Understanding the role of Dead box helicase 1 (DDX1) and its co-factors in alphavirus infection

Natasha Palmalux

December 2024

A thesis submitted in fulfilment for the degree of Doctor of Philosophy in Virology

MRC - Centre for Virus Research - School of Infection and Immunity - College of Medical, Veterinary and Life Sciences - University of Glasgow

Abstract

The tRNA ligase complex (tRNA LC) is essential in tRNA maturation, stress response pathways and viral regulation, among others. It comprises proteins with distinct roles, including a non-canonical GMP-driven RNA ligase, RTCB, a cap-binding protein CGI99, and an ATP-dependent RNA helicase DDX1. In Sindbis virus (SINV), the tRNA-LC relocalises to viral replication organelles and interacts directly with viral RNA (vRNA). However, the functional role of the tRNA-LC in SINV infection remained unclear. This thesis characterises tRNA-LC interaction dynamics and elucidates its antiviral mechanism during SINV infection. Co-precipitation and crosslinking mass spectrometry (XL-MS) demonstrated robust inter-protein interactions, with DDX1, RTCB, and CGI99 forming the core of a tightly coordinated complex. CGI99 emerged as a central component, corroborated by Alphafold3 (AF3) modelling and complex destabilisation following a siRNA-mediated knockdown. Novel interactions, such as those with RPL11 and MYH9, suggest broader functional implications. The tRNA-LC displayed potent antiviral activity, as depletion of DDX1 and CGI99 significantly increased viral protein production and downregulated over 11,000 host genes during SINV infection. These findings indicate that tRNA-LC plays a central role in restricting viral lifecycle, and in its absence, the cellular microenvironment is more favourable to host viral infection. To identify the antiviral mechanism exerted by the tRNA-LC, I assessed how its RNA and protein interaction landscapes were altered during infection. The interaction landscape of the tRNA-LC analysed via iCLIP2 and protein-protein interaction analysis revealed a transition from cellular mRNA to vRNA binding during infection, primarily targeting the 5' UTR and the start of the coding sequence. Enhanced interactions with ribosomal factors suggested involvement in translation regulation. Using a SINV replicon system, DDX1 was identified as regulating both viral replication and translation. This study proposes that the tRNA-LC inhibits vRNA processes by blocking essential viral factors such as replicase or translation components accessing the vRNA, offering new insights into its antiviral mechanisms and potential applications against positive-strand RNA viruses.

Acknowledgements

It takes a village to raise a child, and it took a village to get me to where I am today and for this thesis to be a reality. This thesis is so much more than a body of scientific results. The journey to get to this point would not have been possible without the incredible support from the people who advised, criticized, laughed, cried, gossiped, and ranted with me over the last four years.

The MRC provided invaluable funding to pursue this research, and the CVR as a whole has supported the application of my advances. Without the expertise of the institute's technical and research groups, none of this would have been possible. Thank you to everyone who works in and out of the golden building that houses our institute. I would love to name everyone, but it might become longer than the thesis itself, so hopefully, you know who you are. Some definitely need to be named...

Alfredo thank you for teaching me to be a scientist. To be critical but also imaginative and think up all the possibilities. Your passion and excitement for science is contagious. You are a deep well of knowledge. I have always been impressed by how you find a silver lining in even the worst data that I have shown you, even if it is just my labelling. It's not always been easy, from an unorthodox start via zoom, unsuccessful repeat experiments and confusing phenotypes. Where I had doubts, you found... a silver lining. Thank you for letting me join your lab and sharing with me your tips and tricks, hopefully some have been instilled in me. Thank you for helping me become more resilient and a better scientist.

Marko, once upon a time it was just the two of us. Benches galore, unrestricted space, no bookings required, and reagents just where we left them. Ah, the days. Thank you for getting me on my researcher feet and teaching me how to be independent and resourceful.

I need to thank the Castello lab as a whole. Past and present. The shared opinions and support have been great. The lab meetings were both nerve-wracking and helpful. Can't find a group of harsher critics, yet now conferences hold no fear. I worry the organisation of the labs will fall to tethers, but hopefully, I have taught you all something, too! Namah and Rozeena as the "youngest" members, the baton is passed to you.

Ana. It's funny to think back to where we started. I was once bright-eyed, bushy-tailed and excited to start at the CVR working under you. At the time, the Genomics group was a 4-person strong, mighty group and me, the newest member. Now I am a little hardened, and my tail is a little ruffled, and your group has more than doubled in size. You are an incredible person and you are doing amazing work. Throughout my journey in research you have been

there. You were the best boss I had ever had and when I started my PhD, you remained the most amazing support. I am incredibly appreciative of you being part of my advisory team. I have loved our informal chats and your scientific advice and ideas. Thank you for being the incredible, thoughtful person that you are.

I would be amiss if I didn't mention my second group, the one I never left (at least not all of me), the CVR Genomics group (and superuser attaches). Lily, Kathy, Danny, Jenna and Alice. Thank you for introducing me to the CVR, and teaching me the world of sequencing, but mostly, thank you for being great friends. Chats over library preps, where the 5min incubation turned to 10 or maybe more. The crazy COVID months, where I truly appreciated just who I was working side by side with. More recently thank you for the catch-ups and welcoming me in any mood I came in (I am sorry!). Kathy our lab mum and Lily our (my) hillwalking mum. We sadly lost the real mothering hen of our small group, Kiki. An incredible woman who you could speak to all day, who genuinely supported and loved everyone she worked with. You were an incredible scientist, insightful, passionate but most of all kind. Kiki, I am saddened you will not read this, but you were one of the first people I wanted to share good news with and would have loved to tell you "I did it".

To the people who shared lab benches, TC hoods and office space with me, thank you for putting up with me. Nicole, thank you for your external perspective and insightful ideas in virology research, cloning advice, general company during late nights in the lab and chats about life. Zaydah, I will be honest about something. The daily hugs weren't always just for you. I actually really loved them too. "It's been a steep learning curve" but a little less steep when you have someone to lean on.

To the wise duo, Enzo and Azman. Enzo your calm, collected way of approaching any aspect of lab work is definitely something I can learn from. Your resilience and tenacity but most of all your passion for the work we do is inspirational. I worry about your sleep diffusion and your caffeine intake and would love for you to pipette a little faster, but I think that's a me problem, not a you problem. You have been the calm in the storm of my PhD, providing invaluable advice, conference company as well as an amazing friendship. Azman, thank you for being the person I could bother constantly and be an annoying sibling to. I miss our pointless debates about unrealistic possibilities and life problems. Your structured ideologies have definitely made me evaluate a few aspects of my everyday, including the importance of lentejas. Thank you for keeping me sane throughout this PhD, keeping it light and more than bearable. I am grateful

our paths have crossed.

Louie. Our journeys intertwined just at the right moment. I think we met at a point where we really just needed one another. I sometimes feel I have needed you more than the other way round but I am incredibly grateful for having you in my life and throughout this PhD. You are an amazing, intelligent, imaginative and an incredibly talented woman in STEM. You have supported me in the good and dark times, helped me with every aspect of my PhD and all things said, I don't think I would have gotten to this point without your incredible support. I think we pushed and learned from each other, and distracted each other just the right amount. I am fortunate to call my work partner also my very dear friend. Thank you.

Outside of the little world of academia I am infinitely grateful to my friends Becky, Ioana and Rosalie. Becky our therapeutic hikes, dog walks and coffee debriefs were essential to keeping me sane and also remind me of the world outside my PhD. Ioana and Rosalie, oceans aparts yet somehow we never lost touch and your support throughout my academic journey has been incredible. You don't need to understand what I do to be amazing friends and be incredibly supportive.

My parents who against all odds ended up with a daughter in academia. I broke the family trend and went on to be the first PhD in the family. You both kept saying the importance of education and maybe I took it a little too literally. Thank you for encouraging me, urging me on and being my amazing parents. I wouldn't be here without you. Your trust in my abilities is humbling and sometimes unfounded, yet it's been exactly what I needed.

To my home pack. It's no hidden secret that the animals I share my life with have had an incredible contribution to my PhD. Forced mental health walks by a wagging tail or purring company on my lap as I typed at my computer. The simple joy that you bring to my life has kept me afloat. Kiara, Sofia, Kovu and most recently Zorro. These little four legs all deserve their own thank you.

David, there are very few words to say that you don't already know. Thank you for accepting me in every shape I come in. You have kept me grounded, supported me, encouraged me day in and day out. I definitely could not have made it without you standing by me. I love you.

Contents

List of Tables

List of Figures

Author's Declaration

I, Natasha Palmalux, declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Abbreviations

Abbreviation	Full Term
4SU	4-Thiouridine
aa	Amino acid
ADP	Adenosine diphosphate
AEBSF	4-(2-aminoethyl)benzene sulfonyl fluoride hydrochloride
AF2	Alphafold2
AF3	Alphafold3
Å	Angstrom
AP-MS	Affinity purification-mass spectrometry
ASW	Ashwin protein
ATP	Adenosine triphosphate
bp	Base pair
BP	Biological process
BSL	Bio-safety level
CBC	Cap-binding complex
CC	Cellular Complex
CDS	Coding sequence
CHIKV	Chikungunya virus
circRNA	Circular RNA
CLAMP	Crosslink-assisted messenger RNP purification
CLIP	Crosslinking and Immunoprecipitation
CMV	Cytomegalovirus
Co-IP	Co-immunoprecipitation
Da	Dalton
DENV	Dengue virus
DDX	DEAD/H-box helicase
DDX1	DEAD-box helicase 1
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid

DSS	Disuccinimidyl suberate
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
E1	Envelope protein 1
E2	Envelope protein 2
EEEV	Eastern equine encephalitis virus
eIF	Eukaryotic initiation factor
EM	Electron microscopy
ENSEMBL	Genome browser database
ER	Endoplasmic reticulum
FAM98A	Family with sequence similarity 98, member A
FAM98B	Family with sequence similarity 98, member B
FASN	Fatty acid synthase
FBS	Foetal bovine serum
FDR	False discovery rate
GEMIN5	Gem-associated protein 5
GFP	Green fluorescent protein
GMP	Guanosine monophosphate
GO	Gene ontology
gRNA	Genomic RNA
GTP	Guanosine triphosphate
HCV	Hepatitis C Virus
HEK293	Human embryonic kidney cells
HeLa	Henrietta Lacks cells
HIV	Human immunodeficiency virus
hnRNP	Heterogeneous nuclear ribonucleoprotein
hpi	Hours post infection
hpt	Hours post transfection
IAV	Influenza A virus
IRES	Internal ribosome entry site
ISG	Interferon-stimulated genes
IFIT	Interferon-induced proteins with tetratricopeptide repeats

IFN	Interferon
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IP	Immunoprecipitation
JAK-STAT	Janus kinase-signal transducer and activator of transcription pathway
kbp	Kilo base pairs
kDa	Kilo Dalton
KD	Knockdown
KH	K Homology domain
KO	Knockout
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LNA	Locked nucleic acid
lncRNA	Long non-coding RNA
LTR	Long-terminal repeat
m7G	7-Methylguanosine
M	Molar
MAYV	Mayaro virus
MF	Molecular Function
min	Minutes
ml	Millilitre
mM	Millimolar
mRNA	Messenger RNA
MYH9	Myosin heavy chain 9 protein
MZT	Maternal-to-zygotic transition
NaCl	Sodium chloride
NEB	New England Biolabs
nm	Nanometres
NPC	Nuclear pore complex
ns	Non-structural
ONNV	O'nyong'nyong virus
ORF	Open reading frame
PCR	Polymerase chain reaction

PDB	Protein Data Bank
PKR	Protein kinase R
PRR	Pattern recognition receptor
PTM	Post-translational modification
RBD	RNA-binding domain
RBP	RNA-binding protein
RFP	Red fluorescent protein
RIC	RNA interactome capture
RIG-I	Retinoic acid-inducible gene I
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RRM	RNA recognition motif
RRV	Ross River virus
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
SARS-CoV	Severe acute respiratory syndrome coronavirus
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SESV	Southern elephant seal virus
SFV	Semliki Forest virus
sgRNA	Subgenomic RNA
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SINV	Sindbis virus
SPVD	Salmon pancreatic disease virus
tRNA	Transfer RNA
tRNA-LC	tRNA ligase complex
UPR	Unfolded protein response
UV	Ultraviolet
vRIC	Viral RNA interactome capture
vRNA	Viral RNA
vRNP	Viral ribonucleoprotein
VEEV	Venezuelan equine encephalitis virus
VIR-CLASP	Viral crosslinking and solid phase purfication

VSVG	Vesicular stomatitis virus glycoprotein
WCL	Whole cell lysate
WEEV	Western equine encephalitis virus
WT	Wild-type
XBP1	X-box binding protein 1
XRN1	5’-3’ Exoribonuclease 1
ZAP	Zinc finger antiviral protein
ZIKV	Zika virus
ZnF	Zinc finger

1 Introduction

1.1 Alphaviruses

Alphaviruses are enveloped, single-stranded, positive-sense RNA viruses which are primarily transmitted through arthropod vectors, predominantly mosquitoes. They are known to cause various diseases in humans and animals [1, 2]. In humans, typical alphavirus infection can result in rash, arthritis, encephalitis, and death. They have garnered increasing attention due to their public health implications, causing worldwide outbreaks with substantial morbidity. Within the past decade, infection by Chikungunya virus (CHIKV), a member of the genus Alphavirus, has spread far further than its original discovery in sub-Saharan Africa in the 1950s [3]. CHIKV has been responsible for two worldwide epidemics with millions of cases in the last 15 years. In 2005, approximately six million cases were reported across 40 countries affecting a large part of East Africa, India and Southeast Asia, as well as southern Europe [4–9]. In 2013, the second large-scale epidemic had approximately two million cases across 50 countries, affecting a large proportion of the Caribbean and disseminated across the American continent [10, 11]. The expansion of the CHIKV vector, *Aedes albopictus*, poses the emerging threat and risk of further dissemination worldwide, enabled by climate change, international trade and travel, globalisation and habitat loss [12]. Understanding alphavirus-host interaction and lifecycle is essential for developing effective therapeutic interventions. This thesis focuses on the human-virus interactions.

1.1.1 Taxonomy and geographic distribution of alphaviruses

Alphaviruses were amongst the first arboviruses to be isolated, characterised and assigned a taxonomic status. They are zoonotic pathogens transmitted through arthropod vectors infecting rodents, primates, and birds. Alphaviruses are mainly carried by *Aedes*, *Culiseta*, and *Culex* mosquito species, in which no pathological effect is observed [13, 14]. Part of the Togaviridae family, there are 30 recognised alphavirus species spread across different phylogenetic groups [15]. These are then further classed as encephalitic (also known as New World), arthritogenic (also known as Old World) or aquatic alphaviruses by their geographical origin, distinct symptomatic manifestations and vector preferences. New World viruses are characterised by their infection of the nervous system causing meningitis and encephalitis with potential long-term neurological effects [16]. Meanwhile, Old World viruses cause

musculoskeletal disease characterised by fever, rash, arthralgia, myalgia, myositis and acute and chronic polyarthritis [17]. A summary of a selection of alphaviruses, their classification, geographical dissemination and vector preferences is outlined in Table 1.1.

Encephalitic alphaviruses identified and isolated in the 1930s in the Americas include Eastern equine encephalitis virus (EEEV) (1933 in New Jersey and Virginia [18]), Western equine encephalitis virus (WEEV) (1930 in California [19]) and Venezuelan equine encephalitis virus (VEEV) (1935 in Colombia, Trinidad and Venezuela [20]).

Arthritogenic alphaviruses, which were discovered later, include Chikungunya virus (CHIKV) (1955 in Tanzania) [3], Ross River virus (RRV) (1949 in Oceania including Australia), O'nyong'nyong virus (ONNV) (1959 in East Africa [21]), Mayaro virus (MAYV) (1954 in Trinidad [22]), Semliki Forest virus (SFV) (1942 in Uganda [23]), Sindbis virus (SINV) (1952 in Egypt [24]) and others. SINV is categorised as an arthritogenic alphavirus, however its genome similarity marks it as being more closely related to encephalitic viruses in North America. In mice, SINV has been observed to cause encephalitis, unlike other arthritogenic alphaviruses, where neuronal progression of the disease is only observed in rare cases [25, 26]. SINV is commonly employed as a model virus in alphavirus research due to its broad host range and ability to infect many cell types. It is categorised as a biosafety class 2 agent due to its lesser pathogenesis, allowing it to be utilised in most laboratories.

Mosquitoes are the primary vectors of alphaviruses, with the exception of aquatic strains: salmon pancreatic disease virus (SPDV), which infects salmon and trout, causing mortality in farmed fish [27, 28], and Southern elephant seal virus (SESV). Both of these viruses have been found within lice, *Lepeophtheirus salmonus* for SPDV, and *Lepidohthirus macrorhini* for SESV [29], which suggests an arthropod-borne cycle. However, no direct arthropod-dependent transmission has been demonstrated.

The genetic diversity within the Alphavirus genus, revealed through molecular phylogenetic analyses, underscores these viruses' adaptability, facilitating their emergence in new geographic areas and species spread (Table 1.1).

Alphavirus	Classification	Geographical distribution	Vector
CHIKV	Arthritogenic	Africa, India, Southeast Asia	*Aedes albopictus, Aedes aegypti*
RRV	Arthritogenic	Australia, South Pacific Islands	*Aedes vigilax, Aedes camptorhynchus, Aedes polynesiensis, Culex annulirostris*
MAYV	Arthritogenic	South America	*Aedes albopictus, Aedes aegypti, Anopheles atroparvus*
SFV	Arthritogenic	Africa	*Aedes albopictus, Aedes aegypti*
SINV	Arthritogenic	South and East Africa, Europe, Isreal, Philippines, Australia	*Culex, Culiseta*
VEEV	Encephalitic	Florida, Central and South America	*Psorophora confinnis, Psorophora columbiae, Aedes sollicitans, Aedes taeniorhynchus, Culex*
EEEV	Encephalitic	Atlantic and Gulf coasts, United States (Indiana, Michigan, Wisconsin)	*Culiseta melanura, Coquillettidia perturbans, Aedes vexans*
SPDV	Aquatic	Europe, Scandinavia, and the United States	No vector transmission confirmed

Table 1.1: Alphavirus classification host receptor and vector summary

1.1.2 Pathogenesis of Alphaviruses

The pathogenesis of alphaviruses is a complex interplay between viral factors and host immune responses, which ultimately determines the clinical outcomes of infection. Alphaviruses that affect humans and their subsequent pathogenesis are split into two distinct groups: arthritogenic and encephalitic [30].

Arthritogenic alphavirus infections can range from mild to severe, with manifestations including fever, rash, and debilitating arthritis in multiple joints, which persists months to years after the resolution of acute infection. The Chikungunya virus, for example, is classed as the most dangerous due to its notorious prolonged arthralgia, known as "breakbone fever" [31,32]. SINV is the infectious agent for diseases like Pogosta, Ocklebo, and Karelian fever, all of which are hallmarked by severe arthralgia-like disease [17]. Infections of the encephalitic alphavirus exhibit higher mortality despite being comparatively rarer clinically. Patients may experience neurological symptoms due to the virus's ability to invade the central nervous system. This neurotropism is particularly concerning, as it can lead to encephalitis, resulting in significant morbidity and mortality [16].

Infection with alphaviruses results in a brief period of viremia (usually 5-7 days). Disease severity and persistence of symptoms are associated with the extension of virus replication and the presence of inflammatory mediators in the plasma of patients [33,34]. In vertebrates, initial sites of alphavirus replication include skeletal muscle and Langerhans cells in the skin, leading to infection of the draining lymph node, although a range of cell types have been shown to be susceptible to infection [35–37]. Central nervous system invasion may also occur through endothelial cells or via infected monocytic cells in the blood. Widespread infection of these cells and the associated inflammatory immune response account for the acute symptoms caused by these viruses [17].

Infection with alphaviruses is primarily controlled by the host's immune response, beginning with an early activation of the innate immune system. Type I interferons (IFN-α/β) play a crucial role in controlling viral replication during the initial stages of infection [38]. IFNs trigger antiviral responses in infected and neighbouring cells, through the induction of antiviral proteins that limit viral replication and promote the clearance of infected cells [39]. IFN-α/β signalling, for example, induces the expression of interferon-stimulated genes (ISGs) that inhibit viral replication at multiple levels and recruit immune cells to the site of infection, thereby containing viral spread within the host [38,40]. However, despite inducing IFN production, alphaviruses

are able to antagonize the type I IFN response. This will be explored in more detail later in this introduction.

As infection progresses, adaptive immunity takes over, with B cells producing virus-specific antibodies and T cells targeting infected cells. Neutralising antibodies bind to viral particles, preventing their entry into host cells, which is crucial for clearing the virus from the bloodstream and controlling infection in tissues. Studies show that the presence of specific IgG and IgM antibodies correlates with reduced viral loads and milder symptoms, especially in infections with viruses like CHIKV [41]. Cytotoxic T cells also play a role by recognizing and eliminating infected cells, thereby limiting viral reservoirs in tissues such as the central nervous system or joints [42].

1.1.3 Alphavirus lifecycle

After inoculation into the vertebrate host, alphaviruses enter permissive and susceptible host cells to manufacture new virions. The alphaviruses are noted to have highly efficient infection [43]. The alphaviral replication process is described in detail here, and a summary can be found in Figure 1.1. The steps outlined are common to most alphaviruses.

The mature virion is composed of a spherical capsid that encases a single strand of vRNA and is enveloped by a lipid bilayer coated with viral glycoprotein spikes [44]. The glycoprotein spike, heterodimer E1 and E2 proteins enable host cell receptor recognition and induction of endocytosis of the viral particle into the cell [45]. As the virus-containing endosome matures, it acidifies, which is critical for a major conformational change of the glycoproteins, and the viral particle undergoes fusion with the endosomal membrane via the fusion peptide of the E1 glycoprotein [46]. The nucleocapsid core is subsequently released into the cytoplasm, which quickly disassembles to release the vRNA for translation [47]. The vRNA is polyadenylated and resembles host mRNA, allowing for direct translation by host translational machinery.

Upon entering the cell, alphaviruses rely on their positive-sense genomic RNA (gRNA) strand for efficient viral synthesis. The vRNA genome ($\tilde{1}$2 kb) encodes four nonstructural proteins (nsP1, nsP2, nsP3 and nsP4) and five structural proteins (capsid, E3, E2, 6K/TransFrame and E1) [45]. Together, these proteins mediate viral transcription, replication and host-cell antagonism. The gRNA is divided into two open reading frames (ORFs). The 5' ORF encodes a nonstructural (ns) polyprotein, nsP1234, which is directly translated from the gRNA upon release in the cytosol. The second ORF, which encodes structural proteins, is

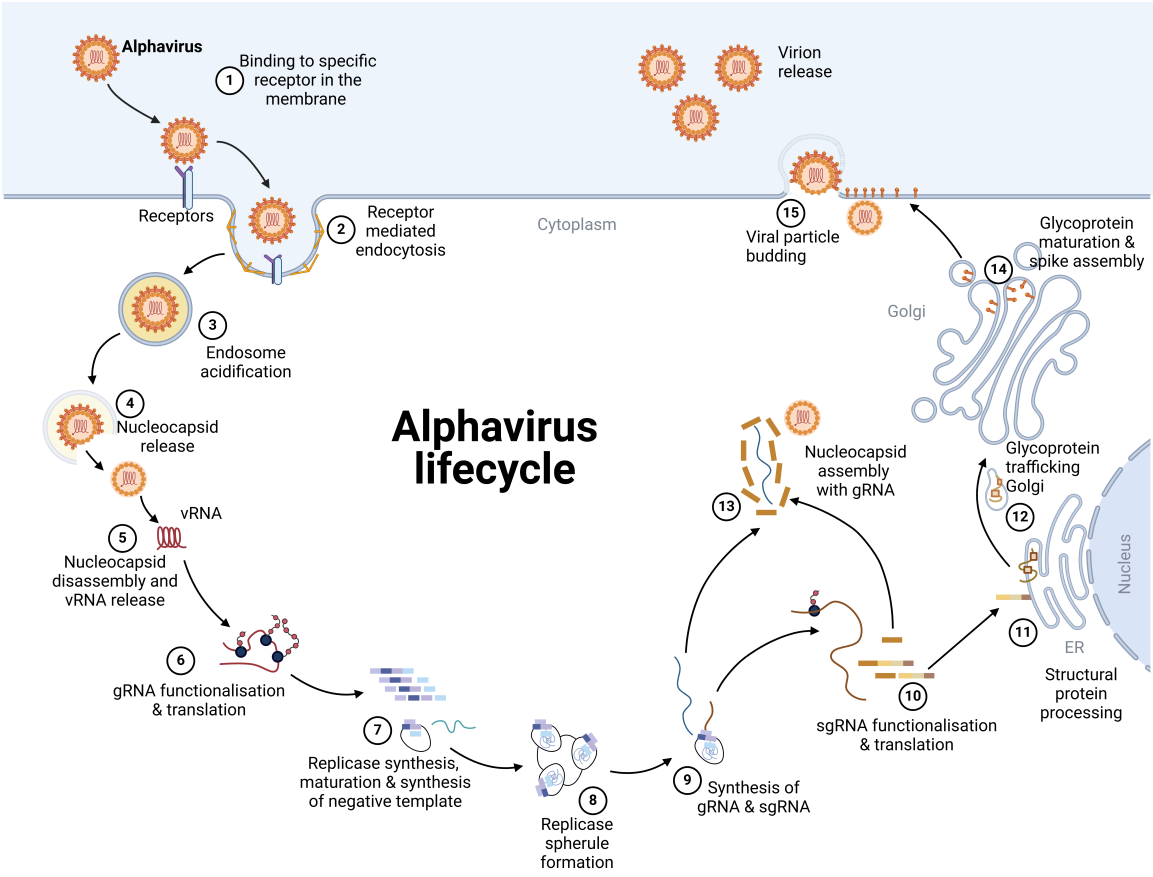

Figure 1.1: **Alphavirus lifecycle schematic**

translated from a subgenomic RNA (sgRNA) transcribed from a negative-strand RNA template during replication [43, 45]. Genomic organisation and the different RNA intermediaries during the viral replication cycle are detailed in Figure 1.2.

Following the release of the gRNA, the alphaviral ns polyprotein, nsP123, is translated from the exposed vRNA as well as an nsP1234 polyprotein due to slippage at the Opal stop codon [48, 49]. The nsP2 component of the polyprotein (and in isolation as a monomer) is a protease. The sequential processing of the three cleavage sites by nsP2 within the nsP1234 regulates the synthesis of different vRNA species [50]. NsP4 is proteolytically cleaved from the polyprotein, releasing it to function as the RNA-dependent RNA polymerase (RdRp). Together, nsP123 and nsP4 assemble to make the initial replicase complex that synthesises the negative strand-RNA [51]. The negative-strand RNA serves as the template for the replication and transcription of the positive-strand gRNA and sgRNA [52]. Its synthesis triggers the next cleavage of the polyprotein, releasing nsP1. The nsP1-P23-P4 complex serves for the synthesis of predominantly the negative-strand RNA. The final cleavage forms the fully processed mature nsP1-nsP2-nsP3-nsP4 replicase complex, which produces positive-strand

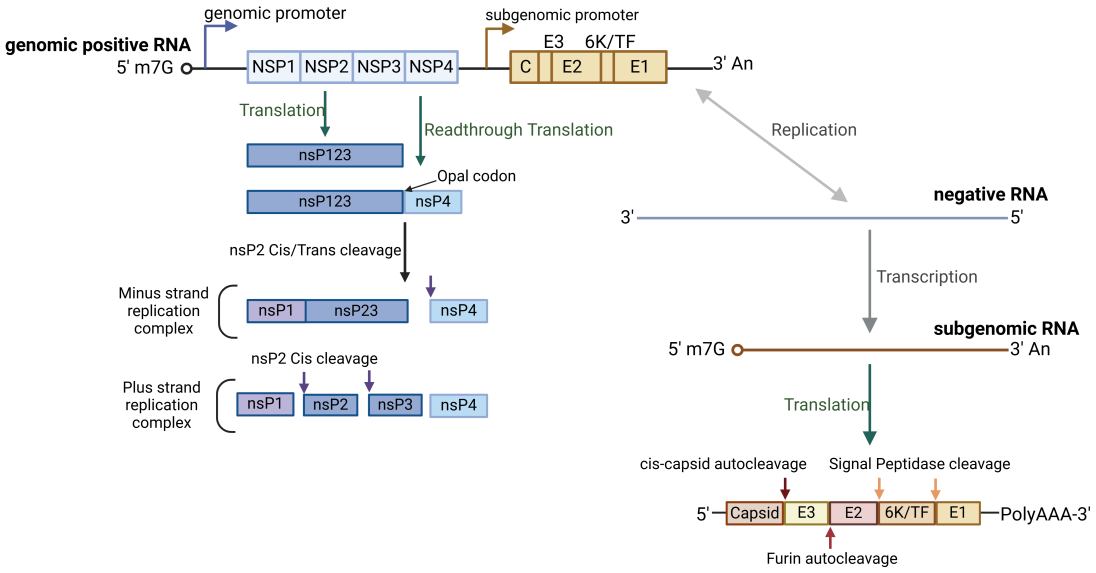

Figure 1.2: **vRNA replication and translation schematic**

RNA exclusively, with a preference for subgenomic RNA synthesis [53]. The vRNA synthesis occurs in invaginated spherules derived from cellular membranes, protecting the dsRNA structure formed during replication from host detection. Synthesised gRNA and sgRNA exit the replication spherules for host-mediated translation, nucleocapsid packaging and assembly.

The structural polyprotein is translated from the sgRNA strand as a polyprotein. The capsid protein autoprotelytically cleaves itself off the actively translating polypeptide and interacts with free gRNA to start forming the nucleocapsid core [54]. Meanwhile, the remaining structural proteins mature through the engagement of host proteases in the ER, such as furin [55]. These proteins then undergo glycosylation, palmitoylation, and disulfide bond rearrangement before being trafficked to and displayed on the host cell surface. The capsid protein of the nucleocapsid core and the cytoplasmic endodomain of the E2 protein then interact, driving the budding of the newly formed virions from the infected host cell [56].

Alphavirus treatment is symptomatic and supportive, with no specific antiviral available. Understanding the alphavirus life cycle and how the virus interacts and modulates the host cell is critical for identifying therapeutic targets and improving clinical outcomes. Research into the molecular mechanisms of viral-host interactions provides valuable insights that could inform vaccine development and therapeutic strategies.

1.2 The emergent roles of RNA-binding proteins (RBPs)

Viruses are obligate obligate intracellular pathogens, that are heavily reliant on host metabolic capabilities to replicate and spread [57]. They are small pathogens that only encode a few proteins. It is thus essential that viruses hijack cellular proteins to facilitate every step of the viral lifecycle: entry, replication, translation, packaging, and assembly of the viral particles [58–60]. Capturing and understanding these host-virus interactions has been a research topic of significant interest. vRNA is an essential molecule within the RNA virus lifecycle, as it functions as a genome, template for replication and transcription, and messenger (m)RNA for protein synthesis. Therefore, cellular RNA-binding proteins (RBPs) have been identified as prime targets for facilitating or inhibiting viral replication [61].

1.2.1 RBP structure and function

RBPs are essential to cellular RNA processing RBPs are a group of proteins that interact with RNA to form ribonucleoprotein (RNP) complexes. These highly dynamic complexes are essential in regulating RNA fate at every stage of the RNA lifecycle [62]. A review published in 2015, compared RBPs to the "mRNA's clothes" [63]. RBPs "dress" and "undress" different regions of the mRNA, guiding its maturation, processing and functional state. RBPs interact with RNA from its synthesis and maturation to its degradation, as described in this section. They associate with RNA immediately after transcription. The nuclear cap-binding complex (CBC), composed of CBP20 and CBP80, mediates the attachment of the 7-methylguanosine cap (m7G) to the 5' end [64]. The cap structure aids in RNA stability, splicing regulation and nuclear export. Furthermore, it is a crucial binding site for other RBPs, such as eukaryotic initiation factors (eIFs), that enable translation initiation. Capped RNA undergoes splicing followed by 3' end polyadenylation. These processes are closely controlled by the spliceosome and the poly(A) polymerase in combination with regulatory RBPs. A host of RBPs influence splice site selection, contributing to alternative splicing, which gives rise to diverse protein products from the same gene sequence, such as Nova and Fox proteins, among others [65,66]. Many RBPs participate in more than one of these processes. For example, NOVA1 is known to regulate both poly(A) and splice site selection [67–69].

After nuclear export, RBPs play an essential role in guiding translation initiation by preparing the mRNA for ribosomal engagement. Translation occurs through recognition of the cap by the heterotrimeric factor eIF4F, followed by the interaction of the preinitiation 43S

complex with the mRNA. The eIF4F complex recognises and assembles on the mRNA's 5' cap, replacing the nuclear CBC. The eIF4F complex consists of three main components: eIF4E, which binds directly to the 5' cap structure; eIF4G, a scaffold protein that interacts with other initiation factors and RBPs; and eIF4A, an RNA helicase that unwinds secondary structures in the 5' UTR of the mRNA. The unwinding of the RNA within the 5' RNA region allows for the small (40S) ribosomal subunit, aided by eIF1 and eIF1A, to linearly scan the leader sequence for the start codon [70, 71]. Upon reaching the AUG start codon, eIF2, which is bound to GTP and methionine-tRNA (Met-tRNAi), pairs with the start codon to establish the reading frame. The hydrolysis of eIF2-GTP stabilises Met-tRNAi at the P-site of the ribosome [72]. Subsequently, the 60S large ribosomal subunit is able to bind, forming the complete 80S ribosome ready for translation elongation. With the translation machinery fully assembled, the mRNA is primed for sequential codon recognition, peptide chain elongation, and the synthesis of the encoded protein. Initiation of translation can also occur by other mechanisms independent of cap recognition, such as internal initiation. In this case, initiation takes place at an internal sequence located at the 5' untranslated region (5'-UTR) of the mRNA, known as the internal ribosome entry site (IRES) [73].

Through this highly orchestrated process, RBPs and initiation factors regulate both the efficiency and fidelity of translation initiation, enabling cells to precisely control protein production in response to cellular needs [74]. Upon completing its role in protein synthesis, mRNA undergoes degradation by proteins such as XRN1, a highly conserved exoribonuclease [75]. This ensures rapid clearance of spent transcripts and maintenance of cellular homeostasis.

RBPs regulatory role in RNA metabolism RBPs play pivotal inhibitory roles in RNA metabolism, adding a crucial layer of regulation to RNA processing and quality control. Beyond their roles in RNA splicing, transport, and stabilisation, RBPs serve as "gatekeepers" at several stages of the RNA lifecycle, particularly during RNA maturation and export from the nucleus. RBPs carefully assess pre-mRNA for structural fidelity and proper processing. Aberrant pre-mRNA transcripts are retained within the nucleus in order to prevent the accumulation and expression of defective transcripts. The exosome complex, a multi-protein RBP complex, is central to nuclear quality control [76]. It is responsible for degrading defective pre-mRNA and noncoding RNAs, inhibiting their release into the cytosol. Further downstream of the RNA

lifecycle, the non-sense mediated decay pathway is a regulatory mechanism involved in the quality control of translation. RBPs within this pathway, like UPF1, detect premature stop codons and recruit the degradation machinery before the ribosome can translate a faulty, truncated protein [77]. Finally, RBPs play a critical role in the innate immune response. Select RBPs are capable of detecting foreign RNA sequences (i.e. vRNA) and inhibiting their replication and translation. These aspects of innate immunity will be explored in more detail later on in connection with virus-host interactions (section 1.3).

RNA can regulate RBP function RBP-RNA interactions are not unidirectional. Recent studies have indicated that RNA can regulate protein function, a process known as "riboregulation". In riboregulation, RNA can recruit specific proteins, modulate their interaction and activity, and even alter their localisation and condensation properties, effectively acting as a regulatory molecule that influences protein dynamics within the cell. For instance, vault RNA1-1 (vtRNA1-1) has been shown to modulate autophagy by directly binding to the autophagy receptor protein p62. Through this interaction, vtRNA1-1 affects p62's oligomerisation, which in turn controls autophagy initiation and progression in cells [78]. Another compelling example involves SHMT1, a metabolic enzyme critical in one-carbon metabolism. RNA binding selectively inhibits SHMT1's ability to catalyze the conversion of serine to glycine by inducing a conformational shift in the enzyme's structure [79]. Alternatively, RNA can recruit and act as a scaffold in membraneless organelles. NEAT1 is a long noncoding RNA that drives the formation of paraspeckles through the recruitment of core paraspeckle proteins, including SFPQ, NONO, and PSPC1 [80]. The formation of paraspeckles sequesters proteins and affects downstream gene regulation. These examples showcase how RBPs can be altered by RNA binding, affecting structural and enzymatic functions as well as their cellular localisation.

1.2.2 The expanding repertoire of RBPs and their RNA binding specificity

The conventional understanding of RBP describes their binding affinity to RNA as dependent on sequence and/or structural motifs in RNA via a limited repertoire of defined RNA-binding domain (RBD) characteristics [81], such as a DEAD box helicase domain [82], RNA recognition motif (RRM) [83] or a K-homology (KH) domain [84]. However, recent advances in biochemical, structural and cellular methodologies and technologies have expanded our knowledge of Protein-RNA interactions that do not abide by our conventional RBP understanding [85].

Technical advances uncover novel RBPs Hundreds of novel RBPs were discovered through *in vitro* methods using immobilised RNA probes followed by proteoarrays or mass spectrometry [86, 87]. Although effective in identifying new RBPs, these methods held major caveats. The absence of a cellular context prevented physiological RNA binding dynamics from being observed. To address this, RNA interactome capture (RIC) was developed as an *in vivo* technique that focuses on native protein-RNA interactions [88, 89]. In RIC, proteins are covalently crosslinked to RNA in live cells by ultraviolet (UV) irradiation. UV irradiation only cross-links closely associated proteins and nucleic acids (virtually zero distances; $\leq$ 2 Å). Polyadenylated RNA and its associated protein are then captured using oligo(dT) beads. Denaturing washes then remove any non-covalent interactions, leading to the protein eluates being analysed by liquid chromatography-mass spectrometry (LC-MS/MS). With the publication of the first RIC datasets in 2012, the number of known RBPs increased dramatically: with 860 classified RBPs in HeLa cells [88] and 791 in HEK293 cells [89], of which 543 were shared across datasets. This number has continued to increase with the expansion of this and related methods to diverse species and cellular contexts.

The limitation of poly(A) RNA specificity has led to some alternative techniques being developed, allowing for the capture of bulk RNP. These techniques include RNA labelling paired with click chemistry [90, 91], solid phase purification [92], and organic phase separation-based methods [93]. These methodologies have since been applied across different organisms, expanding the RBPome dataset further. Hentze et al (2018) compiled all published RNA interactomes into RBP supersets for *Homo sapiens* (1,914 RBPs in total), *Mus musculus* (1,393), *Saccharomyces cerevisiae* (1,273), *Drosophila melanogaster* (777), *Arabidopsis thaliana* (719) and *Caenorhabditis elegans* (593) [85].

RNA binding specificity Many of the newly identified RBPs lacked known RBDs. Amongst the original RIC dataset by Castello and colleagues, a third of the identified RBPs did not contain conventional RBDs [88]. Orthogonal methods were employed to validate the discovery of these RBPs, which in turn broadened our understanding of how RBDs interact with RNA. These orthogonal methods included immunoprecipitation of GFP–RBP fusion proteins and detection of co-isolated poly(A) RNA with fluorescent oligo(dT) probes [88, 94]. Furthermore, RBDmap and RBS-ID methodologies were developed to define RNA-binding interfaces on a proteome-wide scale. RBDmap builds on RIC but employs two successive rounds of oligo(dT)

capture, interspersed with protease digestion steps to specifically identify the tryptic peptides that are crosslinked to RNA [95]. RBS-ID uses a combination of crosslink enrichment and extensive digestion of crosslinked RNA fragments using hydrofluoride to fully cleave RNA into mono-nucleosides, which can then be detected as variable modifications in mass spectra to reach single amino acid resolution at the RBP site [96].

To complement the advances in our understanding of the RBPome, an equal number of high-throughput advances have been made to understand how RBDs give rise to RNA binding specificity. Classical RBDs often recognise short RNA sequence motifs defined by conserved amino acids within their structure. However, unorthodox RBDs may behave differently. A combination of *in vitro* and *in vivo* methods has been established to investigate the RBP-RNA interface.

In vitro methods rely on measuring the relative affinity of purified RBDs to a pool of RNA oligos. Although successfully employed in SELEX (systematic evolution of ligands by exponential enrichment) [97] and RNACompete [98] amongst others, they lack cellular context. The absence of cellular context may lead to missing key cellular functions in proteins that bind two different sequence regions, NOVA, for example, [67]. Furthermore, RNA secondary structures are not accounted for, and RBPs involved in dsRNA binding, such as ADAR [99], would not be captured.

Advances in RNA sequencing allow for the *in vivo* study of RBP-RNA-bound regions. To this effect, crosslinking and immunoprecipitation (CLIP)-based sequencing methods are commonly employed [100,101]. UV-crosslinked RNA and associated protein are purified for specific RBP-RNA complexes, followed by 5' radioactive labelling of RNA by T4 polynucleotide kinase (PNK), reverse transcription PCR or sequencing. *In vivo* techniques account for many pitfalls explored in *in vitro* experiments; however, they, too, have their caveats. These include the requirement of high amounts of starting material, crosslink and precipitation efficiency of the RBP, with the potential loss of low-level signal [102].

Individually, *in vitro* and *in vivo* techniques have their respective issues; however, combining methods can provide a more accurate understanding. An exhaustive study to expand on RBP RNA preferences and systematically map different aspects of RBP activity was published in 2020. Researchers employed eCLIP with four orthogonal approaches, including *in vitro* evaluation of RNA affinity for the same RBPs, chromatin association by ChIP–seq, functional assessment of transcriptome changes by RBP depletion and RNA-seq and subcellular

localisation using immunofluorescence [103]. Nostrand et al. (2020) produced 1223 replicated data sets for 356 RBPs; of these, 150 RBPs indicated a strong correlation across the different experimental techniques.

Binding prediction using machine learning Computational tools are on the rise to keep up with the quantity of experimental output and allow for automation in analysis pipelines. For example, a tool called PRIESSTESS (Predictive RBP-RNA InterpretablE Sequence-Structure moTif regrESSion) is a computational method that captures sequence and structure specificity from *in vitro* RBP-RNA binding data, specifically addressing the diversity of RBP binding [104]. It automates data analysis and produces readily interpretable models, with all available motif predictions. Furthermore, a new focus of research has been to generate machine learning models for binding prediction algorithms trained on CLIP and RBP datasets [105]. One such tool is ASCRB. Developed by Li et al. (2023), ASCRB was developed to predict RNA binding sites using five feature coding schemes trained on circRNA. CircRNA is a non-coding RNA with a specific circular structure, which plays a key role in various life activities by interacting with RNA-binding proteins through its binding sites [106].

A host of factors define the specificity of individual RBPs, from sequence motifs, sequence context, and secondary structures to protein-protein dynamics. No singular experimental technique covers all potential outcomes, so orthogonal methods must be employed in tandem. The advancement of computational tools, however, holds the potential of gathering all factors in a singular location, enabling more accurate modelling.

1.2.3 Changing cellular landscape, changing RBPome

The binding of RBPs to RNA is highly dynamic, with RNA interactomes being context-dependent and responsive to various stimuli. One of the earliest studies investigating the changing RBPome was conducted in *Drosophila melanogaster* during the maternal-to-zygotic transition (MZT) [107]. Comparative RNA RIC on samples from early and late embryos revealed significant changes in the RNA interactome during development. Parallel whole-proteome analysis determined whether these changes were attributable to alterations in protein abundance. This enabled the classification of identified RBPs into three groups: RBPs with unaltered binding, RBPs whose RIC abundance correlated with proteome-level changes, and RBPs that exhibited altered RIC levels independently of protein abundance. In total, 116 RBPs

were classified as dynamic, belonging to the last category [107].

A similar study performed in zebrafish a year later identified 24 and 53 RBPs as significantly changing during the same developmental transition [108]. Among these, Hnrnpa1 was further analysed using iCLIP. The study revealed that Hnrnpa1 shifts its RNA-binding preferences during zebrafish MZT, transitioning from binding the 3' untranslated regions (3' UTRs) of maternal mRNAs in the cytosol to interacting with nuclear noncoding RNAs from Chromosome 4, including *pri-mir-430*. These two studies in MZT transition exemplify the power of combining RBP and RNA CLIP methods to understand dynamic cellular environments. They highlight the versatile behaviour of RBPs and their potential roles in regulating cellular changes, particularly during critical developmental processes.

Viruses drastically change cellular environments, causing full re-arrangement of cellular functions. Virus's parasitic nature requires them to uptake cellular factors to add to their limited repertoire of virus-encoded proteins, such as RBPs. RBPs have been identified as being involved in almost every step of the viral life cycle, including genome replication, viral protein synthesis, and assembly of virus progeny [109, 110]. RBPs can also restrict viral progression as the vRNA is a target of the antiviral innate immune response, and specialised proteins can detect unusual molecular signatures [111, 112]. Understanding how the dynamic RBPome is modulated during viral infection is paramount in understanding the viral lifecycle and subsequent development of targeted therapeutics.

1.3 Global analysis of host-virus interactions in alphavirus infection

1.3.1 Protein-wide compositional analysis of viral ribonucleoproteins (RNPs)

vRNAs, central to viral infection, undergo many of the same processes as cellular RNAs, including translation, localisation, and decay, but they are also involved in virus-specific functions such as replication and packaging. To mediate and regulate each of these stages, vRNA assembles with viral and cellular RNA-binding proteins (RBPs) to form viral ribonucleoproteins (vRNPs) [113, 114]. Comparative RIC studies allow for the capture of full RBPome changes in this changing environment. Although the relative proportion of mRNA in this context is vRNA during advanced stages of infection, a subset is also cellular. This "contaminant", albeit informative, obscures the capture of the vRNPs. Building on RBPome advancements, various complementary methodologies have been pioneered over the last few

years to specifically study vRNPs [115].

Comparative proteomics The emergence of novel proteomics and orthogonal methods described previously can be applied to the changing RBPome that occurs in viral infection. Indeed, to date, three comparative RIC studies have been carried out in SINV [113], SARS-CoV2 [116], and Influenza-A virus (IAV) [117]. During advanced infection, it has been reported that vRNA can represent 70% (in SINV [113]), 20-80% (in SARS-CoV2 [116, 118]) and 50% (in IAV) [119] of total RNA present in the cell. Therefore, the capture of total mRNAs and their interacting proteins will also capture vRNPs and can aid in identifying RBPs potentially hijacked during viral infection.

In the RBPome of SINV-infected cells, a significant change in captured RBP occurs as the infection progresses. Garcia-Moreno et al. (2019) investigated different time points of infection and noted that the most remarkable changes occurred at later stages of infection [113]. A quarter of the RBPome changes at 18h post-infection (hpi) which also corresponds to vRNA representing 70% of the transcriptome. The significant changes in RBP binding can be largely attributed by changes in the transcriptome and RNA availability rather than protein abundance (total proteome). Crucially, RBPs further tested in this study were observed to relocate to viral factories and potentially have functional changes. Two proteins were highlighted in this study as modulators of SINV viral fitness. The transcriptome remodelling during infection was largely attributed to the 5' to 3' RNA degradation machinery. The exonuclease XRN1 and its interactor, PATL1, are stimulated at 18hpi. The subsequent knockout of XRN1 inhibited viral infection completely. Contrastingly, Gem-associated protein 5 (GEMIN5), an RBP that catalyses the formation of the spliceosome and binds the RNA cap, was also stimulated at 18hpi [113]. The overexpression of GEMIN5 caused a delay in viral subgenomic gene expression and inhibited capsid translation. ICLIP of GEMIN5 indicated a switch of RNA binding preferences from cellular to vRNA. While the protein binds to the 3' UTR of some host mRNAs, it was found to interact with the 5' ends of SINV. GEMIN5 affects viral protein expression through this interaction at the 5' end. This study can be used as a springboard for new research avenues. The newly associated dynamic RBPs can be explored further to characterise their functional role and mechanism of interaction. However, not all RBPs identified may correspond to direct viral modulation. Some proteins, although stimulated by SINV, did not relocate to viral factories, such as NGDN, HNRNPA1 and the mitochondrial translation elongation factor TUFM,

suggesting that they don't interact directly with vRNA.

Direct vRNA interactors To explore specific vRNA interactors, novel methods were developed to capture vRNPs. The high-throughput capture of vRNPs follows a similar workflow across different methodologies: infection, protein–RNA crosslinking, specific vRNA isolation, and proteomic analysis. Key differences in each of these steps have created a host of different variations of protocols, and an equal number of variations in results [115]. The stringency of the proteomic captures relies heavily on the strict vRNA capture. A common tool is the use of 4-thiouridine (4SU) RNA labelling, a nucleotide analogue that is taken up by mammalian cells and is incorporated into nascent RNA when added to culture media [120]. 4SU can be used to label *de novo* synthesised vRNAs when added at specific infection time points. Host transcriptional shut-off at later time points signifies that newly synthesised RNA corresponds to vRNA. Crosslinking of 4SU is performed at a higher wavelength (365nm), excluding natural unlabelled RNA-protein crosslinking from occurring. Subsequent 4SU biotinylation and streptavidin purification, capture specific 4SU/RNA-protein. Alternatively, crosslinking is followed by the isolation of RNA using single probes or tailing anti-sense probe sets. These oligos may contain locked nucleic acids (LNAs) for improved double-strand RNA (dsRNA) invasion.

Three proteome-wide approaches have been recently used to elucidate the composition of alphavirus RNPs. These include viral crosslinking and solid-phase purification (VIR-CLASP), crosslink-assisted messenger RNP purification (CLAMP), and vRNA interactome capture (vRIC). Kim et al (2020) developed a method to capture interactions between incoming genomic vRNA and cellular proteins, revealing hundreds of early host-virus interactions in CHIKV [114]. They employed VIR-CLASP, a method that relies on the infection of unlabeled host cells with 4SU-labelled viral genomes. The subsequent UV irradiation crosslinking and solid-phase purification enable the sole capture of incoming viral genomes containing 4SU. Amongst the detected RBPs, they uncovered a previously unreported viral RBP, the fatty acid synthase (FASN), an enzyme that generates palmitic acid, interacting directly with CHIKV vRNA. FASN was further characterised as regulating vRNA via its enzymatic activity [114]. Viral replication sites are rich in palmitic acid, which could justify the uptake of this cellular protein at specific stages of the virus lifecycle. Overall, this method uncovered hundreds of cellular RBPs functionally important in the initial steps of CHIKV infection. However, the later

events of viral replication or interactions between vRNA and host proteins are not captured through this method.

To overcome the limitation of VIR-CLASP, CLAMP focuses on capturing vRNPs at later time points during infection [121]. This technique treats cells with actinomycin to halt cellular transcription before 4SU addition. Crosslinking is performed using formaldehyde, and the vRNPs complexes are purified via the capture of biotinylated sulfhydryl groups in 4SU, by HPDP-biotin conjugates and streptavidin precipitation. Initially used to study the vRNPs in SINV [121], it was later employed as a comparative analysis in three alphaviruses (CHIKV, SINV and VEEV) [122]. The comparative analysis was able to identify 108 conserved RBPs across the three viruses. HnRNP K protein was identified in both studies and is common across the three viruses. The protein was evaluated using CLIP-seq and was identified as interacting with distinct sites on the sgRNA [121]. The disruption of this binding site decreased viral titer in mammalian cells and intriguingly increased structural protein expression. Both studies validated hnRNP K beneficial role for Old-World alphavirus infection. Although informative, CLAMP datasets had a very low incidence of *bona fide* RBPs, likely due to the promiscuous nature of formaldehyde crosslinking and/or limited specificity in the purification of vRNA [115].

To complement the gap in capturing post-replicative alphavirus RNPs, a third method was developed: vRIC. vRIC was originally applied to SARS-CoV-2 [116] and later applied to SINV [123]. This method employs the use of 4SU-labeled vRNA followed by oligo(dT) capture. In SINV, vRIC captured 400 cellular RBPs. These vRNPs were characterised against cellular RNPs to detect notable differences. Enrichment of post-translational modification (PTM) enzymes, such as kinases, was detected in the vRNP fractions. Meanwhile, a net difference in translational initiation factors in the two groups was observed, supporting the previously reported non-canonical cap-dependent translation mechanisms of vRNA in SINV [124–126].

The global capture of dynamic RBPs captured both in comparative RIC and in vRNA-specific RIC methods expands the repertoire of known RBPs involved in viral infection. The identified RBPs can subsequently be further characterised to understand their individual mechanism of viral modulation.

1.3.2 Protein-Protein interaction analysis

The study of the RBPome in viral infection and, subsequently, vRNPs does not fully account for protein-protein (P-P) interaction dynamics at play. Prior to RNA capture methods, the study

of viral-to-cellular protein interaction dynamics was fundamental in understanding how viruses modulate their environment. The most common technique involved the insertion of a reporter protein or an epitope tag into the target protein, which enabled specific capture of tagged proteins and their interacting partners for mass spectrometry analysis. Two studies inserted a GFP-tag into nsP3 of SINV and captured its interactors at different time points of infection [127, 128]. 10 common proteins were identified in both studies. However, technical differences in controls used in parallel meant that a further 20-25 proteins were identified only in one or the other study. The most notable proteins were G3BP1/2 and other nsPs. The comparative study of nsP3 interactors at different times of infection indicated a specific early and persistent recruitment of G3BP and a later recruitment of 14-3-3 proteins [128]. Almost two decades later, another nsP3 co-precipitation was carried out employing an intercalated mScarlet tag [123]. This latter study identified a staggering 378 protein interactors, with G3PB1/2 as one of the most highly enriched proteins alongside the other nsPs. The significant increase in the number of detected enriched proteins reflects the advancement of technologies capable of detecting lower abundance interactors. The interaction with nsPs was consistently detected in parallel studies, which similarly isolated nsP2 [129] and nsP4 [130] . The crossover of protein interactants amongst the nsPs indicated they may be integral parts of the replicase complex.

Varjak et al. (2013) endeavoured to specifically study the P-P interactants of the replicase complex in SFV using an alternative method. Functional intact replicase complexes were captured using dextran-covered magnetic nanoparticles, which later aided in magnetically isolating the nanoparticle-containing lysosomes [131]. This method identified 78 cellular proteins, many of which were previously identified in the above studies, as well as novel proteins. Interestingly, a third of the proteins were characterised as RBPs. Comparatively, the most recent nsP3 enrichment identified a quarter of interactors as binding SINV vRNA when cross-referencing the proteins to the SINV vRIC dataset [123].

P-P interactants of structural proteins have also been studied, although to a lesser extent. Many of their associated cellular interactants are associated specifically with their cellular location. The two spike proteins, E2 and E1, interact with cell surface receptors and components of the actin cytoskeleton to facilitate viral entry and egress. E2 has been shown to interact with Dendritic Cell-Specific Intercellular adhesion molecule-3-grabbing Non-integrin (DC-SIGN) [132] and Heparin sulfate [133], among others. The E3 glycoprotein contains the signal peptide, which interacts with the endoplasmic reticulum (ER) membranes, directing the

structural polyprotein to the lumen. Here, it is cleaved by host cellular proteases, including furin and signalase, to render E3, E2, 6K and E1 proteins [134]. 6K is a transmembrane viral protein that acts as a form of viroporin. Precipitation of biotin-labelled proteins on immobilised streptavidin-agarose suggested that 6K is associated with glycoproteins at the cell surface, instrumental in its role in virion budding from the cell [135]. The capsid protein interacts with a host of cellular factors. In VEEV, for example, capsid protein forms a tetrameric complex with CRM1 and importin α/β that obstructs nuclear pore complex function [136].

1.3.3 Alphaviral modulation of the cellular environment to favour viral infection

The capture of viral-host P-P interactions and vRNPs reveals the sophisticated strategies viruses use to hijack host cellular machinery. Numerous host proteins have been identified that either facilitate or inhibit viral replication, underscoring the dual role of host factors in viral fitness. Alphaviruses, in particular, are adept at manipulating their environment to favour infection. This section delves into how alphaviral vRNA and viral proteins coordinate with host components to orchestrate a productive infection. The particular role of helicases will be described in section 1.3.4.

Host cellular shut-off Alphaviruses have developed a mechanism to create a favourable environment for infection, starting with the host cellular transcription and translation shut-off [137]. This inhibition would interfere with the innate immune system and, subsequently, the antiviral response. It further serves the virus by repurposing protein-synthesising machinery to translate sgRNA, increasing its own viral output. The transcriptional shut-off is managed by the entry of nsP2 into the nucleus in Old World alphaviruses [138–140]. It subsequently targets the RNA polymerase II subunit RPB1 for degradation, which results in general host cell transcriptional shut-off and subsequent cytopathic effects in mammalian cells [139]. In New World alphaviruses, the mechanism of transcriptional inhibition is led by the capsid protein. Capsid in VEEV has been reported to form a complex with nuclear import and export factors which obstruct the nuclear pore and, consequently, nuclear trafficking [136].

The abrogation of cellular protein synthesis is orchestrated by a combination of the initiation factor eIF2α phosphorylation [126, 141], competition of viral mRNA for translation machinery [45, 142] and the modification of the cytoplasmic ionic environment [143, 144]. Protein kinase R (PKR) senses dsRNA, an intermediary state during virus replication, and phosphorylates

eIF2α. This action renders eIF2 unable to be recycled back into its active GTP-bound state, resulting in a general translational shut-off [145]. Translational shut-off via dsRNA recognition by PKR can effectively block viral replication [126]. However, the translation of alphaviral structural proteins from their subgenomic messenger is unaffected by the phosphorylation of eIF2α. A stable RNA hairpin loop structure in the 26S promoter of the subgenomic mRNA from SINV and SFV stalls the ribosome on the correct AUG, providing resistance to eIF2α phosphorylation and thereby enhancing translation of the viral subgenomic mRNA [126, 146].

In alphaviruses, translation occurs in a non-canonical manner. Most cellular mRNAs contain a blocked cap structure at their 5' end and are translated by the canonical cap-dependent scanning mechanism. This involves recognition of the cap by eIF4F, followed by the interaction of the preinitiation 43S complex with the mRNA. The eIF4F complex comprises the cap-binding factor eIF4E, the helicase and ATPase enzyme eIF4A, and the scaffolding protein eIF4G [70]. However, SINV sgRNA is translated without the participation of crucial eIFs such as eIF2 or eIF4A [125]. SINV translation has been found to be resistant to eIF4G cleavage, which normally disrupts cap-dependent translation [124]. The cleavage of eIF4G varies depending on the context. In apoptosis, it is part of the cellular process to shut down protein synthesis during programmed cell death [147]. In viruses with uncapped mRNAs, like picornaviruses, cleaving eIF4G allows them to hijack the host translation machinery and preferentially translate their own RNAs [148]. In alphavirus infections, while direct cleavage doesn't occur, the disruption of eIF4F complex function may still contribute to host translation shutdown and promote viral protein synthesis [143]. These mechanisms highlight how viruses have evolved different strategies to manipulate host translation machinery.

Stress granule (SG) manipulation Biomolecular condensates are prevalent in cells and critical for various cellular functions, including RNA metabolism, embryonic cell fate specification, and neuronal activity [149–151]. These condensates are found throughout eukaryotic cells, including in the nucleus, cytoplasm, and on membranes. Stress granules (SGs), one of the best-characterised biomolecular condensates, are RNA–protein assemblies formed in response to a variety of environmental cues [152]. During viral infection, SG formation and disassembly are tightly regulated by the cellular translation status [126]. In the early phase of many viral infections, the activation of the PKR pathway by the double-stranded vRNA also activates the formation of SGs enriched with translation initiation factors

such as eIF3b. However, in later infection stages, many viruses instead suppress SG formation or disassemble SGs altogether and utilise the stored proteins [153]. The prevalent protein identified across all SINV nsP3 P-P were the G3BP proteins. NsP3 in alphaviruses has been shown to suppress the formation of stress granules by manipulating G3BP1. nsP3 has a conserved N-terminal macrodomain that hydrolyses ADP-ribose from ADP-ribosylated proteins and a C-terminal hypervariable domain that binds the essential SG component G3BP1 [154]. The importance of the nsP3–G3BP interaction became apparent in a deletion mutagenesis study in SFV, where the binding domain of G3BP1 was deleted, and viral fitness was subsequently reduced [155]. G3BP proteins were further characterised as pro-viral factors in CHIKV [156, 157]. The depletion of the proteins directly correlated with a reduction in viral protein expression and progeny viral titer. The sensitivity in CHIKV infection to depletion of G3BP was due to an Arginine residue at the P4 position of the cleavage site between the nsP1 and nsP2 [157]. This particular residue is not present across all alphaviruses, and in the case of SINV, it was observed to be partially resistant to G3BP deletion.

Innate immunity vRNA is the target of the antiviral innate immune response because it typically contains unusual molecular signatures that specialised RBPs can recognise. Pattern recognition receptors (PRRs) are specialised proteins that detect viral elements. These pathogen-associated molecular patterns include triphosphate ends, unmethylated caps, sequence biases, and long dsRNA tracts produced during viral replication [111, 112]. PRRs initiate a cascade of innate antiviral responses, amongst them, the IFN response is the most recognised. However, many of the innate immune responses have co-evolved with viruses. This signifies a dynamic co-evolution between host and virus, where the host develops mechanisms to suppress viral replication and progression, while viruses evolve evasion strategies to counteract these defences, ensuring their survival and replication within the host [158].

IFNs activate neighbouring cells via transmembrane receptors, which cascade down to the nucleus, resulting in the upregulation of IFN-stimulated genes (ISG) with antiviral activity. One of the IFN targets is viral translation. PKR is a well-recognised ISG induced by IFN-I. As previously described, it induces the phosphorylation of eIF2α, a translation inhibitor. In alphaviruses, this is overcome through the use of non-canonical translation mechanisms, circumventing the requirement of eIF2 [125].

IFN further induces proteins with tetratricopeptide repeats (IFITs) as antiviral factors that interact with alphaviruses. They actively block viral translation by binding to specific regions of the vRNA [112]. In most viruses, IFIT1 binds to unmethylated cap structures of vRNA, which prevents the binding of translation initiation factors. IFIT1 recognizes specifically vRNA lacking 2'-O-methylation of the 5' cap, a modification that is common in cellular RNAs but not in many viral RNAs. However, alphaviruses, which have a 5' cap lacking 2'-O-methylation, have evolved to evade IFIT1 restriction by encoding stable secondary structures (like stem-loops) within their 5'UTR. Mutations within the 5'UTR that disrupt these RNA structural elements enable the antiviral restriction by IFIT1, demonstrating the importance of these structures in IFIT1 evasion [159].

The zinc finger antiviral protein (ZAP or ZC3HAV1) is an ISGs involved in inhibiting alphaviruses. It has been shown to bind to vRNA, restricting replication and translation through inducing RNA degradation. The vRNA specificity is achieved through recognition of CpG dinucleotides, which are underrepresented in mammalian transcriptomes, suggesting that ZAP has evolved to exploit this feature for distinguishing self from non-self RNA [160]. Another well-characterized ISG, the 2'-5'-oligoadenylate synthetase (OAS), activates RNase L upon sensing viral dsRNA, leading to widespread RNA degradation, including that of vRNA. Alphaviruses form replication organelles, restricting access of host factors to detect the formation of dsRNA during viral replication, thus reducing the efficiency of OAS-associated pathways at later stages of infection [161].

