Unravelling a co-nsP-iracy: Investigating the functional characteristics of Chikungunya virus non-structural protein 3

Siu Yi Lee

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Faculty of Biological Sciences
School of Molecular and Cellular Biology

August 2022

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

In Chapter 3, nanopore sequencing was performed by Joseph Charles James Ward & Kate Loveday.

In Chapter 4, bioinformatic analysis of nsP3 binding to the host genome was performed by Euan Andrew McDonnell.

Acknowledgements

Firstly, I would like to thank my supervisors Mark and Nic for their continued intellectual support throughout this project and for being my science parents when emotional support was needed. Thank you for encouraging me to believe in my abilities, giving me the confidence to grow scientifically and professionally.

Thank you is also given to every past and current members of the Harris group, particularly the CHIKV wizards Ray, Jenny, and Grace. Thank you to Carsten and Las-Anya who have always lifted me when I was down, with their (occasionally mean-spirited) jokes. Thank you also to Joe and Kate, who have been there since the very beginning of my PhD with insightful scientific discussions and personal support, especially throughout the pandemic. I would also like to thank the wider Garstang 8.61 virology gang.

The biggest thanks goes to my mum, who has loved and supported me unconditionally, and taught me to never give up even when the tunnel is pitch black. My academic achievements would never have been possible without you.

媽咪,

多謝你永遠的支持, 無條件的愛, 和為我做的一切

沒有你我不可能完成這個博士學位

我們的壞天氣日子過去了, 太陽出來了

你永遠是我最愛的, 是我最重要的.

Thank you to Beany who has supported me throughout the final year PhD madness, especially during the thesis-writing period and for always telling me you're proud of me. Thanks must be given to my soulmate Alysha, for as long as I can remember you have been behind me and beside me, cheering me on.

Abstract

Chikungunya virus (CHIKV) is a re-emerging alphavirus transmitted to humans by the *Aedes* species of mosquito. Infection with CHIKV causes chikungunya fever, which can lead to debilitating chronic joint disease. Despite the rising potential as a threat to global health, no effective vaccine nor antiviral agents for prophylaxis or treatment are available.

The CHIKV non-structural protein 3 (nsP3) is essential to the virus lifecycle and is required for genome replication. However, to date, the exact role of this protein remains unclear. In *in vitro* studies, nsP3 has been shown to bind promiscuously to RNA but the precise nature of such interactions *in vivo* has not been explored. To investigate this, individual nucleotide UV-crosslinking and immunoprecipitation (iCLIP) combined with next generation sequencing were employed. This analysis revealed that nsP3 preferentially binds to the CHIKV genome, over host RNA. The nature of nsP3 binding to viral RNA also appeared stochastic. Further bioinformatics analyses were then applied to assess how nsP3 is able to recognise target RNA for binding. However, neither a consensus binding motif nor specific preferences for secondary RNA structures were identified, suggesting that these may not be key drivers of RNA binding recognition for this protein.

In parallel, this project also aimed to investigate a polyproline motif located in the C-terminal HVD of nsP3 which is important for RNA synthesis and interactions with host amphiphysin in multiple alphaviruses but the relevance in CHIKV has not been fully investigated. To address this, a panel of mutations targeting conserved residues of the motif were generated in both CHIKV sub-genomic replicon and full-length infectious virus. Most of the mutants were able to replicate and generate infectious virus in the respective systems, except one mutant (P398A/P401A) which was able to replicate in the subgenomic system but presented reduced replication and a complete inability to produce infectious virus. Further analysis revealed that the mutation responsible for the phenotype was not the intended mutation in the polyproline motif, but instead located to the 3′ UTR, which is essential for negative strand synthesis and subsequent genome replication. The results indicates that the polyproline motif is not absolutely essential for the CHIKV lifecycle.

Table of Contents

List of Figures

List of Tables

Abbreviations

26S RNA	CHIKV Subgenomic RNA
4-SU	4-thiouridine
6-SG	6-thioguanosine
APS	Ammonium persulfate
ATPSβ	ATP synthase β subunit
AUD	Alphavirus unique domain
BCA	Bicinchonic acid
BHK	Baby hamster kidney
BS	Binding site
BSA	Bovine serum albumin
BSL3	Biosafety level 3
C	Core protein
CHIKV	Chikungunya virus
CL	Crosslink
CLIP	Crosslink and immunoprecipitaion
CME	Clathrin-mediated endocytosis
CNS	Central nervous system
CPV	Cytopathic vacuoles
CRISPR	Clustered regularly interspaced short palindromic repeats
CSE	Conserved sequence element
DENV	Dengue virus
DEPC	Diethyl pyrocarbonate
DMSO	Dimethyl sulfoxide
DNA	Deoxiribose nucleic acid
dsRNA	double stranded RNA
E1	Envelope protein 1
E2	Envelope protein 2
E3	Envelope protein 3
eCLIP	Enhanced CLIP
ECSA	East/central/southern African
EDTA	Ethylenediaminetetraacetic acid
EEEV	Eastern equine encephalitis virus
ER	Endoplasmic reticulum
FBS	Foetal bovine serum
FDA	Food and Drug Administration
FHV	Flock House Virus
Fluc	Firefly luciferase
GBS	Guillain-Barré Syndrome
GFP	Green fluorescent protein
GLB	Glasgow lysis buffer
GST	Glutathione-S-Transferase
h	Hours

HCL	Hydrochloric acid
HCV	Hepatitis C virus
HITS	High-throughput sequencing
HITS-CLIP	High-throughput sequencing CLIP
HIV	Human immunodeficiency virus
HSP	Heat shock protein
HVD	Hypervariable domain
ICRES	"Integration of Chikungunya research" virus strain
INF	Interferon
irCLIP	Infared-CLIP
ITC	Isothermal titration calorimetry
kDa	Kilodaltons
LB	Luria-Bertani
MAR	Mono-ADP-ribose
MAYV	Mayaro virus
MC	Methyl cellulose
MEME	Multiple EM for motif elicitation
MOI	Multiplicity of infection
MOPS	3-(N-morpholino) propanesulfonic acid
mRNA	Messenger RNA
MRI	Magnetic resonance imaging
NaOH	Sodium hydroxide
NEAA	Non-essential amino acid
NEB	New England Biolabs
NMR	Nuclear magnetic resonance
NP	Nucleoprotein
nSP1	Non-structural protein 1
nSP2	Non-structural protein 2
nSP3	Non-structural protein 3
nSP4	Non-structural protein 4
nsPs	Non-structural proteins
NTPase	Nucleoside-triphosphatase
ONNV	O'nyong-nyong virus
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAR	Poly-ADP-ribose
PAR-CLIP	Photoactivable ribonucleoside
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PDB	Protein data bank
PFA	Paraformaldehyde
PFU	Plaque forming units
pH	Potential hydrogen
PHB-1	Prohibitin-1

PLB	Passive lysis buffer
PNK	T4 Polynucleotide kinase
PPM	Polyproline motif
PVDF	Polyvinylidene difluoride
QC	Quality control
qRT-PCR	Quantitative, reverse transcriptase PCR
RAR	Rear and release
RBP	RNA binding protein
RdRp	RNA dependent RNA polymerase
RNA	Ribose nucleic acid
RNase	Ribonuclease
RNAi	RNA interference Canonical RNAi
RNP	Ribonucleoprotein
RRV	Ross river virus
SARS	Severe acute respiratory syndrome
SDS	Sodium dodecyl sulphate
SFV	Semliki forest virus
SG	Stress granule
SGR	Subgenomic replicon
SH3	SRC Homology 3 Domain
SHAPE	Selective 2′ Hydroxyl Acylation analysed by Primer Extension
SINV	Sindbis virus
siRNA	Small interfering RNA
SSIV-RT	Super Script IV Reverse Transcriptase
ssRNA	Single stranded RNA
TAE	Tris acetate-EDTA
TBE	Tris-borate-EDTA
TF	Transframe
TST	Twin-strep-tag
UMI	Unique molecular identifier
UTR	Untranslated region
UV	Ultraviolet
UVA	Ultraviolet-A
UVB	Ultraviolet-B
UVC	Ultraviolet-C
VEEV	Venezuelan equine encephalitis virus
VLP	Virus like particle
VOPBA	Virus overlay protein binding assay
WEEV	Western equine encephalitis virus
WT	Wildtype

Chapter 1 Introduction

1.1 Chikungunya virus

1.1.1 Identification and classification of Chikungunya virus

Chikungunya virus (CHIKV) was first isolated from a patient in 1952 during an outbreak that resembled dengue-like fever in Tanzania. The word 'Chikungunya' translates to 'that which bends up' in the local dialect of Makonde, describing the physical contortion of joints observed in patients infected with the virus (Robinson, 1955; Zeller et al., 2016). CHIKV is a member of the *Togaviridae* family, under the *Alphavirus* genus. Viruses in this genus are enveloped, and contain a single-stranded, positive sense genome. Alphaviruses can be further divided into 'New World' or 'Old World' subsets according to geographical location and disease pathology (Garmashova et al., 2007). Full genome phylogenetic analysis of 29 recognised species of alphaviruses suggests that the origin of these viruses arose from the southern oceans then spread throughout the New and Old World, however further study on alphaviruses in the aquatic ecosystem is required (Forrester et al., 2012).

