system as well as viral latency [32]. Targets that facilitate immune evasion include IRAK1 and MyD88 (by miR-K12-9 and miR-K12-5, respectively) whose down-regulation limits TLR/IL-1R signalling. MICB mRNA is targeted by miR-K12-11 to promote cytotoxic T lymphoctye (CTL) evasion [52]. The cellular miR-155 ortholog miR-K12-11 has been suggested to prime B cells for transformation [52]. Anti-apoptotic targets include TGF-β type II receptor (TβRII), TNF-like weak inducer of apoptosis (TWEAK) receptor (TWEAKER) and caspase 3 [111–113]. Mitogenic targets include the DUSP1 (targeted by miR-K12-11 and -1) whose suppression promotes MAPK signalling and cell invasion [32]. Cumulatively the impact of viral miRNAs on cells is one of high transforming potential via targeting of immune, anigogenic, cell cycle and apoptotic-related functions.

1.2.11 The contribution of lytic replication to oncogenesis

While latency has been most strongly associated with oncogenesis, various lytic genes as well as the process of lytic replication in general has been observed to be critical for oncogenesis. While this may be partly explained by lytic replication's proposed necessity in maintaining a stably infected population of cells to facilitate persistence, more direct oncogenic effects have been proposed. Support for a role of lytic replication and gene expression as a contributor to KSHV-induced oncogenesis includes observations of the high expression of several lytic genes including ORF75, ORF74, K14 and K15 in KS and PEL. Further support comes from the observation that ectopic expression of several lytic genes can promote tumourigenesis including K1, vIL6 and ORF74 [114–118]. Moreover treatment of KS patients with drugs that inhibit lytic replication can lead to regression of lesions [14]. Finally, lytic replication has been shown to induce the accumulation of DNA damage, despite p53 being activated, indicating dysregulation to cellular DDR

and surveillance mechanisms [56, 119]. Overall this implicates lytic replication in driving KSHV-driven cancers, however the exact contributions, mechanisms and importance of each remain unclear.

Two possibly complementary mechanisms have been proposed for how lytic replication drives oncogenesis. The first is the phenomenon of abortive lytic replication, whereby predominantly latently infected populations of cells undergo a stunted lytic cycle that doesn't culminate in virion production (Fig 1.10). Such gene expression programs have been observed in KSHV-infected CD19+ T- and KS spindle cell lines *in vitro*, alongside an endothelial lineage of cells transduced with a KSHV bacterial artificial chromosome (mECK36), the latter of which had the capacity to form KS-like tumours in mice [120, 121]. Little is known about the mechanisms that drive abortive lytic replication, particularly in an *in vivo* context, however deletion of RTA's co-factor RBPJk, fatty acid depletion and prevention of RTA's inhibition of PARP1 results in forms of abortive lytic replication that result in no virion production [122, 123]. Moreover it is entirely possible that lytic gene expression during *de novo* infection of uninfected cells in a cell population facilitates this abortive lytic cycle (Fig 1.11). The alternative proposal is suggested to be due to the release of pro-oncogenic lytic factors in paracrine from the fraction of cells in a population undergoing lytic replication, which interact with latently infected cells to drive their transformation (Fig 1.11). Whether either of these mechanisms occur exclusively or mutually is not known and very little research exists into their importance in an *in vivo* setting.

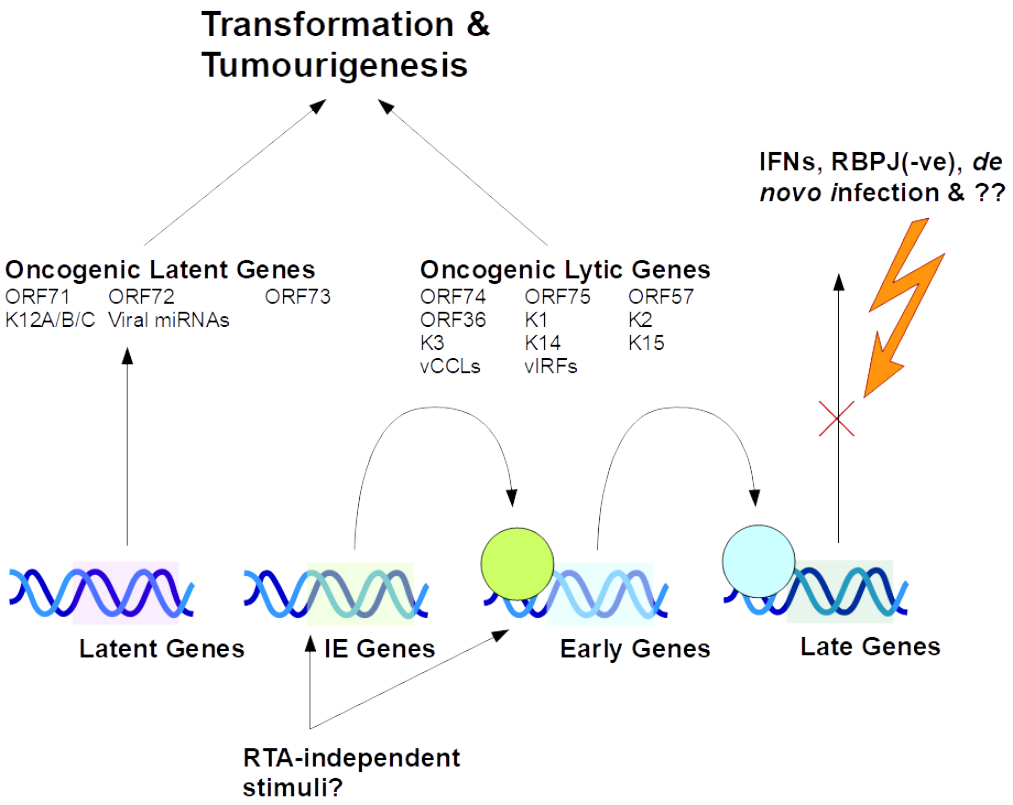


Figure 1.10: Proposed mechanismsfor abortive lytic replication.

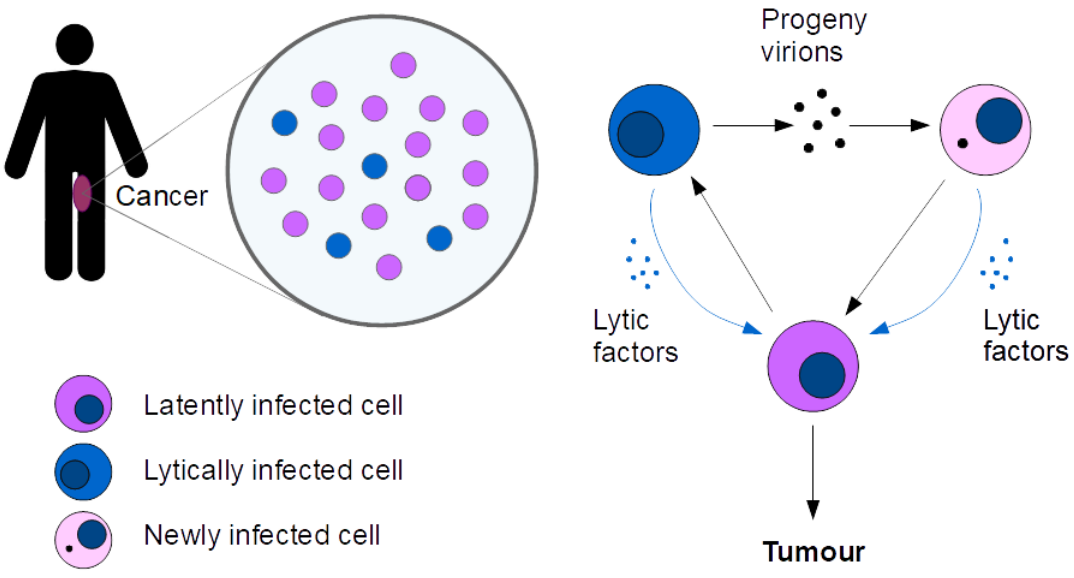


Figure 1.11: Contribution of paracrine signalling by either lytically replicating cells or *de novo* infected cells to the oncogenesis of latently infected cells.

1.2.12 Lytic oncogenes

vGPCR ORF74 (vGPCR) is a constitutively active human IL8R homolog capable of binding IL-8 as well as CC and CXC-family cytokines, with long- proposed pro-oncogenic roles [52]. It is capable of activating a range of pathways including PKC, PI3K, RhoA/Rac, NFkB, MAPK and PLC to induce the activity of TFs including AP1, NFAT, HIF-1α and CREB to promote VEGF, PDGF and pro-inflammatory cytokine-expression (Fig 1.7) [14, 124]. Because of this up-regulation of secreted factors, vGPCR is believed to predominantly promote oncogenesis in a paracrine manner. Interestingly, the ectoptic expression of vGPCR in mice has been showen to result in vascularised lesions that resemble KS [117]. Like vFLIP, vGPCR signalling is also proposed to promote EndMT via induction of Notch signalling (Fig 1.7) [99], In a more unique study, Krause *et al.,* 2016 proposed an additional oncogenic mechanism driven by ORF74 termed the "miR-34 effect" [125]. This is based on the observation of a pair of cell lines ectopically expressing ORF74, where they observed increased expression of miR-34a by vGPCR, which suppressed the expression of several genes involved in genome maintenance [125]. Which of these mechanisms is the most present in driving true oncogenesis is not clear, however.

K15 Like vGPCR, K15 is a transmembrane protein that localises to lipid rafts and thus acts as a conduit for the transmission of signals from the extracellular environment to intracellular processes. It is the most 3' proximal gene encoded downstream of the latency locus and is expressed at low levels in latently infected PEL cells and strongly induced during lytic reactivation [52]. It contains SH2 and SH3 and TRAF-binding domains on its cytoplasmic tail, that facilitate constitutive phosphorylation by Src kinases results in IL-6, -8 and -1β/α, CCL2/20, CXCL3 and COX2 activity via activation of Ras/MAPK, JNK/SAPK, NFAT, AP1 and NF-kB pathways (Fig 1.7) [52, 126, 127]. It also suppresses BCR-induced calcium influx to promote latency and is able to recruit PLCγ to regulate calcineurin-NFAT pathways which promotes angiogensis [126, 128, 129]. While K15 has been implied to be involved in oncogenesis, compared to the more predominant proposed viral oncogenes (ORF73, ORF72, ORF71, ORF74 etc), it remains relatively poorly understood.

vIL6 K2 (vIL-6) is an homolog of cellular IL-6 and as such targets the same processes, having pro-inflammatory and immunomodulatory effects. In fact it is hyper-active as it can activate the gp130 receptor independent of gp80 co-receptor to activate JAK/STAT, PI3K/Akt and MAPK pathways inducing C/EBPL, c-jun and host IL-6, alongside inhibiting IFN signalling [130]. Like K15, vIL6 (K2) is expressed in low levels during latency and

is up-regulated during lytic replication. It can also induce VEGF expression and consequential neoangiogenesis and tubule formation (Fig 1.13) [92]. Of the lytic oncogenes, it has a realtively well-established theoretical basis for its role in driving oncogenesis.

K1 K1 is a cellular BCR homologue expressed on cell and ER membranes that influence a range of processes. its oligomerisation triggers autophosphorylation of its intracellular immunoreceptor tyrosine-based activation motif (ITAM) motif, facilitating the recruitment of several SH2-containing proteins including PI3K/Akt/mTOR, PLCγ, Vav, Syk, Lyn, Ras-GAP and Grb2 and facilitates the activation of NFkB, NFAT, Oct-1 and AP-1 transcription factors (Fig 1.7) [52]. Such extensive interactions likely contribute to its oncogenic potential, for example Akt activation inactivates pro-apoptotic forkhead (FKHR) transcription factors, HSP40/90 binding is anti-apoptotic and it induces VEGF production via MMP9 activation, alongside the expression of inflammatory cytokines including IL-6, -8, -10, -1β and GM-CSF [14, 52]. Moreover Lee *et al.,* 2000 proposed a mechanism whereby K1 down-regulates surface expression of the BCR by interfering with its trafficking [131]. Overall K1's high interactivity implicates it in having potent cell and organismal modulatory activities.

ORF57 ORF57 is expressed early in lytic replication and promotes the transcription, splicing and nuclear export of many viral mRNAs. One pro-oncogenic mechanism was proposed to be due to ORF57's sequestration of the hTRex export complex away from sites of cellular transcription, resulting in the formation of R-loops [68]. This was observed alongside an enrichment of a range of proteins involved in non homologous end joining (NHEJ) [68]. The authors also suggested another mechanism linking chromosome instability to ORF57's sequestration of hTREX, whereby sequestration of UAP56 (a component of hTREX) results in premature sister chromatid separation, resulting in micronuclei formation. Both R loop and micronuclei formation are known to be trigger DNA damage [132, 133]. However the relevance of this to driving *in vivo* transformation and tumourigenesis remains to be confirmed.

vIRFs The vIRFs (vIRF-1, -2, -3 and -4) are believed to primarily contribute to oncogenesis by inhibiting the transcription of pro-inflammatory factors, in particular IFNs, inhibiting apoptosis, proper immune responses and cell cycle arrest [14]. These tend to antagonise host IRF function by binding them and preventing their transcriptional activity, however the exact mechanism varies depending on the isoform [14]. VIRF-1 in particular has been shown to bind and repress p53 transcription thus inhibiting apoptosis, namely via the

ATPM/p53 DNA damage response pathway (Fig 1.7) [134]. However the importance of each individual vIRF to oncogenesis isn't clear.

ORF36 ORF36 (vPK) is an analog of Ribosome S6 Kinase (RSK) which acts downstream of mTOR. It phosphorylates RSP6K to up-regulate protein synthesis, promoting growth, cell proliferation and angiogenesis (Fig 1.7) [135].

vCCLs KSHV encodes 3 viral CC-chemokine ligands (vCCLs), vCCL1 (K6), vCCL2 (K4) and vCCL3 (K4.1) which are cellular MIP1α, MIP1β and CCL2 homologues, respectively [99]. They bind their cellular homologues and, depending on the exact combination of viral and host proteins, either inhibit or promote the activities of the latter [14]. Moreover they have been proposed to limit Th1 cellular responses and favour Th2 to limit efficient anti-viral and anti-cancer responses [136]. Thus their functions are primarily immunomodulatory and so they likely promote oncogenesis by facilitating immune evasion and modulation of the immune environment.

vBcl-2 ORF16 (vBcl-2) and host Bcl-2 are important to promoting KS development, interacting directly with cellular factors such as Beclin-2 to prevent apoptosis and autophagy in KSHV-infected cells [14, 137].

1.3 Kaposi Sarcoma

KS was first described 150 years ago by Moritz Kaposi, however after over 100 years little progress was made into what caused it [138]. However, the HIV-1 epidemic of the late 20th century provided a "smoking gun" as it's strong association with AIDS-endemic regions not only led to its classification as an AIDS defining illness but implicated a pathogen-mediated cause [138]. This was confirmed in 1994 with the isolation of KSHV from KS lesions and the observation of its consistent detection in such lesion tissue [139]. Subsequent research implication of the virus's replication and associated gene expression as the driving factor in lesion development [138]. However because of the relative recency of this association, alongside KS's predominant occurrence in low and middle-income regions of the world, the interface between the virus and the host remains poorly studied; in-particular the influence of the latter. Moreover further factors are involved than immunosuppression and the influence of viral replication and gene expression, as evidenced by, for example, various presentations of KS that occur in apparently immunocompetent individuals. Thus the specific mechanisms that underlie the development of KS remains to be properly elucidated. The proceeding section will discuss clinical presentations of KS, it's epidemiology and the influence of host factors, the contribution of cell types and associated angiogenic and immunological factors, additional KS-induced pathologies and how they may relate to KS and a final discussion on the state of unbiased transcriptomic methods applied to KS.

1.3.1 Clinical Presentations

Compared to other similar lesions and cancers, KS is considered relatively slow-progressing and capable of cycles of growth and recession, however this depends on the exact presentation and additional factors [140]. Accordingly KS has been grouped into four major classifications. Two are associated with immunosuppression, with the most common and well-studied being AIDS-associated epidemic KS, alongside iatrogenic KS, that results from chemotherapy that suppresses the immune system [38]. The other two, classical KS and endemic KS, are not directly associated or co-morbid with obvious immune suppression or AIDS but may show some association with demographics and population genetics [38, 141, 142]. The epidemic variant tends to be the most aggressive form, however the endemic form is progressive and is capable of systemic dissemination [140]. Iatrogenic KS can be aggressive but usually fully regresses upon cessation of immunosuppression and thus is readily treatable [143]. Classical KS is usually indolent and generally presents

in elderly men of Mediterranean and Ashkenazi Jewish descent. Additionally, a fifth form of KS termed "nonepidemic KS" has been proposed that resembles classical KS but is present in young men who have sex with men (MSM) [144].

Initial presentation of KS is as a discoloured dark red/brown patch on the skin, that progresses to a raised flat plaque stage, to a defined nodular stage and then to further less well-defined and potentially divergent stages [38]. These cutaneous lesions tend to be localised to the lower portions of the body in which chronic lymphoedema is often associated [38, 141]. With increasing progression, lesions coalesce and spread to mucosae, with subsequent stages progressing to lymphadenopathic (involvement of proximal lymph nodes), florid (extensive invasion of local surrounding tissues) and/or infiltrative (invasion of more distant tissues, including musculoskeletal tissue and visceral organs) [137, 140]. This presentation in internal regions is associated with a more advanced and aggressive form of the disease that has poor prognosis, which is usually indicative of a co-occurrence of other KSHV-associated diseases [94].

KS lesions are characteristically highly vascular and this confers their distinctive colour. This is in-part due to KSHV being able to infect and transform blood/lymphatic endothelial cells (BECs/LECs) and mesemchymal cells (alongside more, less defined cell types), causing their expansion and an increase in vascular and lymphangiogenic signals, promoting new blood/lymphatic vessel in-growth. Moreover KSHV-infected endothelial cells shift towards a unique spindle-shaped morphology characteristic of the pathology, termed "KS spindle cells", which are considered the major transformed cell type in KS lesions [87]. Early patch stage lesions generally comprise fewer defined spindle cells, but with high vasculsarisation and immune infiltration [141]. Interlaced bundles of spindle cells and defined vascular spaces are present in plaque stage lesions, alongside increased inflammatory infiltration, in particular lymphocytes, plasma cells, dendritic cells and macrophages [141]. Late stage nodular lesions predominantly comprise intersecting fascicles and sheets of KS spindle cells that resemble fibrosarcoma, alongside the continued presence of neoangiogenesis [135, 141]. Notably, limited differences in terms of gene expression have been observed between the differing cutaneous morphological stages [67].

Further secondary co-morbidities can occur due to the presence of KS lesions that relate to their particular physiological location [141]. For example ulcerative KS presentations on mucosal tissue can cause pain, edema and cellulitis, while gastrointestinal KS can cause bleeding, diarrhoea, malabsorption leading to weight loss and pulmonary lesions can be associated with dyspnea, cough and hemoptysis [140, 141]. Bone involvement is associated with loss of function of the affected limbs due to damage of the muscles, bones and

surrounding connective tissue [140]. KS and KS-associated diseases have been linked to cardiovascular diseases, possibly due to the excessive dysregulation of angiogenesis innate to KS [120]. Other than physiological impacts, such as the stigmatism associated with the often very visible lesions, can lead to social isolation and psychological distress [137].

1.3.2 Epidemiology

While KSHV is capable of causing illness in immunocompetent individuals, its infection and thus maladies are most frequently associated with immunocompromisation. As previously stated, as an AIDS-defining illness it is strongly associated with HIV-1 and the two's epidemiology show high overlap. This includes high prevalence in Sub-Saharan Africa reaching as high as 90% in some areas where it is considered endemic (Fig 1.12). Conversely its seroprevalence is as low 10% in most populations of the US, Asia and Northern Europe[135]. Such geographic segregation into endemic/non-endemic regions can be misleading however as differing populations within such regions show differing seroprevalences. For example, seroprevalence is greater in some areas of the Mediterranean (20-30%), while adults of the Kazak and Uyghur population in Xinjiang province of China (where KS is also considered endemic) show higher seroprevalence (20-40%) than most regions of the rest of the country [145]. Moreover KS tends to show a bimodal distribution in terms of prevalence with age, with a peak for children of both sexes aged 4-9 followed by one in the adult population [140].

Given its association with HIV-1 and AIDS, KS increased in frequency during the worldwide AIDS epidemic. As such in non-classically endemic regions, like HIV-1, it tended to present more frequently in "men who have sex with men" (MSM) as well as in those who exhibit IV drug usage [146]. However KS is generally more prevalent in men regardless of sexual orientation and this has been suggested to be due to variations in hormones and sex-specific factors [147]. Given the relatively global prevalence of the AIDS-associated epidemic form of KS, it garnered more attention from more developed nations and as such is the most studied form [38]. Moreover given the strong association with AIDS, treatment with anti-retrovirals is typically the most effective and frequently applied method of therapy. Despite this however, epidemic KSV remains a leading cancer and cause of death for people at high risk of or living with HIV-1 [16, 38, 148**?**].

Transmission is believed to be most frequently via saliva during, for example, close sexual contact, and like most sexually transmitted infections (STIs), risky sexual behaviour such as the number of sexual partners or duration of sex has been linked to KSHV's

transmission. KSHV is frequently detectable in saliva and its secretion from oral epithelial cells is known to be 2 to 3-fold higher in titer relative to other anatomical locations [85]. Interestingly, the study by Jung *et al.,* found that the establishment of a productive KSHV infection of 3D oral epithelial organoid model was only successful upon damage to the superficial and intermediate layer to expose the basal layer [85]. They suggested that this may be due to the presence of entry receptors, integrin β1 and CD98 being mostly limited to such basal layers and that KSHV infection of the oral epithelial layer may be similar to the human papilloma virus (HPV) which shows a similar dependency on access to the basal layer. This may provide one mechanistic explanation as to how KSHV is able to be transmitted from one host to the next.

Other mechanisms of transmission have been proposed but remain poorly characterised. Blood-to-blood transmission is not confirmed as a major route of transfer, but some studies have indicated that use of blood products and blood transfusion can lead to transmission. This uncertainty may be because viraemia in KSHV-infected individuals is generally quite low and uncommon, and notably KSHV is undetectable in serum in about 50% of patients with KS [67, 149]. Similarly, blood sucking insects have been suggested as transmission vectors, but there is little evidence to support this notion [149]. The importance of different transmission routes does vary between endemic and non-endemic regions as for example one report linked blood-to-blood transmission as being relevant in Ugandan adults, while mother-to-child transmission has been suggested to be more important in other endemic areas [140].

Host factors associated with KS are generally believed to be those that promote either the immunosuppression that facilitates KSHV's enhanced survival and replication, or the progression of the disease. The latter case includes processes that contribute to the accelerated angiogenesis, cellular differentiation and chronic inflammation innate to KS [140]. Little is known in terms of contributing genetics, however assessment of this is difficult due to the inherent demographic biases to susceptible populations [120]. While older reports suggested that the HLA-DR5 genotype predisposed to KS, subsequent studies failed to confirm this [149]. Potential clues to the influence of genetic determinants may lie in KSHV's high seropositivity and bias observed in elderly male patients of Eastern European, Mediterranean and Jewish ancestry [120]. Other than HIV-1, further co-morbidities have been associated with KSHV transmission including infection by malaria and other parasites alongside skin diseases and diabetes [13]. Additionally co-infection with other herpesviruses (including EBV and HSV-1) have been suggested to enhance the oncogenesis of KSHV.

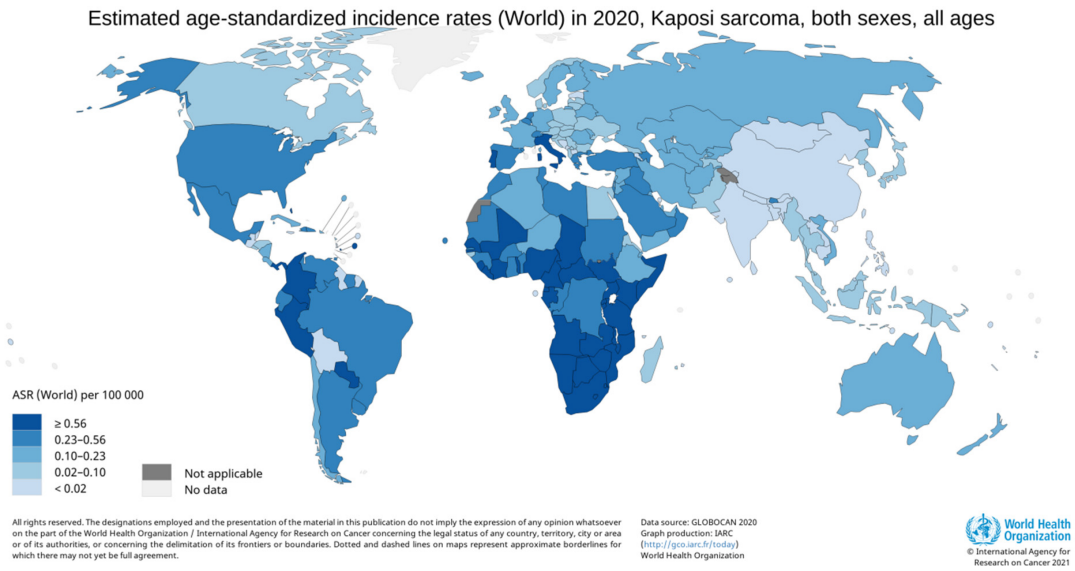


Figure 1.12: Global distribution of Kaposi Sarcoma. Figure adapted from [152].

Further more esoteric factors have also been linked to KSHV transmission including skin hygiene as well as recreational inhalation of nitric oxide (NOS) and nitrite compounds [12, 13, 149]. This is suggested to be due to NOS-mediated immune system suppression and modulation of endothelial cell function, alongside downstream metabolites being known mutagens [149]. One interesting suggestion is the proximity of populations to volcanic soils and clays rich in iron and aluminosilicate, which have been suggested to accumulate in leukocytes, conferring modest immunosuppression and a productive inflammatory environment [150]. Moreover addition of iron salts to KS-like spindle cell culture has been shown to promote their growth via increased expression of pro-inflammatory cytokines such as IL-6 [149]. Iron reserves are also typically lower in women than men, while lesions in women are often lost before or during pregnancy, which may indicate a non-hormonal contributor to the sex disparity observed in KS [151]. However many of these factors remain unconfirmed and consequently largely hypothetical.

1.3.3 The Cellular Origins of KS

Despite having differing clinical characteristics, all forms of KS show similar histopathological characteristics [99, 141]. Central to this is the KS spindle cell, which is considered the major tumour cell within lesion tissue [135]. This is in-part because spindle cells generally increase in abundance with progression through patch, plaque and nodular morphologies as well as numerous histological and *in vitro* studies [135, 140].

KS does not fit the classical dogma of monoclonality that is often attached to cancerous neoplasms, which states that cancers tend to evolve via an iterative process of clonal ex-

pansions, resulting in a majority of cells being from the same clonal line and thus progress towards monoclonality. Instead KS generally presents polyclonally, or as an oligoclonal expansion of virally-infected KS proto-spindle cells which undergo extenisve morphological changes and as such their initial origin is still debated [12]. In-part because of the above, the cellular origin of spindle cells is variable and still debated. This is in part to the fact that KS can infect a range of host cell types including endothelial cells, dendritic cells, monocytes, B cells, fibroblasts and epithelial cells. In addition markers of the lymphatic endothelium (ERG, PROX1, VEGFR3, LYVE-1 and podoplanin) and vascular endothelium (CD34, CD31 and CD36) are readily detectable on these spindle cells [13, 135]. Moreover spindle cells are poorly differentiated and express further markers of varying cell types including markers of smooth muscle, mesemchymal and dendritic cells, alongside macrophages, thus do not fully represent one apparent cellular origin [153]. Instead recent research points towards KSHV's capacity to drive differing cell types towards a common phenotype (KS spindle cells), that is assumed to promote and be promoted by KSHV's replication and gene expression. Namely KSHV is known to infect and drive the differentiation of BECs towards a LEC-like state and *vice versa* for LECs [154**?** – 156]. Conversely, KS spindle cells have characteristics similar to mesemchymal cells [157]. Moreover, during *in vitro* experiments on both primary and immortalised cell lines, KSHV has been demonstrated to instigate both endothelial-to-mesemchymal transition (EndMT) and mesemchymal-to-endothelial transition (MEndT), both of which have been suggested to contribute to the phenotype and progression of KS lesions [87]. This has led to recent suggestions that the cell-of-origin of KS spindle cells are mesemchymal cells that undergo MEndT to become more LEC/BEC-like, but not fully commit and end in an undefined state that exhibit the spindle cell phenotype [87, 141, 154, 158]. Conversely however, B cells and monocytes represent major reservoirs of KSHV in infected hosts and may represent the initial viral source, if not the cellular source of KS *in vitro*.

1.3.4 Angiogenesis and lymphangiogenesis in KS

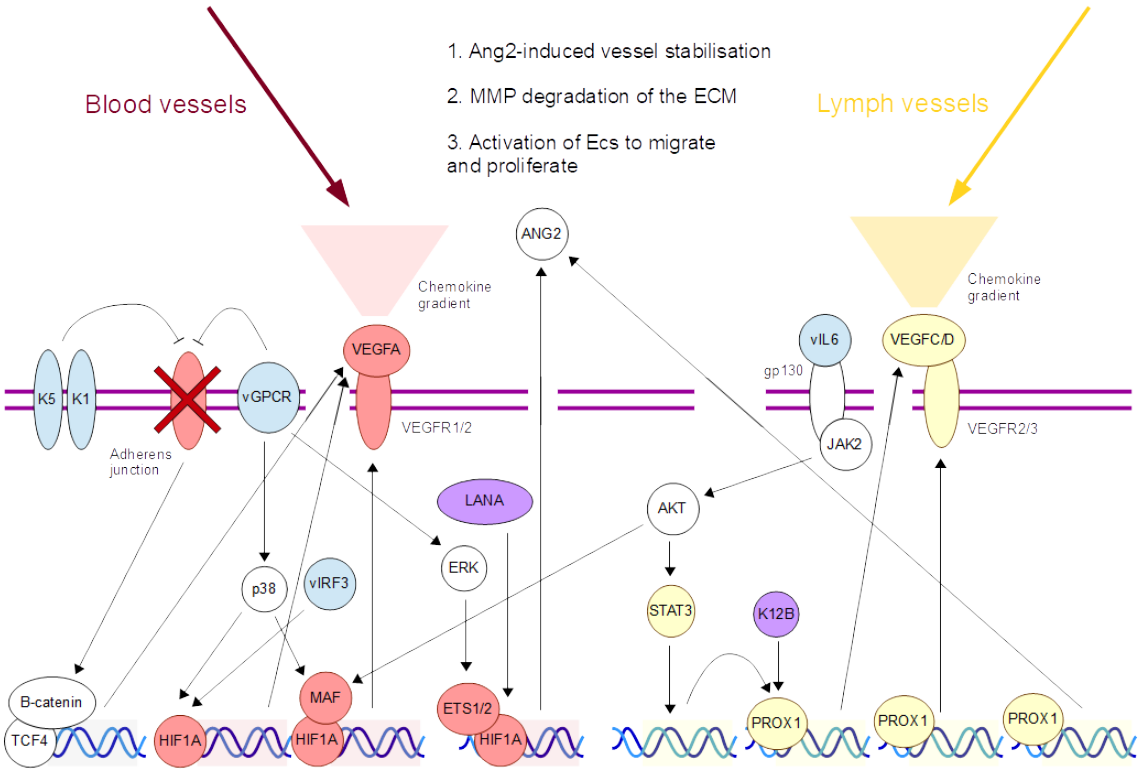


Figure 1.13: KSHV's interplay with angiogenesis and lymphangiogenesis. Red genes indicate those that generally promote angiogenesis, yellow are those that predominantly promote lymphangiogenesis, white indicate promoting both or not specifically related to either while others indicate viral genes. Figure adapted from [159].

KSHV Gene	Effect
gB	Increased VEGF-A secretion.
K8.1	Increased VEGF-A secretion.
KI	Increased VEGF-A secretion; Disrupted VE-cadherin signaling.
LANA-1	HIF-1 stability.
vIRF3	HIF-1 stability.
vGPCR	HIF-1 stability; Increased VEGF-A secretion; Increased angiopoietin-1 secretion; Increased angiopoietin-like 4 secretion; Disrupted VE-cadherin signaling.
vIL-6	Increased VEGF-A secretion; Increased angiopoietin-1 secretion.
Viral miRNAs	Downregulation of thrombospondin-1; down-regulation of MAF.
vCCLs	Chemoattraction.
K5	Degradation of VE-cadherin; Degradation of PECAM-1.
ORF57	Promotes the biogenesis of circHIPK3 which de-represses DLL4 and Notch signalling by sponging miR-30c.
K12B	Targets PROX1 activity.

Table 1.4: Known pro-angiogenic and lymphangiogenic roles of KSHV genes. Table adapted from [159].

Angiogenesis and lymphangiogenesis are related process that are important for the initial development and progression of most tumours as they provide a readily available source of glucose and metabolites, alongside immunomodulatory and growth factors [160]. Cancer cells generally facilitate this by excessive secretion of factors that trigger non-proliferating proximal endothelial cells (ECs) to differentiation and grow [160]. This appears particularly true for KS, given the characteristic extensive and leaky neovascularisation which causes the accumulation of red blood cells and thus KS's characteristic dark purple colour [99].

Angiogenesis and lymphangiogenesis occur via the sprouting of new vessels from pre-existing vasculature along chemokine gradients and into previously avascular spaces. This process involves several stages. These include degradation of the basement membrane, initial sprouting of vessels by EC proliferation and migration, tube formation and recruitment of smooth muscle cells or pericytes to generate mature new vessels [161]. Angiogenesis is a dynamic process whereby vessels are constantly remodelled; ebbing and extending, shrinking and growing depending on the needs of the tissue [161]. Apoptosis is important for this process to limit aberrant growth and accordingly this process tends to be absent or aberrant in KS lesions [160, 161].

Major angiogenic and lytmphangiogenic pathways include VEGF signalling, whereby secreted VEGF ligands binding cognate receptors (VEGFRs) such as VEGFR1/3 (FLT1/4) and VEGFR2 (KDR), predominantly expressed on ECs (Fig 1.13) [160, 161]. This activity acts in autocrine and paracrine, and typically has pro-mitogenic effects [159]. Moreover differential VEGF-VEGFR combinations have differing effect, as VEGF-A and its cognate receptor VEGFR1 (FLT1) are predominantly pro-angiogenic, while VEGF-C/D and their cognate receptor VEGFR3 (FLT4) are predominantly lymphangiogenic and VEGFR2 is context-dependent (Fig 1.13) [162]. The major downstream consequences of VEGF-VEGFR binding is ERK signalling, which up-regulates the key TFs ETS1 and ETS2, promoting both angiogenesis and lymphangiogenesis (Fig 1.13) [163, 164]. MAF and hypoxia, and PROX1 are key regulators of VEGF-mediated angiogenesis and lymphangiogenesis, respectively (Fig 1.13). Of the viral genes, at least K5, K1, vIL6, vIRF3 the vCCIs, glycoproteins gB and K8.1 and vGPCR all additionally target angiogenesis (Table 1.4) [159].

Another pro-angio- and lymphangio-genic pathway critical to KSHV is Notch signalling, whereby Notch receptor binding to cognate ligands Delta-like (DLL1/4 etc) and Jagged/Serrate (JAG1/2) promotes cleavage and subsequent internalisation of Notch's extracellular (EC) and intracellular (IC) domains, respectively (Fig 1.13). Notch signalling is a potent activa-

tor of lytic replication by activating RTA's co-activator RBPJk, with both Notch3 and DLL4 being highly expressed in KS and KSHV-transformed cells [71, 165]. Similarly Oct-1, a further co-activator of RBPJ is important in angiogensis as are STAT3/6, which are constitutively expressed in KSHV infected cells [160]. Further effectors of Notch include Hey-1 and Hes, the prior of which which promote angiogensis and regulates bone morphogenetic proteins (BMPs), which themselves regulate angiogenesis [160]. BMP signalling has been implicated in KSHV's tumourigenic properties [166].

Wnt signalling is another major driver of both angio- and lymphangio-genesis, predominantly via the activity of the canonical Wnt protein downstream transcription factor β-catenin (Fig 1.13). β-catenin is inactivated by phsophorylation by a range of cellular factors including GSK3β, Axin, CK1/2, protein phosphatase 2A (PP2A) and adenomatous polyposis coli (APC) and activation of upstream Wnt signalling components leads to β-catenin's de-repression by these factors [167]. In fact inhibition of GSK3β by LANA is one major proposed mechanism by which KSHV promotes angiogenesis, while vGPCR is known to activate it via PI3K/Akt-mediated inactivation of GSK3β [168, 169]. However at least two lytic gene products, vIRF-4 and ORF36, are believed to suppress it, likely due to the importance of β-catenin to effective IFN responses [169].

Several ncRNAs have roles in regulating angiogenesis. Many miRNAs, typically termed "angioMiRs", are known to target a range of angiogenic-related processes [170]. These include include miR-574-5p, miR-92a and miR-27a-3p, which all promote α-catenin signalling, which has similar roles to β-catenin signalling. Moreover various lncRNAs and circRNAs also have proposed roles in angiogenesis. For example SLCO4A1-AS1 impairs β-catenin/GSK3β association thus promoting the priors activity [171]. Similarly circRNA_102171 (circSMURF2) has been proposed to suppress CTNNBIP1, an inhibitor of β-catenin, by preventing β-catenin/CTNNBIP1 association, likely via direct circSMURF2-CTNNBIP1 interaction, promoting β-catenin/TCF/LEF transcriptional activity and subsequent migration and progression of papillary thyroid cancer [172]. Additionally circHIPK3 is up-regulated during KSHV lytic replication and binds and suppresses the abundance of miR-30c as well as miR-29b, miRNAs with tumour suppressive functions [71]. This is in order to de-repress their target DLL4 and thus promote Notch-mediated functions, potentially including angiogenesis, alongside promoting cell cycle progression and lytic replication [71]. This was found to be dependent on ORF57 and to a lesser extent host RGG1/2 (Table 1.4) [71].

1.3.5 KS and the immune system

KSHV's relationship with the immune system is of critical important to the development of KS [11]. This is most obviously emphasised via it's predominance in immunocompromised individuals, namely those with AIDS and the obvious defective T cell immune responses conferred by the condition. This is predominantly believed to be due to reduced immune system-mediated restriction of viral persistence and lytic replication, enabling greater rates of viral replication and gene expression. However, KS's 10,000-fold increased prevalence in AIDS individuals has been suggested to also be in somepart a consequence of the extensive systemic modulation of cytokines innate to AIDS [99]. Indeed the establishment of KS involves the development of a pro-inflammatory environment, while complex immune cell infiltration is a common feature of KS lesions [16, 67, 94]. For example, KS patient sera contains elevated IL-6 and TNFα, however there is further complexity as an up-regulation of the immunosuppressive IL-10 has been previously observed as well [136].

Another reason that immune processes are important to KS is their tight coupling to angiogenesis, as many cytokines also induce angiogenesis [136, 141, 173]. At least some of this is mediated by the importance of M2-type macrophages in helping orchestrate angiogenesis, as well as associated processes. One such process is wound healing, which shows some parallels with the development of KS. For example KS is associated with the Koebner phenomenom; whereby KS lesion tend to occur at pro-inflammatory sites such as regions that have undergone injury or trauma [141, 173]. Additionally both KS and wound healing involve EndMT to facilitate migration of cells towards damaged or cancerous tissue, respectively [99]. This process is driven by the expression of inflammatory cytokines IL-1β, TNFα and TGFβ, however the latter appears dispensible for KSHV-mediated EndMT [174].

In-line with the extensive pro-inflammatory environment and increased viral replication and gene expression present in KS lesions, extensive immune inflitration has been observed [67, 94]. This includes extensive invasion by phagocytes (monocytes, macrophages and dendritic cells), T cells, B cells, mast cells and granulocytes (particularly mast cells) [175]. Mast cells in particular are enriched in KS lesions and spindle cells and may be key mediators of the pro-inflammatory environment [94, 175]. Similarly macrophages and monocytes have been frequently proposed to contribute to KS development, particularly the wound-healing pro-angiogenic M2 variant [94].

The specific forms of T cells present in KS lesions may relate to the its development

as well. While the presence of AIDS and similar immunodeficiencies result in reduced Th1 and thus CD8+ cytotoxic T cells, the increased production of vCCLs associated with increased lytic replication in KS lesions has been proposed to increase Th2 cell infiltration [136]. This has been suggested to promote an improper and pro-inflammatory adaptive immune response in KS lesions, exacerbating it's development.

1.3.6 Further KSHV Pathologies

KSHV is a contributor to 2 confirmed additional major pathologies: primary effusion lymphoma (PEL) amd multicentric castelamn's disease (MCD). PEL is a very aggressive non-solid lymphoma that comprises predominantly monoclonal populations of latently infected B cells that expand within internal body cavities, with high copy numbers (50-100) of viral genomes per cell [52]. Like KS, PEL is associated with AIDS as well as HAART and viral load has been suggested to be associated with the progression of PEL [176]. Moreover PEL is notable as many of the stable cell lines that are used in *in vitro* laboratory settings to study KSHV, including KS-1, BCBL-1, BCP-1 and JSC-1 which contain latently infected KSHV, were derived from such tissue [177]. Therefore a majority of knowledge of KSHV's transforming potential on infected cells has been determined in the context of already transformed B cells and so may not directly be applicable to KS.

MCD is associated with but not always caused by KSHV infection as the virus is found in only $\sim$50% of MCD patients. It is characterised as a disseminated lymphadenopathy driven by the uncontrolled growth and proliferation of IgM γ-restricted plasmablasts [52, 160] This tends to occur within B cell follicular mantle zones and involved B cells tend to be large and exhibit nuclei of cells undergoing replication (open phase nucleus), with multiple nucleioli [52, 160]. Interestingly, relative to PEL And KS cells, greater rates of lytic replication tend to be observed in MCD and it is believed to be largely driven by aberrant cytokine activity that is believed to be largely driven by vIL6, whose detectable expression is associated with worse prognosis [52].

Perhaps unsurprisingly considering the immune modulation innate to KSHV-induced diseases, at least two aberrant systemic inflammation-related disorders have been associated with KSHV [52]. One is KSHV-associated inflammatory cytokine syndrome (KICS) that is associated with vIL-6 and viral miRNAs, alongside host IL-6, IL-10 and C-reactive protein [178]. However as it is often associated with KS and is often diagnosed by the exclusion of a diagnosis of MCD, it has been suggested to not be a distinct pathology [52]. The second is immune reconstitution inflammatory syndrome (IRIS), which has been suggested to be due auto/pathogen-antigen recognition due to restoration of functional

CD4+ T cells in patients undergoing HAART [179]. Interestingly IRIS is characterised by the emergence of previously undetectable KS (unmasking KS) or rapid progression of existing KS (paradoxical KS), underlining the importance of immune responses to the pathogenesis of KS, but also that immune suppression isn't always directly stimulatory to KS development [179].

As well as KS, MCD, PEL, KICS and IRIS less concrete or confirmed links have been made between KSHV and other pathologies. For example a recent study indicated an association between KSHV and osteosarcomas in the Uyghur population of the Xinjiang provence of Western China [180]. Further clinical conditions associated with KSHV include skin carcinomas, angiosarcomas, angiolymphoid hyperplasia and eosinophilia, multiple myeloma, AIDS-associated immunoblastic lymphoma, primary central nervous system lymphoma, post-transplantation lymphoproliferative disorders and pulmonary inflammatory myofibrolaastic tumours [120, 160].

1.3.7 Recent developments in transcriptomics applied to KS

While KS has been studied for over 150 years, very limited global analyses have been performed into the dysregulation of gene expression in KS. However over the course of the past few years, several studies have been undertaken involving RNA-Seq profiling of KS tumours biopsies in order to elucidate factors that may drive its development [16, 67, 94]. While these studies are limited by their small scope and size alongside the highly heterogeneous nature of KS lesions, insights some key insights have been gained. The purpose of this section is to highlight how these methods have been applied and what has been learned from them.

The first is Tso *et al.,* 2018 which profiled and compared epidemic KS (AIDS-associated) and matched control biopsy tissues from 4 patients and identified net up-regulation of viral expression that varied substantially within the tumours [16]. They found that viral genes clustered along functional lines, with distinct clusters for those involved in lytic gene expression, latency and immunomodulation and structural genes. TGFB1, S100A9, ERG, ETS1 and SPI1 and ADIPOQ, PPARG, IRS1, ERS2 and SREBF1 were the most activated and inhibited TFs, respectively. Moreover IFNα signalling and CIBERSORT enrichments showed a significant enrichment of B cells, macrophages and NK cells, indicating a pro-inflammatory environment producing immunogenic chemoattractants. Glucose metabolism was reduced, reminiscent of the Warburg effect, however unlike what would be expected from the effect, lipid metabolism and PPARγ signalling was also depleted.

The same group performed a follow up study utilising the same 4 (resequenced) samples

and 18 additional samples, 6 of which were from endemic, non HIV-1 co-infected patients [67]. Interestingly they showed surprisingly similar correlations between the transcriptomics of epidemic and endemic KS samples, with the major difference being a greater magnitude of differential expression between lesion and control for endemic relative to epidemic samples. Only 428 genes were differentially expressed, while just 27 genes were uniquely differentially expressed between the two. The latter included a potent growth factor and regulator of glucose transport and glycogen synthesis IGF1, T cell surface glycoproteins involved in CD8 T-cell stimulation, microtubule associated protein RP/EB family member 3 (MAPRE3), CD8A and CD28 and interferon-induced helicase C domain-containing protein 1 (IFIH1) that were also higher in endemic samples than in epidemic samples [67]. They suggested that these observations was due to the reduced dysregulation of gene expression required to induce lesion formation with AIDS. Moreover non-KS and control tissue showed no significant differentially expressed genes, as did the plaque and nodular lesion comparison. Lesion samples also exhibited latent, predominantly lytic or a mixed gene expression profile. This was similar to an older study by Hosseinipour *et al.,* 2002 that performed comprehensive qRT-PCR profiling of the expression of all known viral genes in 35 ART-/chemotherapy naive epidemic KS patients [148].

Another study by Ramswami *et al.,* 2022 compared the transcriptome profiles between skin and gastroinitestial KS alongside matched controls from patients living in the USA [94]. They found much similarity between the expression profiles and enriched functions when comparing lesions to matched controls, such as the IL-6, HIF1a signalling and the granulocyte adhesion and diapedesis pathway. Moreover, the key VEGF receptor FLT4 as well as CD5L, predominanlty released by macrophages, was up-regulated in KS vs matched controls for skin and GI KS. In contrast to Tso *et al.,* 2018 and Lidenge *et al.,* 2020, no dysregulation of glucose or lipid metabolic processes was observed [16, 67]. They did however observe an up-regulation of VEGFR2/3 and KDR, IL-6, IL-10 and the type I IFN receptor IFNAR2 alongside enrichment of BCR signalling in cutaneous but not GI KS, alongside an enrichment of IL-1α and associated signalling in GI KS. In terms of viral genes they observed that ORF75 correlated positively with latent and genes induced by K2, but not lytic genes themselves [94]. Additionally, the expression of vIRFs was found to be variable across the cohort. By also comparing CIBERSORTx-deconvolved immune cell proportions between matched control and KS lesions, they were able to show a decrease in follicular helper T cells and T regulatory cells in GI but not skin KS, while M1 macrophages were enriched in KS lesions in most skin but not GI patients [94]. Finally in subsequent studies on lymphatic endothelial cell culture, they observed an increase in

the pro-inflammatory IL-6 and IFNγ which favour M1 polarisation, alongside a decrease in cytokines that favour M2 [94].

In another study, Rose *et al.,* 2018 applied their previously developed their UCDS feature schema investigated the landscape of viral gene expression in 41 epidemic KS lesions from 30 patients (11 providing 2 samples) [15, 17]. Interestingly in these lesions they that the majority of transcription originated from the P3 promoter that flanked ORF72 and little from the P1 promoter that is predominantly used in cell lines indicating differential promoter usage between KS and most *in vitro* models [17]. They also identified highly variable splicing between an upstream ORF (uORF) between the P3 promote and ORF72 CDS, the DR5 and DR6 repeats and the canonical Kaposin A (K12A) ORF and variable start codons which they suggested indicated further Kaposins: Kaposin D and E, as well as A, B and C [17]. Like Lidenge *et al.,* 2020 they identified lesions with either predominantly latent, lytic or mixed gene expression profiles, with the prior showing greater total viral expression indicating that a predominant latent gene expression profile was not just common to endemic KS [17, 67]. Moreover they also found no considerable differences between the viral gene expression profiles of nodular, macular (flat, plaque-like) or fungating (ulcerations and necrosis) lesions.

Overall these studies have provided an insight into the dysregulation of gene expression in KS and a firm foundation for further more in-depth research.

1.4 Regulation of cellular and viral gene expression by circular RNA competing endogenous RNA networks

1.4.1 An RNA-centric view of gene expression

The canonical view of gene expression encompassed by the central dogma centers on the model that genes are transcribed as messenger RNA (mRNA), which encode genetic information that is are subsequently translation into protein to confer a phenotypic effect. In this model, RNA acts as the central mediator of "messages" from genomically encoding genetic information to its expression as functionally effective proteins. However further work revealed the existence of "non-coding" RNA (ncRNA) with the discovery of transfer RNA (tRNA) and ribosomal RNA (rRNA), key to the process of translation. This emphasises that RNA has functions independent of translation and thus beyond just being a temporary conduit of information.

It wasn't until the turn of the 1990s that the diversity and regulatory extent of ncRNAs began to be fully appreciated. This was with the discovery and functional characterisation of lncRNAs and miRNAs, which were found to confer regulation of gene expression at levels of gene expression upstream or downstream of translation. Since, the number and diversity of ncRNAs has exploded. More recent additions to the ncRNA family are pseudogenes and enhancer RNAs (eRNA), small nuclear (snRNAs) and small nucleolar (snoRNAs), alongside tRNA-derived fragments [181]. Further esoteric examples of ncRNAs include RNAs that function as enzymes or ribozymes, such as RNase P, Twister, Hammerhead ribozymes and some components of the spliceosome [182, 183]. It is likely that these categories will continue to expand and increase in number as research progresses.

NcRNAs have been found in all cellular organisms studied to date as well as most viruses. Moreover their expression is consistent with them being functional rather than "transcriptional noise" or a result of false positives by detection methods. In fact, while some estimates put the total protein-coding proportion of the human genome as just $\sim$2%, up to 75% of the genome has been proposed to be transcribed, which would indicate that the majority of transcripts sequences exert their function via non-coding means [184]. Moreover the abundances of many ncRNAs, especially tRNAs and rRNAs, but also lncRNAs and miRNAs are comparable and in some cases surpass the abundances of most mRNAs within cellular environments. Importantly, such ncRNAs show differential abundances between varying biological conditions indicating their context-specific expression, with follow-up functional studies identifying mechanisms that associate them with the

conditions of interest. Overall while 100s of ncRNAs of various types have now been well-characterised, as with much of genome-scale biology, the unknowns outweigh the knowns and as such the study of most ncRNA species remains in relative infancy.

1.4.2 Regulation of gene expression by miRNAs

While transcriptional, translation and post-translational control of gene expression are all extensive, one aspect of regulation that has garnered interest in the past 2 decades is that of post-transcriptional control. This occurs at the level of regulation of the stability, localisation and activity of mRNAs. Moreover it is an efficient mechanism of regulation, being that mRNA is the critical and importantly transient thus dynamic mediator of information flow through the process of gene expression. This level of control of gene expression is one key role for many non-traditional ncRNAs (ie other tha tRNAs, rRNAs). Probably the most well-characterised plays in this are miRNAs, that act as the primary effectors of the process dubbed "RNA interference" (RNAi). This is the process whereby miRNAs are able to repress the expression of genes based on their binding to (canonically the 3' UTRs of) mRNAs for which they share sequence complementarity. In doing this, they are able to induce either the destabilisation and/or degradation of bound mRNAs, or inhibit their translation.

The biogenesis of miRNAs is relatively well-understood. They are transcribed from inter- or intra-genic loci as concatenated primary miRNAs (pri-miRNAs) or "mirtons" from bespoke or shared promoters or due to sense or antisense read-through transcription (Fig 1.14) [185]. Such precursors are processed by the ribonuclease DICER in complex with DGCR8 in the canonical pathway, or spliceosomes in the non-canonical pathway of biogenesis (Fig 1.14). Both pathways produce precursor miRNAs (pre-miRNAs) as hairpins, with each strand of the stem encoding potential miRNAs with reverse complementarity [185]. These pre-miRNA hairpins are then translocated out of the nucleus to the cytoplasm in complex with the exportin XPO5 [185]. Pre-miRNAs are then cleaved by the ribonuclease Dicer in complex with TRBP/PACT and ADAR1 to release the 5' and 3' stem strands, either of which can become a miRNA (hence -3p and -5p notation). Which exact stem is processed into a mature miRNA depends on a range of factors that govern this process, termed target-mediated miRNA protection (TMMP) [185].

The mechanisms and determinants by which miRNAs facilitate RNAi have been well-characterised. Importantly, miRNAs encode a 5' "seed" sequence that is relatively short (7-9nt) and shares at least partial complementarity with target transcripts [186]. Such sequence complementarity is the primary determinant of miRNA targeting, while the re-

laxation from full complementarity allows for promiscuity in target sequences. Further factors also facilitate miRNA binding such as 3' end complementarity, RNA secondary structure, ADAR-editing, methylation, intrinsic or protein binding-induced stability, non-canonical (5' and CDS/ORF) binding location and the capacity for sequence bulges in or around the seed sequence [185, 186]. MiRNAs are typically loaded into Argonaute (AGO) proteins to form RNA-induced silencing complexes (RISCs) that enable their canonical pathway of repression of target transcripts. MiRNA association into miRISC complexes also facilitates their transport, proper targeting and stability. The canonical mode of action of miRNAs is that they bind and induce the cleavage and subsequent degradation of target transcripts. This canonical pathway proceeds via miRISC-triggered depolyadenylation, decapping and subsequent exonuclease (XRN1 etc)-mediated degradation [185]. In addition, non-canoncial pathways that don't rely on cleavage and degradation but still induce target repression also exist. These include AGO2 competition with translational pre-initiation complexes for binding to the 5' mRNA cap as well as targeting bound mRNAs to P-bodies, which are sites of translational repression that accumulate messenger ribonucleoprotein (mRNP) complexes [187].

1.4.3 The competing endogenous RNA hypothesis

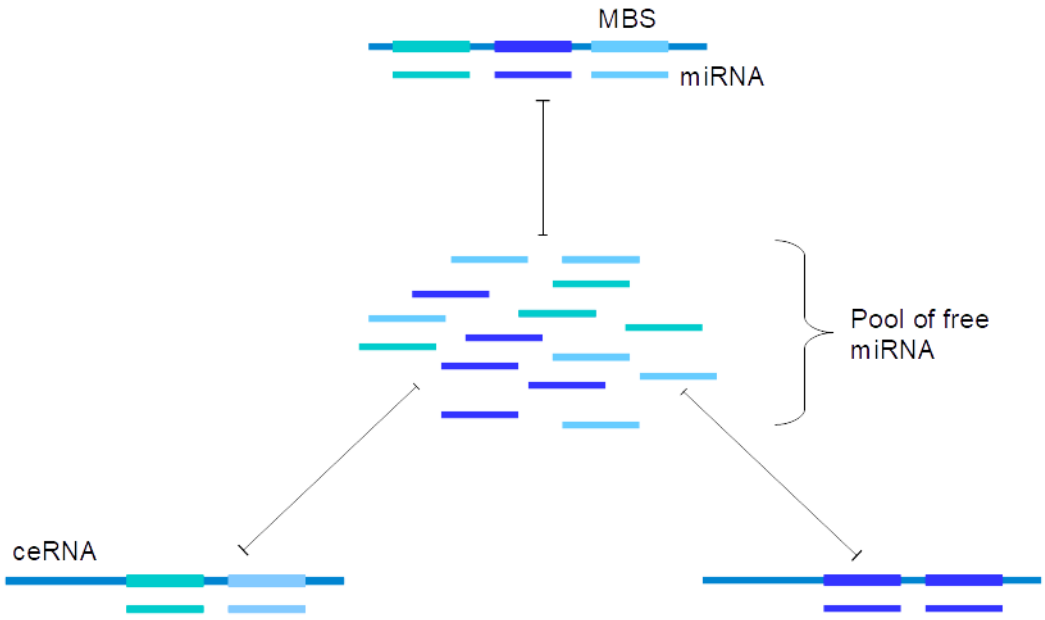


Figure 1.15: The competing endogenous RNA hypothesis. Where MBS is miRNA binding site, ceRNA is competing endogenous RNA, miRNA is micro RNA.

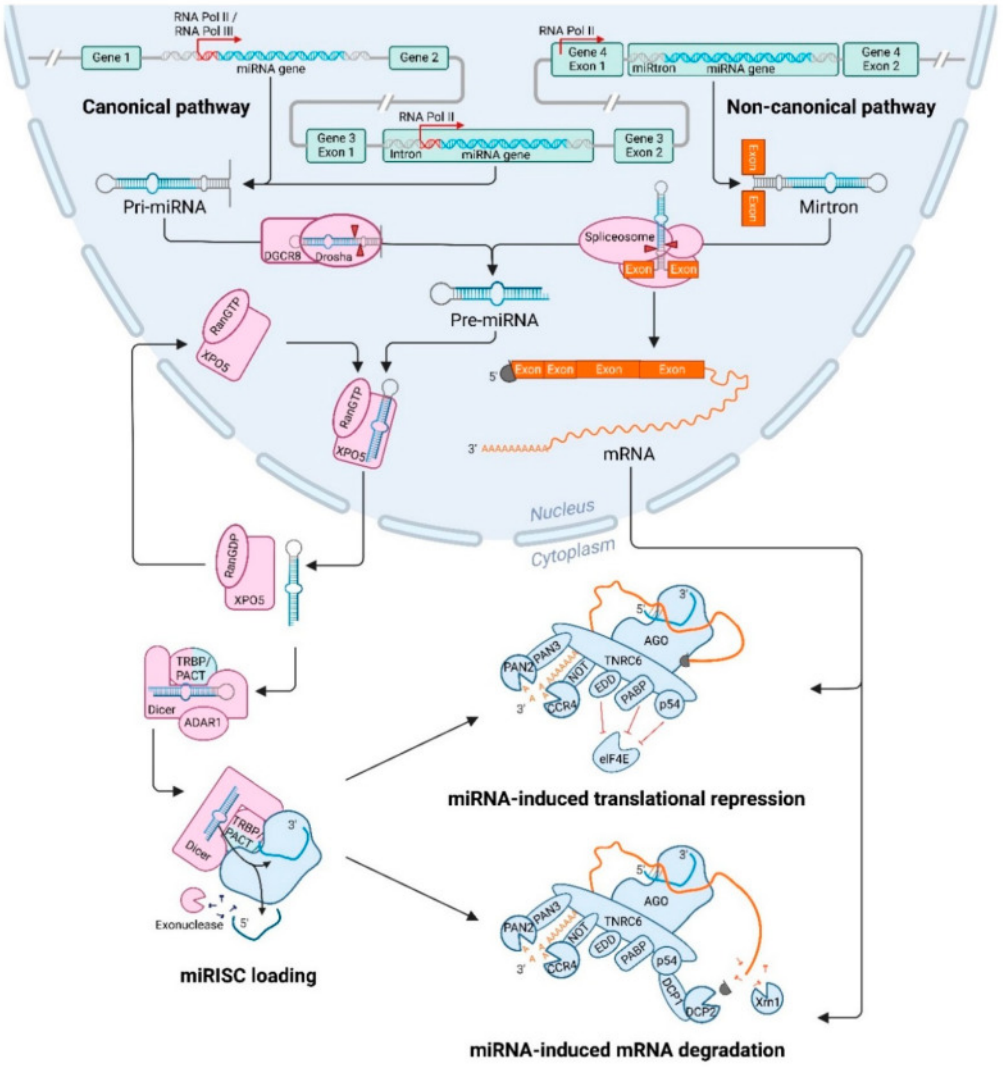


Figure 1.14: Biogenesis of miRNAs. Figure adapted from [185].

The human genome encodes 100s of miRNAs, each of which can target 10s-100s of transcripts and accordingly at least 40% of protein coding genes have been proposed to be directly regulated by miRNA-driven RNAi [185]. Many of these potential miRNA target "regulons" overlap and thus the landscape of miRNA targets in the cell can be considered an interconnecting network with miRNAs as the main mediators. Given this network perspective on miRNA function, Seitz *et al.,* proposed the foundation of the "competing endogenous RNA" (ceRNA) hypothesis, which was later formalised as the competing endogenous RNA (ceRNA) network hypothesis by Salmena *et al.,* [188, 189]. The ceRNA network hypothesis states that non-structural RNAs of sufficient size (primarily mRNAs, lncRNAs, pseudogenes and isolated 3'UTRs) compete to bind a shared pool of miRNAs that is of a finite size and thus, changes in the abundance of one target can influence the abundance of another if they share common miRNAs (Fig 1.15). In this way, traditional RNAi can be viewed as a *cis*-regulatory mechanism of miRNA, while the ceRNA mechanism can be viewed as a *trans*-regulatory mechanism [190]. Thus the ceRNA hypothesis proposes extensive crosstalk between transcripts that adds another post-transcriptional layer of regulation to gene expression [188, 189].

A logical continuation of the ceRNA hypothesis is that targets of miRNA regulation are also able to regulate the abundance of miRNAs by limiting the size of the free pool, in a bi-directional reciprocal relationship. Specifically, transcripts that have the capacity to bind but not be cleaved, destabilised or inhibited by a miRNA have the potential to influence the abundances of a considerable number of mRNA targets that are targeted by common miRNAs. Later terminology characterised many such transcripts "miRNA sponges" due to their capacity to continually sequester miRNA away from the free cellular pool [191]. One archetypal example of a validated miRNA sponge is PTENP1 which forms a ceRNA network with PTEN and sponges several miRNAs, including miR-17, miR-21, miR-19 and miR-26 families that otherwise suppress PTEN mRNA [192]. Similar mechanisms exist for KRAS and its pseudogene KRAS1P as well as the isolated expression of CD44 3'UTR [189]. Since then, many studies have characterised 100s of putative miRNA sponges that regulate 100s of miRNAs and thus potentially 1000s of transcripts. Importantly, while still a hypothesis, many studies have provided evidence for the existence of at least individual ceRNAs acting to regulate the expression of mRNAs, providing evidence for the occurrence, if not the extent, of ceRNA mechanisms that operate in a cellular environment.

1.4.4 Studying ceRNA networks

Due to representing non-covalent and subtle interactions between 10s-100s of interacting components, ceRNA networks are by nature difficult to study by most naive molecular biochemical approaches. Moreover, the biological effect of ceRNA networks are likely minute and difficult to detect or otherwise deconvolve from other regulatory mechanisms. Accordingly many studies that utilise such classical laboratory-based approaches tend to be limited to the investigation of one or at most a handful of individual miRNAs, one miRNA sponge and several target transcripts. A consequence of this is that without accounting for the influence of the other interactors in a ceRNA network, the cause of a change in one element of the network may be wrongly attributed to the direct perturbation of a component (ie a miRNA or miRNA sponge), when this perturbation may instead have indirect effects on whatever variable is used as a read-out [191]. Accordingly, the analysis of such regulatory circuits tends to lend itself to systems-based network analysis methods which utilise high-dimensional 'omics data-sets. Such holistic analysis attempts to capture the full scale of underlying ceRNA networks and consider all components together rather than in a gene-by-gene manner [191]. Network models tend to be constructed by defining links (edges) between miRNAs and their targets using either predictive algorithms, databases of prior knowledge, co-expression or a combination of all three. Networks can then be interrogated via various network analytical methods. These include centrality analyses, which rank network components such as miRNAs, targets and miRNA sponges (nodes) by their relative interconnectivity in the network, with the aim of identifying highly interconnected hub nodes. Such network approaches have proved fruitful in identifying novel ceRNAs and ceRNA regulatory networks in a range of contexts [193]. However given the recency of the ceRNA network model proposal, the true extent of ceRNA network regulation within cells and between biological contexts remains largely unexplored.

1.4.5 CeRNAs and viruses

Many viruses encode their own miRNAs, while infection by many more are known to dysregulate a range of cellular miRNAs. As previously stated, this includes KSHV's two-dozen unique miRNA species that modulate the expression of a range of genes involved in immune modulation, metabolic processes, angiogenesis and oncogenic processes (Fig 1.3) [32]. Additionally viruses encode many other ncRNAs, including lncRNAs, such as those encoded by KSHV, EBV and HCMV, among many others [15, 194].

Given that viruses exploit or are influenced by essentially all levels of gene expression regulation, as well as many encoding their own miRNAs or exploiting host miRNAs, it stands to reason that some also exploit or are influenced by ceRNA network-based regulation. Indeed, one prime example is Hepatitis C virus (HCV), an RNA virus whose genomic RNA sponges miR-122, a highly expressed miRNA in hepatocytes, resulting in global de-repression of the targets of miR-122 including those involved in lipid biogenesis (among other processes) [195]. Most interestingly for KSHV, the related Herpesvirus samiri (HVS) has been shown to express highly abundant miRNA sponges, termed the herpesviral U-rich RNAs (HSUR1 and HSUR2), which resemble snRNAs and sponge miR-27a [196]. MiR-27a is a known anti-viral miRNA and thus HSUR-mediated repression has been shown to promote viral replication [196]. Murine cytomegalovirus (MCMV) encodes similar transcripts that also target miR-27 [197]. Moreover K. Ahn *et al.,* showed that in clinical isolates, the human cytomegalavirus (HCMV) expressed the "miRNAdecay element" (miRDE) from an intergenic region of its genome that specifically destabilises miRNAs of the miR-17-92 cluster [198]. Thus the potential role of ceRNAs in regulating viral replication has been proven in some cases, in-particular for herpesviruses like KSHV.

miRNA	**Targets**	**Functions**
miR-K12-1	Casp3	Apoptosis
	NFkB signaling	KSHV latency
	THBS1	Cell adhesion, migration, and angiogenesis
	p21	Cell cycle arrest
	STAT3	IkBα/NFkB/IL-6 signaling pathway
miR-K12-3	Casp3	Apoptosis
	nuclear factor I/B	KSHV latency
	GRK2	KSHV latency and angiogenesis, dissemination
	THBS1	Cell adhesion, migration, and angiogenesis
	C/EBPβ p20 (LIP)	Influence the secretion of IL-8 and 10 and immune response
miR-K12-4	Casp3	Apoptosis
	Rbl2	KSHV latency
miR-K12-5	BCLAF1	KSHV latency
	MYD88	Regulating the TLR/IL-1R signaling cascade
	Tmskα1	Apoptosis and angiogenesis
miR-K12-6	THBS1	Cell adhesion, migration, and angiogenesis
	Bcr	Angiogenesis
	SH3BGR	Angiogenesis and dissemination
	MAF	Angiogenesis
miR-K12-11	MAF	Angiogenesis [199]

Table 1.5: Validated functions of KSHV viral miRNAs. Adapted from [32].

1.4.6 Circular RNAs

One class of ncRNAs that has garnered a lot of attention in recent years are circular RNAs (circRNAs). These are covalently-closed loops of RNA that are notably distinct from linear species or other circularised RNAs (such as intronic lariats that are a by-product of splicing) in that they have no 3' or 5' termini [200]. As such they lack structures and features present in canonical mRNA species, such as a m7G 5' cap and polyA tails. Moreover their lack of 3' and 5' termini and unique circular shape also confers resistance from exonuclease activity. Accordingly they are highly stable, exhibiting half-lives of 10s of hours, even in the complex environment of serum or the extracellular milieu [201]. They are also broadly expressed and have been observed to be generated from 1000s of human genes and can in theory be expressed from essentially any gene capable of being spliced [202–204].

Biogenesis of circRNAs occurs via a process called "back-splicing", whereby the 3' splice site of a downstream exon is spliced to the 5' splice site of an upstream exon, resulting in a "scrambled" order of exons in a reverse orientation relative to their genomically encoded order (Fig 1.16) [200]. This is proposed to be mediated by the host splicing machinery but expedited by factors that bring back splice sites into proximity [204]. These include inverted repeats in regions flanking the backspliced sites (such as ALU elements) and consequently factors that modify or regulate these sequences, including ADAR1 or hnRNP-C/L [202, 204]. Moreover RNA binding proteins (RBPs), such as human QKI, SFPQ and *Drosophilia melanogaster* muscleblind (MBL) have been found to regulate circRNA biogenesis and are generally proposed to work in a similar manner, by binding to regions of transcripts with reverse complementarity to form stem loops [202]. Several viral factors have also been suggested to promote back-splicing such as KSHV ORF57 [71]. An alternative biogenic mechanism for some circRNA is that they are derived from excised skipped exons present in RNA lariats that result from canonical linear splicing [205]. Additionally like mRNAs, multiple (up to dozens of) circRNA isoforms (ISO circRNAs) can be generated from the same genes (termed circRNA "hotspots"). Examples include the CSPP1, BIRC6, ATM, MTHFD1, SOX5 and ARHGEF12 and XPO1 [206]. Overall while some determinants and mechanisms for circRNA biogenesis have been proposed, most understanding of their biogenesis remains largely hypothetical.