Although viruses can sometimes circumvent individual ISGs, a recent study on VEEV demonstrated the impactful combinatorial role of multiple key ISGs in limiting viral progression [162]. This study identified ZAP, IFIT3, and IFIT1 as dominant effectors that restrict VEEV, while comprising only $< 0.5\%$ of the total ISGs. Moreover, pretreatment with IFNs and the subsequent expression of ISGs significantly reduce viral fitness, emphasising their critical antiviral function. A more nuanced understanding of the IFN response in innate immunity has revealed that its effects extend beyond stimulating ISG transcription. Recent studies have shown that IFN-I can induce changes in transcript processing [163], translation control [164], protein-protein interactions [165], and post-translational modifications outside the JAK/STAT signalling cascade [166]. These diverse regulatory mechanisms not only shape ISG activity but can also enhance the antiviral properties of non-ISGs. This multifaceted regulation provides a potential explanation for how innate immunity remains active even during viral-induced cellular

shutoff. However, the detailed specificities of IFN regulation are beyond the scope of this introduction.

These examples illustrate a range of host strategies to inhibit viral replication via recognition of RNA structural or sequence motifs. The role of RNA-binding restriction factors such as ZAP is particularly relevant to the current study, which aims to understand host proteins with RNA-binding capacity that may act as antiviral restriction factors. Given the evolutionary pressure viruses face to avoid or counteract such host defenses, identifying additional RNA-targeting mechanisms could shed light on both viral evasion strategies and underexplored host restriction pathways.

1.3.4 Alphaviral RNA interactome – Helicases as key RBPs

In the study of alphavirus-host interactions, several RBPs have been identified as critical players in mediating host responses to infection [61, 167]. These RBPs, which can modulate viral replication and host immune defences, often interact directly with vRNA, enabling the host cell to sense and respond infection. Among these RBPs, helicases have emerged as a particularly important class due to their ability to bind vRNA and potentially modulate vRNA structures, thus influencing viral replication and host antiviral signalling [168, 169].

Helicases are a family of enzymes known for their role in unwinding RNA or DNA duplexes, an activity crucial for numerous cellular processes, including transcription, translation, and RNA metabolism [170]. DEAD/H box proteins (DDX) form the largest helicase family, with 41 members in humans, and are characterised by the presence of an Asp-Glu-Ala-Asp/His (DEAD/H) motif. DDX proteins have essential physiological roles in cellular RNA metabolism [82]. In the context of alphavirus infection, several DDX proteins have been identified as significant players within the vRNA interactome. These helicases not only bind vRNA but often recognise conserved RNA structures, enabling them to act as regulators of the viral life cycle and modulators of the host antiviral response [169]. Their functional variability arises from the diverse cellular roles that helicases fulfil and the unique strategies alphaviruses employ to exploit, repurpose, or evade cellular machinery for their own benefit [171]. In the SINV vRIC dataset, 14 DDX proteins are enriched in the vRNPs, suggesting their integral role in vRNA regulation [123].

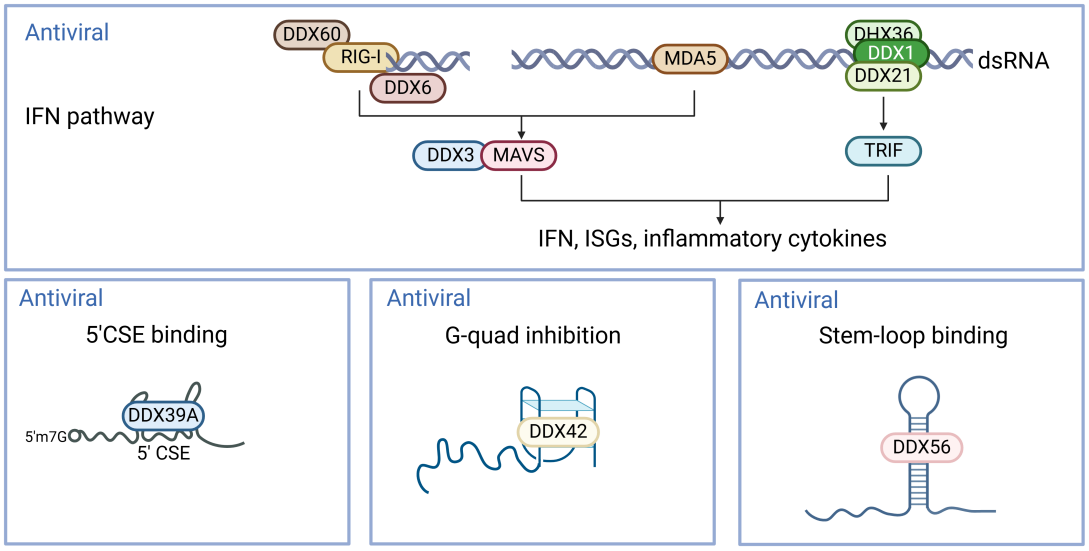

Figure 1.3: **Antiviral role of DEAD/H helicases proteins in alphavirus infection schematic** Examples of antiviral roles of DDX proteins in alphavirus infection. IFN pathway helicases include RIG-I, MDA5 that recognise and bind dsRNA. The former is supported by DDX60 and DDX6. Combined, they trigger MAVS coordinated with DDX3, which triggers a IFN cascade. Alternatively, dsRNA is detected by the DDX1/DDX21/DHX36 complex and triggers TRIF. Among the IFN-independent antiviral DDX proteins, DDX39A binds 5'CSE, DDX42 binds conserved G-quadruplex structures, and DDX56 binds viral stem-loops.

Antiviral factors Certain DDX proteins actively contribute to the host's antiviral defences by detecting and responding to foreign vRNA. They can coordinate their antiviral impact by working with innate immunity sensors that stimulate interferon-mediated and other inflammatory responses, inhibiting viral replication and marking infected cells for immune clearance [172]. RIG-I, also known as DDX58, and MDA5 are DDX helicases widely known for their role in the IFN-I pathway [173]. Both proteins detect dsRNA in the cytoplasm and trigger an inflammatory response. Both RIG-I and MDA5 are equally important in the initial response to alphavirus infection. The concentration of these receptors at the time of infection determines the rate, time, and scale of type I IFN induction [174]. While RIG-I and MDA5 have many similarities, they also have different ways of recognising pathogens and host species. RIG-I recognises short RNA ligands with 5'-triphosphate caps, while MDA5 recognises longer genomic RNA and replication intermediates [175]. Other DDX proteins can also act as additional viral sensors or regulators of the RIG-I-associated IFN activation. DDX6 and DDX60 interact directly with RIG-I and enhance its signalling or dsRNA binding, respectively [176,177].

An alternative interferon mediated response has been observed through the TRIF adaptor. A complex of three helicases, DDX1-DDX21-DHX36, has been reported as a dsRNA sensor

that interacts and stimulates the TRIF adapter and subsequent type I IFN signalling [178]. Although this complex has not been reported as a regulator in alphaviruses, TRIF has been identified as a key regulator of RRV, and in its absence, viral production increases [179]. DDX1 and DDX21 have both been identified in the SINV vRIC, indicating potential interesting interaction which requires further research.

Some DDX proteins have been identified as antiviral factors independently of inteferon pathway. In CHIKV, DDX39A was observed to re-localise to the cytoplasm and inhibit viral replication [180]. Upon further investigation, DDX39A was identified as binding CHIKV vRNA and interacting with the 5' conserved sequence element (5'CSE). The most conserved structural RNA element across the alphavirus genus is the 5'CSE, which is important for the replication of CHIKV, SINV, and VEEV. The binding of this conserved region coupled with the antiviral effect suggests DDX39A hinders the recognition and binding of the viral replicase. Similarly to DDX39A, DDX42 and DDX56 have been identified as antiviral factors independent of IFN signalling, driven by their binding specificity. DDX42 was previously identified as binding specifically to G-quadruplex motifs [181]. In parallel, the viruses DDX42 inhibited contained G4 structures within their genome, such as CHIKV, which strongly suggests a functional link [182]. Meanwhile, DDX56 antiviral activity is associated with its binding of a stem-loop encoded in the CHIKV genome, causing vRNA destabilisation and affecting replication [183].

Pro-viral factors DDX proteins are often linked to antiviral defences. However, they can paradoxically support the viral lifecycle in some instances. Viruses have evolved mechanisms to exploit helicases such as DDX3 and DDX5. DDX3 is a multi-faceted helicase involved in transcription and translation and regulates cellular processes like cell cycle progression, apoptosis and innate immunity [184]. In VEEV, DDX3 was identified with DDX1 to interact with the nsP3 protein [185]. The knockdown of both DDX1 and DDX3 proteins resulted in a decrease in infectious viral titers. The nsP3-DDX3 interaction was further characterised as being crucial for viral translation initiation through its association with translation machinery. DDX5 has recently been characterised as a pro-viral protein in SINV infection [186]. Depletion of the protein negatively impacted the viral replication cycle, while its overexpression had a pro-viral effect. DDX5 as one of the factors associated with the SINV replication complex by dsRNA-IP coupled to mass spectrometry [187]. Furthermore, the DDX5 co-factor DDX17, was also shown to display a pro-viral phenotype. The two proteins were identified as close

interactors with nsP2, which modulates viral-induced host-transcriptional shut-off [186]. DDX5 is involved in host transcriptomic regulation [188], and thus, we can speculate that the nsP2-DDX5 association is linked with restricting cellular transcription during infection.

Altogether, helicases are capable of not only detecting vRNA but also playing active roles in blocking viral replication directly, either by destabilising vRNA structures or promoting degradation. However, viruses have developed methods to recruit these proteins to benefit their own lifecycle. Overall, helicases are extremely diverse and adaptable proteins that play a central role in alphavirus regulation.

1.4 Discovery of DDX1 as a host regulator of virus infection

An interesting protein that deserves attention is the DDX1 helicase. It has multiple functions within host cells and, interestingly, has recently been identified in viral interactomes. The regulatory role and mechanism behind its modulation of alphavirus is yet unexplored. The RNA helicase has, however, been identified to play diverse roles in viral infections, interacting with various viral proteins and affecting viral replication and host immune responses.

1.4.1 Proteome-wide approaches reveal a functional link between DDX1 and RNA viruses

In the comparative RIC of SINV, DDX1 appears as a highly enriched RBP at the late stages of infection [113]. The increased binding activity correlates with increased vRNA in the cell. Furthermore, DDX1 co-localises with viral factories during infection suggesting a close regulatory role in viral infection. In SINV vRIC, DDX1 appears as a highly enriched protein that interacts directly with vRNA [123]. Beyond SINV, DDX1 has been captured to interact directly with a variety of viruses from different viral families. In the comparative analysis performed by Iselin et al. (2022), where different viral interactome capture methods were benchmarked against one another, DDX1 appeared in every study [115]. This suggests that DDX1 is a direct interactor with the vRNA of SINV, CHIKV, ZIKV, DENV and SARS-CoV-2.

Functional studies have identified DDX1 in regulating various virus infection outcomes. Depending on the virus studied, DDX1 has been identified to either have a pro- or anti-viral function. DDX1 was first identified as a viral regulator in HIV-I infection. The helicase was identified as a key cellular co-factor of Rev. Fang et al. (2004) found that DDX1 is required for

efficient Rev function and proper nuclear localisation of Rev in mammalian cells [189]. Small interfering RNA (siRNA) targeting DDX1 provided strong evidence that DDX1 is required for both Rev activity and HIV production from infected cells [190]. A later study evaluated the mechanism through which DDX1 modulates Rev-mediated replication. DDX1 was found to act through the Rev Response Element (RRE) RNA to accelerate the nucleation step of the Rev-RRE assembly process [191]. This interaction is essential for efficient vRNA export and subsequent virus production.

In VEEV, DDX1 was observed to interact directly with nsP3, and further characterisation indicated that the absence of DDX1 significantly inhibited viral titers [185]. Confocal microscopy revealed that DDX1 and VEEV nsP3 co-localise in infected cells, further supporting the interaction between these proteins during VEEV infection. While the exact mechanism is not fully elucidated, it is proposed that the nsP3-DDX1 complex may interact with the host translational machinery, which is essential for the viral life cycle.

In coronavirus infectious bronchitis virus (IBV) and SARS-CoV-1, DDX1 was observed to interact directly with nsP14 [192]. In both viruses, the viral protein is essential for efficient vRNA synthesis and may be involved in RNA proofreading. Manipulation of DDX1 expression, either by siRNA-induced knockdown or by overexpression of a mutant DDX1 protein, indicated that DDX1 promotes viral proliferation. The interaction with nsP14 suggests DDX1 is aiding in the coronavirus RNA replication stage of infection [192]. Furthermore, DDX1 has been observed as a regulator in SARS-CoV-2 infection. The protein was observed to relocate to viral factories when studied by confocal microscopy. The exact role is debatable, as differing phenotypes have been observed in the absence of the protein [116, 193]. However, this will be explored further in this thesis in relation to the technical application of DDX1 knockdowns.

Two other positive-strand RNA viruses have identified DDX1 as an inhibitory factor in their lifecycle. In contrast to its pro-viral role in coronaviruses, DDX1 interacts with Transmissible Gastroenteritis Virus (TGEV) nsP14 and induces the host's innate immune response [166]. It was observed through the knockdown of DDX1 by targeted siRNA, that nsp14-induced IFN-β production was significantly decreased. Similarly, in the Foot-and-Mouth Disease Virus (FMDV), the porcine DDX1 has also been characterised to stimulate IFN-β activation and ISG expression [194]. DDX1-dependent inhibition of FMDV replication relied, in part, on its ATPase/helicase activity, as observed when employing a DDX1 catalytic mutant. Helicases have a crucial regulatory role in innate immunity, and DDX1 is shown to be no different in the

context of these viruses.

A complex, including DDX1, was identified to regulate innate immunity and IFN signalling. Indeed, DDX1, DDX21, and DHX36 form a complex that binds viral dsRNAs and induces IFN signalling through the TRIF adapter [178]. This complex was then further identified as regulating Influenza A virus (IAV) infection. The absence of DDX1, in this instance, increased viral proliferation [195]. Moreover, in Newcastle Disease Virus (NDV), Avian Influenza Virus (AIV) and Vesicular Stomatitis Virus (VSV), chicken DDX1 (chDDX1) was observed to be significantly upregulated after infection [196]. The upregulation was in direct correlation with the expression of IFN-β, IFN-stimulated genes (ISGs), and proinflammatory cytokines. The knockdown of chDDX1 increased the viral yield of NDV and VSV while the overexpression of chDDX1 inhibited viral replication. *In vitro* precipitation of chDDX1 with poly(I:C) indicated a strong and direct interaction, suggesting that chDDX1 acts as an RNA PRR during IFN activation.

Overall, DDX1 has been reported to have a multitude of functions in regulating viral infections. Understanding the protein and RNA interactors can help elucidate its effect on the viral lifecycle.

1.4.2 DDX1 as part of the tRNA ligase complex (tRNA LC)

Within the comparative RIC and vRIC datasets generated in SINV and SARS-Cov-2, besides DDX1 explored above, two additional notable proteins were reported to increase their binding affinity: RTCB (HSPC117) and FAM98A [113, 116, 123]. Furthermore, immunofluorescence analysis revealed RTCB, like DDX1, re-localised in proximity to viral replication organelles of of SINV and SARS-Cov-2 infected cells. This indicated their potential modulation of vRNA and viral fitness. The presence of RTCB and FAM98A was intriguing as, together with DDX1, they have been reported as being part of an essential cellular complex named the tRNA ligase complex (tRNA-LC) [197].

The tRNA-LC is an established complex consisting of 4 principal proteins: DDX1, RTCB, FAM98A or FAM98B and CGI99 (RTRAF), and 2 transiently associated proteins, Ashwin (ASW) and Archease [197, 198]. Individually only DDX1 and CGI99 have been reported as involved in viral regulation, the latter as a modulator of IAV infection. CGI99 was identified in interacting with the PA subunit of the influenza polymerase and plays a significant role in transcription regulation by modulating RNA Polymerase II (RNAP II) activity, an essential

factor in mRNA synthesis [199]. The silencing of CGI99 significantly reduced RNAP II-mediated transcription. This suggests that CGI99 has a role in cellular transcription and is potentially hijacked by IAV viral proteins to repurpose CGI99 for its own vRNA transcription. Intriguingly, a later study by the same group identified CGI99 as incorporated into IAV virions [200]. Both human and avian influenza viruses of various subtypes increased CGI99 protein levels. Confocal microscopy identified CGI99 as colocalising and interacting with vRNPs in the nucleus and cytoplasm during infection. Furthermore, the tRNA-LC member was detected in purified vRNP purifications from virions as well as observed as colocalising inside virions by immunogold labelling and electron microscopy [200]. It is worth noting that the above researchers employed antibody-based approaches, which would not capture the presence of other tRNA-LC members.

The identification of three components of the tRNA-LC within the SINV vRNP screening suggests that the complex is involved altogether in the modulation of viral infection. The specificities of SINV regulation by DDX1 and the tRNA-LC will be explored in detail in this thesis.

1.4.3 DDX1 and the tRNA-LC roles in cell biology

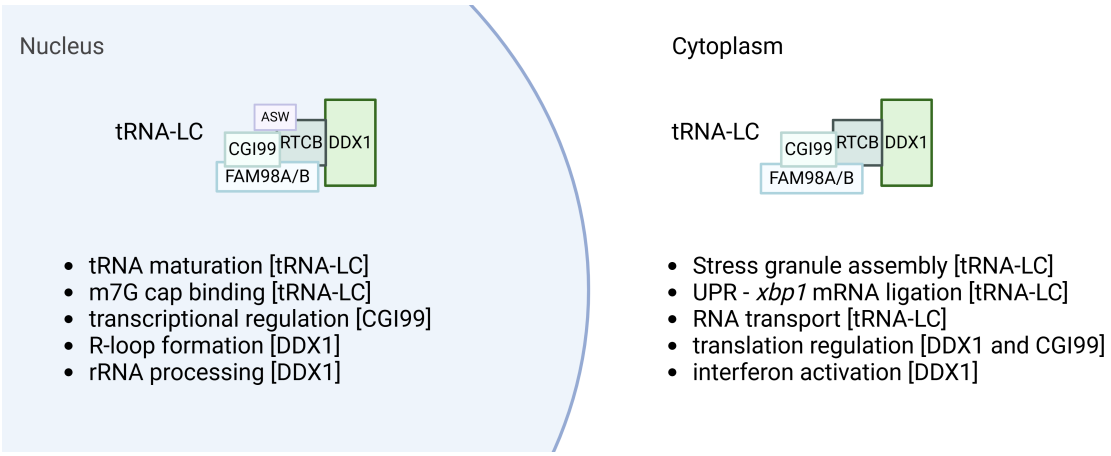

Figure 1.4: **Overview of tRNA-LC cellular functions schematic** Brackets signify whether the function was associated with the tRNA-LC or solely to a specific protein

DDX1 is a versatile protein with functions spanning RNA metabolism, stress response, and viral and immune regulation summarised in the schematic of Figure 1.4. Its diverse roles make it an important player in maintaining cellular homeostasis and responding to various cellular challenges. The close association with the tRNA-LC signify that many of these processes are

associated with this ligase complex.

DDX1 and the tRNA-LC play important roles in RNA-related processes such as mRNA and tRNA processing. The RTCB ligase, and subsequently the tRNA-LC, was identified in 2011 as the ligase responsible for tRNA maturation. Researchers identified RTCB as the protein responsible for the unusual ligation of RNA with 3' ends containing a 2',3'-cyclic phosphate (2',3'-cP) with an RNA fragment with a 5'-hydroxyl (5'-OH) [197]. These RNA termini occur specifically during tRNA intron excision. Archease was found to be required for tRNA-LC ligase activity and jointly with DDX1, facilitated the formation of an RTCB-guanylate intermediate essential for RNA ligation [198]. Indeed, Archease depletion impaired pre-tRNAs' maturation, as seen in depleted RTCB experiments. An RTCB kinetic experiment showed the stalled formation of ligation products after initial enzyme addition. This was subsequently rescued by adding Archease, indicating its presence is essential for enzymatic turnover. The RTCB ligase activity depends on the formation of Rtcb-guanylate intermediates that subsequently can transfer GMP to the 3'-end of the spliced RNA molecule, permitting ligation. This activity was further shown to be enabled by ATP binding to DDX1. Depletion of DDX1, or mutagenesis affecting ATP binding and hydrolysis, impacted the turnover of RTCB activity. Overall, the activity of RTCB depends on Archease and DDX1, which facilitate its guanylation after a single turnover, thus enabling another round of catalysis [198].

The tRNA-LC proteins have been further characterised in modulating RNA transport. The tRNA-LC is located both in the nucleus and cytoplasm. tRNA-LC proteins have been observed shuttling back and forth depending on transcriptional requirements in the nucleus [201]. RTCB, DDX1 and CGI99 were characterised as essential components involved in RNA-transporting granules in neurons [202]. These granules contribute to maintaining RNA stability, indicating that the tRNA-LC proteins are essential for this function.

The tRNA ligase has been linked not only to tRNA processing but also to other cellular functions, particularly its involvement in the unfolded protein response (UPR). Indeed, RTCB is critical in the activation of X box-binding protein 1 (XBP1) mRNA, a key step in UPR [203,204]. During ER stress, the sensor protein IRE1α undergoes dimerisation and phosphorylation, which activates its RNase function, enabling the removal of an intron from *XBP1* mRNA. This splicing event allows XBP1 to produce a functional transcription factor that upregulates genes involved in protein folding, degradation, and ER quality control [205]. The splicing by IRE1α, creates RNA termini identical to the ones observed during tRNA maturation. Studies show

that depletion of RTCB or co-factor, Archease, leads to incomplete XBP1 mRNA splicing, emphasising RTCB's role in this essential cellular process [203]. Furthermore, DDX1 has been associated with stress granule formation. It is recruited to stress granules when cells are exposed to various environmental stressors such as oxidative stress [206]. Proteomic analysis of SG components indicates that DDX1 is located in the SG core [207]. During stress conditions, DDX1 binds to and protects specific target mRNAs in the cytoplasm. The amount of target RNAs bound to DDX1 increases when cells are exposed to stress, and the overall levels of these RNAs are increased during stress in a DDX1-dependent manner [206].

DDX1 has been shown to function as both a pro- and antiviral factor across various viral infections, either facilitating viral replication and translation or inhibiting progression by promoting innate immune responses. DDX1 and the tRNA-LC have an extremely versatile role in cellular processes, which viruses could repurpose or alter to support infection. The specific roles of DDX1 and the tRNA-LC in alphavirus infection remain unexplored. This thesis will examine how these proteins may be diverted to influence infection dynamics in SINV, expanding our understanding of DDX1 and tRNA-LC in viral contexts.

1.5 Aims

The tRNA-LC proteins have been identified as dynamic RBPs in the SINV comparative RIC experiment [113] and direct interactors of SINV vRNA [123]. Furthermore, individual proteins of the complex, such as DDX1, have been highlighted as crucial players in aiding or inhibiting a range of viruses. I hypothesise the tRNA-LC multi-faceted functions in cellular biology are repurposed during SINV infection to restrict viral fitness. This thesis aims to extend our understanding of the tRNA-LC's behaviour under homeostatic conditions and determine how and to what end this behaviour might be altered during SINV infection.

1. AIM 1: Understand the interwoven nature of the tRNA-LC *in cellulo*. Chapter 4 captures the co-dependency of the tRNA-LC proteins in complex formation. It furthermore expands on our understanding of the complex's structure using experimental and computational methods.

2. AIM 2: Explore the role of the tRNA-LC during SINV infection. Chapter 5 explores how the knockdown of key tRNA-LC proteins affects viral fitness and whether the enzymatic activity of the complex can elucidate a functional role in virus infection. Transcriptomic changes that occur during viral infection in the absence of key tRNA-LC proteins are further explored.

3. AIM 3: Capture the tRNA-LC interactome in the presence or absence of infection. Chapter 6 investigates the wider protein and RNA interaction interface of the tRNA-LC. This expanded dataset hints at the mechanism behind viral modulation. Furthermore, viral replication and translation can be pried apart to elucidate which process the tRNA-LC is involved in regulating.

2 Materials

2.1 Reagents & Consumables

Reagent	Supplier	Identifier
10x Tris/Glycine/SDS Running buffer	BioRad	1610732
2X Phusion HF PCR Master mix	NEB	M0531L
4–15% Mini-PROTEAN TGX Precast Protein Gel	Bio-Rad	4561084
5' Deadenylase	NEB	M0331S
96-well, Cell Culture-Treated, Black Flat-Bottom Microplate	Greiner	10369081
Acetic acid glacial	Fisher Chemical	A/0360/PB17
AEBSF	BioChemica	A14210100
Ampure XP beads	Beckman Coulter	A63881
Ammonium persulfate	Thermo Fisher	17874
BamHI	NEB	R3136
Benzonase Nuclease	Millipore	E1014
Colour prestained protein standard, broad range	NEB	P7719S
DH5α E. coli	NEB	C2987H
DMEM	Thermo Fisher	11995065
DMEM (no phenol red)	Thermo Fisher	21063029
DMSO	Sigma-Aldrich	D2260
DpnI	NEB	R0176S
DTT	Sigma Aldrich	D1532
DSS	Thermo Fisher	A39267
Dynabeads MyOne Silane	Thermo Fisher	37002D
EDTA	Millipore	324503
Ethanol absolute		
ERCC spike-in	Thermo	4456740

Table 2.1: Reagents & Consumables (Part 1)

Reagent	Supplier	Identifier
EvaGreen	Biotium	31000
FastAP alkaline phosphatase	Thermo Fisher	EF0654
FBS	Sigma	F9665-500ML
GFP-Trap agarose beads	Chromotek	gta-20
HiScribe T7 Arca	NEB	E2065
Hygromycin B	Thermo Fisher	J60681.MD
IGEPAL CA-630	Sigma-Aldrich	I8896
Lipofectamine 3000	Invitrogen	L3000008
Lipofectamine RNAiMax	Invitrogen	13778075
Lipofectamine messenger Max	Invitrogen	LMRNA003
Luna Universal One-step RT-qPCR kit	NEB	E3005E
NuPage LDS Sample Buffer	Invitrogen	NP0007
NuPAGETM4-12% Bis-Tris 1mm gel	Invitrogen	NP0322BOX
NotI	NEB	R3189
Oligo(dT)25 beads	NEB	S1419S
Opti-MEM reduced serum Medium	Gibco	31985062
PBS	Life technology	10010056
Penicillin/streptomycin	Sigma	P4458-100ML
PFU Turbo DNA polymerase	Agilent	600252
Phenol:Chloroform:Isoamyl Alcohol	Sigma-Aldrich	P3803
Pierce control agarose resin	Thermo Fisher	26150
PNK	New England Biolabs	M0201L

Table 2.2: Reagents & Consumables (Part 2)

Reagent	Supplier	Identifier
ProNex	Promega	NG2001
Protein Lobind tubes 1.5 ml	Eppendorf	0030108116
Proteinase K	Roche	3115828001
Qubit HS RNA Assay Kit	Invitrogen	Q32855
Qubit HS dsDNA Assay Kit	Invitrogen	Q33230
Qubit Protein Broad Range Assay Kit	Invitrogen	A50669
RecJf endonuclease	New England Biolabs	M0264S
RiboLock RNase Inhibitor	Thermo Fisher	EO0381
Ribonucleoside Vanadyl Complex	NEB	S1402S
RNase I	Thermo Fisher	AM2294
SilverQuest staining kit	Invitrogen	LC6070
Sodium Pyruvate	Thermo Fisher	11360070
SpeedBeads Magnetic Carboxylate Modified Particles	Sigma-Aldrich	GE651521 05050250
SpeI	NEB	R3133
Superscript IV	Thermo Fisher	18090010
T4 DNA ligase	NEB	M0202S
T4 RNA ligase	Thermo Fisher	EL0021
Triton X-100	Promega	H5141
TruSeq Stranded mRNA Library Prep	Illumina	20020594
Trypan Blue Stain (0.4%)	Invitrogen	T10282
TrypLE Express Enzyme	Gibco	12604013
TurboDNase	Thermo Fisher	AM2238
Tween-20	Sigma Aldrich	P1379
XhoI	NEB	R0146S

Table 2.3: Reagents & Consumables (Part 3)

2.2 In-House Buffers

Buffer	Composition
RIPA lysis buffer	10 mM Tris HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.1% SDS, 1% Triton X-100, 1% sodium deoxycholate
IP lysis buffer (mild lysis)	10 mM Tris HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5 % IGEPAL
IP wash buffer	10 mM Tris HCl pH 7.5, 150 mM NaCl, 0.05 % IGEPAL, 0.5 mM EDTA
iCLIP 5x PNK buffer	350 mM Tris HCl pH 6.5, 50 mM MgCl2, 25 mM DTT
iCLIP higher salt buffer	1M NaCl, 1 mM MgCl2, 50 mM Tris HCl pH 7.5, 0.05% NP40, 0.2% SDS, 1 mM DTT, 1x AEBSF
iCLIP medium salt buffer	250 mM NaCl, 20 mM Tris HCl pH 7.5, 1 mM MgCl2, 0.05% NP40, 1x AEBSF
iCLIP PK-SDS solution	10 mM Tris HCl pH 7.4, 100 mM NaCl, 1 mM EDTA, and 0.2% SDS
iCLIP PNK buffer	20 mM Tris HCl pH 7.5, 10 mM MgCl2. 0.2% Tween-20
Denaturing buffer 1	50mM Tris HCl pH 7.5, 2M NaCl, 1mM EDTA pH 8, 1% NP40, 0.1% SDS, 0.5% sodium deoxycholate, 1x AEBSF
Denaturing buffer 2	50mM Tris HCl pH 7.5, 1M NaCl, 1mM EDTA pH 8, 1% NP40, 0.2% SDS, 0.5% sodium deoxycholate, 1x AEBSF
Denaturing buffer 3	10mM Tris HCl pH 7.5, 150M NaCl, 0.5mM EDTA pH 8, 0.05% NP40, 4M Urea
Denaturing buffer 4	10mM Tris HCl pH 7.5, 150M NaCl, 0.5mM EDTA pH 8, 0.05% NP40, 10mM DTT

Table 2.4: In-House Buffers

2.3 Antibodies & Dyes

Reagent	Supplier	Identifier
beta-actin primary antibody	Sigma	A1978
GFP primary antibody	chromotek	3h9-100
RFP primary antibody	chromotek	5F8-100
IRDye 680RD Donkey anti-Rabbit IgG	LI-COR Biosciences	926-68073
IRDye 680RD Goat anti-Mouse IgG	LI-COR Biosciences	926-68070
IRDye 800CW Donkey anti-Rabbit IgG	LI-COR Biosciences	926-32213
IRDye 800CW Goat anti-Human IgG	LI-COR Biosciences	926-32232
IRDye 800CW Goat anti-Rat IgG	LI-COR Biosciences	926-32219
oligo(dT) stellaris probe	Biosearch tech	N/A
DDX1 primary antibody	Atlas Antibody	HPA008320
RTCB primary antibody	CusAb	CSB-PA897546LA01
CGI99 primary antibody	Abcam	ab188326
FAM98A primary antibody	Avivasbio	ARP55265
EiF2-alpha phospho primary antibody	Cell Signaling	9721S
SINV Capsid primary antibody	Lab of L. Carrasco	N/A

Table 2.5: Antibodies & Dyes

2.4 Plasmids

Plasmid	Backbone	Source
pcDNA5-FRT-TO-DDX1-GFP	pcDNA5-FRT-TO	Castello Lab
pcDNA5-FRT-TO-RTCB-GFP	pcDNA5-FRT-TO	Castello Lab
pcDNA5-FRT-TO-CGI99-GFP	pcDNA5-FRT-TO	Made in this project
pcDNA5-FRT-TO-FAM98A-GFP	pcDNA5-FRT-TO	Castello Lab
pcDNA5-FRT-TO-FAM98B-GFP	pcDNA5-FRT-TO	Biobasic
pcDNA5-FRT-TO-ASW-GFP	pcDNA5-FRT-TO	Biobasic
pcDNA5-FRT-TO-ANGEL2-GFP	pcDNA5-FRT-TO	Made in this project
pOG44 Flp-recombinase expr vector	pOG44	Thermo Fisher
pT7-SVmCherry	pTE3'2J1	Castello lab
pT7-SVnsP3Scarlet	pTE3'2J1	Castello lab
pT7-SVwt	pTE3'2J1	Lab of L. Carrasco
SINV-reporter-luc	pMC-GTU	Lab of A. Merits
SINV-p1234-luc	pUC57	Lab of A. Merits
pT7-SINV-nsp3TAA-luc	pTE3'2J1	Made in this project
pT7-SINV-nsp3-luc	pTE3'2J1	Made in this project
pLKO-tet-on-shDDX1	pLKO-tet-on	Castello lab
pLKO-tet-on-shCGI99	pLKO-tet-on	Made in this project
pLKO-tet-on-shFAM98A	pLKO-tet-on	Castello Lab
pLKO-tet-on-shRTCB	pLKO-tet-on	Made in this project
pLKO-tet-on-shFAM98B	pLKO-tet-on	Made in this project
VSVG	pLP	Castello Lab
PAX2		Castello Lab

Table 2.6: Plasmids

2.5 Oligonucleotides

Oligonucleotide	Sequence
L01clip2.0	NNNNATCACGNNNNNAGATCGGAAGAGCGTCGTG
L02clip2.0	NNNNCGATGTNNNNNAGATCGGAAGAGCGTCGTG
L03clip2.0	NNNNTTAGGCNNNNNAGATCGGAAGAGCGTCGTG
L04clip2.0	NNNNTGACCANNNNNAGATCGGAAGAGCGTCGTG
L05clip2.0	NNNNACAGTGNNNNNAGATCGGAAGAGCGTCGTG
L06clip2.0	NNNNGCCAATNNNNNAGATCGGAAGAGCGTCGTG
L07clip2.0	NNNNCAGATCNNNNNAGATCGGAAGAGCGTCGTG
L08clip2.0	NNNNACTTGANNNNNAGATCGGAAGAGCGTCGTG
L09clip2.0	NNNNGATCAGNNNNNAGATCGGAAGAGCGTCGTG
L10clip2.0	NNNNTAGCTTNNNNNAGATCGGAAGAGCGTCGTG
L11clip2.0	NNNNATGAGCNNNNNAGATCGGAAGAGCGTCGTG
L12clip2.0	NNNNCTTGTANNNNNAGATCGGAAGAGCGTCGTG
L13clip2.0	NNNNAGTCAANNNNNAGATCGGAAGAGCGTCGTG
L14clip2.0	NNNNAGTTCCNNNNNAGATCGGAAGAGCGTCGTG
L15clip2.0	NNNNATGTCANNNNNAGATCGGAAGAGCGTCGTG
L16clip2.0	NNNNCCGTCCNNNNNAGATCGGAAGAGCGTCGTG
L17clip2.0	NNNNCAACTANNNNNAGATCGGAAGAGCGTCGTG
L18clip2.0	NNNNGTCCGCNNNNNAGATCGGAAGAGCGTCGTG
L19clip2.0	NNNNGTGAAANNNNNAGATCGGAAGAGCGTCGTG
L20clip2.0	NNNNCACCGGNNNNNAGATCGGAAGAGCGTCGTG

Table 2.7: iClip barcodes

Oligonucleotide	**Sequence**
L3-App	AGATCGGAAGAGCGGTTCAG
P3Solexa-FWD	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTC CTGCTGAACCGCTCTTCCGATCT
P3Solexa_s - FWD	CTGAACCGCTCTTCCGATCT
P5Solexa-REV	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTA CACGACGCTCTTCCGATCT
P5Solexa_s -FWD	ACACGACGCTCTTCCGATCT
RToligo - REV	GGATCCTGAACCGCT
ddx1-FWD	GAACCTTCCCGGGAGTTAGC
ddx1-REV	AGTCTTCCCGGAGTACCTACA
cgi99-FWD	GACCATGTTCCGACGCAAGT
cgi99-REV	CAAAGAACTTGGGCCAGTCG
archease-FWD-1	ATCAAGGCCAAGTATCCGCC
archease-REV-1	TCTCCCCATGCGTGTAACTG
archease-FWD-2	CAGAAGGCGATCAAGGCCA
archease-REV-2	CAGAGTATCTCCCCATGCGT
SINV-nsp2-FWD	GGAGGGGCTCCAGGCGGACATCG
SINV-nsp2-REV	GCTCCTCTTCTGTATTCTTGGCG
SINV-capsid-FWD	GAACGAGGACGGAGATGTCATCG
SINV-capsid-REV	CAGCGCCACCGAGGACTATCGC
beta-actin-FWD	CGCGAGAAGATGACCCAGAT
beta-actin-REV	TCACCGGAGTCCATCACGAT
gapdh-FWD	ATGGGGAAGGTGAAGGTCG
gapdh-REV	GGGGTCATTGATGGCAACAATA

Table 2.8: Oligonucleotides

2.6 Silencing targets

Target	Sequence
shDDX1	GAUGUGGUCUGAAGCUAUUAA
siDDX1-PF	GAUGUGGUCUGAAGCUAUUAA
siDDX1-2	GAGCCACAUUAGAACUGAU
siDDX1-3	GGAGUUAGCUGAACAAACU
siCTRL	TTCTCCGAACGTGTCACGT
siCTRL (pool)	Dhermacon [D-001810-10-05]
siDDX1 (pool)	Dhermacon [L-011993-01-0005]
siCGI99 (pool)	Dhermacon [L-020723-01-0005]
siRTCB (pool)	Dhermacon [L-017647-00-0005]
siArchease (pool)	Dhermacon [L-017915-01-0005]

Table 2.9: Silencing targets

2.7 Cell lines

Cell line	Source	Modified?
HEK293	ECACC #85120602	Parental
HEK293T	Castello lab	Parental
HEK293 Flp/In T-Rex (FITR)	Thermo Fisher #R78007	Parental
VeroE6	Castello lab	Parental
BHK	Castello lab	Parental
BSR-T7	Lab of R. Elliot	Modified
HEK293 FITR DDX1-GFP	Castello lab	Modified
HEK293 FITR DDX1-K52A-GFP	Castello lab	Modified
HEK293 FITR RTCB-GFP	Castello Lab	Modified
HEK293 FITR RTCB-C122A-GFP	Castello Lab	Modified
HEK293 FITR CGI99-GFP	Made in this project	Modified
HEK293 FITR FAM98A-GFP	Made in this project	Modified
HEK293 FITR FAM98B-GFP	Made in this project	Modified
HEK293 FITR ASW-GFP	Made in this project	Modified
HEK293 FITR shDDX1	Made in this project	Modified
HEK293 FITR shCGI99	Made in this project	Modified
HEK293 FITR shFAM98A	Castello Lab	Modified

Table 2.10: Cell Lines

2.8 Instruments & Equipment

Instrument	Supplier
Clariostar platereader	BMG LABTECH
Countess II FL Automated Cell Counter	Thermo Fisher
EVOS M5000	Thermo Fisher
GelDoc Imaging System	BioRad
NanoDrop Microvolume Spectrophotometer	Thermo Fisher
Odyssey CLx Imaging System	LiCor
NextSeq500	Illumina
TapeStation 4000	Agilent
PCR machine	ABI
PowerPac Basic Power Supply	BioRad
QuantStudio real-time PCR system	Applied Biosystems
Qubit Fluorometer	Thermo Fisher
Transblot Turbo Transfer System	BioRad
254nm UV Crosslinker	Roth Selection
GloMax	Promega
NextSeq 550	Illumina

Table 2.11: Instruments & Equipment

3 Methods

3.1 Cell Biology

3.1.1 Maintenance of Cells

All cells were kept at 37°C with 5% CO_2. Cells were maintained in Dulbecco's modified eagle medium (DMEM) with 10% fetal bovine serum (FBS) and 1x penicillin-streptomycin. In addition to standard culture conditions, HEK293 Flp/In T-REx stable cell lines were kept in 150 µg mL^{-1} hygromycin B and 5 µg mL^{-1} blasticidin and HEK293/HeLa shRNA stable cell lines were kept in 1 µg mL^{-1} puromycin. Cells were passaged regularly to maintain a confluency of $<$90%. To split cells, media was removed and cells were washed once with 1x PBS. 1x TrypLE Express enzyme was then added and cells were incubated for 2 min at 37°C with 5% CO_2. Cells were resuspended in 10% FBS DMEM and seeded at the desired density. When required, cells were counted using Trypan Blue Stain (0.4%) and the Countess II FL automated cell counter.

3.1.2 Stable Transfection of HEK293 Flip/In T-REx

2×10^6 HEK293 Flip/In T-REx parental cells were seeded in a T25 flask in DMEM (10% FBS, 1x P/S). 24 hours after seeding and 1 hour before transfection, tissue culture media was changed to fresh DMEM (10% FBS, 1x P/S). Transfection was performed using a Lipofectamine 3000 kit following manufacturer guidelines. Two tubes of reaction mix were prepared as below, mixed, and incubated at RT for 10-15 minutes. The DNA-lipid complex was then pipetted directly onto cells.

Component	**Amount**
Opti-MEM Medium	750 µL
Lipofectamine 3000 Reagent	45 µL DNA

Table 3.1: Lipofectamine 3000 transfection Tube 1

Component	Amount
Opti-MEM Medium	750 µL
pOG44 Flp-recombinase expression vector	5.33 µg
pcDNA5-FRT-RBP plasmid	0.67 µg
P3000 Reagent	2 µL µg^{-1} DNA

Table 3.2: Lipofectamine 3000 transfection Tube 2 - FITR

After 24 hours, the media was changed to fresh DMEM (10% FBS, 1x P/S). After a further 24 hours, cells were split into a T75 flask, and 150 µg mL^{-1} hygromycin B was added to cell culture media for cell selection.

3.1.3 siRNA Knockdowns

7.5×10^4 HEK293 Flip/In T-REx parental cells per well were seeded in a 24 well plate in 1ml DMEM (10% FBS, 1x P/S). 24 hours after seeding, 500 µL media was removed from each well and replaced with 500 µL serum-free DMEM. Transfections were performed using Lipofectamine RNAiMAX transfection reagent following the manufacturer's guidelines (Thermo). First, 3 µL RNAiMax reagent was combined with 50 µL OptiMEM. In a separate tube 1 µL 25 µM siRNA was combined with 50 µL OptiMEM. The two mixtures were then combined and incubated for 5min at RT before being pipetted drop-wise onto cells. Cells were incubated for 48 hours before being infected with SINV at an MOI of 0.1. For infection, 500 µL media was removed from each well and 500 µL of serum-free media (with virus) was added. Cells were incubated for 18 hours. To harvest, cells were washed in 1x PBS and lysed in RIPA buffer before being processed for western blot. Protein concentrations were measured using Qubit Protein BR assay, and loading volumes were normalised to the lowest protein concentration.

3.1.4 Generation of viruses

SINV was either generated directly from the plasmid or through the expansion of a pre-existing virus stock. For production from plasmid, the plasmid was first linearised by restriction digest with XhoI.

Component	Amount
DNA	1 µg
10X rCutSmart Buffer	5 µL (1X)
XhoI	1 µL
Nuclease Free Water	Up to 50 µL

Table 3.3: Linearisation reaction mix

Plasmid was incubated in the above mix for 15 minutes at 37°C, then at 65°C for 20 minutes. mRNA was then synthesised by incubating linearised plasmid at 37°C for 30 minutes with T7 RNA polymerase mix.

Component	Amount
DNA	1 µg
2X ARCA/NTP Mix	10 µL
T7 RNA Polymerase Mix	2 µL
Nuclease Free Water	Up to 20 µL

Table 3.4: T7 RNA polymerase mix

DNase I was then added and the mixture was incubated for 15 minutes at 37°C. RNA was purified using RNeasy Mini Kit and quantified using nanodrop. 2.5 µg mRNA was transfected into BHK21 cells at 70-90% confluency in a 10cm dish. Lipofectamine 3000 reagents were mixed as above with the adjustment of Tube 2 as below:

Component	Amount
Opti-MEM Medium	750 µL
RNA	2.5 µg
P3000 Reagent	2 µL μg^{-1} DNA

Table 3.5: Lipofectamine 3000 transfection Tube 2 - Virus RNA

Cells were incubated for 24-48h (37°C, 5% CO_2). Cells were monitored regularly and pH was buffered with HEPES if media became too acidic. The supernatant was harvested when nearly all cells were showing cytopathic effects. To harvest, supernatant was transferred to a centrifuge tube and HEPES solution was added to a final concentration of 500 mM. The supernatant was centrifuged at 1000rpm for 5 minutes to remove cell debris and was then

passed through a 0.45 µm filter. Single-use aliquots were stored at -80°C. For expansion stocks, a T175 flask of BHK21 cells were infected with virus at an MOI of 0.1 in 15ml DMEM. Cells were incubated for 48 hours (37°C, 5% CO_2) and harvested as described above. No more than one round of expansion was performed for tagged virus stocks.

3.1.5 Titration of viruses

SINV was titrated by plaque assay. 1×10^5 Vero cells per well were seeded in a 24-well plate in DMEM (2% FBS, 1x P/S) and incubated overnight. Serial dilution of virus stock or viral supernatant was prepared in DMEM (2% FBS, 1x P/S). Media was carefully removed from 24-well plate and 100 µL of virus stock serial dilution was added to each well. Cells were incubated for 1 hour at 37°C, before 0.5 µL of prepared 0.6% Avicell overlay was added. Cells were gently shaken to distribute the overlay evenly. After 3 days of incubation (37°C, 5% CO_2), Avicell was removed and 500 µL 10% Formaldehyde was added to each well. Plates were incubated in the fume hood for 1 hour, then rinsed twice with PBS. Coomassie blue stain was added and plates were incubated for 1 hour, before plates were washed in water and plaques were counted. Titer was calculated as follows:

$$\text{PFU/ml} = \frac{\text{Average number of plaques}}{\text{Dilution of stock}\times\text{Volume of inoculum}}$$

3.1.6 Platereader assay

4×10^4 HEK293 cells or 2×10^4 HeLa cells per well were seeded on a black, clear-bottomed 96 well plate in 100 µL colourless DMEM (5% FBS, 1x P/S). When using an inducible cell line, media for the 'induced' condition was supplemented with 1 µg mL^{-1} doxycycline. 24 hours after seeding, 100 µL colourless DMEM (0% FBS, 1x P/S) containing virus at the desired MOI was added. Cells were then incubated at 37°C with 5% CO_2 in a CLARIOstar fluorescence plate reader. mCherry/ mScarlet signal was measured every 15 min over a 24-hour period to give a read-out of virus replication.

3.2 Biochemical and molecular biology techniques

3.2.1 Western blotting

Samples were prepared by mixing with NuPAGE 4x LDS sample buffer (Invitrogen) and heated to 95°C for 5 minutes. Proteins were then loaded into 1.5mm SDS- polyacrylamide gels (10%),

prepared using the TGX FastCast acrylamide kit (Bio-Rad). Proteins were separated by SDS-PAGE for 55 minutes at 180V and transferred to nitrocellulose membranes using the Trans-Blot Turbo Transfer System (Bio-Rad). Membranes were blocked for 1 hour at RT or overnight at 4°C. 5% skim milk prepared in 0.1% PBS-T was used for blocking and for preparing antibody dilutions. Membranes were incubated with primary antibody either for 1 hour at RT or overnight at 4°C. After three washes in 0.1% PBS-T, membranes were incubated with the relevant secondary antibodies for 1 hour at RT. Membranes were then washed three more times in 0.1% PBS-T before imaging. Imaging was performed on the LI-COR Odyssey Fc or LI-COR Odyssey CLx.

3.2.2 Silver staining

Samples were prepared by mixing with NuPAGE 4x LDS sample buffer (Invitrogen) and heated to 95°C for 5 minutes. Proteins were then loaded into 1.5mm SDS-polyacrylamide gels (10%) and prepared using the TGX FastCast acrylamide kit (Bio-Rad). Proteins were separated by SDS-PAGE for 55 minutes at 180V and then washed briefly in ultrapure water. Staining was performed using the SilverQuest silver staining kit, according to their basic staining protocol.

3.2.3 Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

Samples were extracted using the QIAGEN Rneasy kit. RNA was quantified using Nanodrop. 100ng of RNA was added to selected primers and LunaScript One-Step RT-PCR Kit. Following the manufacturer's guidelines, PCR cycling was set up as follows:

Step	Number of Cycles	Temperature (℃)	Time
Reverse Transcription	1	55	10 min
RT inactivation / Initial Denaturation	1	98	1min
Denaturing	35	98	10 s
Annealing		62	30 s
Extension		72	30s
Final extemsion	1	72	5min

Table 3.6: PCR mutagenesis thermocycling conditions

3.2.4 Cloning

DNA sequence for Renilla luciferase was amplified from a pRL-null vector (Promega) using Renilla FWD and Renilla REV or Renilla-TAA REV primers in PCR. The PCR product was separated on a 0.8% agarose gel, and the Renilla band was cut out and purified using a Qiagen QIAquick gel extraction kit. The purified Renilla fragment and pt7-SINV-WT vector were digested using SpeI restriction enzyme as per manufacturer instructions. The digested products were separated on a 0.8% agarose gel and then purified using the QIAquick gel extraction kit. The purified vector was then dephosphorylated using antarctic phosphatase and 20 ng of Renilla fragment was ligated into 50 ng vector backbone using T4 DNA ligase as per manufacturer instructions. The assembled plasmid was transformed into DH5α E. coli and resulting colonies were screened for successful transformation via PCR. After confirmation with sequencing, the pT7-SINV-nsp3TAA-luc plasmid was prepared using the Qiagen Plasmid Maxiprep kit.

3.2.5 Dual luciferase assay - replicase

7.5×10^3 HEK293 cells were seeded were seeded in a 24-well plate in DMEM (10% FBS, 1x P/S) and incubated overnight. 24 hours after seeding, half the media was removed from each well and replaced with serum-free DMEM. Transfections of siRNAs were performed using Lipofectamine RNAiMAX. After 48h of incubation, half the media was removed from each well and replaced with serum-free DMEM. Transfection of replicase replicons were prepared using Lipofectamine 3000. Lipofectamine 3000 reagents were mixed as above with the adjustment of the volume so the total volume of tube 1 transfectant was 100 µL and Tube 2 consisted of:

Component	**Amount**
Opti-MEM Medium	50 µL
DNA replicase - SINV-P1234 / or SINV-P1234-GAA	0.125 µg
DNA reporter - SINV Fluc-Gluc	0.25 µg
P3000 Reagent	2 µL μg^{-1} DNA

Table 3.7: Lipofectamine 3000 transfection Tube 2 - Translation replicon

Cells were incubated for 18h (37°C, 5% CO_2). Supernatant was removed and 100 µL of prepared Promega dual fluorescence assay 1x Passive lysis buffer was pipetted on cells. Plate

was placed on rocker at RT for 15min. 20 µL of lysed cells were pipetted into a clear bottomed black 96 well plate, along with 100 µL of prepared Luciferase Assay Buffer II. Plate well was measured using a Promega Glo-Max for Firefly luciferase luminescence. After the first reading, 100 µL of prepared Stop & Glo reagent was added to wells and a second reading was taken to measure Renilla luciferase.

3.2.6 Dual luciferase assay - translation

***In vitro* transcription** pT7-SINV-nsp3TAA-luc and pT7-SINV-nsp3-luc were linearised using XhoI as previously outlined. In vitro transcription was performed using the HiScribe T7 Arca kit, followed by DNase treatment. RNA was purified using QIAGEN RNeasy kit and quantified by Nanodrop.

Transfection 1.5×10^4 HEK293 cells were seeded were seeded in a 12-well plate in DMEM (10% FBS, 1x P/S) and incubated overnight. 24 hours after seeding, half the media was removed from each well and replaced with serum-free DMEM. Transfections of siRNAs were performed using Lipofectamine RNAiMAX. After a further 48h of incubation, half the media was removed from each well and replaced with serum-free DMEM. Transfection of translation replicons was prepared using Lipofectamine messengerMax following the manufacturer's guidelines. First, 6 µL messengerMax reagent was combined with 100 µL OptiMEM. In a separate tube 1 µg of in vitro transcribed RNA was combined with 100 µL OptiMEM. The two mixtures were then combined and incubated for 5min at RT before being pipetted drop-wise onto cells. Cells were incubated for 4 hours.

Dual fluorescence harvest Supernatant was removed and 200 µL of prepared Promega dual fluorescence assay 1x Passive lysis buffer was pipetted on cells. The plate was placed on a rocker at RT for 15min. 20 µL of lysed cells were pipetted into a clear-bottomed black 96 well plate, along with 100 µL of prepared Luciferase Assay Buffer II. followed by 100 µL of prepared Stop & Glo reagent. Plate wells were measured using a Promega Glo-Max for Renilla luciferase luminescence.

3.3 DSS crosslinking protein-protein

Sample harvesting 8×10^6 HEK293 Flp-In T-REx DDX1-eGFP and HEK293 Flp-In T-REx parental cells were seeded per condition in a 10cm dish in DMEM (10% FBS, 1x P/S). Cells were further induced with 1 µg/ml of doxycycline overnight. After 24h incubation, cells were washed in 5ml PBS, followed by a 10min incubation at RT with 2ml diluted dissolved DSS crosslinker (4mM DSS dissolved in PBS). Crosslinker activity was subsequently quenched with 100 µL 1M Tris pH 7.4 (final concentration 50mM Tris to 4mM DSS) and was further incubated at RT for 5min. Cells were harvested using a cell scraper and transferred into a centrifugation tube. Samples were centrifuged for 5 min (1000rpm, 4°C) to sediment cells. Sedimented cells were resuspended in RIPA lysis buffer and incubated for 15min at 4°C. Lysates were spun for 15 minutes (max speed, 4°C). The supernatant was transferred to another fresh tube, snap frozen on dry ice and stored at -80°C.

Immunoprecipitation Thawed samples were immunoprecipitated using GFP-trap agarose beads followed by five RIPA washes. To elute, beads were resuspended in 50 µL 1% SDS and incubated for 5 minutes at 55°C with rotation (1100rpm). Samples were then spun down for 2 minutes (RT, 2500xg) and the eluate was transferred to a fresh tube. Elution was repeated once more, and eluates were combined.

Mass Spectrometry Eluates were transferred to the Rosalind Franklin Centre for mass spectrometry. Analysis of peptides was carried out using an Ultimate 3000 nano-LC 1000 system coupled to an Orbitrap Fusion Ascend Tribrid Mass Spectrometer (Thermo Fisher Scientific).

3.4 RNA sequencing

Sample harvesting 1.5×10^4 HEK293 cells were seeded were seeded in a 12-well plate in DMEM (10% FBS, 1x P/S) and incubated overnight. 24 hours after seeding, half the media was removed from each well and replaced with serum-free DMEM. Transfections of siRNAs were performed using Lipofectamine RNAiMAX. Cells were incubated for 48 hours before being infected with SINV at an MOI of 0.1. For infection, 500 µL media was removed from each well and 500 µL of serum-free media (with virus) was added. Cells were incubated for 18 hours. To harvest the cells, wells were washed in PBS and resuspended in RLT buffer. RNA was

extracted from the lysed cells using the Qiagen RNeasy extraction kit.

Library preparation and sequencing RNA was quantified by Qubit Fluorometer 4 (Life Technology) using the HS RNA assay kit and dsDNA HS assay kit. RNA quality was verified via Tapestation 4200 (Agilent) using HS RNA Screen Tape assay. Mild DNAse treatment was performed to obtain $\leq$ 5% DNA contamination in the examples. Total RNA (500 ng) was mixed with the ERCC spike-in control according to the manufacturer's guidelines (1:100 dilution). The obtained mix was used to prepare libraries for sequencing using the the Illumina TruSeq Stranded mRNA Library Prep kit and SuperScript II Reverse Transcriptase according to the manufacturer's instructions. The PCR amplified dual indexed libraries were cleaned up with Agencourt AMPure XP magnetic beads, quantified using Qubit Fluorometer 4 and Qubit dsDNA HS Assay Kit. Their size distribution was assessed using a 4200 TapeStation System with a High Sensitivity D1000 Screen Tape assay. Libraries were pooled in equimolar concentrations and sequenced using a 75bp single read high-output cartridges on an Illumina NextSeq 550 sequencer. A Q score of $\geq$30 was presented in at least 94% of the sequencing reads generated.

3.5 Protein-protein interaction analysis

Sample harvesting 8×10^6 HEK293 Flp-In T-REx DDX1-eGFP and HEK293 Flp-In T-REx parental cells were seeded per condition in a 10cm dish in DMEM (10% FBS, 1x P/S). Cells were further induced with 1 µg/ml of doxycycline overnight. For the infected condition, cells were infected with a MOI of 3 for 8h or 18h. When harvesting, cells were washed in 5ml PBS, then resuspended in 1ml PBS and transferred to a 1.5ml Eppendorf. Samples were spun down for 5 minutes (1500rpm, 4°C) and PBS was removed. 1000 µL IP lysis buffer was added, and samples were incubated on ice for 30 minutes. Samples were centrifuged for 3 minutes (2000rpm, 4°C) to seperate the nuclei. The cytosol fraction in the supernatant was transferred to a fresh tubes and was spun for 15 minutes (max speed, 4°C). The supernatant was transferred to another fresh tube, snap frozen on dry ice and stored at -80°C.

co-IP Samples were first incubated 25 µL with agarose control beads for 30 minutes with rotation at 4°C to pre-clear them. They were then incubated with 30 µL GFP-Trap beads for 2 hours. Samples were spun down (2500xg, 4°C, 5 minutes) and the supernatant was

removed. Beads were resuspended in 1ml IP Wash Buffer and divided between two tubes, one for Benzonase treatment and one to be left untreated. Tubes were then spun down again, as above, and beads were resuspended in either 500 µL IP Wash Buffer supplemented with 1 µL mL^{-1} Benzonase or in wash buffer without Benzonase. Benzonase-treated samples were incubated for 15 minutes at 37°C and untreated samples were stored on ice. All samples were then washed five more times in 500 µL IP Wash Buffer. To elute, beads were resuspended in 50 µL 1% SDS and incubated for 5 minutes at 55°C with rotation (1100rpm). Samples were then spun down for 2 minutes (RT, 2500xg) and the eluate was transferred to a fresh tube. Elution was repeated once more, and eluates were combined.

Mass Spectrometry Eluates were transferred to the Rosalind Franklin Centre for mass spectrometry. Analysis of peptides was carried out using an Ultimate 3000 nano-LC 1000 system coupled to an Orbitrap Fusion Ascend Tribrid Mass Spectrometer (Thermo Fisher Scientific).

3.6 iCLIP2

Sample harvesting For IP samples, 5×10^6 HEK293 Flp-In T-REx RBP-eGFP cells were seeded per condition in a 10cm dish in DMEM supplemented with 2% FBS. Cells were induced with 1 µg/ml of doxycycline overnight. For the infected condition, cells were infected with 3 MOI of SINV 9 hours before harvesting. For the IFN-treated condition, cells 500U/ml IFNA2 was added to media 20 hours before harvesting. For Size Matched Input (SMI) samples, parental HEK293 Flp-In T-REx cells were seeded and treated in the same way. Next, cells were washed twice in 1X PBS and crosslinked at 0.3 J/cm2 UV light irradiation at 254 nm. Cells were then lysed in 1 ml RIPA buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.5% wt/vol Na deoxycholate and 0.2 mM AEBSF). Lysates were incubated for 30 min on ice, then homogenized by passing through a 27G needle five times. Finally, lysates were cleared by centrifugation (18000xg for 10 min at 4℃), snap-frozen in dry ice, and stored at -80℃ until use.

iCLIP2 4 U TurboDNase and 5 U RNase I were added to lysates. These were mixed by vortexing and incubated for 3 min at 37℃ (1100 rpm). 200 U RiboLock RNase Inhibitor was then added, and lysates were incubated for a further 3 min on ice. Lysates were pre-cleared

with 25 µL of pre-equilibrated control agarose beads for 30 min at 4 ℃ with gentle rotation followed by centrifugation at 2500xg for 2 min. Lysates were transferred to fresh tubes and incubated with 25 µL of pre-equilibrated GFP_Trap agarose bead slurry for 2 h at 4 ℃ with gentle rotation. Beads were washed twice with 900 µL of cold high-salt wash buffer, twice with 900 µL of cold medium-salt wash buffer, and twice with 900 µL of cold PNK wash buffer. RNA 3'-end dephosphorylation was performed at 37 ℃ for 40 min (1100 rpm) in 5x PNK buffer with 5 U PNK, 0.25 U FastAP alkaline phosphatase, 0.5 U TurboDNase and 20 U RNasin. Beads were washed once with 500 µL cold PNK wash buffer, twice with 900 µL of cold high-salt wash buffer, and twice with 900 µL cold PNK wash buffer. L3-IR-App adapter [208] ligation was performed overnight in the dark (16 ℃, 1100 rpm) in a mixture composed as outlined in Table 3.8. Beads were then washed once with 500 µL PNK wash buffer, twice with 900 µL of high-salt wash buffer, and twice with 900 µL PNK wash buffer.

Component	**Concentration/Amount**
L3-IR-App adapter	125 nM
T4 RNA ligase	30 U
Ribolock RNasin	20 U
T4 PNK	4 U
PEG8000	22.5%
DMSO	5%
10x Ligation buffer	1x

Table 3.8: L3-IR-App adapter ligation mix

IP and SMI samples were denatured in 1X NuPage LDS Sample Buffer with 100 mM DTT for 5 min at 70 ℃. IP samples were spun down for 2 min at 2500xg and eluate was transferred to a fresh tube. Elution was repeated once more and eluates were combined. Samples were separated on a 4–15% Mini-PROTEAN TGX Precast Protein Gel for 65 min at 180 V. Protein-RNA complexes were transferred onto an iBLOT2 nitrocellulose membrane using the Bio-Rad Trans-blot Turbo (1.5mm gel setting) and visualized on a LI-COR Odyssey Fc imaging system. The region corresponding to the RBP-eGFP band and above was cut (for both IP and SMI samples) and digested using 350 µg Proteinase K in 180 µL PK-SDS solution for 60 min at 50 ℃ (1100 rpm). RNA was purified by adding 1X volume of Phenol:Chloroform:Isoamyl Alcohol pH 6.6-6.9, incubating for 10 min at 37 ℃ (1100 rpm) and centrifugation at 16000xg for 5 min in

MaxTract tubes. RNA was cleaned using Zymo RNA Clean & Concentrator-5.

For SMI library preparation, SMI control samples were then treated with 5 U PNK, 0.5 U FastAP, and 20 U RNAsin in PNK buffer pH 6.5 for 40 min at 37℃ (1100 rpm). RNA was cleaned up with Dynabeads MyOne Silane. L3-IR-App adapter ligation was performed with 45 U T4 RNA ligase in 1X T4 RNA Ligase Reaction Buffer with 2% DMSO, 27% PEG8000, and 133 nM L3-IR-adapter for 75 min at room temperature, followed by MyOne bead purification. SMI samples were then treated with 25 U 5' Deadenylase and 15 U RecJf endonuclease in 1X New England Biolabs buffer 2 with 20 U RNAsin and 20% PEG8000 for 1 h at 30℃ and then 30 min at 37℃ (1100 rpm), followed by a MyONE clean-up. RNA from IP and SMI samples were reverse transcribed using Superscript IV reverse transcriptase and hydrolyzed by adding 1.25 µL of 1 M NaOH for 15 min at 85℃, before neutralization with 1.25 µL of 1 M HCl.

cDNA was purified using MyOne silane beads. L#clip2.0 adapters with barcodes for multiplexing [209] were ligated to cDNA by mixing 2 µL of 10 µM adapter with 5 µL of cDNA, adding 1 µL of DMSO, and incubating at 75℃ for 2 min before placing on ice. Then, ligation mix (45 U T4 RNA ligase in 1X RNA ligase buffer with 22.5% PEG8000) was added to the cDNA-bead solution and incubated overnight at 20℃ (1100 rpm). cDNA was cleaned up with MyONE beads before PCR amplification.