New World viruses are associated with encephalitic diseases, predominantly located in the Americas, and include Eastern Equine Encephalitis virus (EEEV), Western Equine Encephalitis virus (WEEV) and Venezuelan Equine Encephalitis virus (VEEV). Whilst New World viruses can infect humans, the primary host of these viruses are equine species. On the other hand, old world alphaviruses are associated with human rheumatic diseases and include CHIKV, O'nyong-nyong virus (ONNV), Mayaro virus (MAYV), Ross River virus (RRV), Semliki Forest virus (SFV), and Sindbis virus (SINV). Infection with Old World alphaviruses is associated with chronic and debilitating polyarthritis and/or polyarthralgia (Suhrbier et al., 2012). Other Old World alphaviruses exist which exclusively infect insects and fish (*Insect*: Eilat virus, Taï Forest alphavirus, Mwinilunga alphavirus (Hermanns et al., 2020) *Fish*: Salmonid alphavirus (Deperasińska et al., 2018)) (Figure 1.1).
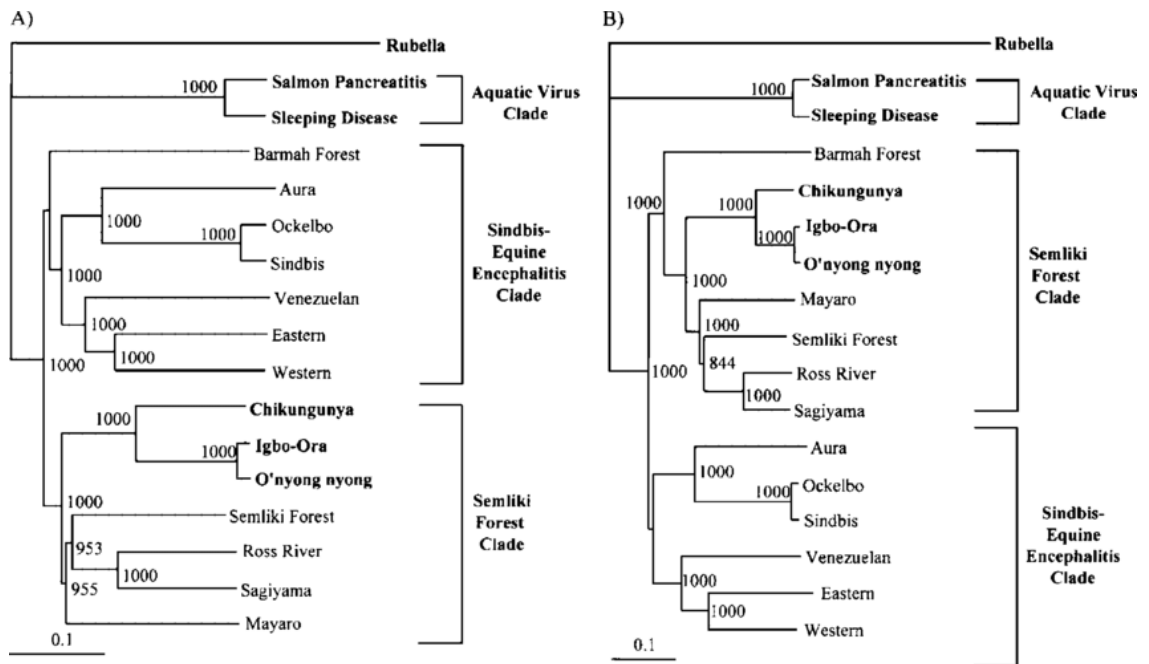
Figure 1.1 Togavirus phylogeny based on the complete genome. Each alphavirus clade is highlighted in bold. A) Phylogenetic tree generated from whole cDNA alphavirus genome sequences. B) Phylogenetic tree generated from whole amino acid alphavirus genome sequences. Bootstrap analysis consisted of 1000 replicates, and values represent the number of times the same branch was observed during phylogenetic reconstruction. Figure obtained from Luers et al., 2005.

1.1.2 Transmission of CHIKV

Human pathogenic alphaviruses are all mosquito-borne, where the presence of an infected mosquito vector is required for transmission. CHIKV transmission is maintained in either sylvatic or urban cycles (Figure 1.2). Sylvatic cycles are often observed in Africa, where the virus is maintained between mosquitoes and non-human primates but can spill over into human populations. On the other hand, urban cycles occur primarily in Asia, where humans are infected through human-mosquito-human interaction (Kumar et al., 2020). It is widely accepted that human-to-human transmission of CHIKV does not occur. Interestingly, studies have revealed the possibility of mother-to-child transmission (Ramful et al., 2007), and CHIKV has also been detected in the breast milk of infected mothers (Campos et al., 2017). However, there is no definitive report of vertical transmission.

CHIKV is spread by mosquitoes of the *Aedes* genus, most notably *Aedes aegypti* and *Aedes albopictus* which are commonly referred to as the yellow fever mosquito and the Asian tiger mosquito, respectively. However, CHIKV has also been detected in at least 6 other species of *Aedes* mosquitos (Coffey et al., 2014), and occasionally, in mosquitoes belonging to the *Anopheles* and *Culex* genera (Pialoux et al., 2007). Comparatively, *A. albopictus* has a wider geographical distribution, and is more resilient than *A. aegypti*. This is because the eggs of *A. albopictus* are tolerant to a dry climate and can remain viable outside of the rain season. Introduction of *A. albopictus* to a naïve environment can occur when vegetative eggs are dispersed during international import and export. *A. albopictus* can therefore be implicated in both urban and rural transmission cycles (Pialoux et al., 2007).
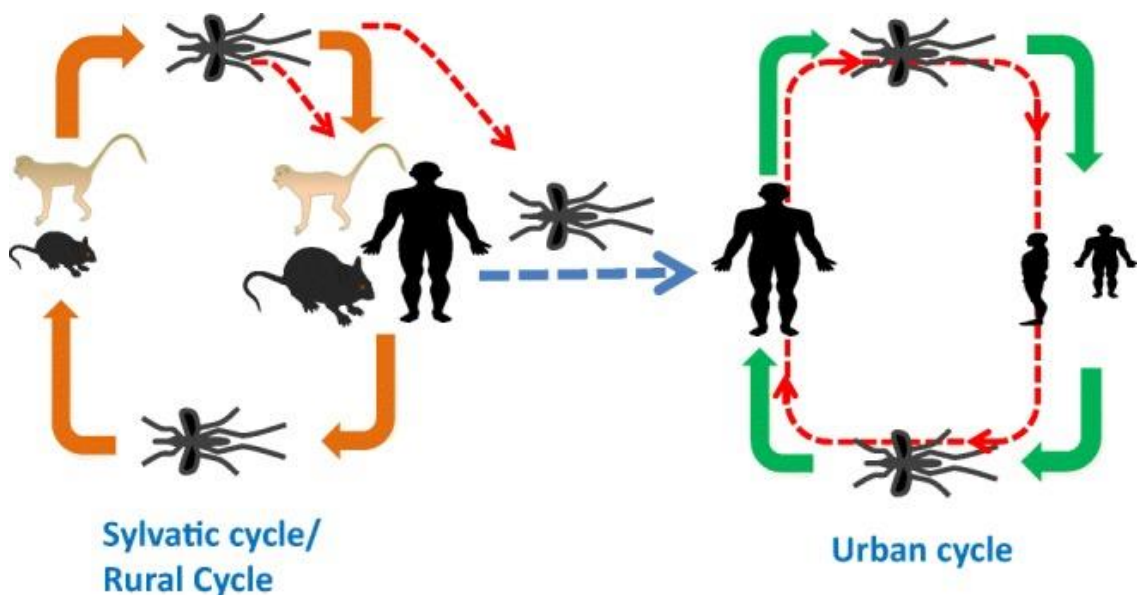
Figure 1.2 Transmission cycles of CHIKV. The virus is maintained in two cycles: a sylvatic cycle observed primarily in Africa, virus transmission occurs between non-human primates, rodents, bats and other vertebrates. In the urban cycle, observed primarily in Asia, virus is directly transmitted to humans and maintained in a human-mosquito-human cycle. Figure obtained from Kumar et al., 2020.

4

1.1.3 Epidemiology and Global Expansion

Differences in transmission cycles, combined with geographical separation and availability of mosquito vectors has allowed CHIKV to evolve into four genotypes (Figure 1.3): West African (WA), East/Central/South African (ECSA), Asian and East Indian/Indian Ocean Lineage. Since its isolation in 1952, CHIKV caused several sporadic and local outbreaks in Asia and Africa. This was the case until 2004, where a CHIKV virus belonging to the ECSA lineage emerged from costal Kenya and spread to approximately 60 countries, causing explosive epidemics (Lanciotti and Valadere, 2014). The most notable outbreak occurred in La Réunion, an island of the Indian Ocean. During this epidemic, lasting around one year (2005-2006), a third of the population was infected. This outbreak also marked the first CHIKV transmission via *A. albopictus* (Constant et al., 2021). Genetic analysis revealed that the responsible virus was an ECSA variant with an amino acid substitution of alanine to valine at position 226 of the envelope glycoprotein protein E1 (Schuffenecker et al., 2006). This variant, termed E1-226V was later responsible for the first autochthonous and European outbreak of CHIKV, in Italy. Since then, it has been responsible for multiple re-current outbreaks in Italy and France (Lindh et al., 2018). CHIKV in the Americas was not reported until 2013, where an outbreak occurred in Saint Martin. At present, CHIKV has been identified in over 45 countries and territories in the Americas (Figure 1.4) (Wahid et al., 2017).
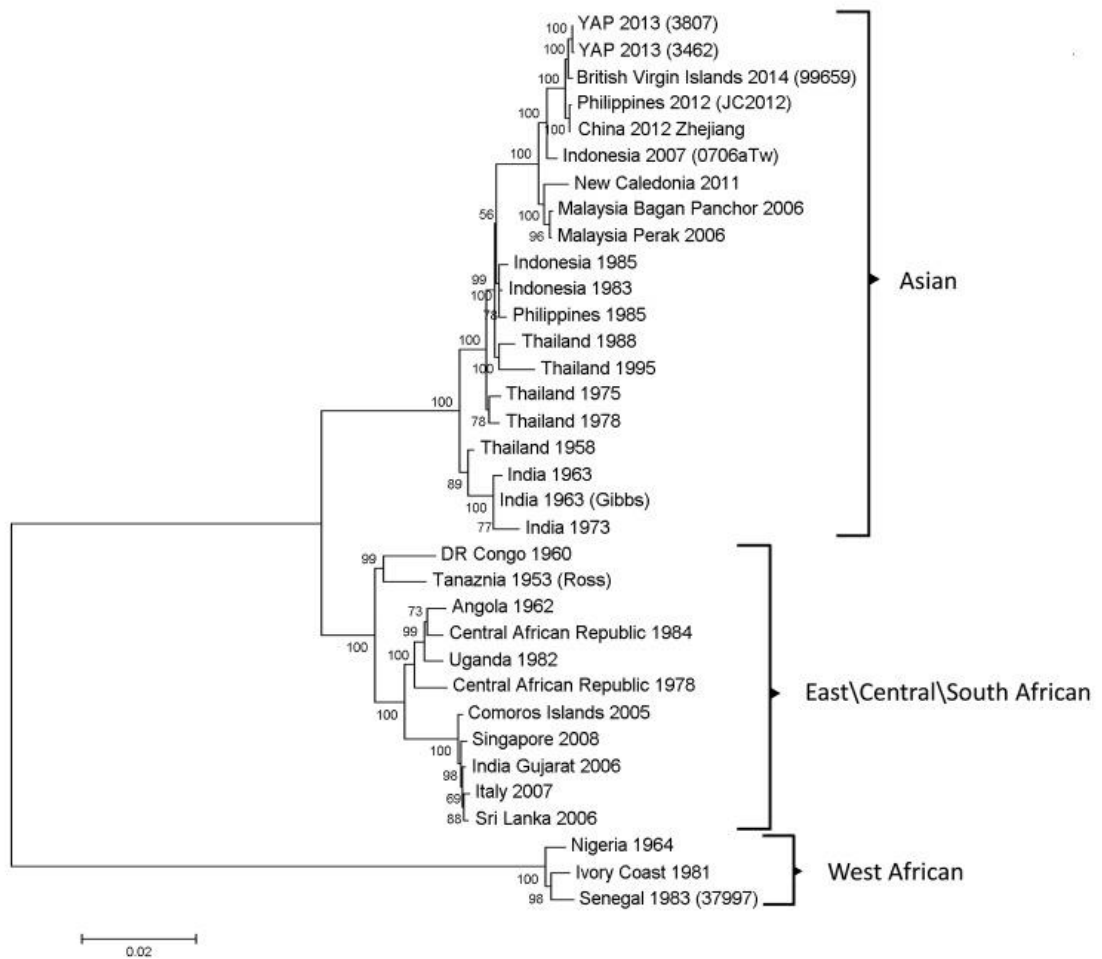
Figure 1.3 Evolutionary relationship between CHIKV Genotypes: Asian and East Indian/Indian Ocean Lineage, East/Central/South African and West African. Figure obtained from Lanciotti and Valadere, 2014.
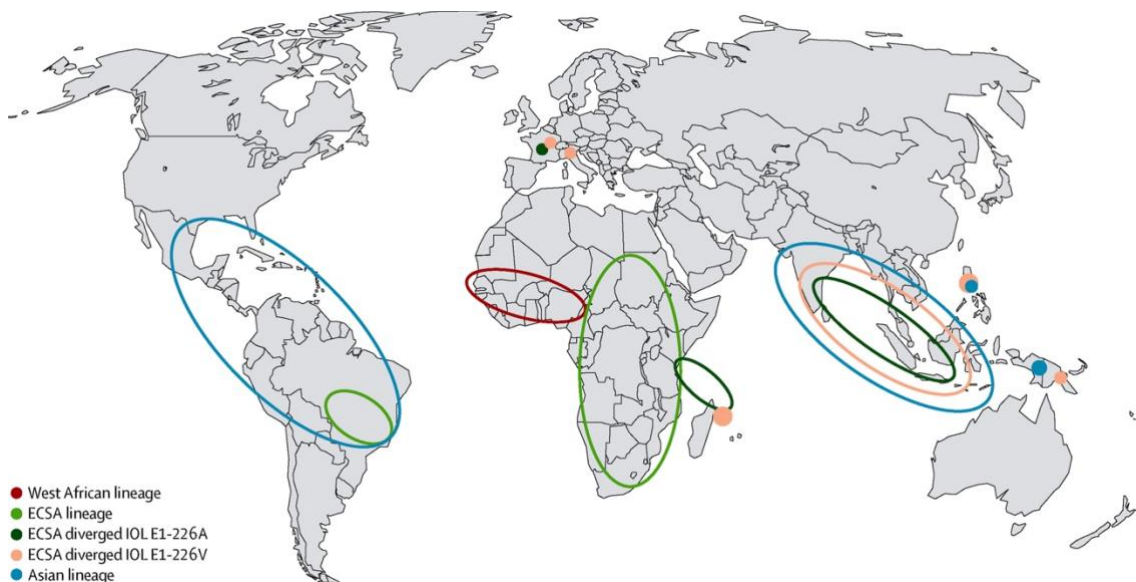
Figure 1.4 Global distribution of CHIKV. The spread of CHIKV across the world, with the responsible lineage highlighted. Red denotes the West African lineage. Light green denotes the East/Central/South African (ECSA) lineage. Dark green denotes the ECSA diverged E1-226A variant, specifically found in the Indian Ocean therefore also referred to as the Indian Ocean Lineage. Orange denotes the ECSA E1-226V variant responsible for the La Réunion outbreak in the Indian Ocean. Blue denotes the Asian lineage. Figure obtained from Burt et al., 2017.