The extensiveness of circRNA expression wasn't apparent until the application of high-throughput NGS methods [203, 204]. Their detection in such data (as well as bespoke microarrays) depends on the reversed-orientation sequence termed the "backsplice" sequence and it's reversed nature relative to the parental gene (Fig 1.16). This backsplice

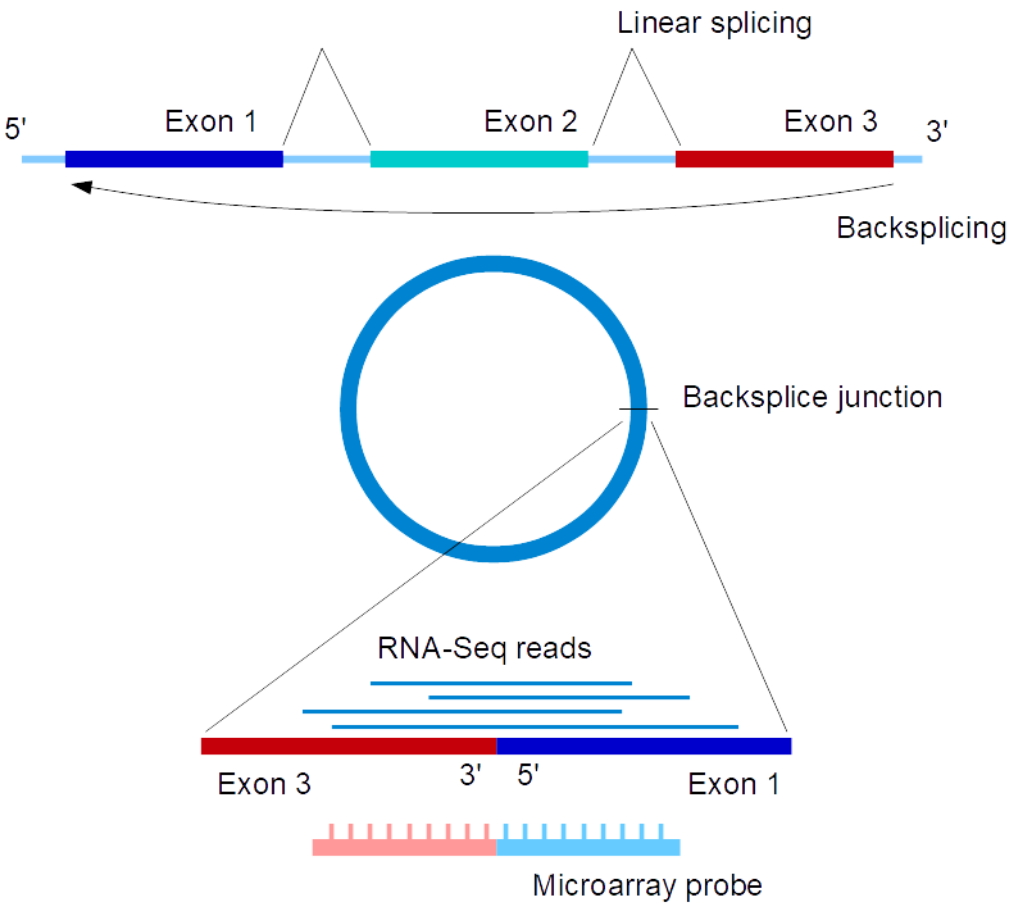


Figure 1.16: Biogenesis and detection of exonic circRNAs.

sequence is in theory infrequent in the transcriptome and thus unique to circRNAs from any one gene, and as such can be used to quantify the relative expression of circRNA at a genome-wide scale. For RNA-Seq, this usually involves mapping reads to artificial references comprising predicted sequences for circRNA backsplice sequences, while for microarrays probes are generated that are complementary to these sequences (Fig 1.16). Using such methods, multiple studies have shown that both on a global scale (ie the expression profile) and at a gene, case-by-case level, circRNA expression varies between different diseases states and physiological conditions [203]. Importantly, the abundance of many linear RNAs and their respective circRNA generated from the same parent genomic loci show poor correlation, indicating that any alteration in circRNA abundance is likely specific, affecting just the circular RNA and not as a result of general up-regulation of the parent genomic loci [200]. Thus high-throughput methodologies have been crucial in the extensive characterisation of circRNA expression, which has enabled follow-up studies showing their context-specific expression, providing evidence for their existence relating to functional entities and not "transcriptional noise" as once assumed.

1.4.7 Functions of circRNAs

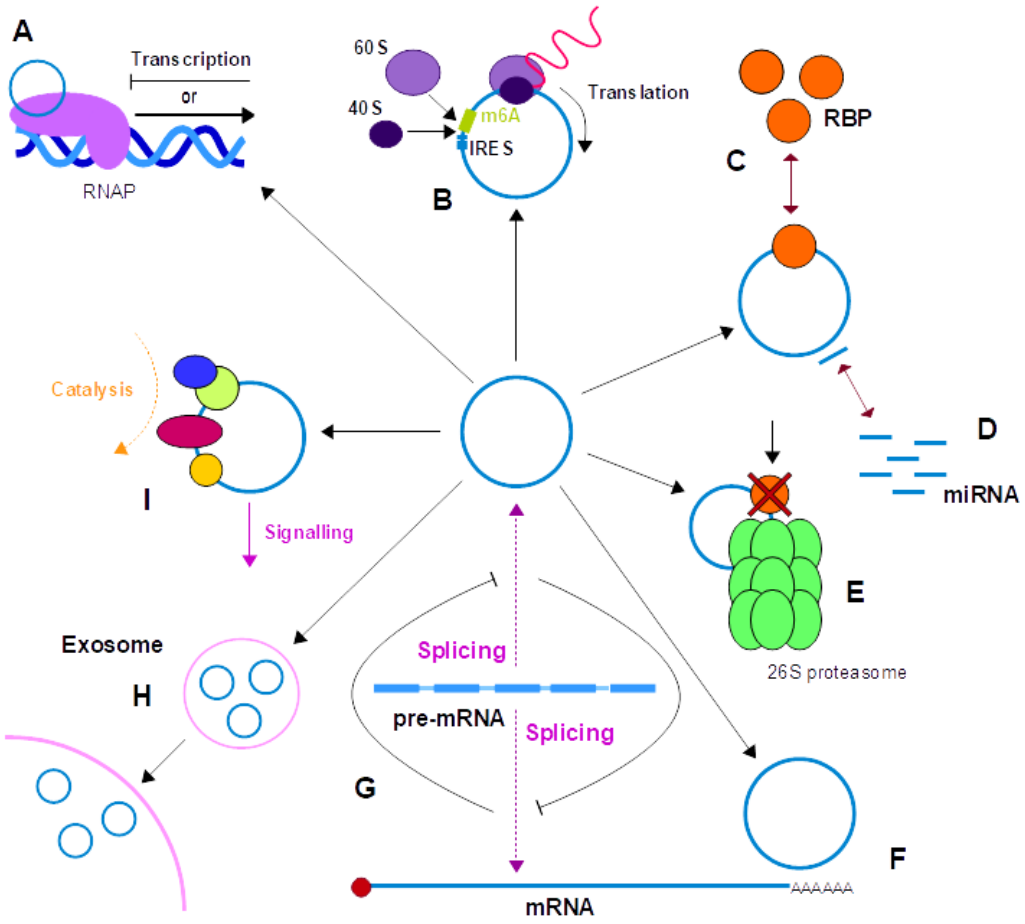


Figure 1.17: Functions of circRNAs. (A) Some circRNAs (in-particular EIciRNAs) regulate transcription of linear mRNA from their parental genes [207]. (B) Some circRNAs can be translated via ribosome loading to M6A sites or IRES-like elements [208]. (C) CircRNAs can act as protein sponges [209]. (D) CircRNAs can sponge miRNAs to de-repress their target transcripts [210]. (E) CircRNAs can expediate the proteasomal degradation of proteins by binding them and bringing them into physical proximiate to the 26S proteasome. (F) CircRNAs can bind to mRNA and stabilise them. (G) Some circRNAs can compete for splicing components or sequester pre-mRNA components from mRNA [210]. (H) CircRNAs can be secreted in exosomes [211]. (I) circRNAs can act as scaffolds or hubs for protein signalling or activity [212].

As with the linear products of genes, circRNAs can be categorised into different biotypes. Several of these biotypes have been described, with the most abundant and extensively studied being exonic circRNAs, which are entirely composed of exons. Further types of circRNAs are less well studied and include intronic circRNAs (ciRNAs), derived from introns and exon-intron circRNAs (EIciRNAs) which contain both introns and exons [207, 213]. Additional circRNAs biotypes include intergenic circRNAs, which aren't attributable to a known gene and those encoded sense or anti-sense overlapping with known genes [206]. Some circRNAs modulate the expression of linear products from

their parent genes. For example circRNAs and EIciRNAs are both nuclearly located and are both proposed to regulate expression from their parental genes by interacting with RNAPII and splicing components (Fig 1.17A) [207, 213]. Much of this relates to their predominant nuclear localisation. More-generally, circRNA biogenesis has been suggested to directly compete with linear splicing thus antagonise mRNA biogenesis from their parental genes (Fig 1.17G) [214].

A majority of interest is focused on cytoplasmically-localised circRNAs which predominantly include exonic circRNA, likely due to these biotypes containing spliced exons with functional exporting signals. Of most relevance to this study is the role of (exonic) circRNAs in binding miRNA and this is further discussed in the next section (Fig 1.17D). However further functions have been described. For example, several circRNAs have been observed to act as protein sponges, sequestering proteins from their functional targets thus acting analagous to miRNA sponges (Fig 1.17C). One example is generalised to many cytoplasmic circRNAs, which act as endogenous inhibitors for protein kinase R (PKR) and are degraded by RNase L upon viral infection (or poly I:C treatment) in order to facilitate PKR activation during innate immune responses [209]. Another mechanism is facilitated by the nuclearly-localised circANRIL, which acts as a competitive inhibitor of pre-rRNA to bind PES1, inhibiting maturation and thus ribosome biogenesis [215]. Additionally several exonic (for example circFOXO3) and at least one intronic (ciINS2) circRNAs can act a scaffold for proteins (Fig 1.17G) [202, 212].

Perhaps unsurprisingly, some reports have suggested that at least a subset of cytoplasmic exonic circRNAs may be translated, which is proposed to be mediated by RNA methylation or small IRES-like sequences (Fig 1.17B) [208, 216]. Therefore like many lncRNAs, their classification as "non-coding" is not strictly true. Nonetheless, one interesting example is a circRNA generated from the β-catenin gene (circβ-catenin), whose circularisation was found to be result in a truncated β-catenin translation product [217]. This shortened isoform was found to stabilise full-length β-catenin by acting as a competitive decoy for GSK3β inactivating phosphorylation. This was found to attenuate Wnt/β-catenin pathway signalling and promote Huh7 liver cancer cell growth *in vitro*, alongside exograft metastasis and tumourigenesis in nude mice [217]. Interest in the translation of circRNA is both interesting from an academic perspective but also for biotechnological reasons as they may represent efficient mechanisms to produce large quantities of protein-products in a stable and programmable manner [**?**].

More esoteric functions of exonic circRNAs have also been described. These include circRNAs that have been found that regulate gene expression by binding to mRNAs to

stabilise them (Fig 1.17F). Moreover, one unique mechanism employed by circPABPC1 involves physically interacting with the 26S proteasome and ITGB1, to bring the latter into proximity with the former and thus inducing the degradation of the latter (Fig 1.17E) [218]. It seems likely that further research into circRNAs will reveal more examples of members with novel functions.

Different circRNA biotypes show differential localisation into the nucleus, cytosol, cytoplasm, ribosome and exosomes [219]. Various factors appear to regulate circRNA localisation, other than the exact biotype of the circRNAs. In terms of exonic circRNAs, length and GC content plays a key role in transportation as longer circRNAs or circRNAs with lower GC contents tend to be cytoplasmic, while the opposite tend to be nuclear [202, 219]. Additionally circRNAs of different lengths are exported via different mechanism [202]. Moreover the predominantly cytoplasmic exonic circRNAs have been proposed to localise to specific sub-cellular regions in a manner similar to mRNA (localisation motifs and RNA bound protein complexes) to facilitate their spatial concentration [219]. Finally, circRNAs can be secreted into exosomes and such presence in blood and interstitual fluid, alongside their stability, has led to their proposed roles as biomarkers [220] (Fig 1.17F). Therefore as with most linear gene products, the localisation of circRNAs is dependent on their biotype and this has consequences for their possible functions.

1.4.8 Circular RNAs as competing endogenous RNAs

As previously stated, circRNA's are proposed to act as miRNA sponges in the context of ceRNA networks. Exonic circRNAs in-particular are generally considered the major miRNA sponge candidates due to their cytoplasmic localisation and stability [202]. Indeed, all validated miRNA sponge circRNAs are exonic, implying that this is a mechanism exclusively carried out by these circRNA biotypes [221]. Stability is a crucial feature of circRNAs relating to their role as miRNA sponges, as while they are generally poorly and slowly expressed, their longevity means that they can accumulate to abundances that can have appreciable biological effects [201, 202]. This, combined with the idea that they may concentrate in localised regions, means that they may be able to fine-tune and direct ceRNA networks and may even introduce thresholding effects on miRNA activity [222].

Examples of circRNAs acting as ceRNAs have been investigated and consistently validated. A prime example is the mouse circRNA circSry, which contains 16 miR-138 binding sites and has been shown to regulate spermatogenesis via modulation of miR-138's target γH2AX [223]. Additionally the first example of a circRNA that could bind a miRNA was the circular CDR1 antisense transcript (circCDR1 or ciRS-7) which contains over

70 binding sites for miR-7 and was found to be up-regulated in colorectal cancer (CRC) tumours [224, 225]. Another circRNA, circHIPK3, sponges a range of miRNA, including miR-193a, miR-29a, miR-29b, miR-30c, miR-124, miR-152, miR-7, miR-338, miR-379 and miR-584, miR-375 and miR-10b-5p [202, 226–228]. Moreover circ-9119 has been found to sponge miR-26a and miR-136 to de-repress TLR3 and RIG-I expression, which suppressed inflammatory responses upon poly I:C treatment of murine Leydig and Sertoli cells [229]. These examples introduce another feature of circRNAs that makes then "ideal" as miRNA sponges, that they can bind multiple miRNAs and thus can can influence many different miRNAs involved in different processes and/or sequester multiple copies of identical miRNAs. Therefore the regulatory potential of circRNAs in as miRNA sponges in ceRNA networks is extensive and supported by case-by-case functional studies.

1.4.9 Circular RNAs and viruses

The first circRNA discovered was actually generated by a virus. This was in a study from 1986 by Kos *et al.,* that applied electron microscopy on the purified ssRNA genome of hepatitis delta virus's (HDV), finding them to be structured as covalently closed loops [230]. However this can be considered a somewhat special case that relates to the uniqueness of the virus. Instead, a more recent and extensive area that circRNAs are of interest in the context of viruses is their role as miRNA sponges. This is largely due to their capacity to regulate 100s of genes meaning that miRNA sponges may represent attractive targets to facilitate viral manipulation of cellular behaviour. These include host circRNAs that show specific disruption during viral infection, alongside the identification of virally-encoded circRNAs by viruses.

Most study into host circRNAs dysregulated by viruses have focused on exonic circRNAs,and in-particular their role as miRNA sponges. One study identified 516 dysregulated circRNAs during early HIV infection that regulated a 21 miRNA, 903 mRNA network that regulated HIV-1 replication by modulating the abundance of IL-5, CCNK and CDKN1A [231]. Another example by Wei *et al.,* 2018 identified 162 differentially expressed circRNAs between chronic HBV and healthy liver samples, with 30 being ISO circRNAs generated from 3 genes, MTHFD1, SOX5 and ARHGEF12 [224]. A similar study used a circRNA microarray to detect circRNAs in clinical HBV-infected hepatocellular carcinoma (HCC) samples that determined the importance of the up-regulated circRNA_100338 in suppressing miR-141-3p, de-repressing targets including the oncogene MTSS1 [224]. These two examples are of relevance to KSHV As they implicate circRNAs

as having putative roles in the development and/or progression of virally-driven cancers.

The only true backspliced virus circRNAs have been identified to be produced predominantly by dsDNA viruses. These include predominantly herpesviruses (γ, KSHV, EBV, rhesus lymphocryptoviruse (rLCV) and MHV68, β, HCMV and α poultry mareks disease virus (MDV)), the carp *Cyprinivirus* Cyprinid herpesvirus 2 (CyHV-2), human β-coronaviruses (SARS-CoV-1, -2 and MERS-CoV-2), human papillomavirus (HPV), human Merkel Cell Polyomavirus (MCV) and rat polyomavirus 2 (Fig 1.18) [23, 23, 26, 232–237]. In addition, there is evidence that some RNA viruses encode circRNAs, including HBV, Respiratory syncytial virus (RSV), Zika virus, Zaire Ebolavirus and HIV (Fig 1.18) [238–240]. In support of circRNA's existence as functional entities that do not arise from mis-splicing, in at least EBV and rLCV, Underleider *et al.,* showed conservation of two latently expressed circRNAs between these two viruses [241]. Moreover orthologous, diverse but low abundance circRNAs have been detected from the OriLyts from KSHV, EBV, rLCV and MHV68 [20, 241]. These examples give tantalising leads into extensive biogenesis of circRNAs by viruses.

Some viral circRNAs have also been functionally characterised. One interesting example is MCV's circMCV-T, expressed during lytic replciation from the T antigen locus and encoded antisense to the viral miRNA MCV-miR-M1 (Fig 1.18) [237]. Such overlap means that the circRNA and miRNA share sequence complementarity and because MCV-miR-M1 inhibits MCV lytic viral replication, promoting latency, its sponging by circMCV-T has been found to help drive the virus into its lytic replication cycle. Conversely HPV encodes a viral circRNA generated from its E7 oncoprotein locus (circE7) that is translated to generate E7 protein that promotes invasion, metastasis and cell growth of infected cells (Fig 1.18) [242]. This circRNA was modified with an N6-methyladenosine (m6A) site that facilitated ribosome loading, a mechanism that is proposed to be important to facilitate the translation of viral and cellular eukaryotic circRNAs [243]. Both these examples exemplify unique mechanisms by how viruses utilise self-encoded circRNAs to regulate their life-cycle or promote their pathogenesis.

CircRNAs roles as miRNA sponges have also been shown to be important during viral infection and virus-associated disease. Several circRNAs have been identified expressed from Epstein Barr Virus (EBV) during both latency and lytic replication (Fig 1.18) [20, 232, 241, 244, 245]. One example is expressed from the latency gene circLMP2A and was found to suppress and promote the expression of miR-3908 and TRIM59, respectively [245]. TRIM59 represses p53 and thus circLMP2A was found to reduce p53 abundance, which promoted cancer stemness of infected cancer cells [245]. In addition

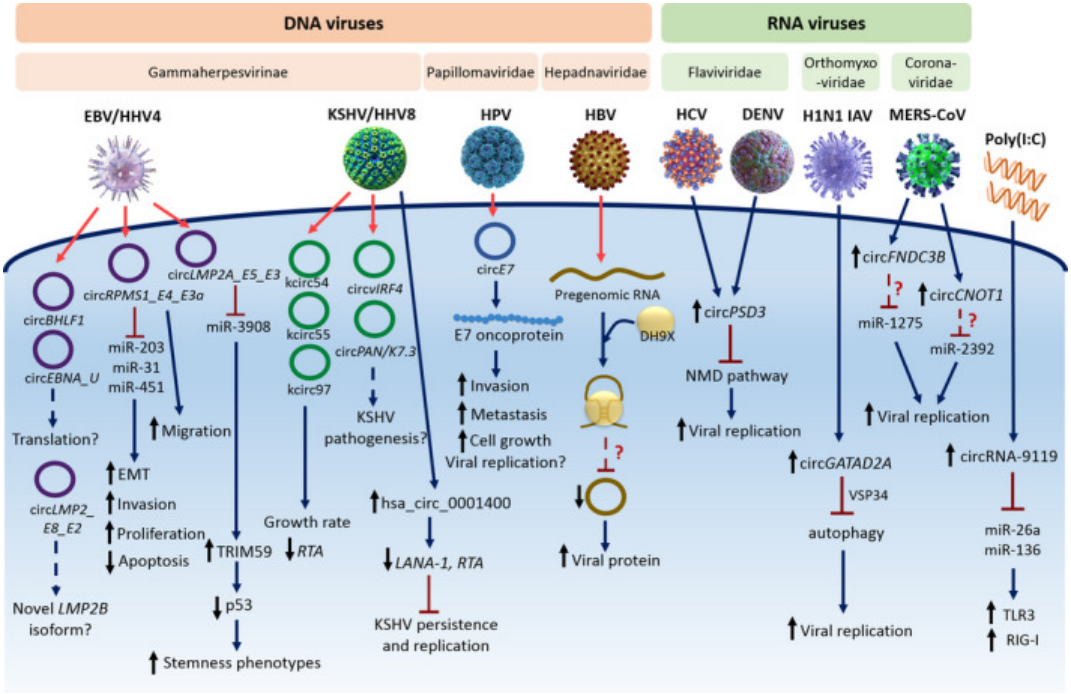


Figure 1.18: Examples of viral circRNAs and circRNAs important to viral infection.

circRNAs have recently been found to be produced by the β-herpesvirus HCMV, which were predicted to sponge a range of miRNAs to modulate DNA replication [246]. In a more large-scale study, Chen *et al.,* 2022 mined viral circRNAs from 41 samples comprising (KSHV-, EBV-, rLCV-, MERS-CoV-, HBV- and HPV-) virally infected cell lines and lympohmas, lung adeno-, cervical and gastric cacrinomas [232]. They identified $\tilde{3}912$ viral circRNAs and by integrating with miRNA target prediction algorithms and databases, identified novel miRNA-circRNA-miRNA axes including 21 KSHV circRNAs with miR-6848-5p, 18 with miR-3085-5p, 10 with either miR-6858-5p or miR-8063 [232]. This indicates that the interactions between viruses and ceRNA networks may be extensive in scale, frequent between viruses and biologically relevant to virus-associated disease.

1.4.10 KSHV-encoded circRNAs

Like HCMV's miRDE and HVS's HSURs, KSHV's lncRNAs were initially proposed to act as miRNA sponges, given the high abundances that they reached, particularly PAN. Surprisingly however such transcripts have not been predicted or reported to bind miRNA, making their role as miRNA sponges unlikely [247]. However both host and viral circRNAs have emerged as having roles in ceRNA networks as miRNA sponges, regulating processes relevant to KSHV biology and pathology, including cellular proliferation, motility and metastasis, apoptosis, cell cycle progression and angiogenesis [216].

Several KSHV-encoded putative circRNAs have been detected and a handful have been

validated. The most abundant and well-studied are those generated from the vIRF4 locus and to a lesser extent the PAN/K7.3 locus (Fig 1.18) [23, 24, 244]. Notably however, there is some ambiguity for the latter circRNAs as they are co-expressed with linear PAN, circRNAs backsplice reads attributed to them were detected from antisense overlapping loci and they have not been detected in all studies into KSHV-generated circRNAs [23, 24, 27, 244]. However both circvIRF4, circPANs and circK7.3s have been detected in abundances greater than LANA in KS, PEL and MCD tissue, in at least two studies, indicating their expression *in vivo* [23]. They have also been found to be incorporated into virions [26]. Further detected KSHV circRNAs include those overlapping with ORF34 (kcirc54), ORF35/36 (kcirc55) and ORF36/37 (kcirc57), alongside one encoded within the T0.7 region which also encodes miR-K12-10 and -12, which show 5' and 3' complementarity, leading to the proposal that circT0.7 may be able to bind these viral miRNAs analogously to MCV's circMCV-T and MCV-miR-M1(Fig 1.18) [23, 26, 26].

Most detected KSHV circRNAs are exonic, however intron-retained circRNAs have been detected produced from KSHV's vIRF4 and vLCV's RPMS1 gene [20, 26]. Limited function of each circRNA has been determined except for kcirc54 and kcirc55 which were found to suppress PAX2 and promote EGRR3 and FGF13 mRNA upon over-expression, relative to a circGFP control, alongside targeting genes enriched for cell cycle progression as well as p53 signalling [23, 24]. Additionally, like for some lncRNAs, KSHV circRNAs have been suggested to directly bind and inhibit mRNA that they share complementarity for, and Tagawa *et al.,* found enrichments for cell cycle progression and Wnt/Notch signalling in such mRNAs [23]. Moreover the presence of detected circRNAs originating from regions overlapping the OriLyts of γ herpesviruses may suggest a role in viral genomic replication. It should be noted that such detection is predominantly derived from RNA-Seq data alone and thus is relatively poor evidence, without follow-up biochemical validation and functional studies.

In terms of biogenesis, KSHV ORF57 is believed to be involved as its knockdown has been shown to decrease the number of detectably differentially expressed circRNAs, relative to unaltered controls [27]. Similarly, vIRF1 has been found to promote the expression of at least one host circRNA [235]. Moreover there is some evidence suggest and involvement that the RBPs MYBL2, SRSF3, FUS, QKI and RBM33 [24]. However much of this research is yet to be formalised.

1.4.11 KSHV and host circRNAs

A handful of studies have investigated the dysregulation of host circRNAs during infection by KSHV and, compared to viral circRNAs, such circRNAs have better functional characterisation as miRNA sponges [23, 24, 71]. Alongside identifying and characterising vIRF4, one set of studies by Tagwa *et al.,* identified hsa_circ_0001400 (encoded by RELL1, circRELL1) that was up-regulated upon KSHV infection and first found to suppress viral gene expression [23]. It was later found to also be induced upon infection by EBV and HCMV and found to specifically inhibit lytic transcription, promoting latency of KSHV in primary endothelial cells [248]. Loss of function mutation studies indicated that circRELL1 promotes PI3K/Akt, pro-inflammatory (TNFα, IL-6) and T cell signalling while suppressing PTEN and GADD45 signalling, alongside directly interacting with and stabilising the mRNA of the mTOR complex gene TTI1 [248]. An additional circRNA found to be up-regulated in KSHV infected cells is circARGEF1, which was found to be induced by vIRF1 binding to ARGEF1's promoter in-complex with LEF1 [235]. CircARGEF1 was then found to sponge the tumour suppressor miRNA miR-125a-3p, resulting in subsequent de-repression of GLRX3 to promote pro-oncogenic processes such as angiogenesis, cell motility and proliferation [235]. In another study, by comparing the transcriptomes of WT and vFLIP-mutant KSHV-infected lytically replicating iSLK cells, Sheng *et al.,* identified a ceRNA network comprising 16, 90, 40 and 6 differentially expressed circRNAs, lncRNAs, mRNAs and miRNAs, respectively [249]. This network was found to be enriched for riboflavin and thiamine metabolism. The authors then went on to validate 2 ceRNA axes, lncRNA AL031123.1/hsa-miR-378i/SPEG/FOXQ1 and hsa_circ_0070049/hsa-miR-378i/SPEG/FOXQ1, whereby vFLIP mutation down-regulated the constituent ceRNAs, up-regulating miRNAs and thus suppressing SPEG and FOXQ1 [249]. Therefore while few in number, the studies that exist have provided some interesting examples of which host circRNAs are dysregulated during KSHV infection as well as providing some functional association of relevance to KSHV's life cycle and pathology.

Finally in a recent study by Harper and other members of our group, we have shown that circHIPK3 is up-regulated upon induction of and essential for KSHV lytic replication in a manner dependent on ORF57 [71]. Next circHIPK3 was found to sponge miR-30c and -29b, with the latter found to de-repress its target the Notch ligand DLL4 which was found to be important for KSHV replication in B cells, alongside having pre-defined roles in KSHV pathogenesis and angiogenesis [71]. However this study was relatively small-scale in that it focused on just a single circRNA and at most 3 miRNAs and as such a more broad perspective may provide further scope for the extent of circRNA dysregulation

during KSHV reactivation.

1.5 Aims of the study

For the KSHV ceRNA network chapters (Sections 3 & 4):

- Quantify changes to the expression of circRNAs, miRNAs and mRNAs upon the induction of lytic reactivation of KSHV.
- Model the relations between these features in the form of a predicted ceRNA network.
- Analyse this network in order to determine the purpose of its dysregulation and how this may influence KSHV.

For the KS biopsy sample co-expression network chapter (Section 5):

- Build a co-expression network model for KS biopsy tissue from publicly available bulk RNA-Seq data.
- Interrogate this network model by applying module analyses in order to characterise the transcriptome of KS lesions.
- Identify a set of co-expressed and differentially co-expressed hub genes as candidate drivers of KS lesion development.

2 Methods

Given the nature of this thesis being largely in two parts (split between the first and second results chapters [Section 3 & 4] on circRNA-miRNA ceRNA networks, and the third results chapter [Section 5] on co-expression networks applied to RNA-Seq data derived from KS lesion tissue biopsies), the following Methods chapter is split into 3 sections. The first (Section 2.1), comprises methods common to all results chapters, the second (Section 2.2) comprises methods unique to the first and second results chapters (Section 3 & 4) and the third (Sections 2.3) contains methods unique to the third results chapter (Section 5).

2.1 Common Methods

2.1.1 Annotation and Genome Data Accession

The *Homo sapiens* GRCh38 genome build .fasta and annotation .gtf files were downloaded from Gencode (v43). The NC_009333.1 GK18 KSHV genome build was downloaded from the NCBI website while the .gtf annotations were based on the specific unique coding sequence (UCDS) feature annotation provided by Prof Tim Rose of the University of Washington (described more in Section 2.1.2) [15, 18]. These annotations were used for all analyses, unless otherwise stated.

2.1.2 Gene Expression Data Normalisation

The UCDS schema views the counts of downstream ORFs as a linear combination of the number of reads mapping to that ORF relative and the number of reads mapping to the upstream ORFs sharing the same polycistronic transcript [18, 250]. To account for this, read counts for upstream ORFs were subtracted from downstream ORFs that shared polycistronic transcripts, with annotations derived from Bruce *et al.,* and Rose *et al.,* and each pairwise subtraction detailed in Table 2.1. Mapping regions between ORFs were shrunk to be at least 75 bp apart to limit read mapping ambiguity. Additional changes included separating ORF73 into two separate features; ORF73a and ORF73b in order to mitigate difficulty aligning to the internal repetitive region present in the native ORF.

Data were then normalised via a series of steps common to most transcriptomic data analyses. Host and viral count data were subject to TPM-normalisation, using the sum of gene-wise exon and UCDS feature lengths for host and viral genes as feature lengths, respectively. Alternatively, "Counts-per-million" (CPM)-normalisation was performed us-

Downstream ORF	Upstream ORF
ORF8	ORF7
ORF9	ORF8
ORF10	ORF9
ORF11	ORF10
K2	ORF2
K4.1	K4.2A
K4.2A	K4.2
PAN	K7
ORF17.5	ORF17
ORF19	ORF20
ORF22	ORF21
ORF23	ORF24
ORF31	ORF30
ORF32	ORF31
ORF33	ORF32
ORF35	ORF34
ORF36	ORF35
ORF37	ORF36
ORF38	ORF37
ORF42	ORF43
ORF45	ORF46
ORF46	ORF47
ORF47	ORF48
K8a	ORF50
ORF57	ORF56
ORF58	ORF59
ORF60	ORF61
ORF64	ORF63
ORF65	ORF66
ORF66	ORF67
ORF67	ORF67A
ORF69	ORF68
ORF71	ORF72
ORF74	K14
ORF75	K15a

Table 2.1: Table of KSHV UCDS upstream ORFs to subtract (right) from downstream ORFs (left).

ing the cpm() function from the Bioconductor package edgeR (3.40.2) [251]. In most cases, normalised data were pseudocount, Log2 transformed ($log2(x+1)$, where x is the gene-wise TPM/CPM).

2.1.3 Hierarchical Clustering Analyses

Correlation heatmaps were constructed predominantly using Spearman's rank correlation unless otherwise specified. Hierarchical clustering analyses for heatmap figures were performed using the default settings of the R package ComplexHeatmap and the hclust()

functions, using agglomerative average linkage clustering [252]. Euclidean distance was used primarily as a distance metric, unless otherwise specified.

2.1.4 Principal Components Analyses

Gene expression values were either transformed via CPM- and TPM-normalisation for ceRNA miRNA/RNA-Seq and KS biopsy RNA-Seq data, respectively, prior to principal components analyses (PCA). For all PCAs, centered and scaled Log2 gene-wise expression values were used as input. For the full and lesion-only comparisons, all detected genes were used as input, while for module eigengene (ME) computation, PCAs were performed on the module-wise gene expression matrices, as further detailed in Section 2.3.7. Raw circRNA microarray data was quantile-normalised prior to PCA. PCAs were performed using the prcomp() function of the factoextra R package.

2.1.5 Centrality analysis

Hub scores were calculated using the Influential R package (v2.2.6) [253]. Hub scores for ceRNA networks (Section 4.4) were calculated as directed, not allowing loops and measuring both weighted in-degree and out-degree. Hub scores for WGCNA hub networks (Section 5.7) were calculated with the undirected setting and incorporating weight. Geodesic paths were calculated using the centiserve package (v1.0.0), selecting for a maximum of 2-steps, ignoring edge weights and for both in- and out-degree. K-core decomposition was performed using igraph, selecting for K = 1, 2, 3 and 4 and set to mode="all" (ignoring edge direction).

2.1.6 Gene Set Annotations

Gene sets were derived from a range of sources. These include the gene ontology (GO) database, primarily using the biological process (BP) sub-categorisation. Endothelial cell (EC), fibroblast and various immunological cell type-specific gene signature gene sets were accessed via the MCPcounter R package, or from the SCtype database of single-cell marker genes [254, 255]. Additional single-cell gene sets were defined from Tabtieng *et al.,* [89].

Additional custom host gene sets were constructed. These comprise lymphatic EC (host LEC) and blood EC (host BEC) cell-specific gene signatures that were derived from a literature search of genes that were validated to be up-regulated in LECs relative to BECs and vice versa. Additionally, a set of 163 host genes whose expression was shown to be

specifically and directly activated by ORF50 in a study by Papp *et al.,* [60]. The 163 genes were filtered based on the presence of called RTA peaks in their promoters with concomitant significant up-regulation in a differential expression analyses. Such genes were termed "core RTA-inducible genes" or "Core-RIGs". Additionally, a set of 167 host genes, found to be significantly enriched (FDR $<$ 0.25) in a CRISPR-Cas9 screen of latently infected TIME cells were derived from Holmes *et al.,*. These genes were termed "Essential to infection" or "Holmes 2020" [57].

Viral gene sets were also defined. These predominantly derived from Bruce *et al.,*, updated with observations from more recent studies. These sets used included grouping viral genes based on 9 functional categories (Table 2.3), the maximal expression timing of genes post lytic reactivation in BCBL1 cells (Table 2.4) and which viral genes were known to be directly RTA-activated (Table 2.5) [15]. An additional annotation of viral genes was based on more classical segregation into classical and relaxed latent, immediate early, early and late lytic genes (Table 2.2). Additionally gene sets were defined as those that showed cell line-specific enrichment from Bruce *et al.,* (defined by cell line-specific Z-tests for significant deviation from mean expression across cell lines, $p<0.05$) [15].

Gene Classification	Viral Genes
Total	All 93 viral genes in the network (below)
Classical latent	ORF71, ORF72, ORF73, K12A, K12A, K12B, K12C, viral miRNAs (mIRregion)
Relaxed Latent	ORF74, ORF75, K1, K2, K14 K15, vIRF-3
Immediate early lytic	ORF45, ORF50, ORF57, K4.2A, K8
Early Lytic	ORF2, ORF4, ORF4A, ORF6, ORF7, ORF9, ORF10, ORF11, ORF16, ORF17, ORF17.5, ORF18, ORF21, ORF24, ORF29A, ORF29B, ORF30, ORF31, OF34, ORF35, ORF36, ORF37, ORF38, ORF40. ORF40A, ORF41, ORF44, ORF46, ORF48, ORF49, ORF54, ORF56, ORF59, ORF60, ORF61, ORF62, ORF63, ORF64, ORF66, ORF67, ORF67A, PAN, vIRF-1, vIRF-2, vIRF-3, PAN, K1, K3, K4, K4.1, K4.2, K5, K6, K7, OLAP
Late Lytic	ORF8, ORF19, ORF20, ORF22, ORF23, ORF25, ORF26, ORF27, ORF28, ORF32, ORF39, ORF42, ORF43, ORF47, ORF4B, ORF52, ORF53, ORF55, ORF58, ORF65, ORF68, ORF70, ORF7

Table 2.2: Table of traits relating to viral gene expression.

Gene Function	Viral Genes
Capsid	ORF17.5, ORF25, ORF26, ORF43, ORF62, ORF65, ORF66
Envelope & Membrane	ORF8, ORF22, ORF23, ORF28, ORF39, ORF47, ORF48, ORF53, ORF58, K4.2A, K8.1
Gene Expression	ORF11, ORF18, ORF30, ORF31, ORF34, ORF37, ORF49, ORF50, ORF57, PAN, K8, K8a
Immune Modulation	ORF4, ORF4A, ORF4B, ORF10, ORF45, ORF63, ORF74, ORF75, K1, K2, K3, K4, K4.1, K4.2, K5, K6, K14, vIRF-1, vIRF-2, vIRF-3, vIRF-4
Mitogenesis & Cell Cycle	ORF16, ORF71, ORF72, ORF73, K7, K12A, K12B, K12C, K15
Replication	ORF2, ORF6, ORF7, ORF9, ORF20, ORF36, ORF40, ORF40A, ORF41, ORF44, ORF46, ORF54, ORF56, ORF59, ORF60, ORF61, ORF70, OLAP
Tegument	ORF19, ORF21, ORF32, ORF33, ORF52, ORF64
Virion Assembly & Egress	ORF17, ORF24, ORF29A, ORF29B, ORF38, ORF42, ORF55, ORF67, ORF67A, ORF68, ORF69
NA	mIRregion (viral miRNAs), ORF35, ORF27

Table 2.3: Table of traits relating to the function of viral genes.

Maximal Expression Timing	Viral Genes
Latent	ORF71, ORF72, ORF73, K12A, K12A, K12B, K12C, mIR-region (viral miRNAs)
Primary (0-10hrs)	ORF11, ORF57, ORF50, K8, ORF4, vIRF-4, ORF10, K6, K5, K4, K2, ORF45, ORF74, K14, ORF16, K7, ORF9, ORF6, ORF69, ORF55, ORF67
Secondary (10-24hrs)	ORF66, ORF65, ORF26, ORF53, ORF47, ORF38, ORF48, ORF8, ORF18, ORF49, ORF34, vIRF-1, vIRF-2, vIRF-3, ORF4B, ORF4A, K3, K4.1, K4.2, ORF27, ORF70, ORF61, ORF59, ORF46, ORF2, ORF56, ORF52, ORF64, ORF42, ORF38, ORF17
Tertiary (48-72hrs)	ORF43, ORF25, ORF62, ORF23, ORF22, K8.1, ORF28, ORF37, ORF30, ORF31, K1, ORF75, K15, ORF20, ORF44, ORF41, ORF40A, ORF36, ORF54, ORF21, ORF32, ORF63, ORF19, ORF29A, ORF29B, ORF24
NA	ORF17.5, ORF58, K4.2A, PAN, K8a, ORF35, OLAP, ORF60, ORF7, ORF40, ORF33, ORF67A, ORF68

Table 2.4: Table of traits relating to the expression timings of viral genes.

RTA-Responsiveness	Viral Genes
Responsive	K12A, PAN, K8a, ORF57, K2, K14, ORF6, ORF8, vIRF-1, K3, ORF59, ORF56, ORF52, K8.1, K8a, ORF37, K1, K15, ORF21
Non-Responsive	ORF73, ORF72, ORF71, mIRregion, K12A, K12B, K12C, ORF17.5, ORF58, K4.2A, ORF35, OLAP, ORF60, ORF7, ORF40, ORF33, ORF67A, ORF68, ORF11, ORF50, K8, ORF4, vIRF-4, ORF10, K6, K5, K4, ORF45, ORF74, ORF16, K7, ORF9, ORF69, ORF55, ORF67, ORF66, ORF65, ORF26, ORF53, ORF47, ORF39, ORF48, ORF18, ORF49, ORF34, vIRF-3, vIRF-2, ORF4B, ORF4A, K4.1, K4.2, ORF27, ORF70, ORF61, ORF46, ORF2, ORF64, ORF42, ORF38, ORF17, ORF43, ORF25, ORF62, ORF23, ORF22, ORF28, ORF30, ORF31, ORF75, ORF20, ORF44, ORF41, ORF40A, ORF36, ORF54, ORF32, ORF63, ORF19, ORF29B, ORF29A, ORF24

Table 2.5: Table of traits relating to responsiveness of viral genes to RTA.

Cell Line	Viral Genes
Vero	K6, OLAP, ORF4B, K1, K3, K4, K4.1, K4.2, K5, K7, ORF23, ORF4A, ORF56, ORF58, ORF60, ORF64, ORF66, ORF70, ORF74
BCBL-1	K14, K2, K6, K8a, OLAP, ORF10, ORF32, ORF38, ORF45, ORF49, ORF4B, ORF50, ORF54, ORF71, ORF73, PAN, vIRF-1
TIME	K4.1, ORF74, ORF56, K2, K8a, ORF10, ORF32, ORF50, ORF54, K12A, K12B, K12C, K15, K4.2A, K8.1, mIRregion (viral miRNAs), ORF11, ORF25, ORF26, ORF29A, ORF34, ORF35, ORF36, ORF37, ORF41, ORF46, ORF47, ORF57, ORF6, ORF61, ORF62, ORF63, ORF67A, ORF7, ORF72, ORF8, ORF9, vIRF-2, vIRF-3, vIRF-4
BEC	K4.1, ORF26, ORF23, ORF28, ORF29B, ORF31, ORF67
LEC	mIRregion (viral miRNAs), K5, ORF68

Table 2.6: Table of traits relating to the highest expressed viral genes in a panel of cell lines.

2.2 Competing endogenous RNA network specific methods

2.2.1 Sample Design

Data used in Section 3 & 4 were generated from RNA extracted from TREx-RTA-BCBL1 cells prior to (latently infected) or a set number of hours post-treatment with Dox (lytically replicating) [256]. Three primary sets of data were used: miRNA-Seq, total RNA-Seq and a circRNA microarray dataset. To enrich for the miRNA fraction, extracted RNA was run on an agarose gel and bands that corresponded to RNA $<$ 200 nucleotides in length were extracted, while total RNA was rRNA-depleted for the total RNA-Seq and RNA was treated with RNase R to deplete linear RNAs and thus enrich for circular transcripts for the circRNA microarray. The small and total RNA-Seq were generated by the Whitehouse Lab at the University of Leeds, namely Belinda Banquero and Becky Foster, respectively and were sequenced at the Next-Generation Sequencing (NGS) facility at St James's Hospital, Leeds. RNA for the circRNA microarray was extracted by Katie Harper and sent to Arraystar for RNase R-depletion, hybridisation and within-array normalisation. Total RNA-Seq samples were taken pre-Dox treatment (0hr) and 8 and 20hrs post-Dox treatment. miRNA-Seq samples were taken pre-Dox treatment (0hr) and 16 and 24 hrs post-Dox treatment. CircRNA microarray samples were taken pre-Dox treatment (0hr) and 20hrs post-Dox treatment. The total RNA-Seq and circRNA microarrays were done in duplicate, while miRNA-Seq was done in triplicate. Note that none of the total RNA-Seq, circRNA microarray or miRNA-Seq data were taken from the same cells or during the same experiments, however experimental culture conditions and extraction methods were maintained as constant. These data were used for both Section 3 & 4.

2.2.2 Annotation and Genome Data Accession

Viral transcript 3' UTR annotations were derived from McClure *et al.,* 2013 [66]. Host and viral miRNA annotations were derived from miRBase [257]. All other genome annotation and sequence files were identical to those used in Section 2.1.1.

2.2.3 Differential expression analyses

For miRNA-Seq, all genes were filtered to remove those with zero counts in all samples, while for total RNA-Seq, genes with less than 3 counts in any sample were excluded from the analysis. Differential expression of miRNA-Seq and total RNA-Seq was performed using DESeq2 (v1.38.3), using DESeq2's standard "median-of-ratios" (MOR) normali-

sation approach [258]. Additionally the differential gene elimination strategy (DEGES), implemented in the TCC package, v1.38.0, was also used to identify non-differentially expressed genes on which scaling factors were calculated from [259]. DEGES was performed with 5 iterative rounds of 5000 genes. KSHV features were excluded prior to calculation of size factors and were re-introduced prior to normalisation by scaling factors. DEGES scaling factor calculation, library normalisation and DESeq2 differential expression were performed independently at the gene- and transcript-level.

CircRNA microarray samples were initially processed by Arraystar, then normalised via quantile-normalisation using the normalize.quantiles() function of the Bioconductor package preprocessCore (1.60.2). CircRNA microarray differential expression was performed using the limma package (v3.54.2).

Raw p-values for all differential expression results (circRNA microarray, total and small RNA-Seq) were transformed by Benjamini-Hochberg (BH) method to adjust for multiple testing and circRNA and small RNA-Seq differential expression results filtered by BH-adjusted P-value (FDR) < 0.05, while total RNA-Seq differential expression results were filtered with an FDR$<$0.01. Differentially expressed circRNAs were filtered to remove those with $|\text{log2FC}| < 0.65$ and log2AE > 8.

2.2.4 KSHV host shut-off factor SOX binding site analyses

SOX cleavage site data and python scripts to identify putative SOX sites were downloaded from the supplementary data of Gaglia *et al.,* [77]. Single nucleotide SOX cleavage site locations were extended by 25 nt each side and converted to a BED file from which fasta sequence files were assembled and these input to the “create_pwm_wbkg.py” script as in Gaglia *et al.,* to generate a SOX motif position weight matrix (PWM). This PWM was then scanned against transcripts using the “score_motif_wholetransc.py”.

2.2.5 Competing endogenous RNA network construction via miRNA target prediction

MiRNA-Target prediction between miRNAs and targets (circRNAs, viral and host protein-coding gene 3’ UTRs) was performed using the miRanda and RNAhybrid miRNA target prediction algorithms [260]. MiRanda was run with defaults while RNAhybrid was performed allowing for a maximum target length of 25000 and constraining the binding helix to be between 2’ and 9’ relative to the miRNA 5’ end. Putative miRNA-target interaction pairs were taken as the intersection between the miRNA-target pairs predicted by mi-

Randa and RNAHybrid with a miRanda score $>$ 150 and RNAhybrid minimal-free-energy (MFE) $<$ -10. Gene-level 3' UTRs were defined as being any 3' UTRs in transcripts originating from a gene that were detected in the samples and as differentially expressed (FDR$<$0.05). Cellular 3' UTR sequences were accessed from biomaRt (v2.54.0). CircRNAs used for network construction were as described in Section 2.2.3, additionally retaining only exonic circRNA and then filtering those less than 200nt and greater than 10000nt. Differentially expressed protein coding/viral genes were those identified using the DESeq2 MOR method as in Section 2.2.3. CeRNA network construction Kamada-Kawai layout algorithm implementation and visualisation was primarily performed using the igraph R package (v1.4.1).

To confirm the specificity of the differentially expressed miRNAs for circRNAs permutation tests were employed. This involved randomly selecting 310 (the same number as the number of differentially expressed candidate exonic circRNAs chosen for network construction) detected (above expression thresholds, see Section 2.2.3) circRNAs and performing miRNA target predictions. 2500 permutations were performed and p-values calculated as the proportion of permutations where the number of miRNA binding sites in randomly selected exonic circRNAs exceeded that for the 310 differentially expressed exonic circRNAs.

2.3 Kaposi sarcoma co-expression network specific methods

2.3.1 Description of the data-set

The data-set generated by Lidenge *et al.,* 2020 under accession code GSE147704 was used for the present study [67]. The study carried about by these authors was concerned with comparing the gene expression profiles of KS lesion samples relative to matched controls and comparing differential expression between epidemic and endemic KS lesion samples. The data-set comprises RNA-Seq data from 27 patients; 3 (male) with no KS presentation or HIV-1 infection, or any other known relevant morbidities (hereby termed "healthy" biopsy samples) and 24 (16 males and 8 females) from patients with cutaneous Kaposi's Sarcoma (KS) (Table 2.7). From the 24 KS patients, biopsies were taken from KS lesions (hereby termed "lesion" or "KS lesion" biopsy samples) and un-involved control tissue (hereby termed "control" biopsy samples) and sequenced, resulting in a matched set of 48 lesion and control samples. Control samples were predominantly taken controlaterally, except for patients 3139, 037 and 3122, where, due to KS involvement on the contralateral surface, samples were taken ipsilaterally [67].

Patients could be further stratified based on whether patients had HIV-1 co-infection or not, otherwise known as "epidemic" KS and "endemic" KS, respectively, with 18 patients being epidemic, 6 patients being endemic (Table 2.7). Additionally, lesion samples could be stratified based on whether biopsies were taken from differing morphologies: 13 nodular, 10 plaque and one patch (Table 2.7). However as the single patch lesion represented the lone resident of its group, it and its paired control sample were discarded, leaving 23 patients, each with paired lesion and control samples.

Additional patient-wise traits were recorded, including months of detectable KS (ranging from 0.8 to 48 months and median 4), months of anti-retroviral therapy (ART) therapy (ranging from 0 to 121.7 and median 0), detectable serum KSHV levels, which ranged from below detectable levels (BDL) to 28,000 cps/mL and HIV-1 serum levels in epidemic samples, which ranged from 0.7 cps/mL to 100,000,000 cps/mL and 0 in all endemic samples. Patients had a minimum age of 24, a median age of 32 and maximum age of 66.

Library sizes ranged from 8-40 million reads and were sequenced at a mean read length of 75bp. Data were split between 4 batches, with most traits being balanced between them (Fig 2.1). This was except for the endemic samples, which were entirely within batch 2 and as such confounded. The possible consequences of this are explored further further in Section 5.2.

Sample Type	N. Samples	HIV-1	Description
Healthy	3	None	Biopsy Skin samples taken from patients with no KS, no AIDS or any known morbidities
KS lesion biopsy	18	Epidemic	Biopsy samples taken from KS lesions present on the skin of patients with epidemic KS (AIDS co-morbidity)
KS lesion biopsy	6	Endemic	Biopsy samples taken from KS lesions present on the skin of patients with endemic KS (no AIDS co-morbidity)
Non-KS control biopsy	18	Epidemic	Biopsy samples taken from non-KS region of skin on patients with epidemic KS (AIDS co-morbidity)
Non-KS control biopsy	6	Endemic	Biopsy samples taken from non-KS region of skin on patients with endemic KS (no AIDS co-morbidity)

Table 2.7: Features and Layout of Samples in Data-Set

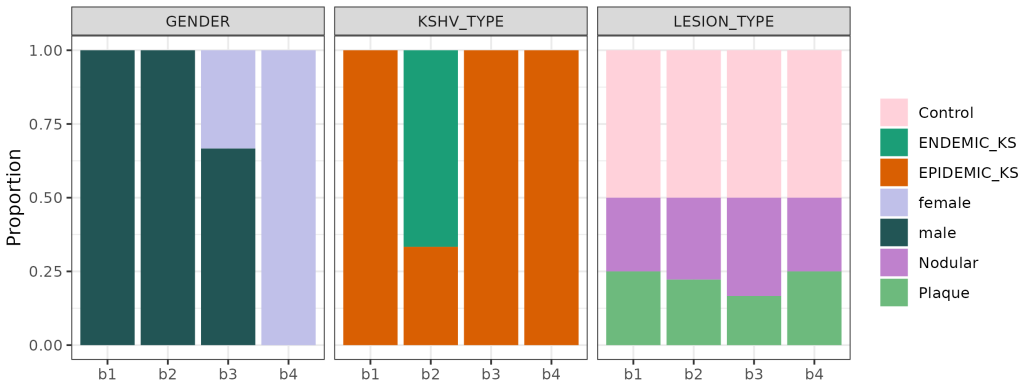


Figure 2.1: Distribution of sample-wise traits between sequencing batches. The data-set generated by Lidenge *et al.,* 2020 was annotated with several sample-wise measured traits. The proportion of samples belonging to individual sample traits is presented. Samples are split between batched (b1, b2, b3 & b4).

2.3.2 Dataset justification and prior recapitulation of Lidenge *et al.,*'s findings

This dataset was selected for study due to it being the first publicly available cohort of KS tumour biopsy tissues samples that was of sufficient size to enable reliable calculation of empirically-defined gene expression profile similarity measures and thus co-expression analyses. Moreover, the presence of matched lesion and control samples facilitated the inclusion of information about genes different (in terms of expression and co-expression)

specifically in lesion vs non-lesion tissue, while accounting for between-patient heterogeneity. These data were used for all analysis in Section 5.

Much of the analysis performed by Lidenge *et al.,* was recapitulated during the course of this study and some of the results are demonstrated as figures in the Appendix (Section 7). We were able to reconstitute all of their findings and slightly adapt them for the purposes of the present study. These included the following:

1. PC1 of a paired lesion/control sample PCA separated by healthy and control samples (Fig S5).
2. Endemic lesion samples showed a greater magnitude of dysregulation of gene expression relative to control samples, both in terms of both log2 fold-change of differential expression and also variance components in the paired samples PCA (Fig 5.2 & S5).
3. Healthy and control samples showed no detectable differences in gene expression (data not shown, see Lidenge *et al.,*).
4. Lesion samples in the dataset showed predominantly latent, lytic or mixed expression profiles (data not shown, see Lidenge *et al.,*).

Note that because of finding 3, we discarded the healthy biopsy samples from subsequent analyses.

2.3.3 RNA-Seq Library Data Accession and Processing

Lidenge *et al.,* RNA-Seq .fastq files were accessed from the NCBI Sequence Read Archive (SRA) via its European Nucleotide Archive (ENA) proxy server, using the Enabrowsertools command line tool. These data were generated using the Illumina's TruSeq RNA Library Prep Kit to generate stranded and single libraries, which were sequenced on an Illumina's NextSeq HighOutput v2 platform. Sample metadata was accessed from the source paper, Lidenge *et al.,*, or using the R package GEOquery (v2.66.0). All subsequent data quality control, processing, alignment and quantification was performed using a bespoke Snakemake pipeline using various UNIX command-line based tools, unless otherwise stated. Firstly, .fastq data were quality filtered using fastp (v0.23.2), with minimum quality and length 20 and 25, trimming any 3' poly nucleotide tracts of 10 or greater and using the Illumina universal adapter sequence "AGATCGGAAGAG" to remote adapter sequences [261]. Fastqc and MultiQC were used to generate .fastq quality reports.

Alignment was then performed using the STAR read aligner [262]. Firstly, separate genome indices were constructed for the host genome and viral genomes respectively, using a bash script with the STAR index function, using arguments –sjdbOverhang of 74 (mean read length) and –genomeSAindexNbases 14 or 7 for host and viral genomes, respectively. RNA-Seq libraries were then aligned against the host genome index, using arguments –alignIntronMax 250000, –outSAMmultiNmax 5 and unaligned reads were then extracted as .fastq files. These unmapped reads were then aligned against the viral genome with options –alignIntronMax 10000 and –outSAMmultNmax 1. Both rounds of alignment used the options –outMultimapperOrder Random, –outSAMtype and BAM SortedByCoordinate. Genes-level counts were then quantified using HTSeq-count, with arguments -m "intersection-nonempty", -s "no", -i "gene id" and feature set to "gene" [263].

2.3.4 Differential expression analyses

Differential expression analyses were performed using the DESeq2 and Limma R packages [258, 264]. These included comparing all lesion samples to matched control biopsies and comparing paired lesion and control samples for endemic and epidemic patients separately. For the prior, un-transformed RNA-Seq count data were used, while batch, sex, HIV-1 co-infection (epidemic or endemic KS), months of ART, months of KS, HIV serum level, age and HIV-1 co-infection were all included as covariates/factors as model parameters. DESeq2 was used to perform this differential expression, filtering with an FDR$<$0.001 and $|\text{log2FC}| > 0.5$. For the separate endemic and epidemic comparisons, log2-transformed TPM expression data were used. This was because this comparison was to test the efficacy of including the endemic/epidemic KS labels in the model formulation during batch correction and Combat, which was determined as optimal, takes as input and outputs Gaussian-distributed data, necessitating log2-transformation (explained more in Section 5.2). This was except for ComBat-Seq, which takes RNA-Seq integer count data as input. Note that aside from the comparison on non batch-adjusted data, batch was not included as a factor in these comparisons, however age and months of KS were. For the epidemic-only comparison, HIV-1 level was additionally included in the models. Genes were determined significant if they had an FDR < 0.01. All numeric covariates (HIV-1 serum level, age, months ART treatment) were Z-scaled prior to inclusion in all models. For all comparisons, inclusion of patient in model formulations was not possible due to confounding with other factors, predominantly batch. All DESeq2-based differential expression analyses were performed using the apeglm log2FC-shrinkage es-

timator [265].

2.3.5 Network Construction

Gene co-expression network construction and analysis was performed using the WGCNA R package [266]. Firstly, the lesion sample expression data were filtered to remove the lower 5% and 25% of genes by mean and variance of expression, respectively. Next, each gene was correlated using Spearman's rank and subsequent pairwise gene-gene correlations. Next, the resultant correlation matrix was subject to the "signed hybrid" transformation, which makes use of a hybrid soft- and hard-thresholding on the gene-gene correlation edges. Specifically, any negative correlations were set to zero, while positive correlations were retained. Next, all non-zero correlation values were transformed by raising them to the power of the soft-thresholding parameter β, optimally chosen as 6 (See Fig 5.4). Correlations were retained and raised to the power of a constant β. Finally, the resultant signed-hybrid network was transformed by computation of the topological overlap matrix (TOM). TOM calculation further transforms the pairwise gene similarity for gene i and j, with the TOM between such pairs of genes being denoted as ω_{ij} and is calculated as in (Eq 1). Broadly, it is a measure of how connected the neighbours of pairs of nodes are and so its calculation further weights the connectivity of nodes that share a similar neighbourhood.

$$\omega_{ij} = \frac{\sum_u a_{iu}a_{uj} + a_{ij}}{min\{\sum_u a_{iu}, \sum_u a_{ju}\} + 1 - a_{ij}} \tag{1}$$

Where a_{ij} is the soft-thresholded edge correlation between gene i and j and a_{un} is the adjacency (edge-weight) between pairs of genes u and n, where u is a neighbour of both i and j and $n \in \{i, j\}$.

2.3.6 Initial module partitioning

Initial cluster partitioning proceeded by first calculating a dissimilarity matrix, D^{Ω} as $D^{\Omega} = 1 - \Omega$, where Ω is the weighted topological overlap matrix calculated in the previous section (Section 2.3.5). Genes were then clustered using average linkage agglomerative clustering, implemented in the 'hclust()' function of the 'stats' R package. Clusters were then defined using the dynamic hybrid method for cluster assignment, using the 'cutTreeDynamic()' function of the WGCNA R package [266, 267]. This method is a hybrid between hierarchical clustering and a modified *k*-medioid approach that proceeds in two steps; the first iterates from the lowest to highest point of the dendrogram, merging

branches based on fulfilment of 4 main criteria. All criteria were set to default except for minimum module size which was set to 75 genes. The second step ignores the dendrogram and attempts to group the unassigned genes into identified modules based on their relative proximity to the modules, as assessed via the *k*-medioids approach. This was performed with the options deepSplit = 2, minClusterSize = 75 and pamRespectsDendro = FALSE. Every other option was default. For further details on the methods outlined in this section, readers are directed to Langfelder *et al.,*, specifically the supplementary PDF provided with the paper [266].

2.3.7 Module eigengenes and module merging

In order to summarise the expression profile of modules, which could include 1000s of genes, module eigengenes (MEs) were calculated as PC1 of a PCA performed on the module-wise gene expression matrix. The resultant sample-wise eigenvalues in the PC1 eigenvector were taken to represent the relative activity of the module in the sample.

Correlation of genes with MEs provides one metric for how closely related the expression profile of genes are with the module as whole. Thus as an additional quality-control step for module assignments, modules were further merged based on the correlations of their constituent genes with MEs. Broadly, this was dependent on testing whether a threshold number of genes correlated more strongly with a ME of a external module, relative to the gene's correlation with the ME of its parents module.

2.3.8 Weighted Gene Co-expression Network Analysis Hub Gene Identification

Hub genes were defined via a pair of WGCNA-specific metrics. These include module membership (MM), which defines the correlation of the hub genes with module eigenegenes and can be further split into parental MM (pMM), or the MM of the gene with the module it is a constituent of, and global MM (gMM), which corresponds to the mean absolute MM of the gene to all network modules. In effect, pMM can be viewed as a module-wise global measure of centrality while gMM can be viewed as a network-wise global measure of centrality [266]. The second was connectivitiy, which is the sum of the edge weights connecting to that gene, which can further be split into intramodular connectivity (iC) and global connectivity (gC), which correspond to the connectivity of a gene to genes within its parent module and to all genes in the network, respectively [266]. Similarly to MM, iC and gC represent comparatively modular and global measures of network centrality, respectively.

In accordance with the above, two sets of candidate hub genes were defined; intramodular hub (module hub genes) genes and global network hub genes (global hub genes). Host hub genes were defined as the intersection of the top 10% of host genes by pMM and iC for modular hub genes and the top 10% of host genes by gMM and gC. Viral hub genes were defined as the intersection of the top 35% of viral genes by pMM and iC for modular hub genes and the top 35% of viral genes by gMM and gC. Candidate hub genes were further filtered based on whether they were significantly differentially expressed between lesion and control samples, explained in more detail in Section 2.3.4). This resulted in a final set of hub genes. Subsequent hub scores were then calculated as outlined in (Section 2.1.5).

2.3.9 Simplification of GO terms with Revigo

The results of many ontological analyses can result in a long list of terms from which it can be hard to garner biological insight. However as some ontological databases, particularly the Gene Ontology (GO) database, are arranged in a hierarchical tree manner, functional redundancy is present. Specifically terms are arranged as a directed acyclic graph, with "child" terms branching off from "parental" terms. The Revigo method, implemented in the R package rrevigo (v1.5.4) simplifies ontology terms based on whether the distance metric ("semantic similarity", chosen as the "Relevance" or "Rel" measure) between a child and parent term exceeds a provided threshold (chosen as Rel $>$ 0.7). Using this distance metric and threshold, as well as a GO term-wise vector of numeric values to rank terms by (chosen to be the -log10 adjusted p-value for ontological enrichment), Revigo clusters terms and chooses the most representative term for that cluster. For further detail on semantic similarity and simplification of GO terms with Revigo, readers are directed to Supek *et al.,* 2011 [268].

2.3.10 Differential Correlation Network Analysis

Differential network analysis was performed using the differential gene correlation analysis (DGCA) R package, testing for differential correlations between genes in lesion relative to control biopsy samples [269]. Spearman's ρ correlation was used to construct the initial lesion- and control-sample correlation matrices. The DGCA method then transforms the correlations via Fisher's Z-transform to convert them to normally-distributed Z-scores (z) as in Eq 2 where. Variance (σ^2) is then determined from these distributions as in Eq 3, where n is the sample size of the calculated correlation, assuming bivariate normality. Next, the difference between Z-scores calculated as in Eq 4, where $\sigma^2_{z_x}$ is the

variance of the Z-score in condition x, where $x \in \{lesion, control\}$.

$$z = atanh(\rho) = \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \tag{2}$$

$$\sigma^2 = \frac{1.06}{n-3} \tag{3}$$

$$\delta_z = \frac{(z_1 - z_2)}{\sqrt{|\sigma^2_{z_{lesion}} - \sigma^2_{z_{control}}|}} \tag{4}$$

Raw two-sided p-values were then calculated from the gene-wise differences in z-scores (d_z) from the standard normal distribution. 100 Sample permutations were performed in order to determine empirical p-values over a range of λ tuning values and the proportion of null p-values determined by fitting a cubic spline. Q-values were then calculated from these null p-value proportions. Differential edges were filtered at FDR $<$ 0.05.

Differentially correlated edges were then classified based on the direction of the change, with 9 total classes as shown in Table 2.8. Edges are classified based on whether the gene-wise correlations remain positive (+/+), remain negative (-/-), are unchanged (show a non-significant difference) (0/0), change from positive to none (+/0), change from positive to negative (+/-), change from negative to none (-/0), change from none to positive (0/+), change from negative to positive (-/+) or none to negative (0/-).

	$\rho > 0, p < \alpha$	$p < \alpha$	$\rho > 0, p < \alpha$
$\rho > 0, p < \alpha$	+/+	+/0	+/-
$p < \alpha$	0/+	0/0	0/-
$\rho < 0, p < \alpha$	-/+	-/0	-/-

Table 2.8: DGCA differentially correlated edge classes.

3 Identification of circRNAs, miRNAs and linear RNAs differentially expressed during KSHV reactivation

3.1 Chapter Introduction

Over 30 years of research has emphasised the importance of post-transcriptonal regulation of mRNA as a key step in gene expression. One of the most studied and characterised mechanisms is RNA interference (RNAi), predominantly mediated by miRNAs. Canonically RNAi is facilitated by miRNAs binding in complex with Argonaute proteins (miRISC complexes), to miRNA binding sites (MBSs) in 3' untranslated regions (UTRs) of messenger RNAs (mRNAs) [185]. This induces repression via recruitment of deadenylation and decapping enzyzmes, resulting in exonuclease-degradation [185].

MiRNA-based repression has been shown to be primarily mediated by changes in the abundances of the miRNAs themselves, whereby cognate targets are repressed or derepressed upon up- and down-regulation of miRNAs between differing biological conditions [185, 186]. This includes host and viral transcripts and accordingly miRNAs have emerged as key mediators of viral infection [32]. Additionally many viruses encode their own miRNAs that have been shown to modulate a range of cellular and viral processes [15, 18, 237].

One such virus is Kaposi's Sarcoma associated herpesviruse (KSHV). KSHV encodes 25 viral miRNAs that modulate the abundance of various host transcripts. Like all known herpesviruses KSHV can undergo distinct latent and lytic gene expression profiles; with minimal viral gene expression, but including the viral miRNAs, and the increased expression of the majority of viral transcripts, respectively [32, 67]. Accordingly while many genes are dysregulated during latent infection, lytic replication involves large-scale changes in host gene expression and modulation of the cellular environment to facilitate virion production [270]. This is mediated by the expression of specific viral factors that directly target such processes or indirectly by modulating host factors that regulate these processes. One example is the viral endonuclease SOX (ORF37), which degrades the majority of host transcripts during lytic reactivation [79].

One common method to investigate miRNA-mediated repression of transcripts is to measure the abundances of miRNAs and target transcripts in a range of conditions and compare changes to infer co-regulation. While this was historically done laboriously through measuring the abundance of genes and transcripts via classical molecular biology approaches (qRT-PCR etc), the advent of 'omics technologies including microarrays but particularly RNA-Seq, has enabled unbiased global profiling of their abundances. Such approaches allow efficient "top-down" global modelling of their interactions from a systems biology perspective. This perspective has led to the suggestion that miRNAs act in

consorted networks that can modulate the abundances of transcripts to facilitate complex gene expression programs [188, 189].

Accordingly one insight in recent years was that various transcripts can bind to miRNAs and not be appreciably degraded by them, thereby sequestering them away from the free pool of miRNAs that are available to repress host transcripts [191]. Therefore such transcripts, termed "miRNA sponges", can positively modulate the abundances of transcripts targeted by such sponged miRNAs and all 3 comprise the elements of sponge-miRNA-mRNA axes. Such miRNA sponge-miRNA-mRNA networks are often termed "competing endogenous RNA" (ceRNA) networks [189]. Circular RNAs (circRNAs), which are formed via reversed-orientation "back-splicing" of exons to form covalently closed loops of RNA, have emerged in as prime miRNA sponge candidates and as key regulators of ceRNA networks [202, 271]. Importantly recent work over the past 5 years has highlighted the importance of circRNAs to KSHV's life cycle and pathogenesis [23, 24, 26, 244].