Pre-amplification was performed using 2X Phusion HF PCR Master mix with P5Solexa_s and P3Solexa_s primers for six cycles, followed by ProNex size-selective purification. Optimal qPCR cycles were determined by Real-Time qPCR, using EvaGreen, 2X Phusion HF PCR Master mix, and P5/P3 Solexa primers. Final PCR products were purified using two consecutive rounds of ProNex Size selection.

Sequencing Samples were quantified using Qubit DNA HS and library size was measured using High Sensitivity D1000 TapeStation. Each group of samples was pooled equimolarly and then mixed at the following proportions: 60% IP library pool, 40% SMI library pool. Sequencing was performed on a NextSeq 550 sequencer with a 75 cycle High-output kit v2.5.

3.7 Data analysis

General GO enrichment analyses were performed using clusterProfiler in R [210]. A p-value cut-off of 0.01 and a q-value cut-off of 0.05 were used and the Benjamini-Hochberg method was used to correct for multiple testing. Non-unique GO terms were collapsed using clusterProfiler's

simplify function.

Principal component analyses were performed using the base R package on log2 transformed values. Data was filtered to exclude NAs and normalised and batch corrected to match limma analysis processing where appropriate.

XL-MS analysis Protein identification and quantification were performed using Andromeda search engine implemented in MaxQuant [211]. Peptides were searched using the Human Uniprot database. False discovery rate (FDR) was set at 1% for both peptide and protein identification and 'match between runs' was turned on. Otherwise, default parameters were used. Filtered proteins found in the peptide list were further analysed using the software pLink 2.0 (v2.3.4) [212] to identify the cross-links with an integrated false discovery rate (FDR) of 5% at the spectrum level.

RNA sequencing The RNA-Seq reads were aligned to the Homo sapiens genome (GRCh38.110) downloaded via Ensembl using Hisat2 [213]. After the alignment, FeatureCount [214] was used to count reads that mapped to gene annotation files. Differential expression analysis was performed on sample groups using the R package DESeq2 [215]. DESeq2 estimates variance-mean dependence in data counts from high-throughput sequencing data and tests for differential expression based on a model using the negative binomial distribution. The External RNA Controls Consortium (ERCC) spike-ins were added to the sample before library preparation. The ERCC spike-in sets were used for the normalisation of gene expression value.

Protein-protein interaction Protein identification and quantification were performed using Andromeda search engine implemented in MaxQuant [211]. Peptides were searched using the Human Uniprot database with viral proteins added. False discovery rate (FDR) was set at 1% for both peptide and protein identification and 'match between runs' was turned on. Otherwise, default parameters were used.

Data analysis of ProteinGroups file from MaxQuant was performed in R. For DDX1-GFP / parental comparison and benzonase / no-benzonase, no normalisation was performed. For mock / SINV comparison, data was normalised using the 'vsn' package. For benzonase / no-benzonase and mock / SINV comparisons, data was filtered to include only proteins that reached the 0.01% FDR threshold in at least one DDX1-GFP v parental comparison. Rows

were filtered to remove any with $>$2 NA values in each condition under study. Minimum value imputation was performed for on/off changes (all NA values in one condition and $<$2 NA values in the other). Only values for replicates corresponding to non-NA values in the other condition were imputed. Fold-changes and p values were calculated using the limma package [216] and FDR was calculated from p values using the fdrtool package.

To generate interaction networks, protein IDs were imported into string-db [217] and interactions were filtered to include only those with a confidence score $>$0.4. Networks were plotted using Cytoscape.

iCLIP Raw FASTQ files were demultiplexed using the Je Suite [218] and adapters were trimmed using Cutadapt [219]. STAR was used to align reads to a concatenated human (GRCh38, ENSEMBL Release 106) and SINV (pT7-SVwt) genome in end-to-end alignment mode [220]. Only uniquely aligned reads were retained for downstream analysis. PCR duplicates were collapsed using unique molecular identifiers (UMIs) with the Je Suite. The crosslink truncation site for each read (-1 from the 5' start site of the read) was extracted using BEDTools [221].

Peak calling was performed with HTSeq-clip and the R/Bioconductor package, DEW-seq [222]. HTSeq-clip was used to generate a sliding window annotation of the human and SINV genome (50nt window, 20nt step size) and calculate the frequency of crosslink truncation sites within each window. DEW-Seq was then used to calculate the differential enrichment of each window relative to size-matched input control samples, with a cut-off of $\geq$2 log2 fold change and $\geq$0.01 adjusted p-value. Multiple hypothesis correction was performed using the Independent Hypothesis Weighting (IHW) method [223]. Overlapping windows were merged to form binding regions.

PCA was performed using DESeq2 [215]. Following size correction and variance stabilisation, the 1000 most variable sliding windows were selected and used for PCA plotting.

Binding site properties, including gene name, biotype, and gene feature, were extracted from the ENSEMBL genome annotation using the GenomicRanges package. Metagene analyses were performed using functions from the cliProfiler package.

Sequences for motif prediction and secondary structure prediction were defined for each binding site as a 50-nucleotide region, centered on the peak in BigWig signal. For motif prediction, a gene and gene region-matched background sequence was extracted for each

binding site to allow for differential enrichment analysis. Enrichment analysis was performed using STREME from the MEME suite [224]. Universalmotif was used for motif processing and motifStack was used to cluster and plot motifs.

SINV genome coverage in reads per million was calculated using BEDTools [221]. Percent of total signal was then calculated at each position in the IP and SMI samples. SMI signal was subtracted from IP signal for plotting.

4 Revealing the composition and *in situ* properties of the tRNA ligase complex

4.1 Introduction

Proteins within the cellular environment form intricate assemblies that underpin a wide range of biological processes, driving the regulation of pathways and networks essential for cellular function. Our understanding of life at the molecular level relies heavily on elucidating the structures and mechanisms of macromolecules and their interactions. This field has advanced through structural biology techniques such as X-ray crystallography, nuclear magnetic resonance (NMR), and electron microscopy (EM). These methods have provided atomic-scale insights into protein architecture and function. Despite their power, traditional structural techniques often face limitations. Strict requirements for high-purity samples, difficulties with unstable proteins or complexes, mass constraints, and challenges in achieving high resolution can limit their applicability [225]. Integrative approaches have emerged as valuable complements to address these issues. These methods include homology modelling, site-directed mutagenesis, biochemical interaction assays, co-immunoprecipitation (Co-IP), affinity purification-mass spectrometry (AP-MS), and cross-linking mass spectrometry (XL-MS) [226]. Collectively, they enhance our capacity to investigate dynamic and complex protein systems.

XL-MS, first utilised in pioneering studies in the early 2000s, has revolutionised the study of protein conformations and interactions [227–229]. This technique uses cross-linking reagents to covalently bond amino acid residues in close proximity, followed by mass spectrometry. By imposing spatial constraints based on the cross-linkers spacer-arm length, XL-MS provides invaluable structural data about proteins, complexes, and interaction networks. Advances in MS-compatible cross-linkers [230], instrumentation [231], and computational tools have expanded the biological applications of XL-MS, making it a versatile tool for exploring large protein assemblies and conformational states [226].

The rise of computational tools, particularly AlphaFold2 (AF2) [232] and its successors, AlphaFold3 (AF3) [233], AlphaFold-Multimer [234, 235], has further transformed structural biology. These algorithms leverage deep learning to predict protein structures with remarkable accuracy, even for proteins lacking homologous templates. AlphaFold2 has enabled the modelling of over 200 million protein sequences from UniProt, providing structural insights

at an unprecedented scale [236].

However, each technique has inherent limitations. For example, AlphaFold excels at predicting ordered protein domains but struggles with flexible or disordered regions [237, 238]. XL-MS depends on the characteristics of the cross-linker used, the depth of biological coverage in the MS, and the ability to capture interaction stoichiometry accurately [239]. Similarly, Co-IP relies heavily on antibody specificity and may overlook transient or weak interactions. Integrating these diverse techniques can overcome their individual limitations, offering a more comprehensive and accurate representation of molecular events. For example, the combination of AF2 and XL-MS has been employed to distinguish the open and closed structures of luciferase and glutamine-binding periplasmic protein [240]. Furthermore, novel protein-protein interactions were discovered in mitochondria through the use of XL-MS and subsequently examined using AF2 structural predictions [241]. The 3-dimensional architecture of larger protein complexes is more challenging; however, Khan et al. (2022) were able to capture the 11-protein multi-tRNA synthetase complex using XL-MS, and the resulting crosslinks were mapped onto AF2-predicted models for structural context [242]. In both studies, AF2 predictions provided a visual framework for interpreting crosslinking data and highlighted possible limitations of the models in capturing dynamic or flexible regions.

Proteins operate within complex assemblies, such as the tRNA-LC (tRNA ligase complex), a highly stable 200 kDa protein complex established in 2011 [197]. This complex consists of the RTCB ligase protein, the DEAD-box helicase DDX1, the RNA transcription, translation, and transport factor (RTRAF/CGI99), FAM98B, and Ashwin. Over the last decade, investigations into the tRNA-LC have clarified its composition and stability. Co-IP studies initially highlighted RTCB's role in forming a stable complex with these proteins in cells, while knockdown experiments revealed that loss of DDX1, FAM98B, or CGI99 destabilises the assembly [197]. *In vitro* XL-MS with disuccinimidyl suberate (DSS) cross-linking elucidated the interaction interfaces among complex members in both cytoplasmic and nuclear forms of the tRNA-LC [243]. The cytoplasmic complex includes CGI99, FAM98B, DDX1, and RTCB, while the nuclear form adds ASW to this core. This study identified the existence of a stable sub-complex which excluded RTCB. However, the cellular context is absent and consequently omits the capture of novel interactions or environmental factors that may drive complex assembly. Partial structural resolutions have been achieved for several components, including RTCB [243, 244], CGI99 [243] and DDX1 [245]. Unfortunately, the complete complex resolution remains elusive.

Furthermore, recent studies have highlighted the dynamic conformational changes within RTCB during Archease-mediated recycling, showing that Archease promotes the formation of a covalent RTCB-GMP intermediate through GTP and metal ion coordination [244]. The dynamic modelling in RTCB indicates a potential dynamic conformation of the larger complex *in cellulo*.

This chapter focuses on elucidating the tRNA-LC's structural dynamics *in situ* using an integrative approach. First, I performed protein-specific co-IP of each individual component of the complex. Systematic knockdown experiments with DDX1, RTCB, and CGI99 further characterised the proteins' interdependencies and their roles in maintaining a stable tRNA-LC. Finally, I employed XL-MS coupled with AF3 to map the structural interfaces and hierarchical interactions within the tRNA-LC, shedding light on its tightly regulated assembly.

4.2 Results

4.2.1 The tRNA-LC is formed by tightly interacting proteins with DDX1 as one of the core components

tRNA-LC has largely been studied *in vitro*, highlighting a need to elucidate the properties of the complex in a cellular model. In order to characterise the individual proteins that compose the tRNA-LC, the generation of tagged proteins that can positively and consistently be captured was necessary. I generated stable cell lines with inducible expression for each member of the tRNA-LC: DDX1, RTCB, FAM98A, FAM98B, CGI99 and Ashwin. To simplify detection and enable biochemical characterisation, they were fused to the enhanced green fluorescent protein (eGFP). The linker between the protein and the tag consists of Glycine (Gly) and Serine (Ser) rich amino acids for independent folding of the protein and tag. To generate stable cell lines, I utilised a HEK293 Flp-In T-Rex cell line, which contains a single integrated *flippase* (Flp) Recombination Target (FRT)/lacZ-Zeocin construct at a transcriptionally active genomic locus, with a Tetracycline (Tet) repressor under the cytomegalovirus (CMV) promoter. I co-transfected the cloned pcDNA5/FRT plasmids containing the tagged tRNA-LC proteins with the pOG44 plasmid containing the Flp recombinase into the parental cell lines. Together, these two plasmids allow the insertion of the gene of interest at the single FRT site, replacing the parental zeocin resistance with hygromycin resistance for cell selection. Selected cells are expected to be isogenic because of the single integration site, resulting in homogenous levels of protein expression. The obtained cell lines were tested using doxycycline, which is similar to tetracycline in terms of tetR binding but has longer stability in cells. Western blotting against GFP of the induced cell lines indicated that all fusion proteins were expressed at their expected molecular weights (Figure 4.1). The expression was lower for FAM98B than for the other proteins, possibly due to lower protein stability.

Different protein-protein interaction studies have consistently reported that DDX1, RTCB, CGI99, FAM98B, and ASW interact, forming the tRNA-LC [197, 198, 243, 246]. The presence of a GFP tag in my constructs enabled me to perform immunoprecipitations (IPs) with very high specificity and affinity using the GFP-Trap, allowing me to further characterise the native interactions of each protein *in cellulo*.

To preserve *native* interactions, lysis and IPs were performed in buffers with physiological salt concentrations (150mM). To assess the quality of the IP, I analysed input (whole cell

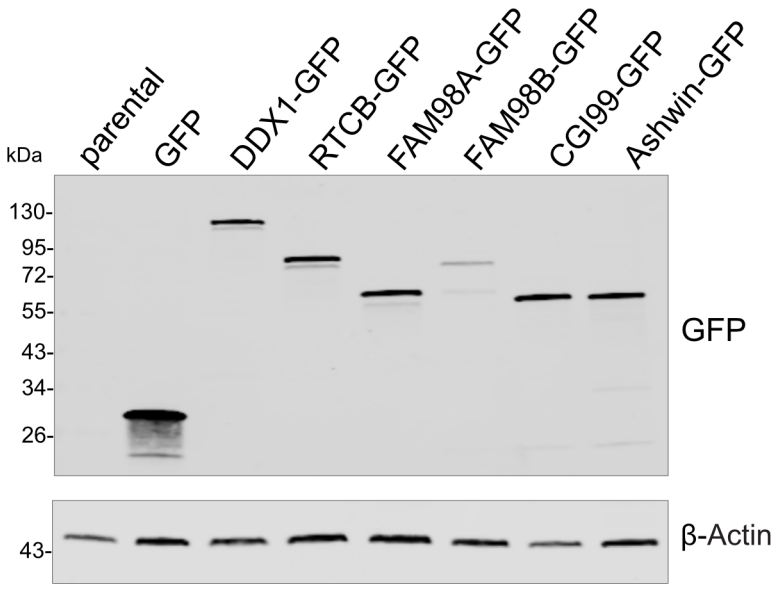

Figure 4.1: **tRNA-LC inducible expression cell lines**

Western blot of HEK293 FITR inducible expression cell lines after 24h of doxycycline induction. In order: parental, GFP only, DDX1-GFP, RTCB-GFP, FAM98A-GFP, FAM98B-GFP, CGI99-GFP and Ashwin-GFP cells.

lysates) and eluates using the standard protein staining method, silver staining (Figure 4.2A). The banding pattern indicated enrichment of a predominant polypeptide at the expected molecular weight of the targeted protein. Interestingly, the protein banding pattern showed consistency across all of the proteins except for FAM98A. Moreover, the molecular sizes of these bands match those of the known components of the tRNA-LC. These results suggest that the IP enriched the eGFP-tagged protein, along with the rest of the tRNA-LC components. To verify the specific enrichment of tRNA-LC proteins, I performed a western blot on the inputs and eluates (Figure4.2B). DDX1, RTCB and FAM98B were highly enriched in all IPs with the exception of the FAM98A IP. In addition, DDX1-GFP, RTCB-GFP and CGI-99-GFP enriched for all tested tRNA-LC proteins (DDX1, RTCB, FAM98A, FAM98B and CGI99) detected by western blot. Meanwhile, IPs performed in cell lines erexpressing FAM98A, FAM98B, and ASW captured fewer tRNA-LC partners. This suggests that DDX1, RTCB, and CGI99 are core tRNA-LC components.

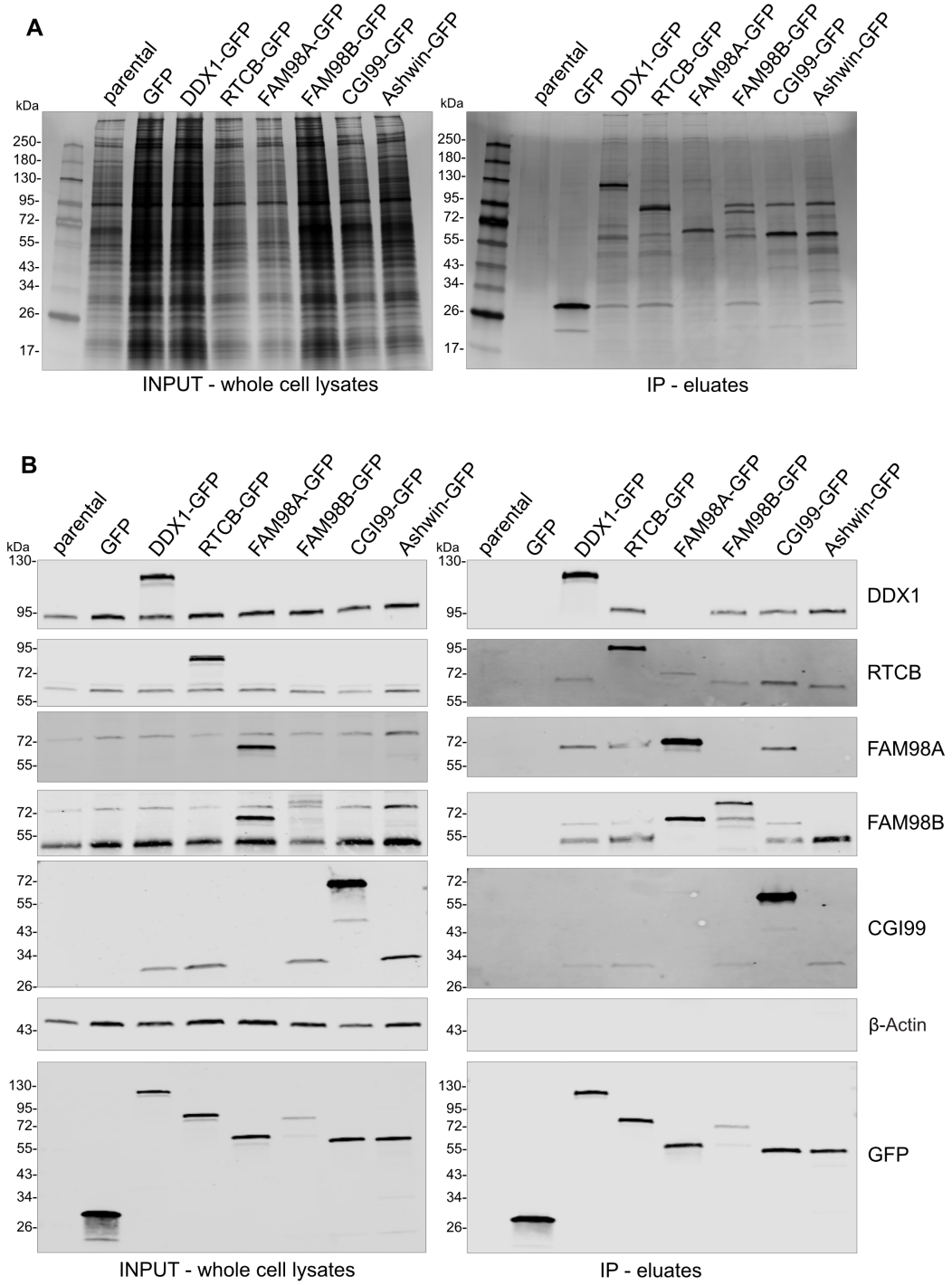

Figure 4.2: **Proteins of the tRNA ligase complex**

A HEK293 FITR inducible expression cell lines of parental, GFP only, DDX1-GFP, RTCB-GFP, FAM98A-GFP, FAM98B-GFP, CGI99-GFP and Ashwin-GFP cells induced for 24h with doxycycline and harvested for IP. Silver stain of whole cell inputs (left) and IP eluate (right) washed in mild buffer (150mM NaCl). **B** Western blots of whole cell inputs (left) and IP eluates (right). Antibodies for tRNA-LC proteins: DDX1, RTCB, FAM98A, FAM98B, CGI99, β-actin and GFP. Note: the FAM98B antibody binds to FAM98A at lower specificity.

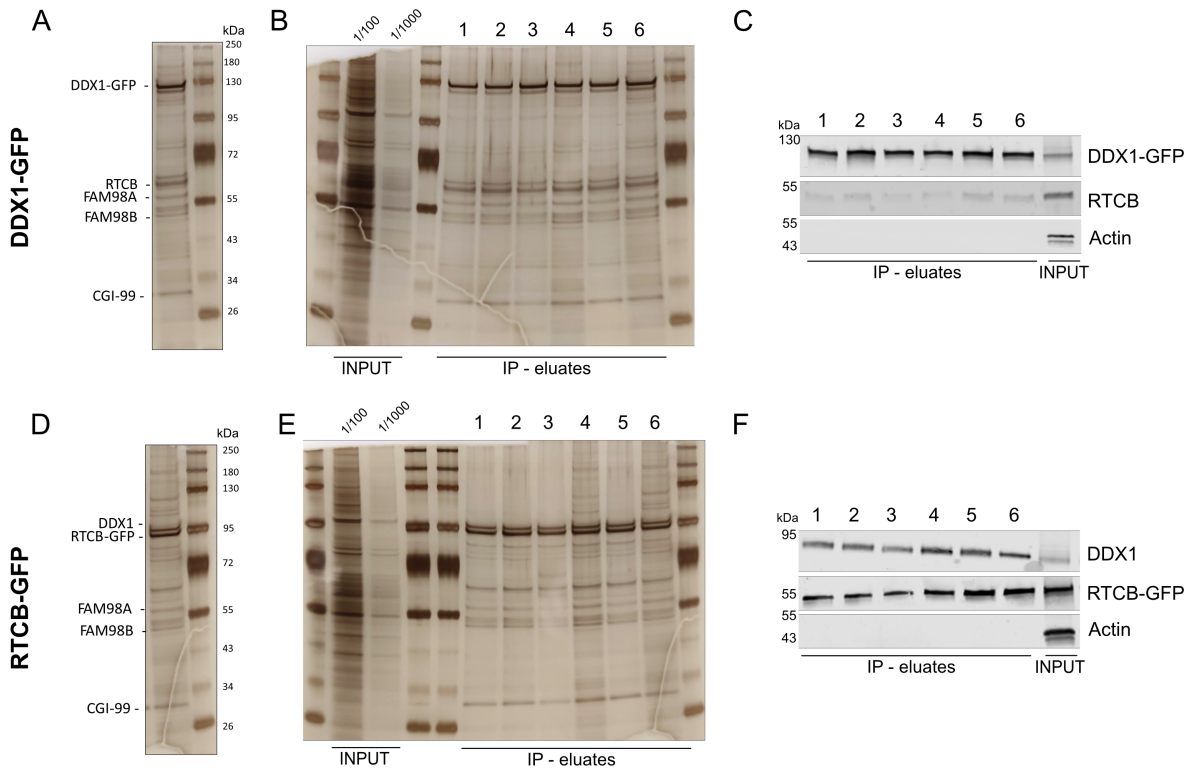

Figure 4.3: **DDX1 and RTCB denaturation by stringent wash buffers**

A and D HEK293 FITR DDX1-GFP and RTCB-GFP cells induced for 24h with doxycycline and harvested for IP and washed in mild buffer (150 mM NaCl). Silver stain of eluates with protein ladder. Protein bands enriched during IP labelled with expected protein sizes of tRNA-LC as indicated in [197]. **B-C** HEK293 FITR DDX1-GFP and **E-F** RTCB-GFP cells induced for 24h with doxycycline and harvested for IP. INPUT samples refer to samples prior IP and ELUATES are samples after IP. Six different buffers were used for washes during IP as follows: *(1)* 2M NaCl, 0.1% SDS. *(2)* 1M NaCl, 0.2% SDS. *(3)* 150mM NaCl, 4M Urea. *(4)* 150mM NaCl, 8M DTT. *(5)* 1M NaCl, 0.1% SDS (RIPA buffer). *(6)* 150mM NaCl (Mild wash buffer). **B** Silver stain of DDX1-GFP samples in all 6 IP wash buffers. **C** Western blot of DDX1-GFP of IP eluates and INPUT sample, with antibody against GFP (detecting DDX1-GFP), RTCB and β-actin. **E** Silver stain of RTCB-GFP samples in all 6 IP wash buffers. **F** Western blot of RTCB-GFP of IP eluates and INPUT sample, with antibody against GFP (detecting RTCB-GFP), DDX1 and β-actin.

DDX1 and RTCB, together with CGI-99, are considered core members of the tRNA-LC based on *in vitro* assays [197, 198, 243]. Under physiological salt concentrations, all members of the tRNA-LC co-precipitated (Figure 4.2 and Figure 4.3A and D). To assess the strength of the interactions between the components of the tRNA-LC, I next tested whether these interactions were preserved under progressively more stringent wash conditions. Focusing on DDX1 and RTCB as central components, I used six different buffers with a variety of reagents that impair protein-protein interactions, including denaturing agents (Figure 4.3 B-C and E-F). The concentration of the reagents was selected considering the maximal tolerance of GFP-

Trap agarose. Even under the most stringent wash conditions, including high sodium chloride (NaCl, ionic strength), high SDS (chaotropic detergent), high DTT (reducing agent), and urea (denaturing agent), the bands at the molecular weight of the tRNA-LC proteins were still visible. These results indicated that the tRNA-LC is a very stable and remarkably sturdy complex.

4.2.2 tRNA-LC inter-protein dependency

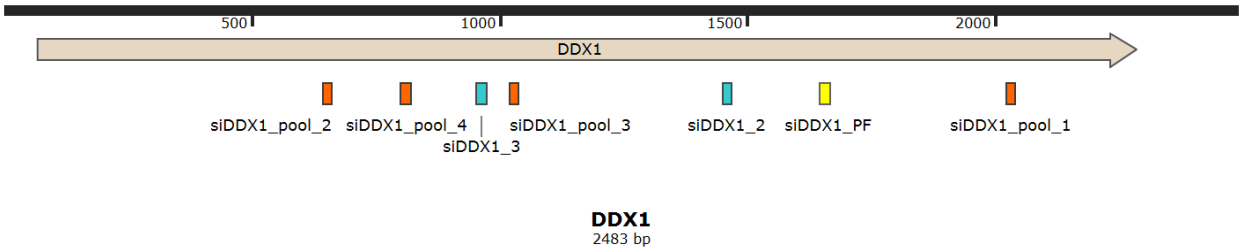

Figure 4.4: **siRNA targeting DDX1 schematic**

Schematic of DDX1 mRNA with indicated coding sequence (amber arrow). SiDDX1 target sequences labelled across gene. In yellow is the siDDX1 from the Proudfoot lab that is established both as a si/shRNA. In blue are single siDDX1 targets. In orange, are 4 siDDX1 targets within a single pool.

The co-precipitation of the tRNA-LC proteins indicated the strong interaction of its components. Original characterisation of the complex in 2011 showed in Hela cells that the knockdown of RTCB and CGI99 affected secondary proteins of the complex [197]. This suggested that the stability of the components of the tRNA-LC is tied to their assembly into a higher-order molecular machinery. I thus sought to characterise the tRNA-LC protein dependency in HEK293 cells using a siRNA knockdown system. I first targeted DDX1 only, using three different single gene targets, depicted in Figure 4.4. The first siRNA is a well-established target sequence in an shRNA system previously used in the Castello and Proudfoot labs [113, 247]. The second and third siRNAs have been employed in various studies aiming to characterise DDX1 in different contexts [193, 248]. The knockdown of all 3 siRNA led to a similar reduction of the other proteins in the tRNA-LC, of which two were significant (siDDX1_1 and siDDX1_3). This indicated that the effects of DDX1 knockdown were robust and reproducible. RTCB and CGI99 levels were also tested, resulting in a milder reduction of the other proteins of the complex, specifically CGI99 by siDDX1_1 and siDDX1_2 (Figure 4.5).

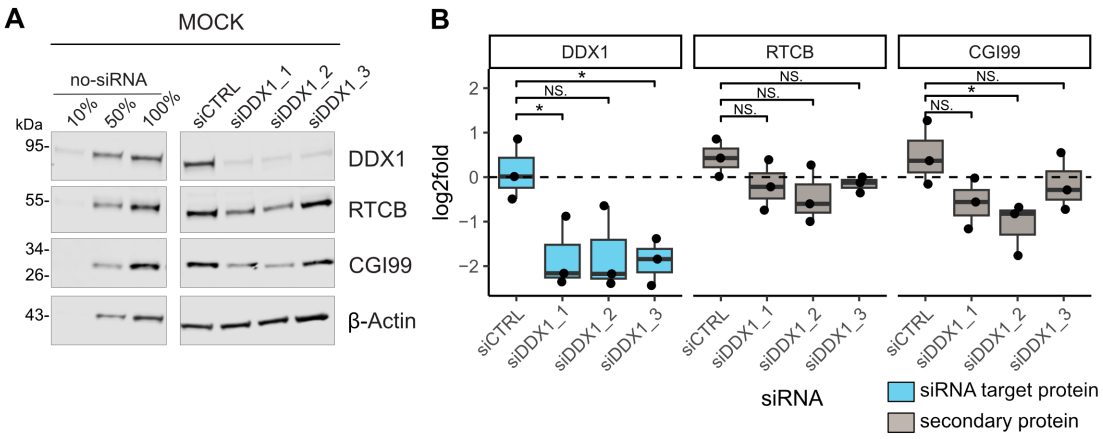

Figure 4.5: **Knockdown of DDX1 by single target siRNA**

A Western blot images of single target siDDX1 after 48h transfection. No-siRNA refers to mock transfected cells (water only). A dilution of 10% and 50% of the lysates was performed to to infer the quantitative accuracy of the assay. Antibodies for DDX1, RTCB, CGI99 and β-actin were done to study the knockdown on DDX1 and secondary proteins of the tRNA-LC. **B** Relative quantification of western blot protein signals from 3 independent biological replicates. Normalised to β-actin signal levels, followed by a log2 fold change quantification against the signal of the no-siRNA control sample. Statistical significance was tested using two-tailed T-test (P $\leq$0.05 *, ns: not-significant).

The knockdown of DDX1 by the single target siRNA was not permissive to other components of the tRNA-LC as previously observed in HeLa cell [197]. To corroborate whether the observed effect was due to technical differences in siRNA or to biological differences in cell lines, I next used the previously established siRNA pools with higher knockdown efficiency (Dhermacon [197]), focusing on the three key tRNA-LC proteins: DDX1, CGI99 and RTCB. A pool of four siRNAs per gene increases knockdown efficacy. The lower concentration of each sequence target within the pool reduces the potential off-target effect and decreases the likelihood of cell toxicity.

I determined the effect of each siRNA pool by western blotting and quantified the signal of the protein bands (Figure 4.6), revealing a pronounced reduction in the level of the targeted protein (indicated in blue in Figure 4.6B). Knockdown of the target protein had a secondary effect on the other proteins in the tRNA-LC. The knockdown of DDX1 induced a significant decrease in CGI99 protein levels. The increased significance in CGI99 knockdown with the siDDX1 pool, as oppose to the single siRNA target, suggests that effective depletion of this protein is required to cause effects in the other proteins of the complex. With the siCGI99 and particularly siRTCB pools, I observed a significant decrease in DDX1, RTCB and CGI99. This suggests that the stability of tRNA-LC proteins is dictated by the assembly of the complex. I

hypothesise that the absence of tRNA-LC constituents necessary for complex assembly results in protein degradation. Moreover, CGI99 and RTCB had a stronger overall effect on complex stability, suggesting a more central role in the complex organisation.

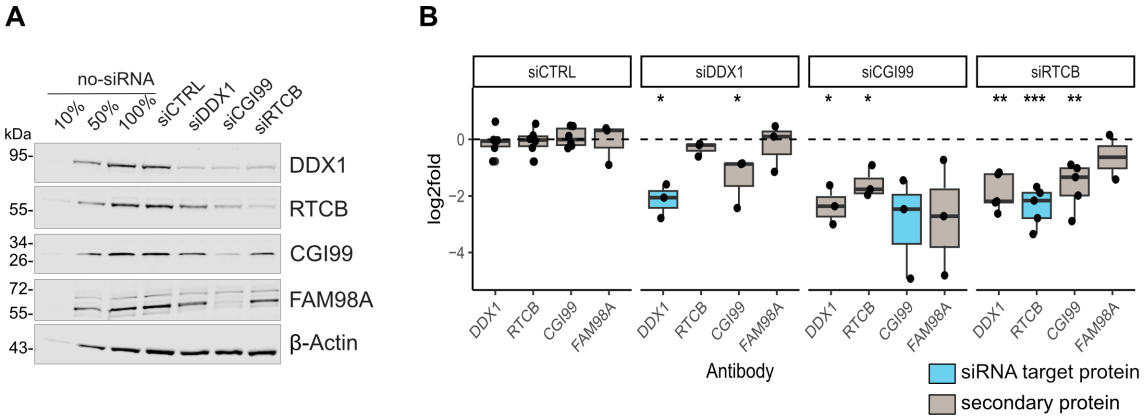

Figure 4.6: **Knockdown of tRNA-LC proteins by an siRNA pool**

A Western blot images of siRNA pools targeting DDX1, CGI99 and RTCB after 48h transfection. No-siRNA refers to mock transfected cells (water only) and siCTRL refers to siRNA scramble pool. Antibodies for DDX1, RTCB, CGI99 and FAM98A and β-Actin were used to study the knockdown of secondary proteins of the tRNA-LC. **B** Relative quantification of western blot protein signals from 3 independent biological replicates. Normalised to β-actin signal levels, followed by a log2 fold change quantification against the signal of the no-siRNA control samples. Statistical significance was tested using two-tailed T-test comparing to siCTRL sample (P $\leq$0.001 ***, P $\leq$0.01 **, P $\leq$0.05 *).

4.2.3 Elucidating the tRNA-LC interfaces *in cellulo* using disuccinimidyl suberate (DSS) and mass spectrometry reveals a CGI99-centric complex

The tRNA-LC has been analysed by XL-MS *in vitro*, with minimal components and lacking other cellular proteins that may engage with it in cells [243]. Although this has provided valuable insight into the direct interaction dynamics of the established complex components and their structural hierarchy, it does not account for novel protein interactors not previously characterised. Moreover, the lack of full-length proteins in some instances may lead to the loss of key interactions *in cellulo*. Here, I sought to elucidate the protein-protein interactions of the tRNA-LC, focusing on the core component DDX1, and aiming to obtain all potential protein interaction partners, binding interfaces, complex topology, and dynamics. In collaboration with Dr Marko Noerenberg, DDX1-GFP was expressed and crosslinked to its partners in cells using DSS, followed by GFP-trap IP with stringent washes. In parallel, the parental cell line with an empty FITR cassette was used as a background control. In collaboration with Prof Shabaz Mohammed and Dr Yana Demyanenko at the Rosalind Franklin Institute, the eluates were

analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS).

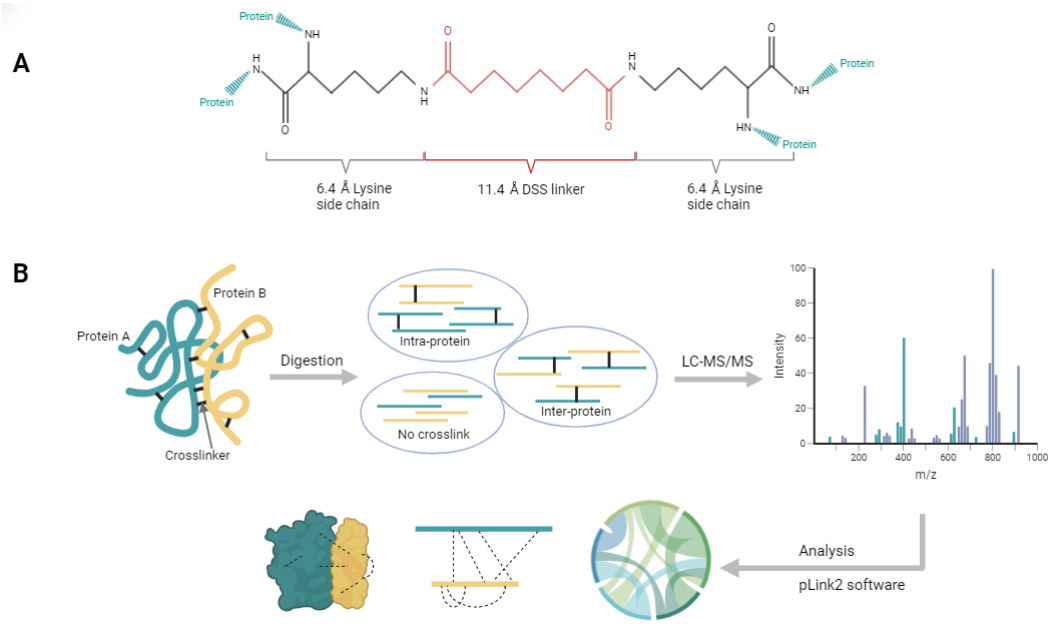

Figure 4.7: **Schematic of XL-MS/MS analysis pipeline**

A Schematic DSS crosslinker with lysine side chains. **B** Schematic of XL-MS/MS experimental and analysis pipeline.

DSS is a molecular crosslinker and cell-permeable, allowing for *in vivo* crosslinking that allows for the study of intracellular interactions. DSS reacts specifically with lysines on proteins, forming covalent bonds with a reaching distance of 26.2Å (distance outlined in schematic Figure 4.7A). Importantly, the DSS-mediated crosslinks are non-cleavable and are stable under the conditions used in the MS analysis. This allows for the search of hybrid peptides (sequences mapping to 2 proteins) plus the crosslinker to identify the interfaces. The workflow of this experiment is outlined in Figure 4.7B. Crosslinking peptides can occur between proteins (inter-protein) and are indicative of protein-protein interaction dynamics. Intra-protein crosslinks occur between adjacent lysines within the same protein and can be used to aid protein folding by implementing distance constraints to AI structural models. They can also reflect multimeric states of a given protein, and in this case, they would be inter-molecular even if they are mapping to a single protein.

MS data was analysed using the software pLink 2.0 (v2.3.4) [212] to identify the cross-links with an integrated false discovery rate (FDR) of 5% at the spectrum level. I identified 124 cross-linked peptides, constituting 1.9% of the obtained spectra. Despite the substoichiometric nature of crosslinked peptides, I robustly pinpointed several interactions with a minimum of

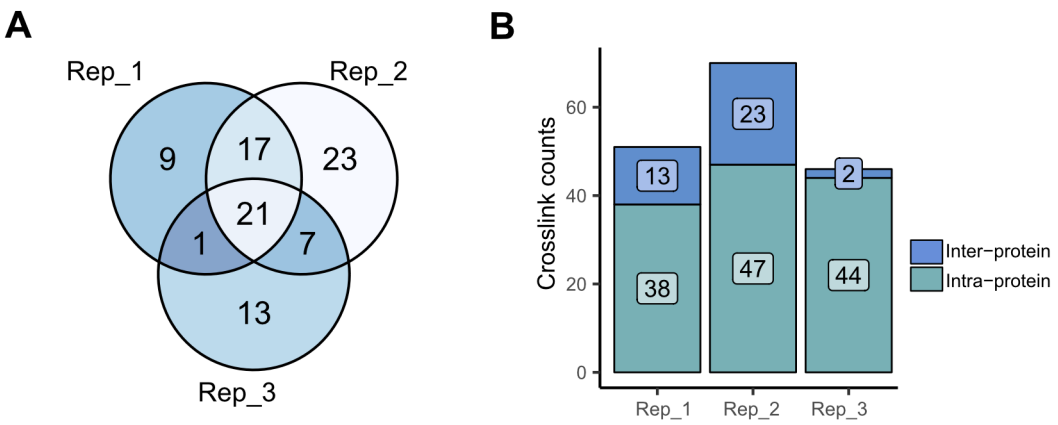

Figure 4.8: **Crosslink peptides identified across replicates**

(**A**) Total count of unique crosslinks identified in each replicate visualised by a venn diagram to show unique/common found crosslinks. (**B**) Stacked barplot of total crosslinks identified in each replicate split across counts of intra-protein sites and inter-protein sites.

two independent replicates. In Figure 4.8A, I illustrate the total obtained crosslinks in each replicate and their interconnections. Notably, some crosslinks were exclusive to specific biological replicates. This is expected as the identification of crosslinking peptides is extremely challenging, and it is suggested that a substantial random sampling between replicates is performed due to the lack of depth. Identification of a crosslink site by two replicates is within the standards in the field and is commonly found as a cutoff across published XL-MS datasets [240, 243, 249]. IP stringency was validated by the lack of crosslinked peptides in the control samples. Naturally, intra-protein crosslinks are more readily captured, as reflected by their higher representation across the different replicates (Figure4.8B). However, the more challenging inter-protein crosslinks were also present.

Inter-protein cross-linked peptide pairs provide evidence of direct, bona fide protein interactions. The unique capabilities of XL-MS allow for the capture of not only stable but also transient protein interactions, although the abundance of cross-linked peptides from transient interactions is typically low. Given the qualitative nature of the crosslinking data in this analysis, any single or multiple crosslinking occurrences that exceed the statistical threshold (5% FDR) were considered valid indicators of direct interactions. To maximise coverage and insight, I pooled crosslinked sites from all three biological replicates and represented them in the chord diagram (Figure 4.9).

Inter-protein sites identified amongst tRNA-LC proteins place CGI99 at the epicentre of the complex, with all components of the tRNA-LC solely interacting with CGI99. CGI99 is a small

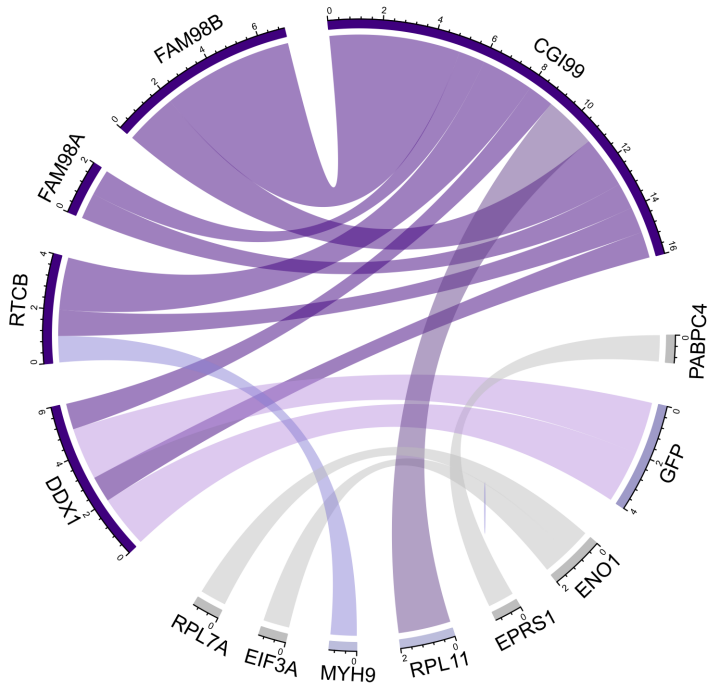

Figure 4.9: **Inter-protein crosslinks identify CGI99 centric crosslinking**

Chord diagram showing inter-protein crosslinks identified in DSS XL-MS data. In dark purple are tRNA-LC proteins. Chord width indicates number of crosslinks within two protein sequences, and the protein sizes indicate the total number of identified inter-protein crosslinks.

protein, hence the number of cross-links identified is not biased by a large molecular weight. The central nature of CGI99's is also supported by its importance in tRNA-LC complex stability previously identified in the siCGI99 western blots. Unsurprisingly, crosslinks between GFP and DDX1 were also identified. The DDX1-GFP protein, although fused through a flexible arm described previously, folds as two independent but linked proteins. The crosslinking identified between GFP and DDX1, although categorised as inter-protein crosslinking, must be considered intra-protein in this case.

Strikingly, two proteins outside the canonical tRNA-LC, MYH9 and RPL11, were found to crosslink with RTCB and CGI99, respectively. This observation suggests potential novel tRNA-LC partners and indicates possible functional engagement with other large complexes, such as the ribosome and intermediary filament networks. In addition to these direct interactions, other proteins were enriched but did not crosslink directly to tRNA-LC members. These proteins were detected in the DDX1-GFP IP, suggesting their indirect association with the tRNA-LC. Among these enriched proteins, EIF3A and RPL7A, both critical components of the translation machinery, formed direct crosslinks with ENO1. ENO1, recently described as a moonlighting

RNA-binding protein (RBP) involved in gene expression regulation [250], may play a role in translational control based on its interactions observed here. These findings hint at a broader functional network involving the tRNA-LC.

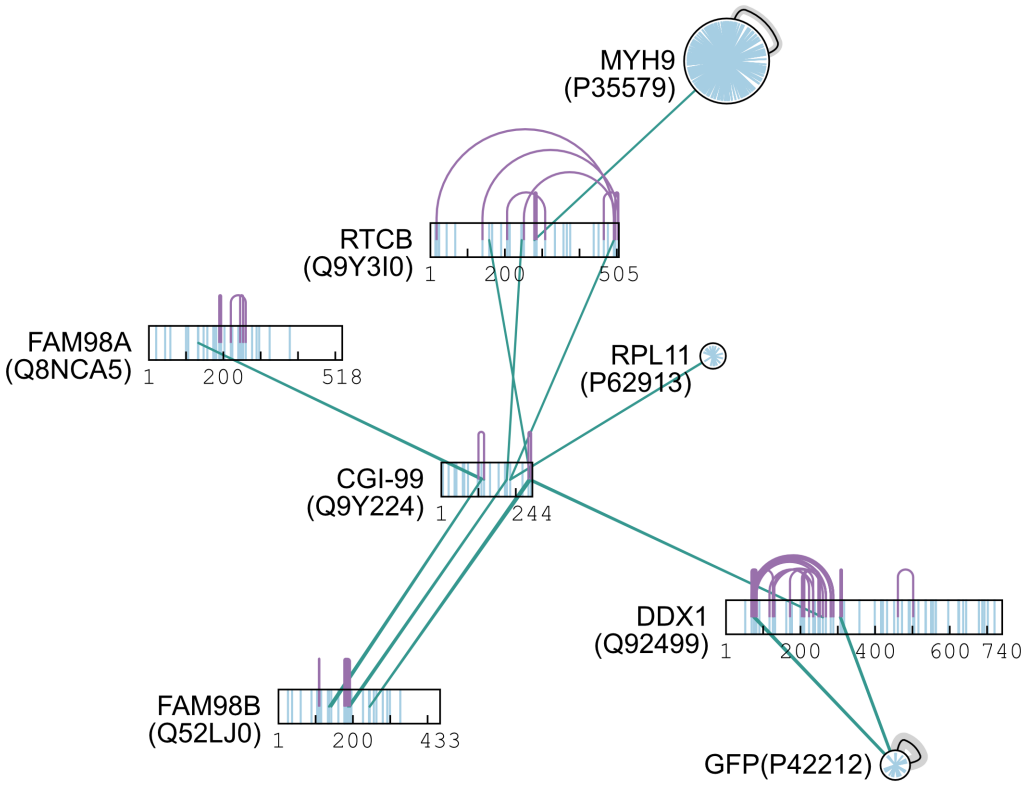

Figure 4.10: **tRNA-LC XL-MS intra- and inter- rotein crosslinking sites**

Intra- and inter-protein cross-links between tRNA-LC proteins, as well as MYH9, RPL11 and GFP identified in XL-MS. TRNA-LC protein lengths are proportional to each other. MYH9, RPL11 and GFP are in circular format. Lysine residues are marked in light blue inside each linear representation of the protein to indicate a potential crosslinking site. Identified crosslinks are labelled: Intra-protein crosslinks in purple, inter-protein crosslinks in green.

By studying the crosslink locations across proteins, it is possible to reveal the protein folding structures and protein-protein interfaces, including those where multiple proteins bind at the same site. In Figure 4.10, I mapped the position of the identified crosslink sites. Theoretically, all protein lysines hold the potential of being crosslinked; however, upon mapping lysine residues across the proteins, very few are captured as crosslinking two peptides (either with the same protein: intra-protein, or different protein: inter-protein). The lack of crosslinking of all lysines can be due to the inability of the crosslinker to access the lysines due to protein structural restriction and space, or the limited data capture. Interestingly, among the inter-protein crosslinking sites, I observed overlapping crosslink sites. Overlapping sites indicate that more than one protein interacts with a given interface. For example, the homologues

FAM98A and FAM98B crosslink exactly at the same amino acids of CGI99 (108aa and 187aa), suggesting mutually exclusive binding (Figure 4.10). DDX1 and FAM98B also compete for the same site on CGI99 (234aa). Furthermore, three proteins (DDX1, RTCB and FAM98B) are identified as crosslinking the same 241aa on CGI99. These crossover interactions can indicate the tRNA-LC flexible reorganisation to allow for different functions through sub-complexes with slightly different compositions. Another indication of alternative complex structures is portrayed by the dual crosslinking of RTCB and RPL11 on CGI99 (at 185aa). RPL11 is a novel direct interactor of CGI99 and could reflect a role for CGI99 in the absence of RTCB.

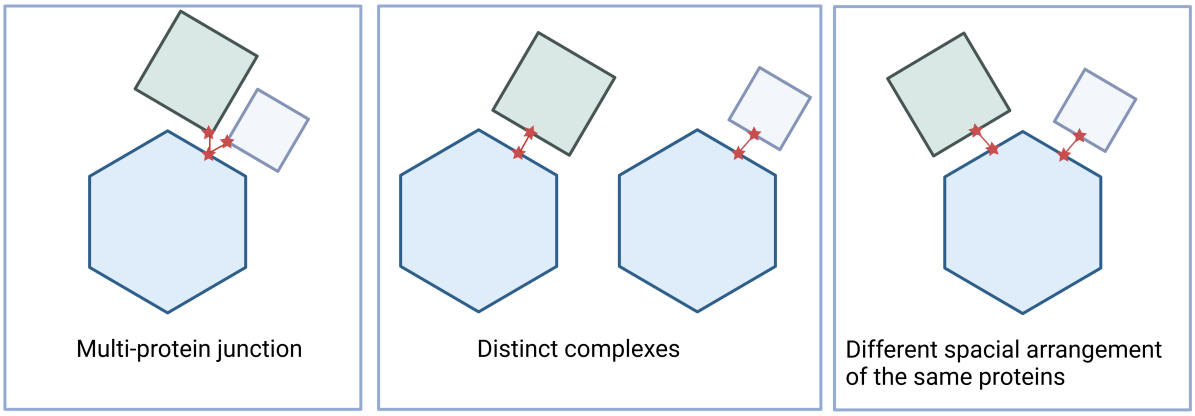

Figure 4.11: **Schematic of multi-protein crosslink interfaces**

Overlapping crosslinking sites can be a product of potentially three reasons outlined in the schematic of Figure 4.11. Firstly, crosslinking can occur at the interface of multiple close amino acids of more than two proteins, generating a multi-protein junction. Secondly, overlapping crosslinks can signify competitive binding and the existence of distinctly different complexes. Finally, conformational changes in the complex assembly can signify that the same protein combination is interacting at different locations. A combination of all three possibilities outlined can also occur. In the case of the tRNA-LC, remodelling is known to occur during RTCB recycling by Archease [244], or by DDX1 during RNA unwinding [251]. The tRNA-LC has also been previously observed to change interaction dynamics in the presence or absence of ASW *in vitro* [243].

4.2.4 XL-MS data correlates with Alphafold3 prediction

Recent advancements in machine learning-based protein folding prediction tools have significantly improved our ability to predict protein structures and the dynamics of complex

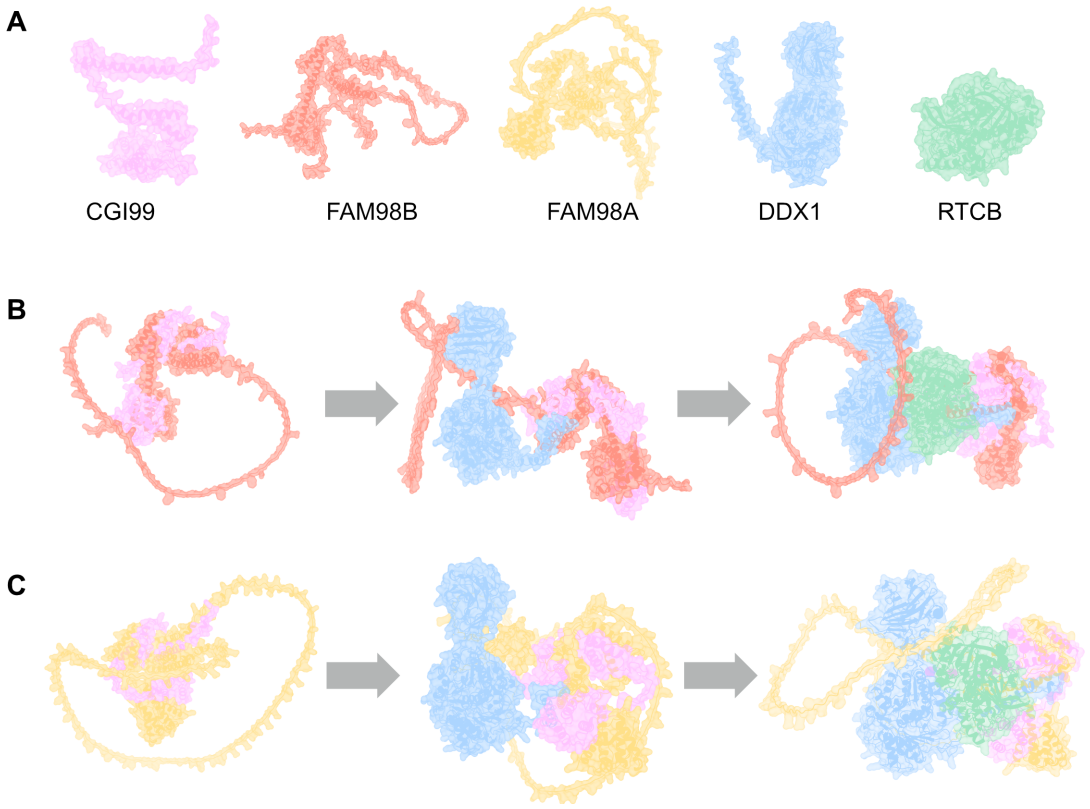

Figure 4.12: **Alphafold3 predicts tRNA-LC proteins and complex assembly**

A Alphafold3 prediction of tRNA-LC proteins: CGI99, FAM98B, FAM98A, DDX1 and RTCB. **B and C** Step-wise complex building of tRNA-LC with FAM98B (B) or FAM98A (C). From left to right: CGI99 and FAM98B/A interaction prediction, DDX1 is added, and finally RTCB is added.

assembly. RTCB, DDX1, and CGI99 have fully or partially resolved structures, which enhances the reliability of prediction softwares [243–245]. In collaboration with Rozeena Arif, I utilised AlphaFold3 (AF3) software [233] to predict the individual structures of the tRNA-LC proteins and subsequently model their potential complex assembly. By incorporating the PDB-deposited resolved structures of RTCB, DDX1, and CGI99, I achieved high accuracy in our protein structure predictions (Figure 4.12A). Reference sequences (from NCBI's RefSeq) were employed to predict the structures of FAM98A and FAM98B. Both proteins exhibited a high prevalence of disordered regions and showed overall structural resemblance for their globular domains, consistent with their sequence similarity.

Complex assembly was conducted using the informed structural dynamics outlined in Kroupova et al. (2021) [243]. Two simulations were performed, incorporating either FAM98A or FAM98B as part of the complex (Figure 4.12B and C). Initially, the interaction dynamics between CGI99 and FAM98A or FAM98B were analysed. In the AF3 prediction of the complex, CGI99 and FAM98A/B form coiled coils in their C-terminal domains, which associate to create

a helical bundle. Additionally, the two proteins form a heterodimer through an independent interface involving their N-terminal domains. Subsequently, a new simulation was conducted to include DDX1. The incorporation of DDX1 into the two-protein complex highlighted the α-helical C-terminal regions of DDX1, FAM98A/B, and CGI99 as key points of interaction. Finally, a simulation of the four-protein structure, including RTCB, was performed. RTCB is predicted to occupy a central position within the complex, as AF3 suggests that it inserts itself into the gap formed between the CGI99/FAM98A/B bundle and the connective C domain of DDX1. Notably, both tRNA-LC structures with either FAM98 proteins displayed similar interaction domains. A simulation, including both FAM98 proteins within the full complex structure (not shown here), indicated identical location and interaction interfaces with only disordered region differences outside of the core complex. The FAM98 proteins' similar structure and overlapping assembly within the tRNA-LC indicated the likely competitive binding within the complex and the formation of distinct structures.

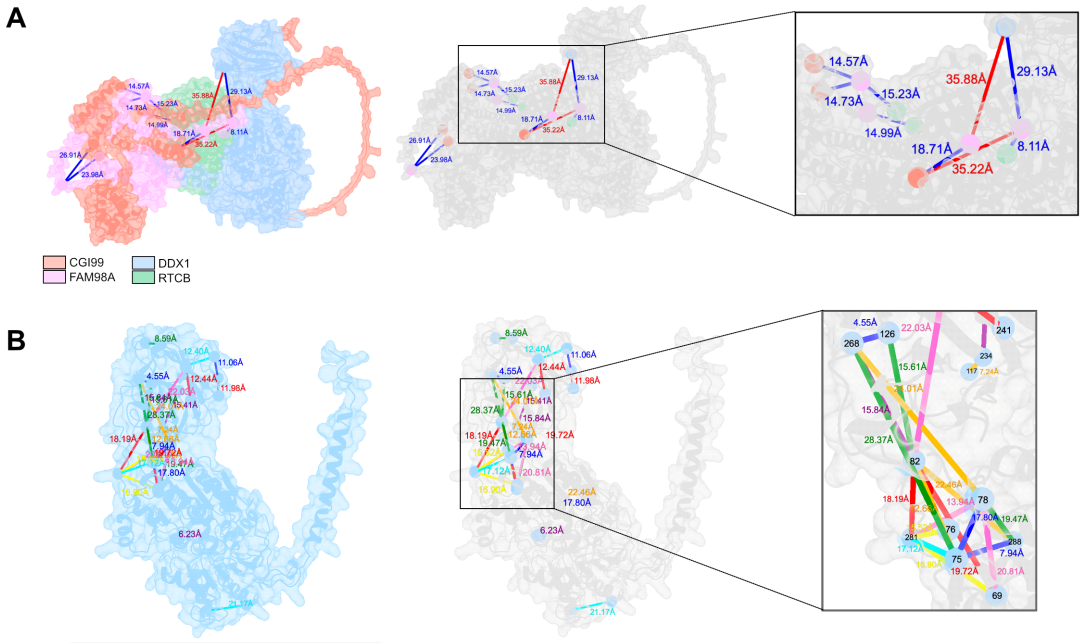

Figure 4.13: **Inter- and intra-protein XL-MS data corroborated in Alphafold3 tRNA-LC prediction**
A Alphafold3 prediction of four protein tRNA-LC: CGI99, FAM98B, DDX1 and RTCB. Experimentally derived inter-protein crosslinks are overlayed and coloured by Cα–Cα distance (*blue*: $\leq$30 Å, *red*: $\geq$30 Å). **B** Alphafold3 prediction of DDX1 with experimentally derived inter-protein crosslinks overlayed. Crosslink distances of Cα–Cα are indicated. In the final panel, the numbers in the blue circle correspond to the lysine residue position.

Structural predictions generated by AF3 can be cross-referenced using known intra- and inter-protein crosslinking sites identified by XL-MS. DSS has a maximum crosslinking distance

of 30Å, as discussed in a mathematical model considering molecular dynamics simulations [240, 252]. This distance is based on the length of two extended lysine chains and the DSS spacer length (as seen in the schematic in Figure 4.7A). In my dataset, I observed that most crosslink sites were within the 30Å restrictions in the inter-protein tRNA-LC prediction (figure 4.13A) and intra-protein DDX1 prediction (Figure 4.13B). This indicated strong experimental and computational correlation. Furthermore, the majority of crosslinks were clustered at the centre of the complex, which our predictions indicate to be the primary interaction interface. Only two crosslinks are outside of the 30Å distance constraint, suggesting a dynamic shift in protein assembly or a transient structure in which the proteins are in closer proximity.

The primary region of intra-DDX1 crosslinks is localised around the protein's N-terminal region, with crosslinked lysines indicated by numbered residues (Figure 4.13B). This region corresponds to the SPRY domain which has a resolved structure, in which there are two layers of stacked concave, anti-parallel β-sheets and a third β-sheet beneath the β-sandwich [253]. This structural feature is rich in lysines [254], which might account for the high density of crosslink sites.

In conclusion, the systematic knockdown of tRNA-LC proteins indicated a strong co-dependency and inherent co-regulated stability. The strong interaction between tRNA-LC components was further validated by the co-IP assay, which demonstrated interactions even in very harsh buffers. These two observations correlate well with my structural predictions that reveal a highly interconnected protein complex with CGI99 at its epicentre.

4.3 Discussion

Over the last decade, insight into the tRNA-LC structure and molecular behaviour has grown; however, many elements remain unclear. Deciphering the complex's intricate structure could enable the individual understanding each protein and its role within the complex.

4.3.1 tRNA-LC co-dependency and inter-Protein Interactions

The tRNA-LC relies heavily on strong inter-protein interactions, specifically between its core components. This was observed in the IP assay of each tRNA-LC member, which consistently co-precipitated the same protein interactors. Published co-precipitation of CGI99 and RTCB in native conditions with and without RNAse treatment reported DDX1, RTCB, CGI99, FAM98B and ASW as their primary interactors and the interaction was not mediated by an RNA bridge [197, 201]. This protein-protein network was "rediscovered" in the DDX1 and RTCB IP carried out here under very stringent conditions. The tRNA-LC proteins remain closely interacting irrespective of the experimental conditions, indicating robust interactions within the complex.

The siRNA knockdown underscores a persistent co-dependency among the tRNA-LC partners, revealing that the absence of core proteins DDX1, RTCB, and CGI99 leads to the destabilisation of the entire complex. Interestingly, the stability of the tRNA-LC varies between HEK293 cells analysed here and published HeLa cells, despite using the same siRNA sequences in both cases [197]. This variation suggests that cellular context plays a critical role in modulating complex stability and inter-protein dependencies. In both cell lines, RTCB knockdown impacts the complex entirely, while DDX1 impacts only itself and CGI99. The main difference between HEK293 and HeLa is in the knockdown of CGI99. The siCGI99 had little to no effect on RTCB in HeLa, whereas in HEK293 it significantly influenced the subsequent protein levels of RTCB and DDX1. CGI99's central role is further confirmed in published HEK293T cells, where the knockdown using a different siCGI99 displays a similar pattern to what was observed here [201]. Interestingly, outside of the cellular context, affinity purifications from insect cells infected with expression constructs lacking one of the tRNA-LC subunits revealed that deletion of any of RTCB, DDX1, FAM98B, or CGI-99 resulted in failure of the remaining four subunits to form a stable complex [243]. This suggests that different cell lines maintain and balance the complex asymmetrically, potentially dependent on the cellular environment and influences. However, I can conclude that in HEK293 cells, CGI99 and RTCB, and to a lesser extent DDX1, are key tRNA-LC scaffold proteins keeping the complex stable.