1.1.4 Pathology and clinical manifestations of CHIKV

Transmission of CHIKV to humans occurs when a CHIKV-infected mosquito takes a blood meal from an individual. Mosquito saliva has been shown to play a crucial role in the enhancement of virus replication, predominantly by modulating the host's immune response. Specifically, it has been demonstrated that the saliva of *Aedes aegypti* can enhance CHIKV replication in human skin fibroblasts via the inhibition of the Type I interferon signalling pathway (Wichit et al., 2017). Following entry and infection of local cells, CHIKV enters the lymphatic and circulatory systems, causing rapid dissemination to major organs (Figure 1.5) (Caglioti et al., 2013). At peak viraemia, virus titre can reach 10^9 virus particles/mL which allows the virus to be easily transmitted to a CHIKV-naïve mosquito during a blood meal (Silva and Dermody, 2017).

Nature Reviews | Microbiology
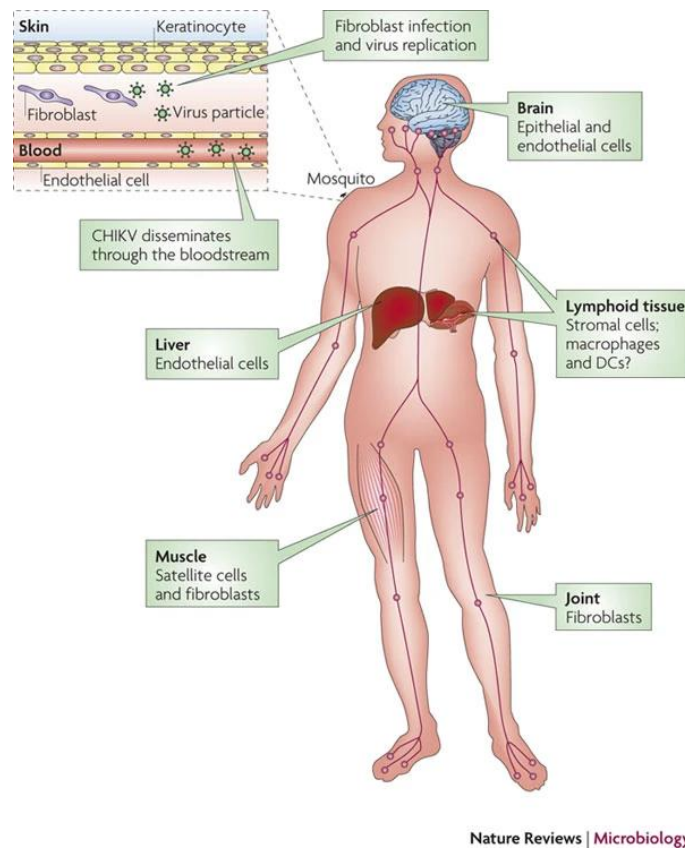
Figure 1.5 CHIKV dissemination throughout the human body. Following entry and infection of local dermal fibroblasts, CHIKV enters the circulatory and lymphatic systems to the liver, muscles, joints, lymphoid tissue, and brain where it systematically infects these organs. Figure obtained from Schwartz and Albert, 2010.

Infection with CHIKV causes chikungunya fever, which is clinically characterised into two stages: acute illness including a sudden onset of fever accompanied with a skin rash and incapacitating poly-arthralgia, and a chronic stage presented as persistent arthropathy. Acute symptoms usually resolve within 2 weeks without medical attention, but rheumatological manifestations may persist chronically for years after the initial infection. CHIKV associated polyarthropathy commonly affect joints in the fingers, wrists, ankles, knees, and shoulders. Although the exact cause of chronic joint disease caused by CHIKV infection has yet been determined, there is some conflicting evidence concerning whether lasting arthralgia is the result of continued CHIKV infection. It has been suggested that chronic musculoskeletal tissue pathology is caused by persistent and active CHIKV infection (Hawman et al., 2013). Conversely, another study

showed that presence of CHIKV RNA or proteins is not found in the synovial fluid of patients with continued arthritis (Chang et al., 2018).

1.1.5 Neurological Manifestation of CHIKV

Further to the typical symptoms highlighted in the previous section, CHIKV infection is associated with respiratory, renal, hepatic, neurological, cardiovascular, and gastrointestinal systems which are collectively referred to as "atypical features" (Mehta et al., 2018). Of the CHIKV-associated atypical features, neurological disorders appear to be the most common complication and account for up to 25% and 60% of atypical and severe atypical cases, respectively (Cerny et al., 2017). Such complications are a major cause of intensive care unit and fatalities in patients with chikungunya fever. Neurological signs present following a symptom-free period of 1-3 weeks, suggest a possible autoimmune process (Cerny et al., 2017).

Aside from the canonical CHIKV permissive cells, acute and chronic infection can occur in neurones/glia or central nervous system (CNS) neural progenitor cells, (Tang, 2012) leading to associated neurological disorders such as encephalopathy and encephalitis which are the most common presentations for arbovirus infection. Encephalopathy is defined as a "clinical state of altered mental status, manifesting as confusion, disorientation, behavioural changes or other cognitive impairment" whilst encephalitis is defined as inflammation of the brain caused by direct viral infection (Mehta et al., 2018). In one study comparing encephalopathic patients and patients presenting with encephalitis, the former cohort had more severe neurological disease despite no significant differences between viral load in serum or cerebrospinal fluid suggesting a role for host factors in neurovirulence (Mehta et al., 2018). Further, encephalitis caused by CHIKV does not show a distinct pattern in MRI imaging usually observed in encephalitis caused by canonical CNS pathogens such as Japanese encephalitis virus (Ganesan et al., 2008).

Ocular manifestations are another commonly reported disorder caused by CHIKV, and include retinitis, conjunctivitis, and optic neuritis (Mittal et al., 2007;

Murthy et al., 2008; Parola et al., 2006). Some manifestation begins during the acute phase of infection, whilst other complications have been reported more than 12 weeks post infection. In one retrospective study where 26 patients diagnosed with ocular complications were followed up after 3 months, 42% had improved visual acuity, 46% remained the same and 12% had worsened (Lalitha et al., 2007).

Guillain-Barré Syndrome (GBS) is a disorder caused by autoimmune insult of the peripheral nerves typically causing sensory-loss, limb weakness, and facial palsy symptoms (Seneviratne, 2000). Cases of CHIKV-associated GBS have risen following outbreaks of chikungunya and are usually secondary to virus infection. GBS is classified into variants according to clinical and pathological features, and specific variants have been assigned to common GBS-associated pathogens such as *Campylobacter jejuni* (Pritchard, 2010). However, acute and severe GBS associated with CHIKV infection presents with a range of variants (Farooq et al., 2018). In one clinical case report, a rare 'pharyngeal-cervical-brachial' variant of GBS was observed in a patient diagnosed with chikungunya fever 4 weeks before admission (Hameed and Khan, 2019).

To note, disorders resulting from direct CHIKV infection of the nervous systems occurs more commonly in neonates, infants and elderly patients whilst middle-aged, previously healthy patients, are affected by auto-immune forms particularly following an asymptomatic interval following infection (Cerny et al., 2017). Other neurological complications caused by CHIKV infection include myelopathy, myelitis, behavioural changes including attention disorders, memory loss, stroke and hearing loss. There is difficulty in characterising neurological disorders associated with CHIKV infection as most studies have been performed retrospectively, in part due to moderate attention being paid to the neurovirulence of the virus until the latter part 2010s; by 2018, only a total of 856 cases of CHIKV-associated neurological diseases were reported in literature (Mehta et al., 2018).