Thus the goals of this chapter are to profile changes in gene expression between latent and lytically infected cells. Small RNA-Seq, total RNA-Seq and circRNA microarrays were utilised to investigate global changes in miRNAs, linear RNAs and circRNAs, respectively. Additionally the influence of SOX on host transcripts was investigated and assessed. The culmination of this was the identification of sets of circRNAs, miRNAs and linear RNAs that showed changes between latency and lytic replication which could be used for subsequent ceRNA network construction.

3.2 Reactivation of KSHV induces a predominant down-regulation in host miRNAs

As miRNAs are at the interface between circRNAs and mRNAs, identifying those showing a change in expression between latent and lytic replication is the first step in characterising a ceRNA network operating this process. To begin, RNA fractions of latently infected (0hr time-point, pre Dox-induction) and lytically reactivated (16 and 24 hr time-points post Dox-induction) TREx-RTA-BCBL1 cell's were sequenced, aligned and quantified. Next differential expression analysis was performed between 0 and 16 and 0 and 24 hrs. From this it firstly became apparent that there was little change in miRNAs prior to and 16 hrs post lytic reactivation, with just 3 miRNAs signficantly differentially expressed. Of these, two were down-regulated (miR-424-3p and miR-148a-5p) and one was up-regulated (miR-3609). In contrast when comparing the 0 and 24 hr time points, a clearer distinction emerged; 25 miRNAs were significantly differentially expressed, with 20 being down-regulated and 5 up-regulated (Fig 3.1 & Table 3.1). Indeed, miRNA clustered based on whether they were up- or down-regulated at 24hrs, relative to 0hrs (Fig 3.2). The top 3 up-regulated miRNAs were miR-196a-3p, miR-3609 and miR-375-3p, while the top down-regulated miRNAs were miR-92a-1-5p, miR-29b-1-5p and miR-27a-5p. Several well-studied miRNAs were differentially expressed, including miR-92a-1-5p and miR-29b-1-5p, but also the miR-30 family members miR-30b and miR-30c and miR-26b-3p [272].

In summary this section has facilitated the identification of a set of miRNAs dysregulated upon reactivation of KSHV from latency to it's lytic replication cycle.

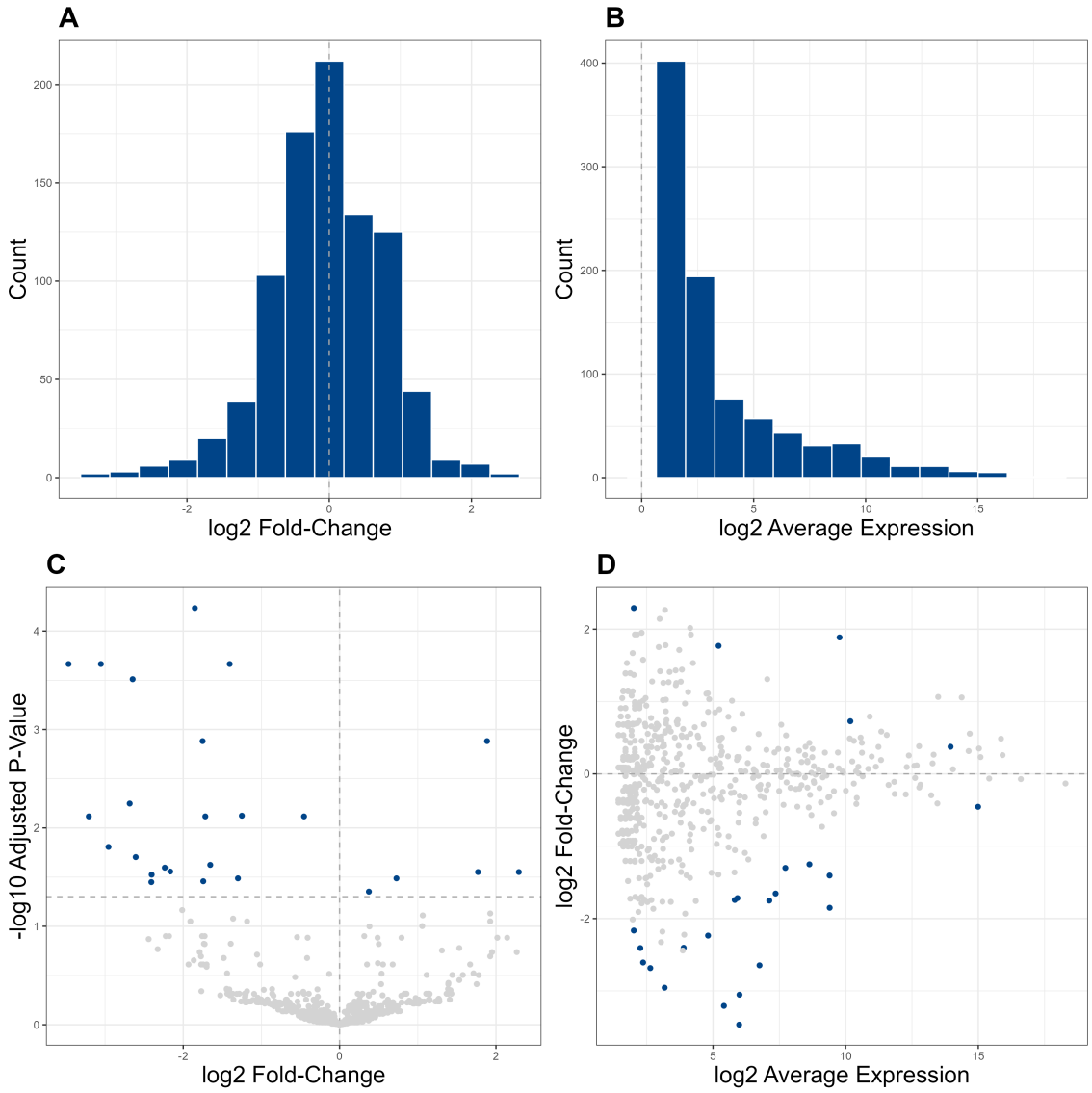


Figure 3.1: Differential expression of miRNAs between latently infected BCBL1 cells and cells 24hrs into lytic replication. Differential expression was performed between triplicate miR-Seq profiles in latency and 24 hours post Dox-mediated induction of KSHV reactivation, using TREx-RTA-BCBL1 cells. Differential expression was performed using DESeq2 (as further detailed in Methods Section 2.2.3). Distribution of log2 fold-changes between latency and 24 hrs post reactivation (A). Distribution of miRNA log2 average expression (log2AE) (B). Volcano plot of differential expression results (C). MA plot of log2FCs vs log2AEs for miRNAs. MiRNAs are coloured according to whether the were significantly (FDR < 0.05) differentially expressed (blue) or not (grey).

miRNA	AE	Log2FC	FDR
miR-196a-3p	2.01	2.29	2.808812e-02
miR-3609	9.77	1.89	1.312897e-03
miR-375-3p	5.21	1.77	2.808812e-02
miR-181d-5p	10.18	0.73	3.257994e-02
miR-301a-3p	13.95	0.37	4.442729e-02
miR-21-3p	14.99	-0.45	7.642672e-03
miR-424-3p	8.63	-1.25	7.519954e-03
miR-3689a-3p	7.73	-1.30	3.257994e-02
miR-148a-5p	9.40	-1.41	2.161286e-04
miR-671-3p	7.36	-1.65	2.376432e-02
miR-26b-3p	5.92	-1.72	7.642672e-03
miR-30b-3p	5.82	-1.74	3.487538e-02
miR-25-5p	7.12	-1.75	1.312897e-03
miR-181a-3p	9.40	-1.85	5.829772e-05
miR-6728-3p	2.01	-2.17	2.775447e-02
miR-30c-1-3p	4.81	-2.24	2.535209e-02
miR-23a-5p	3.89	-2.40	2.989346e-02
miR-488-5p	2.26	-2.41	3.556106e-02
miR-6734-5p	2.37	-2.61	1.980717e-02
miR-128-1-5p	6.75	-2.65	3.082831e-04
miR-590-5p	2.64	-2.68	5.655284e-03
miR-5087	3.18	-2.96	1.561490e-02
miR-27a-5p	6.00	-3.05	2.161286e-04
miR-29b-1-5p	5.41	-3.21	7.642672e-03
miR-92a-1-5p	5.99	-3.47	2.161286e-04

Table 3.1: Statistics associated with differentially expressed miRNAs. MiRNA = microRNA, AE = Average Expression (CPM), log2FC = Log2 Fold-Change, FDR = False Discovery Rate (Benjamini-Hochberg-adjusted)

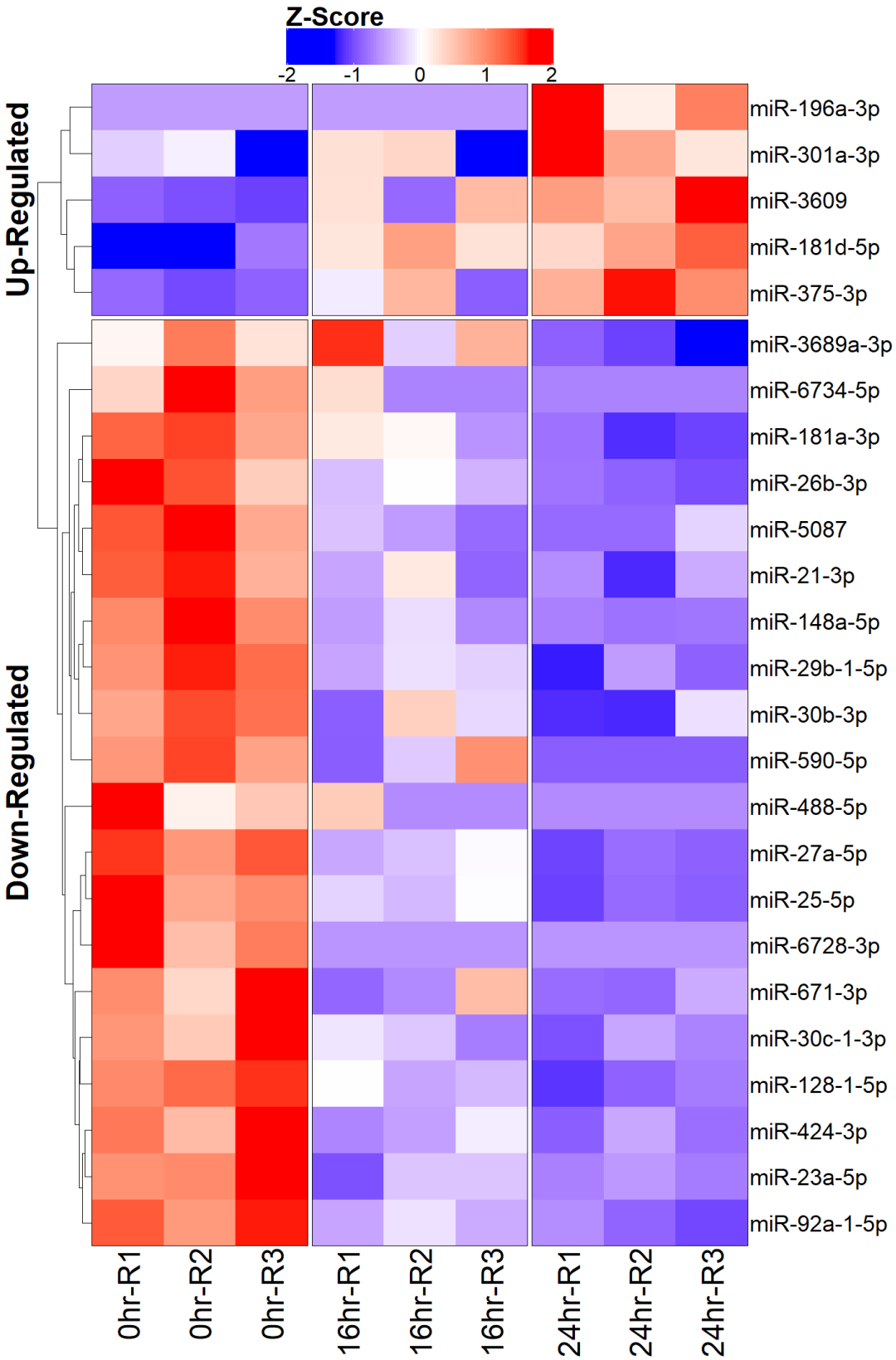


Figure 3.2: Heatmap of the relative miRNA expression during KSHV reactivation. Differentially expressed miRNAs were determined at 24hrs post Dox-induced KSHV reactivation relative to 0hrs (pre-induction/latency) in TREx-RTA-BCBL1 cells (FDR$<$0.05, BH-adjusted). MiRNA expression was CPM-normalised, Z-scaled and clustered using hierarchical agglomerative clustering. Heatmap is coloured by Z-score. Heatmap columns represent samples and triplicate replicates are grouped for 0, 16 and 24hrs post-Dox treatment.

3.3 Total RNA-Seq reveals global mean trend for the down-regulation of host transcripts

Next, analysis focused on the changes to the non-small, linear RNA fraction of the transcriptome. Duplicate total RNA samples were taken prior to and 8 and 24 hrs post lytic reactivation of TREx-BCBL-1 cells. Reads were aligned against the human and then KSHV genomes and transcripts quantified at the gene-level. Subsequent analyses were then performed to investigate changes to gene expression during reactivation.

A principal components analysis (PCA) was performed in order to characterise the major sources of variance in the data-set. This showed that principal component 1 (PC1) was strongly associated with time post-reactivation; indicating that the majority of variance (67.38%) was due to the dysregulation of gene expression with increasing progression through the lytic cycle (Fig 3.3). PC2 explained 12.4% of the variance and was most strongly associated with variance between the two replicates. However there was a notable divergence between replicates at each time point as lytic replication progressed, with the 24hr time-points being the most distinct from eachother (Fig 3.3). This is reflective of a known property of herpesvirus lytic reactivations in that their lytic transcriptional dynamics can diverge between even identically treated biological replicates, resulting in relatively distinct end-states [51]. This is most likely due to the exponential nature of lytic reactivation as being a cascade-like process that is susceptible to subtle stochastic variance in cellular and external environmental states. The discrepancy between the relative proportion of viral genes may explain some of this as the replicate with the greatest proportion of viral reads showed the greatest separation from the latent (0hr) sample (Fig 3.3 & 3.4c).

Two issues needed to be accounted for to provide a proper assessment of differential expression, in order to identify a robust set of differentially expressed genes. First, KSHV expresses the host shuf-off factor SOX early in during lytic replication, which degrades the majority ($\sim$75%) of host transcripts. This represents a global decrease in host transcripts which RNA-Seq cannot innately detect due to it being limited to detecting only proportional abundances of total transcript pools [77]. Second, there is a large up-regulation of viral transcripts during this process, which inflate from occupying 0.15%/0.048% to 13.19%/8.71% (for replicates 1 and 2, respectively) of the total library read depths (Table 3.2 & Fig 3.4a & c). If not accounted for, this may result in an apparent but possibly artificial down-regulation of host transcripts due to competition for read depth (Fig 3.4b & S1a). To investigate changes further, one tool that has been developed is the differentially expressed genes elimination strategy (DEGES) [259]. After 5 iterations of the

DEGES procedure, this resulted in a trend for apparent down-regulation of host genes (342 up-, 4650 down-regulated), with median log2FC = -0.96 (Fig 3.4d). While not an estimate of the true global change in gene expression, this provides some evidence for a net down-regulation of genes that is not confounded by increased competition with viral genes, due to their large up-regulation.

The effect of SOX and increased viral gene expression have competitive effects that precludes determination of true log2FCs. One thing that was observed upon trialling differential expression methods was that unlike other methods DESeq2 resulted a near zero-centered median (0.06) (Fig 3.4e). Conversely, edgeR's TMMapproach resulted in a strong tendency for down-regulation, much like DEGES (Fig S1b). From a pragmatic perspective, this could be interpreted as facilitating the identification of host genes that deviate from the SOX-mediated trend for down-regulation. Interestingly this predominant up-regulation was particularly skewed for protein-coding transcripts (814 up-, 277 down-regulated) and pseudogenes (41 up-, 23 down-regulated), but not lncRNAs (367 up-, 303 down-regulated) (Fig 3.4f). Additionally, one important thing to note is that log2FCs calculated for all differential expression schemas were highly correlated, indicating that the DESeq2 approach didn't scramble the relative rankings of genes and so this aspect is representative of DEGES "truer" estimate of differential expression.

Cumulatively the main focus of this chapter has been to identify a set of differentially expressed protein-coding genes. Despite being confounded by fundamental limitations of short-read RNA-Seq technology, we believe that a set of host genes has been identified that are amenable to downstream network construction.

Set	**0hr (R1)**	**0hr (R2)**	**8hr (R1)**	**8hr (R2)**	**20hr (R1)**	**20hr (R2)**
Total	6155683	8413230	4879405	6028907	1958178	3435050
Host	6146154	8409159	4721313	5850236	1699828	3135660
Viral	4071	9529	158092	178671	258350	299390

Table 3.2: Total-, host- and viral-aligned RNA-Seq library read counts per sample.

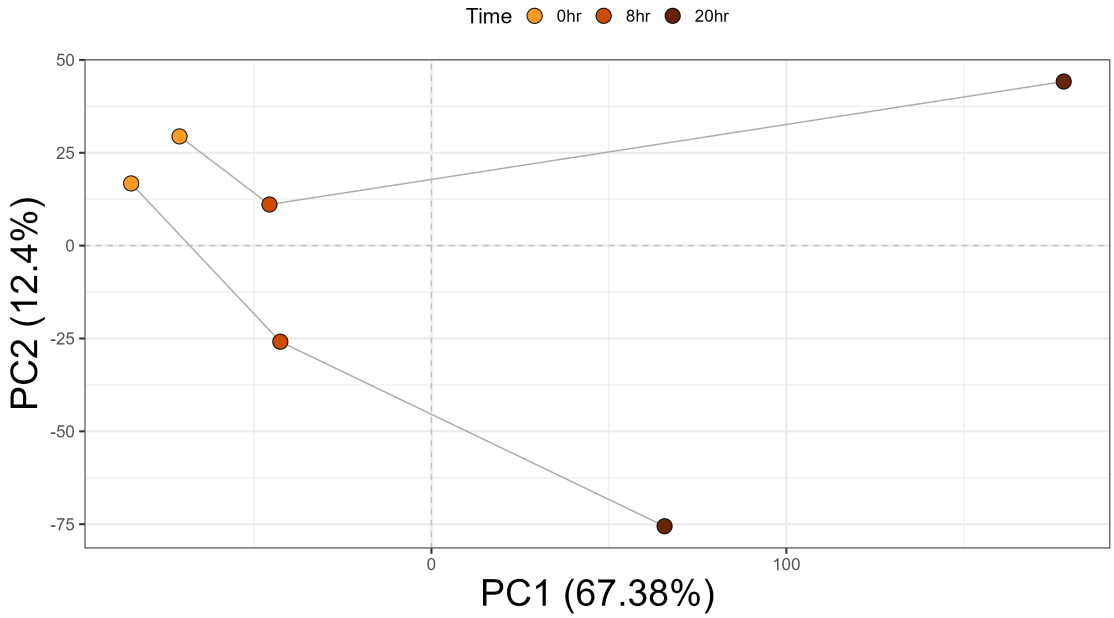
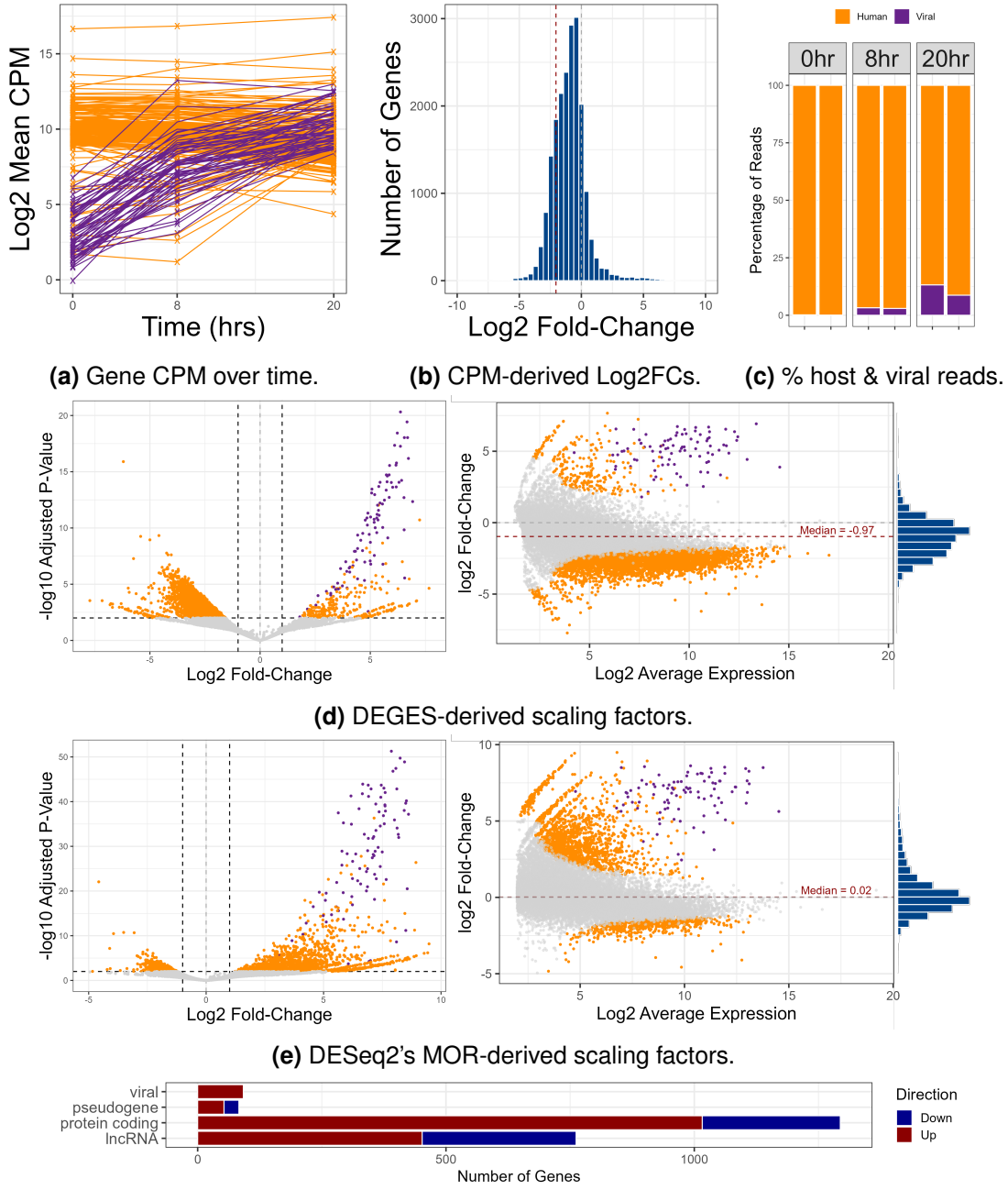


Figure 3.3: Principal components analysis (PCA) performed on duplicate total RNA-Seq profiles prior to and post reactivation. Total RNA-Seq was extracted during latency prior to (0hr) and 8 and 20hrs post-lytic reactivation. PCA analysis were performed on centered and scaled gene expression data. Data points are coloured according to the time in which samples were taken. Percentages associated with axis labels represent the percentage of variance explained by each PC. Lines represent links between data points taken from each replicate.



(a) Gene CPM over time. **(b)** CPM-derived Log2FCs. **(c)** % host & viral reads.

(d) DEGES-derived scaling factors.

(e) DESeq2's MOR-derived scaling factors.

(f) Up- and down-regulated genes by biotype.

Figure 3.4: Differential expression between latent and lytic cells. Differential expression using DESeq2 was performed between 0 (latency) and 24 hours post Dox-mediated induction of TREx-RTA-BCBL1 cells. The mean Log2 CPM expression for the top 250 host genes by variance and all viral genes were extracted and shown for 0, 16 and 24hrs (a). Log2 fold-changes (log2FCs) were manually calculated between 24 and 0hrs, where red line indicates median log2FC (b). Percentage of reads mapping to the host and viral genomes split by replicate and time point (c). Differential expression was performed using the DEGES method-derived (d) and DESeq2 median-of-ratios (MOR)-method (e) scaling factors. Volcano plots showing -log10 adjusted p-value and log2FC are shown in leftmost plots while MA plots showing log2 average expression (log2AE) against log2FC. Genes/genomes are coloured according to whether they were not significantly differentially expressed (FDR$\geq$0.05) (grey) or were significantly differentially expressed host (orange) or viral (purple) genes. Red lines in MA plots (d & e) indicates the median log2FC for all genes, regardless of significance. The number of significantly differentially expressed genes stratified by viral, pseudogene, protein-coding and lncRNAs (f).

3.4 Down-regulated host genes are enriched for KSHV endonuclease SOX's degenerate cleavage motif

Despite the suggestions of the previous section, there is still ambiguity as to whether SOX is responsible for the observed net down-regulation of host genes. To provide confidence in the the assertion that SOX was driving this process, SOX sites in up- and down-regulated genes were compared. One approach to provide some confidence in the impact of SOX was to test whether the down-regulated set of genes were significantly enriched for genes targeted by SOX, relative to the up-regulated set. Gaglia *et al.,* 2015 characterised SOX's degenerate RNA targeting motif and identified a set of 129 genes that were targeted and spared its degrading effect [77]. Cross-referencing of these genes with the DEGES differentially expressed genes identified in the previous section revealed few in common and no statistically significant enrichment in the down-regulated genes (Table 3.3a). This could likely be explained by differences in the transcriptomes between the cells used in their study, iSLK cells (which is derived from an adrenocarcinoma contaminant) and BCBL-1 cells, PEL cells derived from transformed B cells [77].

Gaglia *et al.,* 2015 also defined the targeting motif of SOX, which could be used to predict SOX sites in the set of differentially expressed genes identified in the previous section, thus improving the power and specificity of the previous analysis. To further investigate SOX binding sites, salmon was used to quantify transcript abundances across the samples, retaining transcripts originating from genes identified differentially expressed via the DEGES method [273]. This resulted in 15186 detected transcripts originating from 2925 (58.59%) of the DEGES-derived differentially expressed genes. Next, the position-weight matrix (PWM) for SOX's motif generated by Gaglia *et al.,* 2015 was scanned against these transcripts and the PWM score filtered for sequential values, which resulted in many putative SOX sites in transcripts derived from the differentially expressed genes (Table 3.3b, Fig 3.5). Importantly there was a strong statistically significant enrichment of putative sites in transcripts of down-regulated relative to up-regulated genes for sets of sites up to those filtered with a minimum PWN score greater than 6 (Fig 3.5 & Table 3.3b). Additionally, predicted SOX target genes showed a significantly more negative log2FC than than non-targets ($p<0.00005$, W=2867421, two-sided Wilcoxon rank sum test). Correlation between the PWM score of transcripts and their LFC indicate that the confidence of binding does not correlate with the observed down-regulation of host genes, however this was also observed by Gaglia *et al.,* 2015.

Overall such evidence of an enrichment of SOX cleavage sites in down-regulated genes provides support for the notion that SOX is responsible for the predominant down-regulation

of genes observed via the DEGES method. This helps confirm that SOX-driven down-regulation of genes is the predominant mediator of down-regulation and justifies the decision to use the DESeq2 MOR-method for further analysis.

Total Genes		**N. Genes with SOX**		**% Tx with SOX**		**P**
Up	Down	Up	Down	Up	Down	-
223	4645	3	233	1.29 %	1.85%	0.38

(a) SOX binding sites for differentially expressed genes.

PWM Score	**Total Genes**		**N. Genes with SOX**		**% Genes with SOX**		**P**
	Up	Down	Up	Down	Up	Down	
0	194	2731	110	2730	56.70%	99.96%	1.60e-106
1	-	-	110	2721	56.70%	99.63%	1.94e-95
2	-	-	109	2666	56.19%	97.62%	1.46e-67
3	-	-	108	2536	55.67%	92.86%	9.48e-41
4	-	-	101	2279	52.06%	83.45%	3.70e-22
5	-	-	90	1906	46.39%	69.79%	5.66e-11
6	-	-	75	1434	38.66%	52.51%	1.24e-4
7	-	-	51	916	26.29%	33.54%	0.022
8	-	-	24	407	12.37%	14.90%	0.20
9	-	-	7	89	3.61%	3.26%	0.70
10	-	-	1	16	0.52%	0.59%	0.69

(b) Predicted SOX binding sites in DEGES genes

Table 3.3: Enumeration and percentage of known (a) and predicted (b) SOX sites in transcripts originating from differentially expressed genes, iteratively filtering by minimum PWM score $<$ 1-10. Where SOX = KSHV host shut-off endouclease encoded by ORF37; PWM = position weight matrix; Tx = transcript; P = raw p-values.

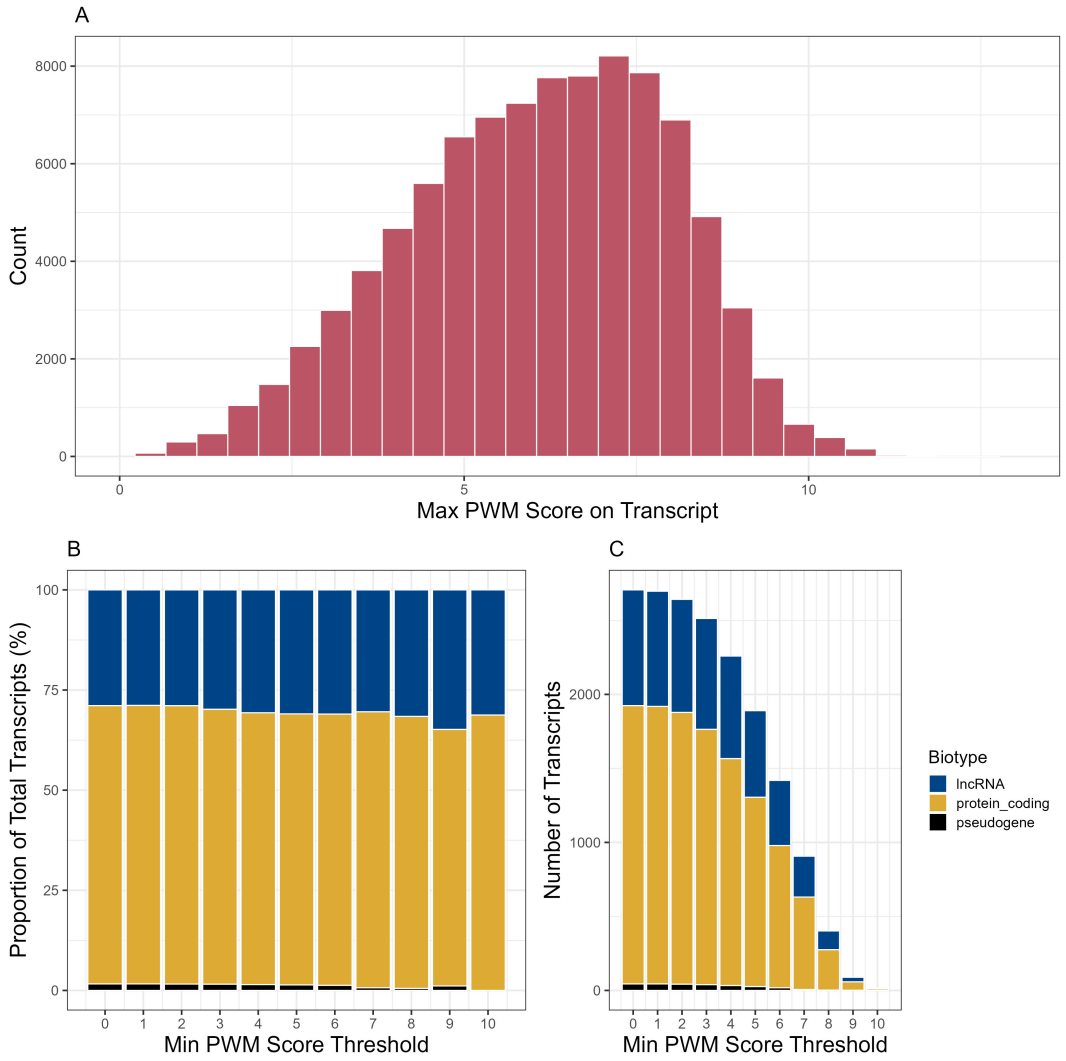


Figure 3.5: Predicted SOX motifs in transcripts produced by differentially expressed genes. Predicted binding sites for KSHV's host shut-off lytic gene SOX (ORF37) in Salmon-detected transcripts produced by differentially expressed genes using the position-weight matrix (PWM), derived from Glaunsinger *et al.,* 2015 (described more in Methods Section 2.2.4). Transcripts were quantified via salmon and cross-referenced with the significantly differentially expressed genes derived from the DEGES approach for differential expression (Results Section 3.3). Distribution of PWM scores over transcripts (A). Proportion (B) and total number (C) of transcripts with SOX sites in differentially expressed transcripts, coloured by biotype, at differing minimum PWM score cut-offs.

3.5 Significantly differentially expressed circRNAs show a bias for up-regulation

In order to investigate the dynamics of circRNAs during lytic reactivation of KSHV, total RNA was extracted from TREx-Rta-BCBL-1 cells prior to- and post- Dox-treatment. These were then subject to RNase R-treatment followed by hybridisation on a circRNA microarray chip comprising probes complementary to 13410 circRNA backsplice sequences.

Just like for the total RNA-Seq and miRNA-Seq, differentially expression was performed in order to identify a set of candidate circRNAs. However, immediately prior to differential expression, circRNAs with a mean intensity across replicates of less than 5 were removed, resulting in 12688 circRNAs. However initial analyses made it apparent that one 0hr replicate was likely an outlier. This was determined via observation of a PCA biplot and initial exploratory differential expression analyses with limma (Fig 3.6a). For the latter, its inclusion prior to analysis resulted in no circRNAs identified as differentially expressed, while its exclusion resulted in the identification of 1344 circRNAs as differentially expressed, including circHIPK3 which is known to be up-regulated during lytic replication and acts as a "marker" circRNA (FDR$<$0.05) (3.6B & C). This latter experimental design was taken forward. To provide a more robust set of circRNA, circRNAs were filtered to retain those with a Log2AE $>$ 8. Additionally, circRNas were filtered based on having a Log2FC $>$ 0.65 as this drastically reduced the number of circRNAs, while retaining circHIPK3 (Log2FC $>$ 0.66), resulting in 412 candidate circRNAs.

The vast majority of differentially expressed circRNAs were exonic (89.86%), with just 1.05% antisense, 3.99% intronic and 5.07% sense-overlapping (Fig 4A). Immediately clear was a predominance for up-regulation of circRNA, with 276 up-regulated and 136 down-regulated (Fig 3.6D). This was particularly driven by the exonic circRNAs, as they in particular were up-regulated (248 up-regulated, 108 down-regulated) (Fig 3.6d). Importantly, like for the miRNAs and unlike for RNA-Seq, such a bias was present just for the differentially expressed circRNAs and wasn't a general bias in the underlying distribution, with all, unfiltered circRNAs having a mean log2FC of just 0.01.

There remains some ambiguity over whether detected circRNAs (particularly those identified via high-throughput methods) represent true biotypes produced from genes or are by-products of linear RNA splicing. However in this experiement, the correlation between circRNA and linear RNA from the same gene was relatively modest (Pearson's r = 0.14), suggesting detected circRNAs were not a result of reads derived from the linear RNA from the parental gene. Additionally gene over-representation analyses (ORA) for

the parent genes of the circRNAs showed no significant enrichment for any gene ontology (GO) terms or KEGG pathways (FDR$\geq$0.1). This indicated that changes in circRNA abundances are not associated with concerted changes in the abundance of the linear output of their parental genes. However it is worth stating that for any one circRNA, this cannot be confirmed without further biochemical analysis.

In summary this section has identified a set of candidate exonic circRNAs for ceRNA network model construction and subsequent analysis.

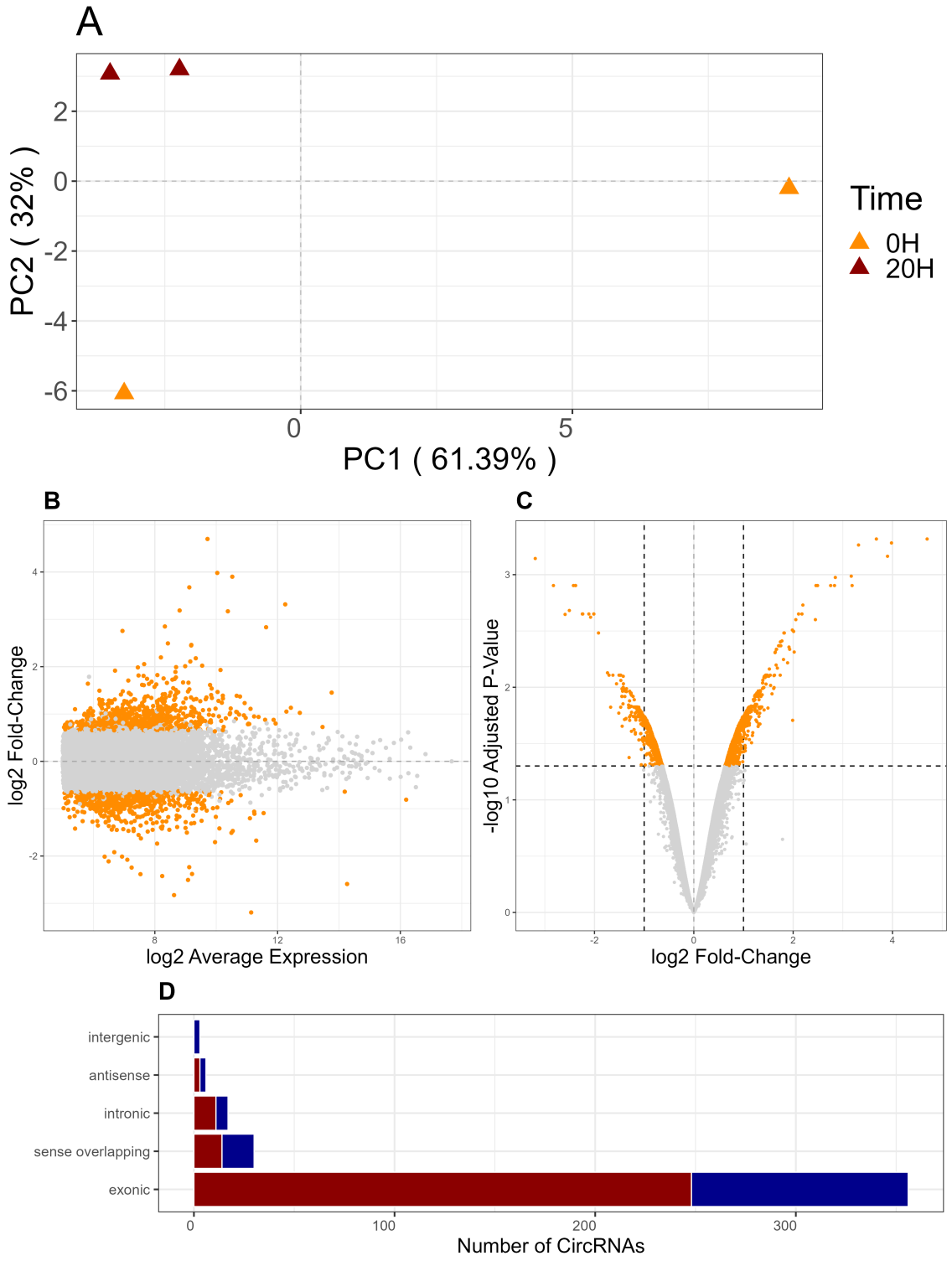


Figure 3.6: Comparative analysis of circRNA expression profiles. CircRNA expression profiles were generated by RNase R-treating total RNA taken prior to (latency) and 20hrs post Dox-mediated induction of reactivation in TREx-RTA-BCBL1 cells. These then hybridised against an arraystar microarray with probes complementary to circRNA back-splice sequences. Principal components analysis (PCA) was performed on the circRNA expression data (A). Differential expression analyses were performed between 0 and 20hrs post induction of reactivation using Limma. MA (B) and volcano (C) plots for differential expression analyses, where circRNAs are coloured according to whether they were significantly differentially expressed (FDR < 0.05) (orange) or not (grey) (B & C). Frequency of up- (red) and down- (blue) regulated circRNAs by circRNA biotype.

3.6 Discussion

Overall, this chapter aimed to identify sets of differentially expressed circRNAs, miRNAs and linear RNAs to serve as a foundation for ceRNA network construction. This revealed predominant up-regulation of specifically exonic circRNAs and a concordant down-regulation of miRNAs. Despite inherent limitations, a set of differentially expressed protein coding genes were identified.

Several miRNAs identified as differentially expressed have been relatively well-characterised with functions relevant to the KSHV and the rest of the present study. These include miR-30c-1-5p and miR-29b-1-5p which were previously identified in Harper *et al.,* 2022 as being sponged by circHIPK3. Additionally, both the miR-30 family members miR-30b and miR-30c as well as miR-29b-1-5p have well-defined tumour suppressive roles and so their down-regulation may contribute to the oncogenic effect of lytic replication [71]. Interestingly, miR-30c-1-5p has previously been observed to be down-regulated in KS lesion tissue suggesting a functional oncogenic role in KS [274]. In contrast, members of the pro-oncogenic family of miRNAs, miR-17-92 and miR-106b-25, miR-92a-1-5p and miR-25-5p, respectively, were down-regulated [272].

Interpretation of differentially expressed linear RNAs was confounded by the impact of SOX degradation of host genes as well as increased competition from largely up-regulated viral gene expression. Nonetheless using DESeq2's MOR method a set differentially expressed genes were identified, however this involved removal of viral genes prior to calculation of library size scaling factors. This could potentially be problematic as there was quite extensive variability in RNA-Seq library sizes, with the smallest (20H_R1) and largest (0H_R2) comprising 1958299 and 8413233 reads, respectively (Table 3.2). This could be expected to result in increased variance of lowly expressed transcripts in libraries with low sample depth, due to the integer nature of RNA-Seq count data. To limit this, genes were filtered that had on average less than 3 counts per each sample. Moreover, both 20hr replicates showed $\sim$4-fold less read depth than 20hrs which may be due to the greater GC percentage of viral transcript, resulting in less efficient PCR amplification of cDNAs originating from viral RNA (Table 3.2) [275, 276].

Differentially expressed circRNAs were also investigated. Robust filtration was performed to select for the most high-confidence circRNAs, with a majority of these were exonic circRNAs. Notably however the reliability of these candidates is still relatively low, given the un-usability of one of the 0hr replicates (Fig 3.6a). While the empirical Bayes (eBayes) sharing of variance from the 20hr lytic samples facilitated differential expression, mea-

surements of variance may not be accurate. However in support of the results having some grounding in known biology, circHIPK3 (hsa_circ_0000284), the circRNA found by Harper *et al.,* 2022 to be up-regulated in latent vs lytic samples, was detected as up-regulated from this data-set [71]. Moreover it is worth stating that, assuming the circRNA samples follow similarly diverging trajectories as observed for the total RNA-Seq, the variance between 20hr replicates is likely greater than between 0hr replicates indicating that circRNA p-values could be expected to be conservative, at least on-average (Fig 3.3). Nonetheless this will require some consideration when with proceeding analyses.

A drawback of relying on circRNA microarray data used, as opposed to next-generation or long-read sequencing approaches is that the length and exon structure of most circRNAs included is largely assumed. The Agilent arraystar microarray probes comprise DNA probes that are complementary to backsplice sequences, derived from pairs of 5' and downstream 3' exon sequences, joined in a "reverse" orientation to the native genomic sequence [277]. While these were also filtered based on detection a series of global circRNA studies from RNA-Seq data, the vast majority of these circRNAs even to date remain unvalidated [203, 204, 271]. Importantly, such approaches assume that the intervening exons are retained in the mature circRNAs and are not spliced out, however this mechanism remains a poorly defined [202]. Accordingly, some reported circRNAs lengths may be less than the length of circRNAs expressed *in vivo*. This obviously impedes circRNA quantification as it results in ambiguity of the detection of circRNAs, but also may impede downstream network construction as miRNA target binding is almost entirely dependent on the sequence representation and thus length of circRNAs. The only way to mitigate this would be to use a different technology, such as long-read sequencing technologies, or short-read RNA-Seq approaches that facilitate inference of circular exon structures from split-aligned reads [278]. Aside from ease, circRNA microarrays do have some advantages relative to sequencing, as CircRNA-Seq suffers from poor reproducibility, the generation of artificial backsplice-like cDNA sequences via reverse transcription and PCR errors and poor backsplice sequence coverage [277].

While interesting by themselves, parallel investigation of miRNA, circRNAs and linear RNAs does not account for their influence to the more integrative nature of post-transcriptional regulation, primarily their roles in ceRNA networks. Nonetheless this chapter facilitated the processing of each dataset in preparation for this downstream, wholistic analysis. The subsequent results chapter will proceed to integrative ceRNA network construction and analysis.

4 Construction and analysis of a competing endogenous RNA network dysregulated between latent and lytic KSHV

4.1 Introduction

Analysing genes in parallel and investigating transcript co-expression can provide key insight into underlying biology. However such approaches can miss the wider high-level interactivity of genes, which is a key facet for how they impart phenotypic effects. Such complex connections can be modelled as networks, whose construction and analysis can allow for the elucidation of the mechanisms that govern biological processes. In this context, such analysis can help to elucidate an additional and poorly-studied layer of the interaction between infected cells and KSHV.

In the language of networks, transcripts and genes are represented as "nodes", while interactions between them represented as links between nodes, or "edges". Networks can be constructed by a range of means, however ceRNA network models are generally produced by defining interactions between miRNAs and targets (mRNAs and circRNAs in this case). These interactions are either inferred by predictive algorithms or provided by prior knowledge databases [187, 260, 279]. Many databases of pre-determined miRNA-target interactions exist, while predictive algorithms tend to leverage known features of miRNA-target binding and are generally based on identifying complementarity between the major binding motifs of miRNAs, termed "seed" sequences and sequences of putatitve targets [187].

The methods in network analysis are extensive and can and have been applied to ubiquitous domains, but in particular have garnered much interest when applied to systems biology. Common methods for interrogating networks generally involve the calculation of measures of connectivity of genes within the network, broadly termed "centrality measures". Such high-ranking genes can be viewed as highly interconnected and are termed "hub" nodes. High connectivity means that they are considered to be highly important as they have the potential to influence the activity of a many other genes.

Given the stability and large potential regulons of circRNAs, modulation of circRNA-mediated ceRNA networks could be an efficient mechanism by which KSHV can impact a broad range of cellular processes could promote or inhibit its replication. Indeed, the past several years has seen the identification of multiple circRNA ceRNA candidates that play direct or indirect roles in KSHV biology. These include host circRNAs such as circARFGEF1 which sponges miR-125a-3p to promote vIRF1-mediated cell invasion and angiogenesis [20, 23, 24, 71, 241, 244]. Moreover we have previously shown that circHIPK3 sponges miR-30c and miR-29b, with the prior interaction promoting the expression of DLL4 during lytic replication and this activity of circHIPK3 is important to KSHV replica-

tion [71]. However many of these studies have been relatively focused on the small-scale, gene-by-gene studies that ignore the wider network context that circRNAs exist in.

To this end, this chapter will focus on the initial construction and robustness testing of a ceRNA network from the data derived from the previous chapter. Enumeration and investigation of genes targeted by shared miRNAs followed, with subsequent centrality analyses to identify core influential hub genes. Finally miRNA- and circRNA-wise regulons were analysed via gene set over-representation analyses (ORA). Cumulatively this enabled the identification of certain circRNAs and miRNAs as key putatitve network effectors, possibly involved in regulating RNA metabolic processes and transcription.

4.2 CircRNA-miRNA-mRNA/viral transcript competing endogenous- RNA network model.

The ceRNA hypothesis states that ceRNA networks comprise ncRNAs (circRNAs, lncRNAs) that bind and negatively regulate miRNAs, which bind and negatively regulate mRNAs. This amounts to directional interactions between circRNA$\rightarrow$miRNA and miRNA$\rightarrow$mRNA that are inhibitory and, in graph theory (the study of networks), consistute a directed graph (Fig 4.1a). Specifically, its structure is comprised of sets of overlapping trees with circRNAs as source nodes, miRNAs as intermediary nodes and mRNA/viral genes as terminal nodes and as there are only 3 primary node types (assuming protein coding/viral genes behave the same) the network can be termed tripartite. This means that edges can only exist that include miRNAs and none exist directly between circRNAs and protein coding/viral genes and none between nodes of the same type, meaing that the network is acyclic. Therefore in full the network can be defined as a directed acyclic tripartite graph.

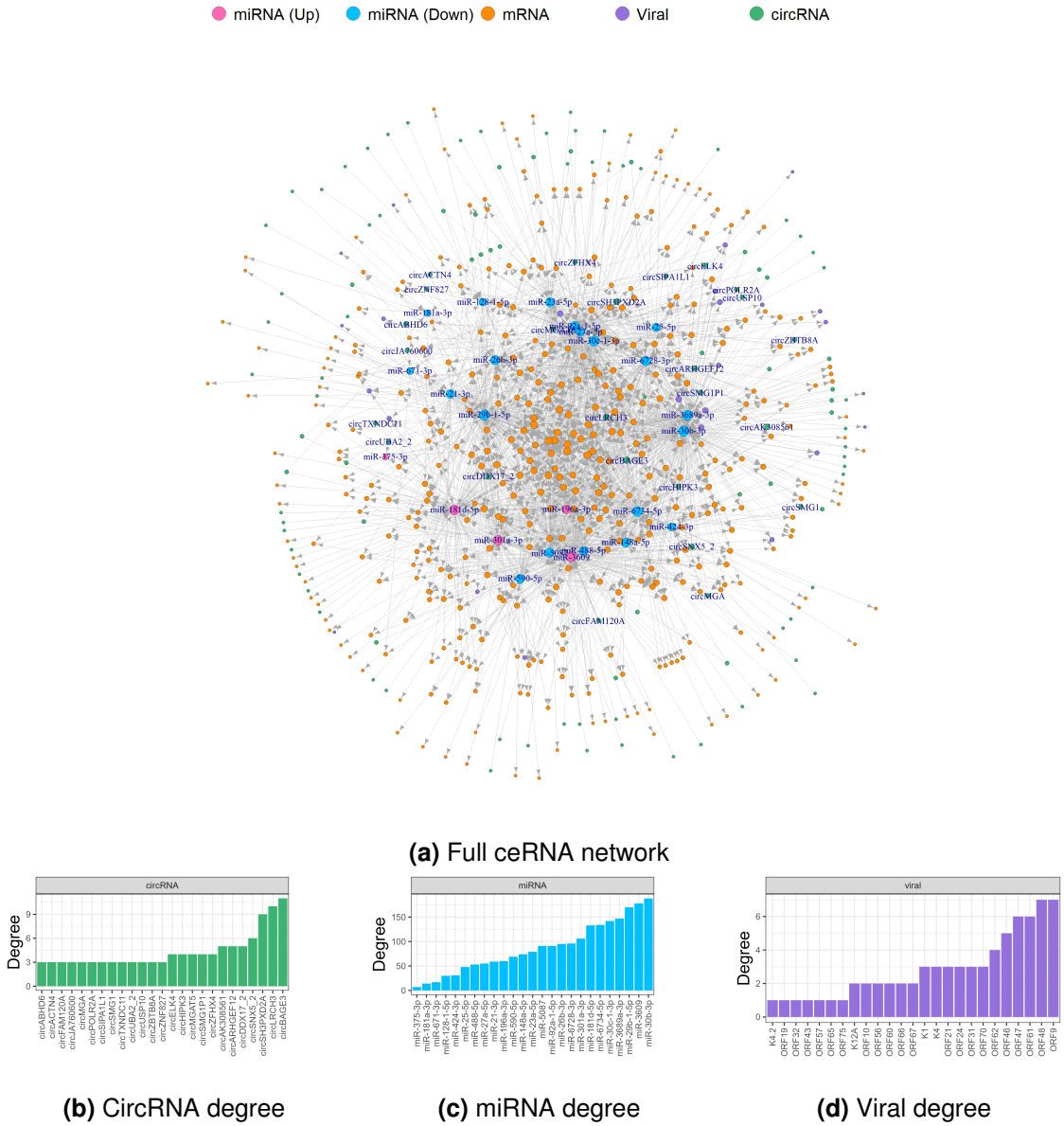
Given this structure, network construction proceeded via the following steps. To further filter circRNAs to focus on the most likely miRNA sponge candidates, the set of 412 differentially expressed circRNAs were further filtered to remove those $<$ 200 and $>$ 10000 nt, alongside excluding circRNAs that weren't exonic, leaving 310 circRNAs. This length filtration step was performed in order to prioritise exonic circRNAs that are more likely to be real and that fall within reasonable length ranges to be capable of sponging miRNA. Next miRNA target predictions were performed between the 3' UTRs of differentially expressed host and viral genes as well as the reduced set of exonic circRNAs. The RNAhybrid and miRanda target prediction algorithms were used and edges defined as the intersection between predictions of these two algorithms [260, 280]. This resulted in a fully interconnected network, with no extraneous disconnected sub-networks (components) (Fig 4.1a). The network comprised 119 circRNAs, all 25 differentially expressed miRNAs, 403 protein coding genes and 25 viral genes (Table 4.1). In accordance with the findings of the previous chapter, the majority of circRNAs (85), protein coding (324) and all viral genes (25) were up-regulated while most miRNAs (20) were down-regulated. None of the network circRNAs were those whose parent gene's protein-coding output was differentially expressed, indicating that they are specifically up-regulated.

To confirm the specificity of the differentially expressed miRNAs for circRNAs permutation tests were employed, randomly selecting 310 circRNAs out of all those detected (also $<$ 200 and $>$ 10000 nt in size) and performing miRNA target prediction. Surprisingly no permutations showed a greater number of sites than for the differentially expressed circRNAs and thus the respective p-values for circRNAs were at most $p<0.0004$. This was

the same for when comparing miRanda and RNA hybrid target predictions individually as well as when comparing the intersection of these approaches. This suggests that the network circRNAs were specifically enriched in predicted binding binding sites for network miRNAs, relative to all detected circRNAs. In addition network circRNAs showed a significantly greater log2FC (median = 2.18) than for non-network circRNAs (median = 0.83) ($p<2.6e-16$, W = 65730, one-sided Wilcoxon Rank Sum test). These observations show that the network circRNAs were significantly enriched for network miRNAs and showed a bias for greater up-regulation, relative to non-network circRNAs.

Degree is a measure of the connectivity and thus influence of a node in the network and thus its calculation provides a relative measure of the importance of each gene in the network. Given that miRNAs are the primary effectors in the network responsible for connecting source circRNA nodes and terminal protein coding/viral genes they tended to have the greatest number connecting edges (degree). In particular, miR-30b-3p, miR-3609, miR-29b-1-5p, miR-3869a-3p and miR-30c-1-3p were highly connected with degrees over 125 (Fig 4.1c). Additionally circBAGE3 ranked the highest in terms of degree, connecting to 10 miRNAs (miR-30b-3p, mirR-30c-1-5p, miR-3689a-3p, miR-3609, miR-6374-5p miR-92a-1-5p, miR-148-5p, miR-301a-3p, miR-27a-5p and miR-26b-3p) (Fig 4.1b). CircBAGE3, circLRCH3, circSH3PXD2A, circSNX5_2 and circDDX17_2 ranked highest out of all circRNAs in terms of degree (Fig 4.1b). Interestingly circHIPK3 ranked 11th by degree and was predicted to bind miR-29b-5p, miR-3968a and miR-5087 but not miR-30c-1.

Thus in summary network model construction was successfully accomplished with miRNA target predictions. Network circRNAs were confirmed to be enriched for network miRNA binding sites. The resultant network structure and node degree distributions suggested the importance of several miRNAs (miR-30b-3p, miR-29b-1-5p, miR-3609, miR-30c-1-3p and miR-3869a-3p) and circRNAs (circBAGE3, circLRCH3, circSH3PXD2A, circSNX5_2 and circDDX17_2).



(a) Full ceRNA network

(b) CircRNA degree

(c) miRNA degree

(d) Viral degree

Figure 4.1: Full circRNA-miRNA-mRNA ceRNA network. Target prediction was performed between all differentially expressed miRNAs above thresholds ($p < 0.05$, $|Log2FC| > 1$, $Log2AE > 0.5$, as identified in Results Section 3.2 and all circRNAs detected above thresholds ($p<0.05$, $|LogFC| > 0.65$, $LogAE > 8$, as identified in Results Section 3.5 and 3' UTRs of all detected host and viral transcripts (as identified in Methods Section 2.3.4) from differentially expressed genes, as identified in Results Section 3.2, using RNAhyrbid and miRanda and edges retained as the intersection between these tools. (a) The resultant network was visualised using the Kamada-Kawai algorithm, colouring nodes based on whether they represent circRNAs, miRNAs, mRNAs or viral transcripts. MiRNAs and circRNAs with degrees > 3 are labelled with their respective names and nodes are scaled by log10(degree) + 1. Degrees for circRNAs (b), miRNAs (c) and viral transcripts (d) in the network for nodes with degree of > 2.

Biotype	Total	Up-reg	Down-reg	Mean Degree
All	572	439	133	7.24
circRNA	119	85	34	1.97
miRNA	25	5	20	73.12
Protein coding	403	324	79	4.53
Viral	25	25	0	2.88

Table 4.1: Frequencies of nodes in competing endogenous RNA network.

4.3 Network miRNAs show variable frequencies of target genes, which show a bias for up-regulation.

Investigating the distribution of target binding by each miRNA may provide insight into the role and importance such miRNAs have in the network. Moreover investigation of the log2FCs of miRNA target genes may provide evidence for the effect of ceRNA network interactions on such miRNAs. Therefore, the number and type of targets of each miRNA were enumerated and features of these targets were investigated.

The number of transcripts predicted by bind each miRNA were determined for each miRNA for all transcripts, protein-coding genes, viral genes and circRNAs (Fig 4.2). As identified previously, the down-regulated miR-30b-3p, miR-29b-1-5p, miR-3689a-3p, miR-6734-5p and miR-30c-1-3p alongside the up-regulated miR-3609 ranked highly by total degree. These 6 miRNAs also ranked highly in terms of total numbers of protein coding genes that they were predicted to target (Fig 4.2B & C). Additionally, MiR-30b-3p and miR-3689a-3p were particularly highly ranked for circRNA binding. Moreover miR-25-5p, miR-30b-3p and and miR-30c-1-3p were particularly enriched for predicted viral targets and interestingly, viral genes did tend to be depleted for up-regulated miRNA predicted binding sites, relative to other biotypes and the network as a whole (Fig 4.2C). This could indicate some form of selective pressure conferred to the virus to not present potential binding sites for these miRNAs in the 3' UTRs of its transcripts.

One key observation from the previous chapter was that the circRNAs and miRNAs and miRNAs and protein coding/viral genes showed reciprocal log2FCs, which provided some circumstantial evidence for ceRNA network activity having an appreciable effect in the network. To further interrogate whether the network supports this, the log2FCs for protein coding and viral genes were compared for genes in the network (Fig 4.3a) and not in the network (Fig 4.3b). This indicated that a slightly greater proportion of these genes were up-regulated in the network relative to non-network genes (Fig 4.3d & c). Importantly, the log2FCs of protein-coding and viral genes in the network were significantly greater than for genes not in the network (one-sided Wilcoxon Rank Sum test, $p<0.01$). This appears to be largely due to a greater proportion of mRNAs wtih log2FCs between $\sim$1.5 and $\sim$5 in the network relative to not (Fig 4.3a & b). This implies that network linear RNAs predicted to be regulated by network miRNAs showed a bias for up-regulation relative to those not predicted to be regulated by them. This provides further support for an association between the fold-change of protein-coding genes and their targeting by miRNA.

Cumulatively this section has profiled the predicted miRNA binding sites in target cir-

cRNA, protein coding and viral genes in the network and ranked miRNAs that have extensive predicted interactions. Moreover miRNA target genes show a statistically significant bias for up-regulation that is reciprocal to the bias for down-regulation of network miRNAs.

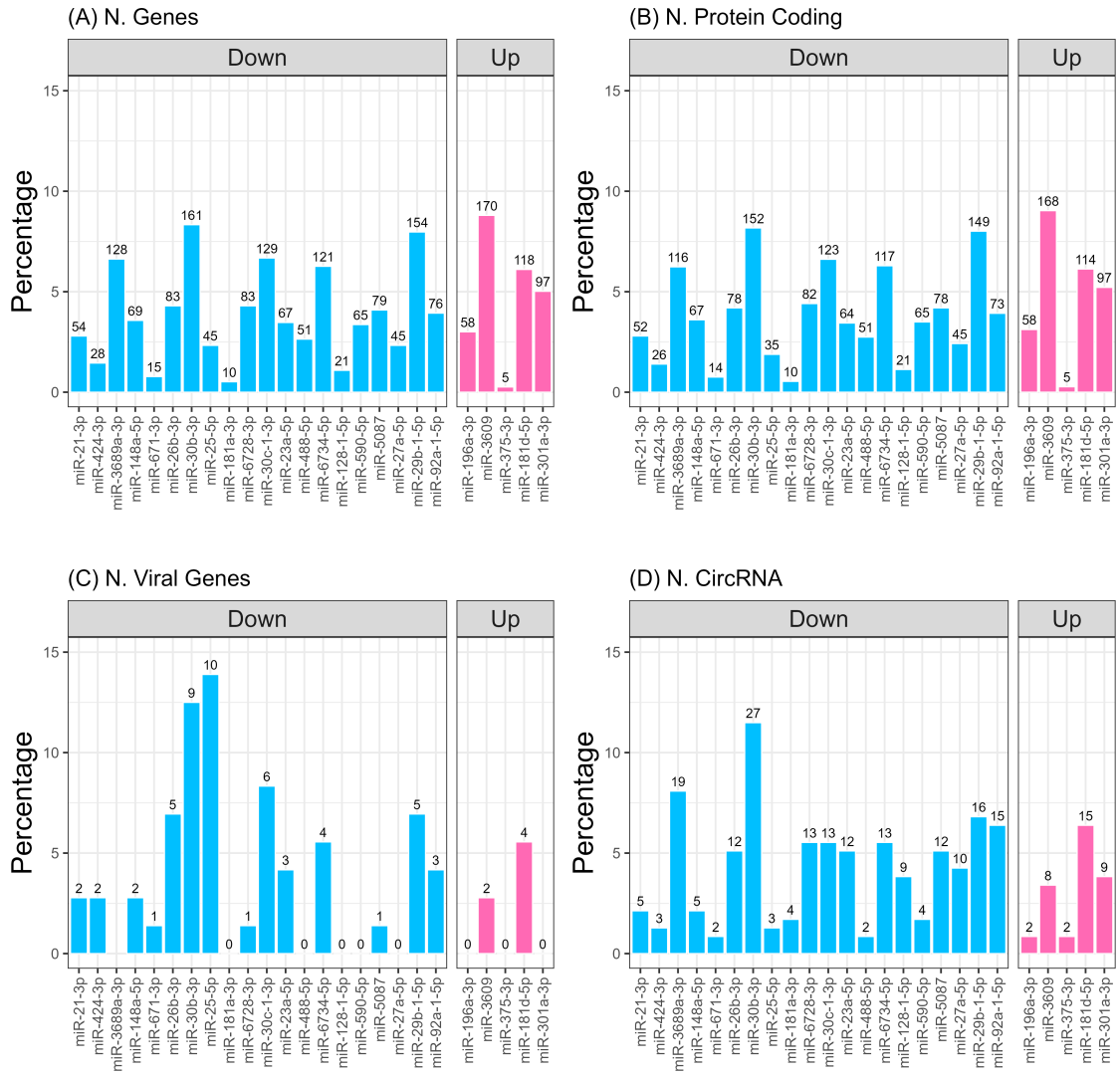
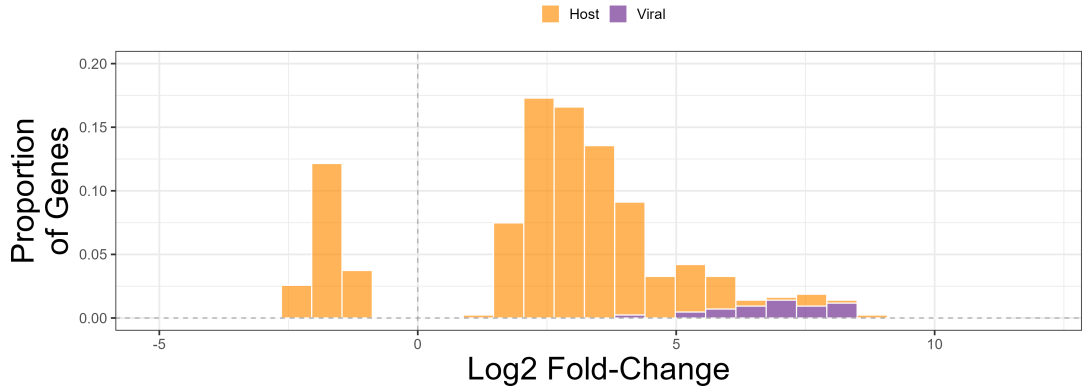
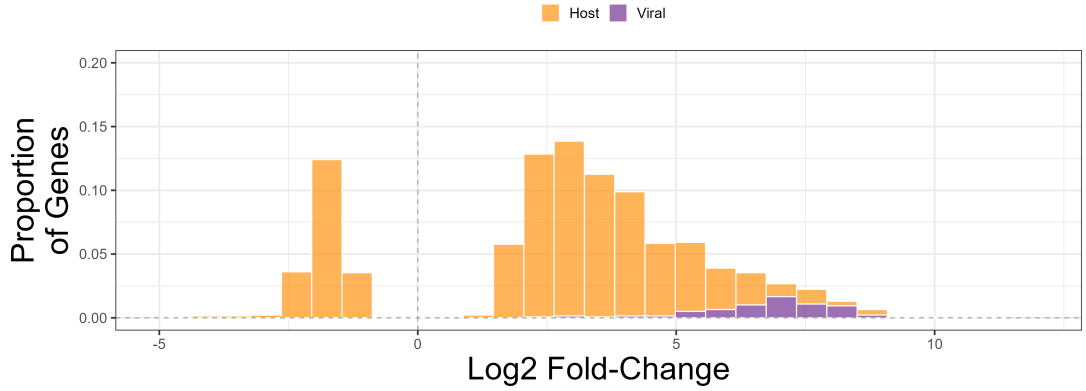


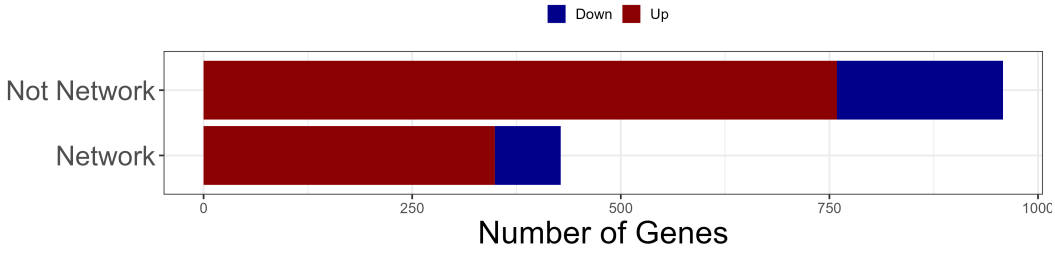
Figure 4.2: Enumeration of predicted miRNA targets. The number of predicted miRNA target genes for all network genes (A) and each individual network biotype (B, C & D). These enumerations are shown for only protein-coding genes (B), viral genes (C) and circRNAs (D). Percentages of total and each biotype-wise targets were determined (x axes) as well as the absolute number of genes (above bars). miRNAs are coloured according to whether they were determined to be up-regulated (pink) or down-regulated (blue).



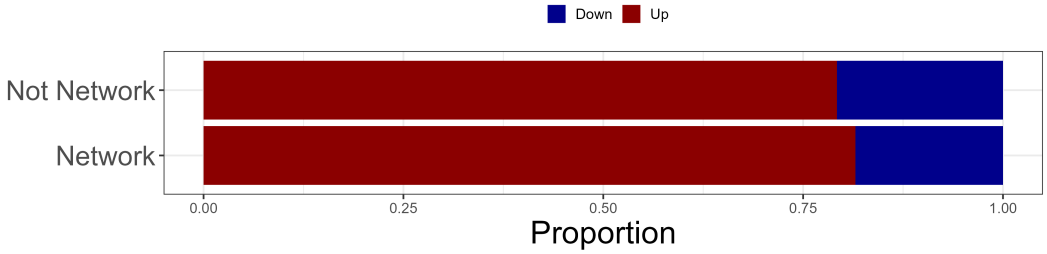
(a) Genes in ceRNA network.



(b) Genes not in ceRNA network.



(c) Number of up- and down-regulated genes.