4.3.2 CGI99, an essential protein in the tRNA-LC structure

Depletion of CGI99 leads to the destabilisation of the tRNA-LC, as demonstrated by siRNA knockdown experiments. XL-MS data further emphasises CGI99 as a structural pillar of the complex, facilitating critical interaction interfaces. Notably, CGI99 primarily exists as a monomer, as suggested by prior studies, indicating that only one CGI99 molecule integrates into the tRNA-LC [201]. This limited molecular ratio may contribute to the complex's susceptibility to destabilisation. XL-MS findings reinforce AF3 predictions, which identify CGI99's C-terminus as a pivotal interaction hub for other tRNA-LC components.

However, in *in vitro* studies, the cross-linking dynamics observed differ somewhat from *in vivo* data, likely due to the absence of cellular factors in the test tube [243]. CGI99-centric crosslinks are significantly less pronounced outside the cellular environment. For example, when analysing minimal component tRNA-LC consisting of DDX1, RTCB, CGI99, and FAM98B, *in vitro* data showed FAM98B crosslinking exclusively to CGI99 and a broader crosslinking amongst DDX1, RTCB, and CGI99. This is distinctly different from the *in vivo* patterns, where DDX1 and RTCB selectively interact with only CGI99. Interestingly, when Ashwin (initially reported as a tRNA-LC component in [197]) was added to the *in vitro* model, the cross-linking landscape significantly shifted, producing an Ashwin-centric pattern, with no other inter-protein crosslinks observed [243]. Ashwin is only reported to be present in the tRNA-LC when localised in the nucleus [255]. The absence of Ashwin in the *in vivo* data may stem from low nuclear tRNA-LC abundance or from the limited MS depth in my data, which limited Ashwin's detection. Furthermore, the significantly higher crosslinks identified in the in vitro study reflect the much higher difficulty of identifying crosslinking in cellullo as the crosslink capture and permeability while maintaining physiologically relevant crosslinks restricts the amount of crosslinks identified.

The in vitro study further analysed the required domains of each protein that would enable complex structure by performing a systematic deletion of different regions. Researchers identified the C-terminal region of each tRNA-LC protein as the essential subunit regions required for the formation of the minimal tRNA-LC. The complex architecture defined in this study aligns with the AF3 prediction generated here, which identified the C-terminal regions as the essential binding platform supported by CGI99. In future work, I would cross-validated the in vitro data to both my in vivo data as well as map the combination on the AF3 simulation.

However, cross-linking dynamics observed *in vitro* differ notably from those seen *in vivo*,

likely due to the absence of cellular factors and physiological conditions in the test tube environment [243]. In particular, CGI99-centric crosslinks appear significantly diminished outside the cellular context. For instance, in the minimal *in vitro* tRNA-LC composed of DDX1, RTCB, CGI99, and FAM98B, crosslinking was primarily observed between FAM98B and CGI99, while broader crosslinking occurred among DDX1, RTCB, and CGI99. This contrasts with the *in vivo* data presented here, where DDX1 and RTCB show more selective interactions with only CGI99.

Interestingly, the addition of Ashwin, previously reported as a nuclear-specific tRNA-LC component [197], to the *in vitro* reconstitution resulted in a dramatic shift in the crosslinking landscape. An Ashwin-centric pattern emerged, with no other inter-protein crosslinks detected [243]. This finding supports the notion that Ashwin-containing forms of the tRNA-LC may be nuclear-specific, consistent with previous reports of its nuclear localisation [255]. The absence of Ashwin in the current *in vivo* dataset could be due to a combination of factors, including low abundance of the nuclear-localised tRNA-LC or limited detection sensitivity due to mass spectrometry depth. Moreover, the higher number of crosslinks identified *in vitro* reflects the technical ease of crosslink capture in purified systems, where crosslinker accessibility and reaction efficiency are maximised [243]. In contrast, *in vivo* crosslinking must balance permeability and physiological integrity, which inherently limits the crosslinking yield and complexity of the data.

The *in vitro* study also mapped domain contributions to complex assembly through systematic deletion analysis [243]. It revealed that the C-terminal regions of each subunit are essential for minimal complex formation. This structural model closely mirrors the AF3 prediction generated here, which similarly identified the C-terminal domains as a key interaction interface. In future work, I aim to integrate these findings by mapping both *in vitro* and *in vivo* crosslink data onto the AF3 structural model to further validate the predicted architecture and explore context-dependent dynamics of the tRNA-LC.

4.3.3 FAM98A/B competitive binding and the existence of a sub-complex

XL-MS data can reveal protein competition for specific interfaces by identifying two or more proteins crosslinking at the same position within the target protein. While this cannot be readily done *in vitro*, my *in vivo* XL-MS data revealed several positions in CGI99 where competition may exist. My data shows that CGI99 is the central component of the tRNA-LC. Notably,

there are overlapping crosslink sites for FAM98A and FAM98B on CGI99. Because there is a single molecule of CGI99 within the tRNA-LC [199], FAM98A/B may be in competition for the same interface. This is supported by the substantial sequence homology between the two proteins and their similar subcellular localisation. However, little is reported about how the two proteins differ and whether the presence of FAM98A or FAM98B changes the role and function of the tRNA-LC. Unfortunately, FAM98A is not routinely reported as a protein within the tRNA-LC and was omitted in the *in vitro* XL-MS study [243]. The silver staining of the FAM98A IP indicated the lowest enrichment of tRNA-LC proteins. Nonetheless, my *in vivo* XL-MS data suggest that it is associated with the tRNA-LC at least to some degree. The AF3 prediction of FAM98A/B indicated intrinsically disordered regions (IDRs) in their C-terminus, which are longer in FAM98A. IDRs are enabled with high intrinsic capacity to establish molecular interactions with other proteins and with RNA [256]. I speculate that these differential IDRs could play a crucial role in establishing distinct interaction networks. The longer length of FAM98A IDR suggests a larger interaction platform than that of FAM98B, and indeed, the IDR of FAM98A has been described as binding RNA [88].

A smaller complex has been identified *in cellulo* involving DDX1, CGI99, FAM98A and FAM98B, and PRMT1, an antitumorigenic agent, [257]. The *in vitro* XL-MS data identified a primary complex involving the same combination of proteins, DDX1, CGI99 and FAM98B [243]. My *in vitro* XL-MS data, unfortunately, cannot distinguish between sub-complexes present *in vivo*, as it is not quantitative . However, the AF3 prediction indicated a semi-flexible arrangement which allows the insertion of RTCB into the tRNA-LC as a final component. The flexible organisation could potentially indicate a sub-complex as identified in other datasets.

4.3.4 XL-MS may capture complex dynamics

AF3 and *in cellulo* XL-MS dataset showed a strong correlation between experimental and AI-based predictions. However, a few crosslinks were further than the 30 Å distance restraint of the DSS crosslinker. These longer-distance crosslinks suggest either the existence of dynamic conformational changes or potential AF3 inaccuracies. Notably, conformational changes have been previously observed for DDX1 and RTCB when interacting with ATP/ADP or GTP/GDP, respectively (deposited DDX1-ADP structure in PDB by Zhang et al. in 2023, and [244]). Conformational flexibility is not unexpected among RNA helicases, including DEAD-box proteins, which undergo structural rearrangements when unwinding RNA [251]. Among DEAD-

box proteins, DDX1 is unique due to the presence of an SPRY domain inserted between the Q-motif and the N-terminal domain. This domain is thought to mediate protein–protein interactions [258] and contribute to enzymatic activity. Furthermore, ATP and RNA binding induce a transition from an open to a closed conformation of the helicase core, stabilising interdomain interactions. Kellner et al. (2015) proposed a model in which ATP or RNA binding alone can partially shift the equilibrium towards the closed state [245]. Interestingly, DDX1 exhibits an unusually high affinity for ADP, approximately threefold greater than its affinity for ATP, potentially locking the enzyme in a dead-end ADP-bound state under physiological conditions. Recycling DDX1 may, therefore, require active nucleotide exchange to restore its functional state [245]. RTCB structure alone or with recycling factor Archease has been resolved, revealing minor conformational changes during activation by Archease [244]. RTCB also undergoes catalytic rearrangements to facilitate GDP release and protein recycling mediated by Archease, consistent with its role in the tRNA-ligase complex. XL-MS has the potential to capture multiple conformations, particularly those that are longer-lived or that result in optimal lysine configurations. Future applications of XL-MS could model all potential configurations of the tRNA-LC, offering valuable insights into its structural dynamics.

Machine learning advances have exemplified how the marriage between experimental and computational tools can aid in future research. Indeed, here, I showed a proof of concept where I could accurately determine experimental crosslinks on the structural prediction of the tRNA-LC. Learning models can actively be trained on experimental data, with as little as a single crosslink improving prediction software drastically. Distance restraints can be leveraged to improve AF models, as shown through the use of Alphalink and, its extension AlphaFold-Multimer [235]. Conformational changes of complexes and proteins have been explored by the Topf group. They have developed a pipeline to model the structure of proteins with multiple conformations, called XLMS-tools [240]. The pipeline consists of two main steps: generation using AF2, followed by a conformer selection using XL-MS data. For conformer selection, mathematical modelling of the monolink probability score and the crosslink probability score, both of which are based on residue depth from the protein surface, is used [240]. This highlights the DDX1 XL-MS unexplored dataset of monolinks. Monolinks are single-ended crosslinks which can be representative of surface structure. The focus of this study was to capture the inter-protein networks of the tRNA-LC. However, future work could explore the surface structure of the complex and whether it aligns with computational prediction.

4.3.5 Novel tRNA-LC interactors identified: RPL11 and MYH9

The strength with *in cellulo* XL-MS is the detection of all potential forms of tRNA-LC as well as other novel interactors, without the limitation of the proteins that "one adds to the test tube" and maintaining subcellular localisation and molecular functions. Two proteins, RPL11 and MYH9, crosslinked with CGI99 and RTCB, respectively. These direct interactions indicate that the tRNA-LC associates with other complexes and is involved in a wider range of functions beyond tRNA maturation and Xbp1 splicing [197, 203]. While I cannot distinguish the prevalence of these interactions or whether they occur on the same complex at the same time, the detection of these direct binders supports an intimate association that is prominent enough to be captured.

The tRNA-LC, specifically CGI99, has cap-binding activity and positively modulates mRNA translation [246]. Furthermore, DDX1 has been identified as a key regulator of insulin translation [259]. It is thus not surprising to identify ribosomal proteins RPL11 and RPL7 in the enriched crosslinking dataset, and the translation initiation factor EIF3A that bridges the 40S ribosomal subunit with the mRNA. Intriguingly, RPL11 crosslinks directly with CGI99, indicating a direct association between the 60S ribosomal subunit and the tRNA-LC. The crosslink of RPL11 or GGI99 overlaps with that of RTCB, indicating that they are either mutually exclusive interactions or form a tripartite protein interface. The existence of the previously described sub-complex, including DDX1, CGI99 and FAM98A/B, supports the exclusion possibility [243, 257]. However, further investigation is required, as this crosslink site could also represent dynamic structural reconfiguration to allow for novel protein interactions to occur. The tRNA-LC includes several molecular functions that can be associated with its individual components: RNA ligation by RTCB [197, 198, 203], cap-binding by CGI99 [246], and RNA unwinding by DDX1 [253, 260, 261]. I speculate that the tRNA-LC reconfigures its structure while performing each task, exposing the necessary protein regions for the appropriate process. Meanwhile, the other proteins may remain "silent", providing structural integrity to the complex. This suggests that in a larger multi-protein assembly, such as the tRNA-LC, conformational changes may occur to allow for different interfaces and molecular functions to carry out their functions, for example, to establish new interactions with the translation apparatus or interact with RNA.

The tRNA-LC operates across both the nucleus and cytoplasm, fulfilling distinct roles such as tRNA maturation, translational control, and ER-mediated stress responses [197, 203, 204, 246, 259]. My findings reveal RTCB's interaction with MYH9, a cytoskeleton-associated motor

protein, suggesting the complex engages directly with the cytoskeleton to mediate transport or regulate mRNA movement. Previous studies also link CGI99 with MYLK2 (myosin light chain kinase II), which modulates myosin activity, and demonstrate the involvement of RTCB, DDX1, and CGI99 in neuronal RNA transport granules [201, 202]. The shuttling of tRNA-LC components between compartments is a dynamic process potentially regulated by proteins like CGI99 and Ashwin. CGI99 interacts with MYLK2 which supports its role in cytoplasmic translocation [201], while Ashwin may direct nuclear localisation due to its nuclear signals and basic amino acid profile [262]. Inhibiting transcription reduces the nuclear translocation of RTCB, DDX1, and CGI99, indicating transcriptional activity is in part responsible for their cellular localisation [201]. Future studies should explore whether these dynamics involve post-translational modifications, such as phosphorylation, or signalling pathways, like the unfolded protein response or neuron translation requirements, to fully elucidate the mechanisms driving tRNA-LC movement.

In conclusion, the tRNA-LC is a strongly bound and intricately organised complex that harbours multiple functions in cell biology. Direct protein crosslinking has identified novel interactors, expanding our knowledge of this complex's biological roles. To contextualise these interactions, Chapter 6 carries out a wider protein-protein and protein-RNA interactome, exploring more broadly the scope of proteins and RNAs that interact with the tRNA-LC.

5 Characterising DDX1 phenotype in infection

5.1 Introduction

The tRNA-LC components have previously been identified in regulating viral lifecycles. The most well-characterised protein in the complex is DDX1, which has been observed to inhibit or promote infection depending on the specific virus. DDX1 promotes HIV, SARS-CoV-1 and VEEV through direct interaction with their respective viral proteins Rev, nsP14 and nsP2 [185, 189, 191, 192]. Meanwhile, DDX1 has been observed to inhibit TGEV, FMDV and IAV through its crucial involvement in IFN-β stimulation [166, 194, 195]. Moreover, DDX1 was captured as a direct interactor with a host of different vRNAs. The comparative analysis performed by Iselin et al. (2022) compared different viral interactome capture methods and identified DDX1 consistently across the different studies [115]. This indicated that DDX1 is a direct interactor with the vRNA of SINV, CHIKV, ZIKV, DENV and SARS-CoV-2.

CGI99 has also been implicated in the regulation of IAV infection. Contrary to DDX1, it was characterised as an essential factor for viral transcription and proliferation. During IAV infection, CGI99 interacts with the viral polymerase subunit PA and contributes to increased viral polymerase activity, enhanced vRNA transcription and augmented viral replication [199]. The absence of CGI99 significantly reduced IAV production. Crucially, CGI99 co-localises with vRNP inside IAV virions [200]. The identification of CGI99 in these IAV regulatory roles did not capture tRNA-LC members. However, the researchers did not specifically test for the presence of tRNA-LC members as they employed a predominantly antibody-based approach focused on CGI99. The previous chapter highlighted the importance of CGI99 in the tRNA-LC structure, and I suspect CGI99 was not singularly associated with IAV vRNPs in the virion, although further investigation is required. To note, the inhibitory effect of DDX1 in IAV infection was associated with a different complex, the DDX1-DDX21-DHX36 complex, involved in triggering the TRIF pathway, which activates type I IFN [178]. The DDX1-DDX21-DHX36 complex was not identified in the HEK293 DDX1 DSS crosslinking in the previous chapter and, as such, may be related to the specific stimulus (IAV infection) or to the cell type (dendritic cells).

More recently, DDX1 and other components of the tRNA-LC were involved in the SINV lifecycle. DDX1, RTCB and FAM98A were captured as having increased RNA binding activity over the course of SINV infection and were further identified as direct vRNA interactors [113, 123]. DDX1 and RTCB have both been observed to re-localise to viral replication

organelles during SINV infection [113]. The roles of the tRNA-LC in SINV infection remain, however, uncharacterised. In this chapter, I employed various knockdown techniques to uncover the regulatory functions of tRNA-LC proteins during SINV infection. Building on the findings of the previous chapter, I hypothesise that the proteins within the complex orchestrate a coordinated regulatory mechanism. To explore this, I examined the enzymatic activities critical to the complex's function, including RTCB's ligase activity facilitated by Archease recycling and DDX1's helicase activity. Additionally, I characterized the effects of previously used siRNAs through whole-cell transcriptome analysis of uninfected and infected cells. This analysis highlighted RNA-level changes resulting from the absence of two core proteins, DDX1 and CGI99, shedding light on their potential roles in infection.

5.2 Results

Proteome-wide analyses have identified components of the tRNA-LC as regulators of virus infection [113] and recent work from our lab has shown that they interact directly with SINV vRNA [123]. However, the impact of the tRNA-LC has not yet been assessed in alphaviruses, and its regulatory role, if any, has not yet been characterised. In order to understand how the tRNA-LC modulates alphavirus infection, I used two chimeric SINV constructs previously generated in the Castello lab. Using a chimeric virus that expresses mScarlet from the fusion with the non-structural protein 3 (SINV$_{nsp3-mScarlet}$), I can observe the early translation of non-structural proteins (Figure 5.1A). This provides a proxy for the early stages of infection. Alternatively, I can use a chimeric virus that expresses mCherry from a duplicated subgenomic promoter (SINV$_{mCherry}$), a proxy for subgenomic translation, to assess late viral gene expression (Figure 5.1B).

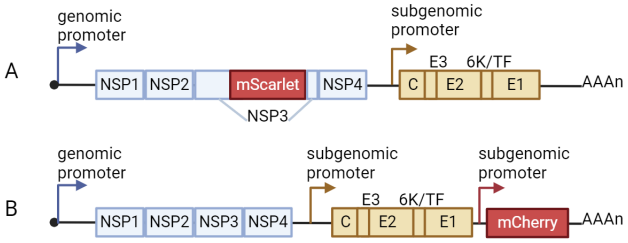

Figure 5.1: **SINV chimeric viruses schematic**

A Schematic of SINV virus with an nsP3 intercalated mScarlet tag (SINV$_{nsp3-mScarlet}$). **B** Schematic of SINV virus with mCherry tag under a duplicated subgenomic promoter (SINV$_{mCherry}$).

5.2.1 HEK293 inducible shDDX1 cell line reveals viral inhibition

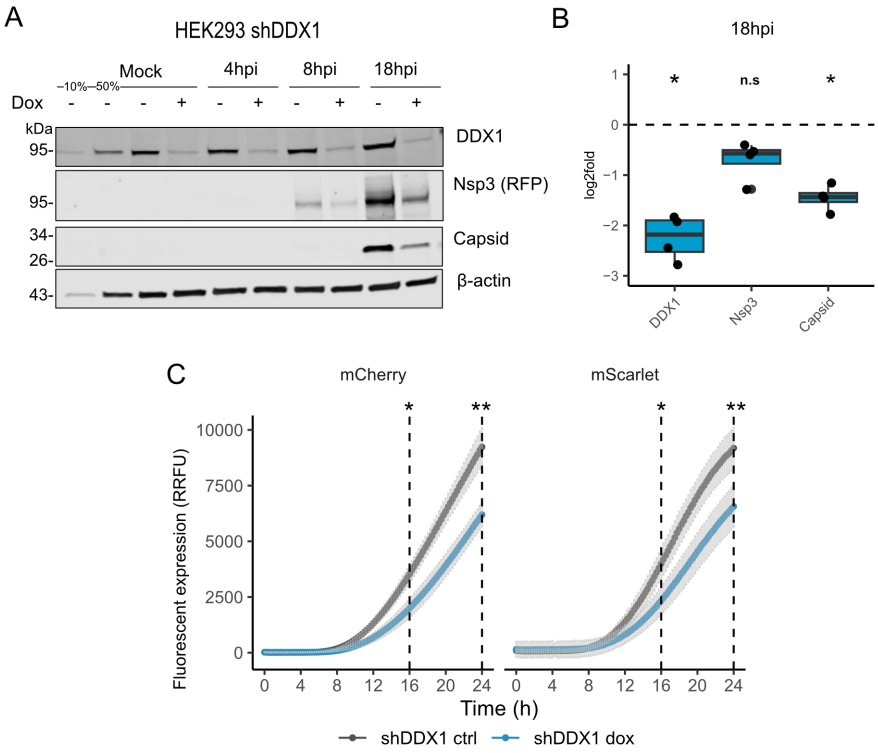

Figure 5.2: **HEK293 shDDX1 indicates viral inhibition in absence of DDX1**

A Western blot of HEK293 shDDX1 cell lines, either induced for 48h (doxycyline induction) or uninduced. Cells were infected with SINV$_{nsp3-mScarlet}$ at an MOI of 0.1 and harvested at Mock, 4hpi, 8hpi and 18hpi. Antibodies for DDX1, RFP (indicative of the viral tagged Nsp3-mScarlet protein), SINV Capsid and β-actin were used. **B** Relative quantification of 18hpi protein signal from western blot protein signals from 4 independent biological replicates. Normalised to β-actin signal levels, followed by a log2 fold change quantification against the signal of the uninduced cell line. **C** HEK293 shDDX1 cell lines were seeded and grown with or without doxycyline (uninduced labelled ctrl, and induced labelled +dox) for 48h. Cells were infected with SINV$_{mCherry}$ (mCherry) and SINV$_{nsp3-mScarlet}$ (mScarlet) at an MOI of 0.1. Fluorescence measurements were taken every 15 minutes for 24 hours by BMG Clariostar plate reader. Fluorescence intensity was normalised as follows: minimum signal level set to 0, and maximum signal set to 10000 for the control uninduced wells; doxycyline induced wells were normalised relative to control. Error bars represent standard deviation across the 3 technical replicates from each condition, and 3 biological replicates from independent plates. Statistical significance was tested using two-tailed T-test comparing to the uninduced control (P $\leq$0.01 **, P $\leq$0.05 *, P $\leq$0.1).

HEK293 cells induce a robust and comprehensive antiviral response against SINV infection [113]. Interestingly, DDX1, RTCB and FAM98A have also been observed to have increased RNA-binding activity in HEK293 cells following SINV infection. To determine whether the knockdown of DDX1 in HEK293 cells altered infection outcomes, I first employed a previously

inducible cell line that expresses short hairpin RNAs (shRNA) targeting DDX1 [116]. The inserted shRNA construct is induced by the addition of tetracycline (or the more stable doxycycline). I infected the shDDX1 line with SINV$_{nsp3-mScarlet}$ after 48 hours of induction. The knockdown of DDX1 was achieved to a high level (around 90% reduction compared to the uninduced). Furthermore, the absence of DDX1 elicited a decrease in the viral protein produced, observed by the lower quantity of nsP3 produced as early as 8hpi and the lower levels of capsid proteins at 18hpi (Figure 5.2A). The relative quantification of the protein signal at 18hpi (Figure 5.2B) indicated a significant decrease of DDX1 and capsid protein. To further validate this effect, I assessed the dynamics of viral growth using a live cell plate reader, Clariostar, which allowed for a time course observation of viral fitness measured by red fluorescent signal emitted from the chimeric viruses as a proxy of viral gene expression. This assay showed a significant difference between the 2 conditions in both viruses, further validating the western blot results (Figure 5.2B).

In addition to assessing protein levels in the HEK293 inducible shDDX1 cell lines, transcriptome analysis was conducted both in the presence and absence of SINV infection. The results (not shown) revealed no significant off-target effects of the shRNA beyond DDX1. As expected, DDX1 was the most significantly downregulated transcript, confirming the specificity of the knockdown. Moreover, the SINV vRNA is significantly downregulated in the doxycycline-induced shDDX1 lines compared to the uninduced control, consistent with the decreased viral protein expression observed in Figure 5.2B.

5.2.2 HEK293 siRNA targeting tRNA-LC indicates a viral increase in the absence of these key proteins

To validate the observed phenotype using the shDDX1 knockdown, I next employed single siRNA knockdown system previously described in Chapter 4. Briefly, I selected three different siRNAs targeting DDX1 (Figure 4.4). Notably, one of the siRNAs corresponds to the sequence in the shDDX1 system developed by the Proudfoot lab (referred to as siDDX1_1).

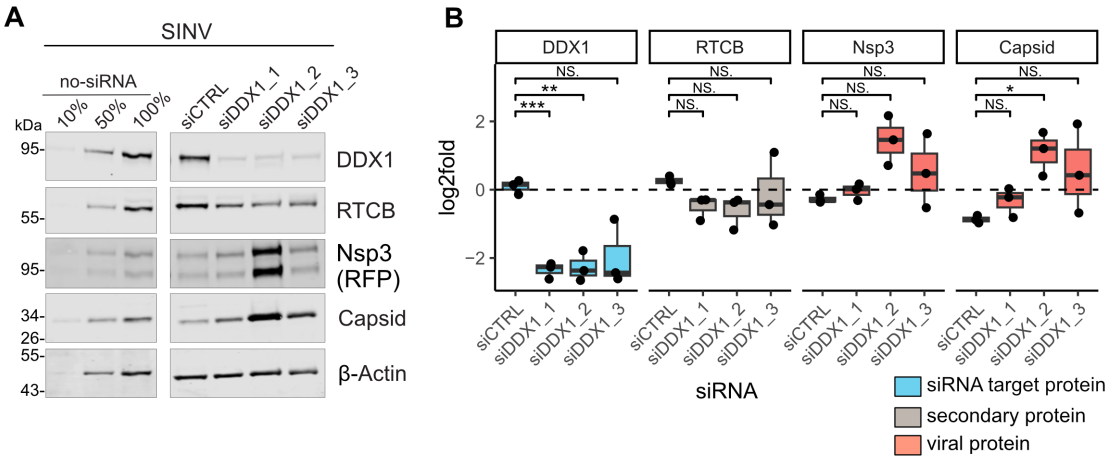

Figure 5.3: **Knockdown of DDX1 by single target siDDX1 suggest viral protein increase**

A Western blot images of single target siDDX1 after 48h transfection followed by SINV infection with an MOI 0.1 $SINV_{nsp3-mScarlet}$ infection for 18h. No-siRNA refers to mock transfected cells (water only). A dilution of 10% and 50% of the lysates was performed to infer the quantitative accuracy of the assay. Antibodies for DDX1, RTCB, RFP (indicative of the viral-tagged Nsp3-mScarlet protein), SINV Capsid and β-actin were used to study the knockdown effect. **B** Relative quantification of western blot protein signals from 3 independent biological replicates. Normalised to β-actin signal levels, followed by a log2 fold change quantification against the signal of the no-siRNA control sample. Statistical significance was tested using two-tailed T-test against siCTRL samples (P $\leq$0.001 ***, P $\leq$0.01 **, P $\leq$0.05 *, NS: non-significant).

As previously, the knockdown experiments revealed consistent DDX1 suppression across all three siRNAs, with only mild secondary effects on the levels of RTCB and CGI99 (Figure 5.3). However, viral protein levels were remarkably different when I challenged the DDX1-knockdown cells with SINV infection. The results were the opposite of the shRNA-induced phenotype previously observed. Interestingly, the depletion of DDX1 by siDDX1-1 (same sequence as the shDDX1) slightly enhanced viral protein production compared to the siCTRL (Figure 5.3B). The two other siRNAs (siDDX1-2 and siDDX1-3) showed a clear upregulation of viral protein expression, with capsid levels in siDDX1-2 being statistically significant. The contradictory results between the shRNA and siRNAs call for further experiments to determine whether DDX1 is a dependency or an antiviral factor.

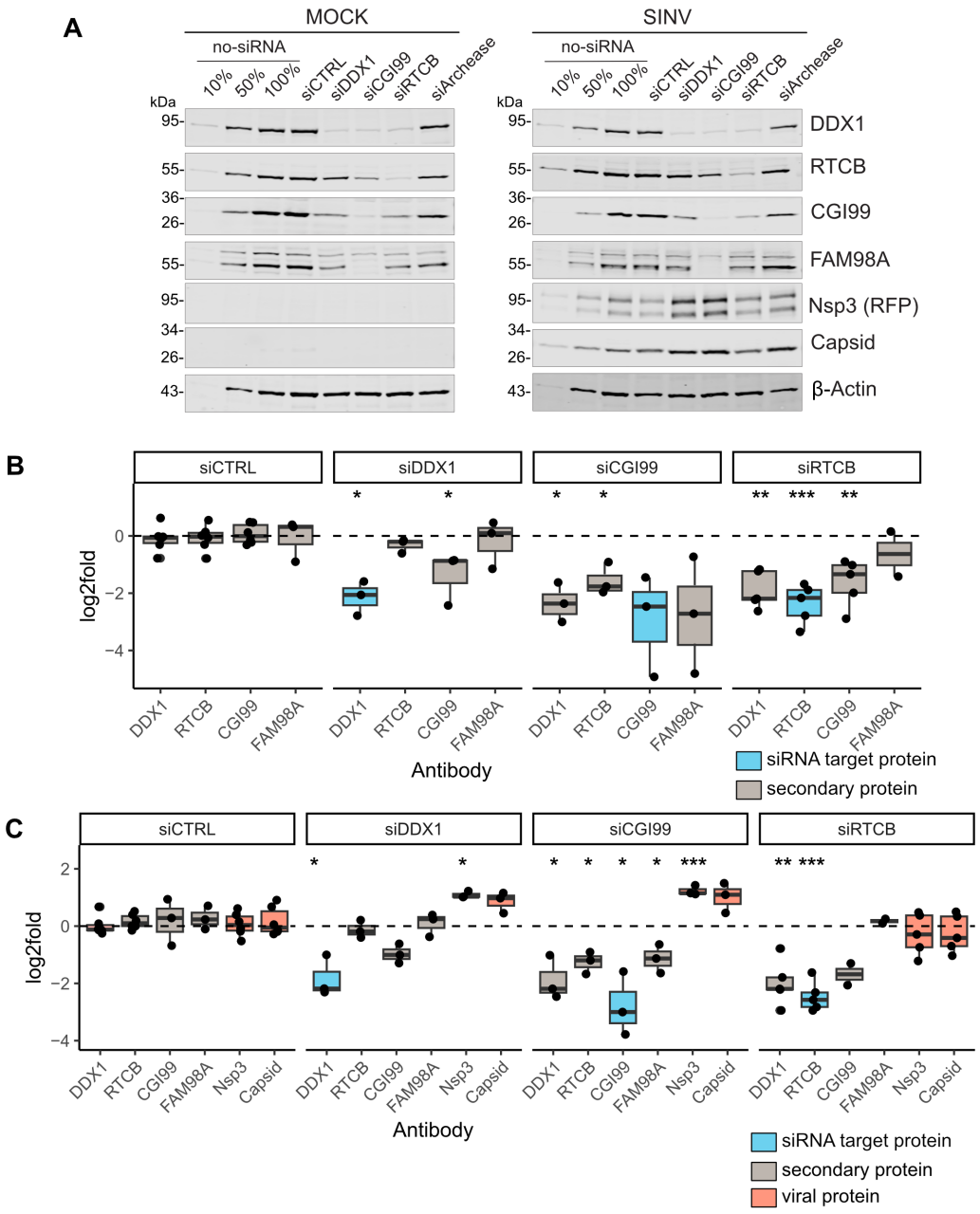

Figure 5.4: **siRNA pools targeting tRNA-LC proteins exhibit antiviral role by DDX1 and CGI99**

A Western blot images of siRNA pools targetting DDX1 (siDDX1), CGI99 (siCGI99), RTCB (siRTCB) and Archease (siArchease) after 48h transfection (MOCK samples) followed by MOI 0.1 SINV$_{nsp3-mScarlet}$ infection for 18h (SINV samples). No-siRNA refers to mock-transfected cells (water only). **B and C** Relative quantification of western blot protein signals from 3-5 independent biological replicates in uninfected mock samples (**B**) and SINV infected samples (**C**) (n=3 for siDDX1 and siCGI99, n=5 for siRTCB). Western blot protein signal normalised to β-actin signal levels, followed by a log2 fold change quantification against the signal of the no-siRNA control sample. Statistical significance was tested using two-tailed T-test against siCTRL samples (P $\leq$0.001 ***, P $\leq$0.01 **, P $\leq$0.05 *). In blue, siRNA target protein; in red, viral protein; and in grey, secondary proteins quantified.

Due to the discrepancy in phenotype when using single DDX1 sequence targets, I decided

to use the siRNA pool for DDX1, CGI99 and RTCB protein knockdown, as previously used in Chapter 4. I challenged the respective knockdowns to SINV and analysed by western blotting tRNA-LC proteins as well as viral proteins (Figure 5.4A). siRNA pools for tRNA-LC proteins led to a pronounced and statistically significant reduction in levels of the target protein and in some cases, other members of the complex, suggesting higher efficiency than individual siRNAs (Figure 5.4B and C). The siDDX1 had a secondary effect on CGI99 protein levels. The siCGI99 reduced the levels of DDX1, RTCB and FAM98A. Meanwhile, siRTCB significantly reduced the relative levels of DDX1 and CGI99. The secondary effect on other proteins in the tRNA-LC indicates a strong co-dependence for complex stability in the hierarchical order previously outlined in Chapter 4. Knockdown of DDX1 and CGI99 led to a significant increase in the amount of nsP3-scarlet and capsid protein (Figure 5.4C), which is consistent for DDX1 with the individual siRNAs. No significant effects in viral protein expression were observed in the siRTCB knockdown samples. Noting that siRTCB also reduced DDX1, my results add further complexity to the tRNA-LC effect in SINV infection. However, an antiviral role is consistently displayed across several siRNAs against DDX1 and CGI99. The siRTCB pool may have confounding effects difficult to control, which may translate to different effects in viral proteins.

The direct knockdown of tRNA-LC proteins exhibited a puzzling array of phenotypes. Using an orthogonal approach to decipher the regulatory role of the complex in SINV infection, I next investigated whether a tangible effect on the catalytic function of the tRNA-LC could be detected. Archease is closely associated with the tRNA-LC and is an essential RTCB cofactor as it promotes catalytic recycling. In conjunction with DDX1, Archease enables the turnover of GTP/GMP on RTCB after a ligation reaction [198].

To investigate if the catalytic function of the tRNA-LC affects SINV infection, I aimed to inhibit the ligase recycling ability through the knockdown of Archease. To assess the impact of Archease depletion on viral fitness, I employed a siRNA pool specifically targeting *Archease*. Using quantitative real-time PCR (qRT-PCR), I measured knockdown efficiency. Two distinct primer sets used in [198] were employed, resulting in varying knockdown levels (80% or 50%, depending on the primer pair) (Figure 5.5A). In both cases, *Archease* mRNA levels significantly decreased compared to control samples treated with siCTRL, which indicates that knockdown occurred. Notably, the knockdown of Archease nearly tripled the quantity of vRNA detected compared to the control siRNA. This suggests that Archease inhibits infection, consistent with

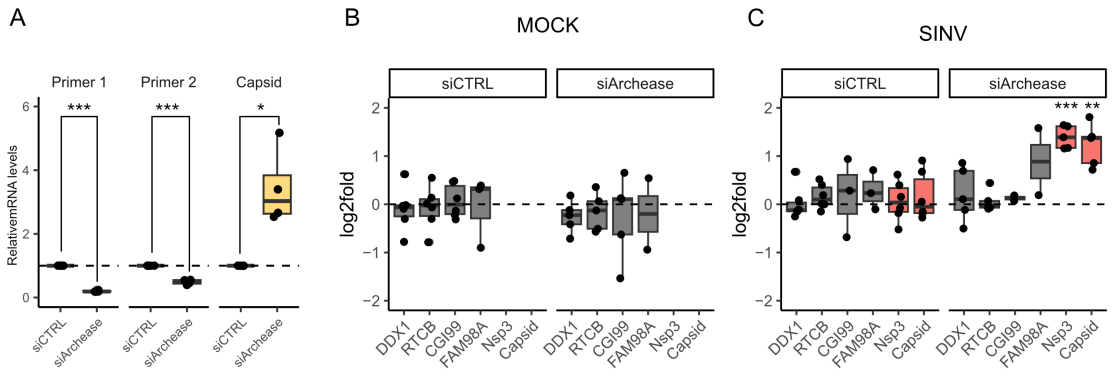

Figure 5.5: **siArchease impact on tRNA-LC and viral fitness**

A Relative mRNA levels measured by qRT PCR of siCTRL and siArchease in SINV infection. RNA isolated from 48h siRNA transfection and 18h SINV infection. Two different primer sets used targeting Archease mRNA (Primer 1 and Primer 2), and primer set targeting SINV Capsid mRNA. Ct values from 4 biological replicates, normalised to housekeeping GAPDH Ct value, were used for $\Delta\Delta$ Cq calculation of siArchease over siCTRL. **B and C** Relative quantification of western blot protein signals from 5 independent biological replicates of siCTRL and siArchease in uninfected mock samples (B) or 18hpi SINV infected samples (C). Protein signal normalised to β-actin signal levels, followed by a log2 fold change quantification against the signal of the no-siRNA control sample. Representative Western blot image in Figure 5.4. Statistical significance was tested using two-tailed T-test comparing to the siCTRL (P $\leq$0.001 ***, P $\leq$0.01 **, P $\leq$0.05 *)

the effects observed with siDDX1 (pool) and siCGI99.

While the interaction between Archease and the tRNA-LC is transient, its importance in activating RTCB may affect the conformation of the complex. Hence, I tested if the knockdown of Archease had an effect on the abundance of the tRNA-LC components. Additionally, I tested whether the knockdown affected viral protein expression. I performed western blot analysis on siArchease lysates, specifically probing for tRNA-LC proteins (Figure 5.5B and C). Under mock and infected conditions, there were no significant changes in the protein levels of DDX1, RTCB, CGI99 and FAM98A when compared to the siCTRL levels. However, I observed a significant increase in SINV nsP3 and capsid. This further indicated that the inhibition of Archease did not affect the stability of the tRNA-LC, yet increased vRNA and protein production.

In total, I have observed CGI99 and DDX1, two essential tRNA-LC proteins, significantly upregulating viral proteins when absent. CGI99 significantly destabilises other proteins of the tRNA-LC, and in the previous chapter, it was observed to be central to the complex's structure. This suggests that the tRNA-LC is a regulator of SINV, inhibiting viral fitness. Furthermore, the absence of Archease, the ligase's essential co-factor, causes a significant stimulation of viral gene expression and translation. This indicated that the catalytic function of the complex may

be necessary for the antiviral effect.

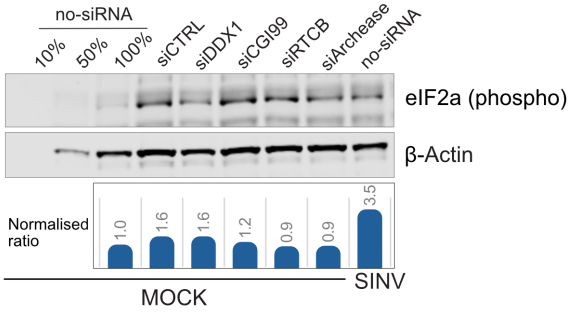

Figure 5.6: **Stress response detected by phosphorylated eiF2α**

Western blot of siRNA transfected samples, probing for phosphorylated eiF2α to verify stress response. The full panel of siRNA in mock conditions and a single no-siRNA sample in SINV condition as a positive control was used. Protein signal of eIF2α-phospho normalised to β-actin signal levels, followed by a ratio calculation to the MOCK no-siRNA control.

siRNA transfections can elicit undesirable secondary responses from the cell, such as off-target silencing or activation of stress response pathways [263]. The integrated stress and the antiviral responses are characterised by the phosphorylation of eukaryotic initiation factor-2α (eIF2α) to arrest protein synthesis [264–266]. I subsequently aimed to verify if the baseline stress/antiviral response is triggered by siRNA transfection. Phosphorylation of eIF2α was assessed by western blot following depletion of various proteins in uninfected cells. SINV is known to induce robust phosphorylation of eIF2α [126, 267, 268], so samples without siRNA under mock and SINV-infected conditions were included as negative and positive controls, respectively (Figure 5.6). As expected, in the absence of siRNA, eIF2α phosphorylation was low under mock conditions (normalized to 1) and increased approximately threefold upon SINV infection. Across all siRNA treatments tested, including the siCTRL, eIF2α phosphorylation levels remained largely unchanged compared to the no-siRNA control. While siRNA transfection can cause cellular stress that might influence experimental outcomes, in this case, the stress appears to be uniform across all conditions. Notably, differential phenotypes were still observed for specific siRNAs. For example, siRTCB did not produce an observable phenotype, and eIF2α phosphorylation levels remained comparable to both siCTRL and no-siRNA controls. These results suggest that the mild phosphorylation of eIF2α induced by siRNA transfectant does not significantly affect SINV infection outcomes.

5.2.3 Assessing the importance of the tRNA-LC catalytic activity in SINV infection

The knockdown of Archease indicated that the tRNA-LC catalytic activity correlates with the antiviral effects of DDX1 and CGI99. Archease has not been reported to be involved in any other pathway other than the RTCB ligase recycling reaction. The cooperation of DDX1 and Archease in ligase recycling and the similarities of their phenotypes represent an interesting connection to explore.

To further characterise DDX1's involvement in viral fitness and specifically its role in the activity of the tRNA-LC, I next expressed a DDX1 catalytic null mutant in cells. DDX1 is an ATP-dependent helicase, and a mutation from a lysine to an alanine at position 52 of the Walker A motif involved in ATP binding and hydrolysis (K52A) renders the helicase catalytically inactive [253]. I employed an inducible cell line established in the Castello laboratory, where the DDX1-K52A mutant with a GFP tag was introduced into a HEK293 FITR cell line. The cell line enabled me to study the impact of viral fitness in the presence of the catalytic null mutant.

I induced the expression of the mutant protein for 24h before challenging the cells with SINV. In parallel, I subjected wild-type DDX1-eGFP (previously used in Chapter 4) to the same treatment for comparison. I assessed the levels of DDX1, RTCB and the viral proteins Nsp3 and Capsid by western blot (Figure 5.7A). Interestingly, I noticed that DDX1-eGFP variants replace the endogenous DDX1. This is consistent with the need for assembly into the tRNA-LC to be stable and thus limited to the other proteins' stoichiometry. DDX1-K52A-eGFP replaced the endogenous more efficiently, which could be explained by the fact that the helicase activity of the tRNA-LC becomes locked in a "no go" conformation (Figure 5.7A). The mutant outcompeting the endogenous functional protein is characteristic of a dominant negative behaviour. The levels of RTCB protein remained stable in the DDX1-GFP cell line in both conditions. However, I noticed a slight reduction in RTCB levels following infection in the DDX1-K52A-eGFP expressing cells. Notably, the viral proteins nsP3 and Capsid levels decreased in DDX1-K52A expressing cells. This indicated that the mutant protein inhibited viral protein synthesis. I further examined the dynamics of chimeric SINV$_{mCherry}$ and SINV$_{nsp3-mScarlet}$ viruses in a live cell plate reader assay (Figure 5.7B). The plate reader indicated mild differences between the cells expressing DDX1 or DDX1-K52A, suggesting lower effects in the fluorescent protein expression. These results with the catalytic mutants were not consistent with the role of DDX1 as an antiviral protein. However, side effects of an assembled tRNA-LC with a "poisoned" helicase may have broader cellular consequences that

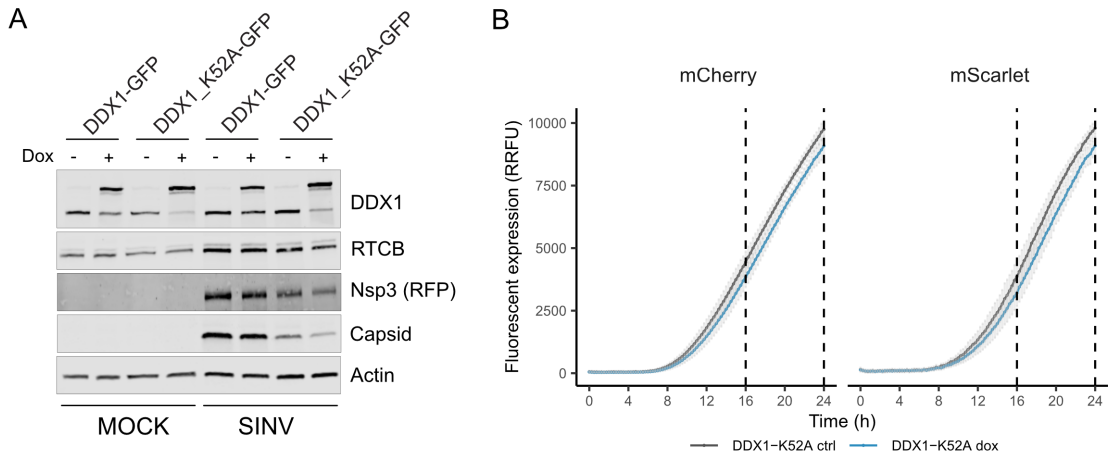

Figure 5.7: **Impact of DDX1 catalytic mutant K52A in infection**

A Western blot image of HEK293 FITR cell lines with either wild type GFP tagged DDX1 (DDX1-GFP) or with catalytic null mutant GFP tagged DDX1-K52A (DDX1_K52A-GFP). Samples were induced for 24h with doxycyline and harvested (MOCK) or subsequently infected with $SINV_{nsp3-mScarlet}$ for 18h (SINV) with MOI 0.1. Antibody against DDX1 indicated endogenous DDX1 and the heavier DDX1-GFP proteins, expressed from doxycyline treatment. Further antibodies used included: RTCB, RFP (as proxy of the viral tagged Nsp3-mScarlet protein) and SINV Capsid.

B HEK293 FITR DDX1-K52-GFP cell lines were seeded and grown with or without doxycyline (uninduced labelled ctrl, and induced lablled +dox) for 24 hours. Cells were infected with $SINV_{mCherry}$ (mCherry) and $SINV_{nsp3-mScarlet}$ (mScarlet) at an MOI of 0.1. Fluorescence measurements were taken every 15 minutes for 24 hours by BMG Clariostar plate reader. Fluorescence intensity was normalised as follows: minimum signal level set to 0, and maximum signal set to 10000 for the control uninduced wells, and doxycyline induced wells were normalised relative to control. Error bars represent standard deviation across the 3 technical replicates from each condition, and 3 biological replicates from independent experiments. Statistical significance was tested using two-tailed T-test compared to uninduced control (P $\leq$0.001 ***, P $\leq$0.01 **, P $\leq$0.05 *, P $\leq$0.1, ns: not-significant) at 16hpi and 24hpi.

are difficult to control and identify.

It was still unknown whether SINV infection had any effect on the tRNA-LC activity. To this effect, I measured the ligation of an *in vitro* transcribed substrate incubated with cellular extracts from uninfected and SINV-infected cells. This assay enabled the discovery of RTCB in 2011 [197], whereby cell lysates were incubated with 3'-phosphorylated (3'P), 5'-OH dsRNA molecules, which, in the presence of RTCB, became covalently linked (schematic in Figure 5.8A). This work was carried out in collaboration with the Martinez laboratory, particularly Dr. Stefan Weitzer, who performed the ligation assay. The radiolabelled dsRNA was incubated with cell lysates from 5 conditions: mock, mock 18h, SINV 4hpi, SINV 8hpi, and SINV 18hpi. The resulting samples were visualised by denaturing gel electrophoresis (Figure 5.8B). I observed

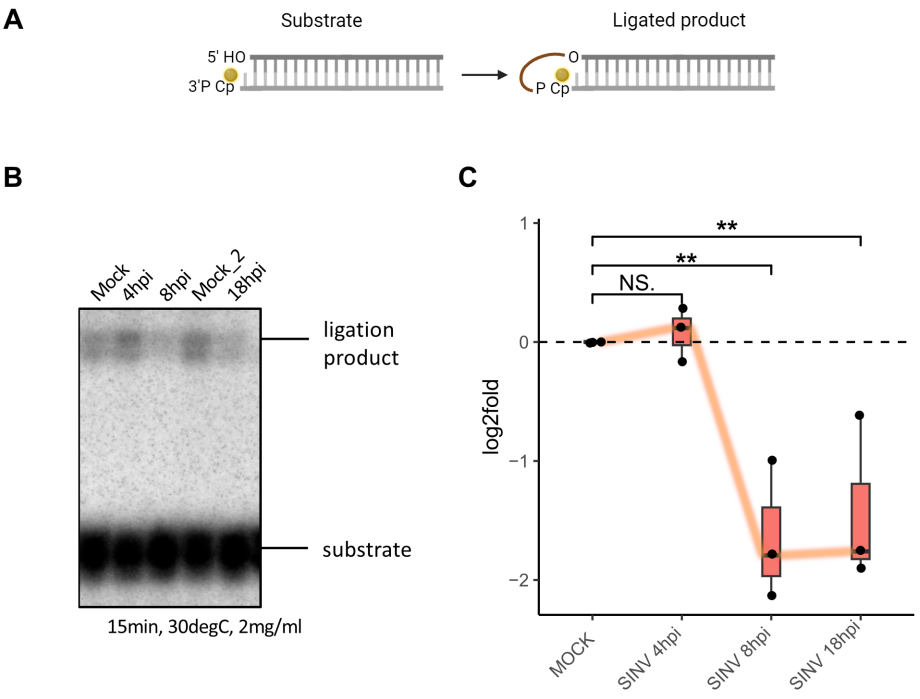

Figure 5.8: **tRNA-LC ligase activity inhibited at late stages of infection**

A Schematic of dsRNA substrate with a 5'-OH and 3'-P RNA oligonucleotides (in gray; with yellow marking the radiolabel), which becomes covalently linked after incubation with cell lysates containing RTCB ligase. **B** Gel electrophoresis image of 2mg/ml HEK293 wild type lysates collected at 4, 8 and 18h post infection and control mock samples collected at points of infection, incubated with dsRNA substrate for 30min. Unligated substrate travels to the bottom of the gel, whilst the interstrand ligation travels higher in the gel (labelled). **C** Relative quantification of signal from interstrand ligation products at each timepoint from 3 biological replicates. Log2fold change calculated relative to respective mock control samples and interlinked to show kinetic. Statistical significance was tested using two-tailed T-test comparing to respective mock samples (P $\leq$0.001 ***, P $\leq$0.01 **, P $\leq$0.05 *, P $\leq$0.1, N.S: not-significant)

that the levels of the ligated product increased slightly at the start of the infection (4hpi). However, as the infection progressed, the amount of ligated substrate significantly decreased (Figure 5.8C). This suggested that the tRNA-LC is inhibited in the late stages of SINV infection. As the inhibition of ligase activity is observed only at the later time points, this effect could be linked to the increase of vRNA levels in the lysates. Since the tRNA-LC binds to vRNA, it might get trapped on the vRNA and subsequently be unable to reassociate with new RNA, in this case, the radiolabelled RNA substrate. DDX1 and RTCB have both previously been characterised to co-localise with SINV viral factories [113] , which may correlate with vRNA saturation. The following chapter will explore the specific RNAs bound by the tRNA-LC in mock and infected conditions. Alternatively, the tRNA-LC is susceptible to oxidative inactivation [269].

During infection, an increase in reactive oxygen species may inhibit the ligase activity of RTCB.

5.2.4 Transcriptomic changes induced by tRNA-LC knockdown in uninfected and infected conditions

The observed tRNA-LC antiviral function may be due to cellular changes in the transcriptome induced by alteration of the abundance of its components. To investigate whether the knockdown of the tRNA-LC, namely DDX1 and CGI99, reflected a transcriptome-wide change in uninfected and infected cells, I next carried out an RNA sequencing experiment. Firstly, I focused on the mRNA changes that occurred upon siDDX1 or siCGI99 transfection in uninfected cells by comparing these samples to the siCTRL. Secondly, I sought to investigate the SINV-specific mRNA changes in each of the siRNA backgrounds by determining the transcriptomic changes occurring with these siRNAs in SINV-infected cells.

RNA was isolated from HEK293 cells transfected with siCTRL, siDDX1, and siCGI99. These RNA samples were harvested in parallel to the protein samples previously analysed in Figure 5.4 , where I observed a significant knockdown of the target proteins DDX1 and CGI99 and a significant increase in viral proteins nsP3 and Capsid. I prepared sequencing libraries specifically enriching mRNA via oligo d(T) capture and then sequenced them using a state-of-the-art Illumina sequencer, NextSeq 550.

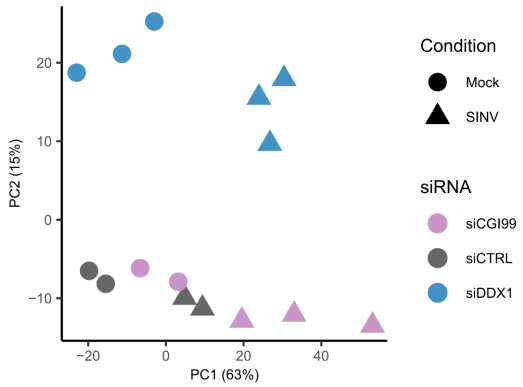

Figure 5.9: **Principal Component Analysis of siRNA RNA sequencing data** PCA of mRNA sequencing data from siRNAs targeting siDDX1, siCGI99 and siCTRL (scramble sequence) in mock or SINV conditions. The first two principal components were plotted as a 2D scatter plot, with the percentage of variance explained by each component given in brackets in the axes' titles. siCGI99 samples are represented in pink, siCTRL in grey and siDDX1 in blue. Circles represent uninfected mock samples, while triangles represent SINV-infected samples.

To increase the quality of my analysis and to capture genuine changes in the transcriptome,

I used an RNA spike-in pool, which allowed for accurate normalisation. During SINV infection, global cellular RNA degradation [113]. Therefore, standard normalisation to housekeeping genes does not accurately reflect the transcriptome changes [270]. An average of 18 million 75bp single reads per sample was obtained with a Q30 score over 94%. The obtained high-quality sequencing data was analysed with the help of the Bioinformatics Group at the CVR, particularly Srikeerthana Kuchi. Unfortunately, three samples had insufficient reads, failed our quality checks, and were omitted from the downstream analysis. To assess the quality of the data, I performed a principal component analysis (PCA) (Figure 5.9). The siDDX1-transfected samples clustered separately from the other groups, and they also separated depending on whether the cells were infected or not. The high divergence of siDDX1 samples over the other samples suggested a significant change in the transcriptome upon depletion of DDX1 that did not occur in the other conditions. Interestingly, siCGI99 samples clustered closer to the siCTRL samples than to the siDDX1 ones. The siCTRL samples did not separate greatly in the PCA. The lack of significant separation suggests that differences between the uninfected and SINV-infected samples were minimal. Previous transcriptomic experiments in the Castello lab were done with MOI 1 and 10, but here, I used MOI 0.1 to maximise the effects of the knockdown as well as have a comparable dataset to the transcriptomic experiments performed in the HEK293 shDDX1 cell lines mentioned earlier. Unfortunately, due to technical differences and analysis pipeline, namely the use of RNA spike-in controls for normalisation, the direct comparison between the two transcriptomic datasets was not possible. The substantially lower MOI is expected to result in a large proportion of non-infected cells, which can explain this phenomenon. Conversely, larger transcriptomic changes were observed between uninfected and infected cells when DDX1 or CGI99 were depleted, suggesting that the lack of these proteins benefits viral replication and spread, increasing the transcriptome differences between mock and infected cells.

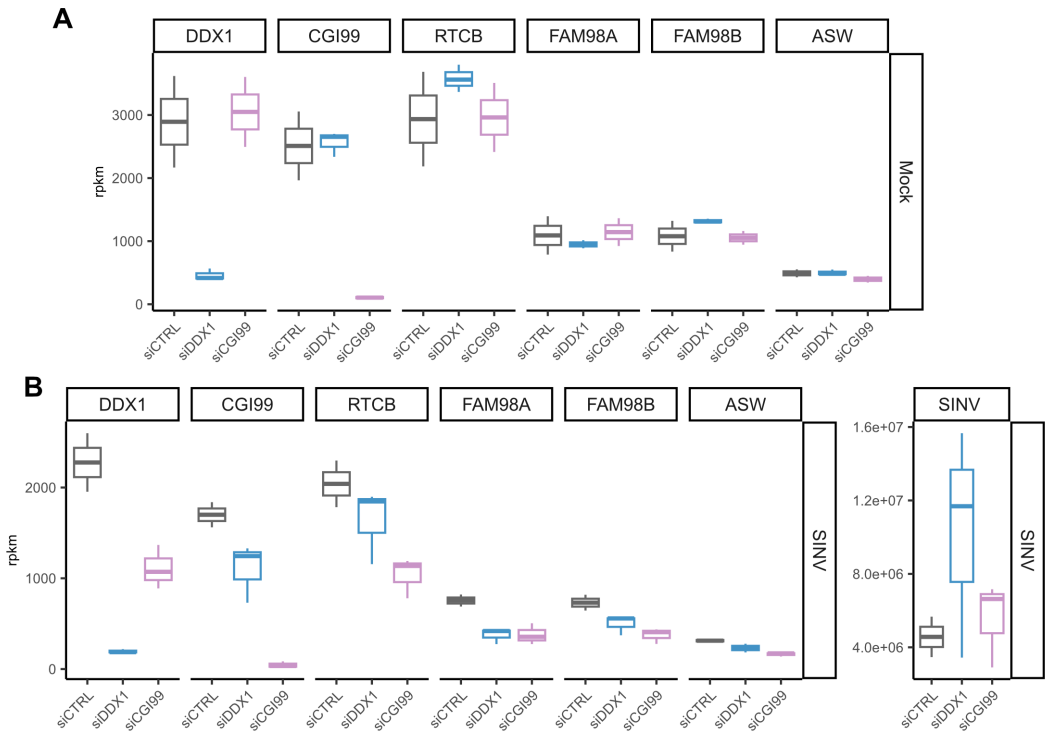

Figure 5.10: **tRNA-LC gene expression in varying conditions**

Normalised mRNA read counts obtained in the respective sample group: siCTRL, siDDX1 and siCGI99 in either **A** Mock or **B** SINV conditions. tRNA-LC mRNAs plotted as well as mRNAs aligning to the SINV genome in (B).

To further assess sample quality, I investigated the reads mapping to components of the tRNA-LC in each group after normalisation (Figure 5.10). As expected, there were minimal read counts for *DDX1* and *CGI99* mRNA in the samples transfected with the siDDX1 and siCGI99, respectively. The mRNA levels of other members of the tRNA-LC (i.e. *RTCB*, *FAM98A/B* and *ASW*) were not altered compared to mock cells (Figure 5.10A). This indicates that the previously observed loss of tRNA-LC proteins upon DDX1 and CGI99 knockdown was not transcriptional but likely due to changes in protein stability. In SINV-infected cells, however, a reduction in tRNA-LC mRNA levels was observed when compared to the siCTRL levels. The degradation of the tRNA-LC mRNAs can reflect the general transcriptome degradation occurring during DDX1/CGI99 knock-down enhanced SINV infection. Notably, I observed increased *SINV* vRNA in the siDDX1 samples and, to some degree, in the siCGI99 samples compared to the siCTRL which is consistent with the increased protein levels previously observed.

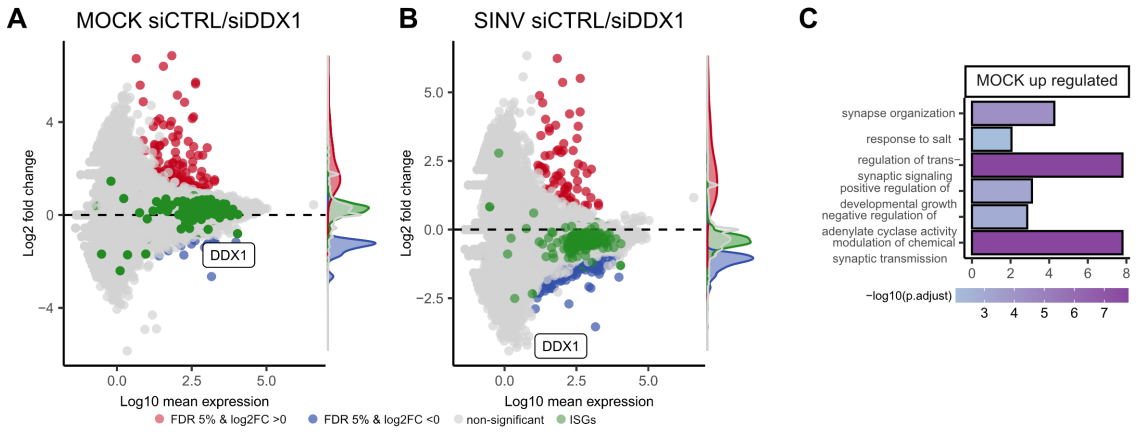

Figure 5.11: **Comparative analysis of siCTRL and siDDX1 samples in mock and SINV**

Comparative analysis of siCTRL and siDDX1 samples in **A** Mock or **B** SINV conditions. MA plots of differentially expressed genes in each condition. Downregulated genes labelled in blue, upregulated labelled in red and not-significantly changed labelled in grey. ISGs labelled in green in each MA plot (list of genes from [Chen et al. 2025, in preparation]). **C** GO enrichment analysis of downregulated genes in SINV siCTRL/siDDX1 comparison. Enrichment analysis was performed with clusterProfiler and overlapping GO terms were removed, using a similarity threshold of 0.25 for Biological Processes analysis. The colour of each bar indicates the significance of the enrichment, with darker purple indicating more significant enrichment.

To assess the impact of DDX1 absence in transcriptome changes, I analysed the differential gene expression in siCTRL and siDDX1 samples in both mock and SINV infected conditions (figure 5.11A and B). In addition to plotting significantly differentially expressed genes, I represented ISGs (plotted in green). In both the Mock and SINV conditions, ISG expression remained unchanged probably due to the low penetrance of infected cells with an MOI 0.1. In uninfected cells, the lack of changed ISGs in DDX1 knockdown compared to siCTRL suggests that DDX1 does not regulate ISG levels . Typically, viral infection leads to changes in ISG regulation [38, 40]. However, the absence of significant changes here likely reflects similar levels of stimulation in both the SINV siCTRL and SINV siDDX1 samples, resulting in no discernible difference. This could indicate that ISG expression levels are equivalent in the siCTRL and siDDX1 conditions during SINV infection.

In Mock, only 16 genes were downregulated, with the most significant being *DDX1* as expected. However, in SINV, I observed 442 downregulated genes, which may reflect that RNA degradation is triggered in the subpopulation of infected cells in the culture. Aspects of SINV-specific changes will be explored in the specific mock and infected comparison within each siRNA background later on.

Data shows a large number of upregulated genes in mock and SINV-infected cells upon knockdown of DDX1. A comparison of both sets revealed that 98% of the genes upregulated in SINV and mock cells are shared between the two conditions. This implies that these changes are inherent to the absence of DDX1 in the cells. I further investigated these genes by a GO enrichment analysis (Figure 5.11C). I identified that a large proportion of these genes are involved in synaptic signalling and organisation.

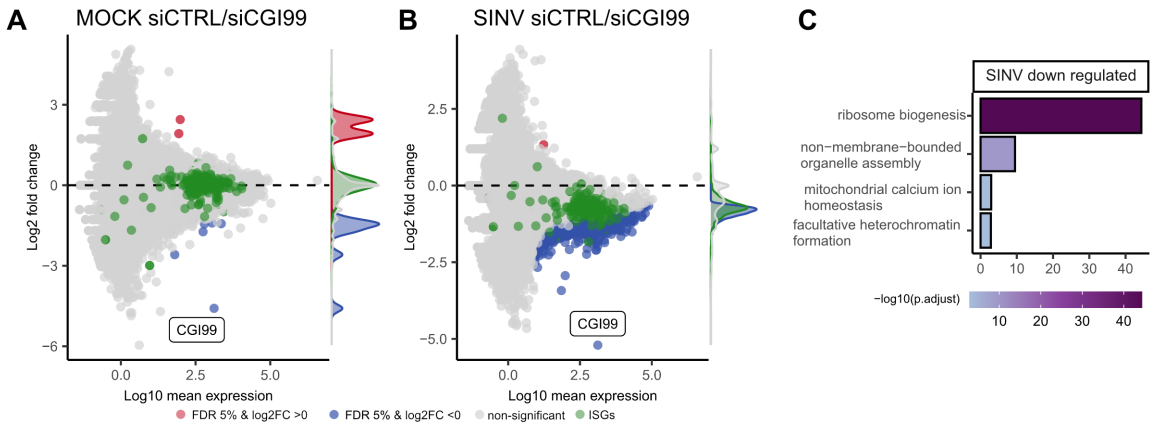

Figure 5.12: **Comparative analysis of siCTRL and siCGI99 samples in mock and SINV**

Comparative analysis of siCTRL and siCGI99 samples in **A** Mock or **B** SINV conditions. MA plots of differentially expressed genes in each condition. Downregulated genes labelled in blue, upregulated labelled in red and not-significantly changed labelled in grey. ISGs labelled in green in each MA plot (list of genes from [Chen et al. 2025, in preparation]). **C** GO enrichment analysis of downregulated genes in SINV siCTRL/siCGI99 comparison. Enrichment analysis was performed with clusterProfiler and overlapping GO terms were removed, using a similarity threshold of 0.25 for Biological Processes analysis. The colour of each bar indicates the significance of the enrichment, with darker purple indicating more significant enrichment.