1.1.6 Diagnosis and treatment strategies

CHIKV has historically been mis-diagnosed as Dengue virus (DENV) due to overlaps in clinical presentations in many endemic areas. DENV infection is more likely to progress to a life-threatening condition thus it remains precedent to diagnose over CHIKV. Laboratory based diagnostic approaches include molecular-based reverse transcription polymerase chain reaction (RT-PCR) or quantitative polymerase chain reaction (qPCR) using primers against the CHIKV structural gene E1 and serology-based Immunoglobulin G (IgG)/Immunoglobulin M (IgM) detection in serum (Grivard et al., 2007). In areas where laboratory diagnosis is not readily accessible, attention is paid to symptoms such as the onset of fever where CHIKV associated febrile disease is more abrupt compared to the gradual progression seen in DENV. Maculopapular skin rash is also commonly associated with both types of infections but occurs at a higher frequency in CHIKV infection (Soto-Garita et al., 2018). In early stages of infection prior to onset of a skin rash, severe arthralgia can be used to distinguish CHIKV from other infections (Burt et al., 2012).

Although there are currently no CHIKV-specific antiviral drugs available for treatment, several non-specific antivirals including arbidol, ribavirin and interferon alpha have been recognised as effective against CHIKV in both *in vitro* and *in vivo* applications (Parashar and Cherian, 2014). Nevertheless, such compounds are not approved for clinical treatment of CHIKV. Therefore, patient management is focused on providing primary supportive care such as rehydration, rest and/or analgesic and anti-inflammatory medications to target pain resulting from polyarthralgia (Soto-Garita et al., 2018).

Various anti-CHIKV compounds have been the focus of research showing promising results in *in vitro* studies which can be categorised by their target: CHIKV entry inhibitors, non-structural protein 1 (nsP1) inhibitors, non-structural protein 2 (nsP2) inhibitors, non-structural protein 4 (nsP4) inhibitors, inhibitors of genome replication, and those with activity against CHIKV but have multiple/unidentified targets (Hucke and Bugert, 2020).

1.1.7 Prophylactics strategies to control CHIKV infection

1.1.7.1 Bite prevention and vector control

As there are no CHIKV-specific antivirals for the treatment of infection, the main management strategy against the virus is to avoid being bitten by an infected mosquito. Travellers are advised to maintain protective measures against mosquitoes such as long-sleeve clothing and using insect repellent. The other main strategy to prevent infection is via the control of vectors, which can be categorised into chemical and non-chemical approaches and can target mosquitoes at any stage of the lifecycle (larvae/pupae/adult).

In the past decade, efforts have been focused on the rear-and-release (RAR) of genetically modified mosquitoes that elicit detrimental effects on vertical offspring. There are three main RAR methods, one of which is using males to control the vector. In this method, male mosquitoes are sterilized via radiation, chemically using double stranded RNA to modify sperm production, genetically to reduce sperm production, or microbiologically using *Wolbachia* bacteria to induce sterility. Mating of a sterile male with a wild female results in infertile eggs which do not mature or hatch, termed cytoplasmic incompatibility (Ritchie and Johnson, 2017).

The use of *Wolbachia* to reduce mosquito virus transmission is another RAR method. *Wolbachia* infection of mosquitoes causes cytoplasmic incompatibility, and effectively sterilizes males by inducing early embryonic death. Theoretically, this method of vector control is extremely effective and could affect 75% of offspring following 1 generation. However, *Wolbachia* infected mosquitoes suffer from fitness cost which can limit vector viability and bacterial persistence in the population (Ritchie and Johnson, 2017). This method of vector control has been performed for *A. aegypti* with a specific focus of downstream effects on CHIKV transmission. In this study, *A. aegypti* infected with *Wolbachia* released in Colombia have reduced vector competence for CHIKV even at the presence of a high viral titre bloodmeal (Aliota et al., 2016).

The last RAR method harnesses gene drives to alter vector population which can be referred to as the 'transgenic' approach. During sexual reproduction, 50% of offspring will inherit one allele of any gene. Gene drives are selfish genetic element which does not conform to this rule thereby significantly increasing the likelihood of inheriting an allele carrying a gene drive element over the wild-type allele (Champer et al., 2016). Multiple genetic elements for gene drive have been exploited, including transposons and meiotic drivers, but the most effective results have been observed in the studies where clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein (Cas9) is used as a gene editing measure. Several studies have reported that Cas9 expression in the germline can be used as a robust genome engineering platform in *A. aegypti* to drive antipathogenic effector genes (Li et al., 2017; Reid et al., 2021). Notably, recent interrogation into vector genome polymorphism has invited questions on the feasibility of using CRISPR-Cas9 systems as a vector control method. Several studies have reported that natural, ubiquitous polymorphisms within Cas9-based gene drives provides mosquitoes with resistant alleles but a large-scale study including over 1200 mosquito genomes found an abundance of conserved target sites. Combined with flexibility in Cas9-based gene drive design, concerns over the practicality of this method can be minimised (Schmidt et al., 2020).

1.1.7.2 Vaccine development

Whilst distinct clades and lineages have arisen due to reasons highlighted in 1.1.3, CHIKV displays limited diversity across different strains. Furthermore, it is widely acknowledged that CHIKV exists as a single serotype meaning immunity raised against one strain of the virus provides life-long protection to strains from other lineages (Langsjoen et al., 2018), making CHIKV a viable candidate for vaccine development. Although development began in the 1960s, no vaccines have been approved for use, but several candidates have progressed into phase 3 clinical trial with promising results.

VRC-CHIKVLP059-00-VP is a virus-like particle (VLP) vaccine assembled from CHIKV proteins expressed in mammalian cells. Virus-like particles structurally

resembles a wild-type virus particle, but do not contain the genetic material thus cannot replicate and are not infectious. The presence of virus structural proteins generates an immune response in the recipient (Akahata et al., 2010). To generate this vaccine candidate, CHIKV structural proteins were inserted into a cytomegalovirus expression vector and transfected into 293A human renal epithelial cells to generate VLPs. Immunization of monkeys with the resulting VLPs elicited high titres of neutralising antibodies against multiple CHIKV strains. Additionally, inoculation of the neutralising antibodies into immunodeficient mice subsequently offered protection against lethal doses of CHIKV (Akahata et al., 2010), The outcome of both Phase 1 (2014) and Phase 2 (2020) clinical trials demonstrated the vaccine to be safe, well tolerated and immunogenic for up to 72 weeks following vaccination (Chang et al., 2014; Chen et al., 2020). The vaccine will next move into Phase 3 clinical trials, where the clinical efficacy of this candidate will be interrogated.

Another common vaccine development method exploits virus attenuation. Live-attenuated vaccines (LAV) are advantageous as they offer broad, life-long protection whilst being relatively low-cost to manufacture. Virus attenuation can be achieved by culturing for an extended period at sub-optimal conditions or selecting and promoting mutations that are less replicative than wild type. Attenuated viruses are less virulent but mimic a natural infection as their structural and genetic material are intact (Hanley, 2011). To date, only two LAVs have reached clinical trials. The first of which was based on two amino-acid substitutions in the E2 envelope glycoprotein (Gorchakov et al., 2012), which later reverted into an infectious strain upon introduction into human subjects. The reversion translated to symptoms of mild arthralgia in subjects during Phase 2 clinical trial, the study was subsequently terminated due to safety concerns (Edelman et al., 2000).

VLA1533 is the second LAV and is arguably the current most promising CHIKV vaccine candidate. Attenuation in this vaccine is achieved through a 183 base pair deletion in the 3' end of gene encoding non-structural protein 3 (nsP3) (Hallengärd et al., 2014). Following Phase 3 clinical trial, the vaccine was shown

to confer seroprotection to more than 98% of participants after one month, and more than 96% of participants after six months with confirmed safety and tolerability profiles. As of March 2022, the vaccine is expected to commence a pre-submission process to the Food and Drug Administration (FDA) in the latter part of 2022 ("Valneva Successfully Completes Pivotal Phase 3 Trial of Single-Shot Chikungunya Vaccine Candidate – Valneva,").

1.2 Molecular Biology of CHIKV

1.2.1 CHIKV virion organisation

Alphaviruses virions have a diameter of 60 – 70 nm, and are all spherical and enveloped. A virion is organised into three concentric layers (Figure 1.6A): the virus genome is encased in an icosahedral nucleocapsid core, contained within a host-derived lipid bilayer, and the outermost layer is the envelope constructed from E1 and E2 organised into trimers of E1/E2 heterodimers (shown in Figure 1.6B). Glycoprotein spikes, along with the nucleocapsid core are aligned to form a *T=4* lattice (Button et al., 2020). E1 is responsible for mediating membrane fusion between the viral and host membranes following entry into the host cell via a fusion loop contained within the protein (Holmes et al., 2020). On the other hand, E2 has been shown to bind and interact with cellular receptor to facilitate entry into the host. Previously, it was thought that E2 alone mediated this crucial step in the virus lifecycle (Voss et al., 2010), but recent evidence has highlighted the importance of both proteins for this interaction (Basore et al., 2019).
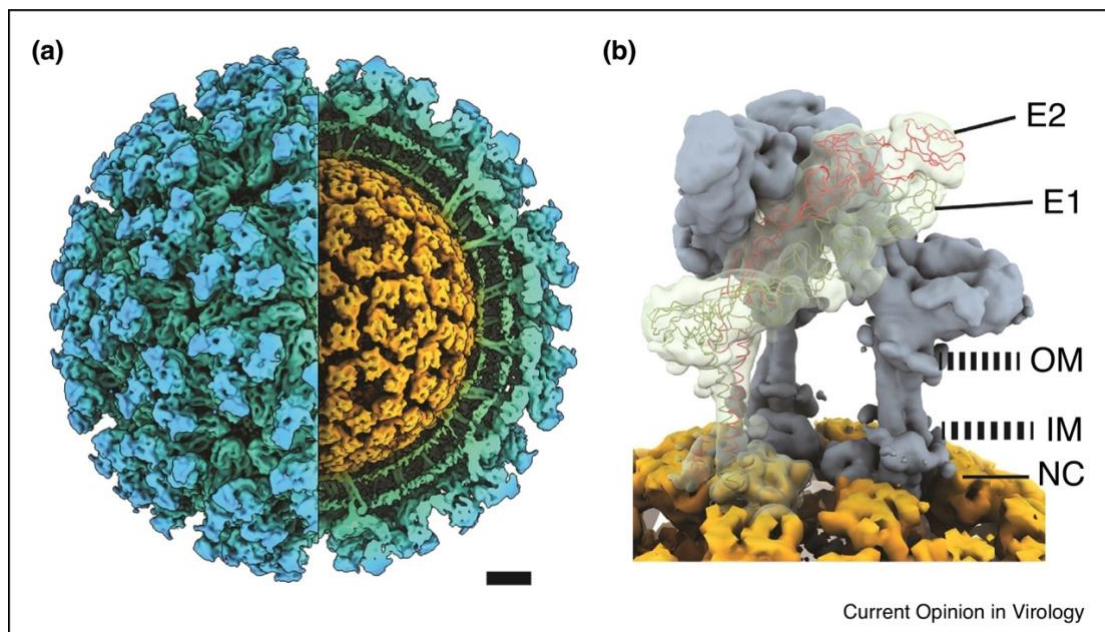
Figure 1.6 The alphavirus virion. A) A cutaway isosurface representation of the alphavirus virion with the three concentric layers highlighted in different colours. Yellow represents the virus nucleocapsid, green is the host-derived lipid bilayer and blue is the envelope layer constructed from viral E1/E2 glycoproteins. B) A single spike trimer, formed from E1/E2 heterodimers. Two heterodimers are shown in grey, and one is shown as an atomic trace. Figure obtained from Button et al., 2020