(d) Proportions of up- and down-regulated genes.

Figure 4.3: Distributions of up- and down-regulated genes. Distribution of log2FCs for host protein-coding and viral genes, coloured accordingly. Distributions are for genes in the ceRNA network in Fig 4.1 (a) and not in the ceRNA network (b). The number (c) and proportion (c) of significantly ($p<0.05$) up- and down-regulated protein-coding and viral genes in in the network and not in the network. Bars in (c & d) are coloured according to the proportion up- and down-regulated genes.

4.4 Centrality-based analyses rank top influential circRNAs and miRNAs in the network.

While node degree is of merit in summarising the influence of nodes in networks it can be relatively naive as it ignores global, higher-order structures of the networks. To mitigate this, further centrality metrics were calculated for the nodes in the network using the Influential R package (see Methods Section 2.3.8) [253]. Specifically "hub" scores were determined (Fig 4.4a & b), alongside the 2-step geodesic paths (Fig 4.4c & d). These metrics were counted for out-degree (Fig 4.4a & c), or the number of directed edges emanating from nodes and in-degree (Fig 4.4b & d), or the number of directed edges that a node is a target of. In a biological context, out-degree can be viewed as nodes that are upstream regulators of networks that exert an effect on down-stream effectors, with in-degree being viewed as the expected influence that downstream nodes "feel" from upstream regulators.

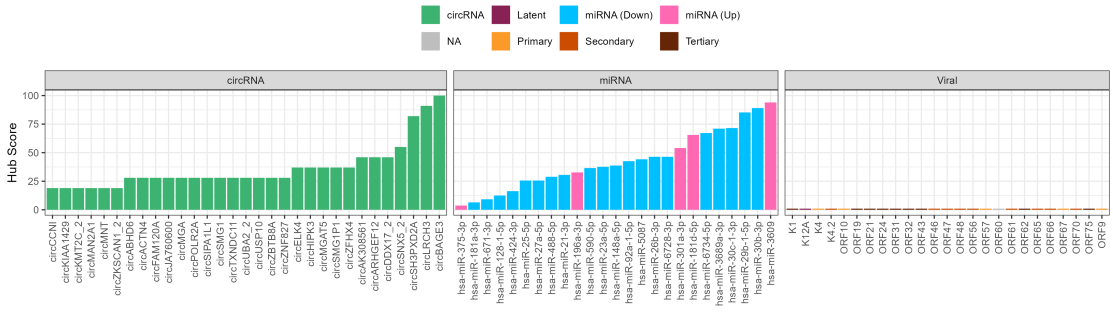
Hubness specifically measures semi-local features of the networks and provides a measure of connectivity of genes, while 2-step geodesic paths enumerate the number of paths that pass through any one node, and as such are a measure of the flow of information through such nodes. MiR-3609, miR-30b-3p, miR-29b-1-5p, mir-30c-1-3p and miR-3689a-3p ranked most highly in terms of out-hub score, underlying the extent by which these miRNAs can potentially influence the expression of protein coding/viral genes. Some of these miRNAs ranked additionally highly in terms of in-hub score, namely miR-30b-3p, miR-3689a-3p and miR-29b-1-5p but also miR-92a-1-5p and the up-regulated miR-181d-5p ranked highly (Fig 4.4b). This indicates that these miRNAs in particular are key downstream effectors of upstream circRNAs. Ranking miRNAs by 2-step geodesic in- and out-paths gave similar results (Fig 4.4c & d). The ranking of out-hub circRNAs was identical to degree-based ranking however this must be the case as edges can only emanate from them (Fig 4.1b & 4.4a). Ranking circRNAs by 2-step geodesic out-paths is consistent with by hubness, except circHIPK3 ranked more highly (Fig 4.4a & c). In-particular circBAGE3 which is within 2 steps of almost all (311) protein coding/viral genes in the network, followed by circLRCH3, circSH3PXD2A, circSNX5_2 and circDDX17_2 (Fig 4.4c). This suggests that these circRNAs and miRNAs strongly influence the activity of many downstream miRNAs and protein-coding genes, respectively.

Strikingly the majority of protein coding genes that ranked highly via in-hub and geodesic in-path were up-regulated, indicating an association between increased targeting by multiple miRNAs and up-regulation of genes (Fig 4.4b & d). Indeed the degree of up-regulated protein coding genes was significantly greater than down-regulated protein

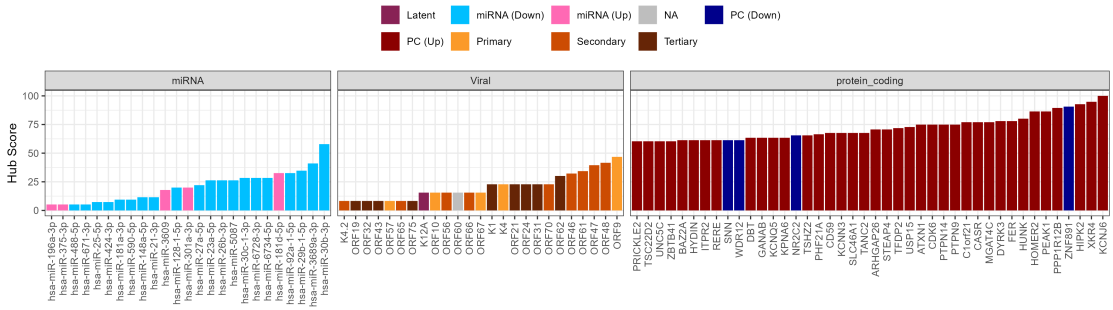
coding genes (one-sided Wilcoxon Rank Sum test, $p<0.0026$). Notable protein-coding genes are further detailed in Table [TABLE] and include XKR4, a phospholipid scramblase involved in phosphatidylserine exposure, HIPK2, an activator of p53 activity, a calcium signalling component CASR, various potassium channel components KCNJ6, KCNN3 and KCNQ5, HUNK, a gene involved in autophagy and CDK6, the primary kinase known to complex with vCyclin [101, 281, 282].

To further investigate the importance of nodes in the network, K-core decomposition was performed (Fig 4.5). Briefly, this procedure aims to find the maximally connected sub-network whereby a maximum number of nodes have a degree of at least K. At K=4 (Fig 4.5B) 12 (8 up-, 4 down) circRNAs were retained, namely those that ranked highly in the centrality metrics (Fig 4.2a & b). These included the highest ranking circRNA circBAGE3, alongside circHIPK3, circLRCH3 and circSMG1P1 (Table 4.2b). Additionally the previously mentioned protein-coding gnees, XKR4, CASR, CDK6, HIPK2, HUNK, KCNJ6, KCNN3 and KCNQ5, among others, were retained in K=4. Viral genes were nearly all depleted upon K-core composition with $K\geq2$ indicating their low "coreness" to the network (Table 4.2a).

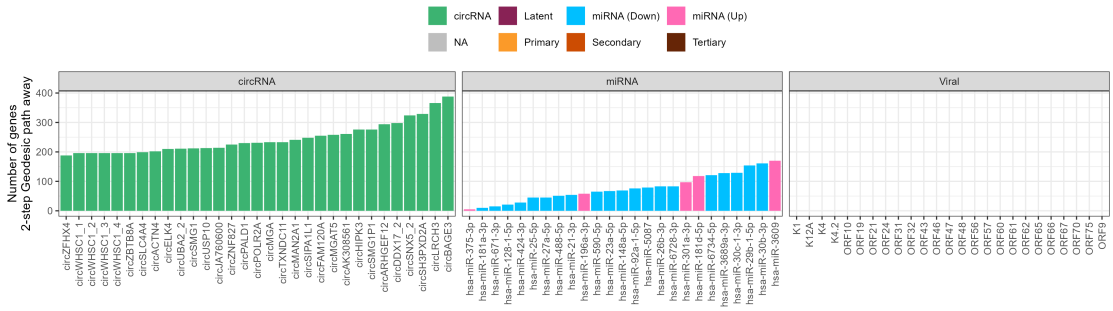
All together, this section has indicated that a handful of circRNAs, particularly circBAGE3, circLRCH3, circHIPK3, circLRCH3 and circSMG1P1, are important potential upstream regulators of the ceRNA network. Additionally some miRNAs and in-particular miR-30b-3p and miR-3689a-3p rank highly in terms of hub score and retention during K-core decomposition, indicating that they are are key transmitters of circRNA-based regulation of protein coding/viral genes in the network. Finally, a set of high-ranking protein coding target genes were identified including those involved in potassium and calcium signalling, gene expression and the cell cycle.



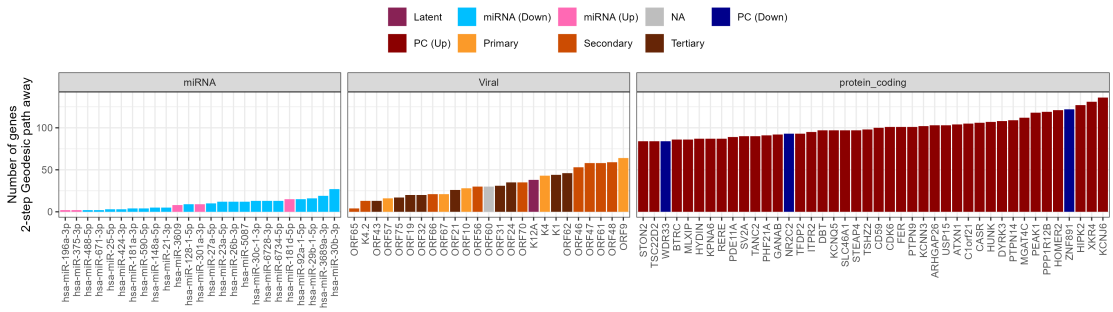
(a) Hub out-degree



(b) Hub in-degree



(c) Geodesic out-degree



(d) Geodesic in-degree

Figure 4.4: Centrality measure of circRNAs, miRNAs and viral transcripts in predicted ceRNA network. Hubness scores were calculated using the IVI package, as further detailed in Methods Section 2.3.8 (a & b). 2-step geodesic paths were calculated as the number of nodes that can be reached with 2 edges of the node in question (c & d). Hubness and 2-step geodesic paths were calculated with respect to direction of edges, with out-degree (a & c) being relative to the source node of a source-target directed edge and in-degree (b & d) being relative to the target node. Transcripts/nodes are ranked by increasing values of their respective centrality measure scores. Host genes are coloured according to whether they were up- (red) or down-regulated (blue), while viral transcripts are coloured according to their maximal expression timing, primary (0-8hrs), secondary (8-24hrs) or tertiary (48-72hrs).

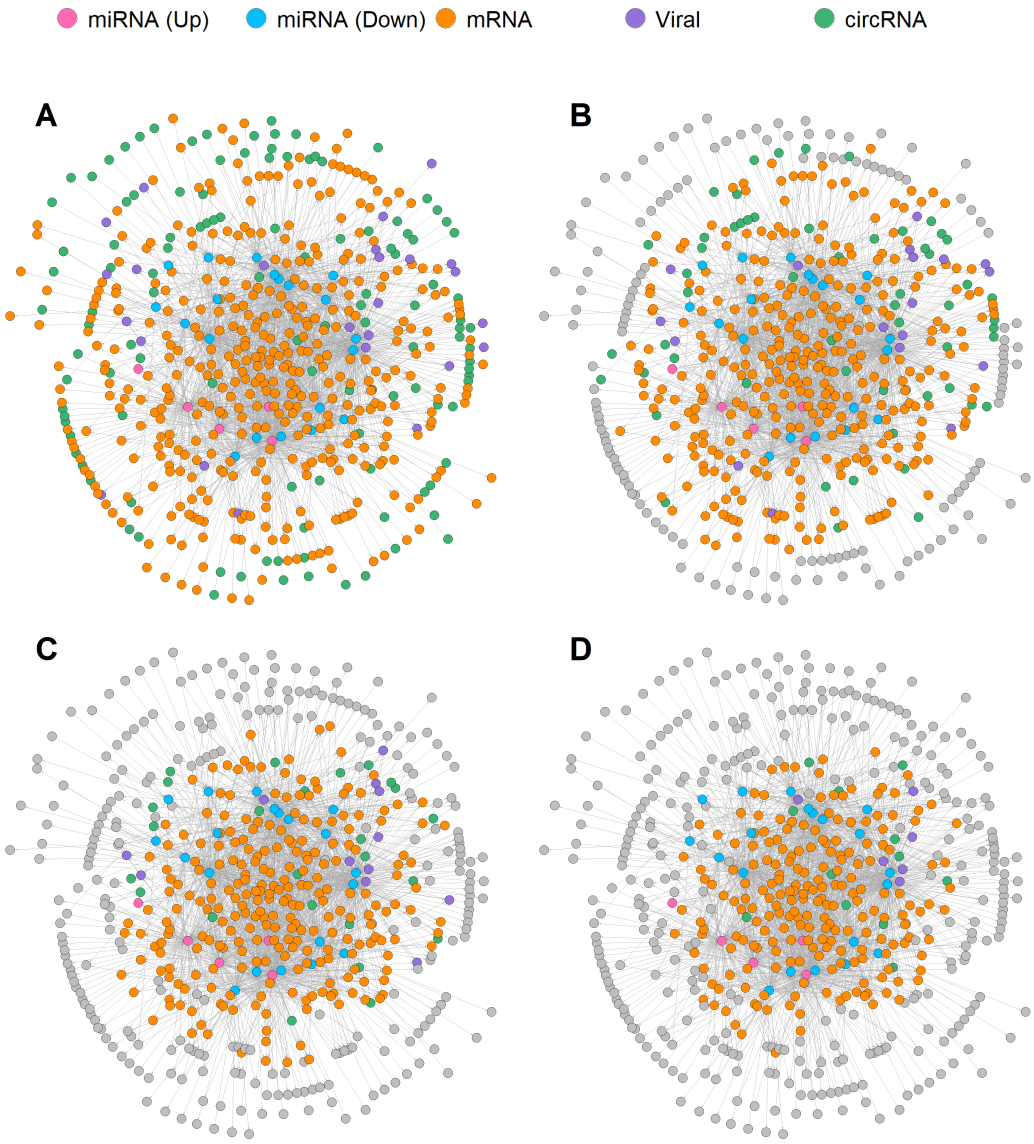


Figure 4.5: K-core decomposition of predicted ceRNA network. In order to isolate the most highly interconnected "core" of the network, K-core decomposition was performed on the full ceRNA network in Fig 4.1, removing nodes with a total (ignoring direction) degree of less than 1 (A), 2 (B), 3 (C) and 4 (D). As with Fig 4.1, nodes are coloured according to whether they are circRNAs, miRNAs (up- and down-regulated), mRNAs and viral transcripts.

K	Total			circRNA			miRNA			Proten coding			Viral
	Total	Up	Down	Total	Up	Down	Total	Up	Down	Total	Up	Down	Total
1	505	376	129	119	85	34	25	20	5	336	261	75	25
2	370	277	93	56	39	17	25	20	5	271	215	56	18
3	268	202	66	25	16	9	25	20	5	206	169	37	12
4	211	157	54	12	8	4	24	20	4	169	139	30	6

(a) Enumeration of nodes in K-core decomposed networks (see Fig 4.5).

ID	Name	Log2AE	Log2FC	FDR	Length
hsa_circ_0000175	circELK4	9.76	0.98	0.017	4577
hsa_circ_0000284	circHIPK3	9.74	0.68	0.042	1099
hsa_circ_0002953	circBAGE3	8.81	1.06	0.016	9817
hsa_circ_0005620	circSH3PXD2A	9.00	0.94	0.029	4865
hsa_circ_0008351	circLRCH3	10.64	0.90	0.041	4727
hsa_circ_0009173	circSNX5_2	8.15	0.83	0.040	4393
hsa_circ_0024615	circARHGEF12	8.25	-0.82	0.034	4205
hsa_circ_0038539	circSMG1P1	8.05	1.62	0.0078	2395
hsa_circ_0056530	circMGAT5	9.10	-0.86	0.028	6470
hsa_circ_0063313	circDDX17_2	9.20	-0.81	0.036	3955
hsa_circ_0084789	circZFHX4	8.46	-0.66	0.047	3139
hsa_circ_0087119	circAK308561	13.46	0.73	0.036	3437

(b) Statistics associated with K=4 K core-decomposed network.

Table 4.2: Statistics associated with K=4 K-core network (a) and K=4 K-core network circRNAs (b). Where K is the minimum number of edges network nodes must have to be retained in K core decomposition, circRNA = circular RNA miRNA = microRNA, log2AE = log2 average expression, log2FC = log2 fold-change and FDR = false discovery rate. Note that K=1 is equivalent to the full network.

4.5 Network targets are enriched for host transcription, nucleotide biosynthetic processes and protein ubiquitinylation.

Gene ontology enrichment analyses are frequently used to interrogate which biological processes subsets of genes are involved in. To this end, gene set over-representation analyses (ORAs) were performed on the full and K-core decomposed networks, as well as on miRNA predicted targets and the targets of miRNAs predicted to bind the same circRNAs.

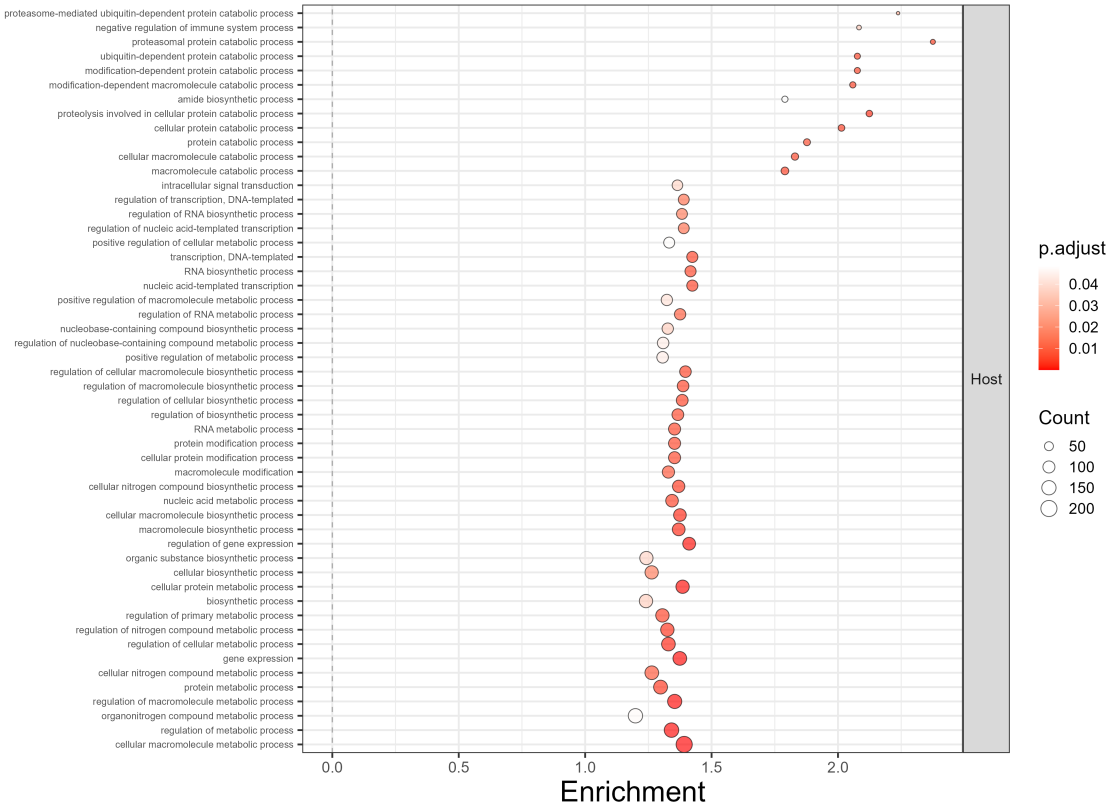
Immediately apparent from the full network enrichment analysis was that it was significantly enriched (FDR$<$0.05) for many terms relating to biosynthetic processes, particularly RNA and macromolecule biosynthesis (Fig 4.6a). Additionally processes involved in and regulating and facilitating transcription and gene expression as well as catabolic protein processes (including "protein ubiquitinylation") (Fig 4.6a). This indicates that the network as a whole encompasses processes that increase RNA biosynthesis to facilitate RNAPII-mediated transcription. Moreover K-core decomposed networks retained these enrichments, indicating that the core of the network is enriched for these processes (Fig S2a).

The network as a whole was also enriched for several viral-derived gene sets, including early lytic, secondary (maximally expressed 8-24hrs post induction), tertiary (maximally expressed 48-27hrs post induction) as well as viral genes that were not responsive to RTA (ORF50) and encoding viral capsid proteins or those involved in replication (Fig 4.6b). This may indicate another role of the network in promoting these viral functions. However K-core decomposed networks were consistently enriched for "Non Responsive", "Early Lytic" and to a lesser extent "Replication" (Fig S2b). This indicates that these viral processes are preferentially targeted too.

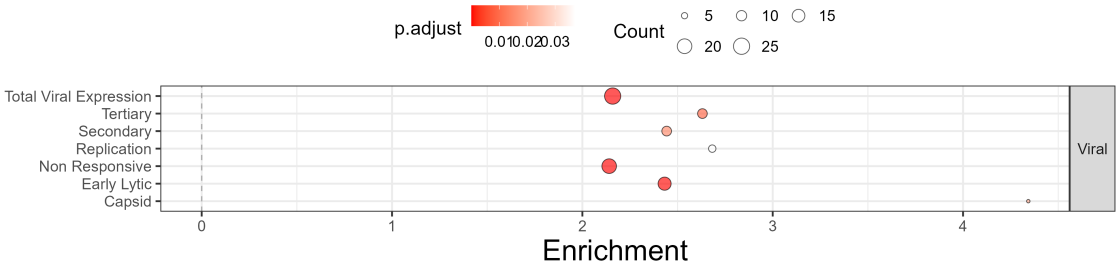
Enrichments for RNA biosynthetic and gene expression processes were present in the targets of miR-29b-1-5p, while miR-30b-3p, miR-3609 and miR-590-5p were enriched for the more general "cellular macromolecule metabolic processes" (Fig 4.7a). Many circRNA regulons (genes that miRNAs predicted to bind circRNAs are predicted to bind) were enriched for similar processes, particularly circDDX17_2, circLRCH3, circSH3PXD2A, circSMG1P1 and circSNX5_2 (Fig 4.8). This suggests these circRNAs as being particularly involved in regulating these proceeses. The sole term "cellular macromolecular metabolic process" was enriched for circAK308561, circARHGEF12, circELK4, circHIPK3 and circBAGE3 (Fig 4.8). Notably however only circHIPK3, circLRCH3, circSMG1P1 and circSH3PXD2A were predicted to bind miR-29b-1-5p, indicating that the enriched terms

were not solely due to the inclusion of miR-29b-1-5p in the network. Moreover, circSNX5_2, circLRCH3 and circDDX17_2 were enriched for terms relating to Ub-mediated protein catabolism indicating these circRNAs potentially regulated this aspect of network genes as well (Fig 4.8).

Therefore this section indicates that the up-regulation of circRNAs during lytic reactivation (and possibly *de novo* infection) facilitates the up-regulation of host genes involved in transcription, RNA biosynthetic processes and protein catabolism, alongside several functions of viral lytic replication.



(a) Total network enriched GO terms.



(b) Network enriched viral terms.

Figure 4.6: Significantly enriched gene ontology (GO) biological process (BP) and viral gene sets in network genes. Over-representation analyses (ORA) were performed on all network protein-coding and viral genes. Enrichment analyses are split by whether Gene ontology GO-BP gene sets or custom viral gene sets were tested for enrichment in full network. Top 50 host terms (a) by enrichment are shown that had FDR$<$0.01 and viral terms (b) are shown for FDR$<$0.05, in one-sided Fisher's exact tests. Dot sizes indicate the number of genes in each gene set that are annotated for that gene set and coloured by FDR-adjusted p-values.

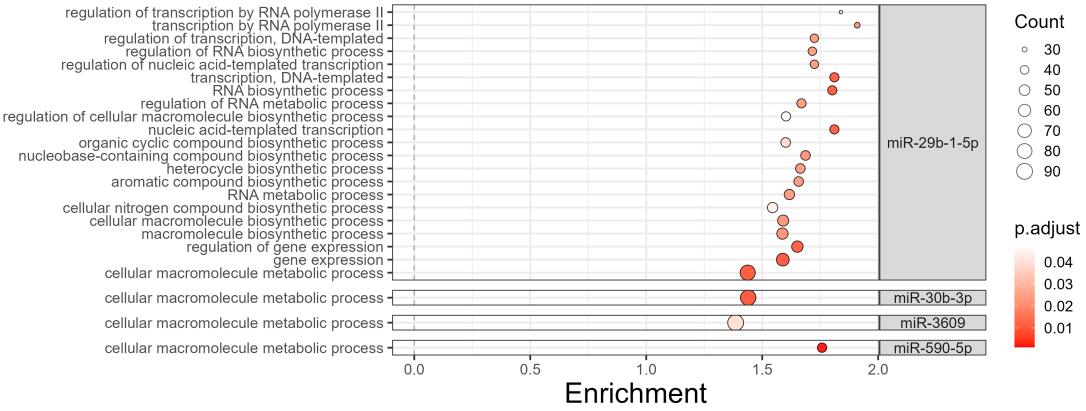


Figure 4.7: Significantly enriched gene ontology (GO) biological process (BP) terms in miRNA regulons. ORAs were performed in a miRNA-wise regulon manner. Terms are shown that had FDR$<$0.05 (one-sided Fisher's exact tests). Dot sizes indicate the number of genes in each gene set that are annotated for that gene set and coloured by FDR-adjusted p-values. Plots are split according to respective miRNA or circRNAs.

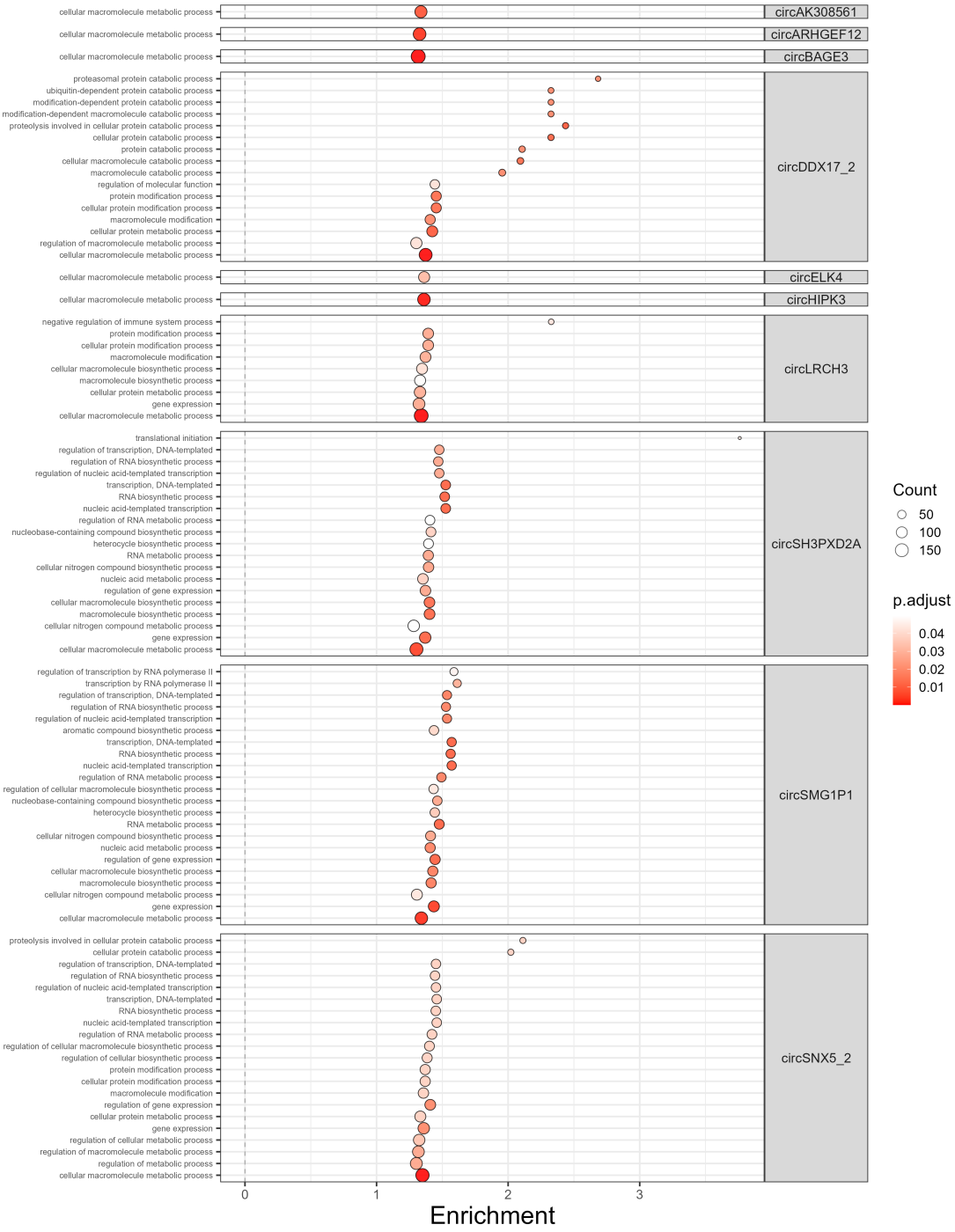


Figure 4.8: Significantly enriched gene ontology (GO) biological process (BP) terms in circRNA regulons. ORAs were performed in a circRNA-wise regulon manner, which area defined as all predicted targets of miRNAs predicted to bind respective circRNAs. Terms are shown that had FDR$<$0.05 (one-sided Fisher's exact tests). Dot sizes indicate the number of genes in each gene set that are annotated for that gene set and coloured by FDR-adjusted p-values. Plots are split according to respective miRNA or circRNAs.

4.6 Discussion

Overall this chapter has facilitated the modelling and subsequent analysis of a ceRNA network based on miRNA-target predictions, between mRNAs, miRNAs and circRNAs differentially during KSHV reactivation to lytic replication (Fig 4.1). Investigation of the distribution of miRNA targets suggested the down-regulated miR-3689a-3p, miR-30b-3p, miR-30c-1-3p and miR-29b-1-5p and up-regulated miR-3609 as the most prolific mediators of the network (Fig 4.2). Finally gene set enrichment analysis suggested that the network was involved in ribonucleotide metabolism, regulation of gene expression and protein ubiquitinylation. The significance of these findings as well as the limitations of analysis are discussed in the remainder of this section.

Of the most highly-ranked miRNAs, miR-30c-1-5p and miR-29b-1-3p were found to be sponged by circHIPK3 in the Harper *et al.,* 2022, while circHIPK3 was present as a component of this network [71]. CircHIPK3 is one of the most well-studied circRNAs and has been shown to be up-regulated in a range of cancers, while miRNA sponging is one suggested mechanism by which circHIPK3 has pro-oncogenic effects [202, 226, 283]. Notably although circHIPK3 was relatively highly ranked in terms of centrality measures and was present in the K=4 K-core decomposed network, it was not among the most high-ranking, particularly in terms of out-hub score (Fig 4.4a & Table 4.2). This highlights circHIPK3 as just a single component of a broader ceRNA network and potentially not the most influential regulator of the network. This underpins issues with approaching ceRNA networks by interrogating individual circRNA/miRNA/target axes in isolation as it may miss the broader regulatory context and complex dynamics of the network. Of note, however, miR-30c-1-3p was not predicted to bind circHIPK3 in the present network, despite the validation in Harper *et al.,* 2022 [71]. CircHIPK3 was however predicted to bind its co-family member miR-30b and miR-3689a-3p which have similar seed sequences, alongside miR-29-1-5p and miR-5087. This underpins the relative low quality of miRNA target prediction via sequence-based methods, which can have a false positive rate as high as 20-40% [186]. While some studies utilise the inclusion of known interactions to mitigate this, inclusion of these edges would have introduced study bias, invalidating much of the comparative centrality analyses.

Gene ontological enrichment analysis suggested an involvement of RNA metabolism and gene expression in the network, alongside protein ubiquitinylation implying that the network and by extension circRNA up-regulation may be to promote these processes (Fig 4.6a & S2a). Together these functions may be of benefit to KSHV as they may provide substrates for and the activation of machinery that facilitate the transcription of viral

genes, alongside modulating ubiquitinylation to alter protein function and abundance. Interestingly KSHV is already known to exploit the host Ub machinery extensively as well as encoding its own Ub ligases, for example RTA, K3 and K5 [284, 285]. Additionally, the targeting of predominantly early lytic and non RNA-responsive genes may indicate a role in de-repressing these genes and driving downstream lytic replication (Fig 4.1) & Table 4.1).

An issue with the present study is that the total RNA-Seq, small RNA-Seq and circRNA microarray datasets were derived from different KSHV lytic induction experiments. Moreover the time frames were not exactly the same, with 0 and 20hrs for the circRNA microarray and total RNA-Seq but 0 and 24hrs for small RNA-Seq. Given the known and observed differential trajectories between KSHV lytic induction experiments, this could result in a-synchronicity between the changes in circRNAs, miRNAs and mRNAs (Fig 3.3). Nonetheless any shifts may well be subtle particularly when averaged over the replicates. Moreover the in-network vs out-of-network permutation tests and distributional comparison tests provide confidence for a biologically *bona fide* relationship between interacting entities.

An issue related to the ambiguity in circRNA length determination is that longer circRNAs are more likely to contain predicted miRNA target sites, biasing longer circRNAs to having greater network degree. Indeed, network circRNAs were significantly larger than non-network circRNAs with mean lengths of 1305 and 534, respectively (W = 17126, $p<1e\text{-}13$, two-sided Wilcoxon rank sum test). This may not be an issue as predicted miRNA binding sites may still be functional, however when taken with the issues with of assumed circRNA length, may lead to spurious associations. Notably, circHIPK3 was uncharacteristically short (1099nt) for a network miRNA and comprised only one exon [226]. One way to mitigate could be to consider a length-normalised metric of miRNA binding sites in circRNAS (ie miRNA binding sites per 1000 nts).

This chapter has facilitated the modeling and analysis of a dysregulated ceRNA network between KSHV undergoing latent and lytic replication cycles. Centrality and K-core analyses were applied in order to rank nodes based on their relative influence in the network, while gene over-representation analyses suggested the involvement of process involved in gene expression and early lytic viral genes. Overall, while subsequent biochemical analyses will likely need to be performed to confirm the findings of such modelling, they may represent a novel mechanism by how KSHV promotes reactivation and subsequent lytic replication. This could add an extra layer to the already complex interactions between KSHV and infected host cells.

5 Kaposi Sarcoma gene co-expression network analysis

5.1 Chapter Introduction

Several recent studies have investigated global changes in gene expression between Kaposi sarcoma (KS) lesion and paired control lesion samples [16, 67, 94]. We recapitulated the analysis of one of these studies, Lidenge *et al.,* 2020, however this analysis (as well as the analysis of the similar studies Tso *et al.,* and Ramaswami *et al.,*) was relatively small in scope. Moreover their methodological approach has some particular limitations, given that the gene-wise differential expression analyses employed are by nature binary comparisons of summarized gene expression values, thus resulting in a substantial loss of information, particularly relationships between genes. In contrast, co-expression analyses empirically measure the relationship between the abundance profiles of genes in transcriptomic data, usually by correlation, which can't be detected by parallel gene-wise differential comparisons. Moreover co-expression analyses can empirically identify novel associations between genes that haven't been identified in previous research or annotated in ontological databases, for example between specific viral and host genes. The main draw back is that co-expression analysis requires enough datapoints to reliably calculate a measure of similarity between data. Importantly however, no publicly available dataset was large enough for this until the publication of the Lidenge *et al.,* data-set.

One of the most frequent co-expression analytical methods is co-expression network analyses (CNA), which extends co-expression analyses to a transcriptome-wide scale and subsequently applies network methods to the resultant data structures [266, 286, 287]. In such networks, nodes represent features (ie genes) while edges represent co-expressions, usually with a some form of a cut-off or thresholding to limit network density. Key to CNA is the "guilt-by-association" (GBA) principle, which is dependent on 2 key assumptions. The first is that features that show high similarity in their abundance profiles are likely involved in similar processes. This allows for genes to be grouped into biologically meaningful "modules" (also communities, clusters) of similarly co-expressed genes involved in similar functions. The second is that features which show a high density of similar profiles with other features likely have a high degree of influence on such neighbouring features [266]. This allows for a mechanism for identifying and ranking influential hub genes in co-expression networks. Moreover, by partitioning networks into modules, the landscape of gene expression in biological samples can be interrogated in more-detail, which can help to separate overlapping and potentially confounding gene-centric processes, including deconvolving cell-type specific influences [288]. This is particularly poignant for KS given the high cellular heterogeneity of KS lesions, both in terms of cel-

lular constituents and replication cycle stage (latent or stages of lytic) of infected cells [16, 67, 87, 89].

Weighted gene CNA (WGCNA) is the most frequently applied method of performing network co-expression analysis and has been applied plentiful times to identify key influential hub genes in many pathological conditions [266, 289–291]. Moreover it has been extensively applied to patient-derived biopsy samples to identify modules of co-expressed genes with relevance to pathological conditions in question, or "disease modules" [267]. This includes in the context of viral infection, including EBV-associated gastric carcinoma [292, 293]. Such disease modules are interrogated via measuring their associations with phenotypic traits or to identify potential highly connected hub genes that may play a role in the condition of interest.

A newer variant of co-expression network analysis is differential co-expression network analysis which, identifies genes that show differential co-expressions between conditions [269]. Such changes can be applied to identify activatory and inhibitory relationships, among others, that are missed by differential and co-expression analyses [269, 294–297]. Importantly, applied to Lidenge *et al.,* such analyses explicitly includes information about genes that change associations between lesion and uninvolved tissue and as such may facilitate the discrimination of important general regulators in KS and those specifically associated with lesion development.

Therefore the goal of this chapter was to construct a co-expression network to model the transcriptomes of KS lesion tissue biopsies in detail. This model was then interrogated to identify modules of co-expressed genes which were then tested for enrichments of various host and viral processes and associated with various sample-wise traits. WGCNA hub genes were then identified and interrogated. Finally, differential correlation network analysis using DGCA was performed on key modules in order to identify differentially correlated hub genes that may play a key role in KS.

5.2 Comparison of batch adjustment methods suggests ComBat as the most effective batch-adjustment method

Batch effect adjustment can confound downstream analysis by preventing the lack of the capacity to discrimination between true biological effects and artefactual differences. Moreover adjusting for batches can improve downstream analyses, including gene co-expression network construction [287, 298]. As the 23 samples were split between 4 batches, the influence of batch effects and their elimination via batch effect adjustment methods were investigated.

A PCA applied to the un-adjusted data revealed some evidence of batching associated with PC3 (Fig 5.1). Specifically, batch 1 and 2 grouped together, with some separation from batch 3 and 4, which grouped together. While this variance due to batch was associated with a relatively low amount of total variance in the data-set (variance explained by PC3 = 6.23%) and in fact much of this variance separates a single sample in batch 3 from the other samples, this still could lead to spurious relationships or mask or confound biologically relevant co-expressions, necessitating batch adjustment.

One issue with the Lidenge *et al.,* dataset is the presence of complete confounding of endemic samples within batch 2 (Fig 2.1). This poses two issues, 1. that the influence of endemic samples cannot be distinguished from the influence of batch and 2. that recent studies have indicated that when trying to correct for such un-balanced experimental designs, batch-adjustment procedures can artificially over-correct for group differences, resulting in artificial changes in effect sizes between levels of the unbalanced group [299, 300]. For the first point, elimination of the influence of endemic samples is likely of benefit so we reasoned that this was not a major point of concern. But for the second point, this could skew the magnitude of co-expressions of genes that varied between epidemic and endemic samples. To investigate this second issue, three batch adjustment methods (ComBat, ComBat-Seq and Limma's removeBatches()) were trialled with (HIV-aware) and without (HIV-naive) the inclusion of endemic/epidemic sample labels as a factor during batch adjustment [264, 301, 302]. PCA was similarly applied to these 6 different batch-adjusted data-sets. This revealed that both adjustment formulations with Combat-Seq resulted in the considerable increased separation of endemic and epidemic samples (and batches), while Limma's method showed considerable batching in the HIV aware but not HIV-naive approach (Fig 5.1). However both approaches using Combat showed limited emphasis of variance due to batch or HIV status and as such this adjustment method was taken forward (Fig 5.1).

Lidenge *et al.,* 2020 showed that the main difference between endemic and epidemic KS lesions transcriptomes was a greater magnitude of differential expression between lesion and control tissue. Such analysis was repeated for un-adjusted and ComBat HIV-naive and -aware data, which showed a negligible decrease in the effect sizes between the differential transcriptomes of endemic vs epidemic lesions (linear regression β coefficients of 1.54, 1.51 and 1.51 for un-adjusted, HIV-naive and -aware approaches, respectively) (Fig 5.2). Genes uniquely differentially expressed in endemic samples for the HIV-naive and HIV-aware adjusted datasets showed a similar number of uniquely differentially expressed genes (539, 486, respectively) to the number for the unadjusted dataset (529). Finally HIV-naive and -aware comparisons showed high overlap with the unadjusted comparison (JSI = 0.66, 0.66, respectively). This was despite the unfavourable usage of Limma on over-dispersed, pseudocount and log2- transformed RNA-Seq data (as explained in Section 2.3.4).

Therefore HIV aware ComBat-adjusted data were taken foward. It is worth noting that all methods resulted in a set of 4 control samples with notably lower mean gene expression of viral genes (Fig S4). These 4 samples are distinguished towards the top left of the PCA plot, as largely separated from the other group of control samples (Fig 5.1). However the lesion samples, which were of most interest, were taken forwards. Therefore batch effect adjustment showed considerable success in removing variance due to batches while maintaining biological signals comparable to un-adjusted data.

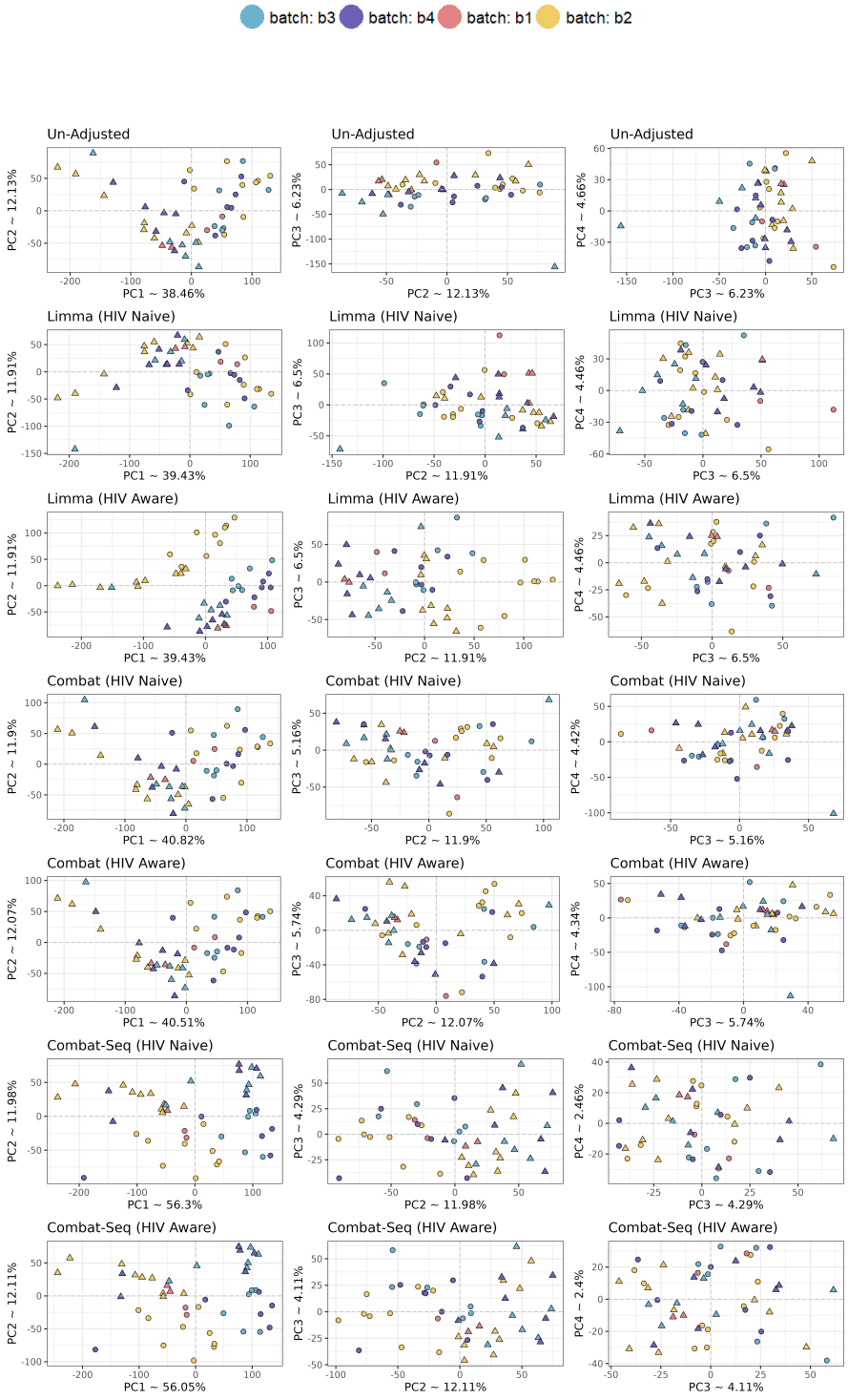
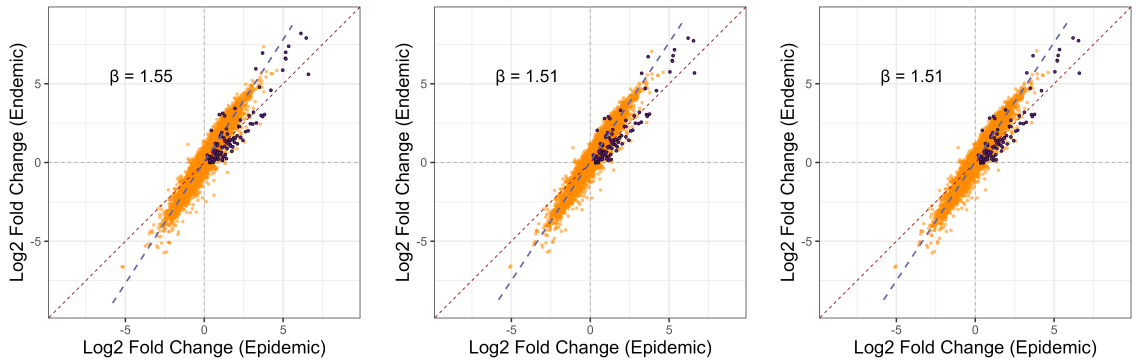


Figure 5.1: Principal components analysis to assess the efficacy of batch adjustment methods. PCA score plots on un-adjusted and batch-adjusted log2 TPM expression data. 3 batch adjustment methods were trialled: Limma's removeBatchEffect(), ComBat and ComBat-Seq, with the exclusion and inclusion (HIV-naive and aware, respectively) of HIV-1 co-infection (epidemic/endemic KS) in the model parameters. Plots are labelled according to the batch adjustment method and HIV-naive and aware labels. PC1-4 are shown plotted in successive pairwise orders horizontally; with PC1 and PC2 (left), PC2 and PC3 (center) and PC3 and PC4 (right). Points represent samples, coloured by batch and shaped according to lesion (circle) and non-lesion (triangle).



(a) Comparison of log2 fold change values, pre-batch adjustment

(b) Comparison of log2 fold change values, post-HIV naive batch adjustment

(c) Comparison of log2 fold change values, post-HIV aware batch adjustment

Figure 5.2: Comparison of endemic and epidemic lesion vs control biopsy tissue differential expression log2 fold-changes. Differential expression on log2-transformed, un- or batch-adjusted expression data was performed using Limma between lesion and control samples for epidemic (x axes) and endemic KS patient samples (y axes) (further detailed in Section 2.3.4). This was performed using non batch-adjusted data (a), batch-adjusted data without the inclusion of HIV co-infection as a factor in model parameters (b) and batch-adjusted data with the inclusion of HIV co-infection as a factor (c). The red line indicates a perfect 1:1 relationship between results (ie intersect of 0, gradient of 1) while the blue line indicates the fit of an ordinary least squares regression line, with slope coefficient for each plot show as β. Genes are coloured according to being host (orange) or viral (purple).

5.3 Initial construction of co-expression network

Co-expression network construction is dependent on calculating similarities in the variance of gene expression profiles of genes. This enables the calculation of a co-expression measure, which is then thresholded to facilitate the determination of edges between nodes (genes).

Lowly expressed genes are more likely to be representative of unwanted noise in expression data while lowly variant genes will show weak and thus likely uninteresting co-expressions. Moreover lowly expressed genes tended to show a greater dependence between their expression and variance (Fig 5.3). So prior to network construction, gene expression data were filtered to remove the bottom 5% and 25% of genes by mean expression and variance, respectively, leaving 16534 (16441 protein-coding and 93 viral) and 12401 (12308 protein-coding and 93 viral) genes, respectively. Such filtration resulted in a near-flat expression mean-variance trend, with no correlation between mean and the standard-deviation (Pearson's r = 0.016) (Fig 5.3).

Correlation is the most common measure of co-expression used to perform WGCNA. Spearman's ρ was chosen to limit the influence of outliers in small data-sets (under 30 samples) used for co-expression network construction, which it has been previously shown to be effective at doing [287]. Spearman's ρ also provides a more reliable parameter estimate as a measure of similarity between pairs of genes that show a non-linear relationship, as well as not relying gene expression data to be normally distributed.

Raw correlation values are not ideal measures of co-expression as they can result in dense networks with many spurious edges [266, 287, 303]. WGCNA employs several key steps to limit this and emphasise only strong measures of co-expressions [266]. First, the "signed hybrid" transformation was applied to the correlation matrix by setting all negative values to 0, as negative values in metrics of co-expression have been suggested to be less reliable measures of true co-expression than positive correlations [266]. Soft thresholding was then performed on the positive correlation values by raising them to the power of a constant value β, which emphasises and minimises stronger and weaker correlations, respectively. The optimal β was chosen via sequentially raising gene expression values by sequential values of β (ranging from 1 to 25, iterating by 1 each time) and measuring of the similarity of the the log node degree distribution to a scale-free distribution as measured by an R^2 value (Fig 5.4). The soft-thresholding β value was chosen as 6, as the maximum R^2 fit value (also the first to exceed R^2 = 0.8) (Fig 5.4). Next, the weighted topological overlap (wTO) transformation was applied to β-transformed Spearman's ρ,

which adjusts edge-weights based on the connectivity of shared neighbour genes [266]. These three transformations limit the influence of weak correlations which are more likely spurious as well as reducing network density (see Fig 5.4 & Section 2.3.5). As such, the resultant non-zero transformed correlation values were chosen as the final network network edges and these values taken forward as edge-weights, to generate an edge-weighted network.

In summary gene expression data were filtered in preparation for WGCNA network construction. Spearman's ρ was chosen as the initial measure of co-expression, the optimal transformation parameter β (β=6) applied and wTO transformation subsequently applied to define final edge weights. This network model could then be interrogated in in further analytical steps.

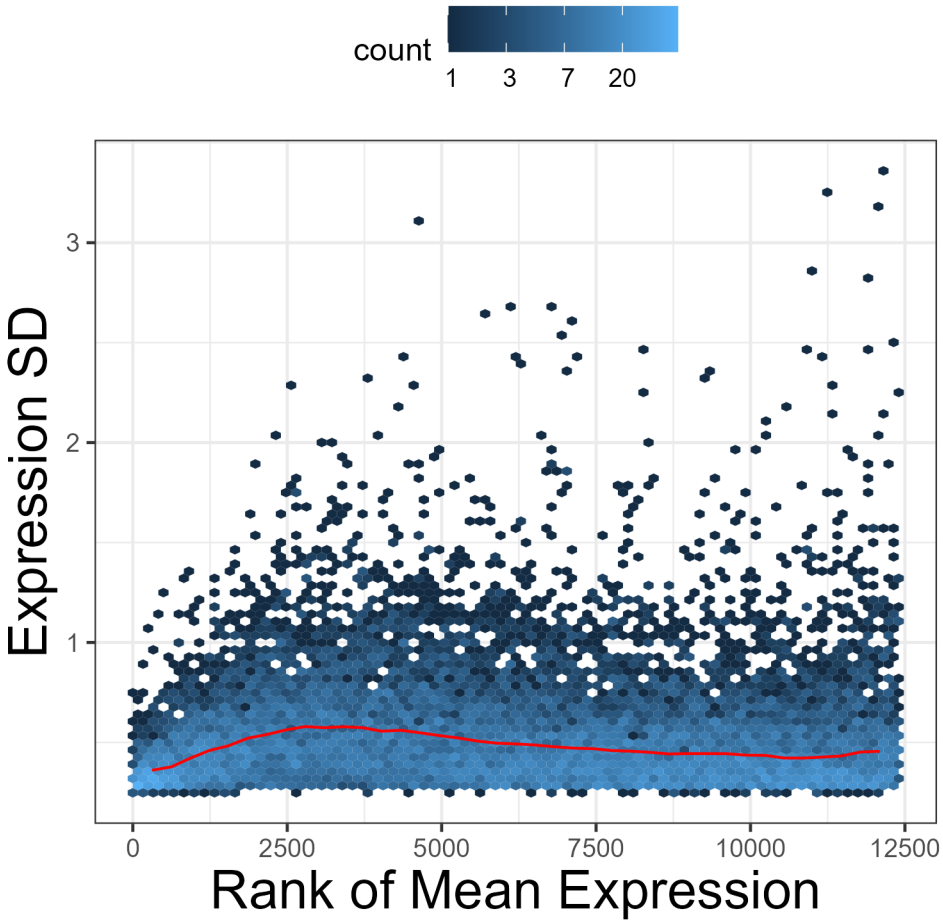


Figure 5.3: Mean-variance trends for expression- and variance-filtered data. Hexplots show the 2D distribution of rank of mean (x axis) and standard deviation (y axis) of log2-gene expression data. Expression values represent RNA-Seq count data that have been subject to TPM and log2 pseudocount-transformation followed by Combat batch-adjustment (detailed further in Section 5.2). Expression data were then filtered to remove the bottom 5% of and 25% of genes by mean expression and expression variance. Red line shows mean trend.

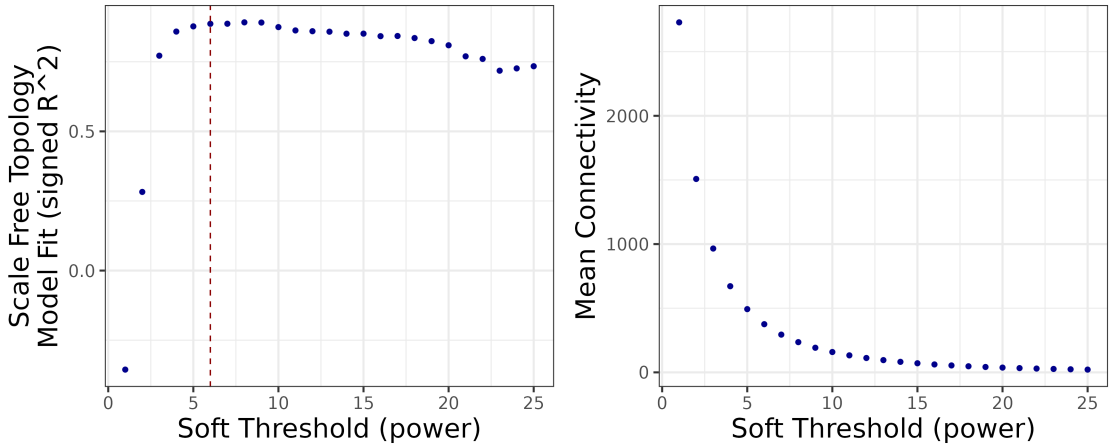


Figure 5.4: Selection and comparison of WGCNA soft-thresholding parameter. Networks were constructed from lesion samples using Spearman's ρ correlation as a co-expression measure, and such pairwise gene-gene co-expression measures were then raised to varying levels of β, from 1-25, increasing by 1 each time. The fit of the resulting network connectivity (mean of node edgeweights) distributions were compared to a perfect, scale free (power law) fit via calculation of R^2. Left shows the distribution of R^2 for the signed network, with respect to increasing values of β, where the red line indicates the chosen soft thresholding parameter β value of 6. Right shows the mean node connectivity of the network when transforming edges by successive values of β.

5.4 Module partitioning identifies modules associated with distinct subsets of viral genes

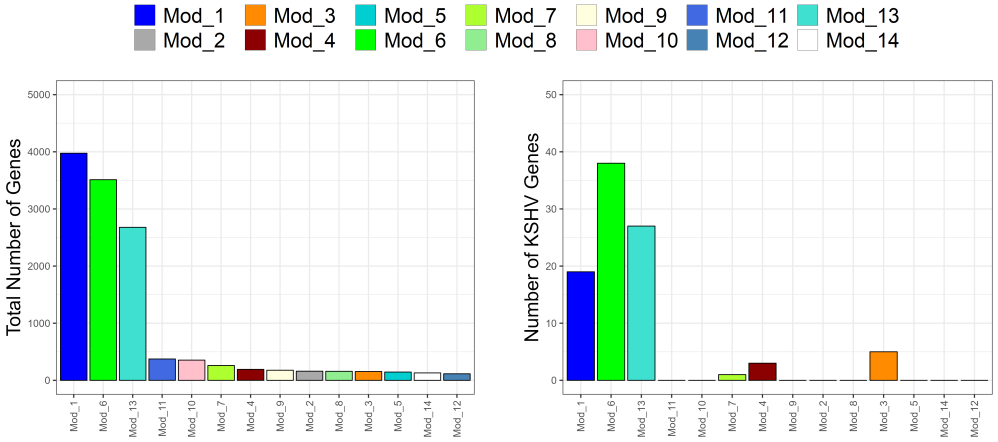
One feature common to most networks is the tendency for nodes to group together in modules (communities or clusters), whose constituents have functional or regulatory similarities. This allows the transcriptome to be deconvolved into discrete sub-networks that enable the isolated interrogation of such discrete co-expressed units. To this end, initial module partitioning was performed using WGCNA's standard hierarchical clustering-based method, outlined in Methods Section 2.3.7. This resulted in 22 modules of varying sizes comprising 12385 genes in total, with 16 genes (SLC6A7, DDX3Y, PF4V1, C4BPA, CELA1, C1QTNF2, DRC1, PRAC1, CXCL5, DEFA4, BNIP3, LEMD1, GPAT2, HLA-C, SPINK8 and KIR3DL2) not being assigned modules and consequently discarded. Decreasing the cutHeight parameter was found to result in grouping of the vast majority ($>$10,000) genes into one module, which proved untenable for downstream analyses. Moreover a deepSplit parameter of 3 was chosen as lower values were found to result in an excess of genes being discarded to the unassigned group of genes, including many viral genes of interest and relevancy to KS. Similarly, setting pamRespectsDendro = TRUE resulted in the discarding of the majority of viral genes and as such this was set to FALSE to retain these genes for downstream analyses.

Initial module identification was followed by a second round of module merging. First module eigengenes (MEs) were calculated as PC1 of a PCA performed on the module-wise gene expression matrix. These MEs represent a summarised expression profile of genes in each module, in each sample. A threshold percentage of 15% and merge percentage of 35% was found to be be optimal parameters for this process, based on empirical observations of clustering of viral gene correlations (Fig S6). This resulted in 14 modules (Fig 5.5a & b). Next, module membership (MM) scores for each viral gene were calculated as their correlation with the ME of their "parental" module (Fig 5.5c, d & e). This provides a measure of module-wise centrality for these genes.

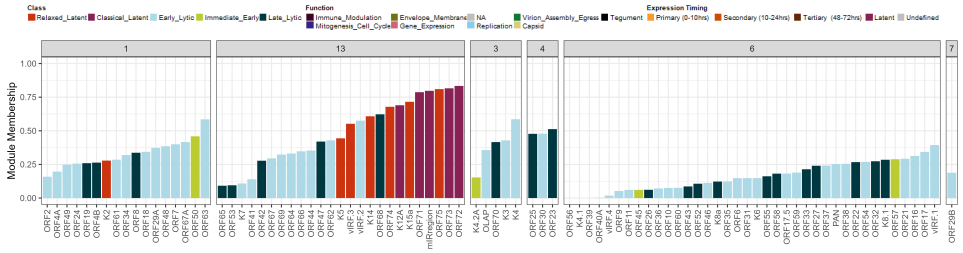
The resultant 14 modules were of varying sizes and contained different combinations of viral genes. The largest module was Mod 1 which comprised 3994 genes (3975 host, 19 viral) (Fig 5.5a). Constituent viral genes were all canonical lytic genes that showed high MM and included the NLRP3 inflammasome homolog ORF63, ORF50 which encodes RTA, the major activator of lytic replication and ORF67A, which is associated with the disassembly of the nuclear lamina prior to virion egress (Fig 5.5d) [250, 304]. K2 (vIL-6), was also present in this module (Fig 5.5c). Interestingly the genes in this module tended to be early lytic or maximally expressed 8-24hrs post lytic induction (Secondary) (Fig

5.5e). The second largest module was Mod 6 which contained 3545 genes (3994 host, 33 viral) and contained genes involved in immunomodulation, gene expression, replication, capsid and the tegument, including vIRF-1, ORF17, ORF16 and ORF57 which showed the highest MM for this module (Fig 5.5c). However several lytic showed low module membership for Mod 6 and the overall size and relatively small variance explained by its ME indicates it is relatively heterogeneous (Fig 5.5 & SS7). The third largest module was Mod 13 which comprised 2705 genes (2678 host, 27 viral) including all the classical latent and remaining relaxed latent genes, which all ranked among the most highly by MM (Fig 5.5b & c). Most constituent genes were involved in mitogenesis and the cell cycle as well as lytic genes involved in immunomodulation, including K5, vIRF-2 and vIRF-3 (Fig 5.5d). The remaining 9 viral genes were distributed between 3 modules, with 5 in Mod 3, 4 in Mod 4 an 1 in Mod 7 (Fig 5.5).

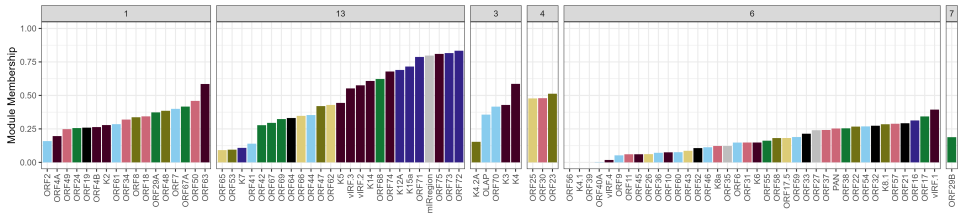
Overall partitioning of the networks into co-expressed modules resulted in 14 modules, of which the majority ($\sim$81%) of genes were present in 3 modules, Mod 1, 6 and 13. These modules contained the majority of viral genes, which partitioned in-line with their functional and expression characteristics.



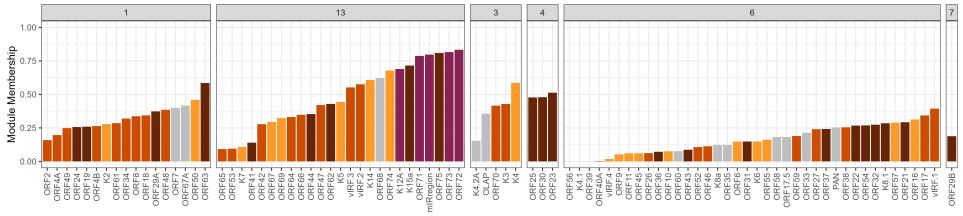
(a) Total number of genes per module. **(b)** Number of KSHV genes per module.



(c) Coloured by viral gene type.



(d) Coloured by viral gene function.



(e) Coloured by maximal viral gene expression timing.

Figure 5.5: Distribution of genes across the network modules. The network was partitioned into 14 modules as outlined in Methods Section 2.3.7. The total number of host and viral genes are shown in (a) and (b), respectively. Viral gene module membership (MM, where MM is the Spearman's ρ correlation between the viral gene and the respective module as outlined in Methods Section 2.3.8) for each viral gene and the module they are partitioned into (parental module) (c, d & e). Viral genes are split according to their parental module. Genes are coloured according to viral gene classification (a), gene function (b) and maximal expression timing (c) as outlined in Methods Section 2.1.6.

5.5 Ontological Analyses on modules characterises host gene expression associated with subsets of viral genes

A key facet of the GBA principle is the assertion that genes involved in similar processes tend to be co-expressed [266]. Therefore modules can be assumed to comprise genes that show relationships between their function, as well as expression. This can also relate to the transcriptomes of differing cell types present in heterogenous samples [288]. Therefore investigating the processes associated with such co-expressed genes can and has been shown to produce biologically meaningful findings. To investigate such module-wise functional enrichments, gene set over representation analyses (ORAs) were performed on each module. Significant terms were then retained (FDR$<$0.001).

Modules showed varying enrichments of host processes as well as various patterns of constituent viral genes. In-line with Mod 13 containing mostly viral mitogenic genes, Mod 13 was enriched for mitogenic processes, relating to chromosome localisation and cell cycle checkpoint signalling (Fig 5.5d & 5.6). As Mod 1 contained ORF50 which is known to be a potent modulator of factors that regulate transcription, it was enriched for processes relating to the initial activation of gene expression, including "histone modification" (Fig 5.6). Further Mod 1 enriched terms represented those controlling further aspects of gene expression, including "posttranscriptional regulation of gene expression", "regulation of translation" and "nuclear-transcribed mRNA catabolic process" (Fig 5.6). Given ORF50's ubiquitin ligase activity, it is interesting to note that processes relating to ubiquitin-mediated proteasomal degradation were also enriched (Fig 5.6) [284]. Genes in both Mod 1 and Mod 13 tended to be up-regulated in lesion relative to control tissue and thus these processes can be assumed to be increased in KS (Fig 5.7b). Mod 6 also contained many lytic genes, including the immediate early genes ORF45 and ORF57, alongside many genes involved in replication and gene expression [68, 305]. From the perspective of the host it was enriched for many immunological related processes, particularly canonically anti-viral immune responses such as Th1-type immune responses (Fig 5.6). Moreover humoral and complement immune processes and superoxide anion generation were enriched (Fig 5.6). Further processes included aspects of metabolism (organic hydroxy, alcoholic, phenol-containing, acid, retinol and arachidonic and fatty acids) (Fig 5.6). Genes in this module showed a tendency to be down-regulated, thus the activity of its enriched processes can be assumed to be decreased in lesion tissue (Fig 5.7b). Mod 5's genes were enriched for various metabolic processes such as fatty acid and lipid metabolism (Fig 5.6). Mod 8 was enriched for "mitochondrion organisation" (Fig 5.6). Genes in this module showed a tendency for down-regulated and so these processes

can be considered to also be down-regulated in lesion tissue (Fig 5.7b). Interestingly down-regulation of lipid metabolism and mitochondrial dysfunction has previously been observed in KS lesions, despite increased lipid metabolism generally being associated with the Warburg effect seen in many cancers [16, 67].

Therefore host module genes were enriched for processes relevant to KSHV and KS. Host gene process enrichments also reflected known functions of their constituent viral genes. Specifically Mod 13 was most closely associated with mitogenic processes, Mod 1 was most associated with lytic processes associated with changes to multiple facets of gene expression and Mod 6 was generally enriched with immune functions indicating that the viral genes present were more immunogenic. All together this shows that co-expression-based transcriptomic partitioning results in modules that show some functional associations between their constituent viral and host genes.