To assess the impact of CGI99 knockdown on transcriptomic changes, I analyzed differential gene expression in siCTRL and siCGI99 samples under both mock and SINV-infected conditions (Figure 5.12A and B). As with the siDDX1, significantly differentially expressed genes (DEGs) were plotted alongside ISGs (highlighted in green). In the mock condition, ISG expression remained largely unchanged, indicating an absence of cellular antiviral response due to the absence of CGI99. However, during SINV infection, a global down-regulation of ISGs was observed, with 57% of the listed ISGs significantly reduced in siCGI99 compared to siCTRL. This effect contrasts with the siDDX1 results (Figure 5.4).

To further investigate these transcriptomic changes, I performed GO enrichment analysis on downregulated genes in SINV-infected conditions (Figure 5.12C). Many of these genes

were associated with ribosome biogenesis and non-membrane-bounded organelle assembly, potentially linked to viral processes. Unlike DDX1 knockdown, the absence of CGI99 caused minimal changes in the transcriptome in mock conditions. Few DEGs were identified, with only two upregulated genes, one being *RAB11B*, a gene associated with synaptic function. The results with mock cells suggest that CGI99 knockdown does not cause ISG downregulation per se, and the virus faces a cell with a similar transcriptome immediately upon infection.

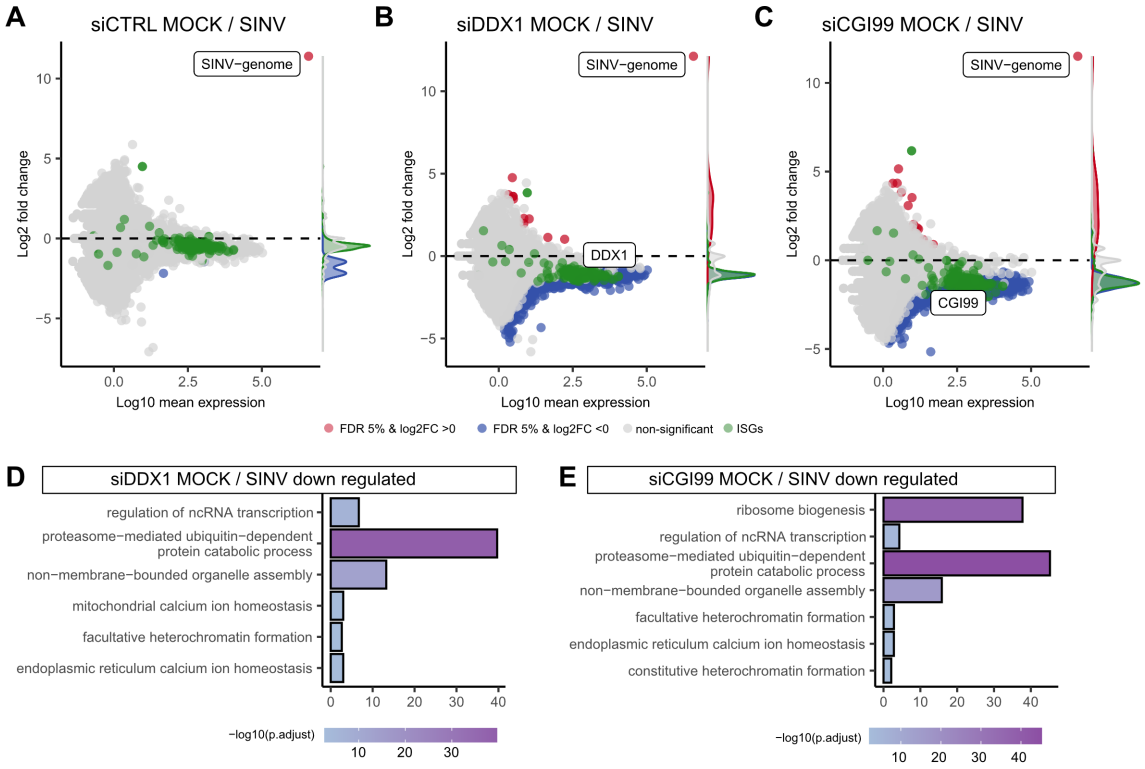

Figure 5.13: **SINV-specific transcriptome changes in each siRNA background**

Comparative analysis of Mock and SINV infected samples in **A** siCTRL **B** siDDX1 and **C** siCGI99. MA plots of differentially expressed genes in each condition. Downregulated genes labelled in blue, upregulated labelled in red and not-significantly changed labelled in grey. ISGs labelled in green in each MA plot (list of genes from [Chen et al. 2025, in preparation]). **D and E** GO enrichment analysis of downregulated genes in siDDX1 (D) and siCGI99 (E) Enrichment analysis was performed with clusterProfiler and overlapping GO terms were removed, using a similarity threshold of 0.25 for Biological Processes analysis. The colour of each bar indicates the significance of the enrichment, with darker purple indicating more significant enrichment.

To study the SINV-specific changes that occur in the absence of DDX1 and CGI99, I investigated the DEGs in Mock and SINV conditions for each of the siRNA targets. The analysis of uninfected and infected samples in each siRNA background permits the study of SINV-specific changes. This analysis negates to some degree the mRNA changes inherent to the siRNA knockdown, such as the upregulated genes identified in the siDDX1 compared to

siCTRL.

In the siCTRL comparison between mock and infected samples, I observed no significant DEGs (Figure 5.13A). However, the overall ISGs are moderately downregulated in infection, as noted by their shift below the median line. The lack of significant changes can be due to the lower infection occurring in the siCTRL samples. I used an MOI 0.1, meaning that at harvest, the isolated bulk RNA may only represent a small fraction of infected cells. Thus, the lack of DEGs is obscured by a predominant uninfected background. Furthermore, in the siCTRL sample set, I only had two biological replicates in each condition, which could impact the statistical significance of the siCTRL comparison. I observed the significant upregulation of the SINV vRNA labelled "SINV-genome", although this is widely expected as no SINV RNA would be present in the mock conditions.

In the mock and SINV infected samples comparisons within the siDDX1 and siCGI99 background, I observed over 12126 and 11128 downregulated genes, respectively (Figure 5.13B and C). A significant number of ISGs, over 80% are represented among the downregulated genes in both siDDX1 and siCGI99 comparisons. Notably, the overall mRNA profile is shifted below the median line, which is a characteristic of host mRNA degradation that occurs during virus infection. The most significantly upregulated gene is the SINV genome. I performed a GO enrichment analysis on the downregulated genes in both siDDX1 and siCGI99 (Figure 5.13D and E). A significant proportion of the genes were shared between the two siRNAs, which translated to shared GO terms such as 'regulation of ncRNA transcription', 'proteasome-mediated ubiquitin-dependent protein catabolic process' and 'non-membrane-bounded organelle assembly'. The identification of genes associated with non-membrane bound organelles may be very relevant in the context of SINV infection, which is known to destabilise stress granules, for example [271].

Overall, the differences between the siCTRL, siDDX1 and siCGI99 comparisons in Mock and SINV-infected samples were striking. While the infection in siCTRL cells indicated no significant changes in expressed genes, the infection in siDDX1 and siCGI99 cells downregulated over 11000 genes. The significantly higher differential expression observed in the siDDX1 and siCGI99 samples can reflect the more favourable infection environment in the absence of the two tRNA-LC proteins. Among the down-regulated genes are a large number of ISGs, which reflects the inhibition of innate immunity and type I interferon response.

In conclusion, I have shown that the effects of the tRNA-LC in infection are complex. Beyond the opposing phenotypes observed in relation to the experimental technique, the most consistent results were displayed by a siRNA pool targeting multiple tRNA-LC targets and Archease. While there is reasonable doubt about the roles of the tRNA-LC in infection, the siRNAs that produced the most reproducible and robust results suggest an antiviral effect. Interestingly, the ligase activity of the complex is inhibited at later stages, which may be linked to the antiviral phenotype observed. The exploration of the transcriptome changes in the absence of DDX1 and CGI99 indicated a significant downregulation of a remarkably large number of genes during SINV infection. These genes were predominantly involved in ribosomal biogenesis, proteasome-mediated ubiquitination, and non-membrane-bounded organelles. However, the most likely explanation is that the lack of DDX1 and CGI99 accelerates SINV infection, leading to the previously described degradation of cellular mRNAs in a larger proportion of cells than in the siCTRL conditions.

5.3 Discussion

The tRNA-LC proteins have recently been implicated in the SINV lifecycle [113]. DDX1, RTCB and FAM98A were identified as having increased RNA binding activity over the course of SINV infection and were further characterised as direct vRNA interactors [123]. However, their roles in SINV infection have not yet been identified. Here, I employed different knockdown techniques to elucidate what these roles may be. The results revealed method-derived differences. While shRNAs targeting DDX1 indicated DDX1 as a dependency factor in SINV infection, the results employing siRNAs targeting various tRNA-LC components consistently showed an antiviral phenotype, namely DDX1 and CGI99. Furthermore, the recycling of RTCB was shown to be detrimental to SINV infection, as illustrated by the depletion of Archease. The mRNA sequencing of the siRNA transfected cell lines targeting DDX1, CGI99 and control scramble sequence revealed a significant increase in infection characterised by the increased downregulation of over 11000 genes. While more experiments will be required to deconvolute the phenotype of the tRNA-LC in infection beyond a reasonable doubt, the consistency of the siRNAs, paired with their low secondary effect in the transcriptome, suggests an antiviral phenotype.

5.3.1 Technical difficulties in resolving tRNA-LC SINV phenotype

To evaluate DDX1's impact on virus fitness, two knockdown techniques were used: cell-stable inducible shRNA and transfected siRNA. Each technique produced different effects despite targeting the same sequence within DDX1: shDDX1 inhibited SINV, while siDDX1-1 with the same sequence increased viral output relative to siCTRL. These results are mirrored in studies on SARS-CoV-2 with the same sh/siRNA target sequences used (shDDX1 in [116] siDDX1-2 and siDDX1-3 in [193]). One study found that shDDX1 expression inhibited viral fitness in A549-Ace2 cell lines [116]. By contrast, another study reported the opposite results in HUH7-Ace2 cells transfected with siRNAs against DDX1 and infected with SARS-CoV-2 [193]. These results suggest that a technical element inherent to the knockdown approach influences the outcome of the infection. I would expect to identify similar patterns of inhibition in both SINV and SARS-CoV-2, which is why the conundrum in the sh/siRNA phenotype is intriguing. The disparity in phenotype observed in these published studies indicates that the obtained results in this chapter are not unique. The discrepancies identified indicate that an alternative knockdown system is required to truly study the behaviour of DDX1. Unfortunately, the complete knock-

out (KO) of DDX1 is not possible. DDX1 is an essential cellular protein, and the KO has been shown to be lethal in mouse embryos [272]. To improve on this experiment, in future work, I would seek to establish an inducible targeted degradation, such as the auxin-inducible degron [273]. The rapid protein depletion could aid in quick viral screens and circumvent the activation of undesirable pathways or technical biases. The complete removal of DDX1 could provide a clear-cut effect on SINV infection.

Beyond the technical application of the knockdown systems, I identified an additional consideration in interpreting siRNA KD results in my experimental system. siRNA transfections can activate cellular stress response pathways [263]. I observed that all transfected siRNA elicited a stress response, detected by the phosphorylation of eIF2-α. This may prime the cells to exhibit a different reaction during infection that may not singularly be from the protein absence but rather from an activated innate immunity response, interferon. This pathway activates the dsRNA-dependent PKR upstream of eIF2-α. Interestingly, SINV and SFV have been shown to be able to circumvent the translation initiation inhibition caused by the phosphorylation of eIF2-α through RNA secondary structures slowing ribosomes at the correct start site [126]. Furthermore, in a study on the impact of autophagy and eIF2-α phosphorylation during SINV infection [268], researchers found that eIF2-α phosphorylation is an essential step in inhibiting cellular translation and increases viral translation. Together, these studies could indicate that in my data, the activation of the PKR stress response by siRNA transfection correlates with increased viral fitness. However, the relative comparison between the siCTRL and the siRNA targeting tRNA-LC indicates this is not the case. The increased viral proteins detected in the siDDX1 and siCGI99 are, in fact, symptomatic of the knockdown of those specific proteins. Furthermore, the complete transcriptomic analysis performed in siRNA-transfected cells indicated that ISG stimulation does not occur in steady-state conditions and, subsequently, is not a factor when analysing the SINV-infected microenvironment. The antiviral effect observed by the knockdown of DDX1 and CGI99 is robustly portrayed as inherent to the absence of these key tRNA-LC proteins rather than an off-target effect caused by transfection stress.

5.3.2 The antiviral role of the tRNA-LC

Isolating the specific behaviour and effects of individual members of the tRNA-LC presents significant challenges. As discussed in Chapter 4, targeting one protein often has substantial

impacts on the stability of other components within the complex. This inter-protein dependency was evident during individual siRNA targeting experiments, which caused knock-on effects on the other proteins in the tRNA-LC. In the context of SINV infection, the same occurs. Interestingly, the mRNA of tRNA-LC components remains stable in the absence of one protein, indicating the dissolution of the complex components occurs at the protein level. In SINV-infected samples, siDDX1 and siCGI99 treatments significantly increased the relative protein levels of nsP3. Notably, the pronounced knockdown of DDX1 by siCGI99 could suggest that the antiviral effects seen in the absence of CGI99 may be attributed to the loss of DDX1, a protein more extensively characterised in viral infections.

The transcriptome of siDDX1 transfected cells indicated an upregulation of genes involved in synaptic functioning. DDX1 and the tRNA-LC have previously been characterised to play a role in synaptic function in neurons [246]. The tRNA-LC is a component of cytoplasmic mRNA-transporting granules that are kinesin-associated in dendrites. These granules are involved in transporting specific mRNAs from the cell body to dendrites, allowing for local mRNA translation at sites distant from the nucleus. This process is crucial for synaptic plasticity and function [246]. The upregulation of these pathways in the absence of DDX1 could indicate a compensatory mechanism to maintain cellular homeostasis.

During SINV infection, a significant remodelling of the transcriptome is observed in siDDX1 and siCGI99 transfected cells. The knockdown of DDX1 and CGI99 instigated a significant downregulation of over 11000 genes. The genes are predominantly associated with non-membrane-organelles. In addition to the tRNA-LC involvement in the non-membrane structure involving mRNA-transporting in dendrites [246], DDX1 has further been characterised to associate with stress granules (SGs) in response to environmental stressors [206]. DDX1 has been observed to co-localise with SG markers such as G3BP1. The interaction between these two proteins has been further characterised as protein-dependent rather than via an RNA stabilising bridge [206]. SGs play a crucial role in viral progression [274]. SGs are part of the cell's integrated stress response, forming as a consequence of translation inhibition during viral infection. They act as cytoplasmic RNA-protein complexes that can suppress vRNA translation, potentially limiting viral replication through sequestration [274]. However, alphaviruses have evolved strategies to manipulate SG formation to their advantage. SINV for example, hinders the ability of vertebrate cells to form SGs via host cellular shut-off [271]. I speculate that the absence of tRNA-LC proteins during SINV infection further hinders the

formation of SGs, enabling SINV to progress faster through the cell.

DDX1 has previously been shown to directly stimulate IFN-β production in a variety of RNA viruses such as TGEV and FMDV [166, 194]. In IAV infection DDX1 has been proposed to form an alternative complex with DDX21 and DHX36, which binds to dsRNA in the cytoplasm. Upon dsRNA sensing, the complex interacts with the TRIF pathway to activate type I IFN [178]. The suppression of DDX1 in all of these studies significantly decreased the ability to mount an interferon response. In my study, the suppression of DDX1 during SINV infection, significantly downregulated ISGs. I hypothesise that the absence of tRNA-LC proteins, specifically DDX1, inhibits innate immune response pathways such as IFN. In the next chapter, I explore the DDX1 protein-protein interaction changes during infection, which could indicate the association of the tRNA-LC with these antiviral complexes.

5.3.3 The importance of the ligase and helicase activities of the tRNA-LC in SINV

The primary role of the tRNA-LC, is the unusual ligation of RNA with 3' ends containing a 2',3'-cyclic phosphate (2',3'-cP) with an RNA fragment with a 5'-hydroxyl (5'-OH) [197]. This role is orchestrated by the function of the RTCB ligase in coordination with DDX1 and Archease. The ligase activity of the tRNA-LC is essential in tRNA maturation as well as the ligation of *XBP1* during UPR [197, 203]. I aimed to understand whether this central function of the complex is important during viral infection. Using an *in vitro* ligation assay, I observed that the ligase activity of the tRNA-LC increases 4h upon infection, followed by a substantial inhibition at later time points. An exciting hypothesis is that RTCB may not be able to engage with the substrate to catalyse its ligation if it is saturated by binding vRNA. Indeed, RTCB has been reported to interact directly with SINV RNA [123], and vRNA represents over 70% of cellular polyadenylated RNA at 18hpi [113]. The high abundance of vRNA and its high concentration at viral replication organelles may potentially overwhelm RTCB, which is known to relocate to these replication foci [113]. The exact RNAs bound by RTCB will be further discussed in the following chapter.

An alternative explanation for the lack of ligation during the above assay is the lack of enzymatic turnover. RTCB may not be undergoing normal guanylation due to a lack of GTP/GMP conversion by Archease and DDX1 [198]. Normal recycling is most likely still occurring early upon infection correlating with the high ligation activity detected at 4hpi. GTP is an essential molecule in the alphavirus life cycle, and its depletion has been shown to hinder

SFV and CHIKV infection by inhibiting nsP1 capping of vRNA [275–277]. In the context of RTCB recycling, it is plausible that cellular GTP is exhausted by viral processes such as translation, leading to a reduction in RTCB recycling at later times post-infection. A repetition of the ligation assay in the presence of different doses of GTP could shed light on this point.

The strong antiviral phenotype of Archease stands in stark contrast to RTCB's lack of phenotype in infection. Archease has solely been identified as a co-factor to RTCB and the tRNA-LC ligase activity [198]. An exciting future direction of this project would be to explore the infection-specific roles of Archease and how it contributes to its pronounced antiviral effect. An important next step would be to identify the protein interaction partners of Archease following SINV infection to determine whether there is a reduction in engagement with RTCB and whether there is increased engagement with other factors that could promote its antiviral function.

The other enzymatic protein in the tRNA-LC is DDX1, whose helicase activity may be functionally important for SINV fitness. Helicases are known to be essential modulators of viral infections [278]. DDX1 may be essential for the unwinding of SINV vRNA during infection to enact its antiviral role. The DDX1-K52A mutant replaces endogenous DDX1, but as with other helicases, the mutation is expected to function as a dominant negative [279]. The overexpression of the catalytic mutant, as opposed to the wild-type DDX1, reduced the the level of viral proteins, although the effect is very mild in the plate reader assay. This suggested that the DDX1-K52A mutant inhibits viral fitness. The catalytic mutant may still assemble into the tRNA-LC. However, the lack of ATPase activity could render the complex inactive overall. Without DDX1 ATP hydrolysis, the tRNA-LC might become unable to function due to "molecular poisoning". Dominant negative phenotypes have previously been observed in proteins that assemble as part of protein complexes, and these mutant subunits can effectively “poison” the assembly [280], as might be the case with the tRNA-LC. DDX1 is also involved with Archease in RTCB recycling [198]. However, the knockdown of Archease induces stimulation of virus infection that is antagonistic with the effects observed with the DDX1 point mutant. Thus, I favour the hypothesis that the ATPase mutant essentially "poisoned" the tRNA-LC assembly and function. To better understand the dynamics of the tRNA-LC complex with the integration of DDX1-K52A and their combined effect in SINV, further investigation is required. Determining if the mutant assembles with the other tRNA-LC components, if the tRNA-LC binds to the vRNA, and whether molecular poisoning causes toxic effects for the cell will shed light on the

phenotype observed.

The previous chapter highlighted the importance of CGI99 in the tRNA-LC structure and stability of the complex. This chapter did not explore CGI99's catalytic activity. However, future work on CGI99's involvement in translation initiation would be interesting, given its antiviral role.

In conclusion, the tRNA-LC antiviral role can stem from a large number of cellular roles, from stimulating IFN production to non-membrane organelle formation, specifically SGs. The next chapter explores the protein-protein interaction of DDX1 during SINV infection, which could further elucidate the origin of its antiviral role.

6 Elucidating the interactome of DDX1 in SINV infection

6.1 Introduction

DDX1 is a nuclear and cytoplasmic protein with diverse cellular functions which have been characterised as part of the tRNA-LC, such as translation activation [246], tRNA maturation [198] or unfolded protein response [203]. Meanwhile, other processes have been associated with DDX1 specifically, such as fatty-acid-dependent insulin regulation, R-loop formation [247, 281], rRNA processing [282], double-strand break repair [283] and immunoglobulin class switching [247]. Furthermore, DDX1 and tRNA-LC proteins have been implicated in inhibiting or facilitating viral infections including HIV, SARS-CoV2 and VEEV [116, 185, 189, 192, 193]. The diverse roles of the tRNA-LC and DDX1 independently have portrayed these proteins as essential RNA regulators.

In Chapter 4, I observed the close interaction of the tRNA-LC proteins and their protein co-dependence in complex stability. I further captured direct interaction with ribosomal-associated proteins implicating the complex in translation regulation. Previous studies have observed two tRNA-LC proteins, namely DDX1 and CGI99, as essential modulators in translation control. Indeed, DDX1 has been identified as binding insulin mRNA and regulating the translation of the protein through its binding in the 5'UTR [259]. Although tRNA-LC proteins were not identified in this study, it is possible that they are also present as part of this functional role. The researchers employed RNA antisense purification coupled with mass spectrometry (RAP-MS) in INS-1 cells. Only selected bands separated by polyacrylamide gel electrophoresis were analysed by MS, which means tRNA-LC proteins of different molecular weights to DDX1 were not captured in this method [259]. Furthermore, CGI99 has been captured as interacting with mRNA cap structures as part of the tRNA-LC, indicating its potential to regulate translation initiation in this region [246]. Altogether, the tRNA-LC may have broad involvement in protein expression and translation control.

A comprehensive study of the DDX1 interactome in rat cells was carried out in the context of alternative splicing (AS) events in pancreatic β cells that lead to insulin secretion [284]. Zhong et al. (2018) employed high-throughput RNA sequencing, CLIP-Seq and co-IP of DDX1. They identified hundreds of alternative splicing genes that are targeted by DDX1. DDX1 was observed as interacting with the spliceosome and regulating AS events in protein-coding sequences. The absence of DDX1 caused the skipping of the spliced region [284]. This

novel regulation by DDX1 explored in this paper expanded our understanding of DDX1-RNA modulation. This study highlighted how the binding location of a given RBP to its target RNA can be essential for its subsequent function.

In the context of DDX1, the cellular localisation of the protein is crucial in its functional implication. In the nucleus, DDX1 has been observed to modulate DNA/RNA hybrids in R-loop formation [281]. DDX1 was captured in the IP of the RNA exosome subunit, EXOSC3, in a neuronal cell line (N2A). Although both proteins are nuclear and cytoplasmic, the fractionation of the IP indicated a compartment-specific interaction between EXOSC3 and DDX1 in the nucleus. The interactome of cellular DDX1-protein and -RNA interactions is crucial in understanding the implication of the protein in relation to cytoplasmic viruses.

In relation to viral infection, DDX1 has been observed to have a multitude of pro- or anti-viral roles in different viruses. DDX1 has been observed to interact directly with viral proteins such as Rev in HIV-I infection [189–191], nsP3 in VEEV and SINV [123, 185] and nsP14 in SARS-CoV and TGEV [166, 192]. These host-virus protein-protein interactions allude to close involvement in the viral lifecycle. Beyond viral protein interactions, it is essential to capture the wider RNP complexes to understand the functional outcome of these interactions. For example, in the VEEV DDX1-nsp3 interaction, researchers focused on the most abundant interactors, DDX1 and DDX3, excluding the further analysis of the ribosomal proteins also identified [285]. The co-precipitation of all these proteins can allude to a cohesive RNPs complex where nsP3 interacts with the host translational machinery through DDX1.

The tRNA-LC's RNA-binding properties have further implied the complex in the direct regulation of vRNAs. Indeed, in the comparative analysis performed by Iselin et al. (2022), where different viral interactome capture methods were balanced against one another, DDX1 appeared in every study [115]. This indicated that DDX1 is a direct interactor with the vRNA of SINV, CHIKV, ZIKV, DENV, and SARS-CoV-2. The functional implications of these interactions require further study and will be explored in this chapter.

The involvement of the tRNA-LC in cytoplasmic processes, particularly in translation control, appears to be context-dependent and warrants further exploration. To date, studies examining the interaction landscapes of tRNA-LC with proteins and nucleic acids have been pivotal in elucidating its cellular functions. However, the diversity of these interactions poses challenges in predicting the mechanisms by which this complex restricts SINV infection. I previously explored the tRNA-LC's intricate structure and its protein components'

interdependence in mediating antiviral functions. The antiviral role of the tRNA-LC in SINV can be further elucidated by investigating the protein-protein and protein-RNA interaction dynamics that occur during SINV infection. These interactions can provide critical insights into the functional mechanisms of the tRNA-LC. In this chapter, I delve into the broader DDX1 interactome to uncover the complexes DDX1 forms and their roles in viral infection. This includes identifying both DDX1-associated proteins and tRNA-LC RNA targets. To this effect, this chapter expands on the previous high-resolution complex analysis by XL-MS, exploring the landscape of interactions broadly established by DDX1 in the cell and the dynamics of these interactions following infection. Additionally, I investigate the RNA binding specificity of tRNA-LC in both cellular and viral RNA contexts.

6.2 Results

6.2.1 DDX1 interactors are altered by virus infection

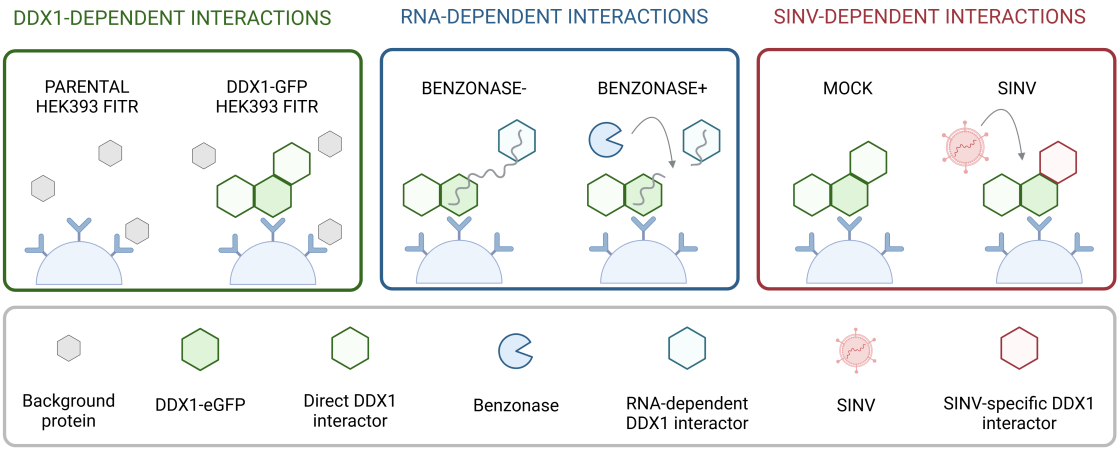

Figure 6.1: **Schematic of DDX1 comparative protein-protein interaction analysis**
Schematic adapted from Dr Louisa Iselin

DDX1 is an extremely versatile cellular protein with roles in SINV infection (as observed in Chapter 5). The composition of the tRNA-LC and their interfaces were established XL-MS of DDX1 (Chapter 4). Further novel cellular proteins involved in translation and cellular transport were also captured, alluding to interactions with cellular proteins and complexes beyond the tRNA-LC itself. XL-MS provides high-resolution data on protein-protein interactions at a cost of depth. The capture of a wider protein-protein interactome could help identify which interaction partners and, by extension, which functions might be important for the tRNA-LC's antiviral activity. To address this, I analysed DDX1 interaction network by performing an IP with the high affinity and specificity GFP-trap IP using the HEK293 FITR cell line expressing DDX1-GFP (Chapter 4). The designed experiment assessed the DDX1 protein network on three levels, summarised in Figure 6.1.

Firstly, I sought to include the controls that would help me to separate bona fide and spurious interactors of DDX1. I performed the IP in both HEK293-DDX1-GFP and HEK293 FITR (parental line) as a negative control. I did not use the HEK293-GFP lines as a control, as data from the Castello lab had previously observed how remarkably clean the IP was (LC-MS/MS data analysed). This may be due to the high expression levels of GFP and subsequent saturation of the GFP-trap agarose beads, probably preventing unspecific binding. The parental line, however, generated a protein pattern in silver stains that had shared bands

with the HEK293-DDX1-eGFP (data from Dr. Wael Kamel), thus representing a more stringent control.

The negative control could be further optimised to better aid in identifying true DDX1 interactors, accurately reflect the background binding capacity of the GFP tag, and ensure that any differential interactions are attributable specifically to DDX1. One improvement would be to use a GFP-expressing control cell line in which GFP expression is tuned to match the levels of the DDX1-eGFP fusion protein, which is achievable, for example, by adjusting doxycycline induction. Additionally, introducing a degradation signal (such as a degron tag) to GFP in control cells could prevent overaccumulation and more closely mimic the stability and turnover of the fusion protein. Furthermore, performing a comparable analysis of the input samples, specifically, the cytosolic fraction of the whole cell proteome, would strengthen the assessment of IP specificity. This would allow for normalisation against protein abundance in the starting material, helping to distinguish between specific enrichments versus highly abundant proteins that may non-specifically associate with the IP matrix. Without this, there is a risk that the IP profile merely reflects a small, diluted subset of the broader proteome, rather than a true enrichment of interaction partners.

Secondly, I wanted to assess whether I observed changes in interacting partners at 8h and 18h post-SINV infection when compared to mock conditions. The tRNA-LC is a nuclear and cytoplasmic complex. However, SINV replication occurs in the cytoplasm, and tRNA-LC members have been observed to migrate to viral replication organelles and interact with vRNA [113]. The IP was thus performed with cytoplasmic extracts to capture the virus-specific interactors. I thus sedimented the nuclei fraction prior to IP using a mild NP40/Igepal mediated lysis and performed the IP with the supernatant corresponding to the cytosolic fraction.

Finally, RNA is an important scaffold mediating interactors, and its removal is required to differentiate between protein-protein and RNA-bridged interactions. To do so, I split the samples into two aliquots that were treated or not with benzonase. Subsequent stringent washes were applied to remove the RNA-dependent protein interactors.

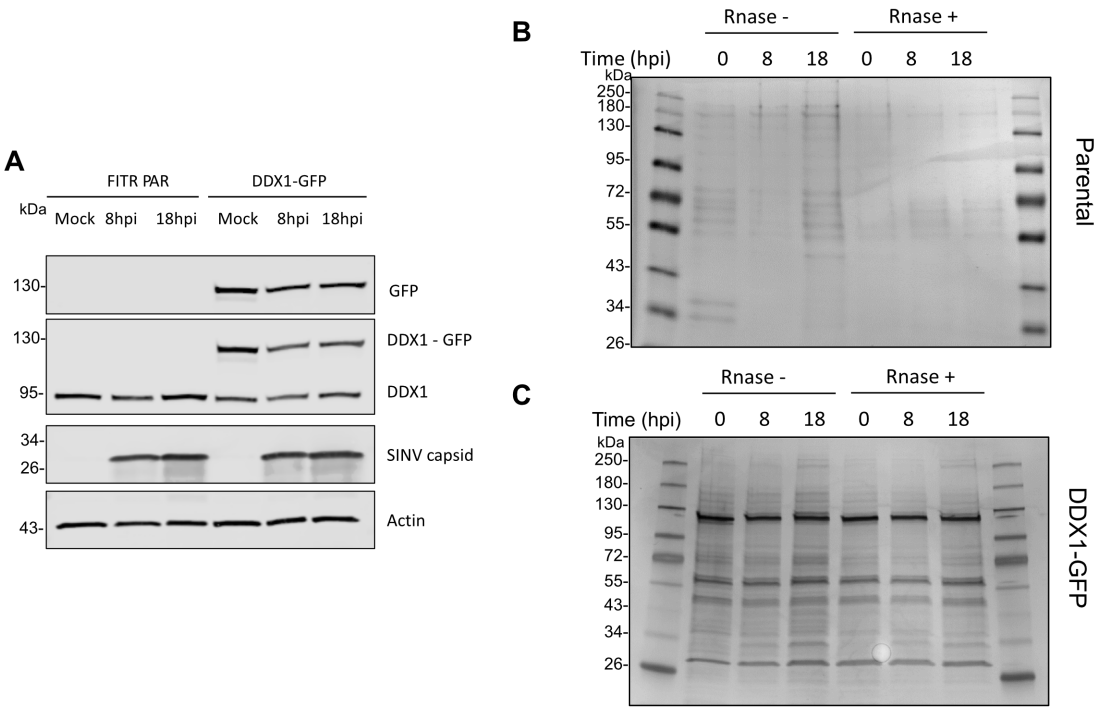

Figure 6.2: **DDX1-GFP protein/protein interactions**

A Western blot images of whole cell lysates from HEK293 FITR parental and DDX1-GFP cell lines in mock, 8h and 18h post SINV infection used for PPI. **B and C** Silver stains of IP eluates in Parental (B) and DDX1-GFP (C) samples. Samples were split during the IP, into RNase treated and untreated as indicated in silver stains.

To assess the quality of the eluates, I performed silver staining. Firstly, the DDX1-GFP samples showed substantially stronger and more complex protein patterns than the negative control, indicating a significant enrichment of proteins in the DDX1-GFP IP (Figure 6.2 B and C). The banding patterns of DDX1-GFP IP were remarkably similar, with the strongest band corresponding to DDX1-GFP itself. This confirms that the prevalent interactors of DDX1 remain consistent over the course of infection. The strongest bands are compatible with the protein sizes of the tRNA-LC, and the banding pattern throughout is similar to the IP silver stains observed in Chapter 4 for the tRNA-LC members (RTCB, CGI99, FAM98B and ASW), suggesting that the tRNA-LC remains compositionally unaltered throughout SINV infection. Overall, this shows that the IP resulted in the selective capture of DDX1-GFP and tRNA-LC components, together with other interacting partners.

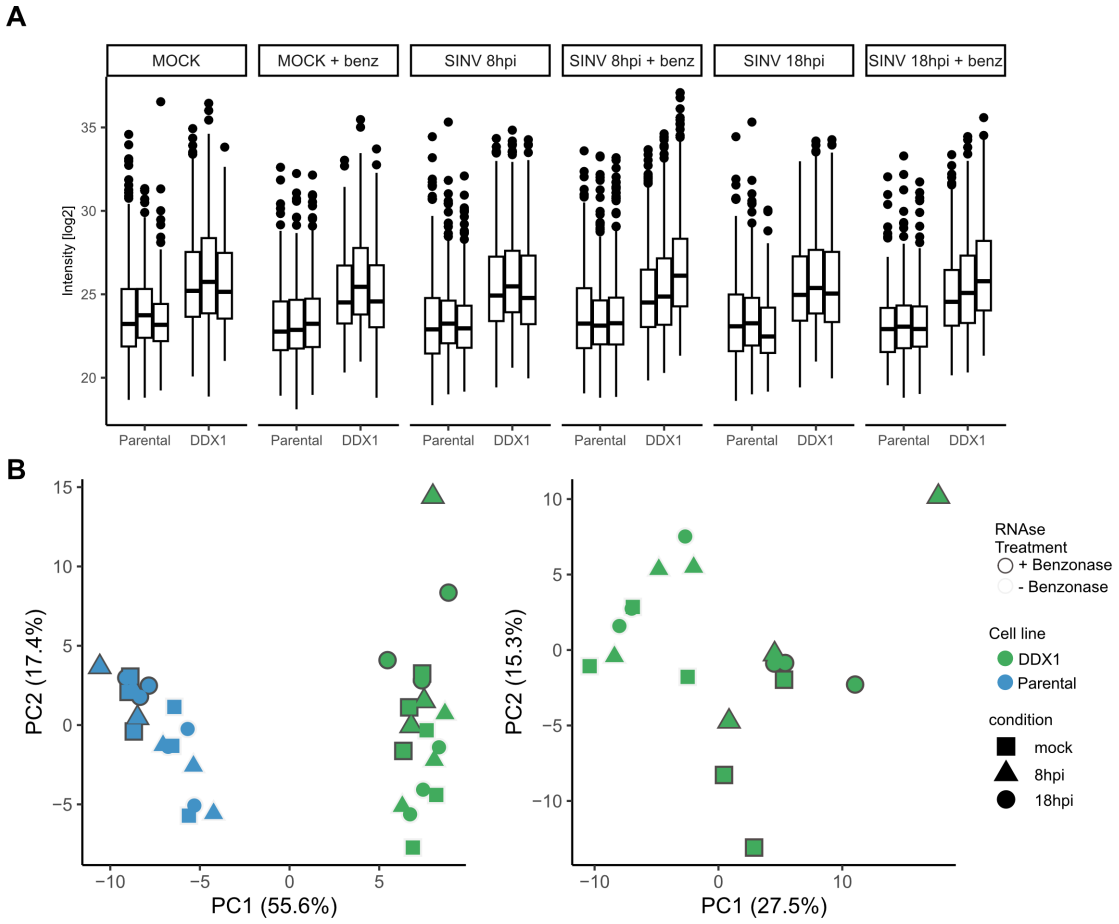

Figure 6.3: **Raw intensity values and PCA quality control of of DDX1 IP LC-MS/MS data**

A DDX1 IP samples were analysed by LC-MS/MS and peptides/proteins were identified and quantified in MaxQuant. Boxplot showing the log2 intensity distribution of triplicate samples for each condition without normalisation. **B** PCA of DDX1 IP samples with and without the parental control included: all samples (left) and DDX1 only (right). In both cases, the first two principal components were plotted as a 2D scatter plot, with the percentage of variance explained by each component given in brackets in the axes titles. DDX1-GFP samples are shown in green and parental control samples in blue. Triangles and circles represent infected samples at 8hpi and 18hpi respectively, and uninfected samples are shown as squares. RNase treatment is indicated by the points' borders; RNase-treated samples have a grey border, while untreated samples do not.

The eluates were processed for proteomics in collaboration with Dr Yana Demyanenko and Prof. Shabaz Mohammed at the Rosalind Franklin Institute on an LC-MS/MS. Peptides and proteins were determined and quantified by MaxQuant 2.0 [211]. I analysed the resulting data with the advice of Dr Louisa Iselin. Protein intensities across the different conditions and replicates showed a substantially higher protein intensity in DDX1-GFP IP than in the negative controls, in agreement with the silver staining (Figure 6.3A). DDX1-GFP samples were consistent in protein intensity distribution across the different conditions and biological

replicates. This observation was reinforced in the principal component analysis (PCA) plots, where I observed the separate clustering of the DDX1-GFP and control samples. The PCA also revealed differences between RNase treated and non-treated samples (Figure 6.3B), which indicates the existence of RNA-dependent and -independent interactions involving DDX1. The separation between the different infection conditions is not as strong, which suggests only minor changes in the interactome over the course of infection.

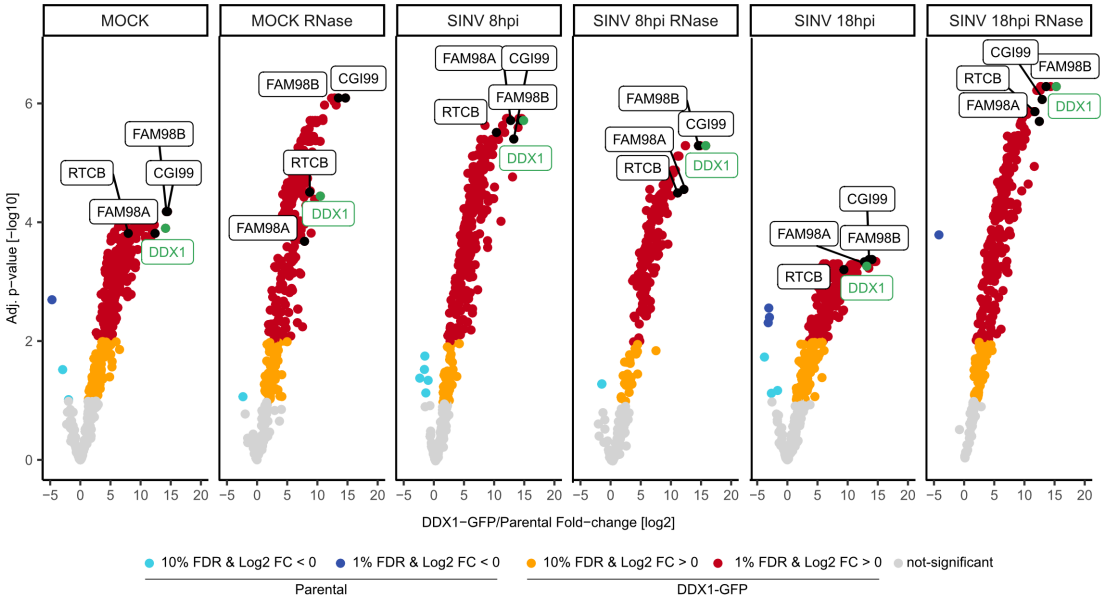

Figure 6.4: **DDX1 enriches a large number of proteins over the parental control**

Volcano plots summarising the results of limma testing of each DDX1-GFP condition against its parental control. Log2 foldchange between the conditions is plotted against -log10 of the adjusted p-value. Proteins enriched in the DDX1-GFP sample over the parental control coloured red (1% FDR) and orange (10% FDR). Proteins enriched in parental control are coloured dark blue (1% FDR) and light blue (10% FDR). Proteins that do not meet this significance threshold are shown in grey. DDX1 in each dataset is labelled in green and the tRNA-LC proteins labelled in black.

To remove contaminant proteins, I assessed the enrichment of DDX1-GFP over the parental control using the limma package in R (Figure 6.4). A significant proportion of the total captured proteins are identified in the stringent 1% FDR cut-off, further supporting the high quality of the DDX1-GFP IP. The tRNA-LC proteins (labelled in black) are consistently enriched in the DDX1 IP fraction, as expected. To ensure maximal stringency, the 1% FDR cutoff was used for downstream analysis, and the proteins within this threshold are considered "DDX1 interactors".

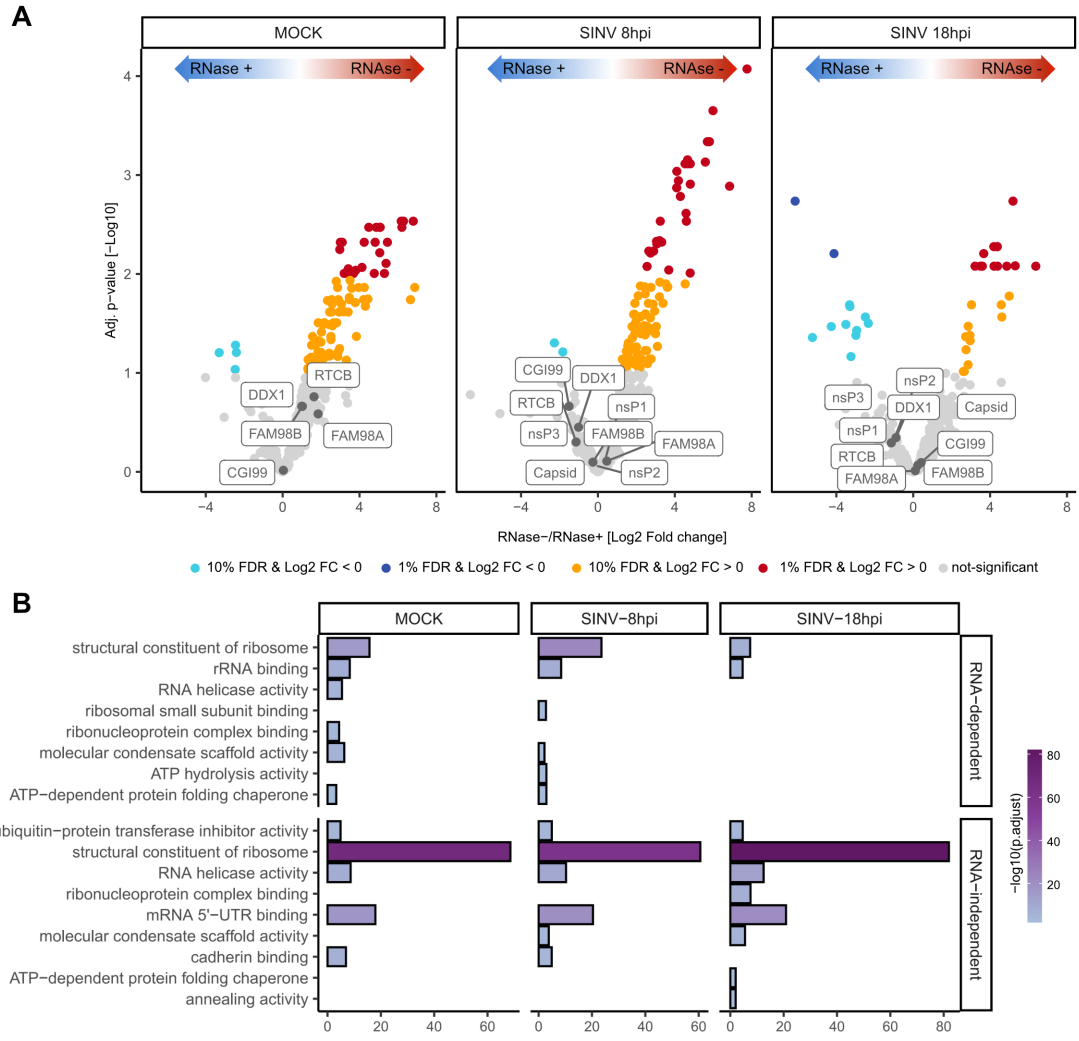

Figure 6.5: **DDX1 protein interactors are significantly enriched via an RNA-bridge**

A Volcano plots summarising the results of limma testing comparing DDX1-GFP mock and SINV 8hpi and 18hpi samples in the presence and absence of benzonase. Testing was performed on a filtered list of proteins identified as significantly enriched (1% FDR) over the parental control in at least one condition. Log2 fold-change between the conditions is plotted against -log10 of the adjusted p-value. Proteins enriched in benzonase-untreated samples are coloured red (1% FDR) and orange (10% FDR). These are classified as RNA-dependent interactions. Proteins enriched in benzonase-treated samples are coloured dark blue (1% FDR) and light blue (10% FDR). These are classified as RNA-independent interactions. Proteins that do not show a significant change are coloured grey. The tRNA-LC and viral proteins in each dataset are labelled. **B** Molecular function GO enrichment analysis of RNA-dependent and independent interactions in each condition. Enrichment analysis was performed with clusterProfiler, and overlapping GO terms were removed, using a similarity threshold of 0.25. The shade of purple reflects the significance of the enrichment, with darker purple indicating more significant enrichment.

DDX1 is an RBP and, as such, interacts with RNA as well as other proteins. RNA associates with proteins forming complex RNPs. In the absence of benzonase, DDX1 interactome would reflect the composition of the RNPs it is part of, even if many of the

interactions are indirect and bridged by RNA. To study which proteins are RNA-bridged, I next explored the differences in RNA-dependent and RNA-independent interactions by performing a limma test comparing samples processed with or without benzonase treatment (Figure 6.5A). I classified proteins as "RNA-dependent" if they had a significantly higher interaction with DDX1 in the presence of RNA. Proteins with no significant change or enrichment in the absence of RNA were classed as 'RNA-independent'. A significant proportion of protein interactors are unchanged by RNase treatment, and, indeed 73% of captured proteins across all conditions are either unchanged or specifically enriched in the absence of RNA. Among the proteins insensitive to nuclease treatment are the tRNA-LC proteins as well as viral proteins (as indicated in Figure 6.5A). The lack of change observed in the tRNA-LC upon RNase treatment correlates with the chapter 4 data where I observed direct crosslinking occurring between these proteins independently of RNA. It is, however, interesting to note that the viral proteins captured were also insensitive to nuclease treatment, indicating that the DDX1-nsP interactions were not mediated by RNA.

To explore the scope of functions associated with DDX1, I performed GO analysis of proteins enriched in the IP in either RNA-dependent or -independent (Figure 6.5B). Interestingly, most of the GO terms are associated with the ribosome and rRNA binding, although most of these instances were RNA-dependent. Meanwhile, the "ubiquitin-protein transferase inhibitor activity" proteins were enriched in RNA-independent proteins, which suggested direct protein association with no RNA mediation. Interestingly, some GO terms appear in one set of conditions in the RNA-dependent group and then shift later in infection to the RNA-independent group, such as "molecular condensate scaffold activity" or "ATP-dependent protein folding activity". The shift in RNA dependency in the different conditions could reflect a shift in RNA-binding partners or a change in function in response to infection. Many GO terms overlap in both protein groups, such as the "structural constituent of ribosome". The large ribosome complex is formed by a large number of proteins, and while some proteins may interact directly with the tRNA-LC, others may do so through the scaffold role of rRNAs and mRNAs. I previously established a direct interaction between the tRNA-LC and a ribosomal factor, RPL11, in Chapter 4. The association with similar complexes in both RNA-dependent and independent groups indicates that parts of DDX1 interactions are stabilised by an RNA bridge, whilst others are directly bound to the tRNA-LC.

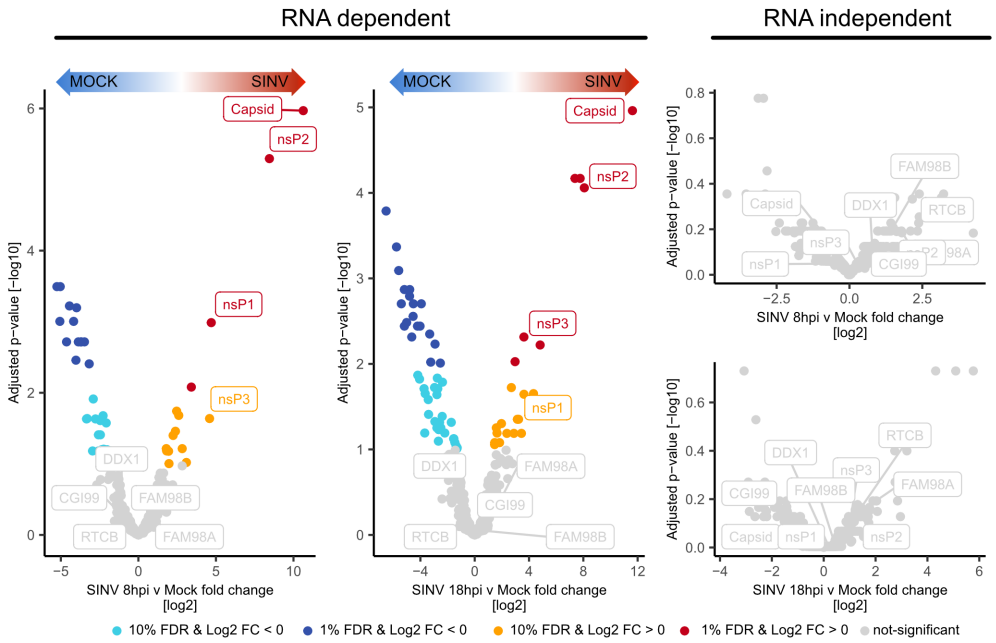

Figure 6.6: **DDX1 alters its interacting partners over the course of infection in an RNA-dependent manner** Volcano plots summarising the results of limma testing for enrichment of DDX1-GFP interaction partners in mock and SINV samples at either 8hpi or 18hpi in benzonase untreated (RNA dependent) or benzonase treated (RNA independent) samples. Testing was performed on a filtered list of proteins identified as significantly enriched (1% FDR) over the parental control in at least one condition. Log2 fold-change between the conditions is plotted against -log10 of the adjusted p-value. Proteins enriched in infected samples are coloured red (1% FDR) and orange (10% FDR). Proteins enriched in uninfected samples are coloured dark blue (1% FDR) and light blue (10% FDR). Proteins that do not show a significant change are coloured grey. tRNA-LC and the viral proteins are labelled.

The primary focus of this study was to determine changes in DDX1 interactome that occur during the course of infection, which could shed light on its regulatory roles. To explore this, I compared the SINV-infected samples to their uninfected counterparts. No changes were observed for RNA-independent interactors of DDX1, which suggests that the core complex and its primary interactors remain unaltered upon infection (Figure6.6). However, significant changes occur at 8hpi and 18hpi for RNA-dependent interactors, with the most significantly enriched proteins being viral proteins (nsP1, nsP2, nsP3 and Capsid). The interaction with viral proteins is very exciting, but because they are absent in mock conditions, it is essential to compare the dataset to the whole cell proteome to determine if they are bona fide interactors or IP "carryovers" due to high cellular abundance.

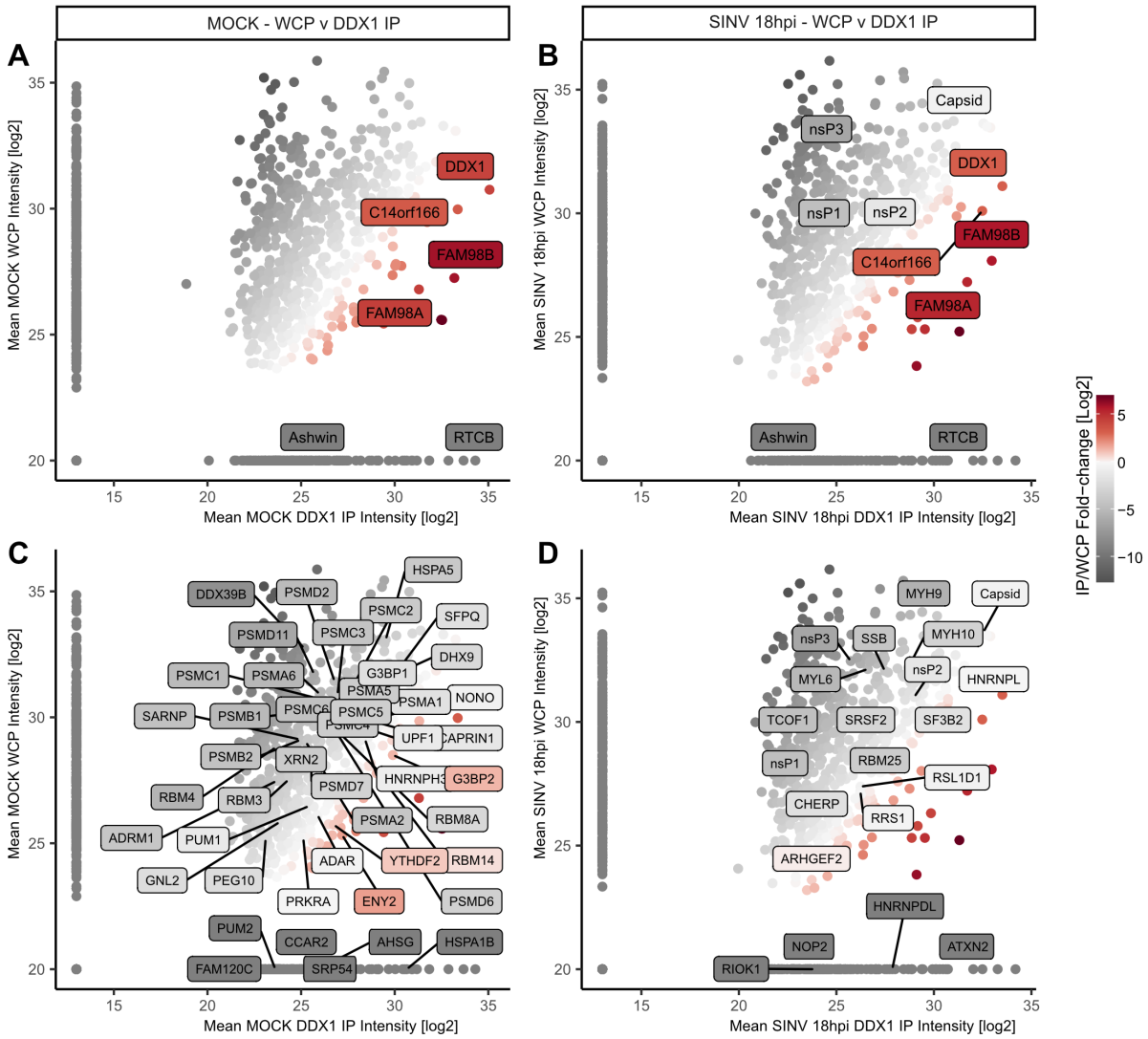

Figure 6.7: **Comparative analysis of whole cell proteome and DDX1-IP**

Scatter plots comparing DDX1-GFP IP log2 intensities in mock and SINV 18hpi without benzonase treatment to whole cell proteome log2 intensities 18hpi. Colour reflects log2 fold-change in MATR3 IP v WCP. Red points are proteins with a higher enrichment in DDX1 IP. Points in dark grey near the axis bar are proteins only present in one or the other dataset (y-axis associated with WCP and x-axis associated with DDX1-IP). tRNA-LC and viral proteins are labelled.

Protein-protein interaction experiments can represent true protein-specific interactions or can reflect a diluted subset of the whole proteome. Protein abundance can influence IP capture as they rely on the dilution through several washes of the whole cell lysate via an antibody-based enrichment interacting with the bait protein. To determine whether the proteins observed as significantly enriched in the SINV infection are true DDX1 interactors, I compared my DDX1-GFP IP analysis to a previously published whole cell proteome (WCP) performed in SINV-infected HEK293 cells in similar conditions [113]. However, this comparison has some

limitations. The WCP includes the entire proteome, including the nuclear fraction, which was omitted in the DDX1-IP dataset. Additionally, differences in the MOI may affect the kinetics of the infection and cellular signalling pathways, which could influence the observed protein interactions. Ideally, this comparison would have been performed using a cellular fraction of the whole cell proteome under identical experimental conditions.

I plotted the protein intensity fold change in either mock or SINV at 18hpi conditions and observed a significant proportion of proteins are enriched in the DDX1 IP (Figure6.7 A and B). The tRNA-LC proteins are notably enriched, with two of the complex components, RTCB and Ashwin, only detected in the IP and not in the WCP. The enrichment of tRNA-LC proteins indicated the stringency of the DDX1 IP and the ability to detect proteins that would be otherwise difficult to capture in the more complex WCP. In the SINV 18hpi condition, the viral proteins are more enriched in the WCP than in the DDX1-IP. The WCP proteome was collected after 18hpi with an MOI of 10, whereas the DDX1-IP I performed was harvested from cells infected for the same length of time with an MOI of 3. The higher MOI used for the WCP would have likely resulted in a high number of replication centres and viral proteins. Therefore, I conclude that it is likely that the interaction between DDX1 and viral nsPs and capsid is real, but it is likely transitory or substoichiometric. A plausible explanation for such a short-lived interaction is that the tRNA-LC might interact with the viral RNA that is pulled out from the replication organelles through the nsP1 pore, creating a temporary RNA-bridged interaction with the viral replicase complex that is lost when the viral RNA is released.

I further investigated the significantly enriched proteins identified in the DDX1-eGFP IP in the comparison between Mock and 18hpi shown in Figure 6.6. On the mock WCP/DDX1-IP comparison, I labelled proteins significantly decreased in DDX1 binding after SINV (Figure 6.6C). A large proportion of proteins appear to be abundant in both IP and WCP datasets. Nonetheless, a subset of proteins is enriched specifically in DDX1-IP, particularly G3BP2, FAM120C, and RBM14.

Proteins significantly enriched in DDX1 binding during SINV infection were labelled on the SINV 18hpi WCP/DDX1-IP comparison (Figure 6.6D). Once again, DDX1-enriched proteins appear to also be highly abundant in the WCP. NOP2 and ATXN2 are two proteins, however, that appear exclusively in the DDX1-IP and are significantly enriched during infection. Intriguingly, MYH9 has a higher enrichment in WCP than in DDX1. However, in my XL-MS data in Chapter 4, I identified MYH9 as a direct interactor with the tRNA-LC in mock conditions.

This indicates that although intensity comparison between WCP and IP can allude to protein abundance "contaminating" the IP, for many proteins, this is probably not always the case, such as MYH9. Therefore, this analysis should be used for extreme cases: i.e. proteins enriched mostly in DDX1 IP (representing strong interactors) and proteins substantially enriched in the WCP (representing contaminants or interactors with low stoichiometry with the bait). The significant capture of proteins such as tRNA-LC proteins in the IP over the WCP agrees with this notion, reflecting selective enrichment of the core members that strongly engage with DDX1.

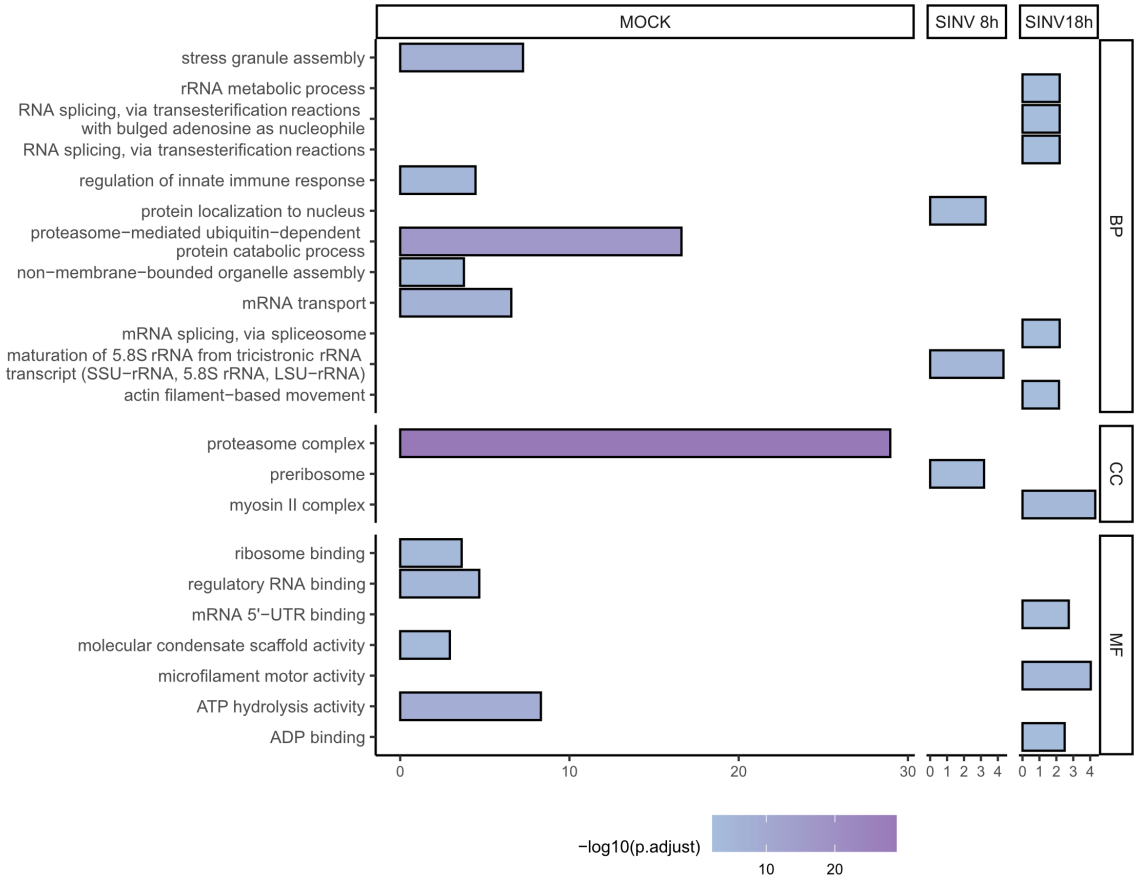

Figure 6.8: **DDX1 functional switch during infection identified by GO enrichment**

GO enrichment analysis of proteins enriched in the mock sample and the SINV infected. Significantly enriched protein lists were taken from Figure 6.6. Mock list was combined from both infection analyses. Enrichment analysis was performed with clusterProfiler, and overlapping GO terms were removed, using a similarity threshold of 0.25. The shade of purple reflects the significance of the enrichment, with darker purple indicating more significant enrichment.