1.2.1.1 CHIKV genome organisation and features

CHIKV has a positive sense, single stranded RNA genome of 11.8 kb that is organised into two open reading frames (ORFs) separated by a junction region (Figure 1.7A). Similarly, to a eukaryotic messenger RNA molecule, the 5' end of the genome is capped with an m^7G cap, and the 3' end is polyadenylated (Solignat et al., 2009).
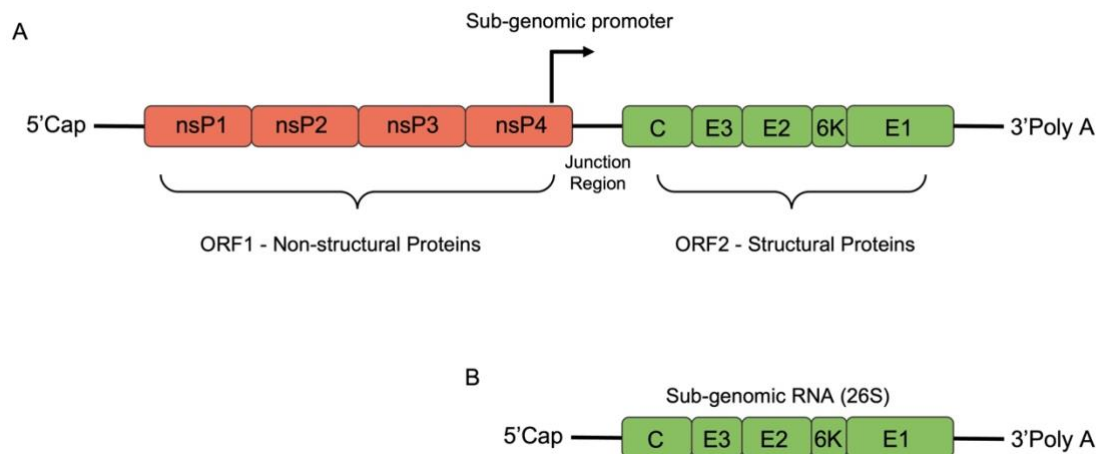
Figure 1.7 The genome organisation of CHIKV. A) The CHIKV genome has a m7G cap is present at the 5' end and a poly-A-tail at the 3' end. The coding region for the non-structural proteins is shown in red, and the coding region for the structural proteins is shown in green, separated by a junction region. The structural proteins are under a sub-genomic promoter. B) The CHIKV subgenomic RNA contains the same features as the full-length genome but only contain the structural protein coding region.

At the 5' end, the first ORF encodes non-structural proteins nsP1, nsP2, nsP3 and nsP4 which form the virus replication complex. nsP1 has methyltransferase and guanylyltransferase activity responsible for 5' genome capping. It is also the only characterised replication complex membrane anchor. The N-terminus of nsP2 has NTPase, helicase and RNA triphosphatase activity whilst the C-terminus has cysteine protease activity for non-strutural polyprotein processing. nsP4 is the virus RNA-dependent RNA polymerase (RdRp). The role of nsP3 remains enigmatic, and no specific function has been assigned, but has been shown to be important in minus-strand synthesis and is able to interact with host proteins and RNA (Silva and Dermody, 2017). An in-depth overview of nsP3 is provided in section 1.2.3.

The second ORF encodes the CHIKV structural proteins which are expressed through an internal subgenomic promoter (Figure 1.7B). In the alphavirus dogma, the resulting subgenomic RNA encodes 5 proteins: C, E3, E2, 6K and E1. Three of the expressed proteins, C, E1 and E2 are components of a mature virion. C is a non-glycosylated nucleocapsid protein whilst E1 and E2 form the viral glycoproteins. The remaining structural proteins are responsible for correct protein and virion formation: E3 facilitates folding of the E2 precursor and formation of E1-E2 heterodimers whilst 6K assists in the transport of structural proteins to the endoplasmic reticulum, the cleavage of the E2 precursor, and has been demonstrated to be important for virus budding (Singh et al., 2018). Whilst the production of 5 structural proteins has been previously accepted as the bona fide products of the subgenomic RNA, bioinformatics analysis later revealed a novel structural protein, transframe (TF), that is produced as a result of a ribosomal frameshift event during translation of the 6K gene (Snyder et al., 2013). Whilst TF is not absolutely required in cell culture, a mutation in this protein reduces release of virus particles (Snyder et al., 2013).

Alphavirus genomes contain many RNA secondary structures in both coding and non-coding regions which play important roles in the virus lifecycle. Although alphaviruses share a significant amount of similarities in their genomic organisation and economy, there is conflicting evidence over the conservation of RNA secondary structures between members of this genus. Whilst specific conserved sequence elements appear to be well conserved between SINV, SFV and VEEV, studies have demonstrated that most RNA secondary structures are not well conserved between alphaviruses and this structural divergence suggests functional specificity to individual viruses (Kutchko et al., 2018; Michel et al., 2007).

For CHIKV, RNA secondary structures have been investigated in both the 5' and 3' ends using selective 2'-hydroxyl acylation analysed by primer extension (SHAPE). Six stem loops within the 5' UTR and the adjacent nsP1 coding region were investigated for their roles in CHIKV replication. Specifically, one RNA structure was observed to be required in the transcription of the positive sense genome in a structure dependent manner. The importance of this stem loop was

also shown to be cell-type, and temperature specific. Further, this study demonstrated that the formation of this stem loop appears stochastic, suggesting versatility of RNA secondary structures within the CHIKV genome (Kendall et al., 2019). At the 3' end, SHAPE followed by 3' Rapid Amplification of cDNA ends (3'RACE) was used to characterise the UTR of a CHIKV isolate from the 2013 Caribbean outbreak. Three 3'UTR isoforms were identified and tested for their ability to replicate in mosquito cells where distinct phenotypes were displayed. However, investigation in an in vivo study highlighted that the 3'UTR variants had no effect on virus infection (Madden et al., 2020).

1.2.2 The CHIKV lifecycle
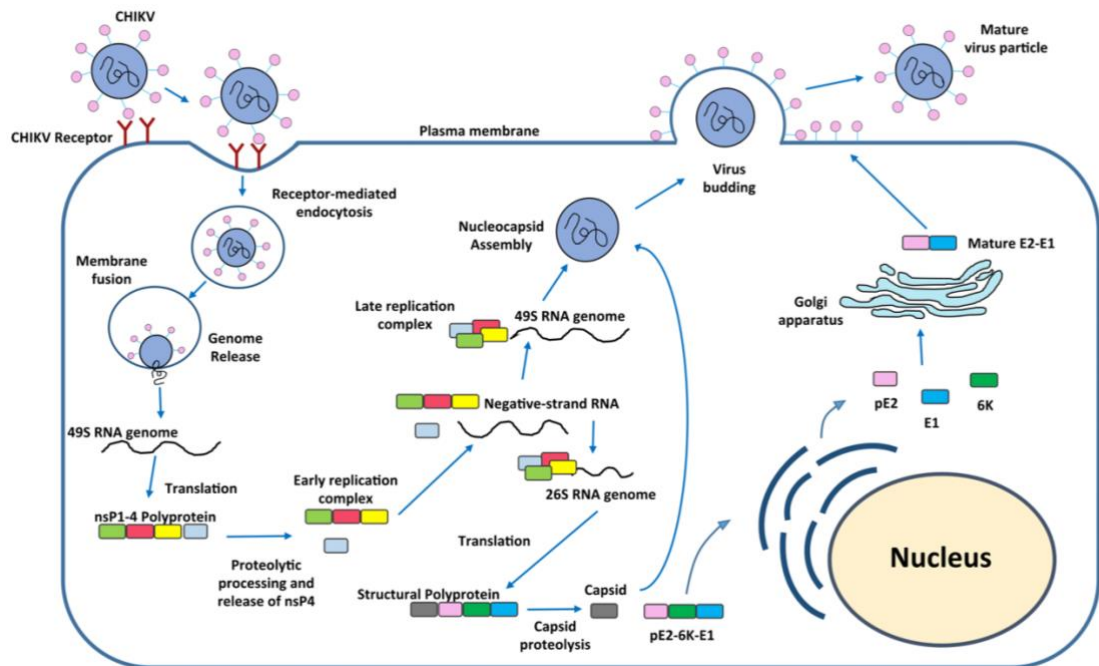
Figure 1.8 The alphavirus lifecycle. CHIKV binds to the target receptor on the host cell and enters through endocytosis. Changes in pH triggers fusion of the viral membrane with the endosomal membrane to release the genome into the cytoplasm. Free genomic RNA (49S) is directly translated into the nsP polyprotein P1234. Proteolytic processing and release of nsP4 from the polyprotein allows nsP4 to interact with P123 to form the early replication complex which synthesises negative-strand RNA. Further auto-proteolysis and processing forms mature nsP1, nsP2 and nsP3. All four mature non-structural proteins interact to produce the late replication complex, synthesising the genomic and subgenomic RNA (26S RNA). The subgenomic RNA is translated into the structural polyprotein which are then post-translationally processed into the CHIKV structural proteins C, E3, E2, 6K and E1. The glycoproteins are eventually trafficked to the plasma membrane. Here they assemble with the nucleocapsid to form a mature CHIKV particle. Figure adapted from (Abdelnabi et al., 2015).

1.2.2.1 Cell attachment & entry

To establish an infection, the virus must first enter the host cell (Figure 1.8). Over the years, efforts have been made to identify the cell receptor required for alphavirus cell entry and several, along with associated attachment factors, have been described (Figure 1.9). Attachment factors allow the virus to contact the target cell then the virus can directly interact with the entry receptor to gain access into the host. Heparan sulphate, C-type lectins and phosphatidylserine receptors have been implicated as alphavirus attachment factors. To qualify as a cell receptor, the virus must directly bind and mediate subsequent internalisation. Further, preventing this interaction using an antibody, receptor docking molecules or mutagenesis must prevent virus infection.
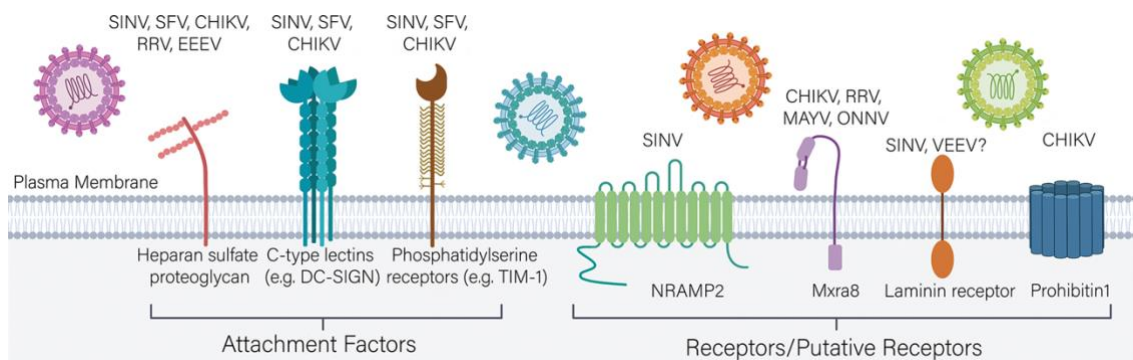
Figure 1.9 Alphavirus receptors and attachment factors. Multiple cell surface molecules have been described for alphaviruses including Heparan Sulphate proteoglycan, C-type lectins and phosphatidylserine receptors which are required for initial attachment. Some have been shown to increase infectivity. Receptors such as NRAMP2 for SINV and Mxra8 for CHIKV, RRV, MAYV and ONNV appear to be true receptors as they satisfy all required criteria. Figure obtained from (Holmes et al., 2020).