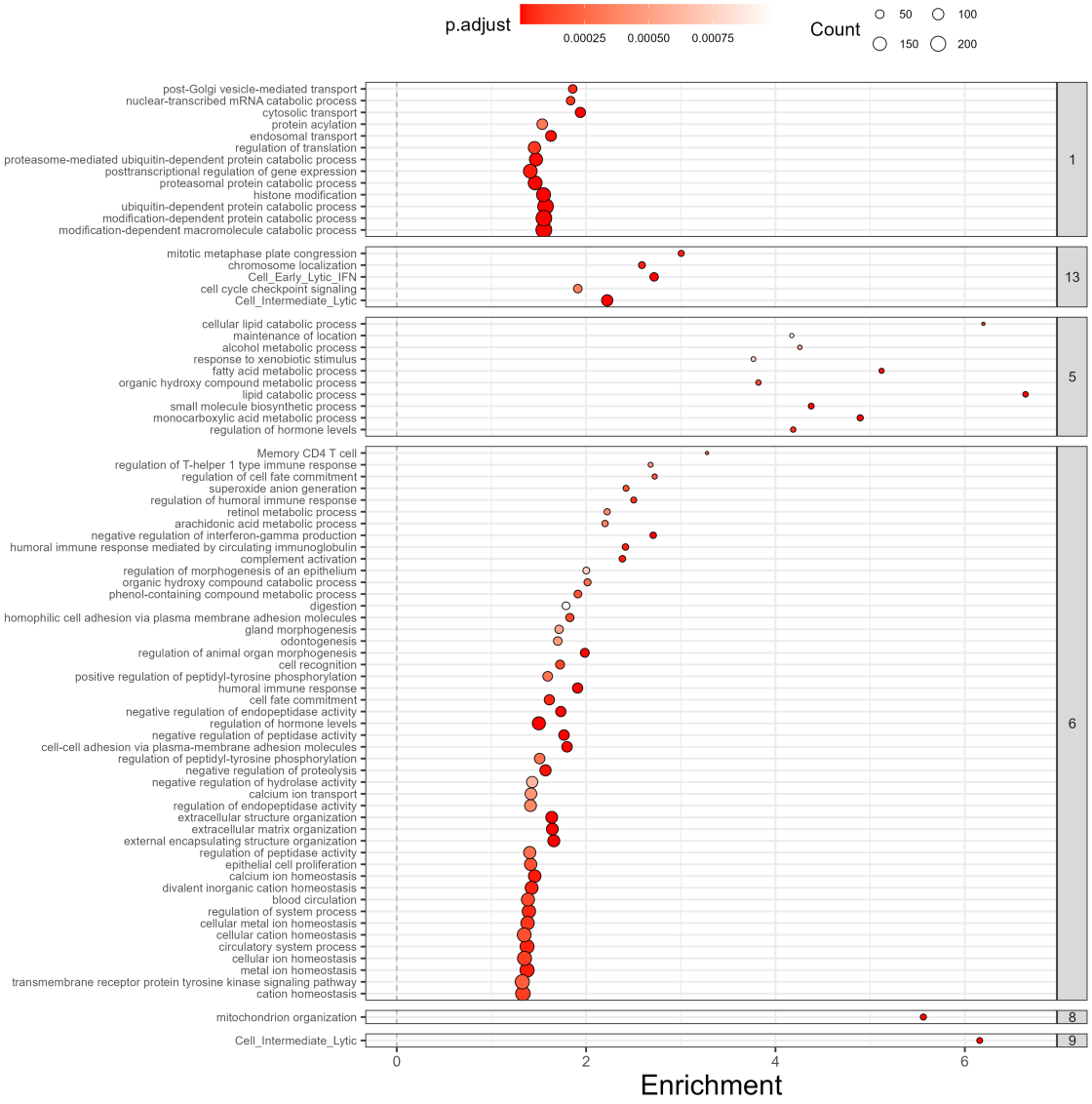
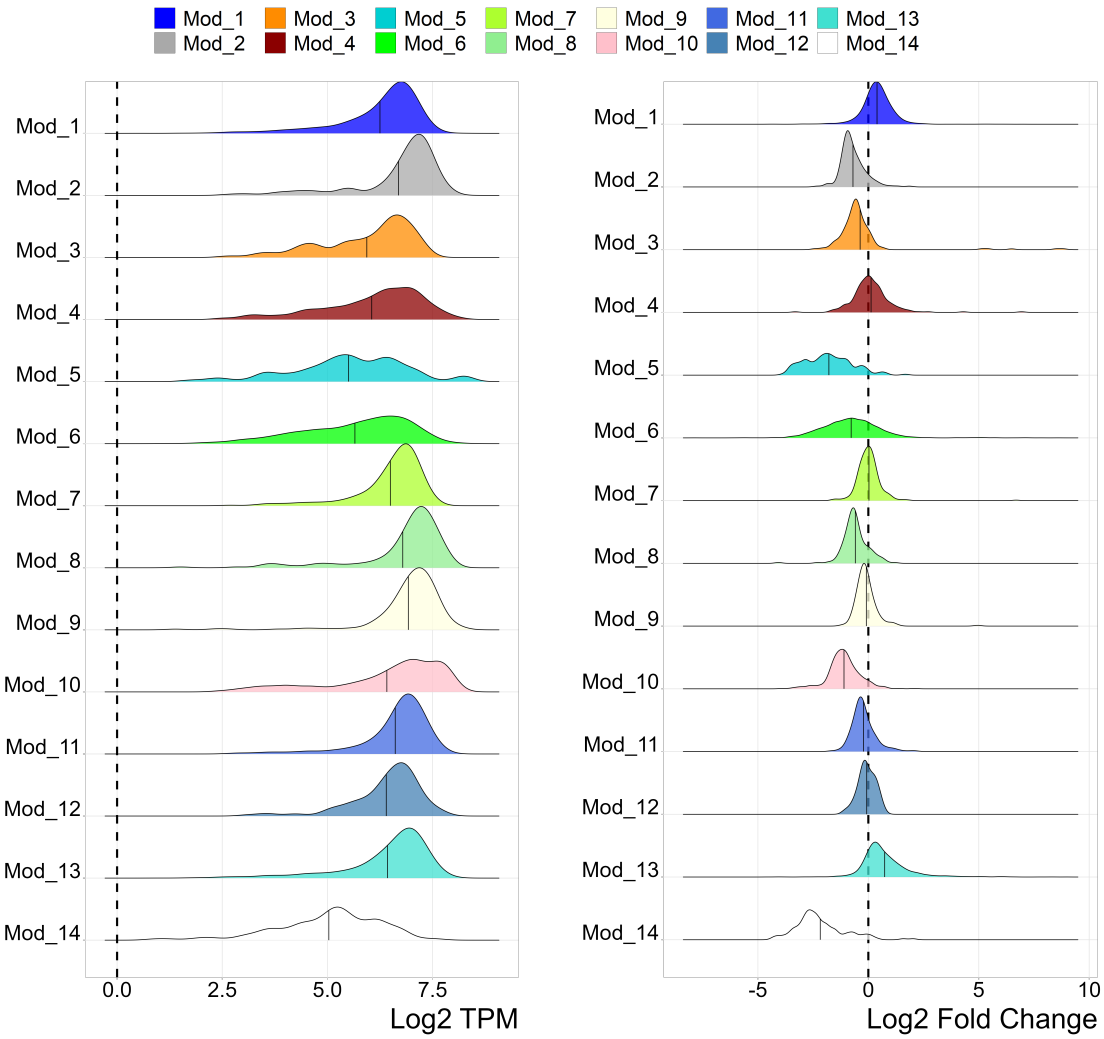


Figure 5.6: Gene ontology and custom gene set enrichments for WGCNA modules. Ontological enrichments for GO biological process terms were calculated for all genes in each module, using a one-sided hypergeometric test. All significant terms were retained ($p<0.001$). The size and ordering of the dots represent the number of genes annotated for that term in that module, while colour indicates the false discovery rate (FDR). Terms are split by module, as annotated on the right hand side.



(a) Log2 mean expression.

(b) Log2 fold-change.

Figure 5.7: Distribution of module gene expression and log2 fold-changes. Genes were partitioned into modules as in 2.3.7 & Section 5.4. Log2 mean expression (a) and Log2FCs (b) distributions are presented, coloured by their respective modules. Log2 mean expression values were calculated from TPM-transformed, Log2-pseudocount-transformed, batch-corrected gene expression data for 23 KS lesion biopsies (See Section 2.1.2). Gene-wise log2 fold-changes (log2FCs) were calculated between lesion and control biopsies as in Section 2.3.4.

5.6 Correlation of module eigengenes identifies primary partitions of the network associated with dysregulaton

Module eigengenes (MEs) are calculated as PC1 of a module-wise PCA performed on lesion tissue. As successive PCs captured negligible proportions of variances in the dataset, MEs can be viewed as a "de-noised" summarised expression profile of all genes in the module. As such the relationship between module eigengenes provides global information on the relationships between the expression of module genes and thus the topology of co-expression networks.

It is generally not recommended to consider the sign of MEs when comparing them as the sign of principal components are arbitrary. However when correlating sample-wise MEs in a pairwise manner, the relationship between modules reflected the relationships between their constituent viral genes expression profiles (Fig 5.8 & S6). This indicates that MEs are representative of the relationships of their constituent viral genes and indicates that these relationships are also present at the level of whole modules of host genes, not just viral genes. This is likely due to the importance of viral gene expression in driving KS [16, 67, 138]. Module eigengenes could be grouped into 3 clusters, X, Y and Z, based on their relative patterns of pairwise correlations (Fig 5.8). Specifically they all showed relatively strong positive intra-cluster correlations and predominantly negative inter-cluster correlations. Cluster X contained Mod 6, the module containing the most viral genes including ORF45 and ORF57 and was associated with immune processes, alongside Mod 10 which was enriched for lipid metabolic processes (Fig 5.6 & 5.8). Cluster Z contained the classical/relaxed latent gene module Mod 13, as well as Mod 1, Mod 4, Mod 7 and Mod 12. The modules in these two clusters generally exhibited a negative relationship with each-other, particularly Mod 13 with Mod 6 and Mod 10 with Mod 1 and Mod 10 (Fig 5.8). Cluster Y was somewhat intermediary and contained Mod 8, which was depleted for mitochondrial function (Fig 5.6 & 5.8).

In re-producing the analysis of Lidenge *et al.*,, we showed that PC1 of a combined control and lesion PCA (PC1_paired) of separated samples by these two groupings and showed PC1 was positively correlated with the expression of most viral genes (Fig S5). Moreover, PC1 and PC2 of a lesion-only PCA (PC1/PC2_lesion) were near-identical to PC1/PC2_paired (Spearman's ρ = 0.98 and 0.98, respectively), indicating similar separation by and redundancy of these pairs of axes (Fig S9). Additionally, PC2_lesion separated samples based on the relative expression of classical/relaxed latent and the remaining lytic genes, which showed positive and negative correlations, respectively (Fig S9). This relationship showed some similarity at the level of modules, primarily as most

viral gene-containing modules correlated positively with PC1_lesion, while Mod_1 and Mod_6 correlated negatively with PC2_lesion and Mod_13 correlated positively (Fig 5.9). This indicates that as well as the classical and relaxed latent genes, a subset of ORF50-centric viral and also host genes are associated with the development of KS. Conversely, Mod_6, which contained lytic genes, was itself negatively correlated with PC1_lesion indicating that constituent viral genes associated host genes were adversely associated with lesion development (Fig 5.9).

While the modules of cluster Y showed variable correlations with PC1_lesion, the three modules that comprised it showed very strong positive correlations with PC2_lesion, as did Mod_2 from cluster X, indicating their association with classical/relaxed latent expression (Fig 5.9).

Cumulatively this section has characterised the relationships of module expression profiles with all other modules as well as with the major axes of variance. This suggested a strong relationship between Mod 1 and Mod 13 and the development of KS. Therefore the relationship between viral genes is constituted at the level of the modules that they are constituents of. However, it is important to note that while module gene expression related to PC2_lesion in a similar manner to the expression of their constituent viral genes, the inclusion of the endemic samples may bias this observation, a point that is expanded further in this chapters discussion section (Section 5.9).

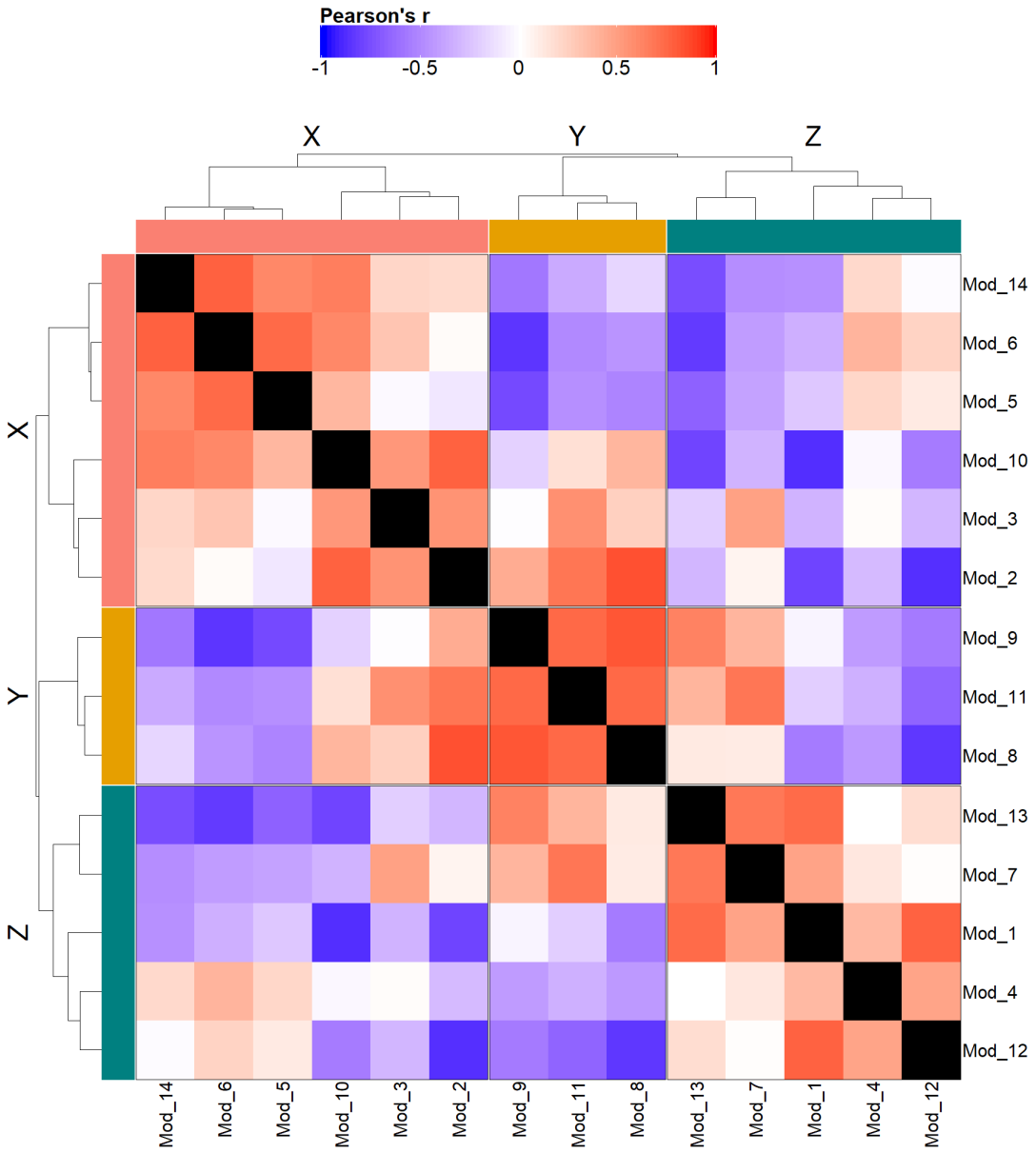


Figure 5.8: Module eigengene correlation matrix heatmap. Pairwise correlations were performed between module eigengenes using Pearson's r. Heatmap cells are coloured by Pearson's r. Diagonal values were set to NA and coloured black. Columns and rows were correlated using agglomerative hierarchical clustering using euclidean distance as a distance measure. 3 clusters were chosen and labelled as X (Mod 2, 3, 5, 6, 10 and 14), Y (Mod 8, 9 and 11) and Z (Mod 1, 4, 7, 12 and 13). Left and top annotations correspond to these clusters.

Figure 5.9: Correlations of module eigengenes with principal components. Module eigengenes (ME) were computed as the first principal component (PC) of a principal components analysis performed on the module-wise expression matrices. MEs were then correlated via Spearmans ρ with PC1 and PC2 for the KS lesion-only PC analysis (PCA) performed on the gene expression data (as in Section 5.9). Figure is split by correlations with PC1 (top) and PC2 (bottom). Modules are shown along the x axis and split according to the module clusters identified in Fig 5.8.

5.7 Investigation of hub gene networks proposes key signalling axis involved in KS

Another key inference of the GBA principle is that highly inter-connected genes (hub genes) in co-expression networks indicate a regulatory relationship between these genes and their neighbours. MM is one metric for assessing this, however MM alone does not strictly consider the network structure and is essentially a module-wise global centrality measure. Conversely, connectivity is a local centrality measure and is the sum of a gene's edge-weights (co-expressions) its neighbours in the network (equivalent to weighted node degree) [266]. By filtering by these two metrics and applying a sufficient cut-off for both, genes could be filtered for local and global importance and as such a set of robust hub genes that may play key roles in KS could be identified.

Like MM, connectivity can be measured only for edges connecting to other genes in a shared module (intramodular connectivity, iC) or for all genes in the network (global connectivity, gC) [266]. Both iC and gC were calculated for all module and hub genes. Notably, this resulted in rankings of viral genes by iC and gC that were very similar to their rankings by MM (Fig 5.5c, d & e).

Hub genes were determined by retaining the intersection of host genes that ranked in the top 10% by iC and parental MM, and further filtered to retain genes that were significantly differentially expressed (FDR«0.001) in lesions vs control tissue, resulting in 341 genes. As this biases to module-wise hub genes, an additional set of hub genes were defined as the intersection between the top 5% of host and top 25% of viral genes by mean MM and gC and that were differentially expressed between lesion and control tissue (FDR«0.001). This set comprised 146 genes, with 43 being common with module hubs (with ORF73 being the sole joint modular and global viral hub gene). All together this resulted in a total hub list of 444 hub genes (426 host, 18 viral), hereby termed "WGCNA hubs". The top 10 edges for each hub gene were then retained to generate a hub network. This resulted in decomposition of the network into two components, one representing Mod 1 and 13 and the other Mod 6 (Fig 5.11). This indicates a high level of co-expression of Mod 1 and Mod 13 genes and as these were most strongly correlated with PC1 and thus lesion development, this component was focused on (Fig 5.9). The top 25 hub genes by this resultant Influential hub score (see Methods Section 2.3.8) for each sub-component are labelled in Fig 5.11.

Gene set over-representation analyses were applied to the resultant set of hub genes. This showed an enrichment of small GTPase signal transduction and associated pro-

cesses, including signalling by Rho and Ras proteins (Fig 5.10). In particular this included NRas but also Ras guanyl releasing proteins (RASGRF), RASGRF2, RASGRP3 and RASSF10 (Fig 5.11). Notably RASGRP3 and a key downstream transcription factor of Ras-signal transduction via the Raf/MEK/ERK MAPK cascade, ETS1, ranked highly by influential hub score in the hub network (Fig 5.11) [306]. The downstream enhancer of ERK signalling that also activates ETS1, ribosomal protein S6 kinase (RPS6K) A3 (RPS6KA3), was also a very high ranking hub gene (Fig 5.11).

Many of the above processes and genes promote angiogenesis via downstream effectors and thus it is not surprising to see that angiogenic processes were enriched in the hub gene set (Fig 5.10). This is due to the presence of a set of pro-angiogenic genes, notably the oncogene SRGAP1, β-catenin (CTNNB1) and CTNNIBP1 (a negative regulator of β-catenin) (Fig 5.11, [307]. Additionally TCF4, a transcription factor that complexes with β-catenin signalling to promote angiogenesis was a high-ranking hub gene (Fig 5.11) [308]. Further related hub genes related to angiogenesis included MIB1, WWC2, PDZRN3, GNA12/13, AMOTL1, CDH5, EPHB1, JAK1, PTPRB, PTPRM, SLC12A6, SMAD1, STARD13, TJP1, UBP1 and VASH1. Moreover QKI and The miRNA processing ribonuclease DICER1, which are associated with angiogenesis, were high-ranking hub genes (Fig 5.11).

Hub genes were further enriched for processes relating to regulation of mitotic chromatid cohesion (Fig 5.10). This was due to the presence of RB1 (pRb), the pRb co-factor E2F3, ATRX, as well as CTNNB1 and CTNNBIP1, SMC1A and SLF2 [167]. SMC proteins are key regulators of DNA stability during mitosis and notably SMC1A, SMC2 and SMC6 were hub genes [309–312]. Moreover cytoskeleton-dependent kinesis and related terms were enriched and relevant hub genes included ACTR2, ROCK1/2, STMN1, MYH10, SEPTIN11, SON, BIRC6, KHLH9 and SVIL (Fig 5.10) [313]. Hub genes involved in regulating microtubules were also present, namely the high-ranking oncogene STMN1 but also several centrosome-associated genes were hub genes, including CENPO, CEP170 and CEP295 (Fig 5.11).

To conclude, host hub genes included candidates known to be involved in KSHV's life cycle and the KS. These included the broad enrichment of processes such as small GTPase signalling and downstream effects including angiogenesis, as well as mitosis-associated processes with a particular prevalence of SMC proteins. Moreover there was evidence for a central role of signalling via the ERK MAPK signalling pathway as evidenced by the presence of several upstream agonists (NRas and associated proteins), downstream transducers such as RPS6KA3 and a key downstream effector, ETS1.

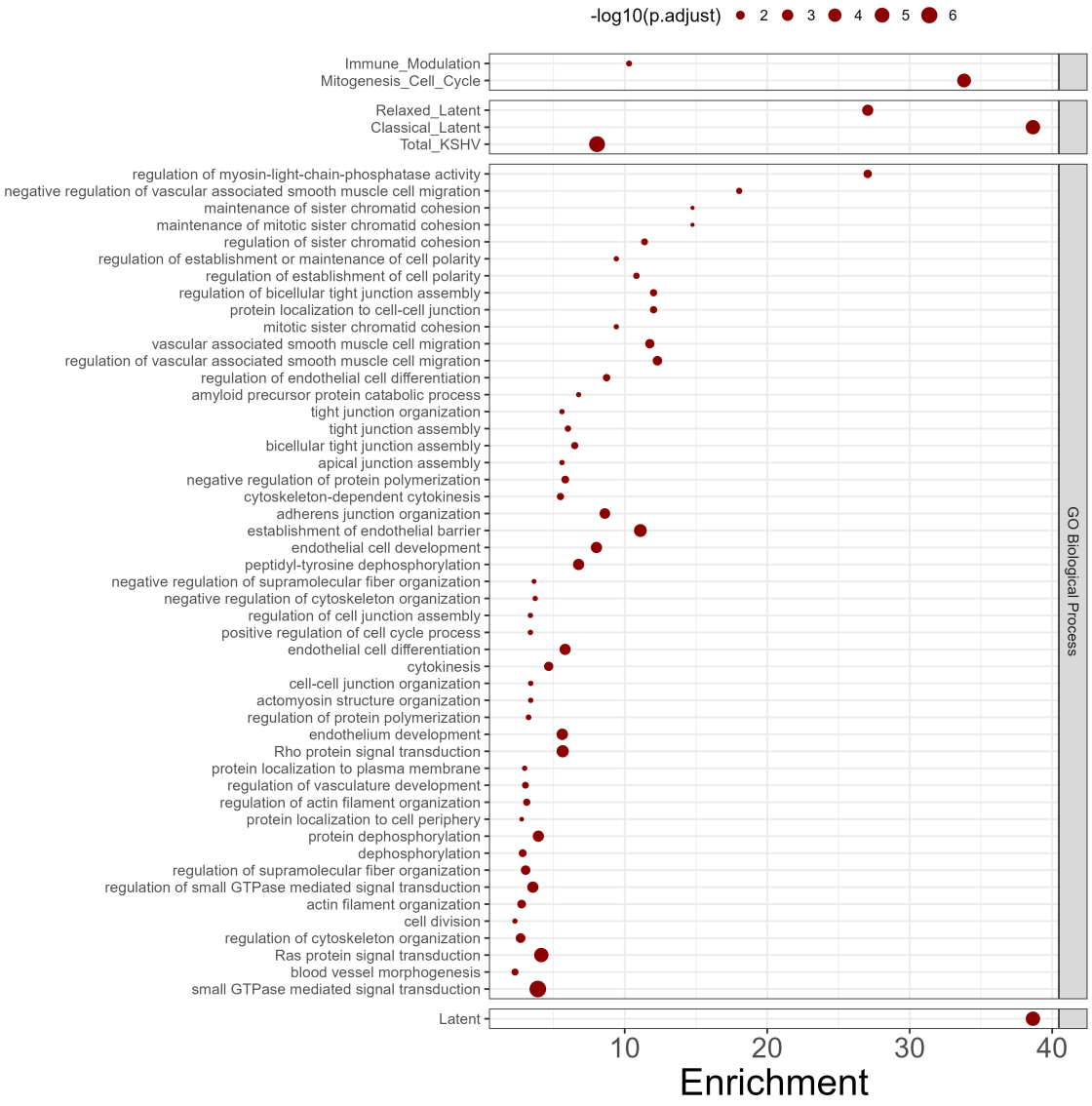
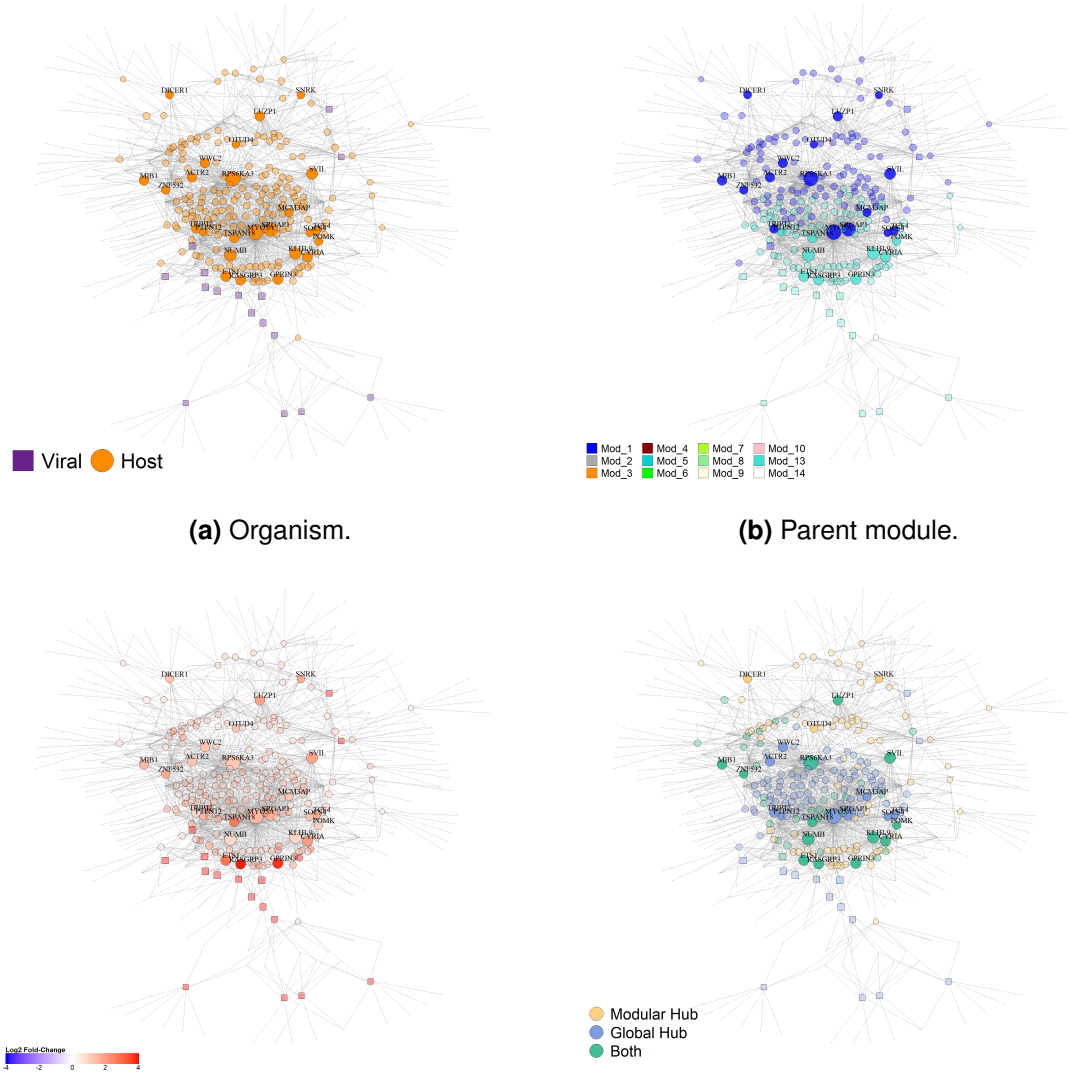


Figure 5.10: Enriched GO biological process ontological terms enriched in hub gene sets. Gene set over-representation analyses were performed on the 465 hub genes using a one-sided hypergeometric test to test for significant enrichment ($p<0.01$) of GO biological process terms as well as all custom host and viral gene sets. Note unlike module-wise ontological enrichments, terms have not been subject to simplification with rrevigo. Terms are split according to their corresponding source. Note that no terms were enriched for down-regulated hub genes.



(a) Organism.

(b) Parent module.

(c) Log2 fold-change.

(d) Hub type.

Figure 5.11: Hub gene network for 1st sub component of co-expression network. Hub genes were classified according to the criteria in Methods Section 2.3.8 and their top 10 correlated genes were extracted from the full network. This resulted in 2 major non-connected components and the first component is shown in this figure. The remaining genes were then plotted as a network, using the Kamada-Kawai layout algorithm. Gene hub centrality was then calculated using the influential package (as in Section 2.3.8). Non-hub neighbour genes are the smallest, grey coloured dots. Hub hub score was then log10-transformed and gene node sizes are scaled by this resultant transformed centrality measure. Additionally, the top 25 genes by hub centrality are labeled and their corresponding nodes opaque, while all other hub genes were set to transparent. Nodes are shaped according to whether they correspond to viral genes (square) or host genes (circle). Nodes are further coloured according to organism (a), their parent module (b), the log2 fold-change (log2FC) between lesion and control tissues (c) (described in Section 2.3.4) and (d) whether hubs were defined as intra-modular or global hub genes (described more in Section 2.3.8). Note that in (c), the log2FC of the viral genes were greater than 4 however including their log2FC in the colour scale overwhelmed the signal from the host genes.

5.8 Differential correlation network analysis proposes novel differentially correlated host hub genes associated with KS

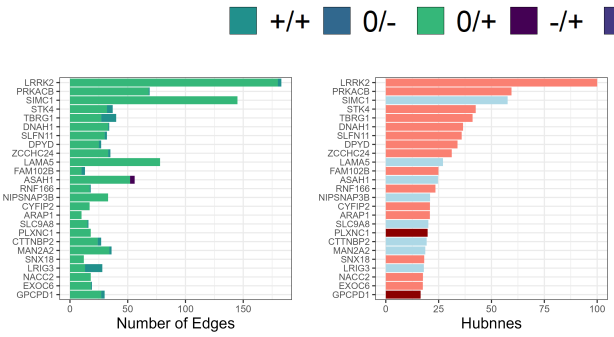
Co-expression can provide information on the importance of genes to the overall transcriptome of biological conditions, however it cannot distinguish between conditions. Differential co-expression network analyses can however identify genes who's changes co-expressions are associated with variations in the levels of a condition of interest. Given the paired lesion-control nature of the dataset, differential gene correlation analysis (DGCA) was performed.

While Mod 13 was of interest considering it contained all the classical and most the relaxed latent genes, those most associated with oncogenesis, the relatively higher expression of these genes in endemic samples means that it's importance may be somewhat over-emphasised by the inclusion of these samples. However considering Mod 1 comprises viral genes with less well-known associations to KS development, including the driver of lytic replication RTA, it was taken forward as the main "disease" module for further analysis.

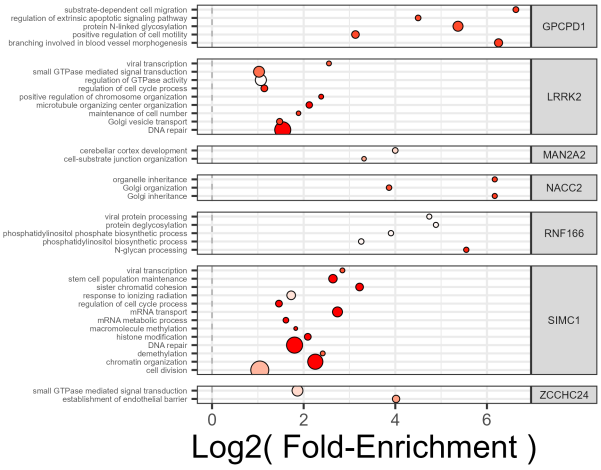
A differential correlation network was constructed for Mod 1 between lesion and control tissue, filtering edges by (FDR$<$0.001). Genes in this network were then ranked according to their hubness, retaining the top 25 genes, hereby termed "DGCA hub genes" (Fig 5.12). Differential edges could be grouped into different classes based on the transition of the edge (Table 2.8). Classes for all Mod 1 differential hub genes almost exclusively showed a shift from nothing to positive (0/+) or retainment of positive (+/+) (Fig 5.12a). Moreover, the log2FCs for all neighbours of these hub genes tended to be up-regulated, irregardless of edge class, indicating an activatory relationship between these hubs and their neighbours (Fig 5.12c). Both of these observations are indicative of net activation of these hub genes, in terms of their effect on their neighbours. Notable Mod 1 genes included LRRK2, which is most well-studied for the role of gain-of-function mutations in Parkinsons disease, but it also has roles in endosome trafficking, inflammation and infection [314, 315]. Neighbour genes of LRRK2 were enriched for small GTPase activity, Golgi vesicle transport, microtubule organising center (MTOC), cell cycle, chromosomes organisation and, interestingly, viral transcription (Fig 5.12b). An additional DGCA hub gene was SIMC1, which associates with SMC5/6 complexes to form known anti-viral effector complexes that, interestingly, have been suggested to be degraded by RTA (Fig 5.12a) [59, 316, 317]. Despite this, its neighbour genes were enriched for viral transcription, alongside chromatin organisation, histone modification and demethylation indicating a pro-gene expression role (Fig 5.12b). Neighbour genes were further enriched for mRNA

transport and metabolic processes, cell cycle processes, DNA repair and chromatid cohesion (Fig 5.12b). An additional notable Mod 1 differential hub gene was PRKACB, a subunit of the cAMP-activated Protein Kinase A (PKA) that regulates many aspects of cellular metabolism and promote lytic replication. However its neighbours were not enriched for any processes.

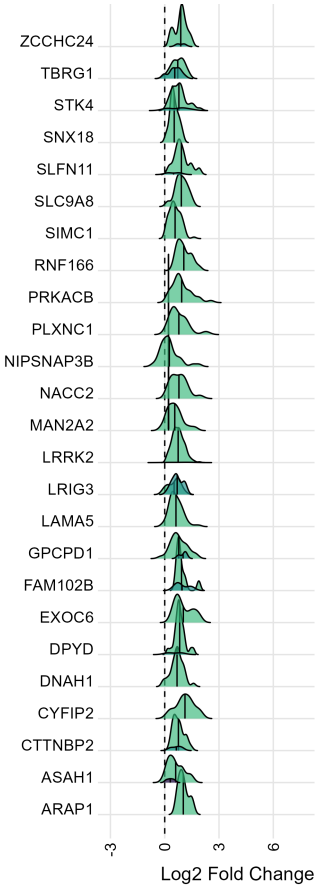
Altogether this section has identified Mod 1 hub genes that show generally activatory relationships with their neighbouring genes. Moreover DGCA hub-wise neighbouring genes were enriched for processes relevant to KSHV as well as KS and oncogenesis in general. We believe that these genes are ideal candidates for further study and their possible relevance to KS is detailed further in the discussion chapter.



(a) Degree of DGCA hubs.



(b) DGCA hub neighbour gene enrichments.



(c) DGCA neighbour gene log2FCs.

Figure 5.12: Differentially correlated hub genes. Differential gene correlation analysis (DGCA) was performed (as in Section 2.3.10) on Mod 1 genes, between lesion and control samples. Significant differentially correlated edges were retained (FDR$<$0.05) and hubness centrality measures computed as in (Section 2.3.8), retaining the top 25 genes by hubness as "DGCA hub genes". In (a), the left-most plots show the number of differentially correlated edges per hub are enumerated in, coloured according to DGCA edge class, while right-most plots indicate the ranked centrality of each genes, coloured by significant or non-significant up-regulation (dark red and light red, respectively), or significant or non-significant down-regulation (dark blue and light blue). (b) shows the Z-scaled relative expression of hub genes across samples, split by module eigengene (ME) clusters (Section 5.6. Heatmap row annotations represent module colours. Neighbours of DGCA hub genes were subject to gene set over-representation analyses on on their up-regulated sets and these are presented in (c), coloured by adjusted p-value. Log2FCs for DGCA hub neighbours, split and coloured by DGCA edge class, are shown in (d).

5.9 Chapter Discussion

In summary, the analysis in this chapter aimed to model the transcriptomes of KS lesion samples as a co-expression network and interrogate the resultant network model to gain insights into key genes that may drive KS. This involved partitioning of the network into modules, followed by ontological and expression analysis of partitioned modules. Subsequent analysis of co-expressed (WGCNA) hub genes suggested possible key signalling hubs and associated process important to lesion development. Finally, extension to differential co-expression network analyses, comparing lesion to paired control tissues, revealed possible key genes involved in the development of otherwise normal to lesion tissue. The subsequent section will aim to discuss these findings in context of the chapter results and methods as a whole.

Batch adjustment was performed which appeared to be successful in removing the influence of batches and maintaining the effect of endemic samples, despite complete confounding (Fig 5.1, 5.2 & 2.1). While this may still somewhat compromise the signal that delineates between endemic and epidemic samples, this was not of interest to this study as this difference was essentially a nuisance parameter. Still, we cannot entirely rule out the influence of endemic samples to multivariate analyses as, such as PCA (both transcriptome-wise and module-wise) and module assignment, particularly as we found Mod 13 to be significantly enriched for the uniquely differentially expressed endemic genes ($p<0.05$). However we did take steps to decrease their impact, including choosing a ranked measure (Spearman's ρ) of co-expression, which negates the greater magnitude of differential expression in endemic samples as well as choosing to focus on Mod 1 for DGCA analysis, the disease module least associated with endemic samples.

3 major modules contained viral genes, Mod 1, Mod 6 and Mod 13, however only Mod 1 and Mod 13 were positively associated with lesion development (Fig 5.5 & 5.9). Mod 6 on the other hand appeared refractory and its gene ontological enrichments were predominantly immune in nature, namely Th1 CD4 T cell responses, which may indicate an association between these viral genes and an appropriate immune response, which could limit lesion development (Fig 5.9 & 5.6). All Mod 6 constituent genes were lytic, which is known to be more immunogenic. Conversely while Mod 1 contained many lytic genes, it was strongly associated with lesion development (Fig 5.9). However these lytic genes were predominantly early lytic and included the key lytic activator gene, ORF50, and was associated with early lytic host functions, such as histone modifications and regulation of gene expression (Fig 5.9 & 5.6). Interestingly this could be evidence for the importance of an abortive lytic cycle in KS lesion development.

WGCNA hub genes were shown to be enriched for processes known to be important to the life cycle and pathology of KSHV as well as KS, such as ERK signalling and angiogenesis. Importantly, hub genes tended to be upstream regulators of these processes, including NRas and associated exchange proteins (RASGRF2/3 and RASSF10) and RPS6KA3, and downstream effectors including ETS1, β-catenin and TCF4. ETS1 and β-catenin both have previous well-defined influences to KS and KSHV, with ETS1 being previously predicted to be activated via Ingenuity Pathway Analysis (IPA) in Tso *et al.,* [16, 318]. The reason these genes and not ERK pathway components themselves were identified as hub genes likely relates to the transience of MAPK signalling, a greater coupling between the gene expression of transcription factors and their regulons as well as the importance of convergence of multiple signals via the upstream regulators [319]. Interestingly, the most high-ranking DGCA hub gene, LRRK2, has a MAPK kinase kinase (MAPKKK) domain that is known to activate downstream ERK signalling and this may implicate it in this signalling axes.

Additional notable processes significantly enriched in WGCNA hub genes were those related to mitosis and chromatid dynamics. This was in-part due to the presence of pRb and it's co-factor E2F3 but also the presence of SMC ATPase proteins, primarily SMC1A, SMC2 and SMC6, components of the SMC1/3 (cohesin), SMC2/4 (condensin) and SMC5/6 complexes, respectively, as well as SLF2, a component of SMC5/6 complexes. These complexes are critical to proper regulation, assembly and separation of chromatids during mitosis, as well as effective DNA repair mechanisms, and thus their dysregulation has obvious oncogenic potential [132, 310, 317]. Moreover, KSHV is known to exploit at least SMC1/3 complexes to manipulate its genome to favour conformations and modifications that promote latency, while SMC2 is known to be transcribed by a TCF4/β-catenin complex to promote oncogenic angiogenesis [50, 312, 320]. Perhaps most-interestingly however, SMC5/6 complexes have recently described broad anti-viral activity by super-compacting viral genomes, while many viral proteins target it for degradation [133, 317, 321–327]. Most interestingly however, SMC5/6 proteins have been shown to targeted to at least polyomavirus and adeno-associated virus (AAV) dsDNA genomes by SIMC1, one of the most high-ranking DGCA hub genes, to instigate their repression (Fig 5.12a) [317]. Therefore why SIMC1 would show positive relationships with lytic ORF50-centric Mod 1 genes and lesion development, alongside an enrichment of viral transcription in its neighbours is unclear (Fig 5.12b). The significance of this and the previous insights are discussed further in the proceeding Discussion chapter.

6 Discussion

KSHV is an oncogenic herpesvirus that extensively interacts with host cells, dysregulating almost all levels of gene expression in order to modulate host and viral processes. This includes at the transcriptional level (eg RTA transactivation etc), post-transcriptional level (eg SOX-mediated destabilisation of host transcripts) and post-translational level (eg RTA, K3 and K5 ubiquitinylation) [21, 60, 77]. One post-transcriptional layer of host-viral interactions that is emerging as important to KSHV is the modulation of gene expression by targeting circRNA-mediated ceRNA networks. Given that little is known about the dysregulation of circRNAs by KSHV, particularly during KSHV lytic replication, one main aim of this study was to identify differentially expressed circRNAs, miRNAs and mRNAs in order to model and analyse the importance of ceRNA network dysregulation during KSHV reactivation.

Host-viral interactions have pathogenic implications and contribute to KSHV's pathologies, which tend to be strongly associated with the replication and gene expression of the virus. For instance KS, a highly angiogenic and atypical malignancy that KSHV derives its name from, is believed to be predominantly driven by increased viral replication that enables the increased expression of viral oncogenes, which ultimately drive cellular transformation. However most work into this has been performed in already-transformed cell line cultures and as such doesn't immediately translate or scale to complex KS lesion tissue. Moreover, the influence of host factors remains poorly understood and the means to investigate them in an unbiased manner has been limited. This was until the recent publication of global RNA-Seq profiling datasets of KS lesions [94, 328**?**]. Therefore, another aim of this study was to interrogate one of these datasets by co-expression network modelling in order to identify key hub genes that may play a crucial role in the development of KS.

6.1 Relevance of network ncRNAs to viral infection and viral onocgenesis

The competing endogenous RNA network constructed assumed a directed structure in which circRNAs inhibit miRNAs, which in-turn inhibit mRNAs. Given that each miRNA can target 10s-100s of mRNAs and each circRNA can target several miRNAs, under this model circRNAs represent potent regulators of such ceRNA networks, with miRNAs being the primary "effector" intermediates. Centrality analyses facilitated the ranking of these circRNAs and miRNAs and the identification of candidate hub circRNAs and miRNAs.

In general, most circRNAs have been poorly characterised due to the recency of their dis-

covery, their ambiguity relative to linear products from their parental genes and the relative difficulties in studying them. Perhaps the most interesting component circRNA of the network was circHIPK3, which is one of the most extensively studied circRNAs with known pro-oncogenic roles (Fig 4.4a & c) [71, 202, 329]. Moreover it has previously been characterised by the Whitehouse lab as being up-regulated during lytic reactivation, where it sponged 2 network miRNAs (miR-30c-1-5p and miR-29b-1-5p) to promote lytic replication [71]. It's detection as up-regulated in this study helps to confirm this occurrence. Importantly, it has both pro-lytic and pro-oncogenic roles, meaning that it's up-regulation likely both promotes KSHV's replication and induces an oncogenic phenotype independent of KSHV, which, because KSHV replication drives most KS-associated cancers, indicates it has strong transforming potential in KS-associated disease [71, 202]. However circHIPK3's relatively low rankings in terms of centrality scores does indicate that, while certainly influential, other circRNAs have the potential to be even moreso (Fig 4.4a & c). These may have been may have been missed by previous studies, such as Harper *et al.,* 2022. One circRNA, circBAGE3, is potentially interesting as it was consistently the most high-ranking circRNA by all centrality metrics (Fig 4.4a & c). However to the best of our knowledge, this circRNA has not been previously studied. On the other hand, multiple isoforms of circARHGEF12, another high-ranking circRNA, have previously been shown to be differentially expressed between liver tissue from healthy patients and those with chronic HBV infection [330]. In contrast however, it was down-regulated during lytic replication (Fig 4.4, 4.5 & 4.1). One alternative isoform (hsa_circ_0024615) has been observed to be down-regulated in gastric cancer and it's role was attributed to predicted sponging of miR-134 and miR-590, de-repressing SMAD4 and CTNNA1/CTNNA2 [331]. Neither of these miRNAs were differentially expressed, but the involvement of CTNNA genes may indicate a role of this circRNA in angiogenesis, however neither CTNNA1 or CTNNA2 were themselves differentially expressed. Nonetheless, its involvement in cancer and in-particular a chronic virally-driven cancer may indicate a role of this circRNA in KSHV's oncogenesis.

While the known prevalence of viral ceRNA interactions at present is limited, some notable parallels can be drawn between the findings of this study and previous work. Notably miR-27a-5p was down-regulated in this study, a known anti-viral miRNA that inhibits at least adeno associated virus (AAV), HCV, HVS, MCMV and Respiratory syncytial virus (RSV) by targeting a range of pathways, including lipid metabolism, the cell cycle, immune and ER signalling [196, 197]. Of most relevance to this study, it is known to be suppressed by both HVS and MCMV-encoded miRNA-destabilising transcripts [196, 197]. Similarly, miR-17 92 cluster member miR-92a was down-regulated during KSHV reacti-

vation and other members of this cluster (miR-17 and miR-20) have been shown to be specifically targeted by the HCMV miRDE transcript [198]. Given that KSHV doesn't encode equivalent transcripts, dysregulation of native cellular ceRNA networks by modulating the abundance of circRNAs that target such miRNAs may facilitate the equivalent suppression of these miRNA by the virus. However these HVS, MCMV and HCMV-encoded transcripts specifically induce degradation of their target miRNAs and, while instigating miRNA sponge-like effects, aren't strictly ceRNAs. Whether miRNA-circRNA binding induces degradation of target miRNAs is still an area of active study.

6.1.1 The dysregulated ceRNA network is enriched for genes that promote viral gene expression and DNA replication

The network protein coding genes were enriched for transcription, gene expression and RNA biosynthetic processes (Fig 4.6a). Given that these were particularly enriched in the up-regulated set of host genes, this would suggest that one function of the ceRNA network is to promote such processes, potentially for the virus. The timing of the differential expression ($\sim$20hrs post induction of reactivation) is at the peak of early lytic gene expression as viral DNA replication has just initiated. This could imply two things depending on whether the changes represent an initiating cascade of gene expression or signify the end of a wave of expression. In the prior case, increased RNA biosynthesis and transcription may facilitate the expression of late viral genes which requires distinct transcription complexes different for early lytic gene expression and are encoded by early lytic genes, as well requiring viral DNA replication to have begun occurring [15]. In the second case, the ceRNA network could act early on in lytic replication to facilitate early lytic gene expression. However the limited differential expression of miRNAs at as late as 16hrs (see Chapter 3.2) indicates that this option is less likely than the first. Notably miR-29-1-5p's targets in-particular were enriched for "RNA metabolic process", "transcription, DNA-templated" and "RNA biosynthetic process" implying that the down-regulation this miRNA was in-part responsible for the up-regulation of these processes (Fig 4.8). In support of an effect of the ceRNA network on late lytic gene expression, Harper *et al.,* 2022 showed that miR-29b-1-5p over-expression specifically suppressed expression of the late lytic genes ORF65 and KSHV DNA abundance at 72hrs post induction. Moreover, the network as a whole was enriched for tertiary (maximally expressed 48-72hrs post induction) and capsid proteins, however it was also enriched for early lytic, secondary (maximally expressed 8-24hrs post induction) and viral genes involved in replication (Fig 4.6b).

A notable protein-coding gene that ranked highly by centrality metrics was HIPK2, which is involved in regulating gene expression. HIPK2 has previously been identified as a potential target of KSHV's miR-K12-4 via quick Cross-Linking, Ligation, and Sequencing of Hybrids (qCLASH) [332, 333]. HIKP2 has roles in regulating transcription by phosphorylating transcription factors and components of the transcriptional machinery (Fig 4.6) [281]. Moreover it can modulate chromatin states by interacting with PRC2 and various histone acetylases/de-acetylases with repressive and pro-transcription roles [334]. Thus de-repression of HIPK2 by the up-regulation and subsequent down-regulation of circRNAs and miRNAs during KSHV reactivation may promote KSHV lytic gene expression and replication by promoting transcription of lytic genes. However HIPK2 promotes cell cycle growth arrest and can induce apoptosis, largely by its phosphorylation of p53, promoting the latter's translocation to the nucleus to activate the expression of response genes [281]. However, paradoxically, p53 activity is required for efficient gammaherpesvirus lytic replication, while at least EBV is known to promote HIPK2's stabilisation to promote lytic replication [56]. Thus de-repression of HIPK2 may serve to promote KSHV lytic replication by promoting p53 activity.

Limited study into the dysregulation of metabolism during KSHV reactivation have been performed. Lagunoff *et al.,* have previously shown that glycolysis, glutaminolysis, and fatty acid synthesis are up-regulated, which is proposed to overlap with KSHV's dysregulation of hypoxia signalling [335]. Both glutaminolysis and glycolysis can contribute to nucleotide production [336]. However findings similar to the present study have been made during lytic replication of other gammaherpesviruses. For example, a metabolomics screen 24 hrs post lytic replication of the mouse gammaherpesvirus model MHV68 showed that 9 of the top 12 most significantly increased metabolites were those produced during nucleotide synthesis [337]. Moreover, cytidine metabolism in-particular has recently been shown to be important for the maintenance of EBV infected transformed B cells [338]. In addition, the alpha herpesvirus HSV1 dysregulates the TCA cycle to promote purine synthesis via shunting metabolites through the pentose phosphate pathway [339]. While these mechanisms are yet to be confirmed for KSHV, it does indicate that dysregulation of nucleotide metabolism is a key factor for herpesvirus lytic replication. In general, such modulation of cellular metabolism is believed to increase the abundance of nucleotides and ribonucleotides for viral genomic replication and gene expression, respectively. Targeting of circRNAs to dysregulate a ceRNA network that regulates it in order to facilitate such metabolic changes is, to the best of our knowledge, a novel finding.

As well as modulation host metabolic pathways, KSHV encodes several enzymes that

play a direct role in nucleotide metabolism. Interestingly, some of most high-ranking viral genes targeted by this network have roles in DNA metabolism and replication including thymidylate synthasee (ORF70), ribonucleoprotein reductase (ORF61) and uracil DNA glycosylase (ORF46) (Fig 4.4b & d) [21]. Given that the timing of the dysregulated network is roughly when viral DNA replication would begin and the network was enriched for viral genes involved in replication, a complementary function of the ceRNA network may be to promote viral lytic DNA replication. Otherwise given the interaction between nucleotide and ribonucleotide metabolism, increased ribonucleotides may be funneled into DNA replication to promote viral DNA replication, or *vice versa*. This indicates that dysregulation of the ceRNA network may also facilitate de-repression of viral genes as well as host genes involved in nucleotide metabolism.

6.2 A paradoxical involvement of SMC complexes in KS

Several SMC proteins were identified as WGCNA hub genes, indicating a central role in KS lesions (Fig 5.11). SMC proteins are highly conserved factors found in all eukaryotes and prokaryotes. They pair with other SMC proteins to form heterodimer ATPases (SMC1/3, SMC2/4, SMC5/6), six of which comprise hexamer rings in-complex with other proteins that act as adaptor or regulatory factors [310]. SMC1/3, SMC2/4 and SMC5/6 complexes are known as cohesin, condensin and SMC5/6 complexes, respectively [310]. In all organisms, they are essential for proper genome replication and chromosome segregation during mitosis of both cancerous and non-cancerous cells. The presence of members of each complex, SMC1A, SMC2 and SMC6, as well as SLF2, an SMC5/6 complex DNA adaptor protein, as WGCNA hub genes implicates these proteins as having a considerable effect on gene expression within KS lesions.

KSHV is known to dysregulate cohesin and SMC5/6 complexes, indicating a link between KSHV's activity and the perturbation of these genes in KS. Cohesin complexes are exploited to regulate the chromatin structure of KSHV's episome, to regulate the latent-to-lytic switch [50, 320, 340, 341]. On the other hand, SMC5/6 complexes have recently described roles in inhibiting viral replication and/or transcription, suppressing at least KSHV, EBV, HBV, HPV, EBV, AAV, polyomavirus and HIV-1, an effect which is proposed to be via inducing the compaction of target genomes [133, 317, 321–327]. Moreover at least KSHV, EBV, HBV, HIV and AAV encode proteins (RTA, BNRF1, HBX, VPR and E4, respectively) that specifically inhibit SMC5/6 complexes, indicating that they are important targets of viral dysregulation [133, 321, 323, 325, 342]. For KSHV RTA and EBV BNRF1, degradation of SMC5/6 complexes specifically enable lytic replication, with

KSHV RTA and EBV BNRF1 degrading them via ubiquitin-targeted proteasomal degradation [133, 323]. Interestingly, the Mod 1 DGCA hub genes, SIMC1, is a paralog of the SMC5/6 adaptor protein SLF1 that complexes with SMC5/6-SLF2, which has been shown to facilitate targeting of SMC5/6 complexes to polyomavirus and AAV genomes, restricting their replication (Fig 5.12a) [316, 317]. Moreover, SLF2 but not SLF1 has previously been shown to be essential for the localisation to and transcriptional silencing of SMC5/6 to non-integrated HIV-1 DNA [324]. This research has highlighted a new mechanism of viral restriction by host cells as well as a novel interface between viruses and host cellular genome replication and maintenance machinery. Importantly it indicates that such mechanisms occur in KS lesion tissue and that the factors involved in them are central to the process of lesion development.

Given the growing body of work indicating a viral restrictive role for SMC5/6-SIMC1 complexes, it seems paradoxical that SIMC1 should be associated with lesion development given the strong association between increased viral replication and this process (Fig 5.12). In fact, SIMC1's DGCA network neighbours were predominantly up-regulated and enriched for "viral transcription" (Fig 5.12c). Moreover, the presence of SIMC1 in Mod 1 is antithetical to previous work due to it's observed inhibition of KSHV lytic replication and RTA's degradation of it's protein product [323]. This all indicates a positive association with SIMC1 expression, lesion development and viral gene expression.

One clue to resolve this may come from the detection of this association in a cancerous lesion. It could be that SIMC1 targets SMC5/6 to KSHV episomes but the protein complex is subsequently degraded by RTA, abrogating its anti-viral impact. If the cell continues to transcribe and translate SIMC1, SMC5/6 and SLF2 mRNA, new complexes may be continually diverted from native DNA repair processes in a failed attempt to repress KSHV gene expression and replication. This could be particularly impactful in the context of KS, where massive up-regulation of viral replication is associated with lesion development. Such diversion of SMC5/6 complexes may promote genomic instability, driving lesion development. Indeed, SIMC1's DGCA neighbours were enriched for processes related to this, primarily "DNA repair", "response to ionizing radiation" and "sister chromatid cohesion". Moreover, SMC5/6 has important roles in genome stability and its deletion has been shown to result in increased nuclear bridges, micronuclei, mitotic catastrophes, all associated with aberrant mitosis, DNA damage and cellular transformation [132, 309]. Such aberrations have previously been observed in KSHV-transformed cells and in KS spindle cells [122, 132, 343]. Further support for this mechanism comes from the observation that SIMC1 has been shown to compete with SLF1 to complex with SMC5/6

complexes, indicating a competitive nature between SMC5/6 complex targeting to native host cellular and viral genomes [317]. Importantly, such an occurrence could explain why SIMC1 is positively correlated with lytic gene expression at the mRNA level. In fact, Han *et al.,* observed a significant increase in SMC5 mRNA abundance upon lytic induction and ectopic ORF50 expression, while SMC6 was increased at the mRNA level in lesion tissue in the current study (log2FC=1.03) [323]. Similar mechanisms may apply to other oncogenic viruses that are known to dysregulate SMC5/6, such as HBV, EBV and AAV and may underpin some of their oncogenic potential. However a major issue with this model is that it does assume that KSHV is incredibly efficient at de-repressing SMC5/6-mediated silencing and so much so that increased SIMC1 mRNA abundance doesn't increasingly suppress viral replication, which could be expected to result in a negative correlation. Moreover, SMC5/6 complex impairment is associated with p53 signalling and cell cycle arrest, however KSHV is known to co-opt and dampen these responses, respectively, to promote lytic replication [344].

Another explanation may be that SMC5/6 complexes promote some form of abortive lytic replication as seen for EBV [133]. Specifically, Yiu *et al.,* observed that deletion of the SMC5/6 destabilising protein BNRF1 resulted in repression of replication compartment formation, late lytic gene expression and infectious virion production. While this could be expected to result in a negative correlation with classically late lytic genes, BNRF1 is itself a late lytic gene and Yiu *et al.,* observed an up-regulation of some EBV's late lytic and many early lytic genes upon suppression of BNRF1 [133]. Moreover it may explain why some viral genes were observed to correlate negatively with PC1_paired and PC1_lesion (Fig 5.9 & S9). In this model, oncogenesis is likely driven by the facilitation of the expression of viral oncogenes during such abortive lytic gene expression programs. Other mechanisms are possible too. For example, secreted factors produced by cells undergoing lytic replication may induce SIMC1's expression as a restriction factor in neighbouring cells that either don't contain the virus or that drive the virus into latency. However this would presumably inhibit further viral infection thus at the tissue level, leading to reduced viral gene expression and thus a negative correlation between SIMC1 and viral gene expression. Another possibility is that SIMC1 has alternative functions that may be independent of SMC5/6 that are oncogenic in some regard. Finally, SIMC1 may function to repress HIV-1 into latency, which may have pro-oncogenic effects.

Overall these models are all highly speculative, but the positive relationship between a viral restriction factor, viral gene expression and preogression of a virally-caused disease in nontheless highly interesting and could indicate a novel mechanism for how a viral

restriction factor is perturbed to promote a viral-driven cancer.

6.3 A role for LRRK2 in KS

LRRK2 was identified as the most high-ranking DGCA Mod 1 hub gene Fig 5.12b. It is a well-studied kinase predominantly known for it's role in Parkinson's Disease (PD); primarily the G2019S gain-of-function mutation in its MAPK kinase kinase (MAPKKK) domain that predisposes towards it [314, 315]. There are some interesting parallels between the occurrence of PD and KS. Primarily, G2019S is found in 20-40% of individuals of the Ashkenazi Jewish demographic, a population that also shows a preponderance for classic KS (further detailed in Table 6.1a) [315, 345, 346]. This could indicate a link between the catalytic activity of LRRK2 and susceptibility to developing KS. In addition, PD shows a 2-fold increased prevalence in men than women, something observed in KS, while the G2019S mutation shows differential impacts on NFAT signalling between male and female mice, a pathway that promotes KSHV's lytic replication [347].

Further studies have provided additional pathogenic roles for LRRK2, notably in chronic inflammation which contributes to PD as well as leprosy [315, 348]. In fact, like PD KS shows some interesting similarities to leprosy which may further implicate LRRK2's involvement in KS (see Table 6.1b for relevant points). Moreover another chronic inflammatory-mediated pathology, Crohns disease, is more prevalent in Ashkenazi Jews and is believed to be associated with LRRK2 [315]. Finally, LRRK2 is known to activate NF-kB and the NLRC4 inflammasome, as well as being suggested to act as a PRR like NOD2 [315, 347, 349–353]. These observations specifically implicate LRRK2 as an important factor in driving complex, chronic immune-driven disorders as well as at least one that is largely driven by chronic infection of an intracellular parasite.

One thing to note however is that LRRK2's neighbour genes were not enriched for directly immune-related processes (Fig 5.12c). Instead enrichments included "positive regulation of chromosome organisation", "regulation of cell cycle process" and "DNA repair" among others, implying against a role of LRRK2 in immunomodulation in KS (Fig 5.12b). Instead, these enrichments suggest a role of its MAPKKK activity in promoting these processes via downstream promotion of MAPK activity. Indeed LRRK2 is known to interact with the MAPK pathway components MKK3, 6, 7 and JIP4 as well as being implicated in activating downstream ERK signalling to promote PD [360]. This may indicate it's role in driving the downstream ERK factors associated with KS, identified in Chapter 5.7, such as such as ETS1. Additionally, it is also known to drive Wnt/β-catenin signalling and one of its differentially correlated neighbours was TCF4, a WGCNA hub gene. Importantly,

(a) KS and Parkinson's disease

- The G2019S gain-of-function mutation is the most common causal variant in Parkinsons disease and is over-represented in the Ashkenazi Jewish population. This population also shows increased susceptibility to KS as well, namely non AIDS-associated classical KS [315, 345, 346].
- Chronic inflammation is a key contributor to neurodegeneration and LRR K2 contributes to this, activating NFkB, the NLRC4 inflammasome and NFATC, all pathways considered important to the development of KS [348].
- PD patients exhibit raised serum inflammatory cytokines (IL-1B, IL-6, IL-8, TNFa, IFNy, MIP-1 and MCP-1) common to KS [315].

(b) KS and Leprosy

- KS and leprosy show clinical similarities such as forming vascularised lesions that tend to occur on older men, on the lower trunk, legs and feet and lesions initially retain sense that is lost with progression [354].
- Both leprosy and KS are driven by complex and innappropriate immune reactions to a replicating intracellular parasite [355].
- Leprosy is associated with increased systemic inflammatory cytokine levels like KS (TNFα, IFNγ and IL1β/6/8/10) [356].
- De-vascularisation of both KS and leprosy is associated with regression [357, 358].
- Like KS, leprosy been linked to the development of carcinomas in-part due to increased pathogenic load; a feature similar to KS [355].
- Both KS and leprosy can be treated with thalidomide and its derivative pamolimide, which promote CD4+/CD8+ co-stimulation and via inducing the degradation of ikaros (IKZF1) and aiolos (IKZF3), which modulate IL-6, VEGF and TNFα [359].

Table 6.1: Similarities between KS, Parkinson's disease (a) and leprosy (b), with respect to the involvement of LRRK2.

TCF4/β-catenin signalling is known to drive SMC2 expression, which may link LRRK2 to SMC dynamics, which as discussed in the previous section, may contribute to lesion development [312]. LRRK2 has also been shown to be phosphorylated by ATM in response to DNA damage, which up-regulated p53 and p21 expression to inactivate MDM2 via phosphorylation, surprisingly promoting cell cycle progression through the G1/S phase checkpoint and cellular proliferation [312]. Interestingly, there have been some proposed associations between PD and DNA damage, with LRRK2 and ATM implicated, while LRRK2 gain-of-function mutations have additionally been linked to mtDNA damage and dysfunction [361, 362]. Given that p53 activity is increased and required for lytic replication, this mechanism may underpin some of the cancer-like effects of KSHV lytic replication [177].

LRRK2 also regulates many aspects of vesicle transport and accordingly, the neighbours of LRRK2 were also enriched for GTPase activity-related terms and "Golgi vesicle transport" (Fig 5.12b) [363]. An aspect of this is its capacity to regulate the actin and microtubule cytoskeletions and the latter seems most important as "organisation of microtubule organising center" was also enriched in its neighbours and it is known to interact

with PAKs, of which PAK2 and 6 were both WGNCA hub genes [363]. Indeed it contains small GTPase activity and it is associated with various GAPs and GEFs as well as Rac1 and Rab5/7 [315, 348]. Interestingly it regulates autophagy and phagocytosis which, unlike other herpesviruses (EBV, HCMV and HSV-1), increases upon induction of lytic replication [364]. This may be because it provides catabolites via degradation of cellular macromlecules, limits antigen presentation to the cell surface or lysosome-resident PRR detection and prevents maturing virions targeting to lysosomes. Moreover LRRK2 activity has been shown to inhibit late stage phagosome/autophagosome fusion with lysosomes which may be of benefit to newly infecting viruses for for viral exit [348]. This may also promote cancer development by inhibiting the process of autophagy which in-turn may inhibit apoptotsis, thus contributing to the neoplastic nature of KSHV-infected cells [365].

LRRK2 has never been conclusively linked to KSHV or its pathogenesis, however a recent machine-learning kinase inhibitor screen did predict LRRK2 as regulating KSHV reactivation [366]. It is worth noting that the authors were unable to validate this prediction via siRNA-targeting of LRRK2, but they suggested that this may be due to the essentiality of LRRK2 to cellular survival resulting in poor knockdown efficiencies [366]. In the context of the present study's findings, LRRK2 would appear to have some association with lytic replication given its presence in Mod 1 and, importantly, an association with the pathogenesis of KSHV, specifically KS. One model could be that it's MAPKKK domain may promote lytic replication as well as having mitogenic and angiogenic effects, which directly contribute to viral replication and initial lesion development, respectively, ultimately driving lesion progression. This would provide a mechanistic link between the overactive MAPKKK domain G2019S mutation, viral replication and lesion development. Such a model is highly speculative however, given the complexity of lesion tissue and the factors that may influence its development.

While it is difficult to ascertain which if any of these mechanistic roles of LRRK2 contribute to KS, it is nonetheless highly interesting that so many show relevance to mechanisms that are associated with KSHV's pathogenesis. Moreover its high ranking by hubness and presence in Mod 1 implicates it as a key factor associated with shift from non-lesion to KS tissue, associated with lytic replication.

6.4 A discussion on the datasets used in this study, their limitations and improvements to the overall study design

Some key problems with the datasets used for this study require considerations. For Chapters 3 and 4, the main issues were two-fold: 1., a violation of a key distributional as-

sumption of most RNA-Seq library normalisation with a confounding source (SOX cleavage of host transcripts and RNA composition bias by increased viral reads) and 2., a lack of replication for the 0hr latent time point for circRNA microarray. We feel that the first point is primarily a problem of analysis that we have largely mitigated and we reiterate that it is not an issue that we have encountered being explicitly accounted for in any studies using differential expression applied to KSHV. The second has obvious consequences as it provides poor estimates of circRNA expression means and variances and thus greatly limits the confidence in the circRNA candidates presented. An obvious solution would be to generate a further biological replicate pair of 0 and 20hr lytic induction experiments. However given that the study of circRNA is still in relative infancy, especially applied to KSHV, circRNA-Seq-based approach (NGS sequencing of RNase R-treated RNA-Seq samples) would likely be more appropriate, as further quality-control and analytical steps can be performed on circRNA-Seq data, alongside enabling the detection of both novel viral and host circRNAs [278]. Additionally, a recent improvement on traditional circRNA-Seq is CircRNA identification using A-tailing RNase R approach and Pseudo-reference alignment (CARP), which would improve confidence in the detected circRNAs [278]. This adds an "A-tailing" step to the standard circRNA enrichment protocol, whereby RNase R-treated samples are subject to polyadenylation and polyA tail depletion, from which circRNAs are quantified and their relative abundances compared to untreated samples, in order to identify "RNase R-sensitive" reads that indicate false positive circRNAs, thus facilitating their elimination [278]. Such robust circRNA detection as well as a pseudo-mapping approach employed also facilitates greater confidence in assigning circRNA identity and length, limiting the issue with circRNA ambiguity previously discussed in Chapter 4.6. This would greatly improve the confidence in any analyses performed, inferences applied and conclusions gained to them.

For Chapter 5, the issue with the data-set was due to confounding of all endemic patients in one batch. We speculate that this was likely due to the practicalities of accumulating a dataset for a relatively infrequent disease in a clinical setting. While we tailored our analysis to minimise the influence of this, it confers ambiguity to the data, both in terms of the difference between KS forms but also the presence/absence of HIV-1. An obvious solution is to re-sequence the samples with a more balanced batch design. Alternatively, different data-sets could be utilised and at least two, Rose *et al.,* and Ramaswami *et al.,* could have been analysed with or in-parallel to the Lidenge *et al.,* dataset [18, 94]. Both data-sets comprise only epidemic KS samples, while Rose *et al.,* is the largest, comprising 41 KS samples sequenced to a greater read depth but no matched controls, whereas Ramaswami does comprise control samples but is split between epidermal (10)

and gastrointestinal (12) KS samples, sequenced in unknown batches and largely unknown experimental designs. However at the time of this study we were limited to using the Lidenge *et al.,* dataset because the Rose *et al.,* samples in the GEO repository only comprised viral-mapped reads and we were unable to acquire the full data from the study authors, while the Ramaswami *et al.,* dataset is yet to be made public at the date of writing. We hope that these data and further data are made available quickly to enable more-intergrative analyses or validation studies to be performed.