To explore the changes in the DDX1 interactome after infection, I performed GO enrichment followed by STRING network analysis on the proteins identified as significantly changed in

DDX1 binding over the course of infection. I identified 32 and 46 proteins with reduced binding at 8hpi and 18hpi, respectively. Among these identified proteins, 26 were shared in both groups. In further analysis, I grouped both sets of proteins with decreased interaction with DDX1 into a larger "Mock" group.

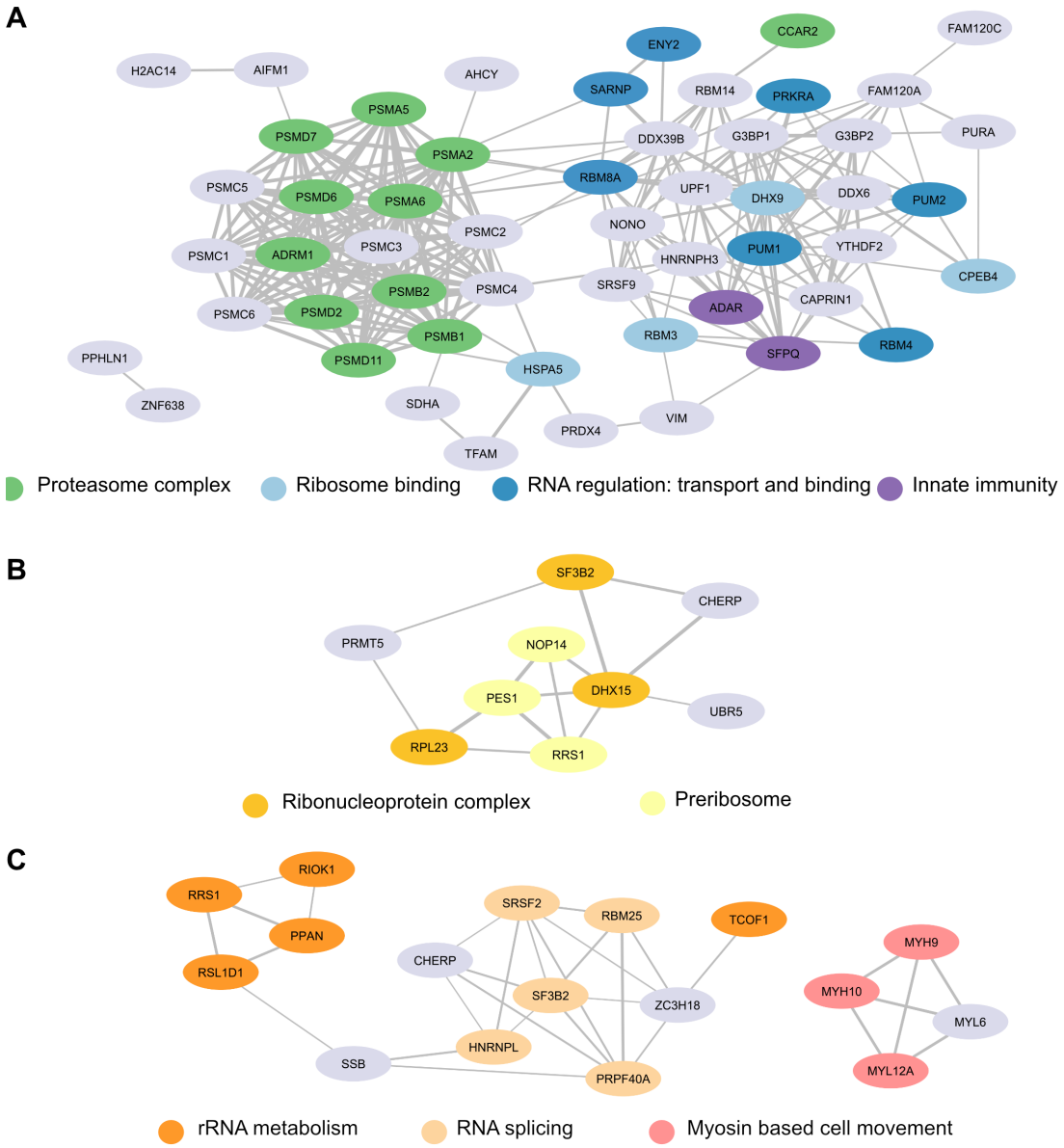

Figure 6.9: **STRING network of DDX1 dynamic interactors**

Networks of significantly changed protein interactors during SINV infection. Proteins lists were taken from the 10% FDR cutoff in 6.6. **A** Proteins with significantly reduced binding and associated to Mock condition. **B** Proteins with significantly increased binding and associated with SINV 8hpi condition. **C** Proteins with significantly increased binding and associated with SINV 8hpi condition. The width of the connecting line reflects the interaction score, with wider lines reflecting stronger interactions. Proteins were coloured based on GO annotations. Network analysis was performed using STRING and results were plotted with Cytoscape.

In the cellular component (CC) GOs, I observed a strong shift from the proteasome complex to the ribosome at 8hpi and an additional shift to the myosin complex at 18hpi (Figure 6.8). This indicated a change in interacting partners, which could reflect a functional alteration. I further observed these complexes in the STRING network analysis (Figure 6.9). In the mock group, DDX1 associates with the proteasome complex with high confidence (Figure 6.9A). I further observed proteins involved in ribosome binding and RNA regulation grouping together.

At 8hpi, the differential interactors were predominantly structural components of the ribosome, instead of proteins involved in ribosomal regulation and biogenesis observed in the mock conditions (Figure 6.9B). Interestingly, I previously observed a direct interaction between the tRNA-LC and RPL11 in uninfected cells when studying tRNA-LC interfaces in Chapter 4. I further identified RPL7, EIF3A and ENO1 in the previous dataset, although no direct crosslinks with the tRNA-LC were observed. The direct association with these proteins in mock conditions indicates that DDX1 also interacts with the translation apparatus in uninfected cells. The increased binding to the translation apparatus observed at 8hpi and 18hpi in my interaction analysis suggests that infection increases the association of DDX1 with the ribosome.

The increased binding of myosin complex proteins at 18hpi is very intriguing (Figure 6.9C). I previously observed a direct interaction between the tRNA-LC and MYH9 in Chapter 4 as well. The increased binding to MYH9 and its associated complex (MYH10 and MYL12A detected in the protein-protein network), suggests an increased association with the cell transport machinery after infection.

Furthermore, among the biological processes (BP) (Figure 6.8), I observed that a significant proportion of proteins dissociating from DDX1 during infection (indicated in the mock group) are proteins previously related to virus infection. These include "stress granule assembly" and "non-membrane-bound organelle assembly", with proteins such as DDX6, G3BP and CAPRIN1. Viruses are known to disassemble molecular condensates such as p-bodies and stress granules that can hijack viral proteins [154]. There were also proteins involved in innate immunity, such as ADAR and SFPQ, which also decreased their binding to DDX1 as the infection progressed (Figure 6.9A). At 18hpi, DDX1 exhibits increased association with spliceosomal proteins as well as 5'UTR mRNA binding factors. This is very relevant as SINV RNAs translate in a non-canonical manner in infected cells, specifically pertaining to initiation mechanisms [125, 286, 287].

The versatile role of DDX1 in cellular regulation is portrayed by the different protein

complexes it associates with. The dynamic changes over the course of infection suggest that DDX1 alters its protein partners during infection, potentially inhibiting the viral lifecycle. Which of these interactions drives DDX1's antiviral activity is unclear. However, I hypothesise the antiviral role of the tRNA-LC is through negatively affecting translation regulation or recruitment of antiviral proteins such as spliceosomal subunits.

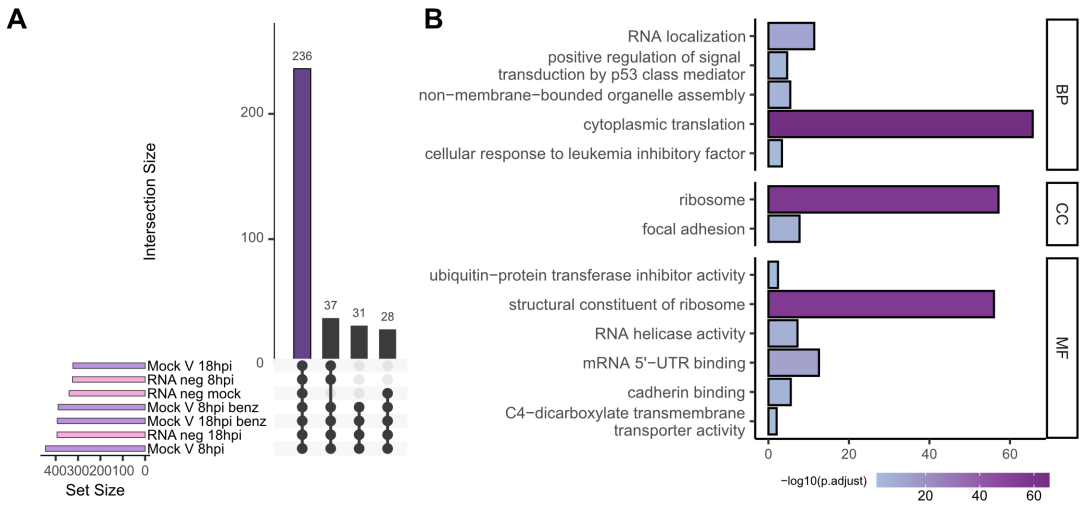

Figure 6.10: **DDX1 core-interactors insensitive to RNase treatment and SINV infection**

A Upset plot of the RNA insensitive and SINV unchanged proteins. The intersection of all groups form the "core" DDX1 interactors labelled in dark purple. **B** Core-DDX1 interactors were selected for GO enrichment analysis. Enrichment analysis was performed with clusterProfiler, and overlapping GO terms were removed, using a similarity threshold of 0.25. The shade of purple reflects the significance of the enrichment, with darker purple indicating more significant enrichment.

The overlap in protein functions in the dynamically changed proteins observed above, as well as the low number of significantly changed proteins, indicated that a much larger proportion of DDX1 protein partners remained unchanged. I thus compared the unchanged protein interactors pulled from the different comparisons. I identified core-DDX1 interactors that are both insensitive to RNase treatment and unchanged during SINV infection, such as the tRNA-LC components. I cross-matched each dataset to identify which proteins consistently bind to DDX1 irrespective of stimuli or treatment (Figure 6.10A). I observed 236 unique proteins that intersected all conditions. I further analysed the large protein dataset by GO enrichment. I observed that the most prevalent GO terms were related to cellular translation (Figure 6.10B). When comparing it to previous GO analyses performed on differentially associated proteins during infection, I observed some functional overlapping such as the binding of 5'UTR of mRNAs and the association with the ribosome. The crossover between core-interactors and

dynamically increased binding partners indicates that DDX1 is not repurposed for a novel role during infection. I suspect the tRNA-LC is most likely modulating its function through the association with regulatory proteins controlling those processes.

Overall, DDX1 is a versatile protein associated with a large number of different protein complexes, mainly tRNA-LC explored previously. When focusing on the main interactors of DDX1, I can conclude that it is heavily involved in protein synthesis, which agrees with CGI99 being a cap-binding protein [246]. During infection, most compositional changes are RNA bridged, which reflects changes in the RNP DDX1 is part of rather than changes in the complexes. For instance, DDX1 was observed to interact with Capsid and nsP2, but these interactions were RNA bridged. The dynamic changes observed for the DDX1 interactome reflect alterations in RNP composition/function. However, the consistency of the differentially associated proteins and their overlapping functions suggests modulatory changes in the core-DDX1 interactome rather than entirely new associations.

6.2.2 Analysis of RNA binding sites of DDX1 and RTCB by iCLIP2 sequencing

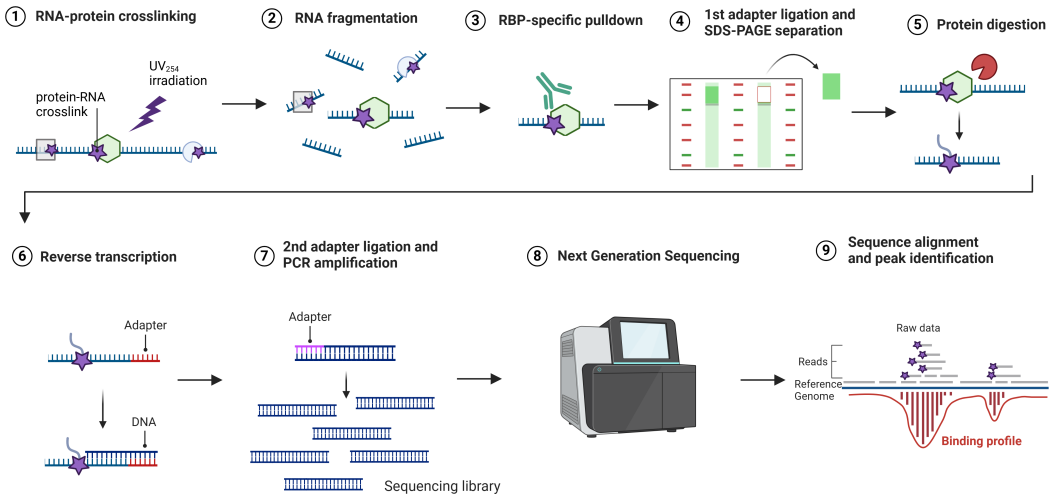

Figure 6.11: **Schematic of iCLIP2 protocol**

Schematic adapted from Dr Louisa Iselin and [101, 209]

DDX1 and RTCB are both tRNA-LC core proteins and RBPs. As a consequence of their RNA-binding nature, a substantial proportion of their interactors are bridged by RNA as observed in the previous protein-protein interactome of DDX1. Previous studies have identified both DDX1 and RTCB as direct interactors with SINV vRNA [123].

I next explored whether they bind to specific sequences on the vRNA using iCLIP2. This

state-of-the-art method allows for the identification of RNA binding sites *in cellulo* at a single nucleotide resolution. The iCLIP2 techniques is outlined in Figure 6.11. To summarise, the protein-RNA interface is stabilised by UV crosslinking, which forms covalent bonds between the appropriate nucleotides and amino acids when placed at "zero distances". The RNA is then fragmented by RNase treatment, and the protein-RNA complexes are immunoprecipitated with very stringent wash buffers to select only the protein of interest and its bound RNA fragment. The complexes are further isolated by SDS-Page gel separation followed by excision in a size-specific manner. At this point, the protein is digested using Proteinase K, leaving fragmented RNA with a protein adduct. Subsequent RNA processing is done for sequencing library preparation. This involves adapter ligation and reverse transcription. The 5' end of the cDNA will, in most cases, terminate due to the presence of the protein adduct representing the crosslinking site. Secondary adapters, containing Unique Molecular Identifiers (UMI) and indexes are ligated and PCR amplified. The use of UMIs allows for the identification of single RNA fragments, removing PCR amplification bias during the normalisation of reads, while indexing allows for multiplexing. To control for background non-specific RNA signal, a parallel pipeline is performed on whole cell lysates without IP enrichment. The samples are run directly on the SDS-PAGE gel and excised at the same size region as the IP samples (indicated as white boxes in Figure 6.12C). They are referred to as Size Matched Input (SMI) samples and contain all the RNA fragments bound by all proteins migrating at the same region of the gel as the protein of interest.

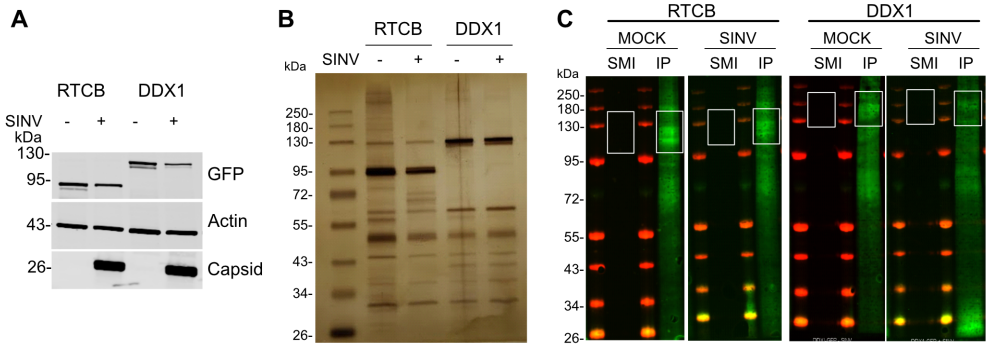

Figure 6.12: **DDX1 and RTCB iCLIP2 quality control and processing**

HEK293 FITR RTCB-GFP and DDX1-GFP at different stages of iCLIP2 processing: **A** Western blot of iCLIP2 input samples at start of iCLIP2 processing. **B** Silver Stain of iCLIP2 samples after IP. **C** Licor Odyssey images of SMI (input) and IP samples after adapter-ligation. The green indicates the adapter ligated RNA. The white boxes represent the areas cut out of the membrane for iCLIP2 downstream processing at the protein size specific regions allowing for additional ~60kDa for RNA bound weight.

To elucidate the RNA binding profiles of DDX1 and RTCB, I performed an iCLIP2 experiment in HEK293 FITR DDX1-GFP and RTCB-GFP cell lines in mock or SINV infected for 16hpi. The sequenced data was analysed with the support of Dr Louisa Iselin. Using an in-house analysis pipeline, I aligned sequencing data to a combined human and SINV genome annotation and defined 'crosslink sites' as the first nucleotide before the start of each read. I then identified regions with significantly enriched crosslink frequency over the SMI using the ht-seqCLIP/ DEWseq analysis pipeline [222]. Unfortunately after sequencing, a couple of the mock DDX1-SMI samples were absent of reads. DDX1 and RTCB are members of the same complex, and, therefore, binding sites may be identical or proximal. Although stringent buffers were used to remove protein partners during iCLIP2 preparation, I cannot exclude cross-contamination to some extent (visible in silver stain Figure 6.12B). It is thus possible that within the RTCB binding sites, a small fraction may be DDX1 binding sites given that DDX1 molecular weight is higher than RTCB and falls within the extracted membrane region (Figure 6.12C). However, I do not expect cross-contamination in the opposite comparison as RTCB-RNA complexes migrate lower in the gel than the DDX1 ones. Due to the absence of mock DDX1-SMI, I employed mock RTCB-SMI in downstream analysis for both proteins as I expect crossover between binding sites and RTCB-SMI to contain DDX1-related background. I performed PCA to assess how the samples compared to each other (Figure 6.13). PCA demonstrates that there are pronounced differences between the IP and SMI samples, as well as between mock and infected conditions in both tRNA-LC proteins. DDX1 and RTCB in SINV condition cluster together closely, reflecting a low degree of technical variation.

The identified binding sites were strikingly different in mock and infected samples. 600 and 1000 unique binding sites were captured in DDX1 and RTCB in mock, respectively, and only 150 in infected cells (Figure 6.14A). This may be caused by the change in the RNA landscape occurring during SINV infection. The cellular transcriptome at late stages of SINV infection has been characterised to consist of 70% vRNA and a vast downregulation and degradation of cellular housekeeping RNAs [113, 270]. RTCB and DDX1 had 30% and 50% target genes overlapping, respectively (Figure 6.14B). The overlapping target genes indicate that a strong crossover of binding preferences exists between the two tRNA-LC proteins, as they may be binding in tandem or in proximity to one another.

The genes the two tRNA-LC components bind to could reflect a wider functional role beyond nuclear tRNA splice junction binding [197] and *Xbp1* mRNA cytoplasmic splicing

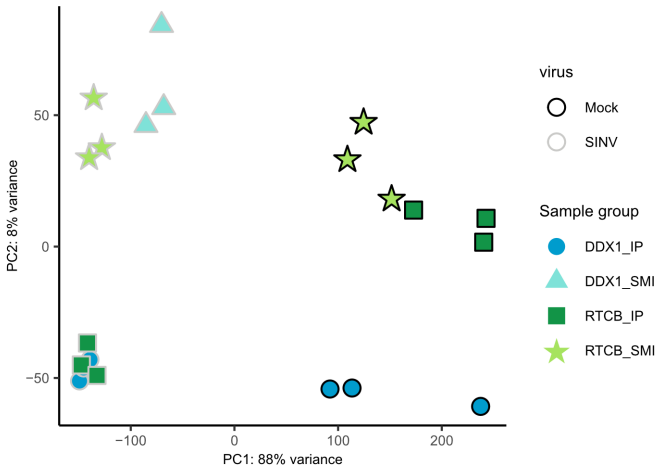

Figure 6.13: **Principal Component Analysis of DDX1 and RTCB iCLIP samples**

The first two principal components were plotted as a 2D scatter plot, with the percentage of variance explained by each component given in brackets in the axes titles. Mock samples are shown with a black ring and SINV-infected samples with a light grey ring. Triangles and stars represent SMI samples for RTCB and DDX1 respectively. Squares and circles represent IP samples for RTCB and DDX1 respectively.

[203]. I previously established a strong involvement in translational machinery and a crucial dependence on RNA in the DDX1 protein interactome, suggesting that a variety of RNAs are bound by the tRNA-LC. I performed a GO enrichment analysis on the set of bound mRNAs overlapping between DDX1 and RTCB (Figure 6.14C). Intriguingly, the highest occurrence of genes corresponded to mitochondrial genes, specifically mitochondrial translation genes. Whether any of these genes are subjected to tRNA-LC mediated ligation or to other functions of the complex, such as translational modulation, requires further investigation. However, the detection of mitochondrial genes indicates mitochondrial localisation of the tRNA-LC.

To understand the properties of RTCB and DDX1 RNA binding, I focused on the uninfected cellular binding profiles. I first categorised the types of RNAs the proteins bound and identified that they predominantly interacted with protein-coding RNAs (Figure 6.15A). Although the tRNA-LC has been found to be involved in tRNA maturation, only a small fraction of the identified sites belonged to mitochondrial tRNA. Capturing tRNA sequences is extremely difficult because of their inherent short sequence (less than 100nt) and complex secondary structures [288]. Furthermore, tRNAs contain the highest density of post-transcriptional modifications among all RNAs [289]. The reverse transcription during library preparation can be prematurely terminated, and adapter ligations can be problematic as the 5'- and 3'-end of mature tRNA form a rigid terminal structure that blocks adapter access [290]. Nonetheless,

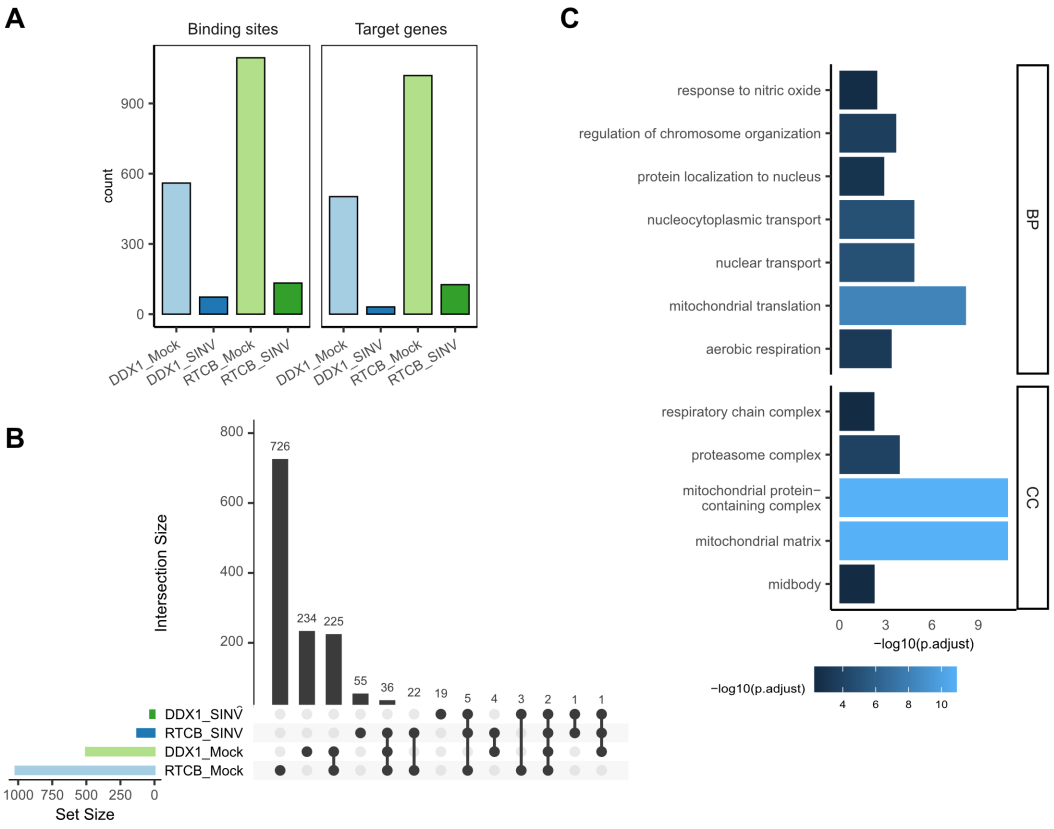

Figure 6.14: **ICLIP2 binding sites show disparity in Mock and SINV conditions**

A iCLIP2 counts of binding sites and unique target genes of DDX1-GFP (blue) and RTCB-GFP (green) in mock and SINV infected conditions. **B** Upset plot of binding sites across samples. **C** GO enrichment analysis of overlapping DDX1 and RTCB mock binding genes. Enrichment analysis was performed with clusterProfiler, using a similarity threshold of 0.35 for BP analysis and 0.5 for CC. The colour of each bar indicates the significance of the enrichment, with lighter blue indicating more significant enrichment.

my iCLIP data suggests that the tRNA-LC binds to a variety of protein-coding RNA, indicating involvement in a much larger host of RNA regulation.

Analysing the binding location within the target genes, indicated that both proteins preferentially associate with coding regions (CDS) followed closely by the 5'UTR (Figure 6.15B). This preference was substantially higher than that of the global set of cellular RBPs when considered together (eCLIP superset in ENCODE). Upon further inspection by plotting the signal distribution within each region, DDX1 and RTCB showed a binding propensity at the end of the 5'UTR and the very start of CDS (Figure 6.15C and D). 5' UTR regions near the cap structure and the start codon are fundamental elements for translation initiation. Furthermore, a core member of the tRNA-LC, CGI99, has been described as being involved in cap-binding activity together with the other members of the complex [246], which could be

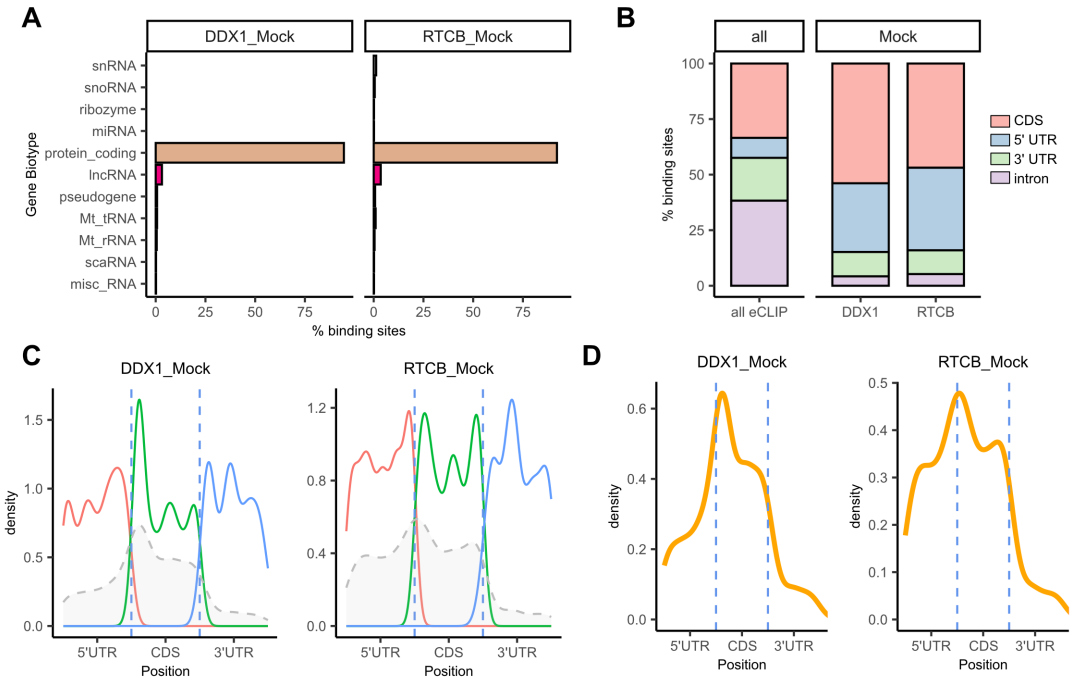

Figure 6.15: **DDX1 and RTCB preferentially bind 5'UTR and start of CDS**

A Bar plots showing the proportion of target RNAs corresponding to different RNA biotypes for shared interactors in DDX1 and RTCB in mock condition. **B** A stacked bar plot showing the distribution of enriched binding sites in mRNAs across gene regions. The average distribution of binding sites across all available eCLIP datasets from ENCODE is included for comparison. **C** Meta-gene profiles of binding site distribution across 5' UTR, CDS, and 3' UTR. Coloured lines reflect binding within each region. The grey shaded area reflects the overall pattern of binding across all regions. **D** A line plot showing the overall binding across all regions

driving the observed binding within the 5'UTR region. Together with the observed interaction with ribosomal proteins, this suggests the tRNA-LC role in translational control.

iCLIP2 can be used to identify binding motifs recognised by the RBPs across the genome. I took a 50 nucleotide (nt) window for each binding site in cellular RNA, taking the signal peak as the window's centre, and searched for enrichment over a set of gene and gene region-matched background sequences using STREME [224]. In DDX1 and RTCB, I detected a G-rich motif that is compatible with the sequence feature of G-quadruplexes (G4) (Figure 6.16A). This is consistent with published work, which has identified DDX1 binding to G4 structures present in intronic switch transcripts to promote class switch recombination at the immunoglobulin heavy-chain (IgH) locus, converting them into S-region R-loops [247]. However, the proportion of sequences containing the motif was moderately low (40% in the lower degenerate similarity threshold and only 2% in the more stringent threshold) (Figure 6.16B). I tested the frequency within the sequences at two similarity thresholds, and in both cases, the frequency was double

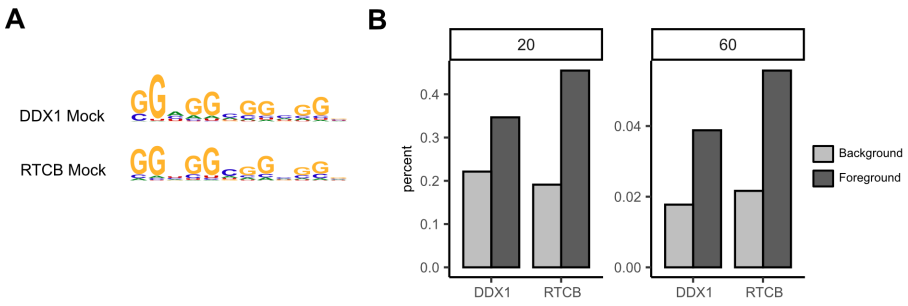

Figure 6.16: **DDX1 and RTCB display a G4 binding motif**

A Most enriched motif in mock binding sites for DDX1 and RTCB. Motif enrichment was performed on 50 nucleotide windows, centred around the signal peak of each binding site. Enrichment was relative to gene and gene region-matched sequences generated for each binding site, and was performed using STREME from the MEME suite. **B** Percentage of input sequence containing a predicted G4 motifs at two different sensitivity threshold using the predicted quadruplexes (PQs) finder package in R [291].

over the background sequences. The tRNA-LC has no discernible motif and may bind G-rich regions rather than a specific motif. The low occurrence of G4 binding sites within the sequences suggests that, in some instances, tRNA-LC binds to a G4 motif. However, this is not the main driver of tRNA-LC-RNA binding.

To explore whether G-rich regions were tRNA-LC binding drivers, I examined single- and di-nucleic acid frequencies within the sequence binding regions as previously performed for motif prediction. Using a 50nt window with the binding site at the centre, I observed DDX1 and RTCB distinct affinity towards Guanine-rich sequences (G) across the full window (Figure 6.17A). A sharp preference for Uracil (U) was also identified in the middle of the window, at the peak signal point. The identification of U at the centre of the binding window is due to the crosslinking bias of the nucleotide. I cross-checked this data by examining in a linear manner the enrichment of the identified set of genes (foreground) over gene region-matched background sequences (background) (Figure 6.17B). There was an even distribution of nucleotides across the 50nt sequence, a consistent peak at the midpoint for U and a clear second peak for G immediately after were observed.

In the di-nucleic acid mapping, DDX1 and RTCB both show a distinct preference for GG-rich regions followed by GC and CG-rich across the binding window (Figure 6.17C and D). Meanwhile, UU is enriched at the centre of the binding window as previously observed in the single nucleotide mapping. Notably, both proteins behave in an identical manner. I further investigated if binding preferences were driven by secondary structure (not shown). However,

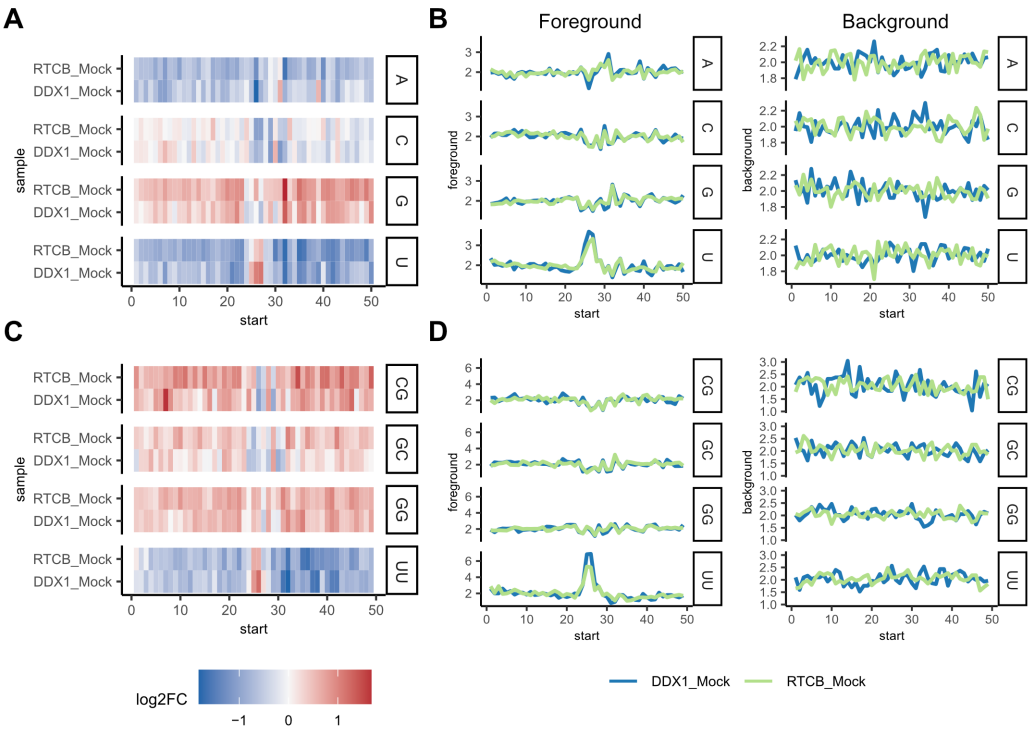

Figure 6.17: **tRNA-LC proteins show a preference for G-rich binding regions**

A Heatmap of nucleic acid distribution in DDX1 and RTCB mock samples in a 50 nucleotide windows, centred around the signal peak of each binding site. **B** Line plot of nucleic acid signal in DDX1 and RTCB mock samples in a 50 nucleotide windows, centred around the signal peak in the binding sites (foreground) and region matched sites (background). **C** Heatmap of dinucleic acid distribution in DDX1 and RTCB mock samples in a 50 nucleotide windows, centred around the signal peak of each binding site. Of the 16 different nucleic acid combination, the top 4 highest enriched dinucleotides are shown. **D** Line plot of dinucleic acid signal in DDX1 and RTCB mock samples in a 50 nucleotide windows, centred around the signal peak in the binding sites (foreground) and region matched sites (background).

no structure correlated significantly to binding site preferences.

Recent work from the Castello lab has shown that RTCB and DDX1 interacts with SINV vRNA [123] and accumulates in viral replication organelles [113], I hypothesised that these proteins interact directly with the vRNA, which becomes the dominant poly(A) RNA in the cell. Indeed, when plotting the binding sites across the viral genome, I observed mapping of the binding regions across the viral genome (Figure6.18). I subtracted the peak signal from the SMI signal (negative control), allowing for a specific tRNA-LC binding pattern to emerge. The strongest peak is identified at the start of the sgRNA, specifically within the sgRNA 5'UTR and the start of the CDS of the structural protein Capsid. The most distinguished peaks are within the sgRNA which could be due to the much higher abundance of the sgRNA at the later stages of SINV infection. Nonetheless, tRNA-LC binding peaks are observed across the

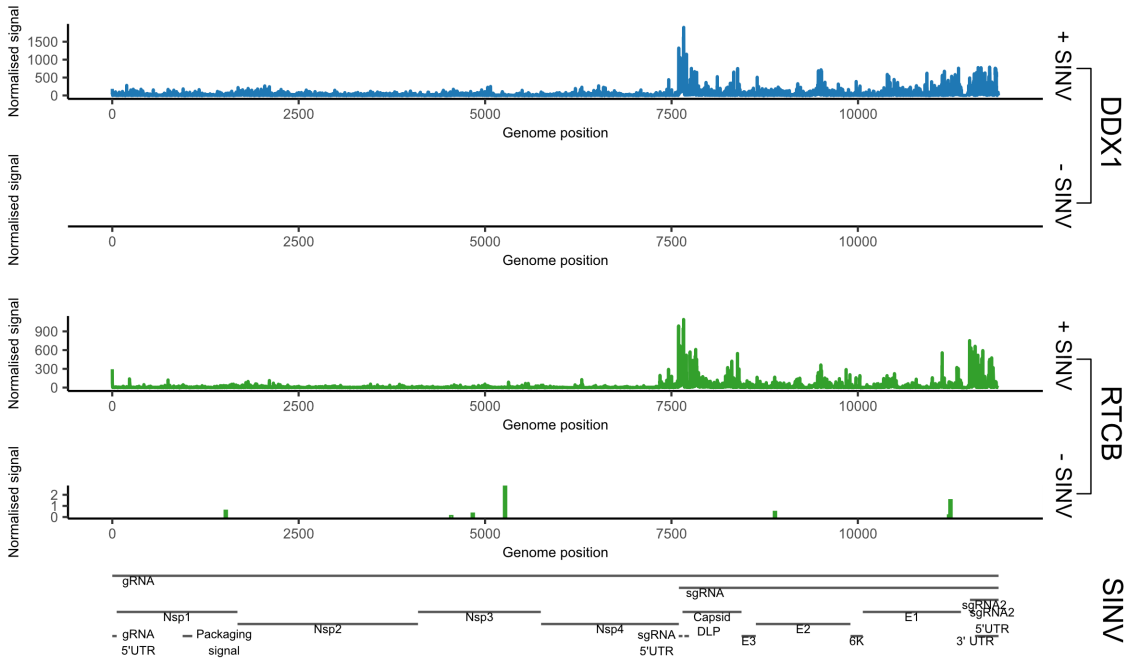

Figure 6.18: **DDX1 and RTCB bind SINV positive-strand vRNA**

Line plots depicting DDX1-GFP and RTCB-GFP signal on the SINV positive and negative strand RNA. The density of crosslinked sites at each position is the binding signal values, normalised by the subtraction of the SMI value. Annotation of the SINV genome is depicted at the bottom

full viral genome, with a preference for the region which overlaps between the gRNA and the sgRNA. This suggests that the tRNA-LC binds at several positions across the viral genome with a preference for the 5'UTR of the sgRNA. Crucially, the binding profiles of both proteins were remarkably similar. Notably, no binding was observed on the negative strand, which is the replication template. This suggests that the tRNA-LC modulates viral gene expression through processes that involve the positive vRNA, which include the synthesis of the negative strand, translation of the sgRNA or/and gRNA, stability of the positive sense RNAs, and formation of the viral particles.

In cellular mRNA, DDX1 and RTCB showed a preferential binding to G-rich and, to some degree, to G4-like binding motifs. I mapped the G frequency as well as the predicted G4 across the viral genome to determine if the binding preferences on cellular and vRNA are similar (Figure6.19). DDX1 and RTCB had very similar binding profiles, and thus, I used DDX1 as a representative example of the tRNA-LC SINV binding profile. I co-aligned the DDX1 binding profile to the G-rich and predicted G4 regions and noted that G-rich regions within the sgRNA overlap tRNA-LC binding sites. However, the binding overlap was not mutually

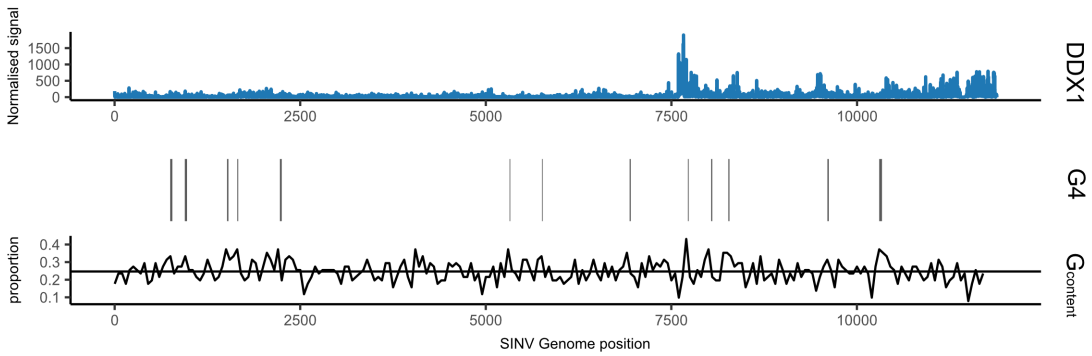

Figure 6.19: **SINV genome contains predicted G4**

Line plot depicting DDX1 across the SINV genome. The density of crosslinked sites at each position is the binding signal values, normalised by the subtraction of the SMI value. Below are the predicted G4 sites, using PQsfinder package in R. Bottom is a line plot depicting the G content across the SINV viral genome, the horizontal line represents the expected average of 25% of nuclei acid content.

exclusive. The binding of the tRNA-LC and the predicted G4 sequences in the viral genome did not occur concurrently. This suggests that although the tRNA-LC binds G-rich regions within the positive-sense vRNAs to some degree, further factors directing the binding specificities are at play.

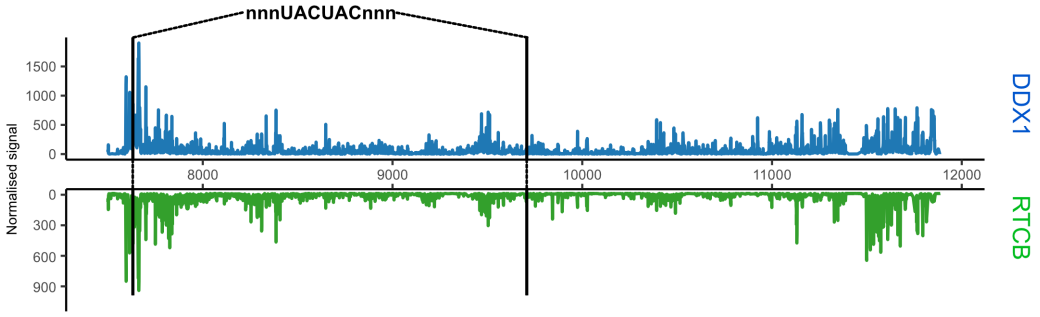

Figure 6.20: **Correlation of tRNA-LC binding on the sgRNA with predicted U2 snRNA and SF3B complex binding motifs**

Line plot depicting DDX1 and RTCB across the subgenomic region of the SINV genome (positions 7597 to 11885 nt of the SINV genome). The density of crosslinked sites at each position is the binding signal values, normalised by subtracting the SMI value. In dashed lines are the U2 snRNA complementary branch sites (UACUAC).

In the protein-protein interaction analysis of DDX1 earlier in this chapter, I observed the significant enrichment of splicing proteins at later stages of infection. SF3B2 was among the enriched interactors at 18 hpi. Notably, Kamel et al. (2024) documented the dynamic shuttling of nuclear factors to viral factories during the course of infection, in particular the

SF3B complex proteins [123]. The SF3B complex was further characterised in this study to inhibit viral gene expression in a splicing-independent manner. The full SF3B complex and U2 snRNA (collectively known as U2 snRNP) suppress viral infection through direct binding to SINV RNA [123]. I hypothesise that the tRNA-LC binding to vRNA is occurring due to its recruitment from other factors rather than sequence specificity. I speculate that the binding of the U2 snRNP is aiding the tRNA-LC vRNA binding and exerting its antiviral role. Kamel et al. (2024) identified three perfect complementary U2 snRNA branch site interacting stem-loop motifs present in the SINV genome [123]. Two of the three sites are within the sgRNA region of the genome. I mapped the U2 snRNA branch site interacting stem-loop motifs regions in correlation to the iCLIP of DDX1 and RTCB (Figure 6.20). I observed that the first site is highly enriched in my iCLIP2 data. The second site is moderately enriched with tRNA-LC binding, although to a lesser extent. This suggests a potential cooperative binding in these regions via protein-protein interactions.

DDX1 and RTCB were demonstrated to bind RNA in similar and proximal regions of one another, portraying their close relationship as part of the tRNA-LC. The two proteins predominantly interacted with 5'UTRs and the start of CDS of cellular RNA. Their binding specificity indicated a preference for G-rich regions. In SINV infection, the tRNA-LC proteins were identified to bind directly to the positive-sense vRNA. A distinct binding in the sgRNA 5'UTR and the start of CDS of viral proteins was observed. Intriguingly, the proteins were not observed to bind to the negative strand. Altogether, the binding pattern in cellular RNAs and vRNAs indicated a potential role in translation initiation.

6.2.3 Decoupling SINV replication and translation

The antiviral role of the tRNA-LC in SINV infection, as established in Chapter 5, was further underscored by the observation that the knockdown of its components, DDX1 and CGI99, led to an upregulation of viral protein production. A notable increase in nsP3 protein levels suggested a potential early involvement of the tRNA-LC during infection. In this chapter, I established the broad protein partners of DDX1, with many of these interactions mediated by RNA as occurs with the viral proteins nsP2 and Capsid. Additionally, the iCLIP2 study of DDX1 and RTCB revealed that tRNA-LC proteins bind directly to vRNA, implying that their antiviral function may arise from direct RNA regulation.

In positive-stranded RNA viruses like SINV, vRNA serves a dual role as both the genome for replication and the mRNA for translation. Therefore, the tRNA-LC interaction with positive-sense vRNA could influence either process. I hypothesise that the tRNA-LC regulates infection by modulating viral translation. This hypothesis is supported by the binding profiles of tRNA-LC in both cellular RNA and vRNA, particularly in the 5' UTR, as well as its close association with translational factors. However, these findings were not functionally tested, requiring further investigation to clarify the precise mechanisms involved in DDX1 regulation of infection.

To study precisely if translation or replication is the target of the tRNA-LC, I employed two different sets of SINV replicons in a DDX1 knockdown background. Firstly, the alphavirus *trans*-replicase system has been developed by the Merits research group, and plasmids were kindly shared with me for this project. This sensitive system has been employed to efficiently study the replicase complex formation, the functional analysis of nsPs and/or the requirements of host factors [157, 292, 293]. Two plasmids are co-transfected into cells, one expressing only the SINV replicase nsPs, which enables the replication of the second mini-genome plasmid via the expression of the replicase complex and specifically the nsP4, RdRp (Figure 6.21A). The expression of the replicase is under the control of the human CMV promoter, widely expressed in human cells. The second plasmid, encoding the SINV mini-"genomic" RNA, contains the sequence encoding the 5'UTR, the N-terminus of nsP1, followed by luciferase *firefly* reporter (Fluc). It further encodes a second luciferase reporter encoding the *gaussia* protein (Gluc) directly after the subgenomic (SG) promoter (Figure 6.21A). For simplicity, the synthesis of the full-length RNA serving as a template for Fluc expression is termed "replication", and the RNA synthesised from the SG promoter serving as a template for Gluc expression is termed "transcription".

To study whether DDX1 and, consequently, the tRNA-LC targets replication or transcription in SINV infection, I used cells transfected with siRNA targeting DDX1 and a scramble control sequence as the background for the *trans*-replicase assay. After 18h post-transfection, I measured the *firefly* and *gaussia* luciferase expression (Figure 6.21B). I observed that in the absence of DDX1, there is significant upregulation of the luciferase expression in both replication and transcription over the siCTRL. This indicates that DDX1 may be involved in inhibiting replication/transcription in the viral lifecycle.

To specifically study the effect of tRNA-LC on SINV translation, I designed and cloned a novel SINV chimeric virus with a *Renilla* luciferase (Rluc) reporter inserted into nsP3 (Figure

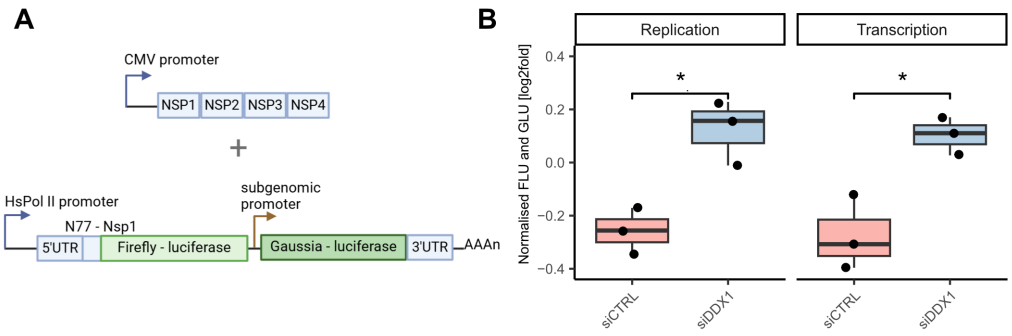

Figure 6.21: **DDX1 knockdown upregulates SINV replication and transcription**

A Schematic of *trans*-replicase plasmids. **B** Luciferase signal (Fluc and Gluc) from three independent biological replicates siRNA treated cells transfected with SINV-P1234 and SINV-Fluc/Gluc normalised to cells transfected with the catalytic mutant control SINV-P1234$_{GAA}$ and SINV-Fluc/Gluc. Log2fold change was calculated over the no-siRNA control and plotted. Statistical significance was tested using a two-tailed T-test (P $\leq$0.05 *).

6.22A). Two variants were created: the first encoded a replicative chimeric virus in which the *Renilla* protein was fused to nsP3, similar to the SINV$_{nsp3-mScarlet}$ described in Chapter 5. The second contained a double-stop codon following the *Renilla* insert, preventing further translation and, importantly, the formation of the replication complex. Both constructs enabled the detection of the initial round of translation upon cytoplasmic entry of the viral RNA. However, the readthrough Renilla RNA version will replicate and produce more viral gRNA and sgRNA, while the luciferase expression of the construct containing the stop codons can only proceed from the translation of the incoming particles. Therefore, the construct with the stop codons after *Renilla* ensures higher confidence in detecting initial viral translation, as it eliminates potential replication-associated signal amplification. I tested these SINV luciferase constructs by transfecting their *in vitro* transcribed RNA into HEK293 cells and measured luciferase expression at 2, 4, 6, and 18 hpt (Figure 6.22B). As expected, the replicative luciferase construct exhibited an exponential increase in signal over time. In contrast, the non-replicative luciferase construct peaked at 4 hpt, followed by a gradual decline in signal intensity.

To assess the role of tRNA-LC in translation, I transfected *in vitro* transcribed SINV-luciferase RNA into HEK293 cells treated with siRNA targeting DDX1, CGI99, or a scramble control sequence, as previously described in Chapter 5. At 4hpt, the absence of DDX1 resulted in a mild but reproducible increase in *Renilla* signal in the replicative SINV-Rluc and a significant upregulation of the non-replicative SINV-Rluc-TAA construct (Figure 6.22C). Meanwhile, the CGI99 knockdown had no effect compared to the control. These results suggest that DDX1 has a mild regulatory role in gRNA translation, while CGI99 appears to be dispensable.

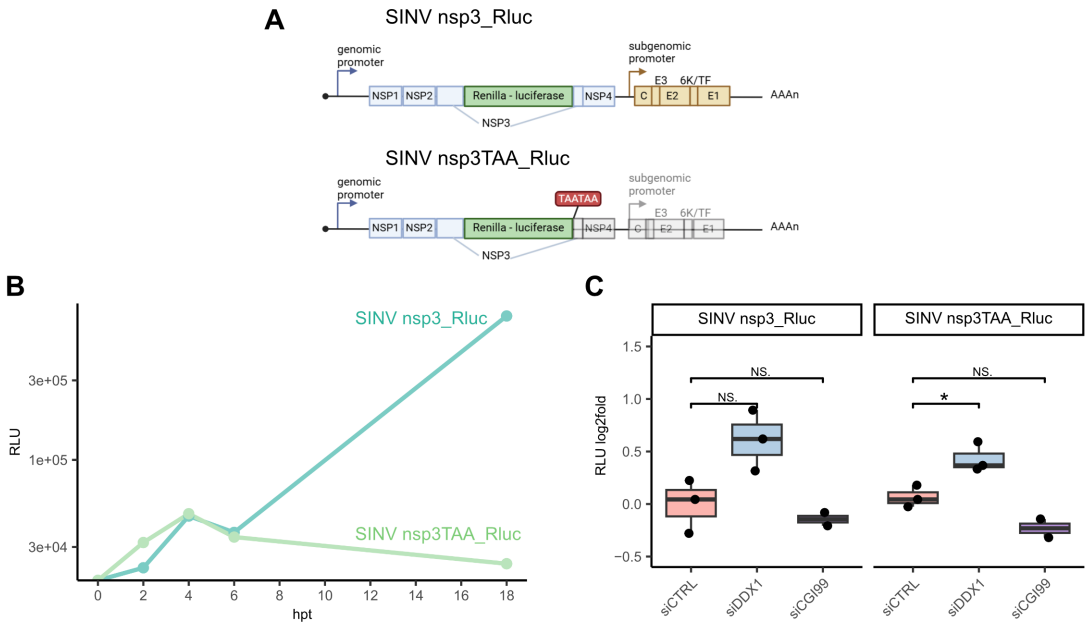

Figure 6.22: **DDX1 knockdown upregulates initial SINV translation**

A Schematic of of SINV-nsp3-Rluc and SINV-nsp3TAA-Rluc. **B** Luciferase (renilla) signal detected in preliminary testing performed in HEK293 cells after transfection of *in vitro* transcribed SINV-nsp3-Rluc and SINV-nsp3TAA-Rluc at 2, 4, 6 and 18hpt. **C** Luciferase signal (Fluc and Gluc) from three independent biological replicates siRNA-treated cells transfected with *in vitro* transcribed SINV-nsp3-Rluc or SINV-nsp3TAA-Rluc at 4hpt. Log2fold change was calculated over the no-siRNA control and plotted for each SINV-Rluc construct. Statistical significance was tested using two-tailed T-test ($P \leq 0.05$ *, ns: not-significant).

Overall, the DDX1 interactome was explored in both uninfected and infected conditions. In the PPI, a core set of proteins was identified, which do not change their association to DDX1 in the different conditions tested. The other tRNA-LC components emerge as core interactors together with ribosomal proteins. During infection, a subset of proteins significantly altered their binding to DDX1 via an RNA bridge, which suggests global changes in tRNA-LC-containing RNPs. The iCLIP2 analysis revealed that the tRNA-LC binds a number of mRNAs in uninfected conditions. However, a significant shift in interactions with vRNA is observed during infection. Binding to the vRNA is predominantly at the 5' end of the sgRNA. The pattern of RNA binding and the PPI alluded to a regulatory function in translation control. Finally, the decoupling of SINV replication and translation revealed that the absence of DDX1 upregulated both processes.

6.3 Discussion

DDX1 is a nuclear and cytosolic protein with a wide range of roles in RNA metabolism, through its involvement in the tRNA-LC as well as other protein complexes. I have previously shown that DDX1 plays an antiviral role in SINV infection, so its interactome could reflect how this function is exerted. Understanding which protein partners the tRNA-LC interacts with over the course of infection can provide invaluable clues on the role it plays during infection. To this effect, I performed a DDX1 PPI to study the protein-protein interaction dynamics that occur over the course of infection. I used native wash conditions (150 mM) to capture physiologically relevant interactions and processed several negative controls in parallel to ensure the capture of bona fide DDX1 interactions. Combined with a state-of-the-art mass spectrometer used for LC-MS/MS analysis allowed for even peripheral and indirect interactions to be captured. Building upon the previous high-resolution complex analysis by XL-MS, this study aimed to understand the landscape of interactions that DDX1 broadly establishes in the cell and its dynamics after infection.

This study further explores the tRNA-LC's RNA binding specificity. I characterised the RNA-specific interactions of two tRNA-LC proteins by iCLIP2 sequencing. The novel RNA binding profiles identified through this technique identify potential RNA regulatory roles in steady-state conditions and during infection. Combined, these two techniques enabled the robust characterisation of the DDX1 and the tRNA-LC interactome in both uninfected and infected conditions.

6.3.1 DDX1 core interactors and implications in steady-state conditions

DDX1 was identified to bind to a large number of proteins that intriguingly did not change in the different conditions tested. I identified a core-DDX1 group of proteins that were characterised by the lack of RNase sensitivity and were unchanged by the cellular remodelling that occurs during virus infection, among them the tRNA-LC proteins. A recent study on RBP RNA dependence by Caudron-Herger and colleagues in 2019, by a technique called R-DeeP, assessed proteins' RNA dependence by comparing their sucrose gradient migration profiles in the presence and absence of RNases [294]. Migration in the presence of RNases suggested that the protein interactors are RNA-dependent. In the case of DDX1, little to no shift in the peaks was observed, suggesting that RNA does not play a role in the interacting partners. This observation is consistent with my findings regarding the core-DDX1 proteins. However, a

few minor differences exist within my dataset. A small fraction of the total proteins binding to DDX1 alter their binding to DDX1 in an RNA-dependent manner, which is pronounced in the infected environment. Upon further investigation, the RNA-dependent proteins associate with the same protein complexes as the proteins associating with DDX1 in an RNA-independent manner, such as the ribosome. This indicates that the RNA is stabilising parts of the same complex bound by DDX1.

DDX1 core-interactors are predominantly interacting with ribosomal proteins. In the iCLIP2 data, the two tRNA-LC proteins predominantly interact with 5'UTRs and the start of CDS of cellular RNA. This pattern matches the binding profiles of translation initiation proteins such as eIF3C and eIF4A2 [295, 296]. The RNA-binding profile coupled with the translational protein interactors suggests an important role of the tRNA-LC in translation initiation. This correlates with a previous study where DDX1 was observed to regulate the translation of insulin mRNA [259]. The latter study carried out a PPI of DDX1 in insulinoma cells and identified similar enrichment of translation proteins, in particular translation initiation proteins. They further identified that DDX1 plays a regulatory role in insulin mRNA translation, whereby RNA binding to the 5'UTR of the transcripts promotes translation. Meanwhile, the removal of DDX1 from this RNA location by phosphorylation downregulated insulin translation [259]. This study is extremely interesting in the context of my data on DDX1 cellular behaviour. The similar binding on cellular transcripts and ribosomal association suggests DDX1 regulates the translation of a variety of proteins in a similar fashion as previously described for insulin. Future experiments could explore whether the translation of the RNA targets identified is influenced by DDX1 binding. Employing phosphorylation-inducing drugs such as palmitate could induce the dissociation of DDX1 from its mRNA targets. I could pair this with click-chemistry [297] to tag and separate the newly synthesised protein to quantify the translation rate in accordance with DDX1-RNA binding. The mRNAs bound by DDX1 in the iCLIP provide a novel and interesting list of target proteins for further investigation.

Among the tRNA-LC proteins, CGI99 has also been reported to be directly involved in translation control [246]. Indeed, CGI99 was co-precipitated with other members of the tRNA-LC in the study of RNA cap protein interactors, employing cap analogue-containing resins. Interestingly, within the CGI99 interactome, eIF4E, the cap-binding factor required for canonical translation, was absent. Researchers suggested that CGI99's cap-binding ability, through its stabilisation by tRNA-LC members, replaced the canonical binding of eIF4E [246]. The two

tRNA-LC proteins interact with similar cellular RNA binding regions, namely the 5'UTR. In the DDX1-IP, besides the enrichment of the tRNA-LC proteins, I observed an enrichment of 5'UTR mRNA binding proteins, which could further support CGI99's cap-associated activity. Among the core-DDX1 protein interactors, I identified eIF proteins such as eIF3A/B/C and eIF4B/G, but not eIF4E. It would be interesting to explore whether the identified mRNA targets are translated in a canonical manner or whether their binding by the tRNA-LC proteins is due to their unusual translational regulation.

6.3.2 DDX1 core interactors and implications in SINV conditions

In Chapter 4, I characterised the tRNA-LC as the primary complex to which DDX1 associates. In the DDX1-PPI, I identified all tRNA-LC members as highly enriched across all conditions, which classified them as core-DDX1 interactors. The coordinated RNA binding preferences identified in the iCLIP2 data further strengthen their close functional interaction. The antiviral role observed in Chapter 5 indicates that the tRNA-LC members play a crucial role in inhibiting viral progression, which may stem from the tRNA-LC working alone or cooperating with other antiviral factors, exerting pressure on the vRNA.