Three cell surface receptors have been described for CHIKV, the first of which is Prohibitin-1 (PHB1), a protein that regulates cell proliferation and preserves mitochondrial integrity (Della-Flora Nunes et al., 2021). The requirement of PHB1 as a CHIKV receptor was first characterised in microglial cells using a two-dimensional virus overlay protein binding assay (VOPBA). Co-localisation of PHB1 and E2 was observed post infection, and co-immunoprecipitation was subsequently confirmed. Treatment with either an anti-PHB1 antibody or siRNA knock down of PHB1 led to a reduction in virus infection (Wintachai et al., 2012).

Later, using the same binding assay, PHB1 was shown to be important as a cell receptor in *Aedes* mosquitoes (Ghosh et al., 2019).

A cell adhesion molecule containing two immunoglobulin G-like domains, Mxra8, is the second characterised receptor for CHIKV-entry. This receptor has also been described as the target for other alphaviruses such as MAYV and ONNV (Zhang et al., 2018). One study used a CRISPR/Cas9-based screen on the mouse genome to confirm Mxra8 as a CHIKV-cell receptor. Subsequent gene editing of both mouse and human Mxra8 led to decreased viral infection and the over expression of this molecule had the opposite effect on infection (Zhang et al., 2018). Later, two research groups were able to solve the crystal structure of Mxra8 in complex with a CHIKV-virion to demonstrate that Mxra8 binds in the canyon spanning between E1 and E2 heterodimers (Song et al., 2019).

Whilst the requirement for Mxra8 has been demonstrated for murine and human infection, an ortholog for this protein does not exist in mosquito. Instead, the most well-characterised receptor for CHIKV entry into mosquito cells is ATP synthase β subunit (ATPSβ). ATP synthases are found on the inner mitochondrial membrane and are responsible for the catalysis of ATP synthesis from ADP and phosphate. In this study, ATPSβ was shown to interact with CHIKV E2 using VOPBA and mass spectrometry. Reducing ATPSβ using both an anti- ATPSβ antibody, and siRNA knockdown led to a significant reduction in both the number of cells infected and production of infectious virus (Fongsaran et al., 2014). The ATPSβ gene is highly conserved between multiple species, but the importance of ATPSβ in human CHIKV infection has not been well characterised. What is clear is that the CHIKV target receptor appear to be possibly host specific.

Using affinity purification and mass spectrometry, a recent study revealed CD147 as a novel CHIKV receptor (De Caluwé et al., 2021, p. 147). CD147, also known as EMMPRIN or basigin, exists in two isoforms. Isoform 1 is exclusively expressed in the retina whilst isoform 2 is ubiquitously expressed and functions to induce metalloproteinases in the extracellular matrix (Zhao et al., 2013). In knockout cell lines without CD147, virus infection was significantly reduced, and

the re-introduction of this protein rescued infectivity. Interestingly, the structure of CD147 and Mxra8 are similar as CD147 also contains two immunoglobulin G-like domains. Furthermore, *in silico* analysis showed that both proteins have similar tertiary structures even though they do not share amino acid sequence homology. This led the authors to speculate that rather than targeting a specific protein to mediate host-cell entry, CHIKV may instead recognise a specific structure which could explain why the CHIKV target receptor remains enigmatic (De Caluwé et al., 2021, p. 147).

1.2.2.2 Internalisation and release of genetic material

The specific events following receptor engagement are not well understood; there is conflicting evidence on how alphaviruses uncoat and deliver genetic material to the cytoplasm of the host. It is generally accepted that alphaviruses exploit host clathrin-mediated endocytosis (CME) to enter the host cell (Kielian et al., 2010; Lee et al., 2013) but this may be host specific. In mosquito cells, CME is the only mechanism shown to be important for CHIKV entry. Transmission electron microscopy was used to reveal invaginations within the plasma membrane which resembled clathrin-coated pits. Further, treatment with receptor mediated endocytosis inhibitor and siRNA targeting clathrin prevented CHIKV entry (Lee et al., 2013). In mammalian cells, evidence for the requirement of CME for CHIKV infection is much more conflicting. In primary human umbilical vein endothelial cells, siRNA prevented CHIKV infection and demonstrates the requirement of CME (Ooi et al., 2013). Additionally, a single virus tracking experiment in monkey kidney cells showed that inhibitor-driven perturbation of CME dramatically decreased CHIKV infection (Hoornweg et al., 2016). On the other hand, siRNA against clathrin heavy chain had no effect on CHIKV infection in HEK293T cells but required epidermal growth factor receptor substrate 15 (Esp15) suggesting that CHIKV infection occurs in clathrin-independent, Esp15-dependent mechanism (Bernard et al., 2010). Interestingly, Esp15 functions in both clathrin-dependent and clathrin-independent endocytosis. There appears to be more evidence for CME requirement than a clathrin-independent entry mechanism in CHIKV infection. It is likely that CHIKV predominantly exploits the CME for infection but may be able to utilise multiple entry mechanism where necessary to establish infection in a host-specific manner.

Regardless of whether clathrin is required for initial uptake into the host cell, what is necessary for CHIKV infection is entry into the endosomal compartment. Here, a decrease in pH promotes fusion of the virus membrane with host membrane by triggering exposure and insertion of the E1 fusion loop. Alphaviruses differ in their pH fusion threshold, for example SFV fusion occurs in early endosomes with a threshold of ~ pH 6.2 whereas a specific strain of SINV requires a lower pH of ~ 5.6 which aligns with the expected pH of in the late endosome (Kielian et al., 2010). For CHIKV, endosomal fusion is another contradictive topic and is based on the requirement of Rab5/Rab7 proteins. Rab conversion from Rab5 to Rab7 is closely linked to early and late endosomal transition. Rab5 is found on early endosomes, whereas Rab7 is only present on late endosomes (Poteryaev et al., 2010). There is evidence to demonstrate that Rab7 positive endosomes are not required for CHIKV infection in mammalian cells suggesting that fusion occurs in early endosomes (Bernard et al., 2010). In mosquito cells CHIKV infection requires both Rab5 and Rab7 positive endosomes to indicate that fusion occurs in both maturing endosomes and late endosomes (Lee et al., 2013).

1.2.2.3 Genome replication

Following fusion with the host membrane, the nucleocapsid uncoats and delivers the virus genome into the cytosol for translation. Once released, the 49S genomic RNA is directly translated into P1234, the polyprotein precursor which is then proteolytically cleaved by nsP2 into P123 and nsP4 forming the short-lived early replication complex (Strauss and Strauss, 1994). Some strains of CHIKV encode an opal stop codon at the end of nsP3, and read-through generates both P123 and P1234 resulting in an altered expression level of nsP4 (Jones et al., 2017). CHIKV strains without the opal stop codon occur naturally and in vivo disruption of this stop codon does not appear to influence viral replication. It is however important for pathogenesis; mice infected with CHIKV containing a codon change to arginine in place of the opal stop codon produced decreased damage and inflammation (Jones et al., 2017). This early viral replicase, P123 and nsP4, function to synthesize full length negative sense intermediate RNA. This occurs in small, membrane associated replication compartments named spherules, which protrude out of the host cell. nsP1 is

thought to be responsible for anchoring the replication complex to these membrane compartments as it is the only non-structural protein known to directly interact with membranes. Using single particle cryo-electron microscopy, it was shown that nsP1 forms a structure within the spherule neck to gatekeep access to this structure (Jones et al., 2021). Further processing of the P123 polyprotein by nsP2 releases the remaining non-structural proteins from the polyprotein and a switch from negative to positive sense RNA production. All four mature non-structural proteins form the late replication complex to synthesise both full length genomic RNA and the 26S subgenomic RNA (Figure 1.8)(Rupp et al., 2015).

1.2.2.4 Translation of structural proteins

The 26S subgenomic RNA is translated into the structural polyprotein which is then cleaved into individual proteins. The first structural protein, C, is released in a self-cleavage event that is dependent on a serine-protease domain contained within the protein. Once cleaved, it is then responsible for nucleocapsid core formation and virion budding (Constant et al., 2021). The remaining structural polyprotein pE2-6K-E1 (sometimes referred to as p62-6K-E1, where pE2 and p62 both denote the precursor to E3 and E2) is then post-translationally processed in the endoplasmic reticulum (ER) where 6K is cleaved at both the N and C terminals to released 6K and E1 from the remaining pE2. Next, pE2 and E1 heterotrimers are formed in the Golgi compartment which is essential for protein folding. pE2-E1 heterotrimers then trimerize to form premature viral spikes. In the trans-golgi network, pE2 is then cleaved by host furin into mature E2 and E3 to form an E1/E2 heterodimer and E3 (Singh et al., 2018).

1.2.2.5 Assembly and release

The specifics of alphavirus assembly and budding are poorly characterised, but studies have suggested a model of how this occurs. In the cytoplasm, core protein C encapsidates the 49S RNA genome. This nucleocapsid core is then trafficked to plasma membrane. Meanwhile, structural proteins E1/E2 are trafficked via the secretory pathway from the trans-golgi network and deposited at the plasma membrane. Here, the nucleocapsid binds the glycoprotein spikes on the plasma membrane to form a complete virus particle which buds out of the

host cell (Jose et al., 2017). Virus budding, at least in SFV, requires the cytoplasmic tail of E2. Specifically, a single tyrosine residue was shown to be important for successful virus budding (Zhao et al., 1994). Additionally, efficient virus budding is enhanced by structural proteins 6K and TF as mutations affecting these proteins leads to reduced virus infectivity and production (Ramsey and Mukhopadhyay, 2017).

1.2.3 Non-structural protein 3 overview

The alphavirus nsP3 has been shown to be an important member of the viral RNA replicase, but the specific function of this protein has not yet been fully characterised.