6.5 Conclusions and further perspectives

The first half of this thesis facilitated the investigation of dysregulated ceRNA networks between the latent and lytic replication states of KSHV. While the unbiased approach facilitated a global perspective on circRNA/miRNA/mRNA ceRNA network dysregulation, further work could specifically investigate individual circRNAs and their associated miRNAs and wider "effectome". Much of this has been performed for circHIPK3 and miR-30b, miR-30c and miR-29b in Harper *et al.,* 2022. However validation of the expression of the other circRNAs, particularly more highly-ranking ones such as circARGHEF12 and circBAGE3 could provide further insight. Additionally, regulatory interactions between miRNAs and their predicted mRNA and circRNA interactions could be established via the use of miRNA mimics and/or cross-linking or co-immunoprecipitation studies.

This study was entirely based off of binary comparisons between static time points from samples that were not matched. As such is limited in its capacity to detect gradual or non-monotomnic changes in gene expression over time which, given circRNA's relatively slow expression but prolonged stability, may be a critical factor. As such circRNA, miRNA and mRNA expression could be measured over time to better characterise their expression profiles and relate changes in their abundance to specific stages of lytic replication.

The second half of this thesis has focused on modelling and investigating co-expression networks derived from KS lesion tissue. Subsequent investigation of the network model via centrality analyses enabled the identification of sets of candidate hub genes with proposed roles in KS (Fig ???5.11). Such hub genes were enriched for angiogenic processes and those related to ERK signalling, chromatid dynamics and mitosis. However such analysis was relatively localised to phenomenological relationships between isolated genes. Instead more integrative differential pathway analysis or subnetwork enrichment approaches could provide evidence for specific dysregulations of groups of interacting of genes, for example those associated with Ras, β-catenin or ERK signalling. This would help confirm not just the involvement but the differential activity of these processes

in KS lesions, relative to un-transformed tissue.

A differential correlation network was constructed for Mod 1 and "DGCA" hub genes identified, with LRRK2 and SIMC1 ranked particularly highly (Fig ???5.12). When contextualised with WGCNA hub gene enriched processes, ERK signalling and chromatid dynamics and mitosis, this highlighted their involvement in KS as key determinants in driving the development of lesions. However such work via solely bioinformatic means cannot confirm the involvement of genes in processes, only propose and rank candidates for further study. Therefore an obvious necessary follow-up would be to perform biochemical validation of proposed candidate genes in cell lines or models of KS. Simple perturbation experiments in cell lines would indicate the importance of the proposed novel factors. However given the complexity of the proposed models of SIMC1's and LRRK2's contribution to KS development as well as the latter's known importance to cell viability, study of such influences would be challenging [366]. To study the the SMC5/6 "depletion" model *in vitro*, measures of DNA damage and genomic instability could be assessed at increasing levels of exogenous RTA expression under a dose-dependent promoter, while measuring SIMC1 at the protein- and transcript-level. This could be compared to a cell-line whereby RTA's capacity to induce SIMC1 degradation is abrogated. Similarly, to preclude any auxiliary effect of RTA, RTA's degradation of SIMC1 could be emulated via somatic-line editing of the SIMC1 gene via CRIPSR or other means to generate endogenous recombinant SIMC1 with a hyper-active ubiquitinylation (or other degradory) signal present to target SIMC1 for proteasomal degradation. This Additionally, as a continuation of the present study, Mutec could be applied to quantify mutational load from the RNA-Seq data and used as a response variable in machine learning models to determine which hub genes contribute most to DNA damage. This could be related to protein-level measurements of SIMC1 activity and contrasted to the relationships at the transcript-level.

The study of LRRK2 by knockdown and especially knockout would likely be difficult due to confounding with decreased cell viability. A weak knockdown approach could be employed, normalising by viable cell count to attempt to limit such confounding. Alternatively, to establish an interaction between KSHV gene products and LRRK2, unbiased viral protein-LRRK2 proteomics could be performed using LRRK2 as bait in a pulldown followed by mass spectrometry. Aside from biochemical studies, given the possible association between LRRK2's G2019S mutation, KS and PD as well as the prevalence of the latter two in the Azshkenazi Jewish population studies, it would be interesting to identify a link between the co-occurence of KS and this mutation and/or PD, in this and similar

genetic backgrounds.

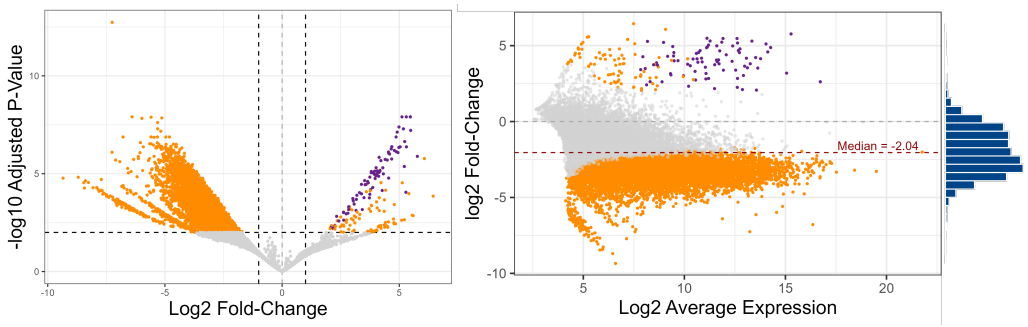
Finally, although co-expression network analysis has been successfully applied to complex tissues in the past, such studies are always limited by their cellular mileau. Many of the proposed genes, such as LRRK2, are known to be particularly associated with immune cells of which previous work indicates extensively infiltrate lesions. Therefore one aspect of experimental follow-up studies could involve determining the importance (or even presence) of these genes in various models of such cell types. Primarily, scRNA-Seq performed on KS lesions would greatly increase understanding as to how the candidate genes proposed in this thesis relate to the development of KS.

On-the-whole, this study aimed to model the transcriptomes of KSHV-infected cells and KS lesions as networks, followed by subsequent interrogation of such networks to propose novel factors that may contribute to the pathogenesis of KSHV and progression of KS. Although such work is largely exploratory in nature, the sets of candidates identified provide a resource that can be used by researchers performing further work into the pathogenesis of KSHV.

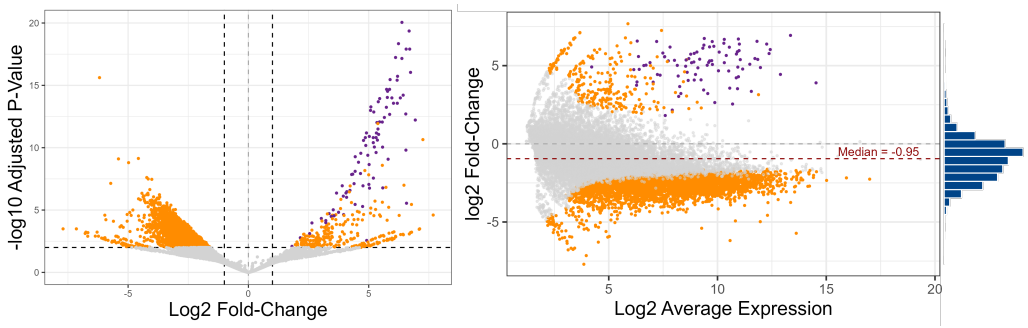
7 Appendix

7.1 Lytic competing endogenous circular RNA network

7.1.1 Differential expression analyses performed with additional methods for scale factor derivation



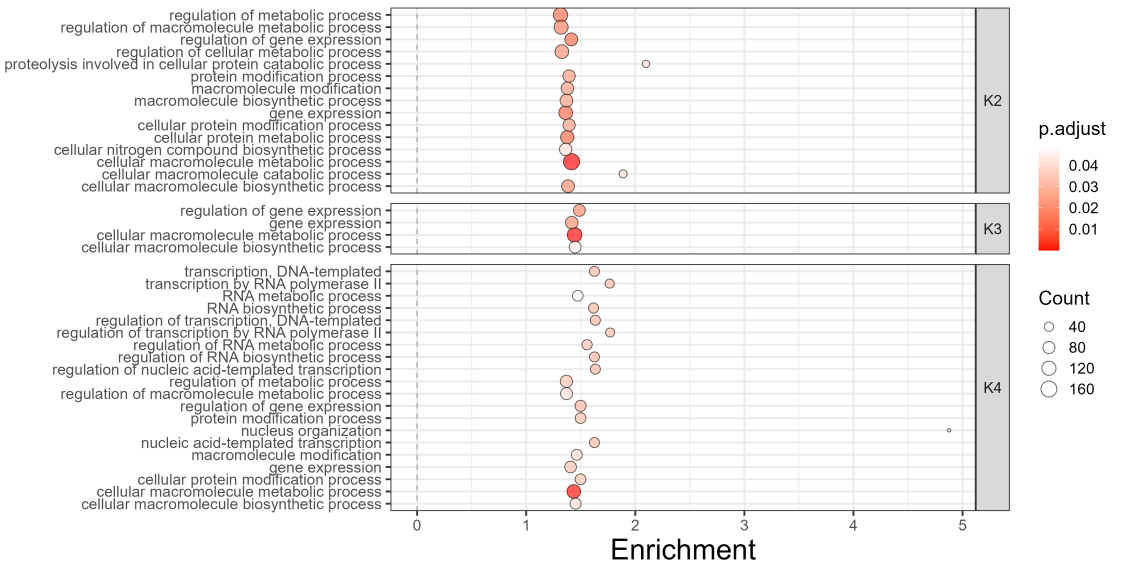
(a) CPM-derived scaling factors.



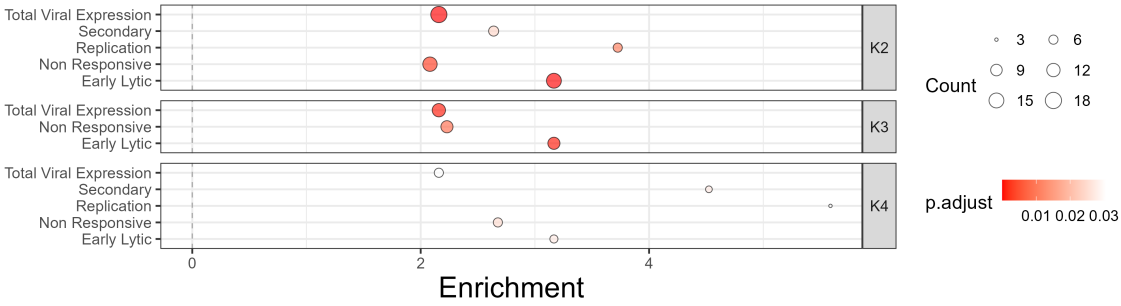
(b) EdgeR TMM-derived scaling factors.

Figure S1: Differential expression between latent and lytic infected cells usign different normalisation methods. Differential expression using DESeq2 was performed between 0 (latency) and 24 hours post Dox-mediated induction of TREx-RTA-BCBL1 cells using CPM- (a) and edgeR's TMM-derived scaling factors. Volcano plots showing -log10 adjusted p-value and log2FC are shown in leftmost plots while MA plots showing log2 average expression (log2AE) against log2FC. Genes/s are coloured according to whether they were not significantly differentially expressed (FDR$\geq$0.05) (grey) or were significantly differentially expressed host (orange) or viral (purple) genes. Red lines in MA plots indicate the median log2FC for all genes, regardless of significance.

7.1.2 Gene ontology analysis on K-core decomposed network protein coding genes



(a) Host GO BP gene set enrichments.

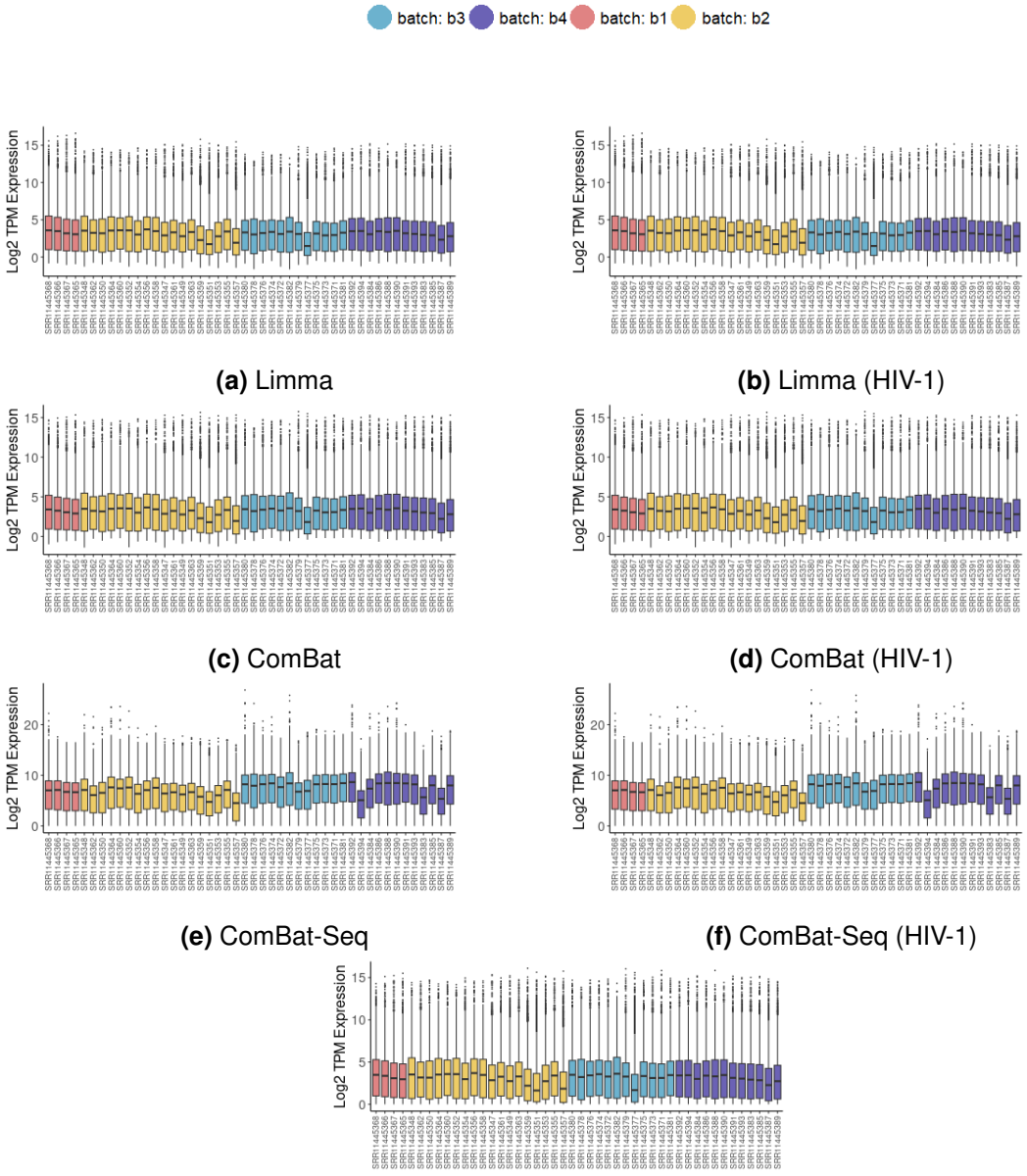


(b) Viral gene set enrichments.

Figure S2: Significantly enriched gene ontology (GO) biological process (BP) (a) and viral gene (b) sets in K-core decomposed network genes. Over-representation analyses (ORA) were performed on all network protein-coding and viral genes for K-core decomposed networks (K=2,3,4). Enrichment results are split according to K. Results are filtered by FDR$<$0.05 via one-sided Fisher's exact tests. Dot sizes indicate the number of genes in each gene set that are annotated for that gene set and coloured by FDR-adjusted p-values.

7.2 Kaposi sarcoma biopsy RNA-Seq co-expression network analysis

7.2.1 KS biopsy sample expression distributions



(a) Limma

(b) Limma (HIV-1)

(c) ComBat

(d) ComBat (HIV-1)

(e) ComBat-Seq

(f) ComBat-Seq (HIV-1)

(g) Un-Adjusted

Figure S3: Sample-wise expression distributions of various batch adjustment methods. The log2 TPM expression of all genes in each sample are presented for various batch adjustment schemas (a-f) as well as un-adjusted data (g). Individual subfigures show data adjusted with Limma's removeBatchEffect() (a & b), ComBat (c & d) and ComBat-Seq (e & f). Left-most plots (a, c & e) show expression data without the inclusion of HIV-1 co-infection (epidemic/endemic) labels as a covariate during batch adjustment, while right-most plots (b, d & f) show expression data with the inclusion of this covariate during batch adjustment. Samples are coloured according to the batch that they were sequenced in.

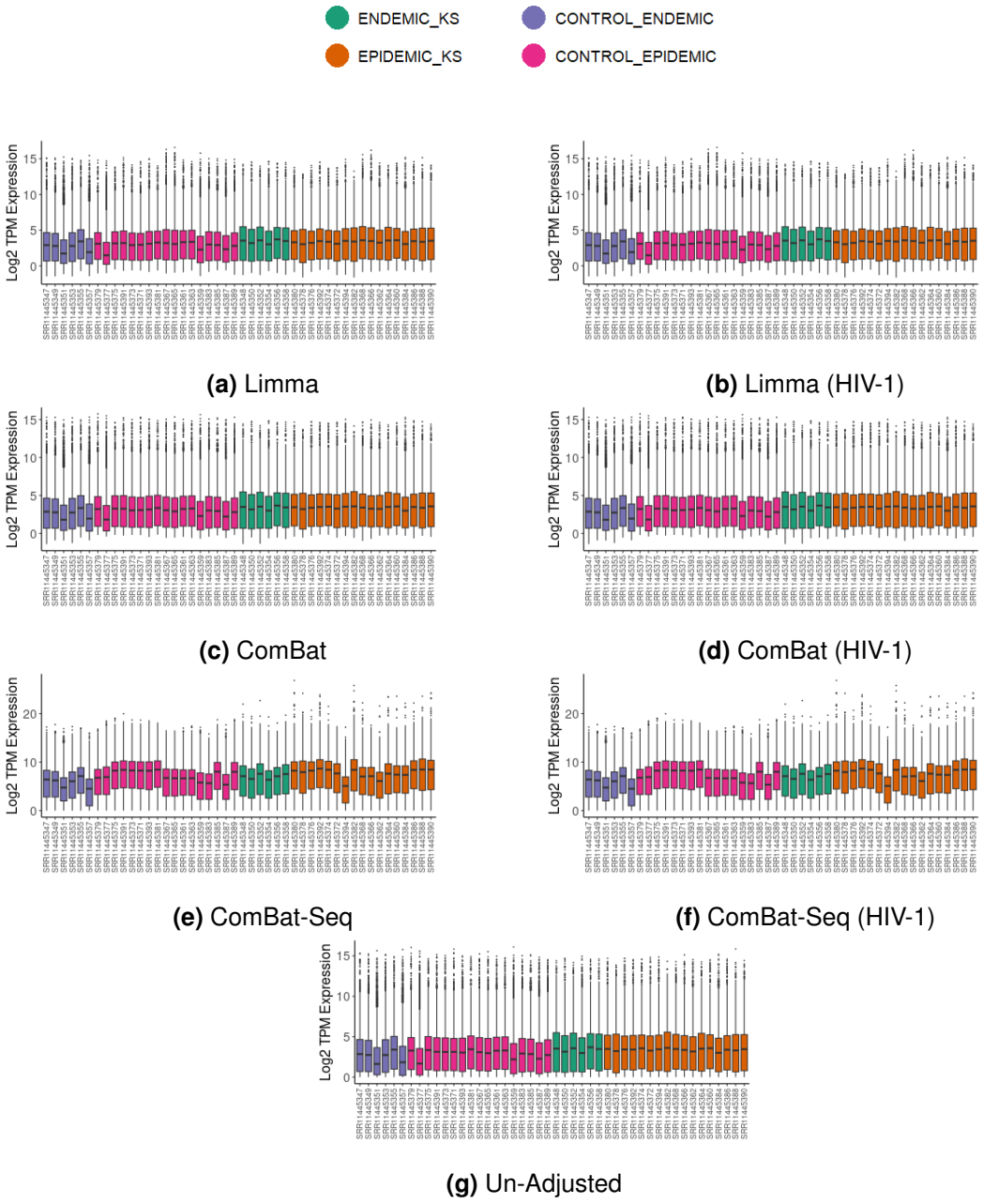
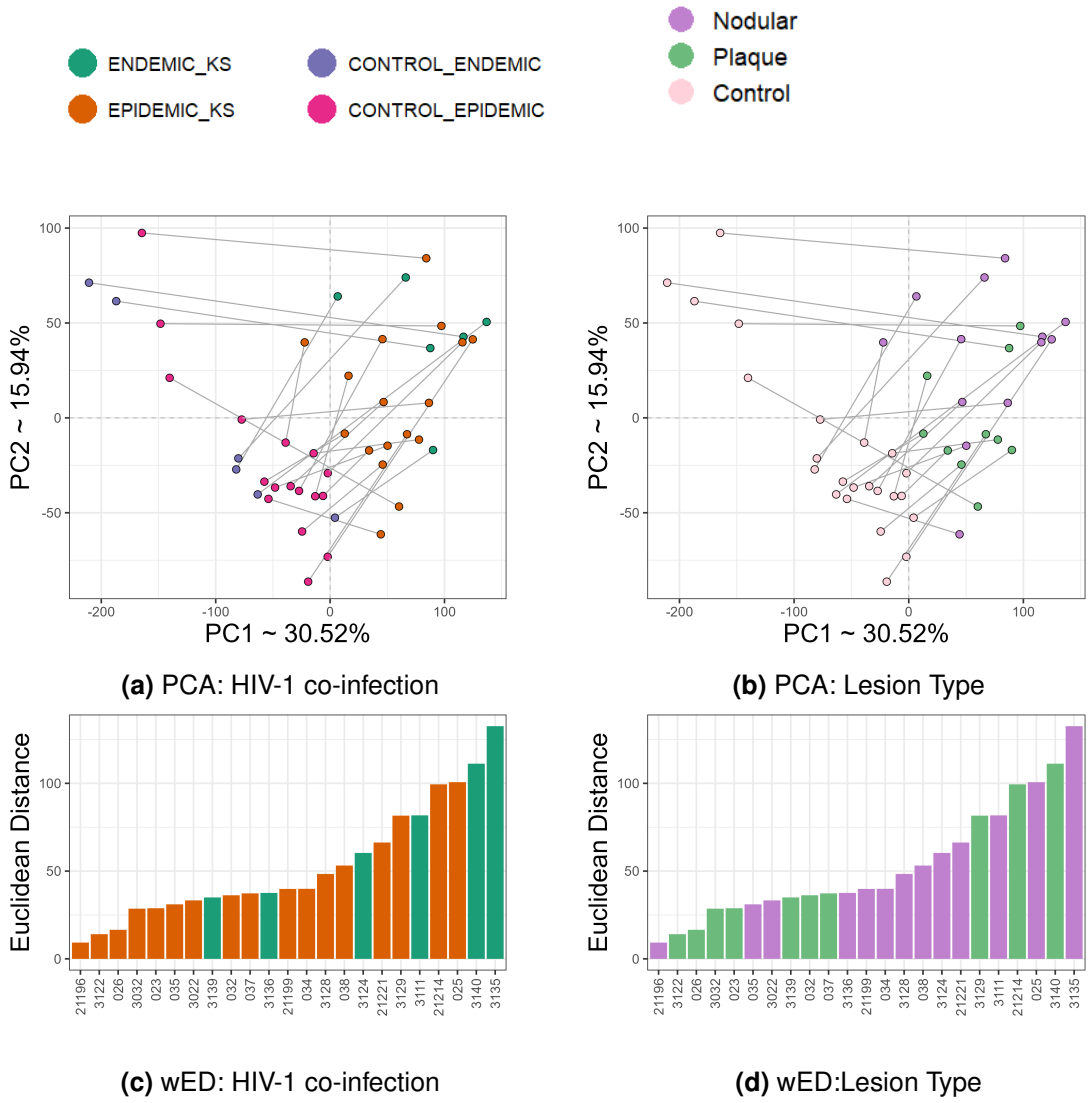


Figure S4: Sample-wise expression distributions of various batch adjustment methods. The log2 TPM expression of all genes in each sample are presented for various batch adjustment schemas (a-f) as well as un-adjusted data (g). Individual subfigures show data adjusted with Limma's removeBatchEffect() (a & b), ComBat (c & d) and ComBat-Seq (e & f). Left-most plots (a, c & e) show expression data without the inclusion of HIV-1 co-infection (epidemic/endemic) labels as a covariate during batch adjustment, while right-most plots (b, d & f) show expression data with the inclusion of this covariate during batch adjustment. Samples are coloured according to whether they are from endemic or epidemic lesion or control biopsies.

7.2.2 PCA applied to paired KS lesion and control samples and corresponding weighted euclidean distance between samples in PC1/PC2-space



(a) PCA: HIV-1 co-infection

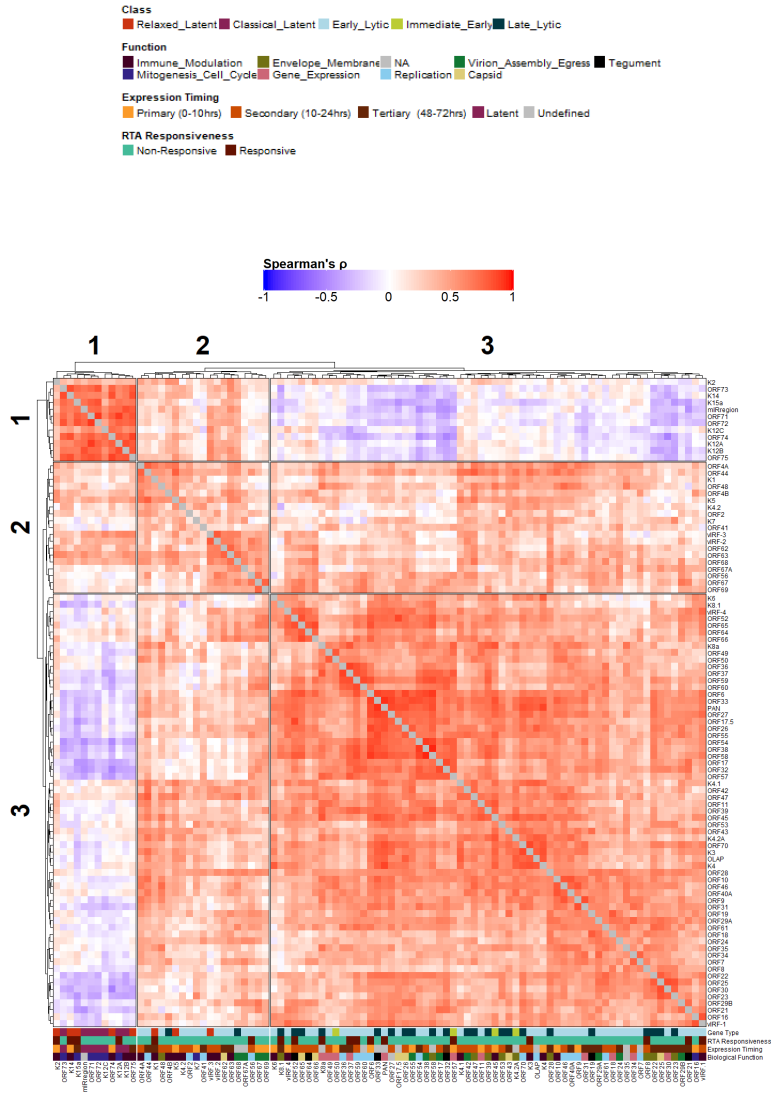
(b) PCA: Lesion Type

(c) wED: HIV-1 co-infection

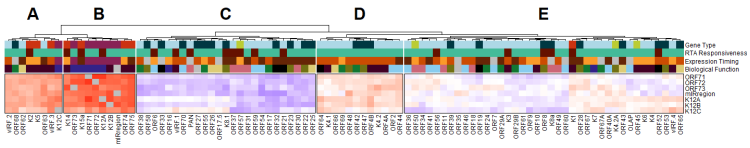
(d) wED:Lesion Type

Figure S5: Weighted euclidean distance between paired KS lesion and control biopsies. Principal Components (PC) score plots showing links between paired KS lesion and control biopsies, coloured by (a) HIV-1 co-infection (epidemic/endemic) or (b) lesion morphology. Points indicate samples and lines indicate links between lesion and control biopsies from the same patient. (c & d) Weighted euclidean distance (wED) between paired lesion and control samples, calculated as the euclidean distance between respective eigenvalues of PC1 and PC2, weighting/scaling each PC-wise distance by the proportion of variance explained by the respective PC. Bars in (c & d) are ordered by increasing wED. Bars in (c) are coloured according to whether the patient had epidemic or endemic KS and bars in (d) are coloured according to lesion morphology. Row labels in (c & d) indicate patient IDs.

7.2.3 Hierarchical clustering of pairwise correlation matrix of viral gene expression



(a) Co-expression clustering of all Pairwise Combinations of Detected KSHV Genes



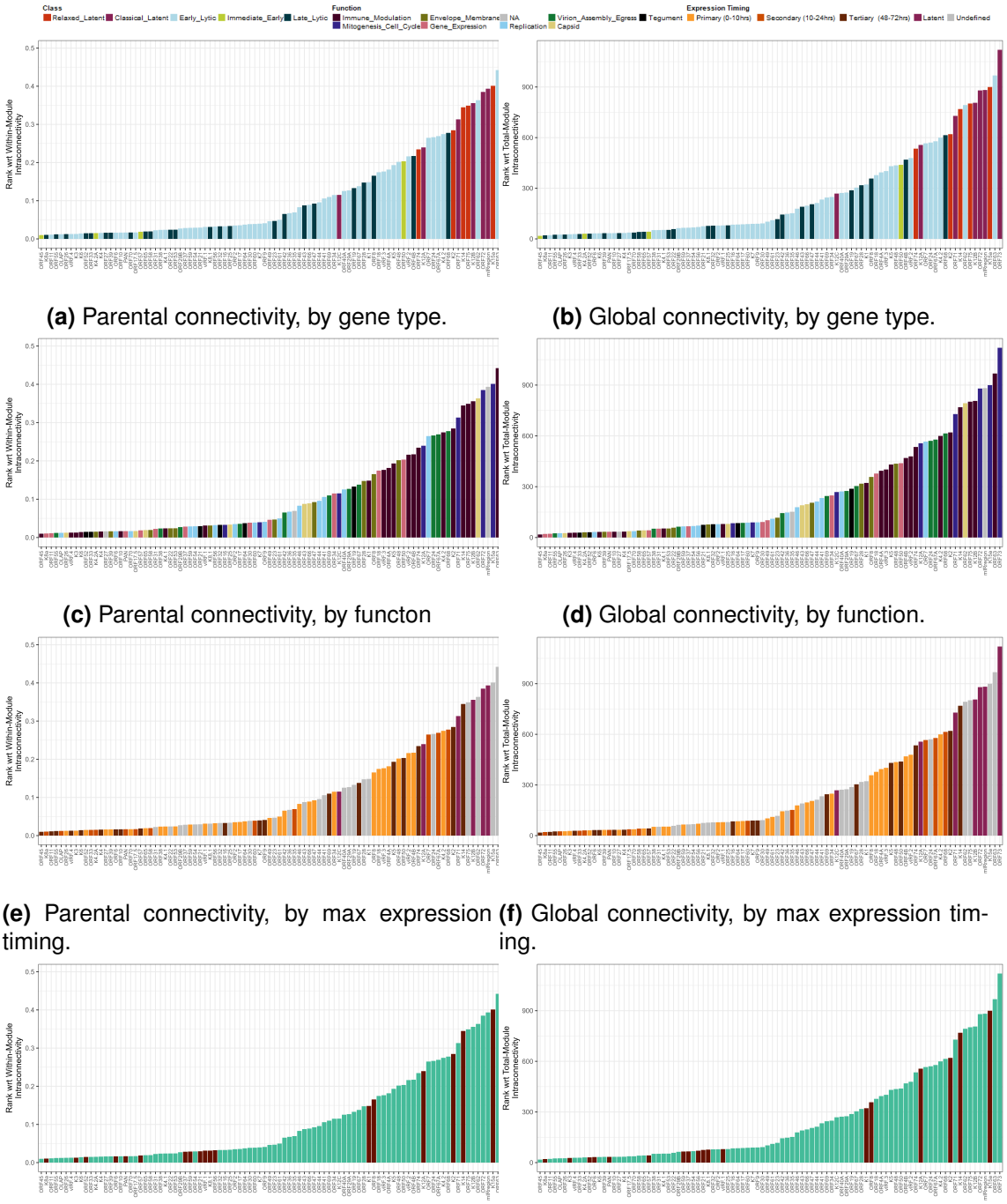
(b) Co-expression clustering of All Detected KSHV Genes Relative to Canonical Latent Genes

Figure S6: Hierachical clustering on pairwise viral gene correlations. Batch-corrected, log2 TPM-transformed viral gene expression profiles were correlated in a pairwise manner using Spearman's ρ correlation. Correlation matrices were clustered using agglomerative average linkage. (a) Pairwise gene-gene correlation heatmap for all detected viral genes. 3 clusters were chosen. (b) Pairwise gene-gene correlation heatmap for all detected genes and only the 7 latent genes (ORF71, ORF72, ORF73, K12A/B/C and viral miRNAs). 5 clusters were chosen. Column-wise gene annotations include viral gene functions, viral gene classifications, direct RTA responsiveness and maximal gene expression timings, as outlined in Section 2.1.6 and in the figure legend.

7.2.4 Variance explained by PC1-4 of module-wise PCA

Figure S7: Percentage of variance explained by module eigengene and successive principal components. A principal components analysis was performed on the module-wise expression matrices, with the 1st principal component (PC1) taken as the module eigengene. The percentage variance explained by each module eigengene and successive PCs (PC2, 3 and 4) are shown, split by each panel.

7.2.5 Viral gene-wise intramodular and global connectivities



(a) Parental connectivity, by gene type.

(b) Global connectivity, by gene type.

(c) Parental connectivity, by functon

(d) Global connectivity, by function.

(e) Parental connectivity, by max expression timing.

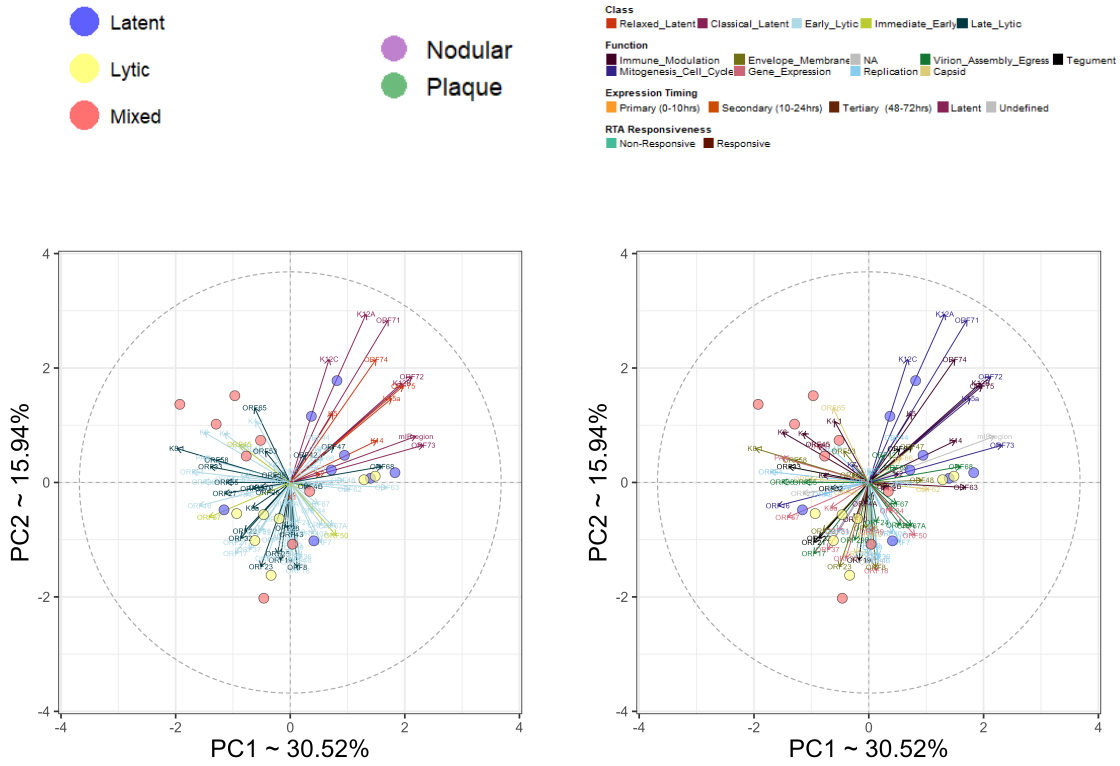
(f) Global connectivity, by max expression timing.

(g) Parental connectivity, by direct RTA responsiveness.

(h) Global connectivity, by direct RTA responsiveness.

Figure S8: Viral gene parental and global connectivity. The intramodular (a, c, e & g) and global (b, d, f & h) connectivity (iC and gC, respectively) were calculated for each gene as outlined in Section 2.3.8. Briefly, connectivity is the sum of a gene's edgeweights, while iC is the connectivity of a gene within the module that it was co-partitioned into, while gC is the connectivity of a gene to the whole network [266]. Genes are ranked in order of decreasing connectivity. Genes are coloured according to gene type (a & b), function (c & d), maximal expression timing (e & f) and direct RTA responsiveness (g & h). Such annotations are further detailed in Section 2.1.6.

7.2.6 Biplots of KS lesion sample-only PCA, showing contribution vectors of viral genes



(a) Gene type.

(b) Function.

(c) Maximal expression timing.

(d) RTA Responsiveness.

Figure S9: Principal components score plots on lesion sample-only samples. A principal components analysis was performed on the 23 KS lesion sample-only, centered (but unscaled) log2 TPM-transformed expression data. Only score plots for PC1 and PC2 are presented. Additionally, factor biplots are presented showing the contribution vector of individual viral genes to PC1 and PC2, colouring according to (a) gene type, (b) gene function, (c) maximal expression timing and (d) direct RTA responsiveness (gene sets outlined in Section 2.1.6). Viral gene contribution arrows are labelled with the viral gene name if their contribution to PC1 or PC2 is greater than the mean contribution for PC1. Samples in the upper two biplots (a & b) are coloured according to the viral gene expression programs, while samples in the lower two biplots (c & d) are coloured according to lesion type. Percentages associated with axes represent the variance explained by each respective PC.

8 Bibliography

[1] Ann Arvin, Gabriella Campadelli-Fiume, Edward Mocarski, Patrick S. Moore, Bernard Roizman, Richard Whitley, and Koichi Yamanishi. Human Herpesviruses. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, pages 10–26, 2007. URL https://www.ncbi.nlm.nih.gov/books/NBK47376/.

[2] Anna Jeffery-Smith and Anna Riddell. Herpesviruses. *Medicine (United Kingdom)*, 49(12):780–784, 12 1996. ISSN 13654357. doi: 10.1016/j.mpmed.2021.09.011. URL https://www.ncbi.nlm.nih.gov/books/NBK8157/.

[3] Sharvan Sehrawat, Dhaneshwar Kumar, and Barry T. Rouse. Herpesviruses: Harmonious Pathogens but Relevant Cofactors in Other Diseases? *Frontiers in Cellular and Infection Microbiology*, 8(MAY):177, 5 2018. ISSN 22352988. doi: 10.3389/FCIMB.2018.00177. URL /pmc/articles/PMC5981231/ /pmc/articles/PMC5981231/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5981231/.

[4] Magdalena Weidner-Glunde, Ewa Kruminis-Kaszkiel, and Mamata Savanagouder. Herpesviral Latency—Common Themes. *Pathogens 2020, Vol. 9, Page 125*, 9(2):125, 2 2020. ISSN 2076-0817. doi: 10.3390/PATHOGENS9020125. URL https://www.mdpi.com/2076-0817/9/2/125/htmhttps://www.mdpi.com/2076-0817/9/2/125.

[5] Don Ganem. KSHV and the pathogenesis of Kaposi sarcoma: listening to human biology and medicine. *The Journal of clinical investigation*, 120(4):939–49, 4 2010. ISSN 1558-8238. doi: 10.1172/JCI40567. URL http://www.ncbi.nlm.nih.gov/pubmed/20364091http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2847423.

[6] Jeffrey I. Cohen. Herpesvirus latency. *The Journal of Clinical Investigation*, 130(7):3361, 5 2020. ISSN 15588238. doi: 10.1172/JCI136225. URL /pmc/articles/PMC7324166//pmc/articles/PMC7324166/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7324166/.

[7] Anthony A. Nash, Bernadette M. Dutia, James P. Stewart, and Andrew J. Davison. Natural history of murine gamma-herpesvirus infection. *Philosophical Transactions of the Royal Society of London. Series B*, 356(1408):569, 4 2001. ISSN 09628436. doi: 10.1098/RSTB.2000.0779. URL /pmc/articles/PMC1088445/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1088445/.

[8] Helmut Fickenscher and Bernhard Fleckenstein. Herpesvirus saimiri. *Philosophical Transactions of the Royal Society of London. Series B*, 356(1408):545, 4 2001. ISSN 09628436. doi: 10.1098/RSTB.2000.0780. URL /pmc/articles/PMC1088444/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1088444/.

[9] Vincent Lacoste, Philippe Mauclère, Guy Dubreuil, John Lewis, Marie Claude Georges-Courbot, and Antoine Gessain. A Novel γ2-Herpesvirus of the Rhadinovirus 2 Lineage in Chimpanzees. *Genome Research*, 11(9):1511, 2001. ISSN 10889051. doi: 10.1101/GR.158601. URL /pmc/articles/PMC311113//pmc/articles/PMC311113/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC311113/.

[10] Devon E. McMahon, Toby Maurer, and Esther E. Freeman. 25 years of Kaposi sarcoma herpesvirus: Discoveries, disparities, and diagnostics. *Journal of Global Oncology*, 6:505–507, 3 2020. ISSN 23789506. doi: 10.1200/GO.20.00027.

[11] Ethel Cesarman, Blossom Damania, Susan E. Krown, Jeffrey Martin, Mark Bower, and Denise Whitby. Kaposi sarcoma. *Nature Reviews Disease Primers 2019 5:1*, 5(1):1–21, 1 2019. ISSN 2056-676X. doi: 10.1038/s41572-019-0060-9. URL https://www.nature.com/articles/s41572-019-0060-9.

[12] Blanca Iciar Indave Ruiz, Subasri Armon, Reiko Watanabe, Lesley Uttley, Valerie A. White, Alexander J. Lazar, and Ian A. Cree. Clonality, Mutation and Kaposi Sarcoma: A Systematic Review. *Cancers*, 14(5), 3 2022. ISSN 20726694. doi: 10.3390/CANCERS14051201/S1. URL /pmc/articles/PMC8909603//pmc/articles/PMC8909603/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8909603/.

[13] Aude Jary, Marianne Veyri, Adélie Gothland, Valentin Leducq, Vincent Calvez, and Anne Geneviève Marcelin. Kaposi’s Sarcoma-Associated Herpesvirus, the Etiological Agent of All Epidemiological Forms of Kaposi’s Sarcoma. *Cancers 2021, Vol. 13, Page 6208*, 13(24):6208, 12 2021. ISSN 2072-6694. doi: 10.3390/CANCERS13246208. URL https://www.mdpi.com/2072-6694/13/24/6208/htmhttps://www.mdpi.com/2072-6694/13/24/6208.

[14] Oliver Manners, James C Murphy, Alex Coleman, David J Hughes, and Adrian Whitehouse. Contribution of the KSHV and EBV lytic cycles to tumourigenesis. *Current Opinion in Virology*, 32:60–70, 10 2018. ISSN 18796257. doi: 10.1016/j.coviro.2018.08.014. URL https://linkinghub.elsevier.com/retrieve/pii/S1879625718300920.

[15] A. Gregory Bruce, Serge Barcy, Terri Dimaio, Emilia Gan, H. Jacques Garrigues, Michael Lagunoff, and Timothy M. Rose. Quantitative Analysis of the KSHV Transcriptome Following Primary Infection of Blood and Lymphatic Endothelial Cells. *Pathogens*, 6(1), 3 2017. ISSN 20760817. doi: 10.3390/PATHOGENS6010011. URL /pmc/articles/PMC5371899//pmc/articles/PMC5371899/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5371899/.

[16] For Yue Tso, Andrew V. Kossenkov, Salum J. Lidenge, Owen Ngalamika, John R. Ngowi, Julius Mwaiselage, Jayamanna Wickramasinghe, Eun Hee Kwon, John T. West, Paul M. Lieberman, and Charles Wood. RNA-Seq of Kaposi's sarcoma reveals alterations in glucose and lipid metabolism. *PLOS Pathogens*, 14(1):e1006844, 1 2018. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1006844. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1006844.

[17] Timothy M Rose, A Gregory Bruce, Serge Barcy, Matt Fitzgibbon, Lisa R Matsumoto, Minako Ikoma, Corey Casper, Jackson Orem, and Warren Phipps. Quantitative RNAseq analysis of Ugandan KS tumors reveals KSHV gene expression dominated by transcription from the LTd downstream latency promoter. *PLOS Pathogens*, 14(12):1–42, 2018. doi: 10.1371/journal.ppat.1007441. URL https://doi.org/10.1371/journal.ppat.1007441.

[18] Timothy M. Rose, A. Gregory Bruce, Serge Barcy, Matt Fitzgibbon, Lisa R. Matsumoto, Minako Ikoma, Corey Casper, Jackson Orem, and Warren Phipps. Quantitative RNAseq analysis of Ugandan KS tumors reveals KSHV gene expression dominated by transcription from the LTd downstream latency promoter. *PLOS Pathogens*, 14(12):e1007441, 12 2018. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1007441. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1007441.

[19] Lucas E. Cavallin, Pascal Goldschmidt-Clermont, and Enrique A. Mesri. Molecular and Cellular Mechanisms of KSHV Oncogenesis of Kaposi's Sarcoma Associated with HIV/AIDS. *PLOS Pathogens*, 10(7):e1004154, 2014. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1004154. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1004154.

[20] Nathan A. Ungerleider, Scott A. Tibbetts, Rolf Renne, and Erik K. Flemington. Gammaherpesvirus RNAs Come Full Circle. *mBio*, 10(2):1–8, 2019. doi: 10.1128/MBIO.00071-19. URL /pmc/articles/PMC6445933//pmc/articles/PMC6445933/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6445933/.

[21] Carolina Arias, Ben Weisburd, Noam Stern-Ginossar, Alexandre Mercier, Alexis S. Madrid, Priya Bellare, Meghan Holdorf, Jonathan S. Weissman, and Don Ganem. KSHV 2.0: A Comprehensive Annotation of the Kaposi's Sarcoma-Associated Herpesvirus Genome Using Next-Generation Sequencing Reveals Novel Genomic and Functional Features. *PLOS Pathogens*, 10(1):e1003847, 1 2014. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1003847. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003847.

[22] Tuna Toptan, Bizunesh Abere, Michael A. Nalesnik, Steven H. Swerdlow, Sarangarajan Ranganathan, Nara Lee, Kathy H. Shair, Patrick S. Moore, and Yuan Chang. Circular DNA tumor viruses make circular RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (37):E8737–E8745, 9 2018. ISSN 10916490. doi: 10.1073/PNAS.1811728115/SUPPL{_}FILE/PNAS.1811728115.SD04.XLSX. URL https://www.pnas.org/doi/abs/10.1073/pnas.1811728115.

[23] Takanobu Tagawa, Shaojian Gao, Vishal N. Koparde, Mileidy Gonzalez, John L. Spouge, Anna P. Serquiña, Kathryn Lurain, Ramya Ramaswami, Thomas S. Uldrick, Robert Yarchoan, and Joseph M. Ziegelbauer. Discovery of Kaposi's sarcoma herpesvirus-encoded circular RNAs and a human antiviral circular RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 115(50):12805–12810, 12 2018. ISSN 1091-6490. doi: 10.1073/PNAS.1816183115. URL https://pubmed.ncbi.nlm.nih.gov/30455306/.

[24] Takanobu Tagawa, Daniel Oh, Jerico Santos, Sarah Dremel, Guruswamy Mahesh, Thomas S. Uldrick, Robert Yarchoan, Vishal N. Kopardé, and Joseph M. Ziegelbauer. Characterizing Expression and Regulation of Gamma-Herpesviral Circular RNAs. *Frontiers in Microbiology*, 12, 6 2021. doi: 10.3389/FMICB.2021.670542. URL /pmc/articles/PMC8278476//pmc/articles/PMC8278476/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8278476/.

[25] Vladimir Majerciak, Mathew Lu, Xiaofan Li, and Zhi-Ming Zheng. Attenuation of the suppressive activity of cellular splicing factor SRSF3 by Kaposi sarcoma–associated herpesvirus ORF57 protein is required for RNA splicing. *RNA*, 20(11):1747–1758, 11 2014. ISSN 1355-8382. doi: 10.1261/rna.045500.114. URL http://www.ncbi.nlm.nih.gov/pubmed/25234929http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4201827http://rnajournal.cshlp.org/lookup/doi/10.1261/rna.045500.114.

[26] Bizunesh Abere, Jinghui Li, Hongzhao Zhou, Tuna Toptan, Patrick S. Moore, and Yuan Chang. Kaposi's Sarcoma-Associated Herpesvirus-Encoded circRNAs Are Expressed in Infected Tumor Tissues and Are Incorporated into Virions. *mBio*, 11(1), 1 2020. ISSN 2150-7511. doi: 10.1128/MBIO.03027-19. URL https://pubmed.ncbi.nlm.nih.gov/31911496/.

[27] Vladimir Majerciak, Beatriz Alvarado-Hernandez, Alexei Lobanov, Maggie Cam, and Zhi Ming Zheng. Genome-wide regulation of KSHV RNA splicing by viral RNA-binding protein ORF57. *PLOS Pathogens*, 18(7):e1010311, 7 2022. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1010311. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1010311.

[28] Kylee Morrison, Mark Manzano, Kevin Chung, Matthew J. Schipma, Elizabeth T. Bartom, and Eva Gottwein. The Oncogenic Kaposi's Sarcoma-Associated Herpesvirus Encodes a Mimic of the Tumor-Suppressive miR-15/16 miRNA Family. *Cell Reports*, 29(10):2961–2969, 12 2019. ISSN 2211-1247. doi: 10.1016/J.CELREP.2019.11.005.

[29] Jennifer S. Cannon, John Nicholas, Jan M. Orenstein, Risa B. Mann, Paul G. Murray, Philip J. Browning, Joseph A. DiGiuseppe, Ethel Cesarman, Gary S. Hayward, and Richard F. Ambinder. Heterogeneity of Viral IL-6 Expression in HHV-8—Associated Diseases. *The Journal of Infectious Diseases*, 180(3):824–828, 9 1999. ISSN 0022-1899. doi: 10.1086/314956. URL https://academic.oup.com/jid/article/180/3/824/817909.

[30] Patrick S. Moore, Chris Boshoff, Robin A. Weiss, and Yuan Chang. Molecular Mimicry of Human Cytokine and Cytokine Response Pathway Genes by KSHV. *Science*, 274(5293):1739–1744, 12 1996. ISSN 00368075. doi: 10.1126/SCIENCE.274.5293.1739. URL https://www.science.org/doi/10.1126/science.274.5293.1739.

[31] Carolina Arias, Ben Weisburd, Noam Stern-Ginossar, Alexandre Mercier, Alexis S. Madrid, Priya Bellare, Meghan Holdorf, Jonathan S. Weissman, and Don Ganem. KSHV 2.0: A Comprehensive Annotation of the Kaposi's Sarcoma-Associated Herpesvirus Genome Using Next-Generation Sequencing Reveals Novel Genomic and Functional Features. *PLOS Pathogens*, 10(1):e1003847, 1 2014. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1003847. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003847.

[32] Jie Qin, Wan Li, Shou-Jiang Gao, and Chun Lu. KSHV microRNAs: Tricks of the Devil. *Trends in microbiology*, 25(8):648, 8 2017. doi: 10.1016/J.TIM.2017.02.002. URL /pmc/articles/PMC6904892//pmc/articles/PMC6904892/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6904892/.

[33] Alejandro Casco, Akansha Gupta, Mitchell Hayes, Reza Djavadian, Makoto Ohashi, and Eric Johannsen. Accurate Quantification of Overlapping Herpesvirus Transcripts from RNA Sequencing Data. *Journal of Virology*, 96(2):1635–1656, 1 2022. ISSN 0022-538X. doi: 10.1128/JVI.01635-21/SUPPL{_}FILE/JVI.01635-21-S0003.XLSX. URL https://journals.asm.org/doi/10.1128/JVI.01635-21.

[34] Adam Grundhoff and Don Ganem. Mechanisms Governing Expression of the v-FLIP Gene of Kaposi's Sarcoma-Associated Herpesvirus. *Journal of Virology*, 75(4):1857, 2 2001. ISSN 0022-538X. doi: 10.1128/JVI.75.4.1857-1863.2001. URL /pmc/articles/PMC114095//pmc/articles/PMC114095/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC114095/.

[35] Benes L. Trus, J. Bernard Heymann, Karin Nealon, Naiqian Cheng, William W. Newcomb, Jay C. Brown, Dean H. Kedes, and Alasdair C. Steven. Capsid Structure of Kaposi's Sarcoma-Associated Herpesvirus, a Gammaherpesvirus, Compared to Those of an Alphaherpesvirus, Herpes Simplex Virus Type 1, and a Betaherpesvirus, Cytomegalovirus. *Journal of Virology*, 75(6):2879, 3 2001. ISSN 0022-538X. doi: 10.1128/JVI.75.6.2879-2890.2001. URL /pmc/articles/PMC115914//pmc/articles/PMC115914/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC115914/.

[36] Danyang Gong, Xinghong Dai, Jonathan Jih, Yun Tao Liu, Guo Qiang Bi, Ren Sun, and Z. Hong Zhou. DNA-Packing Portal and Capsid-Associated Tegument Complexes in the Tumor Herpesvirus KSHV. *Cell*, 178(6):1329–1343, 9 2019. ISSN 0092-8674. doi: 10.1016/J.CELL.2019.07.035.

[37] Pravinkumar Purushothaman, Suhani Thakker, and Subhash C. Verma. Transcriptome Analysis of Kaposi's Sarcoma-Associated Herpesvirus during De Novo Primary Infection of Human B and Endothelial Cells. *Journal of Virology*, 89(6):3093, 3 2015. ISSN 0022-538X. doi: 10.1128/JVI.02507-14. URL /pmc/articles/PMC4337554//pmc/articles/PMC4337554/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4337554/.

[38] Florian Full, Doris Jungnickl, Nina Reuter, Elke Bogner, Kevin Brulois, Brigitte Scholz, Michael Stürzl, Jinjong Myoung, Jae U. Jung, Thomas Stamminger, and Armin Ensser. Kaposi's Sarcoma Associated Herpesvirus Tegument Protein ORF75 Is Essential for Viral Lytic Replication and Plays a Critical Role in the Antagonization of ND10-Instituted Intrinsic Immunity. *PLOS Pathogens*, 10(1):e1003863, 1 2014. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1003863. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003863.

[39] James C. Murphy, Elena M. Harrington, Sophie Schumann, Elton J.R. Vasconcelos, Timothy J. Mottram, Katherine L. Harper, Julie L. Aspden, and Adrian Whitehouse. Kaposi's sarcoma-associated herpesvirus induces specialised ribosomes to efficiently translate viral lytic mRNAs. *Nature Communications 2023 14:1*, 14(1):1–15, 1 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-35914-5. URL https://www.nature.com/articles/s41467-023-35914-5.

[40] Belinda Baquero-Pérez and Adrian Whitehouse. Hsp70 Isoforms Are Essential for the Formation of Kaposi's Sarcoma-Associated Herpesvirus Replication and Transcription Compartments. *PLOS Pathogens*, 11(11):e1005274, 11 2015. ISSN 1553-7374. doi: 10.1371/journal.ppat.1005274. URL https://dx.plos.org/10.1371/journal.ppat.1005274.

[41] Ramina Nabiee, Basir Syed, Jesus Ramirez Castano, Rukhsana Lalani, and Jennifer E. Totonchy. An Update of the Virion Proteome of Kaposi Sarcoma-Associated Herpesvirus. *Viruses 2020, Vol. 12, Page 1382*, 12(12):1382, 12 2020. ISSN 1999-4915. doi: 10.3390/V12121382. URL https://www.mdpi.com/1999-4915/12/12/1382/htmhttps://www.mdpi.com/1999-4915/12/12/1382.

[42] Jinjong Myoung and Don Ganem. Infection of primary human tonsillar lymphoid cells by KSHV reveals frequent but abortive infection of T cells. *Virology*, 413(1):1, 4 2011. ISSN 00426822. doi: 10.1016/J.VIROL.2010.12.036. URL /pmc/articles/PMC3070441//pmc/articles/PMC3070441/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3070441/.

[43] Binod Kumar and Bala Chandran. KSHV Entry and Trafficking in Target Cells—Hijacking of Cell Signal Pathways, Actin and Membrane Dynamics. *Viruses*, 8(11), 11 2016. ISSN 19994915. doi: 10.3390/V8110305. URL /pmc/articles/PMC5127019//pmc/articles/PMC5127019/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5127019/.

[44] Pravinkumar Purushothaman, Timsy Uppal, and Subhash Verma. Molecular Biology of KSHV Lytic Reactivation. *Viruses*, 7(1):116–153, 1 2015. ISSN 1999-4915. doi: 10.3390/v7010116. URL http://www.mdpi.com/1999-4915/7/1/116.

[45] Stephen J. Dollery. Towards Understanding KSHV Fusion and Entry. *Viruses 2019, Vol. 11, Page 1073*, 11(11):1073, 11 2019. ISSN 1999-4915. doi: 10.3390/V11111073. URL https://www.mdpi.com/1999-4915/11/11/1073/htmhttps://www.mdpi.com/1999-4915/11/11/1073.

[46] Daniela Dünn-Kittenplon, Asaf Ashkenazy-Titelman, Inna Kalt, Jean Paul Lellouche, Yaron Shav-Tal, and Ronit Sarid. The Portal Vertex of KSHV Promotes Docking of Capsids at the Nuclear Pores. *Viruses 2021, Vol. 13, Page 597*, 13(4):597, 3 2021. ISSN 1999-4915. doi: 10.3390/V13040597. URL https://www.mdpi.com/1999-4915/13/4/597/htmhttps://www.mdpi.com/1999-4915/13/4/597.

[47] Grant Broussard and Blossom Damania. Regulation of KSHV Latency and Lytic Reactivation. *Viruses 2020, Vol. 12, Page 1034*, 12(9):1034, 9 2020. doi: 10.3390/V12091034. URL https://www.mdpi.com/1999-4915/12/9/1034/htmhttps://www.mdpi.com/1999-4915/12/9/1034.

[48] Zsolt Toth, Bernadett Papp, Kevin Brulois, Youn Jung Choi, Shou Jiang Gao, and Jae U. Jung. LANA-Mediated Recruitment of Host Polycomb Repressive Complexes onto the KSHV Genome during De Novo Infection. *PLOS Pathogens*, 12(9):e1005878, 9 2016. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1005878. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1005878.

[49] Zsolt Toth, Kevin Brulois, Hye Ra Lee, Yoshihiro Izumiya, Clifford Tepper, Hsing Jien Kung, and Jae U. Jung. Biphasic Euchromatin-to-Heterochromatin Transition on the KSHV Genome Following De Novo Infection. *PLOS Pathogens*, 9(12):e1003813, 2013. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1003813. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003813.

[50] Zsolt Toth, Richard J. Smindak, and Bernadett Papp. Inhibition of the lytic cycle of Kaposi's sarcoma-associated herpesvirus by cohesin factors following de novo infection. *Virology*, 512:25–33, 12 2017. ISSN 1096-0341. doi: 10.1016/J.VIROL.2017.09.001. URL https://pubmed.ncbi.nlm.nih.gov/28898712/.

[51] Kawalpreet K. Aneja and Yan Yuan. Reactivation and Lytic Replication of Kaposi's Sarcoma-Associated Herpesvirus: An Update. *Frontiers in Microbiology*, 8:613, 4 2017. ISSN 1664-302X. doi: 10.3389/fmicb.2017.00613. URL http://journal.frontiersin.org/article/10.3389/fmicb.2017.00613/full.

[52] Louise Giffin and Blossom Damania. KSHV: Pathways to Tumorigenesis and Persistent Infection. *Advances in virus research*, 88:111, 2014. ISSN 15578399. doi: 10.1016/B978-0-12-800098-4.00002-7. URL /pmc/articles/PMC4104069//pmc/articles/PMC4104069/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4104069/.

[53] Ashish Kumar, Yuanzhi Lyu, Yuichi Yanagihashi, Zhi-Ming Zheng, Mel Campbell, Yoshihiro Izumiya, Chanikarn Chantarasrivong, Vladimir Majerciak, Michelle Salemi, Kang-Hsin Wang, Tomoki Inagaki, Frank Chuang, Ryan R Davis, Clifford G Tepper, Kazushi Nakano, Chie Izumiya, Michiko Shimoda, Ken-Ichi Nakajima, and Alexander Merleev. KSHV episome tethering sites on host chromosomes and regulation of latency-lytic switch by CHD4. *CellReports*, 39:110788, 2022. doi: 10.1016/j.celrep.2022.110788. URL https://doi.org/10.1016/j.celrep.2022.110788.

[54] Jin Gan, Chong Wang, Yanling Jin, Yi Guo, Feng Xu, Qing Zhu, Ling Ding, Hong Shang, Junwen Wang, Fang Wei, Qiliang Cai, and Erle S. Robertson. Proteomic profiling identifies the SIM-associated complex of KSHV-encoded LANA. *PROTEOMICS*, 15(12):2023–2037, 6 2015. ISSN 1615-9861. doi: 10.1002/PMIC.201400624. URL https://onlinelibrary.wiley.com/doi/full/10.1002/pmic.201400624https://onlinelibrary.wiley.com/doi/abs/10.1002/pmic.201400624https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.201400624.

[55] Timsy Uppal, Sagarika Banerjee, Zhiguo Sun, Subhash C. Verma, and Erle S. Robertson. KSHV LANA—The Master Regulator of KSHV Latency. *Viruses*, 6 (12):4961, 2014. ISSN 19994915. doi: 10.3390/V6124961. URL /pmc/articles/PMC4276939//pmc/articles/PMC4276939/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4276939/.

[56] Enea Gino Di Domenico, Luigi Toma, Valentina Bordignon, Elisabetta Trento, Giovanna D'Agosto, Paola Cordiali-Fei, and Fabrizio Ensoli. Activation of DNA Damage Response Induced by the Kaposi's Sarcoma-Associated Herpes Virus. *International Journal of Molecular Sciences 2016, Vol. 17, Page 854*, 17 (6):854, 6 2016. ISSN 1422-0067. doi: 10.3390/IJMS17060854. URL https://www.mdpi.com/1422-0067/17/6/854/htmhttps://www.mdpi.com/1422-0067/17/6/854.

[57] Daniel L. Holmes, Daniel T. Vogt, and Michael Lagunoff. A CRISPR-Cas9 screen identifies mitochondrial translation as an essential process in latent KSHV infection of human endothelial cells. *Proceedings of the National Academy of Sciences of the United States of America*, 117(45):28384–28392, 11 2020. ISSN 10916490. doi: 10.1073/PNAS.2011645117/SUPPL{_}FILE/PNAS.2011645117.SD02.XLSX. URL https://www.pnas.org/doi/abs/10.1073/pnas.2011645117.

[58] Nenavath Gopal Naik, Thomas Hong Nguyen, Lauren Roberts, Luke Todd Fischer, Katherine Glickman, Gavin Golas, Bernadett Papp, and Zsolt Toth. Epigenetic factor siRNA screen during primary KSHV infection identifies novel host restriction factors for the lytic cycle of KSHV. *PLoS Pathogens*, 16 (1), 2020. ISSN 15537374. doi: 10.1371/JOURNAL.PPAT.1008268. URL /pmc/articles/PMC6977772//pmc/articles/PMC6977772/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6977772/.

[59] Margaret Chang, Helen J. Brown, Alicia Collado-Hidalgo, Jesusa M. Arevalo, Zoran Galic, Tonia L. Symensma, Lena Tanaka, Hongyu Deng, Jerome A. Zack, Ren Sun, and Steve W. Cole. β-Adrenoreceptors Reactivate Kaposi's Sarcoma-Associated Herpesvirus Lytic Replication via PKA-Dependent Control of Viral RTA. *Journal of Virology*, 79(21):13538, 11 2005. ISSN 0022-538X. doi: 10.1128/JVI.79.21.13538-13547.2005. URL /pmc/articles/PMC1262578//pmc/articles/PMC1262578/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1262578/.

[60] Bernadett Papp, Naeem Motlagh, Richard J Smindak, Seung Jin Jang, Aria Sharma, Juan D Alonso, and Zsolt Toth. Genome-Wide Identification of Direct RTA Targets Reveals Key Host Factors for Kaposi's Sarcoma-Associated Herpesvirus Lytic Reactivation. *Journal of virology*, 93(5), 3 2019. ISSN 1098-5514. doi: 10.1128/JVI.01978-18. URL http://www.ncbi.nlm.nih.gov/pubmed/30541837http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6384073.

[61] Praneet Kaur Sandhu and Blossom Damania. The regulation of KSHV lytic reactivation by viral and cellular factors. *Current Opinion in Virology*, 52:39–47, 2 2022. ISSN 1879-6257. doi: 10.1016/J.COVIRO.2021.11.004.

[62] Yuying Liang and Don Ganem. Lytic but not latent infection by Kaposi's sarcoma-associated herpesvirus requires host CSL protein, the mediator of Notch signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 100(14):8490–8495, 7 2003. ISSN 0027-8424. doi: 10.1073/PNAS.1432843100. URL https://pubmed.ncbi.nlm.nih.gov/12832621/.

[63] Lauren R. Combs, Lauren McKenzie Spires, Juan D. Alonso, Bernadett Papp, and Zsolt Toth. KSHV RTA Induces Degradation of the Host Transcription Repressor ID2 To Promote the Viral Lytic Cycle. *Journal of Virology*, 96(12), 6 2022. ISSN 0022-538X. doi: 10.1128/JVI.00101-22. URL /pmc/articles/PMC9215225//pmc/articles/PMC9215225/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9215225/.

[64] Yuanzhi Lyu, Kazushi Nakano, Ryan R. Davis, Clifford G. Tepper, Mel Campbell, and Yoshihiro Izumiya. ZIC2 Is Essential for Maintenance of Latency and Is a Target of an Immediate Early Protein during Kaposi's Sarcoma-Associated Herpesvirus Lytic Reactivation. *Journal of Virology*, 91(21), 11 2017. ISSN 0022-538X. doi: 10.1128/JVI.00980-17/SUPPL{_}FILE/ZJV99918303 8S1.PDF. URL https://journals.asm.org/doi/10.1128/JVI.00980-17.

[65] Fan Xiu Zhu, Teresa Cusano, and Yan Yuan. Identification of the Immediate-Early Transcripts of Kaposi's Sarcoma-Associated Herpesvirus. *Journal of Virology*, 73(7):5556, 7 1999. ISSN 0022-538X. doi: 10.1128/JVI.73.7.5556-5567.1999. URL /pmc/articles/PMC112613//pmc/articles/PMC112613/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC112613/.

[66] Lydia V. McClure, Rodney P. Kincaid, James M. Burke, Adam Grundhoff, and Christopher S. Sullivan. Comprehensive Mapping and Analysis of Kaposi's Sarcoma-Associated Herpesvirus 3 UTRs Identify Differential Posttranscriptional Control of Gene Expression in Lytic versus Latent Infection. *Journal of Virology*, 87(23):12838, 12 2013. ISSN 0022-538X. doi: 10.1128/JVI.02374-13. URL /pmc/articles/PMC3838127//pmc/articles/PMC3838127/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3838127/.

[67] Salum J. Lidenge, Andrew V. Kossenkov, For Yue Tso, Jayamanna Wickramasinghe, Sara R. Privatt, Owen Ngalamika, John R. Ngowi, Julius Mwaiselage, Paul M. Lieberman, John T. West, and Charles Wood. Comparative transcriptome analysis of endemic and epidemic Kaposi's sarcoma (KS) lesions and the secondary role of HIV-1 in KS pathogenesis. *PLOS Pathogens*, 16(7):e1008681, 7 2020. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1008681. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1008681.