The tRNA-LC proteins are associated with a large number of ribosomal proteins. During infection, a significant enrichment of translational factors occurs via an RNA bridge as the infection progresses, in line with the significant increase in vRNA in the cellular environment [113]. Viruses are known to hijack cellular machinery, in particular, to enhance the translation of their own vRNA [298]. The proximity of the tRNA-LC proteins to the translational factors recruited for vRNA translation could reflect the antiviral role of the complex. The iCLIP2 data showed an absence of binding to the viral negative strand, suggesting the tRNA-LC proteins interact specifically with the positive strand, which serves as the template for translation and replication. This lack of interaction with the negative strand may also imply that these proteins do not bind the dsRNA replication intermediate. However, confirming this would require analysis at an earlier timepoint. The 16hpi timepoint used in my dataset may be too late to detect interactions with the negative strand, as it is likely sequestered within replication organelles and thus inaccessible. Interestingly, the specific decoupling of replication and translation indicated that the absence of DDX1 caused an upregulation of both processes. In both cases, the effect was mild, suggesting the tRNA-LC antiviral regulation may be occurring elsewhere, or the combination of inhibiting both processes significantly hinders the

viral lifecycle. The binding of the tRNA-LC was remarkably enhanced in the subgenomic region of RNA, specifically at the 5'UTR of the sgRNA and the start of the polyprotein encoding the structural proteins. The higher prevalence of sgRNA could corroborate the signal strength. However, I hypothesise that a more enhanced regulation will be detected if perturbing specifically the translation within this latter region. In future work, I would employ replicons to report the translation of the sgRNA specifically and capture the later stages of infection.

In the scope of the tRNA-LC working as an independent antiviral factor, the complex has an array of functions that could model its antiviral activity. Focusing on the start of the sgRNA region where the strongest binding of the tRNA-LC proteins was observed, a few possibilities come to light. The 5'-UTR leader sequence contains a m7G cap structure at its 5'-end, promoting RNA stability [299]. This leader sequence confers eIF4F complex independence and is implicated in the inhibition of host translation [124, 142]. The tRNA-LC protein CGI99 is capable of binding cap structures [246]. I can speculate that the direct binding of the cap structures inhibits other translation initiation factors from binding and hinders the host translation shut-off from occurring.

DDX1 helicase activity could unwind critical secondary structures necessary for protein recruitment. At the start of the coding region of the capsid protein, an essential hairpin (stem-loop RNA) enables efficient sgRNA translation. The downstream stable hairpin (DSH) is positioned 27nt downstream of the AUG codon [126]. The DSH confers eIF2-independent translation and signals the precise codon at which translation begins [126, 141]. The absence of this structure hinders correct translation from occurring [286]. DDX1 could dissolve the required structure for correct ribosomal binding, hindering translation, for instance. Finally, I previously hypothesised that RTCB is saturated by vRNA and unable to efficiently recycle its ligase activity (explored in Chapter 5). The iCLIP2 indicates definite binding of RTCB to the vRNA. I can thus speculate that its "locked" form inhibits translational factors from scanning and binding the necessary sequence. Combined, the tRNA-LC is a powerful toolbox that could hinder translation initiation.

6.3.3 Virus infection models select cellular complexes

A subset of DDX1 protein binding partners alter their interaction during infection in an RNA-dependent manner. The previously mentioned RNase insensitivity in steady-state conditions does not entirely corroborate cellular changes that occur in a virus-infected environment. The

iCLIP2 analysis identified a significant shift from cellular RNAs to vRNAs, which could explain the RNA-dependent changes in the PPI.

In mock conditions, DDX1 is identified as being involved with the ubiquitin-proteasome system (UPS), such as the proteasome complex. In the SINV comparative analysis, I observed a notable shift away from proteasome-associated proteins during infection. Intriguingly, the UPS plays a crucial role in the establishment of productive virus infection across various virus species. UPS is hijacked by the virus to degrade cellular proteins inhibitive to its life cycle and, in some cases, used for ubiquitination of viral proteins [300]. Proteasome inhibitors, MG132 and bortezomib, and UPS signalling pathway inhibitors have been shown to inhibit viral replication in alphaviruses [285, 301–303], and other RNA viruses such as SARS-Cov2 [304]. For these reasons, it is interesting that in my dataset, the very proteins that seem essential for efficient viral infection are dissociating from DDX1 over the course of infection. I found UPS-related proteins to interact significantly more with DDX1 in Mock than in SINV conditions. From the literature, it suggests that proteasome-related functions are necessary for infection, whilst I have shown that the tRNA-LC is inhibitory to SINV. Their opposing function in infection can explain this change in interaction. Alternatively, the RNA that bridges these proteins changes over the course of infection, whereby DDX1 binds vRNA, and proteasomal proteins act on cellular RNA to suppress the host viral response.

During viral infection, the cell undergoes a heavy reorganisation in order to establish appropriate localisation of viral particles, proteins, and vRNPs [305]. The movement of vRNPs is thought to be aided by motor proteins such as ones of the myosin complex. The host membranes are rearranged into cytoplasmic structures known as type-1 cytopathic vacuoles (CPVs) [306]. The transport of endocytosed spherules and assembly of CPVs is dependent on cellular cytoskeletal elements such as the myosin complex [307, 308]. In the SINV 18hpi PPI of DDX1, the myosin complex proteins, including MYH9, were significantly enriched. MYH9 was previously identified as a direct interactor by XL-MS in tRNA-LC in Chapter 4. This protein has also been identified in eight vRNA interactomes in different viral species, highlighting its direct involvement in vRNA binding [115] and its potential role in the viral lifecycle. The enrichment of the myosin complex protein during viral infections, as well as specifically in my DDX1 PPI, is intriguing. This suggests that DDX1, in cooperation with the myosin complex, is manipulating vRNP cellular movement. In future experiments, I would seek to employ myosin-specific inhibitors that could disrupt the cellular movement of the tRNA-LC and, subsequently,

determine whether its movement is critical for its antiviral function.

An intriguing association of DDX1 identified as enriched during infection is the increased interaction with nuclear splicing factors. SF3B2 and its associated RNA splicing proteins are among these proteins. Notably, Kamel et al. (2024) documented the dynamic shuttling of nuclear factors to viral factories during the course of infection, in particular the SF3B complex proteins [123]. The full SF3B complex and U2 snRNA (collectively known as U2 snRNP) were observed to suppress viral infection through direct binding to SINV RNA [123]. Intriguingly, in the rat DDX1 interactome study, DDX1 was identified as an essential regulator in alternative splicing events and was characterised to interact with core spliceosomal and spliceosome-associated proteins (namely DDX5, DHX15, SF1, SF3B1 and SF3B2 among others) [284]. In the original RTCB co-IP that identified the tRNA-LC components, the SF3b complexes were also captured [197]. Here, I mapped the predicted binding locations of the U2 snRNP complex to my iCLIP2 data and identified two of the predicted sites as correlating to proximal regions bound by the tRNA-LC proteins.

The connection between U2 snRNP and tRNA-LC proteins warrants future investigation, specifically in their relation to vRNA binding specificity. Cross-referencing a CLIP analysis of the U2 snRNP, for example, may highlight whether the tRNA-LC and U2 snRNA bind cohesively at similar locations.

This particular

Intriguingly, in the rat DDX1 interactome study, DDX1 was identified as an essential regulator in alternative splicing events and was characterised to interact with core spliceosomal and spliceosome-associated proteins (namely DDX5, DHX15, SF1, SF3B1 and SF3B2 among others) [284]. In the original RTCB co-IP that identified the tRNA-LC components, the SF3b complexes were also captured [197].

Several nuclear proteins that have relocalised to the cytoplasm are significantly enriched in the infected DDX1-PPI. SF3B2 and its associated RNA splicing proteins are among these proteins. Notably, Kamel et al. (2024) documented the dynamic shuttling of nuclear factors to viral factories during the course of infection, in particular the SF3B complex proteins [123]. The SF3B complex was further characterised in the latter study, to inhibit viral gene expression in a splicing-independent manner. The full SF3B complex and U2 snRNA (collectively known as U2 snRNP) suppress viral infection through direct binding to SINV RNA [123]. Intriguingly, in the rat DDX1 interactome study, DDX1 was identified as an essential regulator in alternative

splicing events and was characterised to interact with core spliceosomal and spliceosome-associated proteins (namely DDX5, DHX15, SF1, SF3B1 and SF3B2 among others) [284]. In the original RTCB co-IP that identified the tRNA-LC components, the SF3b complexes were also captured [197]. The identification of splicing factors in my DDX1-IP may suggest the antiviral role of the tRNA-LC stems from its association with the SF3B complex and U2 snRNA antiviral function. I hypothesise that the tRNA-LC binding to vRNA is occurring due to its recruitment from other factors rather than sequence specificity. I speculate the binding of the U2 snRNP is aiding the tRNA-LC vRNA binding and exerting its antiviral role. Kamel et al. (2024) identified three perfect complementary U2 snRNA branch site interacting stem-loop motifs present in the SINV genome [123]. Two of the three sites correlate with increased binding sites identified in the iCLIP2 of the tRNA-LC. The connection between U2 snRNP and DDX1 warrants future investigation, specifically in their relation to vRNA binding specificity. Cross-referencing a CLIP analysis of the U2 snRNA, for example, may highlight whether the tRNA-LC and U2 snRNA bind cohesively at similar locations.

In cellular mRNA, the tRNA-LC proteins were identified to bind G-rich sequence region with no discernible motif or secondary structure preference. In infection, the binding profile indicated a mild correlation to G-rich regions. However, the binding profile observed across the viral genome suggests G-rich and predicted G4 regions are not necessarily the binding drivers. The cellular and viral data is interesting since DDX1 has previously been shown to bind directly with RNA G4. Indeed it was shown to be an essential modulator of class switch recombination (CSR) at the immunoglobulin heavy-chain (IgH) locus by binding RNA G4 and aiding in the conversion to R-loops [247]. In this case, the helicase activity of DDX1 is employed to dissolve the secondary structure and enable R-loop formation. The frequency of G4 binding in my captured cellular mRNA is around 2% at the high stringency threshold, indicating that G4 binding occurs in rare mRNAs. The lack of RNA sequence binding specificity of the tRNA-LC indicates that other factors are coordinating the recruitment of the complex to specific regions of the RNA.

DDX1 has previously been reported to interact with a variety of protein complexes beyond the tRNA-LC complex, one of which is the DDX1/DDX21/DHX36 complex. DHX36 is a core interactor of DDX1 (identified in the PPI of DDX1-GFP IP), while DDX21 appears consistently in the core-DDX1 interaction network across all but one comparison, suggesting a stable and robust interaction. Both DHX36 and DDX21 are well-characterized G4-binding proteins

[309–311], and it is plausible that DDX1's interaction with specific G4 sequences is facilitated through cooperation with these proteins. Regarding the DDX1 iCLIP dataset, it is possible that I am detecting the binding of DHX36, given that its molecular weight is within the range that DDX1 could capture during iCLIP2 processing. However, I believe this is unlikely due to the IP conditions used, which were highly stringent (involving 2M salt washes), which should predominantly retain only the strongest interactions, such as those observed with the tRNA-LC proteins. Additionally, published PAR-CLIP data for DHX36 reveals distinct RNA binding domains, with a preference for binding to 3' UTR regions of protein-coding sequences [312], which is notably different to what I observed for DDX1. Both datasets also indicated a sharp peak of binding at the start of the CDS, suggesting an overlap in binding mechanism in this region. Further analysis of both datasets is required to precisely distinguish whether the two proteins are binding the same RNA transcripts and whether there is indeed an overlap in binding preferences. However, the difference in binding preferences between DHX36 (as seen in PAR-CLIP) in the 3'UTR and DDX1 (as captured in my iCLIP2 dataset) in the 5'UTR indicates that the interactions I observed are specifically attributed to DDX1, rather than to DHX36 or other associated proteins.

The DDX1 and tRNA-LC interactome explored in this chapter reveal novel functional roles for the protein complex. In steady-state, the core-DDX1 protein interactors combined with the iCLIP2 data portray an important role in cellular translation more comprehensive than previously known. The novel mRNA binding targets identified here could direct future research avenues.

During infection, the large group of core-DDX1 proteins that remain unchanged during the various conditions indicated that DDX1 and the tRNA-LC are not gaining novel protein-interactors or subsequent functional repurposing during infection. Instead, an enhancement or suppression of a certain subset of protein groups is taking place. The core-DDX1 proteins overlap with dynamically enriched proteins. The observation of these significant changes occurs only in RNA-dependent comparisons, which could be the driver of these changes. Through the iCLIP2 analysis of the tRNA-LC, I observed a significant shift to vRNA binding during infection, which could explain the heightened activity of certain complexes over others. I further investigated in detail viral replication and translation and identified DDX1 as a potential suppressor of both of these activities.

7 General discussion and future directions

7.1 Summary

In this thesis, I have outlined my contribution to understanding the composition and functional importance of the tRNA-LC in uninfected and SINV-infected cells. I have identified that the tRNA-LC may facilitate a wide scope of cellular processes, including translation control. Amongst these roles, I observed that the tRNA-LC may play an antiviral role in SINV infection.

In steady-state conditions, I explored the intricate assembly of the complex and identified CGI99 as a central protein. Co-immunoprecipitation of the tRNA-LC proteins in HEK293 cells indicated consistent precipitation of other members of the complex, even under the most stringent wash conditions. Knockdown of DDX1, RTCB, and CGI99 highlighted the co-dependency of these core members of the tRNA-LC in complex assembly and stability. Loss of each tRNA-LC member individually caused a reduction in greater or lower degree of the other members of the complex, while their RNA levels remained unchanged. These results imply that the stability of the tRNA-LC proteins is linked to their co-assembly.

To further characterise the organisation of the tRNA-LC, I employed XL-MS *in situ*. This allowed me to identify protein interfaces within the native tRNA-LC, which could then be validated by AF3 modelling. The results of this analysis confirmed the central position of CGI99 within the complex and identified novel interaction partners of the complex, including the ribosomal factor RPL11 and the non-muscle myosin MYH9. These novel interactions were further validated in a high quality co-IP-MS experiment. In keeping with the directly crosslinked interactors, I found that DDX1's interactome was enriched in proteins linked to translation and intracellular mobility. Excitingly, the results of iCLIP2 analysis further indicated the role of tRNA-LC in translation control. The tRNA-LC was identified to bind predominantly the 5'UTR and start of CDS of cellular mRNAs. This exciting combination suggests that the tRNA-LC is likely involved in regulating translation initiation, which is consistent with CGI99 cap-binding activity [246]. The identified list of mRNAs bound by the tRNA-LC warrants further research. One possibility is that these interactions speak to a non-canonical translation initiation mechanism that could have far-reaching roles in cell biology.

In parallel, I investigated the role of the tRNA-LC during SINV infection, building on previous studies that identified tRNA-LC components as direct interactors with SINV vRNA and observed their relocalisation to viral replication organelles [113, 123]. While these studies

provided evidence of physical interactions, the regulatory roles of the tRNA-LC components remained unexplored. Using siRNA knockdown, I identified a potential antiviral role for the tRNA-LC in SINV infection. In siCTRL transfected cells, infection with SINV at 0.1 MOI caused imperceptible changes in the cellular transcriptome due to a low penetrance of infection. However, over 11000 genes were downregulated when CGI99 or DDX1 were absent, suggesting that the lack of these proteins enhanced virus infection and spread, leading to a broader RNA degradation, comparable to wild-type cells infected at higher MOI. Further analysis of the DDX1 RNA targets during SINV infection revealed its transition toward vRNA binding and increased associations with ribosomal factors, mediated by stabilising vRNA interactions. This shift implied a regulatory role for DDX1 in both viral replication and translation. Functional assays that decoupled these essential viral processes confirmed the involvement of tRNA-LC in regulating both replication and translation.

7.2 Exploring the tRNA-LC architecture and interactome

The integration of XL-MS with AF3 enabled the capture of the tRNA-LC *in cellulo*, significantly advancing our understanding of the complex assembly and its interactions. DSS crosslinks revealed a CGI99-centric architecture, supporting the pivotal role of this protein in the structural and functional dynamics of the tRNA-LC. The AF3 simulations further highlighted the critical role of the C-terminal region of CGI99 as a platform to establish multiple interactions with FAM98A/B and DDX1.

In situ XL-MS presents several challenges. Fewer crosslinks are detected *in cellulo* cross-linking compared to *in vitro* cross-linking, as observed in my data compared to in vitro data generated by Kroupova et al. (2021) [243]. *In cellulo* crosslinking has an inherently lower crosslinking depth. To increase crosslinking efficiency, higher cross-linker concentration or increased incubation time can be implemented [242]. However, a balance between capturing genuine crosslinks and not pivoting towards crosslinker saturation is critical to avoid the capture of biologically irrelevant interactions. To enrich for low-abundant cross-linked peptides and separate them from unmodified peptides, other approaches can be used, such as the use of biotinylated cross-linkers that can be trapped using avidin beads [313]. However, this approach comes at the cost of using bulkier crosslinkers that may fail to crosslink compact interfaces and may result in longer crosslinking reach. Detecting transient or low-abundance interactions is challenging, and indeed, I was unable to capture the transient interaction occurring between

RTCB and Archease [198, 244]. Improving the enrichment of crosslinked peptides will improve depth, which may facilitate the discovery of transitory interactions. Furthermore, subcellular fractionation could enable the capture of specific nuclear or cytosolic tRNA-LC subpopulations while reducing sample complexity and increasing the likelihood of detecting substoichiometric crosslinked peptides.

The observed crosslinks within the tRNA-LC serve as a proof of concept for the ability to monitor protein complex dynamics. If we overcome the technical difficulties of *in situ* XL-MS depth, our approach could theoretically be extended to larger and more intricate protein assemblies, offering a powerful tool for structural and functional studies of macromolecular complexes. Investigating these dynamics under varied cellular stimuli could further reveal how the composition of cellular and viral complexes is influenced by environmental cues. Emerging quantitative crosslinking approaches, such as those described by Wippel et al. (2022), could refine our understanding by resolving subcomplex formation and detecting subtle intramolecular and intermolecular changes, enhancing the resolution of conformational landscapes [314]. By applying quantitative crosslinking approaches in the context of different viral infection stages, we could capture conformational changes that are mediated by host-virus interactions, such as those occurring in the tRNA-LC.

I also identified a pool of cellular RNAs that the tRNA-LC binds. Interestingly, a proportion of the mRNAs bound by the tRNA-LC are associated with the mitochondria. This agrees with earlier observations of Pazo and colleagues in 2019, where they sequenced RNAs bound by the tRNA-LC protein, CGI99 [246]. Despite these associations, the tRNA-LC has not been previously characterised in the mitochondria, leaving its role at this location unknown. The mitochondrial genome encodes 37 proteins processed inside the mitochondria [315]. Meanwhile, a significant proportion of mitochondrial proteins are encoded in the genome and processed by cellular factors before trans-locating to the organelle. Interestingly, the 22 mitochondrial tRNA (mtRNA) are encoded and processed within the mitochondrial organelle. Maturation of mtRNA does not rely on cytoplasmic proteins and, unlike human tRNAs, does not include an intron requiring splicing [316, 317]. Surprisingly, I found mtRNAs associated with the tRNA-LC in my iCLIP2 dataset. The tRNA-LC ligase activity is essential for tRNA maturation. However, the absence of intron splicing of mtRNA suggests this precise function is not required at this location. The tRNA-LC ligase activity requires precise RNA termini for ligation, specifically 2',3'-cyclic phosphate and 5'-hydroxyl groups. A potential link exists with

the phosphatase ANGEL2, which converts 2',3'-cyclic phosphates into 2',3'-hydroxyl termini [318]. ANGEL2's activity has been shown to be crucial for removing cyclic phosphate groups from mitochondrial RNA cleavage products, enabling downstream RNA processes [319]. While a direct connection between ANGEL2 and the tRNA-LC has not yet been established, ANGEL2's role in preparing RNA substrates hints at possible competitive function or negative regulation of ligation processes by the tRNA-LC within the mitochondria. These findings open intriguing possibilities for exploring the tRNA-LC's role in mtRNA processing and broader mitochondrial RNA regulation. Future investigations could clarify whether the tRNA-LC directly participates in mitochondrial tRNA maturation or interacts indirectly with mitochondrial RNA metabolic pathways. Advances in mitochondrial proteomics have enabled the capture of 1100 proteins, among them tRNA-LC components [320, 321]. Mitochondrial function, either oxygen consumption rate or ATP production measurement after the knockdown of tRNA-LC proteins, could highlight the importance of this complex within this cellular organelle.

The cellular RNAs bound by the tRNA-LC hint at a much larger regulatory role than previously thought. The identified RNAs could direct future research in understanding particular gene regulations and whether their expression requires noncanonical translation or is regulated post-transcriptionally. In cancer research, for example, DDX1 has emerged as a significant player in cancer development and progression, displaying both oncogenic and tumour-suppressive functions depending on the context [322]. In breast cancer, elevated DDX1 RNA expression and increased cytoplasmic DDX1 protein levels correlate with early breast cancer recurrence, highlighting its potential as a biomarker for breast cancer screening [323]. In colorectal cancer, DDX1 promotes tumorigenesis by transcriptionally activating the *LGR5* gene, a critical factor in the tumorigenicity of colorectal cancer cells [324]. Additionally, in testicular germ cell tumours, DDX1 drives the activation of stem cell-related genes, such as *cyclin D2*, indicating further a role in the regulation of gene expression [325]. The RNAs bound identified in my iCLIP may be involved in cancer progression, for example. Crossreferencing the identified RNAs with oncogenic targets may highlight potential novel avenues of research. The tRNA-LC has been shown to stabilise mRNAs in non-membrane organelles [206], actively transport RNA in dendrites [246] and here in translation regulation.

7.3 Implication of the tRNA-LC in viral infection

This thesis aimed to uncover the roles of the tRNA-LC in SINV infection by analysing different aspects of its biology, including the importance of its enzymatic activities, its impact on the transcriptome, and its compositional remodelling upon infection. Binding to the 5' UTR of cellular and viral RNAs suggested that the tRNA-LC probably regulates translation. I performed experiments to uncouple the importance of the tRNA-LC in replication and translation, which suggested that it might impact both processes in a multifaceted function. Different viral interactome capture methods identified DDX1 as a direct interactor with the vRNA of SINV, CHIKV, ZIKV, DENV, and SARS-CoV-2 [115]. These findings suggest broader applications of tRNA-LC regulatory mechanisms to other positive-stranded RNA viruses. Systematic characterisation of each virus with known DDX1-vRNA interaction could identify whether the features identified in SINV are applicable to other viruses.

The broad cellular activities of the tRNA-LC position it as a possible antagonist of viral processes that may be universally required across viral species and families. Alphaviruses, such as CHIKV and VEEV, present a public health concern due to their high transmissibility, pathogenicity, and the expanding range of arthropod vector endemic areas. Tools for studying the replication mechanisms of these viruses have been extensively developed, offering an opportunity to examine whether the tRNA-LC plays a conserved role across the alphavirus genus. Notably, the replicase complexes of these viruses can cross-utilise RNA templates from other alphaviruses [292, 326, 327], which underscores a shared strategy for replication to occur. If the tRNA-LC is indeed involved in facilitating or regulating replication and translation, its role could extend beyond SINV to other viruses within the genus. In future work, I would systematically explore the involvement of the tRNA-LC in the replication and translation of different alphaviruses using a similar luciferase reporter system employed in this thesis. *Trans*-replicase system has been developed to study alphaviruses by the Merits group and could be applied here for different viruses.

To investigate the involvement of the tRNA-LC in the pioneering round of viral translation, novel chimeric viruses would need to be constructed to include a luciferase reporter. However, transfecting vRNA can be challenging, particularly when the proteins of interest are involved in innate immune responses, which introduces additional complexity. Alternative approaches include inhibiting viral replication using chemical compounds, such as protease inhibitors or nucleoside analogues. For instance, a previous study exploring G3BP1's role in translation

initiation in noroviruses utilized 2'-C-methylcytidine (2CMC) to inhibit viral replication [328]. In this work, strand-specific qPCR confirmed the suppression of negative-sense RNA synthesis, while polysome fractionation was used to determine which vRNAs were actively associated with ribosomes for translation. A similar approach could be adapted to assess tRNA-LC involvement in vRNA translation initiation. Identifying reliable protease inhibitors or nucleoside analogues is critical for such experiments. Notably, several studies have highlighted promising candidates [326]. For example, RA-0002034, a covalent fragment with a vinyl sulfone warhead, was recently shown to inhibit the replication of CHIKV and other alphaviruses [329].

The understanding of cellular RBPs interacting with vRNA is aided by precise methods such as iCLIP. iCLIP may identify binding sites of RBPs; it does not, however, reveal the mechanisms or functions mediated by them (in this case, the tRNA-LC) at these locations. Co-IP data can contextualise the potential functional implications of tRNA-LC binding, for example, by suggesting protein partners that may support a particular role. I identified the tRNA-LC associated with ribosomal factors that may recruit tRNA-LC components to the translation start sites. Furthermore, I identified the tRNA-LC binding to the U2 snRNA complex, which could indicate antiviral function associated with this latter complex at specific binding sites [123]. Despite these insights, the precise driver of tRNA-LC recognition of RNA, i.e. whether sequence- or structure-specific, remains elusive. Cooperative binding with other proteins may facilitate the deposition of the complex in specific mRNA locations. Alternative approaches need to be deployed to understand whether the tRNA-LC is recruited by other factors to vRNA. For instance, the recently developed TREX (Targeted RNase H-mediated Extraction of crosslinked RBPs) technique by the Mardakheh group offers a promising avenue [330]. TREX captures *in vivo* RBPs associated with specific RNA regions via antisense tiling DNA oligos and RNase H digestion, followed by a Trizol-aided isolation of protein-RNA complexes. This method has already mapped region-specific interactomes for RNAs such as NORAD and 45S rRNA [330]. Applying TREX to SINV vRNA could uncover proteins binding specifically to the sgRNA 5' UTR, which is the most predominant binding site of the tRNA-LC based on my iCLIP results. TREX could thus provide new insights into how the tRNA-LC and its partners (identified by co-IP) interact with the vRNA. Analysis of the TREX experiment and my co-IP results could uncover cooperative interactions that may aid tRNA-LC binding to vRNA.

8 References

[1] A. M. Powers and John T. Roehrig. Alphaviruses. In John R. Stephenson and Alan Warnes, editors, *Diagnostic Virology Protocols*, pages 17–38. Humana Press, Totowa, NJ, 2011.

[2] J. Erin Staples and Ann M. Powers. 217 - Togaviridae: Alphaviruses. In Sarah S. Long, editor, *Principles and Practice of Pediatric Infectious Diseases (Sixth Edition)*, pages 1145–1147.e3. Elsevier, Philadelphia, January 2023.

[3] Marion C. Robinson. An epidemic of virus disease in Southern Province, Tanganyika territory, in 1952–1953. I. Clinical Features. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 49(1):28–32, January 1955.

[4] F. Simon, H. Tolou, and P. Jeandel. Chikungunya, l'épidémie que l'on n'attendait pas. *La Revue de Médecine Interne*, 27(6):437–441, June 2006.

[5] Kanti Laras, Nono C. Sukri, Ria P. Larasati, Michael J. Bangs, Rizal Kosim, Djauzi, Tony Wandra, John Master, Herman Kosasih, Sri Hartati, Charmagne Beckett, Endang R. Sedyaningsih, H. James Beecham, III, and Andrew L. Corwin. Tracking the re-emergence of epidemic chikungunya virus in Indonesia. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 99(2):128–141, February 2005.

[6] Jean-Paul Chretien, Assaf Anyamba, Sheryl A Bedno, Robert F Breiman, Rosemary Sang, Kibet Sergon, Ann M Powers, Clayton O Onyango, Jennifer Small, Compton J Tucker, and Kenneth J Linthicum. Drought-associated chikungunya emergence along coastal East Africa. *The American journal of tropical medicine and hygiene*, 76(3):405–407, March 2007.

[7] Man-Koumba Soumahoro, Pierre-Yves Boelle, Bernard-Alex Gaüzere, Kokuvi Atsou, Camille Pelat, Bruno Lambert, Guy La Ruche, Marc Gastellu-Etchegorry, Philippe Renault, Marianne Sarazin, Yazdan Yazdanpanah, Antoine Flahault, Denis Malvy, and Thomas Hanslik. The Chikungunya Epidemic on La Réunion Island in 2005–2006: A Cost-of-Illness Study. *PLOS Neglected Tropical Diseases*, 5(6):e1197, June 2011. Publisher: Public Library of Science.

[8] R. Angelini, A. C. Finarelli, P. Angelini, C. Po, K. Petropulacos, G. Silvi, P. Macini, C. Fortuna, G. Venturi, F. Magurano, C. Fiorentini, A. Marchi, E. Benedetti, P. Bucci, S. Boros, R. Romi, G. Majori, M. G. Ciufolini, L. Nicoletti, G. Rezza, and A. Cassone. Chikungunya in north-eastern Italy: a summing up of the outbreak. *Weekly releases (1997–2007)*, 12(47):3313, November 2007. Publisher: European Centre for Disease Prevention and Control.

[9] G. Rezza, L. Nicoletti, R. Angelini, R. Romi, A. C. Finarelli, M. Panning, P. Cordioli, C. Fortuna, S. Boros, F. Magurano, G. Silvi, P. Angelini, M. Dottori, M. G. Ciufolini, G. C. Majori, and A. Cassone. Infection with chikungunya virus in Italy: an outbreak in a temperate region. *The Lancet*, 370(9602):1840–1846, December 2007. Publisher: Elsevier.

[10] J. Erin Staples and Marc Fischer. Chikungunya Virus in the Americas — What a Vectorborne Pathogen Can Do. *The New England journal of medicine*, 371(10):887–889, September 2014.

[11] S. Cassadou, S. Boucau, M. Petit-Sinturel, P. Huc, I. Leparc-Goffart, and M. Ledrans. Emergence of chikungunya fever on the French side of Saint Martin island, October to December 2013. *Eurosurveillance*, 19(13):20752, April 2014. Publisher: European Centre for Disease Prevention and Control.

[12] Laura I. Levi and Marco Vignuzzi. Arthritogenic Alphaviruses: A Worldwide Emerging Threat? *Microorganisms*, 7(5):133, May 2019.

[13] José Henrique Oliveira, Ana Cristina Bahia, and Pedro F. Vale. How are arbovirus vectors able to tolerate infection? *Developmental & Comparative Immunology*, 103:103514, February 2020.

[14] Louis Lambrechts and Maria-Carla Saleh. Manipulating Mosquito Tolerance for Arbovirus Control. *Cell Host & Microbe*, 26(3):309–313, September 2019. Publisher: Elsevier.

[15] Rubing Chen, Suchetana Mukhopadhyay, Andres Merits, Bethany Bolling, Farooq Nasar, Lark L. Coffey, Ann Powers, Scott C. Weaver, and ICTV Report Consortium. ICTV Virus Taxonomy Profile: Togaviridae. *Journal of General Virology*, 99(6):761–762, 2018. Publisher: Microbiology Society,.

[16] Michele A. Zacks and Slobodan Paessler. Encephalitic alphaviruses. *Veterinary Microbiology*, 140(3):281–286, January 2010.

[17] Andreas Suhrbier, Marie-Christine Jaffar-Bandjee, and Philippe Gasque. Arthritogenic alphaviruses—an overview. *Nature Reviews Rheumatology*, 8(7):420–429, July 2012. Publisher: Nature Publishing Group.

[18] Gael Ten Broeck and Malcolm H. Merrill. A Serological Difference Between Eastern and Western Equine Encephalomyelitis Virus. *Proceedings of the Society for Experimental Biology and Medicine*, 31(2):217–220, November 1933. Publisher: SAGE Publications.

[19] Vladimir Kubes and Francisco A. Ríos. The Causative Agent of Infectious Equine Encephalomyelitis in Venezuela. *Science*, 90(2323):20–21, July 1939. Publisher: American Association for the Advancement of Science.

[20] John E. Greenlee. Chapter 19 - The equine encephalitides. In Alex C. Tselis and John Booss, editors, *Handbook of Clinical Neurology*, volume 123 of *Neurovirology*, pages 417–432. Elsevier, January 2014.

[21] H. Shore. O'nyong-nyong fever: An epidemic virus disease in East Africa: III. Some clinical and epidemiological observations in the Northern Province of Uganda. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 55(4):361–373, July 1961.

[22] Charles R. Anderson, Wilbur G. Downs, George H. Wattley, Norman W. Ahin, and Alick A. Reese. Mayaro Virus: A New Human Disease Agent. November 1957. Section: The American Journal of Tropical Medicine and Hygiene.

[23] K. C. Smithburn and A. J. Haddow. Semliki Forest Virus: I. Isolation and Pathogenic Properties. *The Journal of Immunology*, 49(3):141–157, September 1944.

[24] R. M. Taylor, H. S. Hurlbut, T. H. Work, J. R. Kingston, and T. E. Frothingham. Sindbis Virus: A Newly Recognized Arthropod-Transmitted Virus. September 1955. Section: The American Journal of Tropical Medicine and Hygiene.

[25] Diane E. Griffin, Foroozan Mokhtarian, Mahin M. Park, and Robert L. Hirsch. Immune Responses to Acute Alphavirus Infection of the Central Nervous System: Sindbis Virus Encephalitis in Mice. In P. O. Behan, V. Ter Meulen, and F. Clifford Rose, editors,

Progress in Brain Research, volume 59 of *Immunology of Nervous System Infections*, pages 11–21. Elsevier, January 1983.

[26] Pablo M. Irusta and J. Marie Hardwick. Neuronal Apoptosis Pathways in Sindbis Virus Encephalitis. In Covadonga Alonso, editor, *Viruses and Apoptosis*, pages 71–93. Springer, Berlin, Heidelberg, 2004.

[27] Jonathan H. Weston, Michael D. Welsh, Marian F. McLoughlin, and Daniel Todd. Salmon Pancreas Disease Virus, an Alphavirus Infecting Farmed Atlantic Salmon, *Salmo salar* L. *Virology*, 256(2):188–195, April 1999.

[28] Jonathan Weston, Stéphane Villoing, Michel Brémont, Jeanette Castric, Martin Pfeffer, Victoria Jewhurst, Marian McLoughlin, OddMagne Rødseth, Karen Elina Christie, Joseph Koumans, and Daniel Todd. Comparison of Two Aquatic Alphaviruses, Salmon Pancreas Disease Virus and Sleeping Disease Virus, by Using Genome Sequence Analysis, Monoclonal Reactivity, and Cross-Infection. *Journal of Virology*, 76(12):6155–6163, June 2002. Publisher: American Society for Microbiology.

[29] May La Linn, Joy Gardner, David Warrilow, Grant A. Darnell, Clive R. McMahon, Ian Field, Alex D. Hyatt, Robert W. Slade, and Andreas Suhrbier. Arbovirus of Marine Mammals: a New Alphavirus Isolated from the Elephant Seal Louse, Lepidophthirus macrorhini. *Journal of Virology*, 75(9):4103–4109, May 2001. Publisher: American Society for Microbiology.

[30] Iranaia Assunção-Miranda, Christine Cruz-Oliveira, and Andrea T. Da Poian. Molecular Mechanisms Involved in the Pathogenesis of Alphavirus-Induced Arthritis. *BioMed Research International*, 2013:973516, August 2013.

[31] Rivaldo V da Cunha and Karen S Trinta. Chikungunya virus: clinical aspects and treatment - A Review. *Memórias do Instituto Oswaldo Cruz*, 112(8):523–531, August 2017.

[32] Amanda M. Avila-Trejo, Lorena I. Rodríguez-Páez, Verónica Alcántara-Farfán, and J. Leopoldo Aguilar-Faisal. Multiple Factors Involved in Bone Damage Caused by Chikungunya Virus Infection. *International Journal of Molecular Sciences*, 24(17):13087, August 2023.

[33] Thomas E. Morrison, Lauren Oko, Stephanie A. Montgomery, Alan C. Whitmore, Alina R. Lotstein, Bronwyn M. Gunn, Susan A. Elmore, and Mark T. Heise. A Mouse Model of Chikungunya Virus–Induced Musculoskeletal Inflammatory Disease: Evidence of Arthritis, Tenosynovitis, Myositis, and Persistence. *The American Journal of Pathology*, 178(1):32, January 2011.

[34] Jean-Jacques Hoarau, Marie-Christine Jaffar Bandjee, Pascale Krejbich Trotot, Trina Das, Ghislaine Li-Pat-Yuen, Bérengère Dassa, Mélanie Denizot, Elsa Guichard, Anne Ribera, Tawfiq Henni, Frank Tallet, Marie Pierre Moiton, Bernard Alex Gauzère, Sandrine Bruniquet, Zaïnoul Jaffar Bandjee, Philippe Morbidelli, Gérard Martigny, Michel Jolivet, Frederick Gay, Marc Grandadam, Hugues Tolou, Vincent Vieillard, Patrice Debré, Brigitte Autran, and Philippe Gasque. Persistent Chronic Inflammation and Infection by Chikungunya Arthritogenic Alphavirus in Spite of a Robust Host Immune Response. *The Journal of Immunology*, 184(10):5914–5927, May 2010.

[35] Karine Labadie, Thibaut Larcher, Christophe Joubert, Abdelkrim Mannioui, Benoit Delache, Patricia Brochard, Lydie Guigand, Laurence Dubreil, Pierre Lebon, Bernard Verrier, Xavier de Lamballerie, Andreas Suhrbier, Yan Cherel, Roger Le Grand, and Pierre Roques. Chikungunya disease in nonhuman primates involves long-term viral persistence in macrophages. *The Journal of Clinical Investigation*, 120(3):894, February 2010.

[36] Laurence Dupuis-Maguiraga, Marion Noret, Sonia Brun, Roger Le Grand, Gabriel Gras, and Pierre Roques. Chikungunya Disease: Infection-Associated Markers from the Acute to the Chronic Phase of Arbovirus-Induced Arthralgia. *PLoS Neglected Tropical Diseases*, 6(3):e1446, March 2012.

[37] Robert B. Tesh, Douglas M. Watts, Kevin L. Russell, Chitra Damodaran, Carlos Calampa, Cesar Cabezas, Gladys Ramirez, Bruno Vasquez, Curtis G. Hayes, Cynthia A. Rossi, Ann M. Powers, Christine L. Hice, Laura J. Chandler, Bruce C. Cropp, Nick Karabatsos, John T. Roehrig, and Duane J. Gubler. Mayaro Virus Disease: An Emerging Mosquito-Borne Zoonosis in Tropical South America. *Clinical Infectious Diseases*, 28(1):67–73, January 1999.

[38] Kate D. Ryman, William B. Klimstra, Khuong B. Nguyen, Christine A. Biron, and Robert E. Johnston. Alpha/Beta Interferon Protects Adult Mice from Fatal Sindbis

Virus Infection and Is an Important Determinant of Cell and Tissue Tropism. *Journal of Virology*, 74(7):3366, April 2000.

[39] Robert L. Seymour, Shannan L. Rossi, Nicholas A. Bergren, Kenneth S. Plante, and Scott C. Weaver. The Role of Innate versus Adaptive Immune Responses in a Mouse Model of O'Nyong-Nyong Virus Infection. *The American Journal of Tropical Medicine and Hygiene*, 88(6):1170, June 2013.

[40] William M. Schneider, Meike Dittmann Chevillotte, and Charles M. Rice. Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annual review of immunology*, 32:513–545, 2014.

[41] Alyson A. Kelvin, David Banner, Giuliano Silvi, Maria Luisa Moro, Nadir Spataro, Paolo Gaibani, Francesca Cavrini, Anna Pierro, Giada Rossini, Mark J. Cameron, Jesus F. Bermejo-Martin, Stéphane G. Paquette, Luoling Xu, Ali Danesh, Amber Farooqui, Ilaria Borghetto, David J. Kelvin, Vittorio Sambri, and Salvatore Rubino. Inflammatory Cytokine Expression Is Associated with Chikungunya Virus Resolution and Symptom Severity. *PLoS Neglected Tropical Diseases*, 5(8):e1279, August 2011.

[42] Nadia Wauquier, Pierre Becquart, Dieudonné Nkoghe, Cindy Padilla, Angélique Ndjoyi-Mbiguino, and Eric M. Leroy. The Acute Phase of Chikungunya Virus Infection in Humans Is Associated With Strong Innate Immunity and T CD8 Cell Activation. *The Journal of Infectious Diseases*, 204(1):115, December 2010.

[43] Andrew M. Skidmore and Steven B. Bradfute. The life cycle of the alphaviruses: From an antiviral perspective. *Antiviral Research*, 209:105476, January 2023.

[44] Erika J. Mancini, Mairi Clarke, Brent E. Gowen, Twan Rutten, and Stephen D. Fuller. Cryo-Electron Microscopy Reveals the Functional Organization of an Enveloped Virus, Semliki Forest Virus. *Molecular Cell*, 5(2):255–266, February 2000. Publisher: Elsevier.

[45] J. H. Strauss and E. G. Strauss. The alphaviruses: gene expression, replication, and evolution. *Microbiological Reviews*, 58(3):491, September 1994.

[46] S Schmid, R Fuchs, M Kielian, A Helenius, and I Mellman. Acidification of endosome subpopulations in wild-type Chinese hamster ovary cells and temperature-sensitive acidification-defective mutants. *Journal of Cell Biology*, 108(4):1291–1300, April 1989.

[47] Gerd Wengler, Christof Gros, and Gisela Wengler. Analyses of the Role of Structural Changes in the Regulation of Uncoating and Assembly of Alphavirus Cores. *Virology*, 222(1):123–132, August 1996.

[48] G P Li and C M Rice. Mutagenesis of the in-frame opal termination codon preceding nsP4 of Sindbis virus: studies of translational readthrough and its effect on virus replication. *Journal of Virology*, 63(3):1326–1337, March 1989. Publisher: American Society for Microbiology.

[49] E G Strauss, C M Rice, and J H Strauss. Sequence coding for the alphavirus nonstructural proteins is interrupted by an opal termination codon. *Proceedings of the National Academy of Sciences*, 80(17):5271–5275, September 1983. Publisher: Proceedings of the National Academy of Sciences.

[50] Lidia Vasiljeva, Leena Valmu, Leevi Kääriäinen, and Andres Merits. Site-specific Protease Activity of the Carboxyl-terminal Domain of Semliki Forest Virus Replicase Protein nsP2 *. *Journal of Biological Chemistry*, 276(33):30786–30793, August 2001. Publisher: Elsevier.

[51] Y Shirako and J H Strauss. Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis, whereas cleaved products from P123 are required for efficient plus-strand RNA synthesis. *Journal of Virology*, 68(3):1874–1885, March 1994. Publisher: American Society for Microbiology.

[52] J A Lemm and C M Rice. Assembly of functional Sindbis virus RNA replication complexes: requirement for coexpression of P123 and P34. *Journal of Virology*, 67(4):1905–1915, April 1993. Publisher: American Society for Microbiology.

[53] J.A. Lemm, T. Rümenapf, E.G. Strauss, J.H. Strauss, and C.M. Rice. Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. *The EMBO Journal*, 13(12):2925–2934, June 1994. Num Pages: 2934 Publisher: John Wiley & Sons, Ltd.

[54] Adriano Mendes and Richard J. Kuhn. Alphavirus Nucleocapsid Packaging and Assembly. *Viruses*, 10(3):138, March 2018. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[55] Xinyong Zhang, Martin Fugère, Robert Day, and Margaret Kielian. Furin Processing and Proteolytic Activation of Semliki Forest Virus. *Journal of Virology*, 77(5):2981–2989, March 2003. Publisher: American Society for Microbiology.

[56] Thomas A. Wilkinson, Timothy L. Tellinghuisen, Richard J. Kuhn, and Carol Beth Post. Association of Sindbis Virus Capsid Protein with Phospholipid Membranes and the E2 Glycoprotein: Implications for Alphavirus Assembly. *Biochemistry*, 44(8):2800–2810, March 2005. Publisher: American Chemical Society.

[57] Hans R. Gelderblom. Structure and Classification of Viruses. In Samuel Baron, editor, *Medical Microbiology*. University of Texas Medical Branch at Galveston, Galveston (TX), 4th edition, 1996.

[58] Norman E. Davey, Gilles Travé, and Toby J. Gibson. How viruses hijack cell regulation. *Trends in Biochemical Sciences*, 36(3):159–169, March 2011.

[59] Shivani K. Thaker, James Ch'ng, and Heather R. Christofk. Viral hijacking of cellular metabolism. *BMC Biology*, 17(1):59, July 2019.

[60] Wenxian Wu, Xiumei Luo, and Maozhi Ren. Clearance or Hijack: Universal Interplay Mechanisms Between Viruses and Host Autophagy From Plants to Animals. *Frontiers in Cellular and Infection Microbiology*, 11, January 2022. Publisher: Frontiers.

[61] Alfredo Castello, Lucía Álvarez, Wael Kamel, Louisa Iselin, and Janosch Hennig. Exploring the expanding universe of host-virus interactions mediated by viral RNA. *Molecular Cell*, 84(19):3706–3721, October 2024. Publisher: Elsevier.

[62] Stefanie Gerstberger, Markus Hafner, and Thomas Tuschl. A census of human RNA-binding proteins. *Nature Reviews Genetics*, 15(12):829–845, December 2014. Publisher: Nature Publishing Group.

[63] Guramrit Singh, Gabriel Pratt, Gene W. Yeo, and Melissa J. Moore. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annual review of biochemistry*, 84:325, March 2015.

[64] Sanghee Kim, Jun-Yi Yang, Jun Xu, In-Cheol Jang, Michael J. Prigge, and Nam-Hai Chua. Two Cap-Binding Proteins CBP20 and CBP80 are Involved in Processing Primary MicroRNAs. *Plant and Cell Physiology*, 49(11):1634–1644, November 2008.

[65] Joshua T. Witten and Jernej Ule. Understanding splicing regulation through RNA splicing maps. *Trends in Genetics*, 27(3):89–97, March 2011.

[66] Chaolin Zhang, Zuo Zhang, John Castle, Shuying Sun, Jason Johnson, Adrian R. Krainer, and Michael Q. Zhang. Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes & Development*, 22(18):2550–2563, September 2008. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[67] Donny D. Licatalosi, Aldo Mele, John J. Fak, Jernej Ule, Melis Kayikci, Sung Wook Chi, Tyson A. Clark, Anthony C. Schweitzer, John E. Blume, Xuning Wang, Jennifer C. Darnell, and Robert B. Darnell. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, 456(7221):464–469, November 2008. Publisher: Nature Publishing Group.

[68] Jernej Ule, Giovanni Stefani, Aldo Mele, Matteo Ruggiu, Xuning Wang, Bahar Taneri, Terry Gaasterland, Benjamin J. Blencowe, and Robert B. Darnell. An RNA map predicting Nova-dependent splicing regulation. *Nature*, 444(7119):580–586, November 2006. Publisher: Nature Publishing Group.

[69] Jacopo Meldolesi. Alternative Splicing by NOVA Factors: From Gene Expression to Cell Physiology and Pathology. *International Journal of Molecular Sciences*, 21(11):3941, May 2020.

[70] Anne-Claude Gingras, Brian Raught, and Nahum Sonenberg. eIF4 Initiation Factors: Effectors of mRNA Recruitment to Ribosomes and Regulators of Translation. *Annual Review of Biochemistry*, 68(Volume 68, 1999):913–963, July 1999. Publisher: Annual Reviews.

[71] Jailson Brito Querido, Irene Díaz-López, and V. Ramakrishnan. The molecular basis of translation initiation and its regulation in eukaryotes. *Nature Reviews Molecular Cell Biology*, 25(3):168–186, March 2024. Publisher: Nature Publishing Group.

[72] Scot R Kimball. Eukaryotic initiation factor eIF2. *The International Journal of Biochemistry & Cell Biology*, 31(1):25–29, January 1999.

[73] Anton A. Komar, Barsanjit Mazumder, and William C. Merrick. A New Framework for Understanding IRES-mediated Translation. *Gene*, 502(2):75, April 2012.

[74] Richard J. Jackson, Christopher U. T. Hellen, and Tatyana V. Pestova. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*, 11(2):113–127, February 2010. Publisher: Nature Publishing Group.

[75] Roy Parker. RNA Degradation in Saccharomyces cerevisae. *Genetics*, 191(3):671–702, July 2012.

[76] Stepanka Vanacova and Richard Stef. The exosome and RNA quality control in the nucleus. *EMBO Reports*, 8(7):651, July 2007.

[77] Yoon Ki Kim and Lynne E. Maquat. UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond. *RNA*, 25(4):407–422, April 2019. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[78] Rastislav Horos, Magdalena Büscher, Rozemarijn Kleinendorst, Anne-Marie Alleaume, Abul K. Tarafder, Thomas Schwarzl, Dmytro Dziuba, Christian Tischer, Elisabeth M. Zielonka, Asli Adak, Alfredo Castello, Wolfgang Huber, Carsten Sachse, and Matthias W. Hentze. The Small Non-coding Vault RNA1-1 Acts as a Riboregulator of Autophagy. *Cell*, 176(5):1054–1067.e12, February 2019. Publisher: Elsevier.

[79] Sharon Spizzichino, Federica Di Fonzo, Chiara Marabelli, Angela Tramonti, Antonio Chaves-Sanjuan, Alessia Parroni, Giovanna Boumis, Francesca Romana Liberati, Alessio Paone, Linda Celeste Montemiglio, Matteo Ardini, Arjen J. Jakobi, Alok Bharadwaj, Paolo Swuec, Gian Gaetano Tartaglia, Alessandro Paiardini, Roberto Contestabile, Antonello Mai, Dante Rotili, Francesco Fiorentino, Alberto Macone, Alessandra Giorgi, Giancarlo Tria, Serena Rinaldo, Martino Bolognesi, Giorgio Giardina, and Francesca Cutruzzolà. Structure-based mechanism of riboregulation of the metabolic enzyme SHMT1. *Molecular Cell*, 84(14):2682–2697.e6, July 2024. Publisher: Elsevier.

[80] Tetsuro Hirose, Giorgio Virnicchi, Akie Tanigawa, Takao Naganuma, Ruohan Li, Hiroshi Kimura, Takahide Yokoi, Shinichi Nakagawa, Marianne Bénard, Archa H. Fox, and

Gérard Pierron. NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Molecular Biology of the Cell*, 25(1):169, January 2014.

[81] Bradley M. Lunde, Claire Moore, and Gabriele Varani. RNA-binding proteins: modular design for efficient function. *Nature reviews. Molecular cell biology*, 8(6):479, June 2007.

[82] Patrick Linder and Eckhard Jankowsky. From unwinding to clamping — the DEAD box RNA helicase family. *Nature Reviews Molecular Cell Biology*, 12(8):505–516, August 2011. Publisher: Nature Publishing Group.

[83] Antoine Cléry, Markus Blatter, and Frédéric H-T Allain. RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology*, 18(3):290–298, June 2008.

[84] Roberto Valverde, Laura Edwards, and Lynne Regan. Structure and function of KH domains. *The FEBS Journal*, 275(11):2712–2726, 2008. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1742-4658.2008.06411.x.

[85] Matthias W. Hentze, Alfredo Castello, Thomas Schwarzl, and Thomas Preiss. A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology*, 19(5):327–341, May 2018. Publisher: Nature Publishing Group.

[86] Falk Butter, Marion Scheibe, Mario Mörl, and Matthias Mann. Unbiased RNA–protein interaction screen by quantitative proteomics. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26):10626, June 2009.

[87] Tanja Scherrer, Nitish Mittal, Sarath Chandra Janga, and André P. Gerber. A Screen for RNA-Binding Proteins in Yeast Indicates Dual Functions for Many Enzymes. *PLoS ONE*, 5(11):e15499, November 2010.

[88] Alfredo Castello, Bernd Fischer, Katrin Eichelbaum, Rastislav Horos, Benedikt M. Beckmann, Claudia Strein, Norman E. Davey, David T. Humphreys, Thomas Preiss, Lars M. Steinmetz, Jeroen Krijgsveld, and Matthias W. Hentze. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell*, 149(6):1393–1406, June 2012. Publisher: Elsevier.

[89] Alexander G. Baltz, Mathias Munschauer, Björn Schwanhäusser, Alexandra Vasile, Yasuhiro Murakawa, Markus Schueler, Noah Youngs, Duncan Penfold-Brown, Kevin

Drew, Miha Milek, Emanuel Wyler, Richard Bonneau, Matthias Selbach, Christoph Dieterich, and Markus Landthaler. The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Molecular Cell*, 46(5):674–690, June 2012. Publisher: Elsevier.

[90] Rongbing Huang, Mengting Han, Liying Meng, and Xing Chen. Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. *Proceedings of the National Academy of Sciences*, 115(17):E3879–E3887, April 2018. Publisher: Proceedings of the National Academy of Sciences.

[91] Xichen Bao, Xiangpeng Guo, Menghui Yin, Muqddas Tariq, Yiwei Lai, Shahzina Kanwal, Jiajian Zhou, Na Li, Yuan Lv, Carlos Pulido-Quetglas, Xiwei Wang, Lu Ji, Muhammad J. Khan, Xihua Zhu, Zhiwei Luo, Changwei Shao, Do-Hwan Lim, Xiao Liu, Nan Li, Wei Wang, Minghui He, Yu-Lin Liu, Carl Ward, Tong Wang, Gong Zhang, Dongye Wang, Jianhua Yang, Yiwen Chen, Chaolin Zhang, Ralf Jauch, Yun-Gui Yang, Yangming Wang, Baoming Qin, Minna-Liisa Anko, Andrew P. Hutchins, Hao Sun, Huating Wang, Xiang-Dong Fu, Biliang Zhang, and Miguel A. Esteban. Capturing the interactome of newly transcribed RNA. *Nature Methods*, 15(3):213–220, March 2018. Publisher: Nature Publishing Group.

[92] Claudio Asencio, Aindrila Chatterjee, and Matthias W. Hentze. Silica-based solid-phase extraction of cross-linked nucleic acid–bound proteins. *Life Science Alliance*, 1(3), June 2018. Publisher: Life Science Alliance Section: Methods.

[93] Rayner M. L. Queiroz, Tom Smith, Eneko Villanueva, Maria Marti-Solano, Mie Monti, Mariavittoria Pizzinga, Dan-Mircea Mirea, Manasa Ramakrishna, Robert F. Harvey, Veronica Dezi, Gavin H. Thomas, Anne E. Willis, and Kathryn S. Lilley. Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nature Biotechnology*, 37(2):169–178, February 2019. Publisher: Nature Publishing Group.

[94] Claudia Strein, Anne-Marie Alleaume, Ulrich Rothbauer, Matthias W. Hentze, and Alfredo Castello. A versatile assay for RNA-binding proteins in living cells. *RNA*, 20(5):721, May 2014.

[95] Alfredo Castello, Bernd Fischer, Christian K. Frese, Rastislav Horos, Anne-Marie Alleaume, Sophia Foehr, Tomaz Curk, Jeroen Krijgsveld, and Matthias W. Hentze. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell*, 63(4):696–710, August 2016.

[96] Jong Woo Bae, S. Chul Kwon, Yongwoo Na, V. Narry Kim, and Jong-Seo Kim. Chemical RNA digestion enables robust RNA-binding site mapping at single amino acid resolution. *Nature Structural & Molecular Biology*, 27(7):678–682, July 2020. Publisher: Nature Publishing Group.

[97] James L. Manley. SELEX to Identify Protein-Binding Sites on RNA. *Cold Spring Harbor protocols*, 2013(2):156, February 2013.

[98] Debashish Ray, Kevin CH Ha, Kate Nie, Hong Zheng, Timothy R. Hughes, and Quaid D. Morris. RNAcompete methodology and application to determine sequence preferences of unconventional RNA-binding proteins. *Methods (San Diego, Calif.)*, 118-119:3, December 2016.

[99] Ivo Fierro-Monti. RBPs: an RNA editor's choice. *Frontiers in Molecular Biosciences*, 11, August 2024. Publisher: Frontiers.

[100] Jernej Ule, Kirk Jensen, Aldo Mele, and Robert B. Darnell. CLIP: A method for identifying protein–RNA interaction sites in living cells. *Methods*, 37(4):376–386, December 2005.

[101] Markus Hafner, Maria Katsantoni, Tino Köster, James Marks, Joyita Mukherjee, Dorothee Staiger, Jernej Ule, and Mihaela Zavolan. CLIP and complementary methods. *Nature Reviews Methods Primers*, 1(1):1–23, March 2021. Number: 1 Publisher: Nature Publishing Group.

[102] Emily C. Wheeler, Eric L. Van Nostrand, and Gene W. Yeo. Advances and challenges in the detection of transcriptome-wide protein–RNA interactions. *Wiley Interdisciplinary Reviews. RNA*, 9(1):e1436, August 2017.

[103] Eric L. Van Nostrand, Peter Freese, Gabriel A. Pratt, Xiaofeng Wang, Xintao Wei, Rui Xiao, Steven M. Blue, Jia-Yu Chen, Neal A. L. Cody, Daniel Dominguez, Sara Olson, Balaji Sundararaman, Lijun Zhan, Cassandra Bazile, Louis Philip Benoit Bouvrette, Julie Bergalet, Michael O. Duff, Keri E. Garcia, Chelsea Gelboin-Burkhart, Myles Hochman,

Nicole J. Lambert, Hairi Li, Michael P. McGurk, Thai B. Nguyen, Tsultrim Palden, Ines Rabano, Shashank Sathe, Rebecca Stanton, Amanda Su, Ruth Wang, Brian A. Yee, Bing Zhou, Ashley L. Louie, Stefan Aigner, Xiang-Dong Fu, Eric Lécuyer, Christopher B. Burge, Brenton R. Graveley, and Gene W. Yeo. A large-scale binding and functional map of human RNA-binding proteins. *Nature*, 583(7818):711–719, July 2020. Publisher: Nature Publishing Group.

[104] Kaitlin U Laverty, Arttu Jolma, Sara E Pour, Hong Zheng, Debashish Ray, Quaid Morris, and Timothy R Hughes. PRIESSTESS: interpretable, high-performing models of the sequence and structure preferences of RNA-binding proteins. *Nucleic Acids Research*, 50(19):e111, October 2022.

[105] Yun Zuo, Huixian Chen, Lele Yang, Ruoyan Chen, Xiaoyao Zhang, and Zhaohong Deng. Research progress on prediction of RNA-protein binding sites in the past five years. *Analytical Biochemistry*, 691:115535, August 2024.

[106] Lei Li, Zhigang Xue, and Xiuquan Du. ASCRB: Multi-view based attentional feature selection for CircRNA-binding site prediction. *Computers in Biology and Medicine*, 162:107077, August 2023.

[107] Vasiliy O. Sysoev, Bernd Fischer, Christian K. Frese, Ishaan Gupta, Jeroen Krijgsveld, Matthias W. Hentze, Alfredo Castello, and Anne Ephrussi. Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. *Nature Communications*, 7:12128, July 2016.

[108] Vladimir Despic, Mario Dejung, Mengting Gu, Jayanth Krishnan, Jing Zhang, Lydia Herzel, Korinna Straube, Mark B. Gerstein, Falk Butter, and Karla M. Neugebauer. Dynamic RNA–protein interactions underlie the zebrafish maternal-to-zygotic transition. *Genome Research*, 27(7):1184, July 2017.

[109] Zhenghe Li and Peter D. Nagy. Diverse roles of host RNA binding proteins in RNA virus replication. *RNA Biology*, 8(2):305–315, March 2011. Publisher: Taylor & Francis _eprint: https://doi.org/10.4161/rna.8.2.15391.

[110] Peter D. Nagy and Judit Pogany. The dependence of viral RNA replication on co-opted host factors. *Nature Reviews Microbiology*, 10(2):137–149, February 2012. Publisher: Nature Publishing Group.

[111] Roman Barbalat, Sarah E. Ewald, Maria L. Mouchess, and Gregory M. Barton. Nucleic Acid Recognition by the Innate Immune System. *Annual Review of Immunology*, 29(Volume 29, 2011):185–214, April 2011. Publisher: Annual Reviews.

[112] Gregory I. Vladimer, Maria W. Górna, and Giulio Superti-Furga. IFITs: Emerging Roles as Key Anti-Viral Proteins. *Frontiers in Immunology*, 5, March 2014. Publisher: Frontiers.

[113] Manuel Garcia-Moreno, Marko Noerenberg, Shuai Ni, Aino I. Järvelin, Esther González-Almela, Caroline E. Lenz, Marcel Bach-Pages, Victoria Cox, Rosario Avolio, Thomas Davis, Svenja Hester, Thibault J.M. Sohier, Bingnan Li, Gregory Heikel, Gracjan Michlewski, Miguel A. Sanz, Luis Carrasco, Emiliano P. Ricci, Vicent Pelechano, Ilan Davis, Bernd Fischer, Shabaz Mohammed, and Alfredo Castello. System-wide Profiling of RNA-Binding Proteins Uncovers Key Regulators of Virus Infection. *Molecular Cell*, 74(1):196–211.e11, April 2019.

[114] Byungil Kim, Sarah Arcos, Katherine Rothamel, Jeffrey Jian, Kristie L. Rose, W. Hayes McDonald, Yuqi Bian, Seth Reasoner, Nicholas J. Barrows, Shelton Bradrick, Mariano A. Garcia-Blanco, and Manuel Ascano. Discovery of Widespread Host Protein Interactions with the Pre-replicated Genome of CHIKV Using VIR-CLASP. *Molecular Cell*, 78(4):624–640.e7, May 2020.

[115] L. Iselin, N. Palmalux, W. Kamel, P. Simmonds, S. Mohammed, and A. Castello. Uncovering viral RNA?host cell interactions on a proteome-wide scale. *Trends in Biochemical Sciences*, 47(1):23–38, 2022.

[116] W. Kamel, M. Noerenberg, B. Cerikan, H. Chen, A.I. Järvelin, M. Kammoun, J.Y. Lee, N. Shuai, M. Garcia-Moreno, A. Andrejeva, M.J. Deery, N. Johnson, C.J. Neufeldt, M. Cortese, M.L. Knight, K.S. Lilley, J. Martinez, I. Davis, R. Bartenschlager, S. Mohammed, and A. Castello. Global analysis of protein-RNA interactions in SARS-CoV-2-infected cells reveals key regulators of infection. *Molecular Cell*, 81(13):2851–2867.e7, 2021.

[117] Stefano Bonazza, Hannah Leigh Coutts, Swathi Sukumar, Hannah Louise Turkington, and David Gary Courtney. Identifying cellular RNA-binding proteins during infection uncovers a role for MKRN2 in influenza mRNA trafficking. *PLOS Pathogens*, 20(5):e1012231, May 2024. Publisher: Public Library of Science.

[118] Xuefeng Wang, Yudong Zhao, Feihu Yan, Tiecheng Wang, Weiyang Sun, Na Feng, Wenqi Wang, Hongmei Wang, Hongbin He, Songtao Yang, Xianzhu Xia, and Yuwei Gao. Viral and Host Transcriptomes in SARS-CoV-2-Infected Human Lung Cells. *Journal of Virology*, 95(18):e00600, August 2021.

[119] Usama Ashraf, Clara Benoit-Pilven, Vincent Navratil, Cécile Ligneau, Guillaume Fournier, Sandie Munier, Odile Sismeiro, Jean-Yves Coppée, Vincent Lacroix, and Nadia Naffakh. Influenza virus infection induces widespread alterations of host cell splicing. *NAR Genomics and Bioinformatics*, 2(4):lqaa095, December 2020.

[120] Lars Dölken, Zsolt Ruzsics, Bernd Rädle, Caroline C. Friedel, Ralf Zimmer, Jörg Mages, Reinhard Hoffmann, Paul Dickinson, Thorsten Forster, Peter Ghazal, and Ulrich H. Koszinowski. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*, 14(9):1959, September 2008.

[121] Autumn T. LaPointe, Natasha N. Gebhart, Megan E. Meller, Richard W. Hardy, and Kevin J. Sokoloski. Identification and Characterization of Sindbis Virus RNA-Host Protein Interactions. *Journal of Virology*, 92(7):10.1128/jvi.02171–17, March 2018. Publisher: American Society for Microbiology.

[122] Natasha N. Gebhart, Richard W. Hardy, and Kevin J. Sokoloski. Comparative analyses of alphaviral RNA:Protein complexes reveals conserved host-pathogen interactions. *PLoS ONE*, 15(8):e0238254, August 2020.