Originally, nsP3 was considered a stable protein throughout the virus lifecycle as it can usually be found in complex with other viral or host proteins. However, evidence has emerged to demonstrate that in SFV and SINV, a C-terminal degradation signal causes rapid degradation of nsP3 during early stages of infection (Varjak et al., 2010). The function of this temporal regulation is unclear, but there are suggestions that this phenomenon could be responsible for the regulation of nsP4, the viral RdRp, as a mutant SFV carrying two copies of nsP3 produced increased levels of nsP4 (Saul et al., 2015).

The localisation of nsP3 during infection also appears enigmatic. In early infection, nsP3 can be found at the cytoplasmic surface of the cell where the replication complex is associated with the plasma membrane. Later, in some alphaviruses, nsP3 can be found in the cytopathic vacuoles (CPV) formed in the perinuclear area. Aside from those found in complex with non-structural proteins, a portion of mature nsP3 form large aggregates (Cristea et al., 2006; Froshauer et al., 1988; Kujala et al., 2001). The ratio of replication complex associated nsP3 and aggregated nsP3 differs between alphaviruses. Whilst the purpose of the nsP3 aggregates is unclear, it suggests that nsP3 has functions in the lifecycle independent of its canonical requirement in the replication complex.

nsP3 has also been implicated as a determinant of mosquito vector specificity, particularly in the old-world alphaviruses, CHIKV and ONNV. ONNV is the only

alphavirus that is exclusively transmitted by anopheline mosquitoes whilst CHIKV is predominantly transmitted by *Aedes* mosquitoes. Chimeric CHIKV expressing ONNV nsP3 can successfully infect *Anopheles gambiae* mosquitoes. Interestingly, chimeric ONNV expressing CHIKV nsP3 cannot establish infection in mammalian or insect cells, suggesting that the function of nsP3 in these viruses is distinct (Saxton-Shaw et al., 2013).

Neurological complications as a result of CHIKV infection are discussed in 1.1.5. The neurovirulence of some old-world alphaviruses has been attributed to nsP3. Replacement of nsP3 in an avirulent SFV strain with the nsP3 of a neurovirulent strain led to lethal neurovirulence in mice (Tuittila et al., 2000). For SINV, both the integrity of the nsP3 macrodomain and the hypervariable domain appear important for neurovirulence in mice. Mutations targeting the ADP-ribose binding and hydrolase functions of the macrodomain severely impairs the ability of the virus to replicate in neuronal cells and in the central nervous system of mice (Abraham et al., 2018). Furthermore, an 18-amino acid deletion in the C-terminal domain of nsP3 was found to be a crucial determinant of neurovirulence in adult mouse (Suthar et al., 2005). The importance of nsP3 in neurovirulence appear to be specific to old world alphaviruses. For new world viruses, structural protein E2 has been implicated (Atkins and Sheahan, 2016).

1.2.3.1 Structural Organisation of nsP3

nsP3 is a ~60 kDa protein that is organised into three distinct domains: the macrodomain, an alphavirus unique domain, and a hypervariable domain, as shown in Figure 1.10.

Figure 1.10 Schematic of the CHIKV nsP3 highlighting the three domains: the macrodomain at the N-terminal, the Alphavirus Unique Domain (AUD) in the centre, and the Hypervariable domain at the C-terminal.

1.2.3.2 Macrodomain

Macrodomains are found in proteins from all domains of life, and in some viruses. In both its free or protein-linked form, macrodomains are defined as proteins that bind and hydrolyse ADP ribose (Rack et al., 2016). In nsP3, it is located at the N-terminus, and is highly conserved amongst alphaviruses. Structurally, the macrodomain is organised into a six-stranded central β-sheet surrounded by 3 α-helices on one side, and one α-helix on the other as shown in Figure 1.11. This structural arrangement, and correct folding is crucial for virus infection as it forms a binding pocket that houses ADP-ribose located at the top between beta strands 2, 3 and 4 (Malet et al., 2009). The CHIKV macrodomain has been shown to bind to mono-ADP-ribose and function as a mono-ADP-ribosylhydrolase, and as an antagonist of multiple host Type I mono-ADP-ribosyltransferases (ARTDs) including ARTD 8, 10 and 12 (Eckei et al., 2017; Krieg et al., 2020). Type I ARTDs are induced by Type I interferons or pathogen associated molecular patterns. Such enzymes confer antiviral activity by post-translationally modifying targets, in a process called mono-ADP-ribosylation (MARylation) presumably as a signal to the immune system. ARTD10 and ARTD12 have been identified as restriction factors of CHIKV replication. Specifically, these enzymes perform MARylation on nsP2 to inhibit the proteolytic activity of this protein. To ensure correct processing of the CHIKV non-structural polyprotein, nsP3 is able to hydrolyse and remove mono-ADP-ribose modifications on nsP2 (Krieg et al., 2020). Whilst further work is required to completely resolve the relationship between the macrodomain and the host immune response, the control of mono-ADP-ribosylation evidently plays an important part in virus-host interactions. Furthermore, the macrodomain has also been described as a structural recognition site for nsP2 processing of nsP2/3 in the CHIKV non-structural polyprotein (Lulla et al., 2012).
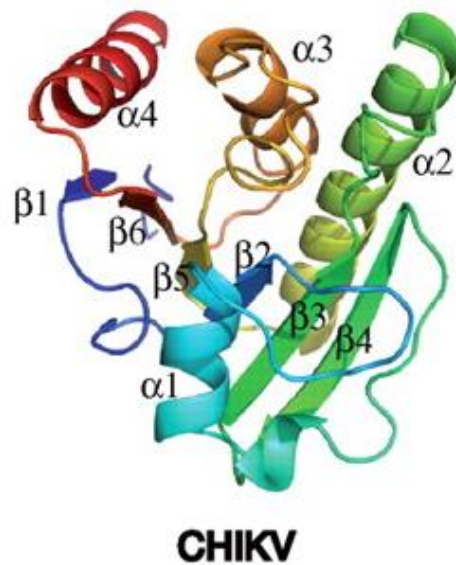
Figure 1.11 Structure of the CHIKV macrodomain in a purple to red gradient from the N-terminus to the C-terminus. Alpha helix and Beta sheet secondary elements are labelled. Figure obtained from Malet et al., 2009.

1.2.3.3 Alphavirus Unique Domain

The alphavirus unique domain (AUD) is located in the centre of nsP3 and is maintained only in alphaviruses. This domain is rich in serine and threonine residues, forming two parallel β-sheets and antiparallel α-helices. Four highly conserved cysteines in the AUD are responsible for binding a structural zinc ion. The importance of the zinc-binding capability of the AUD has led to this domain being named as the zinc binding domain. The integrity of all 4 cysteines is essential for virus replication, and the mutation of any results in failed expression of the non-structural polyprotein, indicating an importance of the AUD in the early stages of the virus lifecycle (Shin et al., 2012). Further to the requirement of intact cysteine residues, other surface exposed amino acids are also important for virus replication; single or double mutations affecting such amino acids leads to reduced replication in a species and cell-type specific manner. One double mutation, P247A/V248A, displayed severely reduced virus production, which was attributed to an inability to transcribe the subgenomic RNA thus reducing expression of the viral capsid protein required for virus assembly (Gao et al., 2019).

In addition to viral interactions, AUD-host interactions have been elucidated using a yeast two hybrid system to screen a human cDNA library. Identified protein partners, SNAPIN and N4BP2L2, were then further confirmed using a GST-pull down assay. Using these two binding partners to build an interactome network revealed other human proteins that play major roles in vesicle transport and transcription, suggesting the ability of AUD to access multiple cellular functions during infection (Ghildiyal and Gabrani, 2021).

1.2.3.4 Hypervariable Domain

The hypervariable domain (HVD), or hypervariable region is an intrinsically unstructured and disordered region. Unlike the macrodomain and AUD, the HVD is not well conserved amongst alphaviruses where the sequence and length varies greatly between viruses. However, some key features of this domain exist in multiple members of the family, including a highly phosphorylated region, a polyproline motif and repeated elements containing FGDF motifs in the C-terminal as shown in (Figure 1.12) (Götte et al., 2018). A fully intact HVD is required for virus replication, but in some alphaviruses deletion or mutation in this domain can be tolerated. In SINV, smaller deletions in the HVD resulted in virus yields that were similar to parental virus but larger deletion or duplications above 95 amino acids led to consistently lower virus yields during early stages of infection (4 hours). By 8 hours, all mutations produced yields that were indistinguishable from the wildtype virus (Lastarza et al., 1994). In SFV, HVD deletions of 129 and 325 bps led to reduced RNA synthesis and multiplication in cell culture, and reduced virulence in mice (Galbraith et al., 2006). However, targeted insertions within the HVD can be used as a molecular tool for studying nsP3 during infection; a GFP insertion between amino acids 388 and 389 in SINV (Liang and Li, 2012), SNAP-tag or mCherry insertions between amino acids 382 and 383 (Remenyi et al., 2017), and a Twin-Strep tag insertion between amino acids 349/350 and 369/370 in CHIKV (Gao et al., 2019) are all well-tolerated.
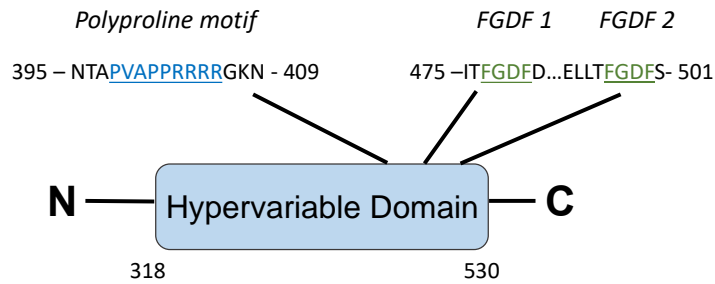
Figure 1.12 Conserved features in the nsP3 Hypervariable domain showing the polyprolife motif (in blue) and two FGDF repeats (in green). Figure adapted from Göertz et al., 2018.

Conserved motifs in the HVD are important for replication and interactions with mammalian and mosquito host proteins, which is discussed in section 1.2.3.5. Furthermore, the requirement of the HVD for transmission in mosquito has also been described. Specifically, mutations impacting either of the FGDF motifs led to significantly lower infection rate compared to wild-type virus in *Aedes aegypti* mosquitoes. Viral load in the mosquito saliva was also reduced when infected with the single FGDF mutation virus, indicating that both FGDF motifs must be intact for optimal mosquito transmission (Göertz et al., 2018).

1.2.3.5 nsP3 and host protein interactions

To establish an optimal environment for replication and production of progeny virus particles, viruses must manipulate, exploit, and divert cellular processes and immune defences in a regulated manner. For this, viruses have evolved multiple mechanisms to target host proteins. Multiple nsP3 protein interactors have been characterised that are important in the virus lifecycle including stress granule related proteins, proteins that function in endocytosis, cell signalling and formation of subcellular structures.