[68] Brian R. Jackson, Marko Noerenberg, and Adrian Whitehouse. A Novel Mechanism Inducing Genome Instability in Kaposi's Sarcoma-Associated Herpesvirus Infected Cells. *PLoS Pathogens*, 10(5):e1004098, 5 2014. ISSN 1553-7374. doi: 10.1371/journal.ppat.1004098. URL https://dx.plos.org/10.1371/journal.ppat.1004098.

[69] Adam Taylor, Brian R. Jackson, Marko Noerenberg, David J. Hughes, James R. Boyne, Mark Verow, Mark Harris, and Adrian Whitehouse. Mutation of a C-terminal motif affects Kaposi's sarcoma-associated herpesvirus ORF57 RNA binding, nuclear trafficking, and multimerization. *Journal of virology*, 85 (15):7881–7891, 8 2011. ISSN 1098-5514. doi: 10.1128/JVI.00138-11. URL https://pubmed.ncbi.nlm.nih.gov/21593148/.

[70] Brian R. Jackson, James R. Boyne, Marko Noerenberg, Adam Taylor, Guillaume M. Hautbergue, Matthew J. Walsh, Rachel Wheat, David J. Blackbourn, Stuart A. Wilson, and Adrian Whitehouse. An Interaction between KSHV ORF57 and UIF Provides mRNA-Adaptor Redundancy in Herpesvirus Intronless mRNA Export. *PLoS Pathogens*, 7(7):e1002138, 7 2011. ISSN 1553-7374. doi: 10.1371/journal.ppat.1002138. URL https://dx.plos.org/10.1371/journal.ppat.1002138.

[71] Katherine L Harper, Timothy J Mottram, Chinedu A Anene, Becky Foster, Molly R Patterson, Euan McDonnell, Andrew Macdonald, David Westhead, and Adrian Whitehouse. Dysregulation of the miR-30c/DLL4 axis by circHIPK3 is essential for KSHV lytic replication. *EMBO reports*, 23(5):e54117, 5 2022. ISSN 1469-3178. doi: 10.15252/EMBR.202154117. URL https://onlinelibrary.wiley.com/doi/full/10.15252/embr.202154117https://onlinelibrary.wiley.com/doi/abs/10.15252/embr.202154117https://www.embopress.org/doi/10.15252/embr.202154117.

[72] Mel Campbell and Yoshihiro Izumiya. PAN RNA: transcriptional exhaust from a viral engine. *Journal of Biomedical Science 2020 27:1*, 27(1):1–10, 3 2020. ISSN 1423-0127. doi: 10.1186/S12929-020-00637-Y. URL https://jbiomedsci.biomedcentral.com/articles/10.1186/s12929-020-00637-y.

[73] Cyprian C. Rossetto and Gregory Pari. KSHV PAN RNA associates with demethylases UTX and JMJD3 to activate lytic replication through a physical interaction with the virus genome. *PLoS pathogens*, 8(5), 5 2012. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1002680. URL https://pubmed.ncbi.nlm.nih.gov/22589717/.

[74] Lei Xu and Yan Yuan. Two microPeptides are translated from a KSHV polycistronic RNA in human cells by leaky scanning mechanism. *Biochemical and biophysical research communications*, 522(3):568, 2 2020. ISSN 10902104. doi: 10.1016/J.BBRC.2019.11.087. URL /pmc/articles/PMC6981015/ /pmc/articles/PMC6981015/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6981015/.

[75] Yoshihiro Izumiya, Chie Izumiya, Albert Van Geelen, Don-Hong Wang, Kit S. Lam, Paul A. Luciw, and Hsing-Jien Kung. Kaposi's Sarcoma-Associated Herpesvirus-Encoded Protein Kinase and Its Interaction with K-bZIP. *Journal of Virology*, 81(3):1072, 2 2007. ISSN 0022-538X. doi: 10.1128/JVI.01473-06. URL /pmc/articles/PMC1797516//pmc/articles/PMC1797516/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1797516/.

[76] Yan Wang, Qiyi Tang, Gerd G. Maul, and Yan Yuan. Kaposi's Sarcoma-Associated Herpesvirus ori-Lyt-Dependent DNA Replication: Dual Role of Replication and Transcription Activator. *Journal of Virology*, 80(24):12171, 12 2006. ISSN 0022-538X. doi: 10.1128/JVI.00990-06. URL /pmc/articles/PMC1676287//pmc/articles/PMC1676287/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1676287/.

[77] Marta Maria Gaglia, Chris H. Rycroft, and Britt A. Glaunsinger. Transcriptome-Wide Cleavage Site Mapping on Cellular mRNAs Reveals Features Underlying Sequence-Specific Cleavage by the Viral Ribonuclease SOX. *PLoS Pathogens*, 11(12), 2015. ISSN 15537374. doi: 10.1371/JOURNAL.PPAT.1005305. URL /pmc/articles/PMC4672902//pmc/articles/PMC4672902/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4672902/.

[78] Britt Glaunsinger, Leonard Chavez, and Don Ganem. The Exonuclease and Host Shutoff Functions of the SOX Protein of Kaposi's Sarcoma-Associated Herpesvirus Are Genetically Separable. *Journal of Virology*, 79(12):7396, 6 2005. ISSN 0022-538X. doi: 10.1128/JVI.79.12.7396-7401.2005. URL /pmc/articles/PMC1143623//pmc/articles/PMC1143623/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1143623/.

[79] Hyunah Lee, Anathe O.M. Patschull, Claire Bagnéris, Hannah Ryan, Christopher M. Sanderson, Bahram Ebrahimi, Irene Nobeli, and Tracey E. Barrett. KSHV SOX mediated host shutoff: the molecular mechanism underlying mRNA transcript processing. *Nucleic Acids Research*, 45(8):gkw1340, 1 2017. ISSN 0305-1048. doi: 10.1093/nar/gkw1340. URL http://www.ncbi.nlm.nih.gov/pubmed/28132029http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5416870https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1340.

[80] Karen Clyde and Britt A. Glaunsinger. Deep Sequencing Reveals Direct Targets of Gammaherpesvirus-Induced mRNA Decay and Suggests That Multiple Mechanisms Govern Cellular Transcript Escape. *PLOS ONE*, 6(5):e19655, 2011. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0019655. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019655.

[81] Ella Hartenian, Aaron S. Mendez, Allison L. Didychuk, Shivani Khosla, and Britt A. Glaunsinger. DNA processing by the Kaposi's sarcoma-associated herpesvirus alkaline exonuclease SOX contributes to viral gene expression and infectious virion production. *Nucleic Acids Research*, 51(1):182–197, 1 2023. ISSN 0305-1048. doi: 10.1093/NAR/GKAC1190. URL https://academic.oup.com/nar/article/51/1/182/6947079.

[82] Zoe H Davis, Nevan J Krogan, Britt A Glaunsinger, Erik Verschueren, Gwendolyn M Jang, Kevin Kleffman, Jeffrey R Johnson, Jimin Park, John Von Dollen, M Cyrus Maher, Tasha Johnson, William Newton, Stefanie Jä, Michael Shales, Julie Horner, and Ryan D Hernandez. Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes Molecular Cell Resource Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes. *Molecular Cell*, 57:349–360, 2015. doi: 10.1016/j.molcel.2014.11.026. URL http://dx.doi.org/10.1016/j.molcel.2014.11.026http://dx.doi.org/10.1016/j.molcel.2014.11.026.

[83] Ren Sun, Su-Fang Lin, Katherine Staskus, Lyndle Gradoville, Elizabeth Grogan, Ashley Haase, and George Miller. Kinetics of Kaposi's sarcoma-associated herpesvirus gene expression. *Journal of virology*, 73(3):2232–2242, 3 1999. ISSN 0022-538X. doi: 10.1128/JVI.73.3.2232-2242.1999. URL https://pubmed.ncbi.nlm.nih.gov/9971806/.

[84] Hiroyuki Nakamura, Michael Lu, Yousang Gwack, John Souvlis, Steven L. Zeichner, and Jae U. Jung. Global Changes in Kaposi's Sarcoma-Associated Virus Gene Expression Patterns following Expression of a Tetracycline-Inducible Rta Transactivator. *Journal of Virology*, 77(7):4205, 4 2003. ISSN 0022-538X. doi: 10.1128/JVI.77.7.4205-4220.2003. URL /pmc/articles/PMC150665//pmc/articles/PMC150665/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC150665/.

[85] Kyle L. Jung, Un Yung Choi, Angela Park, Suan Sin Foo, Stephanie Kim, Shin Ae Lee, and Jae U. Jung. Single-cell analysis of Kaposi's sarcoma-associated herpesvirus infection in three-dimensional air-liquid interface culture model. *PLOS Pathogens*, 18(8):e1010775, 8 2022. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1010775. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1010775.

[86] Hoang Van Phan, Michiel van Gent, Nir Drayman, Anindita Basu, Michaela U. Gack, and Savaş Tay. High-throughput RNA sequencing of paraformaldehyde-fixed single cells. *Nature Communications 2021 12:1*, 12(1):1–11, 9 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-25871-2. URL https://www.nature.com/articles/s41467-021-25871-2.

[87] Ezequiel Lacunza, Anuj Ahuja, Omar A. Coso, Martin Abba, Enrique Mesri, and Julian Naipauer. Single-cell RNA sequencing identifies KSHV-infected Mesenchymal stem cell subpopulations precursors of Kaposi's Sarcoma. *bioRxiv*, page 2023.02.28.530387, 3 2023. doi: 10.1101/2023.02.28.530387. URL https://www.biorxiv.org/content/10.1101/2023.02.28.530387v1https://www.biorxiv.org/content/10.1101/2023.02.28.530387v1.abstract.

[88] Justin T. Landis, Ryan Tuck, Yue Pan, Carson N. Mosso, Anthony B. Eason, Razia Moorad, J. Stephen Marron, and Dirk P. Dittmer. Evidence for Multiple Subpopulations of Herpesvirus-Latently Infected Cells. *mBio*, 13(1), 2 2022. ISSN 21507511. doi: 10.1128/MBIO.03473-21/SUPPL{_}FILE/MBIO.03473-21-SF001.TIF. URL https://journals.asm.org/doi/10.1128/mbio.03473-21.

[89] Tate Tabtieng, Rachel C. Lent, Machika Kaku, Alvaro Monago Sanchez, and Marta Maria Gaglia. Caspase-Mediated Regulation and Cellular Heterogeneity of the cGAS/STING Pathway in Kaposi's Sarcoma-Associated Herpesvirus Infection. *mBio*, 13(6), 12 2022. ISSN 21507511. doi: 10.1128/MBIO.02446-22/SUPPL{_}FILE/MBIO.02446-22-ST003.PDF. URL https://journals.asm.org/doi/10.1128/mbio.02446-22.

[90] Min Xue, Shuihong Yao, Minmin Hu, Wan Li, Tingting Hao, Feng Zhou, Xiaofei Zhu, Hongmei Lu, Di Qin, Qin Yan, Jianzhong Zhu, Shou Jiang Gao, and Chun Lu. HIV-1 Nef and KSHV oncogene K1 synergistically promote angiogenesis by inducing cellular miR-718 to regulate the PTEN/AKT/mTOR signaling pathway. *Nucleic Acids Research*, 42(15):9862–9879, 9 2014. ISSN 0305-1048. doi: 10.1093/NAR/GKU583. URL https://academic.oup.com/nar/article/42/15/9862/2434544.

[91] Wan-Shan Yang, Ting-Yu Lin, Lung Chang, Wayne W. Yeh, Shih-Ching Huang, Tung-Ying Chen, Yi-Ta Hsieh, Szu-Ting Chen, Wan-Chun Li, Chin-Chen Pan, Mel Campbell, Chia-Hung Yen, Yi-Ming Arthur Chen, and Pei-Ching Chang. HIV-1 Tat Interacts with a Kaposi's Sarcoma-Associated Herpesvirus Reactivation-Upregulated Antiangiogenic Long Noncoding RNA, LINC00313, and Antagonizes Its Function. *Journal of Virology*, 94(3):1280–1299, 1 2020. ISSN 0022-538X. doi: 10.1128/JVI.01280-19. URL /pmc/articles/PMC7000985//pmc/articles/PMC7000985/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7000985/.

[92] X. Zhu, Y. Guo, S. Yao, Q. Yan, M. Xue, T. Hao, F. Zhou, J. Zhu, D. Qin, and C. Lu. Synergy between Kaposi's sarcoma-associated herpesvirus (KSHV) vIL-6 and HIV-1 Nef protein in promotion of angiogenesis and oncogenesis: role of the AKT signaling pathway. *Oncogene 2014 33:15*, 33(15):1986–1996, 4 2013. ISSN 1476-5594. doi: 10.1038/onc.2013.136. URL https://www.nature.com/articles/onc2013136.

[93] For Yue Tso, Andrew V. Kossenkov, Salum J. Lidenge, Owen Ngalamika, John R. Ngowi, Julius Mwaiselage, Jayamanna Wickramasinghe, Eun Hee Kwon, John T. West, Paul M. Lieberman, and Charles Wood. RNA-Seq of Kaposi's sarcoma reveals alterations in glucose and lipid metabolism. *PLOS Pathogens*, 14(1):e1006844, 1 2018. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1006844. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1006844.

[94] Ramya Ramaswami, Takanobu Tagawa, Guruswamy Mahesh, Anna Serquina, Vishal Koparde, Kathryn Lurain, Sarah Dremel, Xiaofan Li, Ameera Mungale, Alex Beran, Zoe Weaver Ohler, Laura Bassel, Andrew Warner, Ralph Mangusan, Anaida Widell, Irene Ekwede, Laurie T. Krug, Thomas S. Uldrick, Robert Yarchoan, and Joseph M. Ziegelbauer. Transcriptional analysis identifies overlapping and tissue-distinct profiles between Kaposi sarcoma tumors of the skin and gastrointestinal tract. *bioRxiv*, page 2022.03.18.484923, 3 2022. doi: 10.1101/2022.03.18.484923. URL https://www.biorxiv.org/content/10.1101/2022.03.18.484923v1https://www.biorxiv.org/content/10.1101/2022.03.18.484923v1.abstract.

[95] Emmy W. Verschuren, Nic Jones, and Gerard I. Evan. The cell cycle and how it is steered by Kaposi's sarcoma-associated herpesvirus cyclin. *Journal of General Virology*, 85(6):1347–1361, 6 2004. ISSN 00221317. doi: 10.1099/VIR.0.79812-0/CITE/REFWORKS. URL https://www.microbiologyresearch.org/content/journal/jgv/10.1099/vir.0.79812-0.

[96] Gianna Ballon, Kang Chen, Rocio Perez, Wayne Tam, and Ethel Cesarman. Kaposi sarcoma herpesvirus (KSHV) vFLIP oncoprotein induces B cell transdifferentiation and tumorigenesis in mice. *The Journal of Clinical Investigation*, 121(3):1141–1153, 3 2011. ISSN 0021-9738. doi: 10.1172/JCI44417. URL http://www.jci.org.

[97] Farnaz D. Fakhari, Joseph H. Jeong, Yogita Kanan, and Dirk P. Dittmer. The latency-associated nuclear antigen of Kaposi sarcoma–associated herpesvirus induces B cell hyperplasia and lymphoma. *The Journal of Clinical Investigation*, 116(3):735–742, 3 2006. ISSN 0021-9738. doi: 10.1172/JCI26190. URL http://www.jci.org.

[98] Stoyan A. Radkov, Paul Kellam, and Chris Boshoff. The latent nuclear antigen of Kaposi sarcoma-associated herpesvirus targets the retinoblastoma-E2F pathway and with the oncogene Hras transforms primary rat cells. *Nature medicine*, 6(10):1121–1127, 2000. ISSN 1078-8956. doi: 10.1038/80459. URL https://pubmed.ncbi.nlm.nih.gov/11017143/.

[99] Silvia Gramolelli and Thomas F. Schulz. The role of Kaposi sarcoma-associated herpesvirus in the pathogenesis of Kaposi sarcoma. *The Journal of Pathology*, 235(2):368–380, 1 2015. ISSN 1096-9896. doi: 10.1002/PATH.4441. URL https://onlinelibrary.wiley.com/doi/full/10.1002/path.4441https://onlinelibrary.wiley.com/doi/abs/10.1002/path.4441https://pathsocjournals.onlinelibrary.wiley.com/doi/10.1002/path.4441.

[100] Ying Fan, Sumana Sanyal, and Roberto Bruzzone. Breaking Bad: How Viruses Subvert the Cell Cycle. *Frontiers in Cellular and Infection Microbiology*, 8: 396, 11 2018. ISSN 22352988. doi: 10.3389/FCIMB.2018.00396/BIBTEX.

[101] Pirita Pekkonen, Annika Järviluoma, Nadezhda Zinovkina, Anna Cvrljevic, Sonam Prakash, Jukka Westermarck, Gerard Evan, Ethel Cesarman, Emmy W. Verschuren, and Päivi Ojala. KSHV viral cyclin interferes with T-cell development and induces lymphoma through Cdk6 and Notch activation in vivo. *Cell Cycle*, 13(23):3670, 12 2014. ISSN 15514005. doi: 10.4161/15384101.2014.964118. URL /pmc/articles/PMC4613844//pmc/articles/PMC4613844/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4613844/.

[102] Andrew M. Leidal, Patrick W.K. Lee, and Craig McCormick. Viral subversion of autophagy impairs oncogene-induced senescence. *Autophagy*, 8(7):1138, 7 2012. ISSN 15548635. doi: 10.4161/AUTO.20340. URL /pmc/articles/PMC3429550//pmc/articles/PMC3429550/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3429550/.

[103] Sumitra Muralidhar, Gary Veytsmann, Bala Chandran, Dharam Ablashi, Jay Doniger, and Leonard J. Rosenthal. Characterization of the human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus) oncogene, Kaposin (ORF K12). *Journal of Clinical Virology*, 16(3):203–213, 5 2000. ISSN 1386-6532. doi: 10.1016/S1386-6532(99)00081-5.

[104] Eleonora Forte, Archana N. Raja, Priscilla Shamulailatpam, Mark Manzano, Matthew J. Schipma, John L. Casey, and Eva Gottwein. MicroRNA-Mediated Transformation by the Kaposi's Sarcoma-Associated Herpesvirus Kaposin Locus. *Journal of Virology*, 89(4):2333, 2 2015. ISSN 0022-538X. doi: 10.1128/JVI.03317-14. URL /pmc/articles/PMC4338870//pmc/articles/PMC4338870/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4338870/.

[105] Craig McCormick and Don Ganem. The kaposin B protein of KSHV activates the p38/MK2 pathway and stabilizes cytokine mRNAs. *Science*, 307 (5710):739–741, 2 2005. ISSN 00368075. doi: 10.1126/SCIENCE.1105779/SUPPL{_}FILE/MCCORMICK.SOM.PDF. URL https://www.science.org/doi/10.1126/science.1105779.

[106] Craig McCormick and Don Ganem. Phosphorylation and Function of the Kaposin B Direct Repeats of Kaposi's Sarcoma-Associated Herpesvirus. *Journal of Virology*, 80(12):6165, 6 2006. ISSN 0022-538X. doi: 10.1128/JVI.02331-05. URL /pmc/articles/PMC1472581//pmc/articles/PMC1472581/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1472581/.

[107] Jaehyuk Yoo, Ha Neul Lee, Inho Choi, Dongwon Choi, Hee Kyoung Chung, Kyu Eui Kim, Sunju Lee, Berenice Aguilar, Jinjoo Kang, Eunkyung Park, Yong Suk Lee, Yong Sun Maeng, Nam Yoon Kim, Chester J. Koh, and Young Kwon Hong. Opposing Regulation of PROX1 by Interleukin-3 Receptor and NOTCH Directs Differential Host Cell Fate Reprogramming by Kaposi Sarcoma Herpes Virus. *PLOS Pathogens*, 8(6):e1002770, 6 2012. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1002770. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1002770.

[108] Jennifer A. Corcoran, Benjamin P. Johnston, and Craig McCormick. Viral Activation of MK2-hsp27-p115RhoGEF-RhoA Signaling Axis Causes Cytoskeletal Rearrangements, P-body Disruption and ARE-mRNA Stabilization. *PLOS Pathogens*, 11(1):e1004597, 2015. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1004597. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1004597.

[109] Andrea J. O'Hara, Pauline Chugh, Ling Wang, Eduardo M. Netto, Estrella Luz, William J. Harrington, Bruce J. Dezube, Blossom Damania, and Dirk P. Dittmer. Pre-Micro RNA Signatures Delineate Stages of Endothelial Cell Transformation in Kaposi Sarcoma. *PLOS Pathogens*, 5(4):e1000389, 4 2009. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1000389. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1000389.

[110] Vickie Marshall, Thomas Parks, Rachel Bagni, Cheng Dian Wang, Mark A. Samols, Jianhong Hu, Kathleen M. Wyvil, Karen Aleman, Richard F. Little, Robert Yarchoan, Rolf Renne, and Denise Whitby. Conservation of Virally Encoded MicroRNAs in Kaposi Sarcoma-Associated Herpesvirus in Primary Effusion Lymphoma Cell Lines and in Patients with Kaposi Sarcoma or Multicentric Castleman Disease. *The Journal of Infectious Diseases*, 195(5):645–659, 3 2007. ISSN 0022-1899. doi: 10.1086/511434. URL https://academic.oup.com/jid/article/195/5/645/842338.

[111] Xiufen Lei, Ying Zhu, Tiffany Jones, Zhiqiang Bai, Yufei Huang, and Shou-Jiang Gao. A Kaposi's Sarcoma-Associated Herpesvirus MicroRNA and Its Variants Target the Transforming Growth Factor β Pathway To Promote Cell Survival. *Journal of Virology*, 86(21):11698, 11 2012. ISSN 0022-538X. doi: 10.1128/JVI.06855-11. URL /pmc/articles/PMC3486299//pmc/articles/PMC3486299/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3486299/.

[112] Johanna R. Abend, Dhivya Ramalingam, Philippe Kieffer-Kwon, Thomas S. Uldrick, Robert Yarchoan, and Joseph M. Ziegelbauer. Kaposi's Sarcoma-Associated Herpesvirus MicroRNAs Target IRAK1 and MYD88, Two Components of the Toll-Like Receptor/Interleukin-1R Signaling Cascade, To Reduce Inflammatory-Cytokine Expression. *Journal of Virology*, 86(21):11663, 11 2012. ISSN 0022-538X. doi: 10.1128/JVI.01147-12. URL /pmc/articles/PMC3486292//pmc/articles/PMC3486292/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3486292/.

[113] Guillaume Suffert, Georg Malterer, Jean Hausser, Johanna Viiliäinen, Aurélie Fender, Maud Contrant, Tomi Ivacevic, Vladimir Benes, Frédéric Gros, Olivier Voinnet, Mihaela Zavolan, Päivi M. Ojala, Juergen G. Haas, and Sébastien Pfeffer. Kaposi's Sarcoma Herpesvirus microRNAs Target Caspase 3 and Regulate Apoptosis. *PLOS Pathogens*, 7(12):e1002405, 12 2011. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1002405. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1002405.

[114] Heuiran Lee, Ronald Veazey, Kenneth Williams, Mengtao Li, Jie Guo, Frank Neipel, Bernhard Fleckenstein, Andrew Lackner, Ronald C. Desrosiers, and Jae U. Jung. Deregulation of cell growth by the K1 gene of Karposi's sarcoma-associated herpesvirus. *Nature Medicine 1998 4:4*, 4(4):435–440, 4 1998. ISSN 1546-170X. doi: 10.1038/nm0498-435. URL https://www.nature.com/articles/nm0498-435.

[115] Om Prakash, Zhen Ya Tang, Xiaochang Peng, Roy Coleman, Javed Gill, Gist Farr, and Felipe Samaniego. Tumorigenesis and Aberrant Signaling in Transgenic Mice Expressing the Human Herpesvirus-8 K1 Gene. *JNCI: Journal of the National Cancer Institute*, 94(12):926–935, 6 2002. ISSN 0027-8874. doi: 10.1093/JNCI/94.12.926. URL https://academic.oup.com/jnci/article/94/12/926/2519781.

[116] Yoshiyasu Aoki, Elaine S. Jaffe, Yuan Chang, Karen Jones, Julie Teruya-Feldstein, Patrick S. Moore, and Giovanna Tosato. Angiogenesis and Hematopoiesis Induced by Kaposi's Sarcoma-Associated Herpesvirus-Encoded Interleukin-6: Presented in part at the 40th Annual American Society of Hematology Meeting, December 7, 1998 (Miami Beach, FL). *Blood*, 93(12):4034–4043, 6 1999. ISSN 0006-4971. doi: 10.1182/BLOOD.V93.12.4034.

[117] Silvia Montaner, Akrit Sodhi, Alfredo Molinolo, Thomas H. Bugge, Earl T. Sawai, Yunsheng He, Yi Li, Patricio E. Ray, and J. Silvio Gutkind. Endothelial infection with KSHV genes in vivo reveals that vGPCR initiates Kaposi's sarcomagenesis and can promote the tumorigenic potential of viral latent genes. *Cancer Cell*, 3(1):23–36, 1 2003. ISSN 15356108. doi: 10.1016/S1535-6108(02)00237-4. URL http://www.cell.com/article/S1535610802002374/fulltexthttp://www.cell.com/article/S1535610802002374/abstracthttps://www.cell.com/cancer-cell/abstract/S1535-6108(02)00237-4.

[118] Characterization of Signaling Cascades Triggered by Human Interleukin-6 versus Kaposi's Sarcoma-associated Herpes Virus-encoded Viral Interleukin 61 | Clinical Cancer Research | American Association for Cancer Research. URL https://aacrjournals.org/clincancerres/article/6/3/1180/288114/Characterization-of-Signaling-Cascades-Triggered.

[119] Giuseppe Balistreri, Johanna Viiliäinen, Mikko Turunen, Raquel Diaz, Lauri Lyly, Pirita Pekkonen, Juha Rantala, Krista Ojala, Grzegorz Sarek, Mari Teesalu, Oxana Denisova, Karita Peltonen, Ilkka Julkunen, Markku Varjosalo, Denis Kainov, Olli Kallioniemi, Marikki Laiho, Jussi Taipale, Sampsa Hautaniemi, and Päivi M. Ojala. Oncogenic Herpesvirus Utilizes Stress-Induced Cell Cycle Checkpoints for Efficient Lytic Replication. *PLOS Pathogens*, 12(2):e1005424, 2 2016. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1005424. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1005424.

[120] Lyubomir A. Dourmishev, Assen L. Dourmishev, Diana Palmeri, Robert A. Schwartz, and David M. Lukac. Molecular Genetics of Kaposi's Sarcoma-Associated Herpesvirus (Human Herpesvirus 8) Epidemiology and Pathogenesis. *Microbiology and Molecular Biology Reviews*, 67(2):175–212, 6 2003. ISSN 1092-2172. doi: 10.1128/MMBR.67.2.175-212.2003/ASSET/BA084FE0-79CE-4AC8-B34B-45452FE692BE/ASSETS/GRAPHIC/MR0230008003.JPEG. URL https://journals.asm.org/doi/10.1128/MMBR.67.2.175-212.2003.

[121] Dolores M. Ciufo, Jennifer S. Cannon, Lynn J. Poole, Frederick Y. Wu, Paul Murray, Richard F. Ambinder, and Gary S. Hayward. Spindle cell conversion by Kaposi's sarcoma-associated herpesvirus: formation of colonies and plaques with mixed lytic and latent gene expression in infected primary dermal microvascular endothelial cell cultures. *Journal of virology*, 75(12):5614–5626, 6 2001. ISSN 0022-538X. doi: 10.1128/JVI.75.12.5614-5626.2001. URL https://pubmed.ncbi.nlm.nih.gov/11356969/.

[122] Eriko Ohsaki and Keiji Ueda. Interplay Between KSHV and the Host DNA Damage Response. *Frontiers in Cellular and Infection Microbiology*, 10, 12 2020. ISSN 22352988. doi: 10.3389/FCIMB.2020.604351. URL /pmc/articles/PMC7793933//pmc/articles/PMC7793933/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7793933/.

[123] Barbara A. Scholz, Marie L. Harth-Hertle, Georg Malterer, Juergen Haas, Joachim Ellwart, Thomas F. Schulz, and Bettina Kempkes. Abortive Lytic Reactivation of KSHV in CBF1/CSL Deficient Human B Cell Lines. *PLOS Pathogens*, 9(5):e1003336, 5 2013. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1003336. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003336.

[124] Magdalena Angelova, Mary Beth Ferris, Kenneth F. Swan, Harris E. McFerrin, Gabriella Pridjian, Cindy A. Morris, and Deborah E. Sullivan. Kaposi's sarcoma-associated herpesvirus G-protein coupled receptor activates the canonical Wnt/β-catenin signaling pathway. *Virology Journal*, 11(1):1–10, 12 2014. ISSN 1743422X. doi: 10.1186/S12985-014-0218-8/FIGURES/5. URL https://virologyj.biomedcentral.com/articles/10.1186/s12985-014-0218-8.

[125] Claudia J. Krause, Oliver Popp, Nanthakumar Thirunarayanan, Gunnar Dittmar, Martin Lipp, and Gerd Müller. MicroRNA-34a promotes genomic instability by a broad suppression of genome maintenance mechanisms downstream of the oncogene KSHV-vGPCR. *Oncotarget*, 7(9):10414, 3 2016. ISSN 1949-2553. doi: 10.18632/ONCOTARGET.7248. URL /pmc/articles/PMC4891129//pmc/articles/PMC4891129/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4891129/.

[126] Naira Samarina, George Ssebyatika, Tanvi Tikla, Ja Yun Waldmann, Bizunesh Abere, Vittoria Nanna, Michelangelo Marasco, Teresa Carlomagno, Thomas Krey, and Thomas F. Schulz. Recruitment of phospholipase Cγ1 to the non-structural membrane protein pK15 of Kaposi Sarcoma-associated herpesvirus promotes its Src-dependent phosphorylation. *PLOS Pathogens*, 17(6):e1009635, 6 2021. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1009635. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1009635.

[127] Anika Hävemeier, Silvia Gramolelli, Marcel Pietrek, Ramona Jochmann, Michael Stürzl, and Thomas F. Schulz. Activation of NF-κB by the Kaposi's Sarcoma-Associated Herpesvirus K15 Protein Involves Recruitment of the NF-κB-Inducing Kinase, IκB Kinases, and Phosphorylation of p65. *Journal of Virology*, 88(22):13161, 11 2014. ISSN 0022-538X. doi: 10.1128/JVI.01766-14. URL /pmc/articles/PMC4249085//pmc/articles/PMC4249085/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4249085/.

[128] Nam Hyuk Cho, Young Ki Choi, and Joong Kook Choi. Multi-transmembrane protein K15 of Kaposi's sarcoma-associated herpesvirus targets Lyn kinase in the membrane raft and induces NFAT/AP1 activities. *Experimental & Molecular Medicine 2008 40:5*, 40(5):565–573, 10 2008. ISSN 2092-6413. doi: 10.3858/emm.2008.40.5.565. URL https://www.nature.com/articles/emm200865.

[129] Wei Chen, Changqing Xu, Liuqing Wang, Bing Shen, and Linding Wang. K15 Protein of Kaposi's Sarcoma Herpesviruses Increases Endothelial Cell Proliferation and Migration through Store-Operated Calcium Entry. *Viruses*, 10(6), 6 2018. ISSN 19994915. doi: 10.3390/V10060282. URL /pmc/articles/PMC6024707//pmc/articles/PMC6024707/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6024707/.

[130] Julie Osborne, Patrick S. Moore, and Yuan Chang. KSHV-encoded viral IL-6 activates multiple human IL-6 signaling pathways. *Human Immunology*, 60 (10):921–927, 10 1999. ISSN 0198-8859. doi: 10.1016/S0198-8859(99)00083-X.

[131] Bok Soo Lee, Xavier Alvarez, Satoshi Ishido, Andrew A. Lackner, and Jae U. Jung. Inhibition of intracellular transport of B cell antigen receptor complexes by Kaposi's sarcoma-associated herpesvirus K1. *The Journal of experimental medicine*, 192(1):11–21, 7 2000. ISSN 0022-1007. doi: 10.1084/JEM.192.1.11. URL https://pubmed.ncbi.nlm.nih.gov/10880522/.

[132] Andrés Bueno Venegas, Toyoaki Natsume, Masato Kanemaki, and Ian D. Hickson. Inducible Degradation of the Human SMC5/6 Complex Reveals an Essential Role Only during Interphase. *Cell reports*, 31(3), 4 2020. ISSN 2211-1247. doi: 10.1016/J.CELREP.2020.107533. URL https://pubmed.ncbi.nlm.nih.gov/32320646/.

[133] Stephanie Pei Tung Yiu, Rui Guo, Cassie Zerbe, Michael P. Weekes, and Benjamin E. Gewurz. Epstein-Barr virus BNRF1 destabilizes SMC5/6 cohesin complexes to evade its restriction of replication compartments. *Cell reports*, 38(10):110411, 3 2022. ISSN 22111247. doi: 10.1016/J.CELREP.2022.110411. URL /pmc/articles/PMC8981113//pmc/articles/PMC8981113/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8981113/.

[134] Taegun Seo, Daeyoup Lee, Young Sam Shim, Jon E. Angell, Natesa V. Chidambaram, Dhananjaya V. Kalvakolanu, and Joonho Choe. Viral Interferon Regulatory Factor 1 of Kaposi's Sarcoma-Associated Herpesvirus Interacts with a Cell Death Regulator, GRIM19, and Inhibits Interferon/Retinoic Acid-Induced Cell Death. *Journal of Virology*, 76(17):8797, 9 2002. ISSN 0022-538X. doi: 10.1128/JVI.76.17.8797-8807.2002. URL /pmc/articles/PMC136415//pmc/articles/PMC136415/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC136415/.

[135] Andy Karabajakian, Isabelle Ray-Coquard, and Jean Yves Blay. Molecular Mechanisms of Kaposi Sarcoma Development. *Cancers*, 14(8):1869, 4 2022. ISSN 20726694. doi: 10.3390/CANCERS14081869. URL /pmc/articles/PMC9030761//pmc/articles/PMC9030761/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9030761/.

[136] Hye Ra Lee, Kevin Brulois, Lai Yee Wong, and Jae U. Jung. Modulation of immune system by Kaposi's sarcoma-associated herpesvirus: Lessons from viral evasion strategies. *Frontiers in Microbiology*, 3(MAR):44, 3 2012. ISSN 1664302X. doi: 10.3389/FMICB.2012.00044/BIBTEX.

[137] Amanda de Oliveira Lopes, Pedro Do Nascimento Marinho, Letícia d'Ambrosio de Souza Medeiros, and Vanessa Salete de Paula. Human Gammaherpesvirus 8 Oncogenes Associated with Kaposi’s Sarcoma. *International Journal of Molecular Sciences 2022, Vol. 23, Page 7203*, 23(13): 7203, 6 2022. ISSN 1422-0067. doi: 10.3390/IJMS23137203. URL https://www.mdpi.com/1422-0067/23/13/7203/htmhttps://www.mdpi.com/1422-0067/23/13/7203.

[138] Yuan Chang and Patrick Moore. Twenty Years of KSHV. *Viruses*, 6(11):4258, 11 2014. ISSN 19994915. doi: 10.3390/V6114258. URL /pmc/articles/PMC4246220//pmc/articles/PMC4246220/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4246220/.

[139] Yuan Chang, Ethel Cesarman, Melissa S. Pessin, Frank Lee, Janice Culpepper, Daniel M. Knowles, and Patrick S. Moore. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science (New York, N.Y.)*, 266(5192):1865–1869, 1994. ISSN 0036-8075. doi: 10.1126/SCIENCE.7997879. URL https://pubmed.ncbi.nlm.nih.gov/7997879/.

[140] Perla El Zeinaty, Céleste Lebbé, and Julie Delyon. Endemic Kaposi’s Sarcoma. *Cancers 2023, Vol. 15, Page 872*, 15(3):872, 1 2023. ISSN 2072-6694. doi: 10.3390/CANCERS15030872. URL https://www.mdpi.com/2072-6694/15/3/872/htmhttps://www.mdpi.com/2072-6694/15/3/872.

[141] Janet L. Douglas, Jean K. Gustin, Ashlee V. Moses, Bruce J. Dezube, and Liron Pantanowitz. Kaposi Sarcoma Pathogenesis: A Triad of Viral Infection, Oncogenesis and Chronic Inflammation. *Translational biomedicine*, 1(2), 2010. ISSN 21720479. doi: 10.3823/409. URL /pmc/articles/PMC3472629//pmc/articles/PMC3472629/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3472629/.

[142] M. R. Bongiorno, S. Doukaki, G. Ferro, and M. Aricò. Matrix metalloproteinases 2 and 9, and extracellular matrix in Kaposi's sarcoma. *Dermatologic Therapy*, 23(SUPPL. 2):S33–S36, 3 2010. ISSN 1529-8019. doi: 10.1111/J.1529-8019.2010.01297.X. URL https://onlinelibrary.wiley.com/doi/full/10.1111/j.1529-8019.2010.01297.xhttps://onlinelibrary.wiley.com/doi/abs/10.1111/j.1529-8019.2010.01297.xhttps://onlinelibrary.wiley.com/doi/10.1111/j.1529-8019.2010.01297.x.

[143] Lucia Brambilla, Giovanni Genovese, Emilio Berti, Ketty Peris, Franco Rongioletti, Giuseppe Micali, Fabio Ayala, Silvia Della Bella, Roberta Mancuso, Piergiacomo Calzavara Pinton, and Athanasia Tourlaki. Diagnosis and treatment of classic and iatrogenic Kaposi's sarcoma: Italian recommendations. *Italian journal of dermatology and venereology*, 156(3):356–365, 6 2021. ISSN 2784-8450. doi: 10.23736/S2784-8671.20.06703-6. URL https://pubmed.ncbi.nlm.nih.gov/33179877/.

[144] Ramya Vangipuram and Stephen K. Tyring. Epidemiology of Kaposi sarcoma: review and description of the nonepidemic variant. *International Journal of Dermatology*, 58(5):538–542, 5 2019. ISSN 1365-4632. doi: 10.1111/IJD.14080. URL https://onlinelibrary.wiley.com/doi/full/10.1111/ijd.14080https://onlinelibrary.wiley.com/doi/abs/10.1111/ijd.14080https://onlinelibrary.wiley.com/doi/10.1111/ijd.14080.

[145] Lijun Yan, Vladimir Majerciak, Zhi-Ming Zheng, and Ke Lan. Towards Better Understanding of KSHV Life Cycle: from Transcription and Posttranscriptional Regulations to Pathogenesis. *Virologica Sinica 2019 34:2*, 34(2):135–161, 4 2019. ISSN 1995-820X. doi: 10.1007/S12250-019-00114-3. URL https://link.springer.com/article/10.1007/s12250-019-00114-3.

[146] Eliane Rohner, Natascha Wyss, Zina Heg, Zully Faralli, Sam M. Mbulaiteye, Urban Novak, Marcel Zwahlen, Matthias Egger, and Julia Bohlius. HIV and human herpesvirus 8 co-infection across the globe: systematic review and meta-analysis. *International journal of cancer. Journal international du cancer*, 138(1):45, 1 2016. ISSN 10970215. doi: 10.1002/IJC.29687. URL /pmc/articles/PMC4607648//pmc/articles/PMC4607648/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4607648/.

[147] Rodgers N. Demba, Sylviah M. Aradi, Matilu Mwau, and Walter O. Mwanda. Kaposi's sarcoma-associated herpesvirus protein ORF75 among HIV-1 patients in Kenya. *African Journal of Laboratory Medicine*, 9(1):2225–2002, 2020. ISSN 22252010. doi: 10.4102/AJLM.V9I1.939. URL /pmc/articles/PMC7479412//pmc/articles/PMC7479412/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7479412/.

[148] Mina C. Hosseinipour, Kristen M. Sweet, Jie Xiong, Dan Namarika, Albert Mwafongo, Michael Nyirenda, Loreen Chiwoko, Deborah Kamwendo, Irving Hoffman, Jeannette Lee, Sam Phiri, Wolfgang Vahrson, Blossom Damania, and Dirk P. Dittmer. Viral profiling identifies multiple subtypes of Kaposi's sarcoma. *mBio*, 5(5):1633–1647, 9 2014. ISSN 21507511. doi: 10.1128/MBIO.01633-14/SUPPL{_}FILE/MBO005141991S1.DOCX. URL https://journals.asm.org/doi/10.1128/mBio.01633-14.

[149] Thierry Simonart. Role of environmental factors in the pathogenesis of classic and African-endemic Kaposi sarcoma. *Cancer Letters*, 244(1):1–7, 11 2006. ISSN 03043835. doi: 10.1016/J.CANLET.2006.02.005.

[150] John L. Ziegler, Edward Katongole-Mbidde, Henry Wabinga, and Charles M. Dollbaum. Absence of sex-hormone receptors in Kaposi's sarcoma. *The Lancet*, 345(8954):925, 4 1995. ISSN 01406736. doi: 10.1016/S0140-6736(95)90039-X. URL http://www.thelancet.com/article/S014067369590039X/fulltexthttp://www.thelancet.com/article/S014067369590039X/abstracthttps://www.thelancet.com/journals/lancet/article/PIIS0140-6736(95)90039-X/abstract.

[151] Yanto Lunardi-Lskandar, Joseph L. Bryant, Robert A. Zeman, Victor H. Lam, Felipe Samaniego, Jacques M. Besnier, Philippe Hermans, Alain R. Thierry, Parkash Gill, and Robert C. Gallo. Tumorigenesis and metastasis of neoplastic Kaposi's sarcoma cell line in immunodeficient mice blocked by a human pregnancy hormone. *Nature 1995 375:6526*, 375(6526):64–68, 1995. ISSN 1476-4687. doi: 10.1038/375064a0. URL https://www.nature.com/articles/375064a0.

[152] Sophie Grabar and Dominique Costagliola. Epidemiology of Kaposi's Sarcoma. *Cancers 2021, Vol. 13, Page 5692*, 13(22):5692, 11 2021. ISSN 2072-6694. doi: 10.3390/CANCERS13225692. URL https://www.mdpi.com/2072-6694/13/22/5692/htmhttps://www.mdpi.com/2072-6694/13/22/5692.

[153] Yuqing Li, Canrong Zhong, Dawei Liu, Wenjing Yu, Weikang Chen, Yan Wang, Songtao Shi, and Yan Yuan. Evidence For Kaposi's Sarcoma Originating From Mesenchymal Stem Cell Through KSHV-induced Mesenchymal-to-Endothelial Transition. *Cancer research*, 78(1):230, 1 2018. ISSN 15387445. doi: 10.1158/0008-5472.CAN-17-1961. URL /pmc/articles/PMC5754241//pmc/articles/PMC5754241/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5754241/.

[154] Hsei-Wei Wang, Matthew W B Trotter, Dimitrios Lagos, Dimitra Bourboulia, Stephen Henderson, Ta Ija Mäkinen, Stephen Elliman, Adrienne M Flanagan, Kari Alitalo, and Chris Boshoff. Kaposi sarcoma herpesvirus-induced cellular reprogramming contributes to the lymphatic endothelial gene expression in Kaposi sarcoma. *NATURE GENETICS VOLUME*, 36(7), 2004. doi: 10.1038/ng1384. URL http://www.nature.com/naturegenetics.

[155] Patrick A. Carroll, Elizabeth Brazeau, and Michael Lagunoff. Kaposi's sarcoma-associated herpesvirus infection of blood endothelial cells induces lymphatic differentiation. *Virology*, 328(1):7, 10 2004. ISSN 00426822. doi: 10.1016/J.VIROL.2004.07.008. URL /pmc/articles/PMC3147029//pmc/articles/PMC3147029/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3147029/.

[156] Heather Lindsey. Kaposi sarcoma cells related to lymphatic endothelial cells. *The Lancet Oncology*, 5(8):464, 8 2004. ISSN 1470-2045. doi: 10.1016/S1470-2045(04)01542-6.

[157] Yuhan Liu, Mengting Ding, Qunjun Gao, Anbang He, Yuchen Liu, and Hongbing Mei. Current Advances on the Important Roles of Enhancer RNAs in Gene Regulation and Cancer. *BioMed Research International*, 2018:1–6, 5 2018. ISSN 2314-6133. doi: 10.1155/2018/2405351. URL https://www.hindawi.com/journals/bmri/2018/2405351/.

[158] Julian Naipauer, Daria Salyakina, Guy Journo, Santas Rosario, Sion Williams, Martin Abba, Meir Shamay, and Enrique A. Mesri. High-throughput sequencing analysis of a "hit and run" cell and animal model of KSHV tumorigenesis. *PLOS Pathogens*, 16(6):e1008589, 6 2020. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1008589. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1008589.

[159] Terri A. DiMaio and Michael Lagunoff. KSHV induction of angiogenic and lymphangiogenic phenotypes. *Frontiers in Microbiology*, 3(MAR):102, 3 2012. ISSN 1664302X. doi: 10.3389/FMICB.2012.00102/BIBTEX.

[160] Pravinkumar Purushothaman, Timsy Uppal, Roni Sarkar, and Subhash C. Verma. KSHV-Mediated Angiogenesis in Tumor Progression. *Viruses*, 8 (7), 7 2016. ISSN 19994915. doi: 10.3390/V8070198. URL /pmc/articles/PMC4974533//pmc/articles/PMC4974533/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4974533/.

[161] Jun Jun Olsen, Sebastian öther Gee Pohl, Abhijeet Deshmukh, Malini Visweswaran, Natalie C. Ward, Frank Arfuso, Mark Agostino, and Arun Dharmarajan. The Role of Wnt Signalling in Angiogenesis. *The Clinical Biochemist Reviews*, 38(3):131, 2017. ISSN 18380212. URL /pmc/articles/PMC5759160//pmc/articles/PMC5759160/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5759160/.

[162] Chris Boshoff. Kaposi sarcoma, KSHV and lymphangiogenesis. *Experimental Dermatology*, 16(10):867–868, 10 2007. ISSN 1600-0625. doi: 10.1111/J.1600-0625.2007.00607{_}5.X. URL https://onlinelibrary.wiley.com/doi/full/10.1111/j.1600-0625.2007.00607_5.xhttps://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0625.2007.00607_5.xhttps://onlinelibrary.wiley.com/doi/10.1111/j.1600-0625.2007.00607_5.x.

[163] Steven A. Stacker, Steven P. Williams, Tara Karnezis, Ramin Shayan, Stephen B. Fox, and Marc G. Achen. Lymphangiogenesis and lymphatic vessel remodelling in cancer. *Nature Reviews Cancer 2014 14:3*, 14(3):159–172, 2 2014. ISSN 1474-1768. doi: 10.1038/nrc3677. URL https://www.nature.com/articles/nrc3677.

[164] Yasuhiro Yoshimatsu, Masato Morikawa, Yasuyuki Morishita, and Tatsuhiko Kodama. Ets family members induce lymphangiogenesis through physical and functional interaction with Prox1 Lymphatic endothelial-to-mesenchymal transition View project. *Article in Journal of Cell Science*, 2011. doi: 10.1242/jcs.083998. URL https://www.researchgate.net/publication/51539145.

[165] Ren Liu, Xiuqing Li, Anil Tulpule, Yue Zhou, Jeffrey S. Scehnet, Shaobing Zhang, Jong Soo Lee, Preet M. Chaudhary, Jae Jung, and Parkash S. Gill. KSHV-induced notch components render endothelial and mural cell characteristics and cell survival. *Blood*, 115(4):887, 1 2010. ISSN 15280020. doi: 10.1182/BLOOD-2009-08-236745. URL /pmc/articles/PMC2815507//pmc/articles/PMC2815507/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815507/.

[166] Deguang Liang, Hao Hu, Shasha Li, Jiazhen Dong, Xing Wang, Yuhan Wang, Li He, Zhiheng He, Yuan Gao, Shou Jiang Gao, and Ke Lan. Oncogenic Herpesvirus KSHV Hijacks BMP-Smad1-Id Signaling to Promote Tumorigenesis. *PLoS Pathogens*, 10(7), 2014. ISSN 15537374. doi: 10.1371/JOURNAL.PPAT.1004253. URL /pmc/articles/PMC4092152//pmc/articles/PMC4092152/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4092152/.

[167] Mohadeseh Kosari-Monfared, Novin Nikbakhsh, Sadegh Fattahi, Elham Ghadami, Mohammad Ranaei, Hassan Taheri, Fatemeh Amjadi-Moheb, Gholam A. Godazandeh, Shahryar Shafaei, Maryam Pilehchian-Langroudi, Ali Akbar Samadani, and Haleh Akhavan-Niaki. CTNNBIP1 downregulation is associated with tumor grade and viral infections in gastric adenocarcinoma. *Journal of Cellular Physiology*, 234(3):2895–2904, 3 2019. ISSN 1097-4652. doi: 10.1002/JCP.27106. URL https://onlinelibrary.wiley.com/doi/full/10.1002/jcp.27106https://onlinelibrary.wiley.com/doi/abs/10.1002/jcp.27106https://onlinelibrary.wiley.com/doi/10.1002/jcp.27106.

[168] Seho Cha, Myung Suk Kang, and Taegun Seo. KSHV vPK inhibits Wnt signaling via preventing interactions between β-catenin and TCF4. *Biochemical and Biophysical Research Communications*, 497(1):381–387, 2 2018. ISSN 0006-291X. doi: 10.1016/J.BBRC.2018.02.089.

[169] Wendy J. van Zuylen, William D. Rawlinson, and Caroline E. Ford. The Wnt pathway: a key network in cell signalling dysregulated by viruses. *Reviews in Medical Virology*, 26(5):340–355, 9 2016. ISSN 10991654. doi: 10.1002/RMV.1892.

[170] Yarely M. Salinas-Vera, Laurence A. Marchat, Dolores Gallardo-Rincón, Erika Ruiz-García, Horacio Astudillo-De la Vega, Raquel Echavarría-Zepeda, and César López-Camarillo. AngiomiRs: MicroRNAs driving angiogenesis in cancer (Review). *International journal of molecular medicine*, 43(2):657–670, 2 2019. ISSN 1791-244X. doi: 10.3892/IJMM.2018.4003. URL https://pubmed.ncbi.nlm.nih.gov/30483765/.

[171] Jiangtao Yu, Zhiyang Han, Ziquan Sun, Yue Wang, Ming Zheng, and Chunfang Song. LncRNA SLCO4A1-AS1 facilitates growth and metastasis of colorectal cancer through β-catenin-dependent Wnt pathway. *Journal of experimental & clinical cancer research : CR*, 37(1), 2018. ISSN 1756-9966. doi: 10.1186/S13046-018-0896-Y. URL https://pubmed.ncbi.nlm.nih.gov/30201010/.

[172] Wen Bi, Jiayu Huang, Chunlei Nie, Bo Liu, Guoqing He, Jihua Han, Rui Pang, Zhaoming Ding, Jin Xu, and Jiewu Zhang. CircRNA circRNA-102171 promotes papillary thyroid cancer progression through modulating CTNNBIP1-dependent activation of β-catenin pathway 06 Biological Sciences 0601 Biochemistry and Cell Biology. *Journal of Experimental and Clinical Cancer Research*, 37(1):1–9, 11 2018. ISSN 17569966. doi: 10.1186/S13046-018-0936-7/FIGURES/6. URL https://jeccr.biomedcentral.com/articles/10.1186/s13046-018-0936-7.

[173] Gabriela Rusu-zota, Oana Mădălina Manole, Cristina Galeş, Elena Porumb-andrese, Otilia Obadă, and Cezar Valentin Mocanu. Kaposi Sarcoma, a Trifecta of Pathogenic Mechanisms. *Diagnostics*, 12(5), 5 2022. ISSN 20754418. doi: 10.3390/DIAGNOSTICS12051242. URL /pmc/articles/PMC9140574//pmc/articles/PMC9140574/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9140574/.

[174] Fang Cheng, Pirita Pekkonen, Simonas Laurinavicius, Nami Sugiyama, Stephen Henderson, Thomas Günther, Ville Rantanen, Elisa Kaivanto, Mervi Aavikko, Grzegorz Sarek, Sampsa Hautaniemi, Peter Biberfeld, Lauri Aaltonen, Adam Grundhoff, Chris Boshoff, Kari Alitalo, Kaisa Lehti, and Päivi M. Ojala. KSHV-initiated notch activation leads to membrane-type-1 matrix metalloproteinase-dependent lymphatic endothelial-to-mesenchymal transition. *Cell Host and Microbe*, 10(6):577–590, 12 2011. ISSN 19313128. doi: 10.1016/j.chom.2011.10.011. URL http://www.cell.com/article/S1931312811003647/fulltexthttp://www.cell.com/article/S1931312811003647/abstracthttps://www.cell.com/cell-host-microbe/abstract/S1931-3128(11)00364-7.

[175] Owen Ngalamika and Sody Munsaka. Cells of the Innate and Adaptive Immune Systems in Kaposi's Sarcoma. *Journal of Immunology Research*, 2020, 2020. ISSN 23147156. doi: 10.1155/2020/8852221.

[176] C. Simonelli, R. Tedeschi, A. Gloghini, R. Talamini, M. T. Bortolin, M. Berretta, M. Spina, S. Morassut, E. Vaccher, P. De Paoli, A. Carbone, and U. Tirelli. Plasma HHV-8 viral load in HHV-8-related lymphoproliferative disorders associated with HIV infection. *Journal of Medical Virology*, 81(5):888–896, 5 2009. ISSN 1096-9071. doi: 10.1002/JMV.21349. URL https://onlinelibrary.wiley.com/doi/full/10.1002/jmv.21349https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.21349https://onlinelibrary.wiley.com/doi/10.1002/jmv.21349.

[177] Christin E. Petre, Sang-Hoon Sin, and Dirk P. Dittmer. Functional p53 Signaling in Kaposi's Sarcoma-Associated Herpesvirus Lymphomas: Implications for Therapy. *Journal of Virology*, 81(4):1912–1922, 2 2007. ISSN 0022-538X. doi: 10.1128/JVI.01757-06/SUPPL{_}FILE/SUPPLEMENTAL{_}MOVIE{_}LEGEND{_}1.DOC. URL https://journals.asm.org/doi/10.1128/JVI.01757-06.

[178] Alex Ray, Vickie Marshall, Thomas Uldrick, Robert Leighty, Nazzarena Labo, Kathy Wyvill, Karen Aleman, Mark N. Polizzotto, Richard F. Little, Robert Yarchoan, and Denise Whitby. Sequence Analysis of Kaposi Sarcoma–Associated Herpesvirus (KSHV) MicroRNAs in Patients with Multicentric Castleman Disease and KSHV-Associated Inflammatory Cytokine Syndrome. *The Journal of Infectious Diseases*, 205(11):1665–1676, 6 2012. ISSN 0022-1899. doi: 10.1093/INFDIS/JIS249. URL https://academic.oup.com/jid/article/205/11/1665/844770.

[179] Isabelle Poizot-Martin, Sylvie Brégigeon, Romain Palich, Anne Geneviève Marcelin, Marc Antoine Valantin, Caroline Solas, Marianne Veyri, Jean Philippe Spano, and Alain Makinson. Immune Reconstitution Inflammatory Syndrome Associated Kaposi Sarcoma. *Cancers*, 14(4):986, 2 2022. ISSN 20726694. doi: 10.3390/CANCERS14040986. URL /pmc/articles/PMC8869819//pmc/articles/PMC8869819/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8869819/.

[180] Qian Chen, Jiangtao Chen, Yuqing Li, Dawei Liu, Yan Zeng, Zheng Tian, Akbar Yunus, Yong Yang, Jie Lu, Xinghua Song, and Yan Yuan. Kaposi's sarcoma herpesvirus is associated with osteosarcoma in Xinjiang populations. *Proceedings of the National Academy of Sciences of the United States of America*, 118(10), 3 2021. ISSN 10916490. doi: 10.1073/PNAS.2016653118/-/DCSUPPLEMENTAL. URL /pmc/articles/PMC7958238//pmc/articles/PMC7958238/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7958238/.

[181] Xinling Hu, Liu Yang, and Yin Yuan Mo. Role of Pseudogenes in Tumorigenesis. *Cancers*, 10(8), 8 2018. ISSN 20726694. doi: 10.3390/CANCERS10080256. URL /pmc/articles/PMC6115995//pmc/articles/PMC6115995/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6115995/.

[182] Jennifer A. Doudna and Thomas R. Cech. The chemical repertoire of natural ribozymes. *Nature*, 418(6894):222–228, 7 2002. ISSN 0028-0836. doi: 10.1038/418222a. URL http://www.nature.com/articles/418222a.

[183] Thomas R. Cech and Joan A. Steitz. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell*, 157(1):77–94, 3 2014. ISSN 0092-8674. doi: 10.1016/J.CELL.2014.03.008. URL https://www.sciencedirect.com/science/article/pii/S0092867414003389.

[184] Florian A Karreth and Pier Paolo Pandolfi. ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer discovery*, 3(10):1113–21, 10 2013. ISSN 2159-8290. doi: 10.1158/2159-8290.CD-13-0202. URL http://www.ncbi.nlm.nih.gov/pubmed/24072616http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3801300.

[185] Laura Adriana de Rooij, Dirk Jan Mastebroek, Nicky ten Voorde, Elsken van der Wall, Paul Joannes van Diest, and Cathy Beatrice Moelans. The microRNA Lifecycle in Health and Cancer. *Cancers*, 14(23), 12 2022. ISSN 20726694. doi: 10.3390/CANCERS14235748. URL /pmc/articles/PMC9736740/ /pmc/articles/PMC9736740/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9736740/.

[186] Hilary C Martin, Shivangi Wani, Anita L Steptoe, Keerthana Krishnan, Katia Nones, Ehsan Nourbakhsh, Alexander Vlassov, Sean M Grimmond, and Nicole Cloonan. Imperfect centered miRNA binding sites are common and can mediate repression of target mRNAs. *Genome Biology*, 15(3):R51, 3 2014. ISSN 1465-6906. doi: 10.1186/gb-2014-15-3-r51. URL http://www.ncbi.nlm.nih.gov/pubmed/24629056http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4053950http://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-3-r51.

[187] Ángela L Riffo-Campos, Ismael Riquelme, and Priscilla Brebi-Mieville. Tools for Sequence-Based miRNA Target Prediction: What to Choose? *International journal of molecular sciences*, 17(12), 12 2016. ISSN 1422-0067. doi: 10.3390/ijms17121987. URL http://www.ncbi.nlm.nih.gov/pubmed/27941681http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5187787.

[188] Hervé Seitz. Redefining MicroRNA Targets. *Current Biology*, 19(10):870–873, 5 2009. ISSN 0960-9822. doi: 10.1016/J.CUB.2009.03.059. URL https://www.sciencedirect.com/science/article/pii/S0960982209009130?via%3Dihub.

[189] Leonardo Salmena, Laura Poliseno, Yvonne Tay, Lev Kats, and Pier Paolo Pandolfi. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, 146(3):353–8, 8 2011. ISSN 1097-4172. doi: 10.1016/j.cell.2011.07.014. URL http://www.ncbi.nlm.nih.gov/pubmed/21802130http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3235919.

[190] Rémy Denzler, Vikram Agarwal, Joanna Stefano, David P. Bartel, and Markus Stoffel. Assessing the ceRNA Hypothesis with Quantitative Measurements of miRNA and Target Abundance. *Molecular Cell*, 54(5):766–776, 6 2014. ISSN 1097-2765. doi: 10.1016/J.MOLCEL.2014.03.045. URL https://www.sciencedirect.com/science/article/pii/S1097276514002780.

[191] Daniel W. Thomson and Marcel E. Dinger. Endogenous microRNA sponges: evidence and controversy. *Nature Reviews Genetics*, 17(5):272–283, 5 2016. ISSN 1471-0056. doi: 10.1038/nrg.2016.20. URL http://www.nature.com/articles/nrg.2016.20.

[192] Laura Poliseno, Leonardo Salmena, Jiangwen Zhang, Brett Carver, William J. Haveman, and Pier Paolo Pandolfi. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, 465(7301):1033–1038, 6 2010. ISSN 0028-0836. doi: 10.1038/nature09144. URL http://www.nature.com/articles/nature09144.

[193] Julián Naipauer, Martín E. García Solá, Daria Salyakina, Santas Rosario, Sion Williams, Omar Coso, Martín C. Abba, Enrique A. Mesri, and Ezequiel Lacunza. A Non-Coding RNA Network Involved in KSHV Tumorigenesis. *Frontiers in Oncology*, 0:2217, 6 2021. ISSN 2234-943X. doi: 10.3389/FONC.2021.687629.

[194] Ziqiang Wang, Yiwan Zhao, and Yaou Zhang. Viral lncRNA: A regulatory molecule for controlling virus life cycle. *Non-coding RNA Research*, 2(1):38–44, 3 2017. ISSN 2468-0540. doi: 10.1016/J.NCRNA.2017.03.002.

[195] Joseph M. Luna, Troels K.H. Scheel, Tal Danino, Katharina S. Shaw, Aldo Mele, John J. Fak, Eiko Nishiuchi, Constantin N. Takacs, Maria Teresa Catanese, Ype P. de Jong, Ira M. Jacobson, Charles M. Rice, and Robert B. Darnell. Hepatitis C Virus RNA Functionally Sequesters miR-122. *Cell*, 160(6):1099–1110, 3 2015. ISSN 0092-8674. doi: 10.1016/J.CELL.2015.02.025. URL https://www.sciencedirect.com/science/article/pii/S0092867415001919.

[196] Margaret S. Ebert and Phillip A. Sharp. Emerging Roles for Natural MicroRNA Sponges. *Current Biology*, 20(19):R858–R861, 10 2010. ISSN 09609822. doi: 10.1016/j.cub.2010.08.052. URL https://linkinghub.elsevier.com/retrieve/pii/S0960982210010742.

[197] Valentina Libri, Aleksandra Helwak, Pascal Miesen, Diwakar Santhakumar, Jessica G Borger, Grzegorz Kudla, Finn Grey, David Tollervey, and Amy H Buck. Murine cytomegalovirus encodes a miR-27 inhibitor disguised as a target. *Proceedings of the National Academy of Sciences of the United States of America*, 109(1):279–84, 1 2012. ISSN 1091-6490. doi: 10.1073/pnas.1114204109. URL http://www.ncbi.nlm.nih.gov/pubmed/22184245http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3252920.

[198] Sanghyun Lee, Jaewon Song, Sungchul Kim, Jongkyu Kim, Yujin Hong, Youngkyun Kim, Donghyun Kim, Daehyun Baek, and Kwangseog Ahn. Selective Degradation of Host MicroRNAs by an Intergenic HCMV Noncoding RNA Accelerates Virus Production. *Cell Host & Microbe*, 13(6):678–690, 6 2013. ISSN 1931-3128. doi: 10.1016/J.CHOM.2013.05.007. URL https://www.sciencedirect.com/science/article/pii/S1931312813001893.

[199] Amy Hansen, Stephen Henderson, Dimitrios Lagos, Leonid Nikitenko, Eve Coulter, Sinead Roberts, Fiona Gratrix, Karlie Plaisance, Rolf Renne, Mark Bower, Paul Kellam, and Chris Boshoff. KSHV-encoded miRNAs target MAF to induce endothelial cell reprogramming. doi: 10.1101/gad.553410. URL http://www.genesdev.org.

[200] Shibin Qu, Xisheng Yang, Xiaolei Li, Jianlin Wang, Yuan Gao, Runze Shang, Wei Sun, Kefeng Dou, and Haimin Li. Circular RNA: A new star of noncoding RNAs. *Cancer Letters*, 365(2):141–148, 9 2015. ISSN 0304-3835. doi: 10.1016/J.CANLET.2015.06.003. URL https://www.sciencedirect.com/science/article/abs/pii/S0304383515003833.

[201] Longxin Ren, Qingshan Jiang, Liyi Mo, Lijie Tan, Qifei Dong, Lijuan Meng, Nanyang Yang, and Guoqing Li. Mechanisms of circular RNA degradation. *Communications Biology 2022 5:1*, 5(1):1–6, 12 2022. ISSN 2399-3642. doi: 10.1038/s42003-022-04262-3. URL https://www.nature.com/articles/s42003-022-04262-3.

[202] Katherine L. Harper, Euan McDonnell, and Adrian Whitehouse. CircRNAs: From anonymity to novel regulators of gene expression in cancer (Review). *International Journal of Oncology*, 55(6):1183–1193, 12 2019. ISSN 17912423. doi: 10.3892/IJO.2019.4904/HTML. URL http://www.spandidos-publications.com/10.3892/ijo.2019.4904/abstracthttps://www.spandidos-publications.com/10.3892/ijo.2019.4904.

[203] Julia Salzman, Raymond E. Chen, Mari N. Olsen, Peter L. Wang, and Patrick O. Brown. Cell-Type Specific Features of Circular RNA Expression. *PLoS Genetics*, 9(9):e1003777, 9 2013. ISSN 1553-7404. doi: 10.1371/journal.pgen.1003777. URL https://dx.plos.org/10.1371/journal.pgen.1003777.

[204] William R. Jeck, Jessica A. Sorrentino, Kai Wang, Michael K. Slevin, Christin E. Burd, Jinze Liu, William F. Marzluff, and Norman E. Sharpless. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA (New York, N.Y.)*, 19(2):141–157, 2 2013. ISSN 1469-9001. doi: 10.1261/RNA.035667.112. URL https://pubmed.ncbi.nlm.nih.gov/23249747/.

[205] Ke En Tan and Yat Yuen Lim. Viruses join the circular RNA world. *The Febs Journal*, 288(15):4488, 8 2021. ISSN 17424658. doi: 10.1111/FEBS.15639. URL /pmc/articles/PMC7753765//pmc/articles/PMC7753765/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7753765/.

[206] Jia-xiang Zhang, Jian Lu, Hui Xie, Da-peng Wang, Huan-er Ni, Yong Zhu, Le-hao Ren, Xiao-xiao Meng, and Rui-lan Wang. circHIPK3 regulates lung fibroblast-to-myofibroblast transition by functioning as a competing endogenous RNA. *Cell Death & Disease*, 10(3):182, 3 2019. ISSN 2041-4889. doi: 10.1038/s41419-019-1430-7. URL http://www.nature.com/articles/s41419-019-1430-7.

[207] Qi Hu and Tianshou Zhou. EIciRNA-mediated gene expression: tunability and bimodality. *FEBS Letters*, 592(20):3460–3471, 10 2018. ISSN 1873-3468. doi: 10.1002/1873-3468.13253. URL https://onlinelibrary.wiley.com/doi/full/10.1002/1873-3468.13253https://onlinelibrary.wiley.com/doi/abs/10.1002/1873-3468.13253https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13253.

[208] Xiaojuan Fan, Yun Yang, Chuyun Chen, and Zefeng Wang. Pervasive translation of circular RNAs driven by short IRES-like elements. *Nature Communications 2022 13:1*, 13(1):1–15, 6 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-31327-y. URL https://www.nature.com/articles/s41467-022-31327-y.

[209] Chu Xiao Liu, Xiang Li, Fang Nan, Shan Jiang, Xiang Gao, Si Kun Guo, Wei Xue, Yange Cui, Kaige Dong, Huihua Ding, Bo Qu, Zhaocai Zhou, Nan Shen, Li Yang, and Ling Ling Chen. Structure and Degradation of Circular RNAs Regulate PKR Activation in Innate Immunity. *Cell*, 177(4):865–880, 5 2019. ISSN 0092-8674. doi: 10.1016/J.CELL.2019.03.046.

[210] Dawei Rong, Handong Sun, Zhouxiao Li, Shuheng Liu, Chaoxi Dong, Kai Fu, Weiwei Tang, and Hongyong Cao. An emerging function of circRNA-miRNAs-mRNA axis in human diseases. *Oncotarget*, 8(42), 9 2017. ISSN 1949-2553. doi: 10.18632/oncotarget.19154. URL http://www.oncotarget.com/fulltext/19154.

[211] Hongya Xie, Jie Yao, Yuxuan Wang, and Bin Ni. Exosome-transmitted circVMP1 facilitates the progression and cisplatin resistance of non-small cell lung cancer by targeting miR-524-5p-METTL3/SOX2 axis. *Drug Delivery*, 29(1):1257, 2022. ISSN 15210464. doi: 10.1080/10717544.2022.2057617. URL /pmc/articles/PMC9045767//pmc/articles/PMC9045767/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9045767/.

[212] Rares Drula, Radu Pirlog, Monica Trif, Ondrej Slaby, Cornelia Braicu, and Ioana Berindan-Neagoe. circFOXO3: Going around the mechanistic networks in cancer by interfering with miRNAs regulatory networks. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1867(5):166045, 5 2021. ISSN 0925-4439. doi: 10.1016/J.BBADIS.2020.166045.