[123] Wael Kamel, Vincenzo Ruscica, Azman Embarc-Buh, Zaydah R. de Laurent, Manuel Garcia-Moreno, Yana Demyanenko, Richard J. Orton, Marko Noerenberg, Meghana Madhusudhan, Louisa Iselin, Aino I. Järvelin, Maximilian Hannan, Eduardo Kitano, Samantha Moore, Andres Merits, Ilan Davis, Shabaz Mohammed, and Alfredo Castello. Alphavirus infection triggers selective cytoplasmic translocation of nuclear RBPs with moonlighting antiviral roles. *Molecular Cell*, December 2024.

[124] Alfredo Castelló, Miguel Ángel Sanz, Susana Molina, and Luis Carrasco. Translation of Sindbis Virus 26S mRNA Does Not Require Intact Eukariotic Initiation Factor 4G. *Journal of Molecular Biology*, 355(5):942–956, February 2006.

[125] Manuel Garcia-Moreno, Miguel Sanz, and Luis Carrasco. Initiation codon selection is accomplished by a scanning mechanism without crucial initiation factors in Sindbis virus subgenomic mRNA. *RNA*, November 2014.

[126] Iván Ventoso, Miguel Angel Sanz, Susana Molina, Juan José Berlanga, Luis Carrasco, and Mariano Esteban. Translational resistance of late alphavirus mRNA to eIF2α phosphorylation: a strategy to overcome the antiviral effect of protein kinase PKR. *Genes & Development*, 20(1):87–100, January 2006.

[127] Elena Frolova, Rodion Gorchakov, Natalia Garmashova, Svetlana Atasheva, Leoncio A. Vergara, and Ilya Frolov. Formation of nsP3-Specific Protein Complexes during Sindbis Virus Replication. *Journal of Virology*, 80(8):4122–4134, April 2006.

[128] Ileana M. Cristea, John-William N. Carroll, Michael P. Rout, Charles M. Rice, Brian T. Chait, and Margaret R. MacDonald. Tracking and Elucidating Alphavirus-Host Protein Interactions *. *Journal of Biological Chemistry*, 281(40):30269–30278, October 2006. Publisher: Elsevier.

[129] Svetlana Atasheva, Rodion Gorchakov, Robert English, Ilya Frolov, and Elena Frolova. Development of Sindbis Viruses Encoding nsP2/GFP Chimeric Proteins and Their Application for Studying nsP2 Functioning. *Journal of Virology*, 81(10):5046–5057, May 2007. Publisher: American Society for Microbiology.

[130] Ileana M. Cristea, Heather Rozjabek, Kelly R. Molloy, Sophiya Karki, Laura L. White, Charles M. Rice, Michael P. Rout, Brian T. Chait, and Margaret R. MacDonald. Host Factors Associated with the Sindbis Virus RNA-Dependent RNA Polymerase: Role for G3BP1 and G3BP2 in Virus Replication. *Journal of Virology*, 84(13):6720–6732, July 2010. Publisher: American Society for Microbiology.

[131] Margus Varjak, Sirle Saul, Liisa Arike, Aleksei Lulla, Lauri Peil, and Andres Merits. Magnetic Fractionation and Proteomic Dissection of Cellular Organelles Occupied by the Late Replication Complexes of Semliki Forest Virus. *Journal of Virology*, 87(18):10295–10312, September 2013. Publisher: American Society for Microbiology.

[132] William B. Klimstra, Elizabeth M. Nangle, M. Shane Smith, Andrew D. Yurochko, and Kate D. Ryman. DC-SIGN and L-SIGN Can Act as Attachment Receptors for

Alphaviruses and Distinguish between Mosquito Cell- and Mammalian Cell-Derived Viruses. *Journal of Virology*, 77(22):12022, November 2003.

[133] J. Steven Bear, Andrew P. Byrnes, and Diane E. Griffin. Heparin-binding and patterns of virulence for two recombinant strains of Sindbis virus. *Virology*, 347(1):183–190, March 2006.

[134] Merja Sariola, Jaakko Saraste, and Esa Kuismanen. Communication of post-golgi elements with early endocytic pathway: Regulation of endoproteolytic cleavage of semliki forest virus p62 precursor. *Journal of Cell Science*, 108(6):2465–2475, June 1995.

[135] Julian V. Melton, Gary D. Ewart, Ronald C. Weir, Philip G. Board, Eva Lee, and Peter W. Gage. Alphavirus 6K Proteins Form Ion Channels*. *Journal of Biological Chemistry*, 277(49):46923–46931, December 2002.

[136] S. Atasheva, A. Fish, M. Fornerod, and E.I. Frolova. Venezuelan equine encephalitis virus capsid protein forms a tetrameric complex with CRM1 and importin α/β that obstructs nuclear pore complex function. *Journal of Virology*, 84(9):4158–4171, 2010.

[137] Jelke J. Fros and Gorben P. Pijlman. Alphavirus Infection: Host Cell Shut-Off and Inhibition of Antiviral Responses. *Viruses*, 8(6):166, June 2016.

[138] Nishank Bhalla, Chengqun Sun, L. K. Metthew Lam, Christina L. Gardner, Kate D. Ryman, and William B. Klimstra. Host Translation Shutoff Mediated by Non-structural Protein 2 is a Critical Factor in the Antiviral State Resistance of Venezuelan Equine Encephalitis Virus. *Virology*, 496:147–165, September 2016.

[139] Ivan Akhrymuk, Sergey V. Kulemzin, and Elena I. Frolova. Evasion of the Innate Immune Response: the Old World Alphavirus nsP2 Protein Induces Rapid Degradation of Rpb1, a Catalytic Subunit of RNA Polymerase II. *Journal of Virology*, 86(13):7180, July 2012.

[140] Ivan Akhrymuk, Tetyana Lukash, Ilya Frolov, and Elena I. Frolova. Novel Mutations in nsP2 Abolish Chikungunya Virus-Induced Transcriptional Shutoff and Make the Virus Less Cytopathic without Affecting Its Replication Rates. *Journal of Virology*, 93(4):e02062–18, February 2019.

[141] Gerald M. McInerney, Nancy L. Kedersha, Randal J. Kaufman, Paul Anderson, and Peter Liljeström. Importance of eIF2α Phosphorylation and Stress Granule Assembly in Alphavirus Translation Regulation. *Molecular Biology of the Cell*, 16(8):3753–3763, August 2005.

[142] Rohini K. Patel, Andy J. Burnham, Natasha N. Gebhart, Kevin J. Sokoloski, and Richard W. Hardy. Role for Subgenomic mRNA in Host Translation Inhibition during Sindbis Virus Infection of Mammalian Cells. *Virology*, 441(2):171–181, July 2013.

[143] Luis Carrasco, Miguel Angel Sanz, and Esther González-Almela. The Regulation of Translation in Alphavirus-Infected Cells. *Viruses*, 10(2):70, February 2018.

[144] Robert F. Garry. Sindbis virus-induced inhibition of protein synthesis is partially reversed by medium containing an elevated potassium concentration. *Journal of General Virology*, 75(2):411–415, 1994. Publisher: Microbiology Society,.

[145] Heather P. Harding, Isabel Novoa, Yuhong Zhang, Huiqing Zeng, Ron Wek, Matthieu Schapira, and David Ron. Regulated Translation Initiation Controls Stress-Induced Gene Expression in Mammalian Cells. *Molecular Cell*, 6(5):1099–1108, November 2000. Publisher: Elsevier.

[146] René Toribio and Iván Ventoso. Inhibition of host translation by virus infection in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 107(21):9837, May 2010.

[147] M. Bushell, D. Poncet, W. E. Marissen, H. Flotow, R. E. Lloyd, M. J. Clemens, and S. J. Morley. Cleavage of polypeptide chain initiation factor eIF4GI during apoptosis in lymphoma cells: characterisation of an internal fragment generated by caspase-3-mediated cleavage. *Cell Death & Differentiation*, 7(7):628–636, July 2000. Publisher: Nature Publishing Group.

[148] Walter Glaser and Tim Skern. Extremely efficient cleavage of eIF4G by picornaviral proteinases L and 2A in vitro. *FEBS Letters*, 480(2-3):151–155, 2000. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1016/S0014-5793%2800%2901928-1.

[149] Jeongyeon Nam and Youngdae Gwon. Neuronal biomolecular condensates and their implications in neurodegenerative diseases. *Frontiers in Aging Neuroscience*, 15:1145420, March 2023.

[150] Salman F. Banani, Hyun O. Lee, Anthony A. Hyman, and Michael K. Rosen. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*, 18(5):285–298, May 2017. Publisher: Nature Publishing Group.

[151] Andrew S. Lyon, William B. Peeples, and Michael K. Rosen. A framework for understanding the functions of biomolecular condensates across scales. *Nature Reviews Molecular Cell Biology*, 22(3):215–235, March 2021. Publisher: Nature Publishing Group.

[152] Pavel Ivanov, Nancy Kedersha, and Paul Anderson. Stress Granules and Processing Bodies in Translational Control. *Cold Spring Harbor Perspectives in Biology*, 11(5):a032813, May 2019. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[153] Alessia Ruggieri, Eva Dazert, Philippe Metz, Sarah Hofmann, Jan-Philip Bergeest, Johanna Mazur, Peter Bankhead, Marie-Sophie Hiet, Stephanie Kallis, Gualtiero Alvisi, Charles E. Samuel, Volker Lohmann, Lars Kaderali, Karl Rohr, Michael Frese, Georg Stoecklin, and Ralf Bartenschlager. Dynamic Oscillation of Translation and Stress Granule Formation Mark the Cellular Response to Virus Infection. *Cell host & microbe*, 12(1):10.1016/j.chom.2012.05.013, July 2012.

[154] Aravinth Kumar Jayabalan, Srivathsan Adivarahan, Aakash Koppula, Rachy Abraham, Mona Batish, Daniel Zenklusen, Diane E. Griffin, and Anthony K. L. Leung. Stress granule formation, disassembly, and composition are regulated by alphavirus ADP-ribosylhydrolase activity. *Proceedings of the National Academy of Sciences*, 118(6):e2021719118, February 2021. Publisher: Proceedings of the National Academy of Sciences.

[155] Marc D. Panas, Tero Ahola, and Gerald M. McInerney. The C-Terminal Repeat Domains of nsP3 from the Old World Alphaviruses Bind Directly to G3BP. *Journal of Virology*, 88(10):5888, May 2014.

[156] Florine E. M. Scholte, Ali Tas, Irina C. Albulescu, Eva Žusinaite, Andres Merits, Eric J. Snijder, and Martijn J. van Hemert. Stress Granule Components G3BP1 and G3BP2

Play a Proviral Role Early in Chikungunya Virus Replication. *Journal of Virology*, 89(8):4457, February 2015.

[157] Benjamin Götte, Age Utt, Rennos Fragkoudis, Andres Merits, and Gerald M. McInerney. Sensitivity of Alphaviruses to G3BP Deletion Correlates with Efficiency of Replicase Polyprotein Processing. *Journal of Virology*, 94(7):e01681–19, March 2020. Publisher: American Society for Microbiology.

[158] Iván Ventoso, Juan José Berlanga, René Toribio, and Irene Díaz-López. Translational Control of Alphavirus–Host Interactions: Implications in Viral Evolution, Tropism and Antiviral Response. *Viruses*, 16(2):205, January 2024.

[159] Jennifer L. Hyde, Christina L. Gardner, Taishi Kimura, James P. White, Gai Liu, Derek W. Trobaugh, Cheng Huang, Marco Tonelli, Slobodan Paessler, Kiyoshi Takeda, William B. Klimstra, Gaya K. Amarasinghe, and Michael S. Diamond. A viral RNA structural element alters host recognition of non-self RNA. *Science (New York, N.Y.)*, 343(6172):783, January 2014.

[160] Jennifer L. Meagher, Matthew Takata, Daniel Gonçalves-Carneiro, Sarah C. Keane, Antoine Rebendenne, Heley Ong, Victoria K. Orr, Margaret R. MacDonald, Jeanne A. Stuckey, Paul D. Bieniasz, and Janet L. Smith. Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 116(48):24303, November 2019.

[161] Anne-Claire Bréhin, Isabelle Casadémont, Marie-Pascale Frenkiel, Cécile Julier, Anavaj Sakuntabhai, and Philippe Desprès. The large form of human 2',5'-Oligoadenylate Synthetase (OAS3) exerts antiviral effect against Chikungunya virus. *Virology*, 384(1):216–222, February 2009.

[162] Matthew B. McDougal, Anthony M. De Maria, Maikke B. Ohlson, Ashwani Kumar, Chao Xing, and John W. Schoggins. Interferon inhibits a model RNA virus via a limited set of inducible effector genes. *EMBO Reports*, 24(9):e56901, July 2023.

[163] Kuo-Chieh Liao and Mariano A. Garcia-Blanco. Role of Alternative Splicing in Regulating Host Response to Viral Infection. *Cells*, 10(7):1720, July 2021.

[164] Derek Walsh, Michael B. Mathews, and Ian Mohr. Tinkering with Translation: Protein Synthesis in Virus-Infected Cells. *Cold Spring Harbor Perspectives in Biology*, 5(1):a012351, January 2013.

[165] Craig H. Kerr, Michael A. Skinnider, Daniel D. T. Andrews, Angel M. Madero, Queenie W. T. Chan, R. Greg Stacey, Nikolay Stoynov, Eric Jan, and Leonard J. Foster. Dynamic rewiring of the human interactome by interferon signaling. *Genome Biology*, 21(1):140, June 2020.

[166] Yanrong Zhou, Wei Wu, Lilan Xie, Dang Wang, Qiyun Ke, Zhenzhen Hou, Xiaoli Wu, Ying Fang, Huanchun Chen, Shaobo Xiao, and Liurong Fang. Cellular RNA Helicase DDX1 Is Involved in Transmissible Gastroenteritis Virus nsp14-Induced Interferon-Beta Production. *Frontiers in Immunology*, 8, August 2017. Publisher: Frontiers.

[167] Claire E. Westcott, Cierra M. Isom, Deepa Karki, and Kevin J. Sokoloski. Dancing with the Devil: A Review of the Importance of Host RNA-Binding Proteins to Alphaviral RNAs during Infection. *Viruses*, 15(1):164, January 2023. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[168] Mohamed A. M. Ali. DEAD-box RNA helicases: The driving forces behind RNA metabolism at the crossroad of viral replication and antiviral innate immunity. *Virus Research*, 296:198352, April 2021.

[169] Frances Taschuk and Sara Cherry. DEAD-Box Helicases: Sensors, Regulators, and Effectors for Antiviral Defense. *Viruses*, 12(2):181, February 2020. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[170] Eckhard Jankowsky. RNA Helicases at work: binding and rearranging. *Trends in biochemical sciences*, 36(1):19, January 2011.

[171] Inga Jarmoskaite and Rick Russell. RNA helicase proteins as chaperones and remodelers. *Annual review of biochemistry*, 83:697, March 2014.

[172] Devanand Sarkar, Rob DeSalle, and Paul B. Fisher. Evolution of MDA-5/RIG-I-dependent innate immunity: Independent evolution by domain grafting. *Proceedings of the National Academy of Sciences of the United States of America*, 105(44):17040, October 2008.

[173] Jan Rehwinkel and Michaela U. Gack. RIG-I-like receptors: their regulation and roles in RNA sensing. *Nature Reviews Immunology*, 20(9):537–551, September 2020. Publisher: Nature Publishing Group.

[174] Ivan Akhrymuk, Ilya Frolov, and Elena I. Frolova. Both RIG-I and MDA5 detect alphavirus replication in concentration-dependent mode. *Virology*, 487:230, November 2015.

[175] Morgan Brisse and Hinh Ly. Comparative Structure and Function Analysis of the RIG-I-Like Receptors: RIG-I and MDA5. *Frontiers in Immunology*, 10, July 2019. Publisher: Frontiers.

[176] Rocío Daviña Núñez, Matthias Budt, Sandra Saenger, Katharina Paki, Ulrike Arnold, Anne Sadewasser, and Thorsten Wolff. The RNA Helicase DDX6 Associates with RIG-I to Augment Induction of Antiviral Signaling. *International Journal of Molecular Sciences*, 19(7):1877, July 2018. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

[177] Hiroyuki Oshiumi, Moeko Miyashita, Masaaki Okamoto, Yuka Morioka, Masaru Okabe, Misako Matsumoto, and Tsukasa Seya. DDX60 Is Involved in RIG-I-Dependent and Independent Antiviral Responses, and Its Function Is Attenuated by Virus-Induced EGFR Activation. *Cell Reports*, 11(8):1193–1207, May 2015. Publisher: Elsevier.

[178] Zhiqiang Zhang, Taeil Kim, Musheng Bao, Valeria Facchinetti, Sung Yun Jung, Amir Ali Ghaffari, Jun Qin, Genhong Cheng, and Yong-Jun Liu. DDX1, DDX21, and DHX36 Helicases Form a Complex with the Adaptor Molecule TRIF to Sense dsRNA in Dendritic Cells. *Immunity*, 34(6):866–878, June 2011.

[179] Xiang Liu, Adam Taylor, Yee Suan Poo, Wern Hann Ng, Lara J. Herrero, Patrick Chun Hean Tang, Ali Zaid, and Suresh Mahalingam. TIR-Domain-Containing Adapter-Inducing Interferon-β (TRIF)-Dependent Antiviral Responses Protect Mice against Ross River Virus Disease. *mBio*, 13(1):e03363–21, January 2022. Publisher: American Society for Microbiology.

[180] Iulia Tapescu, Frances Taschuk, Swechha M. Pokharel, Oleksandr Zginnyk, Max Ferretti, Peter F. Bailer, Kanupryia Whig, Emily A. Madden, Mark T. Heise, David C. Schultz, and Sara Cherry. The RNA helicase DDX39A binds a conserved structure in chikungunya virus RNA to control infection. *Molecular Cell*, 83(22):4174–4189.e7, November 2023. Publisher: Elsevier.

[181] Katherine G Zyner, Darcie S Mulhearn, Santosh Adhikari, Sergio Martínez Cuesta, Marco Di Antonio, Nicolas Erard, Gregory J Hannon, David Tannahill, and Shankar Balasubramanian. Genetic interactions of G-quadruplexes in humans. *eLife*, 8:e46793, July 2019. Publisher: eLife Sciences Publications, Ltd.

[182] Boris Bonaventure, Antoine Rebendenne, Ana Luiza Chaves Valadão, Mary Arnaud-Arnould, Ségolène Gracias, Francisco Garcia de Gracia, Joe McKellar, Emmanuel Labaronne, Marine Tauziet, Valérie Vivet-Boudou, Eric Bernard, Laurence Briant, Nathalie Gros, Wassila Djilli, Valérie Courgnaud, Hugues Parrinello, Stéphanie Rialle, Mickaël Blaise, Laurent Lacroix, Marc Lavigne, Jean-Christophe Paillart, Emiliano P Ricci, Reiner Schulz, Nolwenn Jouvenet, Olivier Moncorgé, and Caroline Goujon. The DEAD box RNA helicase DDX42 is an intrinsic inhibitor of positive-strand RNA viruses. *EMBO reports*, 23(11):e54061, November 2022. Publisher: John Wiley & Sons, Ltd.

[183] Frances Taschuk, Iulia Tapescu, Ryan H. Moy, and Sara Cherry. DDX56 Binds to Chikungunya Virus RNA To Control Infection. *mBio*, 11(5):10.1128/mbio.02623–20, October 2020. Publisher: American Society for Microbiology.

[184] Tomás Hernández-Díaz, Fernando Valiente-Echeverría, and Ricardo Soto-Rifo. RNA Helicase DDX3: A Double-Edged Sword for Viral Replication and Immune Signaling. *Microorganisms*, 9(6):1206, June 2021.

[185] Moushimi Amaya, Taryn Brooks-Faulconer, Tyler Lark, Forrest Keck, Charles Bailey, Venu Raman, and Aarthi Narayanan. Venezuelan equine encephalitis virus non-structural protein 3 (nsP3) interacts with RNA helicases DDX1 and DDX3 in infected cells. *Antiviral Research*, 131:49–60, July 2016.

[186] Mélanie Messmer, Louison Pierson, Charline Pasquier, Nikola Djordjevic, Johana Chicher, Philippe Hammann, Sébastien Pfeffer, and Erika Girardi. DEAD box RNA helicase 5 is a new pro-viral host factor for Sindbis virus infection. *Virology Journal*, 21:76, March 2024.

[187] Erika Girardi, Mélanie Messmer, Paula Lopez, Aurélie Fender, Johana Chicher, Béatrice Chane-Woon-Ming, Philippe Hammann, and Sébastien Pfeffer. Proteomics-based determination of double-stranded RNA interactome reveals known and new factors involved in Sindbis virus infection. *RNA*, 29(3):361, March 2023.

[188] Guillaume Giraud, Sophie Terrone, and Cyril F. Bourgeois. Functions of DEAD box RNA helicases DDX5 and DDX17 in chromatin organization and transcriptional regulation. *BMB Reports*, 51(12):613, December 2018.

[189] Jianhua Fang, Satoshi Kubota, Bin Yang, Naiming Zhou, Hui Zhang, Roseline Godbout, and Roger J. Pomerantz. A DEAD box protein facilitates HIV-1 replication as a cellular co-factor of Rev. *Virology*, 330(2):471–480, December 2004.

[190] Stephen P. Edgcomb, Andrew B. Carmel, Souad Naji, Geza Ambrus-Aikelin, Jason R. Reyes, Andrew C. S. Saphire, Larry Gerace, and James R. Williamson. DDX1 Is an RNA-Dependent ATPase Involved in HIV-1 Rev Function and Virus Replication. *Journal of Molecular Biology*, 415(1):61–74, January 2012.

[191] Rajan Lamichhane, John A. Hammond, Raymond F. Pauszek, III, Rae M. Anderson, Ingemar Pedron, Edwin van der Schans, James R. Williamson, and David P. Millar. A DEAD-box protein acts through RNA to promote HIV-1 Rev-RRE assembly. *Nucleic Acids Research*, 45(8):4632–4641, May 2017.

[192] Linghui Xu, Siti Khadijah, Shouguo Fang, Li Wang, Felicia P. L. Tay, and Ding Xiang Liu. The Cellular RNA Helicase DDX1 Interacts with Coronavirus Nonstructural Protein 14 and Enhances Viral Replication. *Journal of Virology*, 84(17):8571–8583, September 2010.

[193] Yuan-Qin Min, Mengzhuo Huang, Kuan Feng, Yajie Jia, Xiulian Sun, and Yun-Jia Ning. A new cellular interactome of SARS-CoV-2 nucleocapsid protein and its biological implications. *Molecular & Cellular Proteomics*, 0(0), May 2023. Publisher: Elsevier.

[194] Qiao Xue, Huisheng Liu, Qiaoying Zeng, Haixue Zheng, Qinghong Xue, and Xuepeng Cai. The DEAD-Box RNA Helicase DDX1 Interacts with the Viral Protein 3D and Inhibits Foot-and-Mouth Disease Virus Replication. *Virologica Sinica*, 34(6):610–617, July 2019.

[195] Lei Cao, Xianfeng Hui, Ting Xu, Haiying Mao, Xian Lin, Kun Huang, Lianzhong Zhao, and Meilin Jin. The RNA-Splicing Ligase RTCB Promotes Influenza A Virus Replication by Suppressing Innate Immunity via Interaction with RNA Helicase DDX1. *The Journal of Immunology Author Choice*, 211(6):1020–1031, September 2023.

[196] Zhenyu Lin, Jie Wang, Wenxian Zhu, Xiangyu Yu, Zhaofei Wang, Jingjiao Ma, Hengan Wang, Yaxian Yan, Jianhe Sun, and Yuqiang Cheng. Chicken DDX1 Acts as an RNA

Sensor to Mediate IFN-β Signaling Pathway Activation in Antiviral Innate Immunity. *Frontiers in Immunology*, 12, September 2021. Publisher: Frontiers.

[197] Johannes Popow, Markus Englert, Stefan Weitzer, Alexander Schleiffer, Beata Mierzwa, Karl Mechtler, Simon Trowitzsch, Cindy L. Will, Reinhard Lührmann, Dieter Söll, and Javier Martinez. HSPC117 Is the Essential Subunit of a Human tRNA Splicing Ligase Complex. *Science*, 331(6018):760–764, February 2011. Publisher: American Association for the Advancement of Science.

[198] Johannes Popow, Jennifer Jurkin, Alexander Schleiffer, and Javier Martinez. Analysis of orthologous groups reveals archease and DDX1 as tRNA splicing factors. *Nature*, 511(7507):104–107, July 2014. Publisher: Nature Publishing Group.

[199] Alicia Pérez-González, Ariel Rodriguez, Maite Huarte, Iñigo J. Salanueva, and Amelia Nieto. hCLE/CGI-99, a Human Protein that Interacts with the Influenza Virus Polymerase, Is a mRNA Transcription Modulator. *Journal of Molecular Biology*, 362(5):887–900, October 2006.

[200] Ariel Rodriguez-Frandsen, Susana de Lucas, Alicia Pérez-González, Maite Pérez-Cidoncha, Alejandro Roldan-Gomendio, Alejandra Pazo, Laura Marcos-Villar, Sara Landeras-Bueno, Juan Ortín, and Amelia Nieto. hCLE/C14orf166, a cellular protein required for viral replication, is incorporated into influenza virus particles. *Scientific Reports*, 6(1):20744, February 2016. Number: 1 Publisher: Nature Publishing Group.

[201] Alicia Pérez-González, Alejandra Pazo, Rosana Navajas, Sergio Ciordia, Ariel Rodriguez-Frandsen, and Amelia Nieto. hCLE/C14orf166 Associates with DDX1-HSPC117-FAM98B in a Novel Transcription-Dependent Shuttling RNA-Transporting Complex. *PLOS ONE*, 9(3):e90957, March 2014. Publisher: Public Library of Science.

[202] Yoshimitsu Kanai, Naoshi Dohmae, and Nobutaka Hirokawa. Kinesin Transports RNA: Isolation and Characterization of an RNA-Transporting Granule. *Neuron*, 43(4):513–525, August 2004. Publisher: Elsevier.

[203] Jennifer Jurkin, Theresa Henkel, Anne Færch Nielsen, Martina Minnich, Johannes Popow, Therese Kaufmann, Katrin Heindl, Thomas Hoffmann, Meinrad Busslinger, and Javier Martinez. The mammalian tRNA ligase complex mediates splicing of XBP1 mRNA

and controls antibody secretion in plasma cells. *The EMBO Journal*, 33(24):2922, November 2014.

[204] Yanyan Lu, Feng-Xia Liang, and Xiaozhong Wang. A Synthetic Biology Approach Identifies the Mammalian UPR RNA Ligase RtcB. *Molecular Cell*, 55(5):758–770, September 2014.

[205] Claudio Hetz. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nature Reviews Molecular Cell Biology*, 13(2):89–102, February 2012. Publisher: Nature Publishing Group.

[206] Lei Li, Mansi Garg, Yixiong Wang, Weiwei Wang, and Roseline Godbout. DEAD box 1 (DDX1) protein binds to and protects cytoplasmic stress response mRNAs in cells exposed to oxidative stress. *The Journal of Biological Chemistry*, 298(8):102180, June 2022.

[207] Saumya Jain, Joshua R. Wheeler, Robert W. Walters, Anurag Agrawal, Anthony Barsic, and Roy Parker. ATPase modulated stress granules contain a diverse proteome and substructure. *Cell*, 164(3):487, January 2016.

[208] Brian J Zarnegar, Ryan A Flynn, Ying Shen, Brian T Do, Howard Y Chang, and Paul A Khavari. irCLIP platform for efficient characterization of protein—RNA interactions. *Nature methods*, 13(6):489–492, June 2016.

[209] Andreas Buchbender, Holger Mutter, F. X. Reymond Sutandy, Nadine Körtel, Heike Hänel, Anke Busch, Stefanie Ebersberger, and Julian König. Improved library preparation with the new iCLIP2 protocol. *Methods*, 178:33–48, June 2020.

[210] Tianzhi Wu, Erqiang Hu, Shuangbin Xu, Meijun Chen, Pingfan Guo, Zehan Dai, Tingze Feng, Lang Zhou, Wenli Tang, Li Zhan, Xiaocong Fu, Shanshan Liu, Xiaochen Bo, and Guangchuang Yu. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, 2(3):100141, July 2021.

[211] Jürgen Cox and Matthias Mann. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26(12):1367–1372, December 2008. Publisher: Nature Publishing Group.

[212] Zhen-Lin Chen, Jia-Ming Meng, Yong Cao, Ji-Li Yin, Run-Qian Fang, Sheng-Bo Fan, Chao Liu, Wen-Feng Zeng, Yue-He Ding, Dan Tan, Long Wu, Wen-Jing Zhou, Hao Chi, Rui-Xiang Sun, Meng-Qiu Dong, and Si-Min He. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nature Communications*, 10(1):3404, July 2019. Publisher: Nature Publishing Group.

[213] Daehwan Kim, Ben Langmead, and Steven L. Salzberg. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4):357–360, April 2015. Publisher: Nature Publishing Group.

[214] Yang Liao, Gordon K. Smyth, and Wei Shi. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930, April 2014.

[215] 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, December 2014.

[216] G. K. Smyth. limma: Linear Models for Microarray Data. In Robert Gentleman, Vincent J. Carey, Wolfgang Huber, Rafael A. Irizarry, and Sandrine Dudoit, editors, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, pages 397–420. Springer, New York, NY, 2005.

[217] Christian von Mering, Lars J. Jensen, Berend Snel, Sean D. Hooper, Markus Krupp, Mathilde Foglierini, Nelly Jouffre, Martijn A. Huynen, and Peer Bork. STRING: known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic Acids Research*, 33(Database Issue):D433, December 2004.

[218] Charles Girardot, Jelle Scholtalbers, Sajoscha Sauer, Shu-Yi Su, and Eileen E.M. Furlong. Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinformatics*, 17(1):419, October 2016.

[219] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12, May 2011. Number: 1.

[220] 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):15–21, January 2013.

[221] Aaron R. Quinlan and Ira M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841, January 2010.

[222] Sudeep Sahadevan, Thileepan Sekaran, Nadia Ashaf, Marko Fritz, Matthias W Hentze, Wolfgang Huber, and Thomas Schwarzl. htseq-clip: a toolset for the preprocessing of eCLIP/iCLIP datasets. *Bioinformatics*, 39(1):btac747, January 2023.

[223] Nikolaos Ignatiadis, Bernd Klaus, Judith B. Zaugg, and Wolfgang Huber. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nature Methods*, 13(7):577–580, July 2016. Publisher: Nature Publishing Group.

[224] Timothy L. Bailey, James Johnson, Charles E. Grant, and William S. Noble. The MEME Suite. *Nucleic Acids Research*, 43(Web Server issue):W39, May 2015.

[225] Hector Zamora Carrera. Key Techniques in Structural Biology, Their Strengths and Limitations.

[226] Andrea Graziadei and Juri Rappsilber. Leveraging crosslinking mass spectrometry in structural and cell biology. *Structure*, 30(1):37–54, January 2022.

[227] J.F. Greenblatt, B.M. Alberts, and N.J. Krogan. Discovery and significance of protein-protein interactions in health and disease. *Cell*, 187(23):6501–6517, 2024.

[228] Juri Rappsilber, Symeon Siniossoglou, Eduard C. Hurt, and Matthias Mann. A Generic Strategy To Analyze the Spatial Organization of Multi-Protein Complexes by Cross-Linking and Mass Spectrometry. *Analytical Chemistry*, 72(2):267–275, January 2000. Publisher: American Chemical Society.

[229] Keiryn L. Bennett, Martin Kussmann, Marie Mikkelsen, Peter Roepstorff, Per Björk, Magdalena Godzwon, and Poul Sörensen. Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping—A novel approach to assess intermolecular protein contacts. *Protein Science*, 9(8):1503–1518, 2000. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1110/ps.9.8.1503.

[230] Andrea Sinz. Divide and conquer: cleavable cross-linkers to study protein conformation and protein–protein interactions. *Analytical and Bioanalytical Chemistry*, 409(1):33–44, January 2017.

[231] Shannon Eliuk and Alexander Makarov. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annual Review of Analytical Chemistry*, 8(Volume 8, 2015):61–80, July 2015. Publisher: Annual Reviews.

[232] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021. Publisher: Nature Publishing Group.

[233] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, June 2024. Publisher: Nature Publishing Group.

[234] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with AlphaFold-Multime, 2021.

[235] Kolja Stahl, Robert Warneke, Lorenz Demann, Rica Bremenkamp, Björn Hormes, Oliver Brock, Jörg Stülke, and Juri Rappsilber. Modelling protein complexes with crosslinking

mass spectrometry and deep learning. *Nature Communications*, 15(1):7866, September 2024. Publisher: Nature Publishing Group.

[236] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, Augustin Žídek, Tim Green, Kathryn Tunyasuvunakool, Stig Petersen, John Jumper, Ellen Clancy, Richard Green, Ankur Vora, Mira Lutfi, Michael Figurnov, Andrew Cowie, Nicole Hobbs, Pushmeet Kohli, Gerard Kleywegt, Ewan Birney, Demis Hassabis, and Sameer Velankar. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1):D439–D444, January 2022.

[237] Zhenyu Yang, Xiaoxi Zeng, Yi Zhao, and Runsheng Chen. AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduction and Targeted Therapy*, 8(1):1–14, March 2023. Publisher: Nature Publishing Group.

[238] Anastassis Perrakis and Titia K Sixma. AI revolutions in biology. *EMBO reports*, 22(11):e54046, November 2021. Publisher: John Wiley & Sons, Ltd.

[239] Claudio Iacobucci, Michael Götze, and Andrea Sinz. Cross-linking/mass spectrometry to get a closer view on protein interaction networks. *Current Opinion in Biotechnology*, 63:48–53, June 2020.

[240] Karen Manalastas-Cantos, Kish R. Adoni, Matthias Pfeifer, Birgit Märtens, Kay Grünewald, Konstantinos Thalassinos, and Maya Topf. Modeling Flexible Protein Structure With AlphaFold2 and Crosslinking Mass Spectrometry. *Molecular & Cellular Proteomics*, 23(3):100724, March 2024.

[241] Yuwan Chen, Wen Zhou, Yufei Xia, Weijie Zhang, Qun Zhao, Xinwei Li, Hang Gao, Zhen Liang, Guanghui Ma, Kaiguang Yang, Lihua Zhang, and Yukui Zhang. Targeted cross-linker delivery for the in situ mapping of protein conformations and interactions in mitochondria. *Nature Communications*, 14(1):3882, June 2023. Publisher: Nature Publishing Group.

[242] Krishnendu Khan, Camelia Baleanu-Gogonea, Belinda Willard, Valentin Gogonea, and Paul L. Fox. An optimized protocol for in vitro and in cellulo structural determination of the

multi-tRNA synthetase complex by cross-linking mass spectrometry. *STAR Protocols*, 3(1):101201, March 2022.

[243] Alena Kroupova, Fabian Ackle, Igor Asanović, Stefan Weitzer, Franziska M Boneberg, Marco Faini, Alexander Leitner, Alessia Chui, Ruedi Aebersold, Javier Martinez, and Martin Jinek. Molecular architecture of the human tRNA ligase complex. *eLife*, 10:e71656, December 2021. Publisher: eLife Sciences Publications, Ltd.

[244] Janina Lara Gerber, Suria Itzel Morales Guzmán, Lorenz Worf, Petra Hubbe, Jürgen Kopp, and Jirka Peschek. Structural and mechanistic insights into activation of the human RNA ligase RTCB by Archease. *Nature Communications*, 15(1):2378, March 2024. Publisher: Nature Publishing Group.

[245] Julian N. Kellner and Anton Meinhart. Structure of the SPRY domain of the human RNA helicase DDX1, a putative interaction platform within a DEAD-box protein. *Acta Crystallographica. Section F, Structural Biology Communications*, 71(Pt 9):1176–1188, August 2015.

[246] Alejandra Pazo, Alicia Pérez-González, Juan Carlos Oliveros, Maite Huarte, Juan Pablo Chavez, and Amelia Nieto. hCLE/RTRAF-HSPC117-DDX1-FAM98B: A New Cap-Binding Complex That Activates mRNA Translation. *Frontiers in Physiology*, 10, 2019.

[247] Claudia Ribeiro de Almeida, Somdutta Dhir, Ashish Dhir, Amin E. Moghaddam, Quentin Sattentau, Anton Meinhart, and Nicholas J. Proudfoot. RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. *Molecular Cell*, 70(4):650–662.e8, May 2018.

[248] Yuji Sunden, Shingo Semba, Tadaki Suzuki, Yuki Okada, Yasuko Orba, Kazuo Nagashima, Takashi Umemura, and Hirofumi Sawa. DDX1 Promotes Proliferation of the JC Virus through Transactivation of Its Promoter. *Microbiology and Immunology*, 51(3):339–347, 2007. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1348-0421.2007.tb03907.x.

[249] Francis J. O'Reilly, Liang Xue, Andrea Graziadei, Ludwig Sinn, Swantje Lenz, Dimitry Tegunov, Cedric Blötz, Neil Singh, Wim J. H. Hagen, Patrick Cramer, Jörg Stülke, Julia Mahamid, and Juri Rappsilber. In-cell architecture of an actively transcribing-translating

expressome. *Science*, 369(6503):554–557, July 2020. Publisher: American Association for the Advancement of Science.

[250] Ina Huppertz, Joel I. Perez-Perri, Panagiotis Mantas, Thileepan Sekaran, Thomas Schwarzl, Francesco Russo, Dunja Ferring-Appel, Zuzana Koskova, Lyudmila Dimitrova-Paternoga, Eleni Kafkia, Janosch Hennig, Pierre A. Neveu, Kiran Patil, and Matthias W. Hentze. Riboregulation of Enolase 1 activity controls glycolysis and embryonic stem cell differentiation. *Molecular Cell*, 82(14):2666–2680.e11, July 2022.

[251] Manuel Hilbert, Anne R. Karow, and Dagmar Klostermeier. The mechanism of ATP-dependent RNA unwinding by DEAD box proteins. *bchm*, 390(12):1237–1250, December 2009.

[252] Eric D Merkley, Steven Rysavy, Abdullah Kahraman, Ryan P Hafen, Valerie Daggett, and Joshua N Adkins. Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine–lysine distances. *Protein Science : A Publication of the Protein Society*, 23(6):747–759, June 2014.

[253] Julian N. Kellner, Jochen Reinstein, and Anton Meinhart. Synergistic effects of ATP and RNA binding to human DEAD-box protein DDX1. *Nucleic Acids Research*, 43(5):2813–2828, March 2015.

[254] Kazuo Fujiwara, Hiromi Toda, and Masamichi Ikeguchi. Dependence of α-helical and β-sheet amino acid propensities on the overall protein fold type. *BMC Structural Biology*, 12(1):18, August 2012.

[255] Justus Tegha-Dunghu, Beate Neumann, Simone Reber, Roland Krause, Holger Erfle, Thomas Walter, Michael Held, Phill Rogers, Kerstin Hupfeld, Thomas Ruppert, Jan Ellenberg, and Oliver J. Gruss. EML3 is a nuclear microtubule-binding protein required for the correct alignment of chromosomes in metaphase. *Journal of Cell Science*, 121(10):1718–1726, May 2008.

[256] Andrea Vandelli, Fernando Cid Samper, Marc Torrent Burgas, Natalia Sanchez de Groot, and Gian Gaetano Tartaglia. The Interplay Between Disordered Regions in RNAs and Proteins Modulates Interactions Within Stress Granules and Processing Bodies. *Journal of Molecular Biology*, 434(1):167159, January 2022.

[257] Khondker Ayesha Akter, Mohammed A. Mansour, Toshinori Hyodo, and Takeshi Senga. FAM98A associates with DDX1-C14orf166-FAM98B in a novel complex involved in colorectal cancer progression. *The International Journal of Biochemistry & Cell Biology*, 84:1–13, March 2017.

[258] Roseline Godbout, Margaret Hale, and Dwayne Bisgrove. A human DEAD box protein with partial homology to heterogeneous nuclear ribonucleoprotein U. *Gene*, 138(1):243–245, January 1994.

[259] Zonghong Li, Maoge Zhou, Zhaokui Cai, Hongyang Liu, Wen Zhong, Qiang Hao, Dongwan Cheng, Xihao Hu, Junjie Hou, Pingyong Xu, Yuanchao Xue, Yifa Zhou, and Tao Xu. RNA-binding protein DDX1 is responsible for fatty acid-mediated repression of insulin translation. *Nucleic Acids Research*, 46(22):12052–12066, December 2018.

[260] Quansheng Yang, Mark Del Campo, Alan M. Lambowitz, and Eckhard Jankowsky. DEAD-Box Proteins Unwind Duplexes by Local Strand Separation. *Molecular Cell*, 28(2):253–263, October 2007. Publisher: Elsevier.

[261] Linda C. Miller, Vanessa Blandford, Robyn McAdam, Maria R. Sanchez-Carbente, Frederique Badeaux, Luc DesGroseillers, and Wayne S. Sossin. Combinations of DEAD box proteins distinguish distinct types of RNA: Protein complexes in neurons. *Molecular and Cellular Neuroscience*, 40(4):485–495, April 2009.

[262] Sonali S. Patil, Tara B. Alexander, J. Akif Uzman, Chih-Hong Lou, Himika Gohil, and Amy K. Sater. Novel gene ashwin functions in Xenopus cell survival and anteroposterior patterning. *Developmental Dynamics*, 235(7):1895–1907, 2006. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/dvdy.20834.

[263] Peter C. Scacheri, Orit Rozenblatt-Rosen, Natasha J. Caplen, Tyra G. Wolfsberg, Lowell Umayam, Jeffrey C. Lee, Christina M. Hughes, Kalai Selvi Shanmugam, Arindam Bhattacharjee, Matthew Meyerson, and Francis S. Collins. Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proceedings of the National Academy of Sciences*, 101(7):1892–1897, February 2004. Publisher: Proceedings of the National Academy of Sciences.

[264] Charles O. Brostrom and Margaret A. Brostrom. Regulation of Translational Initiation during Cellular Responses to Stress. In Kivie Moldave, editor, *Progress in Nucleic Acid*

Research and Molecular Biology, volume 58, pages 79–125. Academic Press, January 1997.

[265] Thomas E. Dever. Gene-Specific Regulation by General Translation Factors. *Cell*, 108(4):545–556, February 2002. Publisher: Elsevier.

[266] David Ron. Translational control in the endoplasmic reticulum stress response. *The Journal of Clinical Investigation*, 110(10):1383–1388, November 2002. Publisher: American Society for Clinical Investigation.

[267] James P. White and Richard E. Lloyd. Regulation of stress granules in virus systems. *Trends in Microbiology*, 20(4):175–183, April 2012.

[268] Matthew Jefferson, Benjamin Bone, Jasmine L. Buck, and Penny P. Powell. The Autophagy Protein ATG16L1 Is Required for Sindbis Virus-Induced eIF2α Phosphorylation and Stress Granule Formation. *Viruses*, 12(1):39, December 2019.

[269] Igor Asanović, Emilia Strandback, Alena Kroupova, Djurdja Pasajlic, Anton Meinhart, Pai Tsung-Pin, Nemanja Djokovic, Dorothea Anrather, Thomas Schuetz, Marcin Józef Suskiewicz, Sirelin Sillamaa, Thomas Köcher, Rebecca Beveridge, Katarina Nikolic, Alexander Schleiffer, Martin Jinek, Markus Hartl, Tim Clausen, Josef Penninger, Peter Macheroux, Stefan Weitzer, and Javier Martinez. The oxidoreductase PYROXD1 uses NAD(P)+ as an antioxidant to sustain tRNA ligase activity in pre-tRNA splicing and unfolded protein response. *Molecular Cell*, 81(12):2520–2532.e16, June 2021. Publisher: Elsevier.

[270] Vincenzo Ruscica, Louisa Iselin, Ryan Hull, Azman Embarc-Buh, Samyukta Narayanan, Natasha Palmalux, Namah Raut, Quan Gu, Honglin Chen, Marko Noerenberg, Zaydah R. de Laurent, Josmi Joseph, Michelle Noble, Catia Igreja, David L. Robertson, Joseph Hughes, Shabaz Mohammed, Vicent Pelechano, Ilan Davis, and Alfredo Castello. XRN1 supplies free nucleotides to feed alphavirus replication. December 2024. Publisher: Cold Spring Harbor Laboratory.

[271] Elena I. Frolova, Oksana Palchevska, Francisco Dominguez, and Ilya Frolov. Alphavirus-induced transcriptional and translational shutoffs play major roles in blocking the formation of stress granules. *Journal of Virology*, 97(11):e00979–23, October 2023. Publisher: American Society for Microbiology.

[272] Matthew R. Hildebrandt, Devon R. Germain, Elizabeth A. Monckton, Miranda Brun, and Roseline Godbout. Ddx1 knockout results in transgenerational wild-type lethality in mice. *Scientific Reports*, 5(1):9829, April 2015. Publisher: Nature Publishing Group.

[273] Aisha Yesbolatova, Yuichiro Saito, Naomi Kitamoto, Hatsune Makino-Itou, Rieko Ajima, Risako Nakano, Hirofumi Nakaoka, Kosuke Fukui, Kanae Gamo, Yusuke Tominari, Haruki Takeuchi, Yumiko Saga, Ken-ichiro Hayashi, and Masato T. Kanemaki. The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nature Communications*, 11(1):5701, November 2020. Publisher: Nature Publishing Group.

[274] Gwen Nowee, Julian W. Bakker, Corinne Geertsema, Vera I. D. Ros, Giel P. Göertz, Jelke J. Fros, and Gorben P. Pijlman. A Tale of 20 Alphaviruses; Inter-species Diversity and Conserved Interactions Between Viral Non-structural Protein 3 and Stress Granule Proteins. *Frontiers in Cell and Developmental Biology*, 9, February 2021. Publisher: Frontiers.

[275] S Briolant, D Garin, N Scaramozzino, A Jouan, and J. M Crance. In vitro inhibition of Chikungunya and Semliki Forest viruses replication by antiviral compounds: synergistic effect of interferon-α and ribavirin combination. *Antiviral Research*, 61(2):111–117, February 2004.

[276] Farhana Abu Bakar and Lisa F. P. Ng. Nonstructural Proteins of Alphavirus—Potential Targets for Drug Development. *Viruses*, 10(2):71, February 2018. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[277] Rhian Jones, Gabriel Bragagnolo, Rocío Arranz, and Juan Reguera. Capping pores of alphavirus nsP1 gate membranous viral replication factories. *Nature*, 589(7843):615–619, January 2021. Number: 7843 Publisher: Nature Publishing Group.

[278] Rizwan Ullah, Jia Li, Puxian Fang, Shaobo Xiao, and Liurong Fang. DEAD/H-box helicases:Anti-viral and pro-viral roles during infections. *Virus Research*, 309:198658, February 2022.

[279] Yuliang Wu and Robert M Brosh. Helicase-inactivating mutations as a basis for dominant negative phenotypes. *Cell Cycle*, 9(20):4080–4090, October 2010.

[280] L. Therese Bergendahl, Lukas Gerasimavicius, Jamilla Miles, Lewis Macdonald, Jonathan N. Wells, Julie P. I. Welburn, and Joseph A. Marsh. The role of protein complexes in human genetic disease. *Protein Science*, 28(8):1400–1411, 2019. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pro.3667.

[281] Julia L. de Amorim, Sara W. Leung, Ramona Haji-Seyed-Javadi, Yingzi Hou, David S. Yu, Homa Ghalei, Sohail Khoshnevis, Bing Yao, and Anita H. Corbett. The putative RNA helicase DDX1 associates with the nuclear RNA exosome and modulates RNA/DNA hybrids (R-loops). *Journal of Biological Chemistry*, 300(2), February 2024. Publisher: Elsevier.

[282] Teruhiko Suzuki, Eiji Katada, Yuki Mizuoka, Satoko Takagi, Yasuhiro Kazuki, Mitsuo Oshimura, Mayumi Shindo, and Takahiko Hara. A novel all-in-one conditional knockout system uncovered an essential role of DDX1 in ribosomal RNA processing. *Nucleic Acids Research*, 49(7):e40, January 2021.

[283] Lei Li, Devon R. Germain, Ho-Yin Poon, Matthew R. Hildebrandt, Elizabeth A. Monckton, Darin McDonald, Michael J. Hendzel, and Roseline Godbout. DEAD Box 1 Facilitates Removal of RNA and Homologous Recombination at DNA Double-Strand Breaks. *Molecular and Cellular Biology*, 36(22):2794–2810, October 2016.

[284] Wen Zhong, Zonghong Li, Maoge Zhou, Tao Xu, and You Wang. DDX1 regulates alternative splicing and insulin secretion in pancreatic β cells. *Biochemical and Biophysical Research Communications*, 500(3):751–757, June 2018.

[285] Moushimi Amaya, Forrest Keck, Michael Lindquist, Kelsey Voss, Lauren Scavone, Kylene Kehn-Hall, Brian Roberts, Charles Bailey, Connie Schmaljohn, and Aarthi Narayanan. The Ubiquitin Proteasome System Plays a Role in Venezuelan Equine Encephalitis Virus Infection. *PLOS ONE*, 10(4):e0124792, April 2015. Publisher: Public Library of Science.

[286] Miguel Angel Sanz, Esther González Almela, and Luis Carrasco. Translation of Sindbis Subgenomic mRNA is Independent of eIF2, eIF2A and eIF2D. *Scientific Reports*, 7(1):43876, February 2017. Publisher: Nature Publishing Group.

[287] Miguel Angel Sanz, Esther González Almela, Manuel García-Moreno, Ana Isabel Marina, and Luis Carrasco. A viral RNA motif involved in signaling the initiation of

translation on non-AUG codons. *RNA*, 25(4):431–452, April 2019. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[288] Chiara Cabrelle, Federico Manuel Giorgi, and Daniele Mercatelli. Quantitative and qualitative detection of tRNAs, tRNA halves and tRFs in human cancer samples: Molecular grounds for biomarker development and clinical perspectives. *Gene*, 898:148097, March 2024.

[289] Christian Lorenz, Christina E. Lünse, and Mario Mörl. tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules*, 7(2):35, June 2017. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[290] M. Shigematsu, S. Honda, P. Loher, A.G. Telonis, I. Rigoutsos, and Y. Kirino. YAMAT-seq: An efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Research*, 45(9):e70, 2017.

[291] Vicki S. Chambers, Giovanni Marsico, Jonathan M. Boutell, Marco Di Antonio, Geoffrey P. Smith, and Shankar Balasubramanian. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature Biotechnology*, 33(8):877–881, August 2015. Publisher: Nature Publishing Group.

[292] Laura Sandra Lello, Age Utt, Koen Bartholomeeusen, Sainan Wang, Kai Rausalu, Catherine Kendall, Sandra Coppens, Rennos Fragkoudis, Andrew Tuplin, Luke Alphey, Kevin K. Ariën, and Andres Merits. Cross-utilisation of template RNAs by alphavirus replicases. *PLOS Pathogens*, 16(9):e1008825, September 2020. Publisher: Public Library of Science.

[293] Laura Sandra Lello, Agneta Miilimäe, Liubov Cherkashchenko, Ailar Omler, Rachel Skilton, Rachel Ireland, David Ulaeto, and Andres Merits. Activity, Template Preference, and Compatibility of Components of RNA Replicase of Eastern Equine Encephalitis Virus. *Journal of Virology*, 97(1):e01368–22, December 2022. Publisher: American Society for Microbiology.

[294] Maiwen Caudron-Herger, Scott F. Rusin, Mark E. Adamo, Jeanette Seiler, Vera K. Schmid, Elsa Barreau, Arminja N. Kettenbach, and Sven Diederichs. R-DeeP:

Proteome-wide and Quantitative Identification of RNA-Dependent Proteins by Density Gradient Ultracentrifugation. *Molecular Cell*, 75(1):184–199.e10, July 2019.

[295] Kotaro Fujii, Olena Zhulyn, Gun Woo Byeon, Naomi R. Genuth, Craig H. Kerr, Erin M. Walsh, and Maria Barna. Controlling tissue patterning by translational regulation of signaling transcripts through the core translation factor eIF3c. *Developmental Cell*, 56(21):2928–2937.e9, November 2021.

[296] Dan Li, Jihong Yang, Xin Huang, Hongwei Zhou, and Jianlong Wang. eIF4A2 targets developmental potency and histone H3.3 transcripts for translational control of stem cell pluripotency. *Science Advances*, 8(13):eabm0478, March 2022.

[297] Neal K. Devaraj and M. G. Finn. Introduction: Click Chemistry. *Chemical Reviews*, 121(12):6697–6698, June 2021. Publisher: American Chemical Society.

[298] Martin Bushell and Peter Sarnow. Hijacking the translation apparatus by RNA viruses. *The Journal of Cell Biology*, 158(3):395–399, August 2002.

[299] Jennifer L. Hyde, Rubing Chen, Derek W. Trobaugh, Michael S. Diamond, Scott C. Weaver, William B. Klimstra, and Jeffrey Wilusz. The 5' and 3' ends of alphavirus RNAs – Non-coding is not non-functional. *Virus Research*, 206:99–107, August 2015.

[300] Honglin Luo. Interplay between the virus and the ubiquitin–proteasome system: molecular mechanism of viral pathogenesis. *Current Opinion in Virology*, 17:1–10, April 2016.

[301] Yogesh A. Karpe, Kunal D. Pingale, and Gayatri D. Kanade. Activities of proteasome and m-calpain are essential for Chikungunya virus replication. *Virus Genes*, 52(5):716–721, 2016.

[302] Brian D. Carey, Allison Bakovic, Victoria Callahan, Aarthi Narayanan, and Kylene Kehn-Hall. New World alphavirus protein interactomes from a therapeutic perspective. *Antiviral Research*, 163:125–139, March 2019.

[303] Niloufar A. Boghdeh, Brittany McGraw, Michael D. Barrera, Carol Anderson, Haseebullah Baha, Kenneth H. Risner, Ifedayo V. Ogungbe, Farhang Alem, and Aarthi Narayanan. Inhibitors of the Ubiquitin-Mediated Signaling Pathway Exhibit Broad-Spectrum Antiviral Activities against New World Alphaviruses. *Viruses*, 15(3):655, February 2023.

[304] Gang Xu, Yezi Wu, Tongyang Xiao, Furong Qi, Lujie Fan, Shengyuan Zhang, Jian Zhou, Yanhua He, Xiang Gao, Hongxiang Zeng, Yunfei Li, and Zheng Zhang. Multiomics approach reveals the ubiquitination-specific processes hijacked by SARS-CoV-2. *Signal Transduction and Targeted Therapy*, 7(1):1–13, September 2022. Publisher: Nature Publishing Group.

[305] Derek Walsh and Mojgan H. Naghavi. Exploitation of Cytoskeletal Networks during Early Viral Infection. *Trends in Microbiology*, 27(1):39–50, January 2019. Publisher: Elsevier.

[306] Anne Salonen, Lidia Vasiljeva, Andres Merits, Julia Magden, Eija Jokitalo, and Leevi Kääriäinen. Properly Folded Nonstructural Polyprotein Directs the Semliki Forest Virus Replication Complex to the Endosomal Compartment. *Journal of Virology*, 77(3):1691–1702, February 2003. Publisher: American Society for Microbiology.

[307] Pirjo Spuul, Giuseppe Balistreri, Leevi Kääriäinen, and Tero Ahola. Phosphatidylinositol 3-Kinase-, Actin-, and Microtubule-Dependent Transport of Semliki Forest Virus Replication Complexes from the Plasma Membrane to Modified Lysosomes. *Journal of Virology*, 84(15):7543–7557, August 2010.

[308] Elena I. Frolova, Rodion Gorchakov, Larisa Pereboeva, Svetlana Atasheva, and Ilya Frolov. Functional Sindbis Virus Replicative Complexes Are Formed at the Plasma Membrane. *Journal of Virology*, 84(22):11679–11695, November 2010. Publisher: American Society for Microbiology.

[309] Johanna Luige, Alexandros Armaos, Gian Gaetano Tartaglia, and Ulf Andersson Vang Ørom. Predicting nuclear G-quadruplex RNA-binding proteins with roles in transcription and phase separation. *Nature Communications*, 15(1):2585, March 2024. Publisher: Nature Publishing Group.

[310] Ewan K.S. McRae, Evan P. Booy, Aniel Moya-Torres, Peyman Ezzati, Jörg Stetefeld, and Sean A. McKenna. Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Research*, 45(11):6656–6668, June 2017.

[311] Dhaval Varshney, Jochen Spiegel, Katherine Zyner, David Tannahill, and Shankar Balasubramanian. The regulation and functions of DNA and RNA G-quadruplexes. *Nature Reviews Molecular Cell Biology*, 21(8):459–474, August 2020. Publisher: Nature Publishing Group.

[312] Markus Sauer, Stefan A. Juranek, James Marks, Alessio De Magis, Hinke G. Kazemier, Daniel Hilbig, Daniel Benhalevy, Xiantao Wang, Markus Hafner, and Katrin Paeschke. DHX36 prevents the accumulation of translationally inactive mRNAs with G4-structures in untranslated regions. *Nature Communications*, 10(1):2421, June 2019. Publisher: Nature Publishing Group.

[313] Sebyung Kang, Liyuan Mou, Jason Lanman, Sadanandan Velu, Wayne J. Brouillette, and Peter E. Prevelige Jr. Synthesis of biotin-tagged chemical cross-linkers and their applications for mass spectrometry. *Rapid Communications in Mass Spectrometry*, 23(11):1719–1726, 2009. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rcm.4066.

[314] Helisa H. Wippel, Juan D. Chavez, Xiaoting Tang, and James E. Bruce. Quantitative interactome analysis with chemical cross-linking and mass spectrometry. *Current Opinion in Chemical Biology*, 66:102076, February 2022.

[315] Jan-Willem Taanman. The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1410(2):103–123, February 1999.

[316] Thalia Salinas-Giegé, Richard Giegé, and Philippe Giegé. tRNA Biology in Mitochondria. *International Journal of Molecular Sciences*, 16(3):4518–4559, February 2015.

[317] Vincent Meynier, Steven W. Hardwick, Marjorie Catala, Johann J. Roske, Stephanie Oerum, Dimitri Y. Chirgadze, Pierre Barraud, Wyatt W. Yue, Ben F. Luisi, and Carine Tisné. Structural basis for human mitochondrial tRNA maturation. *Nature Communications*, 15(1):4683, June 2024. Publisher: Nature Publishing Group.

[318] Paola H. Pinto, Alena Kroupova, Alexander Schleiffer, Karl Mechtler, Martin Jinek, Stefan Weitzer, and Javier Martinez. ANGEL2 is a member of the CCR4 family of deadenylases with 2',3'-cyclic phosphatase activity. *Science*, 369(6503):524–530, July 2020. Publisher: American Association for the Advancement of Science.

[319] Paula Clemente, Javier Calvo-Garrido, Sarah F. Pearce, Florian A. Schober, Megumi Shigematsu, Stefan J. Siira, Isabelle Laine, Henrik Spåhr, Christian Steinmetzger, Katja Petzold, Yohei Kirino, Rolf Wibom, Oliver Rackham, Aleksandra Filipovska, Joanna Rorbach, Christoph Freyer, and Anna Wredenberg. ANGEL2 phosphatase activity is

required for non-canonical mitochondrial RNA processing. *Nature Communications*, 13(1):5750, September 2022.

[320] Marcel Morgenstern, Christian D. Peikert, Philipp Lübbert, Ida Suppanz, Cinzia Klemm, Oliver Alka, Conny Steiert, Nataliia Naumenko, Alexander Schendzielorz, Laura Melchionda, Wignand W.D. Mühlhäuser, Bettina Knapp, Jakob D. Busch, Sebastian B. Stiller, Stefan Dannenmaier, Caroline Lindau, Mariya Licheva, Christopher Eickhorst, Riccardo Galbusera, Ralf M. Zerbes, Michael T. Ryan, Claudine Kraft, Vera Kozjak-Pavlovic, Friedel Drepper, Sven Dennerlein, Silke Oeljeklaus, Nikolaus Pfanner, Nils Wiedemann, and Bettina Warscheid. Quantitative high-confidence human mitochondrial proteome and its dynamics in cellular context. *Cell Metabolism*, 33(12):2464–2483.e18, December 2021.

[321] Zakery N. Baker, Patrick Forny, and David J. Pagliarini. Mitochondrial proteome research: the road ahead. *Nature Reviews Molecular Cell Biology*, 25(1):65–82, January 2024. Publisher: Nature Publishing Group.

[322] Frances V. Fuller-Pace. DEAD box RNA helicase functions in cancer. *RNA Biology*, 10(1):121–132, January 2013.

[323] Mengping Yuan, Jinyong Xu, Shuguang Cao, and Shuangshuang Sun. DDX1 is a prognostic biomarker and correlates with immune infiltrations in hepatocellular carcinoma. *BMC Immunology*, 23(1):59, November 2022.

[324] Kiyoko Tanaka, Narumi Ikeda, Kazuya Miyashita, Hideko Nuriya, and Takahiko Hara. DEAD box protein DDX1 promotes colorectal tumorigenesis through transcriptional activation of the LGR5 gene. *Cancer Science*, 109(8):2479–2489, 2018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cas.13661.

[325] K Tanaka, S Okamoto, Y Ishikawa, H Tamura, and T Hara. DDX1 is required for testicular tumorigenesis, partially through the transcriptional activation of 12p stem cell genes. *Oncogene*, 28(21):2142–2151, May 2009.

[326] Leena Pohjala, Age Utt, Margus Varjak, Aleksei Lulla, Andres Merits, Tero Ahola, and Päivi Tammela. Inhibitors of Alphavirus Entry and Replication Identified with a Stable Chikungunya Replicon Cell Line and Virus-Based Assays. *PLOS ONE*, 6(12):e28923, December 2011. Publisher: Public Library of Science.

[327] Francisco Dominguez, Oksana Palchevska, Elena I. Frolova, and Ilya Frolov. Alphavirus-based replicons demonstrate different interactions with host cells and can be optimized to increase protein expression. *Journal of Virology*, 97(11):e01225–23, October 2023. Publisher: American Society for Microbiology.

[328] Myra Hosmillo, Jia Lu, Michael R McAllaster, James B Eaglesham, Xinjie Wang, Edward Emmott, Patricia Domingues, Yasmin Chaudhry, Tim J Fitzmaurice, Matthew KH Tung, Marc Dominik Panas, Gerald McInerney, Nicolas Locker, Craig B Wilen, and Ian G Goodfellow. Noroviruses subvert the core stress granule component G3BP1 to promote viral VPg-dependent translation. *eLife*, 8:e46681, August 2019. Publisher: eLife Sciences Publications, Ltd.

[329] Eric M. Merten, John D. Sears, Tina M. Leisner, Paul B. Hardy, Anirban Ghoshal, Mohammad Anwar Hossain, Kesatebrhan Haile Asressu, Peter J. Brown, Michael A. Stashko, Laura E. Herring, Angie L. Mordant, Thomas S. Webb, Christine A. Mills, Natalie K. Barker, Jamie J. Arnold, Craig E. Cameron, Daniel N. Streblow, Nathaniel J. Moorman, Mark Heise, Timothy M. Willson, Konstantin I. Popov, and Kenneth H. Pearce. Discovery of a cell-active chikungunya virus nsP2 protease inhibitor using a covalent fragment-based screening approach, March 2024. Pages: 2024.03.22.586341 Section: New Results.

[330] Martin Dodel, Giulia Guiducci, Maria Dermit, Sneha Krishnamurthy, Emilie L. Alard, Federica Capraro, Zeinab Rekad, Lovorka Stojic, and Faraz K. Mardakheh. TREX reveals proteins that bind to specific RNA regions in living cells. *Nature Methods*, 21(3):423–434, March 2024. Publisher: Nature Publishing Group.