[213] Zhaoyong Li, Chuan Huang, Chun Bao, Liang Chen, Mei Lin, Xiaolin Wang, Guolin Zhong, Bin Yu, Wanchen Hu, Limin Dai, Pengfei Zhu, Zhaoxia Chang, Qingfa Wu, Yi Zhao, Ya Jia, Ping Xu, Huijie Liu, and Ge Shan. Exon-intron circular RNAs regulate transcription in the nucleus. *Nature Structural & Molecular Biology*, 22(3):256–264, 3 2015. ISSN 1545-9993. doi: 10.1038/nsmb.2959. URL http://www.nature.com/articles/nsmb.2959.

[214] Reut Ashwal-Fluss, Markus Meyer, Nagarjuna Reddy Pamudurti, Andranik Ivanov, Osnat Bartok, Mor Hanan, Naveh Evantal, Sebastian Memczak, Nikolaus Rajewsky, and Sebastian Kadener. circRNA Biogenesis Competes with Pre-mRNA Splicing. *Molecular Cell*, 56(1):55–66, 10 2014. ISSN 10972765. doi: 10.1016/j.molcel.2014.08.019. URL https://linkinghub.elsevier.com/retrieve/pii/S1097276514006741.

[215] Lesca M. Holdt, Anika Stahringer, Kristina Sass, Garwin Pichler, Nils A. Kulak, Wolfgang Wilfert, Alexander Kohlmaier, Andreas Herbst, Bernd H. Northoff, Alexandros Nicolaou, Gabor Gäbel, Frank Beutner, Markus Scholz, Joachim Thiery, Kiran Musunuru, Knut Krohn, Matthias Mann, and Daniel Teupser. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nature Communications 2016 7:1*, 7(1):1–14, 8 2016. ISSN 2041-1723. doi: 10.1038/ncomms12429. URL https://www.nature.com/articles/ncomms12429.

[216] Lorena Verduci, Emilio Tarcitano, Sabrina Strano, Yosef Yarden, and Giovanni Blandino. CircRNAs: role in human diseases and potential use as biomarkers. *Cell Death & Disease 2021 12:5*, 12(5):1–12, 5 2021. ISSN 2041-4889. doi: 10.1038/s41419-021-03743-3. URL https://www.nature.com/articles/s41419-021-03743-3.

[217] Wei Cheng Liang, Cheuk Wa Wong, Pu Ping Liang, Mai Shi, Ye Cao, Shi Tao Rao, Stephen Kwok Wing Tsui, Mary Miu Yee Waye, Qi Zhang, Wei Ming Fu, and Jin Fang Zhang. Translation of the circular RNA circβ-catenin promotes liver cancer cell growth through activation of the Wnt pathway. *Genome Biology*, 20(1):1–12, 4 2019. ISSN 1474760X. doi: 10.1186/S13059-019-1685-4/FIGURES/6. URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1685-4.

[218] Liang Shi, Boqiang Liu, Dan Dan Shen, Peijian Yan, Yanan Zhang, Yuanshi Tian, Lidan Hou, Guangyi Jiang, Yinxin Zhu, Yuelong Liang, Xiao Liang, Bo Shen, Hong Yu, Yan Zhang, Yifan Wang, Xing Guo, and Xiujun Cai. A tumor-suppressive circular RNA mediates uncanonical integrin degradation by the proteasome in liver cancer. *Science Advances*, 7(13):5043–5067, 3 2021. ISSN 23752548. doi: 10.1126/SCIADV.ABE5043. URL /pmc/articles/PMC7990343//pmc/articles/PMC7990343/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7990343/.

[219] Ju Zhang, Xiuli Zhang, Cuidan Li, Liya Yue, Nan Ding, Tim Riordan, Li Yang, Yang Li, Charles Jen, Sen Lin, Dongsheng Zhou, and Fei Chen. Circular RNA profiling provides insights into their subcellular distribution and molecular characteristics in HepG2 cells. *RNA Biology*, 16(2):220, 2 2019. ISSN 15558584. doi: 10.1080/15476286.2019.1565284. URL /pmc/articles/PMC6380345//pmc/articles/PMC6380345/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6380345/.

[220] Lasse S. Kristensen, Theresa Jakobsen, Henrik Hager, and Jørgen Kjems. The emerging roles of circRNAs in cancer and oncology. *Nature Reviews Clinical Oncology 2021 19:3*, 19(3):188–206, 12 2021. ISSN 1759-4782. doi: 10.1038/s41571-021-00585-y. URL https://www.nature.com/articles/s41571-021-00585-y.

[221] Dmitri D. Pervouchine. Circular exonic RNAs: When RNA structure meets topology. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 5 2019. ISSN 1874-9399. doi: 10.1016/J.BBAGRM.2019.05.002. URL https://0-www-sciencedirect-com.wam.leeds.ac.uk/science/article/pii/S1874939918305042.

[222] Shankar Mukherji, Margaret S Ebert, Grace X Y Zheng, John S Tsang, Phillip A Sharp, and Alexander van Oudenaarden. MicroRNAs can generate thresholds in target gene expression. *Nature genetics*, 43(9):854–9, 8 2011. ISSN 1546-1718. doi: 10.1038/ng.905. URL http://www.ncbi.nlm.nih.gov/pubmed/21857679http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3163764.

[223] Yanze Song, Min Chen, Yingfan Zhang, Jiayi Li, Bowen Liu, Na Li, Min Chen, Miaomiao Qiao, Nan Wang, Yuanwei Cao, Shan Lu, Jian Chen, Wen Sun, Fei Gao, and Haoyi Wang. Loss of circSRY reduces γH2AX level in germ cells and impairs mouse spermatogenesis. *Life Science Alliance*, 6(2), 2 2023. ISSN 25751077. doi: 10.26508/LSA.202201617. URL /pmc/articles/PMC9684031//pmc/articles/PMC9684031/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9684031/.

[224] T.-C. Zhou, X. Li, L.-J. Chen, J.-H. Fan, X. Lai, Y. Tang, L. Zhang, and J. Wei. Differential expression profile of hepatic circular RNAs in chronic hepatitis B. *Journal of Viral Hepatitis*, 25(11):1341–1351, 11 2018. ISSN 13520504. doi: 10.1111/jvh.12944. URL http://www.ncbi.nlm.nih.gov/pubmed/29888838http://doi.wiley.com/10.1111/jvh.12944.

[225] Lasse S. Kristensen, Karoline K. Ebbesen, Martin Sokol, Theresa Jakobsen, Ulrik Korsgaard, Ann C. Eriksen, Thomas B. Hansen, Jørgen Kjems, and Henrik Hager. Spatial expression analyses of the putative oncogene ciRS-7 in cancer reshape the microRNA sponge theory. *Nature Communications 2020 11:1*, 11(1):1–12, 9 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-18355-2. URL https://www.nature.com/articles/s41467-020-18355-2.

[226] Qiupeng Zheng, Chunyang Bao, Weijie Guo, Shuyi Li, Jie Chen, Bing Chen, Yanting Luo, Dongbin Lyu, Yan Li, Guohai Shi, Linhui Liang, Jianren Gu, Xianghuo He, and Shenglin Huang. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nature Communications*, 7(1):11215, 9 2016. ISSN 2041-1723. doi: 10.1038/ncomms11215. URL http://www.nature.com/articles/ncomms11215.

[227] Genwen Chen, Yanting Shi, Mengmeng Liu, and Jianyong Sun. circHIPK3 regulates cell proliferation and migration by sponging miR-124 and regulating AQP3 expression in hepatocellular carcinoma. *Cell Death & Disease*, 9(2):175, 2 2018. ISSN 2041-4889. doi: 10.1038/s41419-017-0204-3. URL http://www.nature.com/articles/s41419-017-0204-3.

[228] Jiang Zhou, Baisheng Wang, Xin Bin, Changqing Xie, Bo Li, Ousheng Liu, and Zhangui Tang. CircHIPK3: Key Player in Pathophysiology and Potential Diagnostic and Therapeutic Tool. *Frontiers in Medicine*, 8:170, 2 2021. ISSN 2296858X. doi: 10.3389/FMED.2021.615417/BIBTEX.

[229] Le Qin, Jie Lin, and Xiaoxiao Xie. CircRNA-9119 suppresses poly I:C induced inflammation in Leydig and Sertoli cells via TLR3 and RIG-I signal pathways. *Molecular Medicine*, 25(1):1–13, 6 2019. ISSN 15283658. doi: 10.1186/S10020-019-0094-1/FIGURES/9. URL https://molmed.biomedcentral.com/articles/10.1186/s10020-019-0094-1.

[230] A. Kos, R. Dijkema, A. C. Arnberg, P. H. Van Der Meide, and H. Schellekens. The hepatitis delta (δ) virus possesses a circular RNA. *Nature 1986 323:6088*, 323(6088):558–560, 1986. ISSN 1476-4687. doi: 10.1038/323558a0. URL https://www.nature.com/articles/323558a0.

[231] Yue Zhang, Hui Zhang, Minghui An, Bin Zhao, Haibo Ding, Zining Zhang, Youwen He, Hong Shang, and Xiaoxu Han. Crosstalk in competing endogenous RNA networks reveals new circular RNAs involved in the pathogenesis of early HIV infection. *Journal of Translational Medicine*, 16(1):332, 12 2018. ISSN 1479-5876. doi: 10.1186/s12967-018-1706-1. URL https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-018-1706-1.

[232] Qinjunjie Chen, Haibo Wang, Zheng Li, Fengwei Li, Leilei Liang, Yiran Zou, Hao Shen, Jun Li, Yong Xia, Zhangjun Cheng, Tian Yang, Kui Wang, and Feng Shen. Circular RNA ACTN4 promotes intrahepatic cholangiocarcinoma progression by recruiting YBX1 to initiate FZD7 transcription. *Journal of Hepatology*, 76(1):135–147, 1 2022. ISSN 0168-8278. doi: 10.1016/J.JHEP.2021.08.027.

[233] Min Zhu, Yaping Dai, Xinyu Tong, Yaxin Zhang, Yang Zhou, Jiali Cheng, Yiting Jiang, Ruolin Yang, Xiangyu Wang, Guangli Cao, Renyu Xue, Xiaolong Hu, and Chengliang Gong. Circ-Udg Derived from Cyprinid Herpesvirus 2 Promotes Viral Replication. *Microbiology Spectrum*, 10(4), 8 2022. ISSN 21650497. doi: 10.1128/SPECTRUM.00943-22. URL /pmc/articles/PMC9431488//pmc/articles/PMC9431488/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9431488/.

[234] Alexis S Chasseur, Gabrielle Trozzi, Céline Istasse, Astrid Petit, Perrine Rasschaert, Caroline Denesvre, Benedikt B Kaufer, Luca D Bertzbach, Benoît Muylkens, and Damien Coupeau. Marek's Disease Virus Virulence Genes Encode Circular RNAs. 2022. doi: 10.1128/jvi.00321-22. URL https://journals.asm.org/journal/jvi.

[235] Shuihong Yao, Xuemei Jia, Fei Wang, Liuxue Sheng, Pengxia Song, Yanhui Cao, Hongjuan Shi, Weifei Fan, Xiangya Ding, Shou Jiang Gao, and Chun Lu. CircRNA ARFGEF1 functions as a ceRNA to promote oncogenic KSHV-encoded viral interferon regulatory factor induction of cell invasion and angiogenesis by upregulating glutaredoxin 3. *PLoS Pathogens*, 17(2), 2 2021. ISSN 15537374. doi: 10.1371/JOURNAL.PPAT.1009294. URL /pmc/articles/PMC7888650//pmc/articles/PMC7888650/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7888650/.

[236] Shaomin Yang, Hong Zhou, Mingde Liu, Dabbu Jaijyan, Ruth Cruz-Cosme, Santhamani Ramasamy, Selvakumar Subbian, Dongxiao Liu, Jiayu Xu, Xiaoyu Niu, Yaolan Li, Lizu Xiao, Sanjay Tyagi, Qiuhong Wang, Hua Zhu, and Qiyi Tang. SARS-CoV-2, SARS-CoV, and MERS-CoV encode circular RNAs of spliceosome-independent origin. *Journal of Medical Virology*, 94(7):3203–3222, 7 2022. ISSN 1096-9071. doi: 10.1002/JMV.27734. URL https://onlinelibrary.wiley.com/doi/full/10.1002/jmv.27734https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.27734https://onlinelibrary.wiley.com/doi/10.1002/jmv.27734.

[237] Bizunesh Abere, Hongzhao Zhou, Jinghui Li, Simon Cao, Patrick S. Moore, Yuan Chang, Tuna Toptan, Adam Grundhoff, and Nicole Fischer. Merkel Cell Polyomavirus Encodes Circular RNAs (circRNAs) Enabling a Dynamic circRNA/microRNA/mRNA Regulatory Network. *mBio*, 11(6):1–20, 11 2020. ISSN 2150-7511. doi: 10.1128/MBIO.03059-20. URL https://pubmed.ncbi.nlm.nih.gov/33323517/.

[238] Min Zhu, Zi Liang, Jun Pan, Xing Zhang, Renyu Xue, Guangli Cao, Xiaolong Hu, and Chengliang Gong. Hepatocellular carcinoma progression mediated by hepatitis B virus-encoded circRNA HBV_circ_1 through interaction with CDK1. *Molecular Therapy. Nucleic Acids*, 25:668, 9 2021. ISSN 21622531. doi: 10.1016/J.OMTN.2021.08.011. URL /pmc/articles/PMC8463320//pmc/articles/PMC8463320/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8463320/.

[239] Zena Cai, Yunshi Fan, Zheng Zhang, Congyu Lu, Zhaozhong Zhu, Taijiao Jiang, Tongling Shan, and Yousong Peng. VirusCircBase: a database of virus circular RNAs. *Briefings in bioinformatics*, 22(2):2182–2190, 3 2021. ISSN 1477-4054. doi: 10.1093/BIB/BBAA052. URL https://pubmed.ncbi.nlm.nih.gov/32349124/.

[240] Kazuma Sekiba, Motoyuki Otsuka, Motoko Ohno, Takahiro Kishikawa, Mari Yamagami, Tatsunori Suzuki, Rei Ishibashi, Takahiro Seimiya, Eri Tanaka, and Kazuhiko Koike. DHX9 regulates production of hepatitis B virus-derived circular RNA and viral protein levels. *Oncotarget*, 9(30):20953, 4 2018. ISSN 19492553. doi: 10.18632/ONCOTARGET.25104. URL /pmc/articles/PMC5940377//pmc/articles/PMC5940377/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5940377/.

[241] Nathan A. Ungerleider, Vaibhav Jain, Yiping Wang, Nicholas J. Maness, Robert V. Blair, Xavier Alvarez, Cecily Midkiff, Dennis Kolson, Shanshan Bai, Claire Roberts, Walter N. Moss, Xia Wang, Jacqueline Serfecz, Michael Seddon, Terri Lehman, Tianfang Ma, Yan Dong, Rolf Renne, Scott A. Tibbetts, and Erik K. Flemington. Comparative Analysis of Gammaherpesvirus Circular RNA Repertoires: Conserved and Unique Viral Circular RNAs. *Journal of Virology*, 93 (6), 12 2018. ISSN 0022-538X. doi: 10.1128/JVI.01952-18. URL http://jvi.asm.org/lookup/doi/10.1128/JVI.01952-18.

[242] Jiawei Zhao, Eunice E. Lee, Jiwoong Kim, Bahir Chamseddin, Rong Yang, Yang Xie, Xiaowei Zhan, and Richard C. Wang. Translation and Transforming Activity of a Circular RNA from Human Papillomavirus. *bioRxiv*, page 600056, 4 2019. doi: 10.1101/600056. URL https://www.biorxiv.org/content/10.1101/600056v1.full.

[243] Lixian Yang, Hanying Wang, Qi Shen, Lifeng Feng, and Hongchuan Jin. Long non-coding RNAs involved in autophagy regulation. *Cell death & disease*, 8(10):e3073, 2017. ISSN 2041-4889. doi: 10.1038/cddis.2017.464. URL http://www.ncbi.nlm.nih.gov/pubmed/28981093http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5680586.

[244] Tuna Toptan, Bizunesh Abere, Michael A Nalesnik, Steven H Swerdlow, Sarangarajan Ranganathan, Nara Lee, Kathy H Shair, Patrick S Moore, and Yuan Chang. Circular DNA tumor viruses make circular RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (37):E8737–E8745, 2018. ISSN 1091-6490. doi: 10.1073/pnas.1811728115. URL http://www.ncbi.nlm.nih.gov/pubmed/30150410http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6140489.

[245] Li-ping Gong, Jian-ning Chen, Min Dong, Zhen-dong Xiao, Zhi-ying Feng, Yu-hang Pan, Yu Zhang, Yu Du, Jing-yue Zhang, Yuan-hua Bi, Jun-ting Huang, Jing Liang, and Chun-kui Shao. Epstein–Barr virus-derived circular RNA LMP2A induces stemness in EBV-associated gastric cancer. *EMBO Reports*, 21(10), 10 2020. ISSN 1469-221X. doi: 10.15252/EMBR.201949689. URL /pmc/articles/PMC7534631//pmc/articles/PMC7534631/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7534631/.

[246] Yiming Yang, Di Wang, Kaixiong Tao, and Guobin Wang. Circular RNA circLRCH3 Inhibits Proliferation, Migration, and Invasion of Colorectal Cancer Cells Through miRNA-223/LPP Axis. *OncoTargets and therapy*, 15:541, 2022. ISSN 11786930. doi: 10.2147/OTT.S366605. URL /pmc/articles/PMC9124491//pmc/articles/PMC9124491/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9124491/.

[247] Cyprian C. Rossetto and Gregory S. Pari. PAN's Labyrinth: Molecular Biology of Kaposi's Sarcoma-Associated Herpesvirus (KSHV) PAN RNA, a Multifunctional Long Noncoding RNA. *Viruses*, 6(11):4212, 11 2014. ISSN 19994915. doi: 10.3390/V6114212. URL /pmc/articles/PMC4246217//pmc/articles/PMC4246217/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4246217/.

[248] Takanobu Tagawa, Daniel Oh, Sarah Dremel, Guruswamy Mahesh, Vishal N. Koparde, Gerard Duncan, Thorkell Andresson, and Joseph M. Ziegelbauer. A virus-induced circular RNA maintains latent infection of Kaposi's sarcoma herpesvirus. *Proceedings of the National Academy of Sciences of the United States of America*, 120(6):e2212864120, 2 2023. ISSN 10916490. doi: 10.1073/PNAS.2212864120/SUPPL{_}FILE/PNAS.2212864120.SD05.XLSX. URL https://www.pnas.org/doi/abs/10.1073/pnas.2212864120.

[249] Liuxue Sheng, Chen Chen, Yuheng Chen, Yujia He, Ruoyu Zhuang, Yang Gu, Qin Yan, Wan Li, and Chun Lu. vFLIP-regulated competing endogenous RNA (ceRNA) networks targeting lytic induction for KSHV-associated malignancies. *Journal of Medical Virology*, 94(6):2766–2775, 6 2022. ISSN 1096-9071. doi: 10.1002/JMV.27654. URL https://onlinelibrary.wiley.com/doi/full/10.1002/jmv.27654https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.27654https://onlinelibrary.wiley.com/doi/10.1002/jmv.27654.

[250] Sean M. Gregory, Beckley K. Davis, John A. West, Debra J. Taxman, Shu Ichi Matsuzawa, John C. Reed, Jenny P.Y. Ting, and Blossom Damania. Discovery of a Viral NLR Homolog that Inhibits the Inflammasome. *Science (New York, N.Y.)*, 331(6015):330, 1 2011. ISSN 00368075. doi: 10.1126/SCIENCE.1199478. URL /pmc/articles/PMC3072027//pmc/articles/PMC3072027/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3072027/.

[251] Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139, 1 2010. ISSN 14602059. doi: 10.1093/BIOINFORMATICS/BTP616. URL /pmc/articles/PMC2796818//pmc/articles/PMC2796818/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2796818/.

[252] Zuguang Gu, Roland Eils, and Matthias Schlesner. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18):2847–2849, 9 2016. ISSN 1367-4803. doi: 10.1093/BIOINFORMATICS/BTW313. URL https://academic.oup.com/bioinformatics/article/32/18/2847/1743594.

[253] Abbas Salavaty, Mirana Ramialison, and Peter D. Currie. Integrated Value of Influence: An Integrative Method for the Identification of the Most Influential Nodes within Networks. *Patterns*, 1(5), 8 2020. ISSN 26663899. doi: 10.1016/j.patter.2020.100052. URL http://www.cell.com/article/S2666389920300635/fulltexthttp://www.cell.com/article/S2666389920300635/abstracthttps://www.cell.com/patterns/abstract/S2666-3899(20)30063-5.

[254] Aleksandr Ianevski, Anil K. Giri, and Tero Aittokallio. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nature Communications 2022 13:1*, 13(1):1–10, 3 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-28803-w. URL https://www.nature.com/articles/s41467-022-28803-w.

[255] Etienne Becht, Nicolas A. Giraldo, Laetitia Lacroix, Bénédicte Buttard, Nabila Elarouci, Florent Petitprez, Janick Selves, Pierre Laurent-Puig, Catherine Sautès-Fridman, Wolf H. Fridman, and Aurélien de Reyniès. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biology*, 17(1):1–20, 10 2016. ISSN 1474760X. doi: 10.1186/S13059-016-1070-5/TABLES/4. URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1070-5.

[256] Shizhen Emily Wang, Frederick Y Wu, Honglin Chen, Meir Shamay, Qizhi Zheng, and Gary S Hayward. Early activation of the Kaposi's sarcoma-associated herpesvirus RTA, RAP, and MTA promoters by the tetradecanoyl phorbol acetate-induced AP1 pathway. *Journal of virology*, 78(8):4248–67, 4 2004. ISSN 0022-538X. doi: 10.1128/jvi.78.8.4248-4267.2004. URL http://www.ncbi.nlm.nih.gov/pubmed/15047839http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC374264.

[257] Ana Kozomara, Maria Birgaoanu, and Sam Griffiths-Jones. miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(D1):D155–D162, 1 2019. ISSN 0305-1048. doi: 10.1093/nar/gky1141. URL https://academic.oup.com/nar/article/47/D1/D155/5179337.

[258] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, 12 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8. URL http://www.ncbi.nlm.nih.gov/pubmed/25516281http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4302049http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8.

[259] Jianqiang Sun, Tomoaki Nishiyama, Kentaro Shimizu, and Koji Kadota. TCC: An R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics*, 14(1):1–14, 7 2013. ISSN 14712105. doi: 10.1186/1471-2105-14-219/FIGURES/3. URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-219.

[260] Jan Krüger and Marc Rehmsmeier. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic acids research*, 34(Web Server issue), 7 2006. ISSN 1362-4962. doi: 10.1093/NAR/GKL243. URL https://pubmed.ncbi.nlm.nih.gov/16845047/.

[261] Shifu Chen, Yanqing Zhou, Yaru Chen, and Jia Gu. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17):i884–i890, 9 2018. ISSN 1367-4803. doi: 10.1093/BIOINFORMATICS/BTY560. URL https://academic.oup.com/bioinformatics/article/34/17/i884/5093234.

[262] Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15, 1 2013. ISSN 13674803. doi: 10.1093/BIOINFORMATICS/BTS635. URL /pmc/articles/PMC3530905//pmc/articles/PMC3530905/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3530905/.

[263] Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166, 1 2015. ISSN 14602059. doi: 10.1093/BIOINFORMATICS/BTU638. URL /pmc/articles/PMC4287950//pmc/articles/PMC4287950/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287950/.

[264] Matthew E. Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47, 4 2015. ISSN 13624962. doi: 10.1093/NAR/GKV007. URL /pmc/articles/PMC4402510//pmc/articles/PMC4402510/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4402510/.

[265] Anqi Zhu, Joseph G. Ibrahim, and Michael I. Love. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*, 35(12):2084–2092, 6 2019. ISSN 1367-4803. doi: 10.1093/BIOINFORMATICS/BTY895. URL https://academic.oup.com/bioinformatics/article/35/12/2084/5159452.

[266] Peter Langfelder and Steve Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1):559, 12 2008. ISSN 1471-2105. doi: 10.1186/1471-2105-9-559. URL http://www.ncbi.nlm.nih.gov/pubmed/19114008http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2631488https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559.

[267] Peter Langfelder, Bin Zhang, and Steve Horvath. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, 24(5):719–720, 3 2008. ISSN 1367-4803. doi: 10.1093/BIOINFORMATICS/BTM563. URL https://academic.oup.com/bioinformatics/article/24/5/719/200751.

[268] Fran Supek, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE*, 6(7):e21800, 2011. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0021800. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021800.

[269] Andrew T. McKenzie, Igor Katsyv, Won Min Song, Minghui Wang, and Bin Zhang. DGCA: A comprehensive R package for Differential Gene Correlation Analysis. *BMC Systems Biology*, 10(1):1–25, 11 2016. ISSN 17520509. doi: 10.1186/S12918-016-0349-1/FIGURES/16. URL https://bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-016-0349-1.

[270] Julio C. Ruiz, Olga V. Hunter, and Nicholas K. Conrad. Kaposi's sarcoma-associated herpesvirus ORF57 protein protects viral transcripts from specific nuclear RNA decay pathways by preventing hMTR4 recruitment. *PLOS Pathogens*, 15(2):e1007596, 2 2019. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1007596. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1007596.

[271] Yang Zhang, Xiao Ou Zhang, Tian Chen, Jian Feng Xiang, Qing Fei Yin, Yu Hang Xing, Shanshan Zhu, Li Yang, and Ling Ling Chen. Circular Intronic Long Noncoding RNAs. *Molecular Cell*, 51(6):792–806, 9 2013. ISSN 1097-2765. doi: 10.1016/J.MOLCEL.2013.08.017.

[272] Yong Peng and Carlo M. Croce. The role of MicroRNAs in human cancer. *Signal Transduction and Targeted Therapy 2016 1:1*, 1(1):1–9, 1 2016. ISSN 2059-3635. doi: 10.1038/sigtrans.2015.4. URL https://www.nature.com/articles/sigtrans20154.

[273] Rob Patro, Geet Duggal, Michael I Love, Rafael A Irizarry, and Carl Kingsford. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods*, 14(4):417–419, 4 2017. ISSN 1548-7105. doi: 10.1038/nmeth.4197. URL http://www.ncbi.nlm.nih.gov/pubmed/28263959http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5600148.

[274] Pauline E. Chugh, Sang Hoon Sin, Sezgin Ozgur, David H. Henry, Prema Menezes, Jack Griffith, Joseph J. Eron, Blossom Damania, and Dirk P. Dittmer. Systemically Circulating Viral and Tumor-Derived MicroRNAs in KSHV-Associated Malignancies. *PLOS Pathogens*, 9(7):e1003484, 7 2013. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1003484. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003484.

[275] Adam Price and Cynthia Gibas. The quantitative impact of read mapping to non-native reference genomes in comparative RNA-Seq studies. *PLOS ONE*, 12(7):e0180904, 7 2017. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0180904. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0180904.

[276] Daniel Aird, Michael G. Ross, Wei Sheng Chen, Maxwell Danielsson, Timothy Fennell, Carsten Russ, David B. Jaffe, Chad Nusbaum, and Andreas Gnirke. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, 12(2):R18, 2 2011. ISSN 14747596. doi: 10.1186/GB-2011-12-2-R18. URL /pmc/articles/PMC3188800//pmc/articles/PMC3188800/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3188800/.

[277] Yanggu Shi and Jindong Shang. Circular RNA Expression Profiling by Microarray—A Technical and Practical Perspective, 4 2023. ISSN 2218273X.

[278] Yangping Li, Feng Wang, Peng Teng, Li Ku, Li Chen, Yue Feng, and Bing Yao. Accurate identification of circRNA landscape and complexity reveals their pivotal roles in human oligodendroglia differentiation. *Genome Biology*, 23(1), 12 2022. ISSN 1474760X. doi: 10.1186/s13059-022-02621-1.

[279] Arthur C. Oliveira, Luiz A. Bovolenta, Pedro G. Nachtigall, Marcos E. Herkenhoff, Ney Lemke, and Danillo Pinhal. Combining Results from Distinct MicroRNA Target Prediction Tools Enhances the Performance of Analyses. *Frontiers in Genetics*, 8, 2017. doi: 10.3389/FGENE.2017.00059. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5432626/.

[280] Bino John, Anton J. Enright, Alexei Aravin, Thomas Tuschl, Chris Sander, and Debora S. Marks. Human microRNA targets. *PLoS Biology*, 2(11), 11 2004. ISSN 15449173. doi: 10.1371/JOURNAL.PBIO.0020363.

[281] R. Puca, L. Nardinocchi, D. Givol, and G. D'Orazi. Regulation of p53 activity by HIPK2: molecular mechanisms and therapeutical implications in human cancer cells. *Oncogene 2010 29:31*, 29(31):4378–4387, 5 2010. ISSN 1476-5594. doi: 10.1038/onc.2010.183. URL https://www.nature.com/articles/onc2010183.

[282] Joelle N. Zambrano, Scott T. Eblen, Melissa Abt, J. Matthew Rhett, Robin Muise-Helmericks, and Elizabeth S. Yeh. HUNK Phosphorylates Rubicon to Support Autophagy. *International Journal of Molecular Sciences*, 20(22), 11 2019. ISSN 14220067. doi: 10.3390/IJMS20225813. URL /pmc/articles/PMC6888122//pmc/articles/PMC6888122/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6888122/.

[283] Guangwei Wei, Stephen Ku, Gene K. Ma, Shin'ichi Saito, Amy A. Tang, Jiasheng Zhang, Jian Hua Mao, Ettore Appella, Allan Balmain, and Eric J. Huang. HIPK2 represses β-catenin-mediated transcription, epidermal stem cell expansion, and skin tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 104(32):13040–13045, 8 2007. ISSN 00278424. doi: 10.1073/PNAS.0703213104/SUPPL{_}FILE/03213FIG13.PDF. URL https://www.pnas.org/doi/abs/10.1073/pnas.0703213104.

[284] Amerria Causey, Mathew Constantine, Arie Admon, Alfredo Garzino-Demo, and Elana Ehrlich. Analysis of the ubiquitin-modified proteome identifies novel host determinants of Kaposi's sarcoma herpesvirus lytic reactivation. *bioRxiv*, page 2022.08.14.503934, 8 2022. doi: 10.1101/2022.08.14.503934. URL https://www.biorxiv.org/content/10.1101/2022.08.14.503934v1https://www.biorxiv.org/content/10.1101/2022.08.14.503934v1.abstract.

[285] Kevin Brulois, Zsolt Toth, Lai-Yee Wong, Pinghui Feng, Shou-Jiang Gao, Armin Ensser, and Jae U. Jung. Kaposi's Sarcoma-Associated Herpesvirus K3 and K5 Ubiquitin E3 Ligases Have Stage-Specific Immune Evasion Roles during Lytic Replication. *Journal of Virology*, 88(16):9335, 8 2014. ISSN 0022-538X. doi: 10.1128/JVI.00873-14. URL /pmc/articles/PMC4136276//pmc/articles/PMC4136276/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4136276/.

[286] Georgios A Pavlopoulos, Maria Secrier, Charalampos N Moschopoulos, Theodoros G Soldatos, Sophia Kossida, Jan Aerts, Reinhard Schneider, and Pantelis G Bagos. Using graph theory to analyze biological networks. *BioData Mining*, 4(1):10, 12 2011. ISSN 1756-0381. doi: 10.1186/1756-0381-4-10. URL https://biodatamining.biomedcentral.com/articles/10.1186/1756-0381-4-10.

[287] Alexis Vandenbon. Evaluation of critical data processing steps for reliable prediction of gene co-expression from large collections of RNA-seq data. *PLOS ONE*, 17(1):e0263344, 1 2022. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0263344. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0263344.

[288] Kevin W. Kelley, Hiromi Nakao-Inoue, Anna V. Molofsky, and Michael C. Oldham. Variation among intact tissue samples reveals the core transcriptional features of human CNS cell classes. *Nature Neuroscience 2018 21:9*, 21(9):1171–1184, 8 2018. ISSN 1546-1726. doi: 10.1038/s41593-018-0216-z. URL https://www.nature.com/articles/s41593-018-0216-z.

[289] Jianqiang Li, Doudou Zhou, Weiliang Qiu, Yuliang Shi, Ji-Jiang Yang, Shi Chen, Qing Wang, and Hui Pan. Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design. *Scientific Reports*, 8(1):622, 12 2018. ISSN 2045-2322. doi: 10.1038/s41598-017-18705-z. URL http://www.nature.com/articles/s41598-017-18705-z.

[290] André Voigt and Eivind Almaas. Assessment of weighted topological overlap (wTO) to improve fidelity of gene co-expression networks. *BMC Bioinformatics*, 20(1):1–11, 1 2019. ISSN 14712105. doi: 10.1186/S12859-019-2596-9/FIGURES/4. URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2596-9.

[291] Lyuba V. Bozhilova, Javier Pardo-Diaz, Gesine Reinert, and Charlotte M. Deane. COGENT: evaluating the consistency of gene co-expression networks. *Bioinformatics*, 37(13):1928–1929, 7 2021. ISSN 1367-4803. doi: 10.1093/BIOINFORMATICS/BTAA787. URL https://academic.oup.com/bioinformatics/article/37/13/1928/5906022.

[292] Yun Yun Xu, Ao Shen, and Zhao Lei Zeng. A potential EBV-related classifier is associated with the efficacy of immunotherapy in gastric cancer. *Translational Cancer Research*, 11(7):2084–2096, 7 2022. ISSN 22196803. doi: 10.21037/TCR-22-461/COIF. URL /pmc/articles/PMC9372194//pmc/articles/PMC9372194/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9372194/.

[293] Hassan Karami, Afshin Derakhshani, Mohammad Ghasemigol, Mohammad Fereidouni, Ebrahim Miri-moghaddam, Behzad Baradaran, Neda Jalili Tabrizi, Souzan Najafi, Antonio Giovanni Solimando, Leigh M. Marsh, Nicola Silvestris, Simona De Summa, Angelo Virgilio Paradiso, Vito Racanelli, and Hossein Safarpour. Weighted gene co-expression network analysis combined with machine learning validation to identify key modules and hub genes associated with sars-cov-2 infection. *Journal of Clinical Medicine*, 10(16), 8 2021. ISSN 20770383. doi: 10.3390/JCM10163567/S1. URL /pmc/articles/PMC8397209//pmc/articles/PMC8397209/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8397209/.

[294] Yijie Zhai, Luis M. Franco, Robert L. Atmar, John M. Quarles, Nancy Arden, Kristine L. Bucasas, Janet M. Wells, Diane Niño, Xueqing Wang, Gladys E. Zapata, Chad A. Shaw, John W. Belmont, and Robert B. Couch. Host Transcriptional Response to Influenza and Other Acute Respiratory Viral Infections – A Prospective Cohort Study. *PLoS Pathogens*, 11(6), 6 2015. ISSN 15537374. doi: 10.1371/JOURNAL.PPAT.1004869. URL /pmc/articles/PMC4466531//pmc/articles/PMC4466531/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466531/.

[295] Jakob P. Pettersen and Eivind Almaas. csdR, an R package for differential co-expression analysis. *BMC Bioinformatics*, 23(1):1–7, 12 2022. ISSN 14712105. doi: 10.1186/S12859-022-04605-1/TABLES/1. URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-022-04605-1.

[296] Anton Zhiyanov, Narek Engibaryan, Stepan Nersisyan, Maxim Shkurnikov, and Alexander Tonevitsky. Differential co-expression network analysis with DCoNA reveals isomiR targeting aberrations in prostate cancer. *Bioinformatics*, 39(2), 2 2023. ISSN 13674811. doi: 10.1093/BIOINFORMATICS/BTAD051. URL https://academic.oup.com/bioinformatics/article/39/2/btad051/6998206.

[297] Atsushi Fukushima. DiffCorr: an R package to analyze and visualize differential correlations in biological networks. *Gene*, 518(1):209–214, 4 2013. ISSN 1879-0038. doi: 10.1016/J.GENE.2012.11.028. URL https://pubmed.ncbi.nlm.nih.gov/23246976/.

[298] Franziska Liesecke, Johan Owen De Craene, Sébastien Besseau, Vincent Courdavault, Marc Clastre, Valentin Vergès, Nicolas Papon, Nathalie Giglioli-Guivarc'h, Gaëlle Glévarec, Olivier Pichon, and Thomas Dugé de Bernonville. Improved gene co-expression network quality through expression dataset down-sampling and network aggregation. *Scientific Reports 2019 9:1*, 9(1):1–16, 10 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-50885-8. URL https://www.nature.com/articles/s41598-019-50885-8.

[299] Charlotte Soneson, Michael I. Love, and Mark D. Robinson. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, 4:1521, 2 2016. ISSN 2046-1402. doi: 10.12688/f1000research.7563.2. URL https://f1000research.com/articles/4-1521/v2.

[300] Longjian Zhou, Andrew Chi-Hau Sue, and Wilson Wen Bin Goh. Examining the practical limits of batch effect-correction algorithms: When should you care about batch effects? *Journal of Genetics and Genomics*, 46(9):433–443, 9 2019. ISSN 1673-8527. doi: 10.1016/J.JGG.2019.08.002.

[301] Yuqing Zhang, Giovanni Parmigiani, and W. Evan Johnson. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genomics and Bioinformatics*, 2(3), 9 2020. ISSN 26319268. doi: 10.1093/NARGAB/LQAA078. URL https://academic.oup.com/nargab/article/2/3/lqaa078/5909519.

[302] W. Evan Johnson, Cheng Li, and Ariel Rabinovic. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8 (1):118–127, 1 2007. ISSN 1465-4644. doi: 10.1093/BIOSTATISTICS/KXJ037. URL https://academic.oup.com/biostatistics/article/8/1/118/252073.

[303] Kayla A. Johnson and Arjun Krishnan. Robust normalization and transformation techniques for constructing gene coexpression networks from RNA-seq data. *Genome Biology*, 23(1):1–26, 12 2022. ISSN 1474760X. doi: 10.1186/S13059-021-02568-9/FIGURES/6. URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02568-9.

[304] Antonella Farina, Roberta Santarelli, Rossella Bloise, Roberta Gonnella, Marisa Granato, Roberto Bei, Andrea Modesti, Mara Cirone, Luiza Bengtsson, Antonio Angeloni, and Alberto Faggioni. KSHV ORF67 encoded lytic protein localizes on the nuclear membrane and alters emerin distribution. *Virus research*, 175(2):143–150, 8 2013. ISSN 1872-7492. doi: 10.1016/J.VIRUSRES.2013.04.001. URL https://pubmed.ncbi.nlm.nih.gov/23623980/.

[305] Xin Wang, Nannan Zhu, Wenwei Li, Fanxiu Zhu, Yan Wang, and Yan Yuan. Mono-ubiquitylated ORF45 Mediates Association of KSHV Particles with Internal Lipid Rafts for Viral Assembly and Egress. *PLoS Pathogens*, 11(12):1005332, 2015. ISSN 15537374. doi: 10.1371/JOURNAL.PPAT.1005332. URL /pmc/articles/PMC4674120//pmc/articles/PMC4674120/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4674120/.

[306] Feng-Chun Ye, David J. Blackbourn, Michael Mengel, Jian-Ping Xie, Li-Wu Qian, Whitney Greene, I-Tien Yeh, David Graham, and Shou-Jiang Gao. Kaposi's Sarcoma-Associated Herpesvirus Promotes Angiogenesis by Inducing Angiopoietin-2 Expression via AP-1 and Ets1. *Journal of Virology*, 81(8):3980, 4 2007. ISSN 0022-538X. doi: 10.1128/JVI.02089-06. URL /pmc/articles/PMC1866109//pmc/articles/PMC1866109/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1866109/.

[307] Tingting Huang, Yuhang Zhou, Jinglin Zhang, Chi Chun Wong, Weilin Li, Johnny S.H. Kwan, Rui Yang, Aden K.Y. Chan, Yujuan Dong, Feng Wu, Bin Zhang, Alvin H.K. Cheung, William K.K. Wu, Alfred S.L. Cheng, Jun Yu, Nathalie Wong, Wei Kang, and Ka Fai To. SRGAP1, a crucial target of miR-340 and miR-124, functions as a potential oncogene in gastric tumorigenesis. *Oncogene 2017 37:9*, 37(9):1159–1174, 12 2017. ISSN 1476-5594. doi: 10.1038/s41388-017-0029-7. URL https://www.nature.com/articles/s41388-017-0029-7.

[308] Florence Poy, Maina Lepourcelet, Ramesh A. Shivdasani, and Michael J. Eck. Structure of a human Tcf4–β-catenin complex. *Nature Structural Biology 2001 8:12*, 8(12):1053–1057, 11 2001. ISSN 1545-9985. doi: 10.1038/nsb720. URL https://www.nature.com/articles/nsb720.

[309] Francesco Rossi, Anne Helbling-Leclerc, Ryotaro Kawasumi, Nanda Kumar Jegadesan, Xinlin Xu, Pierre Devulder, Takuya Abe, Minoru Takata, Dongyi Xu, Filippo Rosselli, and Dana Branzei. SMC5/6 acts jointly with Fanconi anemia factors to support DNA repair and genome stability. *EMBO reports*, 21(2):e48222, 2 2020. ISSN 1469-3178. doi: 10.15252/EMBR.201948222. URL https://onlinelibrary.wiley.com/doi/full/10.15252/embr.201948222https://onlinelibrary.wiley.com/doi/abs/10.15252/embr.201948222https://www.embopress.org/doi/10.15252/embr.201948222.

[310] Alan R. Lehmann. The role of SMC proteins in the responses to DNA damage. *DNA repair*, 4(3):309–314, 3 2005. ISSN 1568-7864. doi: 10.1016/J.DNAREP.2004.07.009. URL https://pubmed.ncbi.nlm.nih.gov/15661654/.

[311] Patrizia Sarogni, Orazio Palumbo, Adele Servadio, Simonetta Astigiano, Barbara D'Alessio, Veronica Gatti, Dubravka Cukrov, Silvia Baldari, Maria Michela Pallotta, Paolo Aretini, Felice Dell'Orletta, Silvia Soddu, Massimo Carella, Gabriele Toietta, Ottavia Barbieri, Gabriella Fontanini, and Antonio Musio. Overexpression of the cohesin-core subunit SMC1A contributes to colorectal cancer development. *Journal of Experimental and Clinical Cancer Research*, 38(1):1–16, 3 2019. ISSN 17569966. doi: 10.1186/S13046-019-1116-0/FIGURES/4. URL https://jeccr.biomedcentral.com/articles/10.1186/s13046-019-1116-0.

[312] Verónica Dávalos, Lucía Súarez-López, Julio Castaño, Anthea Messent, Ibane Abasolo, Yolanda Fernandez, Angel Guerra-Moreno, Eloy Espín, Manel Armengol, Eva Musulen, Aurelio Ariza, Joan Sayós, Diego Arango, and Simó Schwartz. Human SMC2 protein, a core subunit of human condensin complex, is a novel transcriptional target of the WNT signaling pathway and a new therapeutic target. *Journal of Biological Chemistry*, 287(52):43472–43481, 12 2012. ISSN 1083351X. doi: 10.1074/JBC.M112.428466.

[313] Céline Van den Broeke, Thary Jacob, and Herman W. Favoreel. Rho'ing in and out of cells: viral interactions with Rho GTPase signaling. *Small GTPases*, 5(MAR), 3 2014. ISSN 2154-1256. doi: 10.4161/SGTP.28318. URL https://pubmed.ncbi.nlm.nih.gov/24691164/.

[314] Jie Qiong Li, Lan Tan, and Jin Tai Yu. The role of the LRRK2 gene in Parkinsonism. *Molecular neurodegeneration*, 9(1):47, 11 2014. ISSN 17501326. doi: 10.1186/1750-1326-9-47/FIGURES/3. URL https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/1750-1326-9-47.

[315] Susanne Herbst and Maximiliano G. Gutierrez. LRRK2 in Infection: Friend or Foe? *ACS Infectious Diseases*, 5(6):809–815, 6 2019. ISSN 23738227. doi: 10.1021/ACSINFECDIS.9B00051/ASSET/IMAGES/LARGE/ID-2019-000517{_}0001.JPEG. URL https://pubs.acs.org/doi/full/10.1021/acsinfecdis.9b00051.

[316] Ashley M. Ngo and Andreas S. Puschnik. Genome-Scale Analysis of Cellular Restriction Factors That Inhibit Transgene Expression from Adeno-Associated Virus Vectors. *Journal of Virology*, 3 2023. ISSN 0022-538X. doi: 10.1128/JVI.01948-22. URL https://journals.asm.org/doi/10.1128/jvi.01948-22.

[317] Martina Oravcová, Minghua Nie, Nicola Zilio, Shintaro Maeda, Yasaman Jami-Alahmadi, Eros Lazzerini-Denchi, James A. Wohlschlegel, Helle D. Ulrich, Takanori Otomo, and Michael N. Boddy. The Nse5/6-like SIMC1-SLF2 complex localizes SMC5/6 to viral replication centers. *eLife*, 11, 11 2022. ISSN 2050084X. doi: 10.7554/ELIFE.79676.

[318] Masahiro Fujimuro, Frederick Y. Wu, Colette Aprhys, Henry Kajumbula, David B. Young, Gary S. Hayward, and S. Diane Hayward. A novel viral mechanism for dysregulation of β-catenin in Kaposi's sarcoma–associated herpesvirus latency. *Nature Medicine 2003 9:3*, 9(3):300–306, 2 2003. ISSN 1546-170X. doi: 10.1038/nm829. URL https://www.nature.com/articles/nm829.

[319] Donghyeon Yu, Johan Lim, Xinlei Wang, Faming Liang, and Guanghua Xiao. Enhanced construction of gene regulatory networks using hub gene information. *BMC Bioinformatics*, 18(1), 3 2017. ISSN 14712105. doi: 10.1186/s12859-017-1576-1.

[320] Horng-Shen Chen, Priyankara Wikramasinghe, Louise Showe, and Paul M. Lieberman. Cohesins Repress Kaposi's Sarcoma-Associated Herpesvirus Immediate Early Gene Transcription during Latency. *Journal of Virology*, 86(17):9454, 9 2012. ISSN 0022-538X. doi: 10.1128/JVI.00787-12. URL /pmc/articles/PMC3416178//pmc/articles/PMC3416178/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3416178/.

[321] Ishak D. Irwan, Hal P. Bogerd, and Bryan R. Cullen. Epigenetic silencing by the SMC5/6 complex mediates HIV-1 latency. *Nature Microbiology 2022 7:12*, 7(12):2101–2113, 11 2022. ISSN 2058-5276. doi: 10.1038/s41564-022-01264-z. URL https://www.nature.com/articles/s41564-022-01264-z.

[322] Kazuma Sekiba, Motoyuki Otsuka, Kazuyoshi Funato, Yu Miyakawa, Eri Tanaka, Takahiro Seimiya, Mari Yamagami, Takeya Tsutsumi, Kazuya Okushin, Kei Miyakawa, Akihide Ryo, and Kazuhiko Koike. HBx-induced degradation of Smc5/6 complex impairs homologous recombination-mediated repair of damaged DNA. *Journal of Hepatology*, 76(1):53–62, 1 2022. ISSN 16000641. doi: 10.1016/j.jhep.2021.08.010. URL http://www.journal-of-hepatology.eu/article/S0168827821020055/fulltexthttp://www.journal-of-hepatology.eu/article/S0168827821020055/abstracthttps://www.journal-of-hepatology.eu/article/S0168-8278(21)02005-5/abstract.

[323] Chunyan Han, Dun Zhang, Chenwu Gui, Liang Huang, Sijia Chang, Lianghui Dong, Lei Bai, Shuwen Wu, and Ke Lan. KSHV RTA antagonizes SMC5/6 complex-induced viral chromatin compaction by hijacking the ubiquitin-proteasome system. *PLOS Pathogens*, 18(8):e1010744, 8 2022. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1010744. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1010744.

[324] Liane Dupont, Stuart Bloor, James C. Williamson, Sergio Martínez Cuesta, Raven Shah, Ana Teixeira-Silva, Adi Naamati, Edward J.D. Greenwood, Stefan G. Sarafianos, Nicholas J. Matheson, and Paul J. Lehner. The SMC5/6 complex compacts and silences unintegrated HIV-1 DNA and is antagonized by Vpr. *Cell host & microbe*, 29(5):792–805, 5 2021. ISSN 1934-6069. doi: 10.1016/J.CHOM.2021.03.001. URL https://pubmed.ncbi.nlm.nih.gov/33811831/.

[325] Ishak D. Irwan and Bryan R. Cullen. The SMC5/6 complex: An emerging antiviral restriction factor that can silence episomal DNA. *PLOS Pathogens*, 19 (3):e1011180, 3 2023. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1011180. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1011180.

[326] Joseph M. Dybas, Krystal K. Lum, Katarzyna Kulej, Emigdio D. Reyes, Richard Lauman, Matthew Charman, Caitlin E. Purman, Robert T. Steinbock, Nicholas Grams, Alexander M. Price, Lydia Mendoza, Benjamin A. Garcia, and Matthew D. Weitzman. Adenovirus Remodeling of the Host Proteome and Host Factors Associated with Viral Genomes. *mSystems*, 6(4):468–489, 8 2021. doi: 10.1128/MSYSTEMS.00468-21. URL https://journals.asm.org/doi/10.1128/mSystems.00468-21.

[327] Christy S. Varghese, Joanna L. Parish, and Jack Ferguson. Lying low-chromatin insulation in persistent DNA virus infection. *Current Opinion in Virology*, 55:101257, 8 2022. ISSN 1879-6257. doi: 10.1016/J.COVIRO.2022.101257.

[328] Salum J. Lidenge, Andrew V. Kossenkov, For Yue Tso, Jayamanna Wickramasinghe, Sara R. Privatt, Owen Ngalamika, John R. Ngowi, Julius Mwaiselage, Paul M. Lieberman, John T. West, and Charles Wood. Comparative transcriptome analysis of endemic and epidemic Kaposi's sarcoma (KS) lesions and the secondary role of HIV-1 in KS pathogenesis. *PLOS Pathogens*, 16(7):e1008681, 7 2020. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1008681. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1008681.

[329] circHIPK3 prevents cardiac senescence by acting as a scaffold to recruit ubiquitin ligase to degrade HuR - ProQuest. URL https://www.proquest.com/openview/fb7fa5c63f402f58689267c596393d71/1?pq-origsite=gscholar&cbl=5263173.

[330] Xiu-Yan Huang, Zi-Li Huang, Yong-Hua Xu, Qi Zheng, Zi Chen, Wei Song, Jian Zhou, Zhao-You Tang, and Xin-Yu Huang. Comprehensive circular RNA profiling reveals the regulatory role of the circRNA-100338/miR-141-3p pathway in hepatitis B-related hepatocellular carcinoma. *Scientific Reports*, 7(1): 5428, 12 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-05432-8. URL http://www.nature.com/articles/s41598-017-05432-8.

[331] Jacob Cardenas, Uthra Balaji, and Jinghua Gu. Cerina: systematic circRNA functional annotation based on integrative analysis of ceRNA interactions. *Scientific Reports*, 10(1), 12 2020. ISSN 20452322. doi: 10.1038/S41598-020-78469-X. URL /pmc/articles/PMC7746713//pmc/articles/PMC7746713/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7746713/.

[332] Lauren A. Gay, Sunantha Sethuraman, Merin Thomas, Peter C. Turner, and Rolf Renne. Modified Cross-Linking, Ligation, and Sequencing of Hybrids (qCLASH) Identifies Kaposi's Sarcoma-Associated Herpesvirus MicroRNA Targets in Endothelial Cells. *Journal of virology*, 92(8), 4 2018. ISSN 1098-5514. doi: 10.1128/JVI.02138-17. URL https://pubmed.ncbi.nlm.nih.gov/29386283/.

[333] Lauren A. Gay, Daniel Stribling, Peter C. Turner, and Rolf Renne. Kaposi's Sarcoma-Associated Herpesvirus MicroRNA Mutants Modulate Cancer Hallmark Phenotypic Differences in Human Endothelial Cells. *Journal of Virology*, 95(7), 3 2021. ISSN 0022-538X. doi: 10.1128/JVI.02022-20. URL /pmc/articles/PMC8092706//pmc/articles/PMC8092706/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8092706/.

[334] Jana Haas, Daniel Bloesel, Susanne Bacher, Michael Kracht, and M. Lienhard Schmitz. Chromatin Targeting of HIPK2 Leads to Acetylation-Dependent Chromatin Decondensation. *Frontiers in Cell and Developmental Biology*, 8, 9 2020. ISSN 2296634X. doi: 10.3389/FCELL.2020.00852/FULL. URL /pmc/articles/PMC7490299//pmc/articles/PMC7490299/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7490299/.

[335] Erica L. Sanchez, Thomas H. Pulliam, Terri A. Dimaio, Angel B. Thalhofer, Tracie Delgado, and Michael Lagunoff. Glycolysis, Glutaminolysis, and Fatty Acid Synthesis Are Required for Distinct Stages of Kaposi's Sarcoma-Associated Herpesvirus Lytic Replication. *Journal of Virology*, 91(10), 5 2017. ISSN 0022-538X. doi: 10.1128/JVI.02237-16. URL /pmc/articles/PMC5411582//pmc/articles/PMC5411582/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5411582/.

[336] Tingting Li and Shou Jiang Gao. Metabolic reprogramming and metabolic sensors in KSHV-induced cancers and KSHV infection. *Cell & Bioscience*, 11(1), 12 2021. ISSN 20453701. doi: 10.1186/S13578-021-00688-0. URL /pmc/articles/PMC8475840//pmc/articles/PMC8475840/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8475840/.

[337] Sarah A. Clark, Angie Vazquez, Kelsey Furiya, Madeleine K. Splattstoesser, Abdullah K. Bashmail, Haleigh Schwartz, Makaiya Russell, Shun-Je Bhark, Osvaldo K. Moreno, Morgan McGovern, Eric R. Owsley, Timothy A. Nelson, Erica L. Sanchez, and Tracie Delgado. Rewiring of the Host Cell Metabolome and Lipidome during Lytic Gammaherpesvirus Infection Is Essential for Infectious-Virus Production. *Journal of Virology*, 97(6), 6 2023. ISSN 0022-538X. doi: 10.1128/JVI.00506-23. URL https://journals.asm.org/doi/10.1128/jvi.00506-23.

[338] Eric M. Burton and Benjamin E. Gewurz. Epstein–Barr virus oncoprotein–driven B cell metabolism remodeling. *PLOS Pathogens*, 18(2):e1010254, 2 2022. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1010254. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1010254.

[339] Livia Vastag, Emre Koyuncu, Sarah L. Grady, Thomas E. Shenk, and Joshua D. Rabinowitz. Divergent Effects of Human Cytomegalovirus and Herpes Simplex Virus-1 on Cellular Metabolism. *PLOS Pathogens*, 7(7):e1002124, 7 2011. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1002124. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1002124.

[340] William Stedman, Hyojeung Kang, Shu Lin, Joseph L. Kissil, Marisa S. Bartolomei, and Paul M. Lieberman. Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *The EMBO Journal*, 27(4):654–666, 2 2008. ISSN 1460-2075. doi: 10.1038/EMBOJ.2008.1. URL https://onlinelibrary.wiley.com/doi/full/10.1038/emboj.2008.1https://onlinelibrary.wiley.com/doi/abs/10.1038/emboj.2008.1https://www.embopress.org/doi/10.1038/emboj.2008.1.

[341] Da Jiang Li, Dinesh Verma, Tim Mosbruger, and Sankar Swaminathan. CTCF and Rad21 Act as Host Cell Restriction Factors for Kaposi's Sarcoma-Associated Herpesvirus (KSHV) Lytic Replication by Modulating Viral Gene Transcription. *PLOS Pathogens*, 10(1):e1003880, 1 2014. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1003880. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003880.

[342] Adrien Decorsière, Henrik Mueller, Pieter C. Van Breugel, Fabien Abdul, Laetitia Gerossier, Rudolf K. Beran, Christine M. Livingston, Congrong Niu, Simon P. Fletcher, Olivier Hantz, and Michel Strubin. Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature*, 531 (7594):386–389, 3 2016. ISSN 1476-4687. doi: 10.1038/NATURE17170. URL https://pubmed.ncbi.nlm.nih.gov/26983541/.

[343] Hongyi Pan, Fuchun Zhou, and Shou Jiang Gao. Kaposi's Sarcoma-Associated Herpesvirus Induction of Chromosome Instability in Primary Human Endothelial Cells. *Cancer research*, 64(12):4064, 6 2004. ISSN 00085472. doi: 10.1158/0008-5472.CAN-04-0657. URL /pmc/articles/PMC5257260//pmc/articles/PMC5257260/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5257260/.

[344] Pravinkumar Purushothaman, Prerna Dabral, Namrata Gupta, Roni Sarkar, and Subhash C Verma. KSHV Genome Replication and Maintenance. *Frontiers in microbiology*, 7:54, 2016. ISSN 1664-302X. doi: 10.3389/fmicb.2016.00054. URL http://www.ncbi.nlm.nih.gov/pubmed/26870016http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4740845.

[345] Benjamin Barankin, Andrei I. Metelitsa, Eric H. Schloss, and Norman R. Wasel. Skin disorders in Ashkenazi Jews: a review. *International Journal of Dermatology*, 44(8):630–635, 8 2005. ISSN 1365-4632. doi: 10.1111/J.1365-4632.2005.02625.X. URL https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-4632.2005.02625.xhttps://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-4632.2005.02625.xhttps://onlinelibrary.wiley.com/doi/10.1111/j.1365-4632.2005.02625.x.

[346] Laurie J. Ozelius, Geetha Senthil, Rachel Saunders-Pullman, Erin Ohmann, Amanda Deligtisch, Michele Tagliati, Ann L. Hunt, Christine Klein, Brian Henick, Susan M. Hailpern, Richard B. Lipton, Jeannie Soto-Valencia, Neil Risch, and Susan B. Bressman. LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. *The New England journal of medicine*, 354(4):424–425, 1 2006. ISSN 1533-4406. doi: 10.1056/NEJMC055509. URL https://pubmed.ncbi.nlm.nih.gov/16436782/.

[347] Andrea Wetzel, Si Hang Lei, Tiansheng Liu, Michael P Hughes, Yunan Peng, Tristan Mckay, Simon N Waddington, Simone Grannò, Ahad A Rahim, and Kirsten Harvey. Dysregulated Wnt and NFAT signaling in a Parkinson's disease LRRK2 G2019S knock-in model. doi: 10.1101/2023.03.31.535090. URL https://doi.org/10.1101/2023.03.31.535090.

[348] Martin Steger, Francesca Tonelli, Genta Ito, Paul Davies, Matthias Trost, Melanie Vetter, Stefanie Wachter, Esben Lorentzen, Graham Duddy, Stephen Wilson, Marco A.S. Baptista, Brian K. Fiske, Matthew J. Fell, John A. Morrow, Alastair D. Reith, Dario R. Alessi, and Matthias Mann. Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *eLife*, 5(JANUARY2016), 1 2016. ISSN 2050084X. doi: 10.7554/ELIFE.12813.001.

[349] Zhihua Liu, Jinwoo Lee, Scott Krummey, Wei Lu, Huaibin Cai, and Michael J. Lenardo. The kinase LRRK2 is a regulator of the transcription factor NFAT that modulates the severity of inflammatory bowel disease. *Nature Immunology 2011 12:11*, 12(11):1063–1070, 10 2011. ISSN 1529-2916. doi: 10.1038/ni.2113. URL https://www.nature.com/articles/ni.2113.

[350] Changyoun Kim, Alexandria Beilina, Nathan Smith, Yan Li, Minhyung Kim, Ravindran Kumaran, Alice Kaganovich, Adamantios Mamais, Anthony Adame, Michiyo Iba, Somin Kwon, Won Jae Lee, Soo Jean Shin, Robert A. Rissman, Sungyong You, Seung Jae Lee, Andrew B. Singleton, Mark R. Cookson, and Eliezer Masliah. LRRK2 mediates microglial neurotoxicity via NFATc2 in rodent models of synucleinopathies. *Science translational medicine*, 12(565), 10 2020. ISSN 19466242. doi: 10.1126/SCITRANSLMED.AAY0399. URL /pmc/articles/PMC8100991//pmc/articles/PMC8100991/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8100991/.

[351] Mansoureh Hakimi, Thirumahal Selvanantham, Erika Swinton, Ruth F. Padmore, Youren Tong, Ghassan Kabbach, Katerina Venderova, Stephen E. Girardin, Dennis E. Bulman, Clemens R. Scherzer, Matthew J. Lavoie, Denis Gris, David S. Park, Jonathan B. Angel, Jie Shen, Dana J. Philpott, and Michael G. Schlossmacher. Parkinson's disease-linked LRRK2 is expressed in circulating and tissue immune cells and upregulated following recognition of microbial structures. *Journal of neural transmission (Vienna, Austria : 1996)*, 118(5):795–808, 5 2011. ISSN 1435-1463. doi: 10.1007/S00702-011-0653-2. URL https://pubmed.ncbi.nlm.nih.gov/21552986/.

[352] Agnès Gardet, Yair Benita, Chun Li, Bruce E. Sands, Isabel Ballester, Christine Stevens, Joshua R. Korzenik, John D. Rioux, Mark J. Daly, Ramnik J. Xavier, and Daniel K. Podolsky. LRRK2 is involved in the IFN-gamma response and host response to pathogens. *Journal of immunology (Baltimore, Md. : 1950)*, 185(9):5577–5585, 11 2010. ISSN 1550-6606. doi: 10.4049/JIMMUNOL.1000548. URL https://pubmed.ncbi.nlm.nih.gov/20921534/.

[353] Bojan Shutinoski, Mansoureh Hakimi, Irene E. Harmsen, Michaela Lunn, Juliana Rocha, Nathalie Lengacher, Yi Yuan Zhou, Jasmine Khan, Angela Nguyen, Quinton Hake-Volling, Daniel El-Kodsi, Juan Li, Azadeh Alikashani, Claudine Beauchamp, Jay Majithia, Kevin Coombs, Derya Shimshek, Paul C. Marcogliese, David S. Park, John D. Rioux, Dana J. Philpott, John M. Woulfe, Shawn Hayley, Subash Sad, Julianna J. Tomlinson, Earl G. Brown, and Michael G. Schlossmacher. Lrrk2 alleles modulate inflammation during microbial infection of mice in a sex-dependent manner. *Science translational medicine*, 11(511), 9 2019. ISSN 1946-6242. doi: 10.1126/SCITRANSLMED.AAS9292. URL https://pubmed.ncbi.nlm.nih.gov/31554740/.

[354] W. H. Jopling. Clinical similarities between kaposi's sarcoma and lepromatous leprosy. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 71(6):554, 1 1977. ISSN 18783503. doi: 10.1016/0035-9203(77)90158-4/2/71-6-554A.PDF.GIF. URL https://academic.oup.com/trstmh/article/71/6/554/1873442.

[355] SHERI L. RATOOSH, PHILIP R. COHEN, and PATRICIA TRONCOSO. Cutaneous Malignancy and Leprosy. *The Journal of Dermatologic Surgery and Oncology*, 20(9):613–618, 9 1994. ISSN 1524-4725. doi: 10.1111/J.1524-4725.1994.TB00154.X. URL https://onlinelibrary.wiley.com/doi/full/10.1111/j.1524-4725.1994.tb00154.xhttps://onlinelibrary.wiley.com/doi/abs/10.1111/j.1524-4725.1994.tb00154.xhttps://onlinelibrary.wiley.com/doi/10.1111/j.1524-4725.1994.tb00154.x.

[356] Monica Dallmann-Sauer, Yong Zhong Xu, Ana Lúcia França da Costa, Shao Tao, Tiago Araujo Gomes, Rhana Berto da Silva Prata, Wilian Correa-Macedo, Jérémy Manry, Alexandre Alcaïs, Laurent Abel, Aurélie Cobat, Vinicius M. Fava, Roberta Olmo Pinheiro, Flavio Alves Lara, Christian M. Probst, Marcelo T. Mira, and Erwin Schurr. Allele-dependent interaction of LRRK2 and NOD2 in leprosy. *PLOS Pathogens*, 19(3):e1011260, 3 2023. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1011260. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1011260.

[357] Cleverson Teixeira Soares, Patrícia Sammarco Rosa, Ana Paula Fávaro Trombone, Luciana Raquel Vicenzi Fachin, Cássio César Ghidella, Somei Ura, Jaison Antonio Barreto, and Andréa de Faria Fernandes Belone. Angiogenesis and Lymphangiogenesis in the Spectrum of Leprosy and Its Reactional Forms. *PLOS ONE*, 8(9):e74651, 9 2013. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0074651. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0074651.

[358] Dirk P Dittmer and Blossom Damania. Kaposi sarcoma-associated herpesvirus: immunobiology, oncogenesis, and therapy. *The Journal of clinical investigation*, 126(9):3165–75, 2016. ISSN 1558-8238. doi: 10.1172/JCI84418. URL http://www.ncbi.nlm.nih.gov/pubmed/27584730http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5004954.

[359] Mark N. Polizzotto, Thomas S. Uldrick, Kathleen M. Wyvill, Karen Aleman, Cody J. Peer, Margaret Bevans, Irini Sereti, Frank Maldarelli, Denise Whitby, Vickie Marshall, Priscila H. Goncalves, Vikram Khetani, William D. Figg, Seth M. Steinberg, Jerome B. Zeldis, and Robert Yarchoan. Pomalidomide for Symptomatic Kaposi's Sarcoma in People With and Without HIV Infection: A Phase I/II Study. *Journal of Clinical Oncology*, 34(34):4125, 12 2016. ISSN 15277755. doi: 10.1200/JCO.2016.69.3812. URL /pmc/articles/PMC5477825//pmc/articles/PMC5477825/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5477825/.

[360] C. H. Hsu, D. Chan, and B. Wolozin. LRRK2 and the Stress Response: Interaction with MKKs and JNK-Interacting Proteins. *Neurodegenerative Diseases*, 7(1-3):68–75, 4 2010. ISSN 1660-2854. doi: 10.1159/000285509. URL https://dx.doi.org/10.1159/000285509.

[361] Laurie Sanders, S.D. Goodson, C. M. Toste, and C.P. Gonzalez-Hunt. ATM mediated DNA double-strand breaks accumulate in LRRK2 G2019S Parkinson's disease. *The FASEB Journal*, 35(S1), 5 2021. ISSN 1530-6860. doi: 10.1096/FASEBJ.2021.35.S1.03622. URL https://onlinelibrary.wiley.com/doi/full/10.1096/fasebj.2021.35.S1.03622https://onlinelibrary.wiley.com/doi/abs/10.1096/fasebj.2021.35.S1.03622https://faseb.onlinelibrary.wiley.com/doi/10.1096/fasebj.2021.35.S1.03622.

[362] Zhongcan Chen, Zhen Cao, Wei Zhang, Minxia Gu, Zhi Dong Zhou, Baojie Li, Jing Li, Eng King Tan, and Li Zeng. LRRK2 interacts with ATM and regulates Mdm2–p53 cell proliferation axis in response to genotoxic stress. *Human Molecular Genetics*, 26(22):4494–4505, 11 2017. ISSN 0964-6906. doi: 10.1093/HMG/DDX337. URL https://academic.oup.com/hmg/article/26/22/4494/4096717.

[363] Joon Y. Boon, Julien Dusonchet, Chelsea Trengrove, and Benjamin Wolozin. Interaction of LRRK2 with kinase and GTPase signaling cascades. *Frontiers in Molecular Neuroscience*, 7(JULY):64, 7 2014. ISSN 16625099. doi: 10.3389/FNMOL.2014.00064/BIBTEX.

[364] Kwun Wah Wen and Blossom Damania. Kaposi sarcoma-associated herpesvirus (KSHV): Molecular biology and oncogenesis. *Cancer Letters*, 289 (2):140–150, 3 2010. ISSN 03043835. doi: 10.1016/j.canlet.2009.07.004. URL http://www.ncbi.nlm.nih.gov/pubmed/19651473http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4342847https://linkinghub.elsevier.com/retrieve/pii/S0304383509004972.

[365] Carolyn Ann Robinson, Gillian K. Singh, Mariel Kleer, Thalia Katsademas, Elizabeth L. Castle, Bre Q. Boudreau, and Jennifer A. Corcoran. Kaposi's sarcoma-associated herpesvirus (KSHV) utilizes the NDP52/CALCOCO2 selective autophagy receptor to disassemble processing bodies. *PLOS Pathogens*, 19(1):e1011080, 1 2023. ISSN 1553-7374. doi: 10.1371/JOURNAL.PPAT.1011080. URL https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1011080.

[366] Annabel T. Olson, Yuqi Kang, Anushka M. Ladha, Chuan Bian Lim, Michael Lagunoff, Taran S. Gujral, and Adam P. Geballe. Polypharmacology-based kinome screen identifies new regulators of KSHV reactivation. *bioRxiv*, 2 2023. doi: 10.1101/2023.02.01.526589. URL /pmc/articles/PMC9915688//pmc/articles/PMC9915688/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC9915688/.