1.2.3.6 G3BP/Rasputin

G3BP1/2a/2b are RNA-binding proteins that function in the organisation of stress granules (SG). During cellular stresses such as heat or cold shock, oxidative stress or virus infection, translation is stalled leading to an accumulation of mRNP and stress protein accumulation in the cytoplasm (Tourrière et al., 2003). This signals the formation of SGs, which functions to either rescue translation or degrade and remove the contained mRNAs (Buchan and Parker, 2009). Alphaviruses require the host translational machinery for initial translation and as such, must antagonise the formation of bona fide SGs as a response to virus infection. The interaction between alphaviruses and G3BP1/2 was first identified in SINV using immunoprecipitation and mass spectrometry but has since been confirmed in multiple studies (Fros et al., 2012; Panas et al., 2015). To visualise and track the trajectory of nsP3 during infection, a GFP-nsP3 fusion SINV was used. Unlike interactions with other host proteins, G3BP was persistently recruited during infection, confirming the importance of this protein for successful infection (Cristea et al., 2006).

The localisation of nsP3:G3BP throughout virus infection appears analogous between alphaviruses. Under normal cellular function, G3BP is distributed throughout the cell but following alphavirus infection G3BP is localised to nsP3 containing replication complexes (Frolova et al., 2006; Panas et al., 2012). Interestingly, in early CHIKV infection, G3BP localises to CHIKV nsP3 containing foci to form SG-like complexes that are compositionally different to normal SGs (Fros et al., 2012). In the same study, G3BP1 and G3BP2 depletion via siRNA knockdown was shown to severely reduce levels of negative-strand RNA. Subsequently, levels of positive-strand RNA, proteins and progeny infectious virus were all decreased (Scholte et al., 2015). In contradiction, silencing of G3BP1 and G3BP2 in SINV led to an enhancement in early replication (Cristea et al., 2010).

The interaction between G3BP and multiple old-world alphaviruses have been mapped to the HVD of nsP3. Specifically, the FGDF motifs binds to the NTF2-like domain of G3BP (Panas et al., 2014). Crystal structures of SFV nsP3 in

complex with G3BP revealed that each FGDF motif binds to a G3BP monomer on separate dimers to connect multiple G3BP dimers and form a G3BP:nsP3 polycomplex (Schulte et al., 2016). Whilst the FGDF: NTF2 interactions has been extensively studied, recent evidence shines light on the importance on the RGG motif towards the C-terminus of G3BP, for the specific recruitment of host translational initiation machinery. In CHIKV, but not SFV, replication and formation of replication complexes were only possible in cells expressing G3BP variants that contain both the NTF-2 and the RGG domains. This suggests that the proviral effect of G3BP1 is equally dependent on both these domains. The RGG was also shown to be important for the association of nsP3:G3BP with the 40S ribosomal subunits proteins rpS3 and rpS6 to enhance translation at virus induced spherules, and internal CPVs. Furthermore, during SFV infection, nsP3:G3BP:40S complexes contain more ribosomes in active polysomes than in cells where formation of this complex is prevented (Götte et al., 2019).

Whilst the physical interaction between nsP3 and G3BP has been mapped to the HVD, recent evidence has revealed that the macrodomain plays a role in regulating this interaction. This study showed that the hydrolase activity of nsP3 reduces G3BP1 MARylation to prevent SG formation and disassemble virus induced SGs. Furthermore, expression of wildtype nsP3 induces the formation of condensates that do not contain translation initiation factors whereas the expression of nsP3 lacking the ability to hydrolyse ADP-ribose leads to the formation of condensates that retain the translation initiation factors. This shows that during alphavirus infection, the hydrolase activity of nsP3 is important to regulate the formation and composition of biomolecular condensates (Jayabalan et al., 2021). Further work is required to completely elucidate the dynamics of G3BP in the alphavirus lifecycle, but manipulation of this protein has a clear multifaceted benefit for the virus.

The nsP3:G3BP interaction is also conserved in the mosquito vector. The homologue of G3BP, Rasputin, localises to nsP3 foci during CHIKV infection in a granular manner mediated by the FGDF motifs in the HVD of viral protein and NTF-2 of G3BP. Silencing of Rasputin using RNAi leads to a reduction in viral production in *in vivo*, but not *in vitro* infection. However, the exact mechanisms

by which G3BP/Rasputin contribute to CHIKV replication remains undetermined (Fros et al., 2015).

1.2.3.7 Heat shock proteins (Hsp)

Heat shock proteins are a family of highly conserved proteins that function in a range of cellular processes including assembly of protein complexes, cell-cycle control and protection of cells against stress or apoptosis. In the alphavirus lifecycle, specifically through interactions with nsP3, the Hsp70 and Hsp90 family have been implicated.

The Hsp70 family consists predominantly of two proteins: Hsc70 and Hsp70. Hsc70 functions mainly in vesicle uncoating and transportation of proteins to various organelles. Similarly, Hsp70 also functions in these processes, but is exclusively expressed under cellular stress. In both mammalian and mosquito cells infected with SINV, Hsc70 was identified in complex with SINV nsP3 (Gorchakov et al., 2008), but the functional relevance of this in mammalian cells has yet been determined. On the other hand, Hsc70 was later identified as a putative CHIKV receptor in *Aedes albopictus* cells using VOPBA coupled with mass spectrometry and confirmed using an anti-Hsc70 antibody. Cells pre-incubated with anti-Hsc70 had a significant reduction in infectious virus production (Ghosh et al., 2017). Multiple virus interactions with Hsp70 proteins have been described but interactions between alphaviruses and this cellular protein are less explored. HIV interacts with both Hsc70 and Hsp70 to transport viral polyproteins to sites of virus assembly (Gurer et al., 2002) whilst Zika virus exploits Hsp70 as an infection factor to mediate entry, replication and egress (Pujhari et al., 2019). An interaction between alphavirus nsP3 and Hsp70 for virus entry has not been described but transporting viral proteins to site of virus assembly could be plausible.

Hsp90 functions in the folding and activation of multiple signalling proteins such as src tyrosine kinases and Raf kinases which are heavily involved in cell proliferation (Verma et al., 2016). This protein is a cellular target of nsP3, and their interaction is important for viral replication. Hsp90ß (the ER predominant,

isomeric form of Hsp90) was shown to colocalise and co-immunoprecipitate with nsP3 in immunofluorescence and GFP-pull downs, respectively. Furthermore, the inhibition of Hsp90 with both synthetic Hsp90 inhibitors and Hsp90 specific siRNA reduced CHIKV replication *in vitro* and *in vivo*. Additionally, inhibiting Hsp90 reduces CHIKV induced inflammation in the limbs of mice (Rathore et al., 2014). A follow-up study using the Hsp90 inhibitor geldanamycin suggests that Hsp90 contribute to CHIKV replication by stabilising the nsP2 protein (Das et al., 2014).

1.2.3.8 Amphiphysin (BIN-1)

Amphiphysins are a family of proteins involved in cellular processes such as membrane trafficking and endocytosis. Interactions between amphiphysins and other cellular ligands are canonically mediated by SH3 domains in amphiphysin and proline rich sequences in binding partners. Alphavirus nsP3 and amphiphysin interactions were first identified using an SH3 bacteriophage display (Neuvonen et al., 2011). As highlighted in section 1.2.3.4, the overall sequence of the nsP3 HVD is poorly conserved. Despite this, several features of this domain are shared amongst multiple alphaviruses, one of which is the polyproline motif. Class II Polyproline motifs follow the consensus sequence PxxPx+ where + denotes a positively charged amino acid (Kay et al., 2000). This motif in nsP3 was found to be the target site for the Src-homology 3 (SH3) domain of amphiphysin-1/amphiphysin-2. In SINV, both mutations of the first proline and a charge reversing mutation targeting the arginine residue abrogated binding to amphiphysin-1. The distribution of amphiphysin was also investigated; in cells infected with SFV and SINV, amphiphysin was found to be recruited to sites of virus replication, gathering at virus infected CPVs. RNAi or mutational disruption of the nsP3:amphiphysin interaction in SINV and SFV led to impaired viral replication in *in vitro* experiments whilst infection of Balb/c mice with SFV with target polyproline motif deletions was associated with reduced mortality (Neuvonen et al., 2011).

Later, studies using isothermal titration calorimetry (ITC) revealed that CHIKV nsP3 has a higher affinity for the amphiphysin SH3 domain than dynamin, the natural cellular binding partner of amphiphysin. Using the CHIKV nsP3:amphiphysin NMR solution complex structure coupled with ITC data, it was revealed that the high affinity binding originates from the electrostatic interactions between the negatively charged binding surface of the amphiphysin SH3 domain and the positively charged residues at the end of the nsP3 polyproline motif (Figure 1.13). As such, the authors proposed that an extended class II polyproline motif 'PxxPxRPXR' is instead required to mediate the ultra-high affinity binding to amphiphysin, perhaps providing an advantage to outcompete dynamin (Tossavainen et al., 2016). Whilst the role of CHIKV nsP3: amphiphysin in the virus lifecycle is poorly understood, it has been speculated that as amphiphysin can induce membrane curvature, such interactions may be important in the formation of spherules observed during infection. Furthermore, over 300 human proteins contain SH3 domains, with functions in a multitude of cellular processes. The ability of alphaviruses to control this group of proteins via the HVD polyproline motif is certainly advantageous.
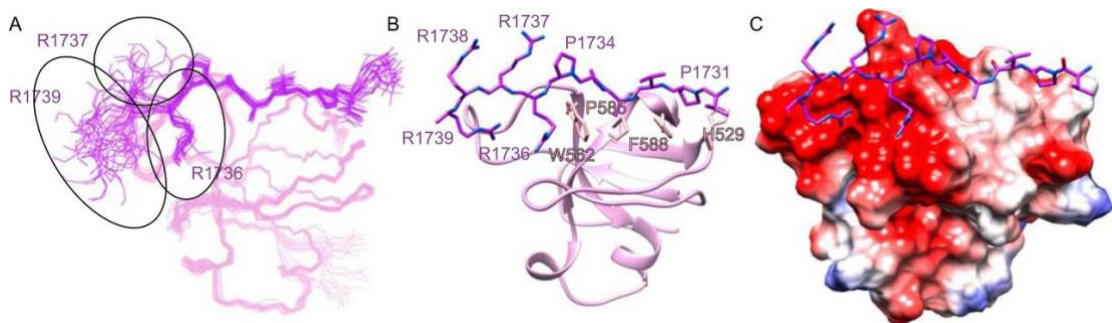
Figure 1.13 Solution structure of the Amphiphysin SH3 domain in complex with the CHIKV nsP3 HVD polyproline motif. A) Interaction between Amphiphysin SH3 with the polyproline motif with the positively charged arginine residues highlighted. B) Interaction between Amphiphysin SH3 represented as a ribbon with the polyproline motif shown as heavy atoms. C) Coulombic surface colouring of amphiphysin SH3 highlighting the large area of negatively charged residues and the positively charged residues of nsP3 HVD polyproline motif interacting with it. Figure obtained from Tossavainen et al., 2016.

1.2.3.9 RNA interference and DEAD-box RNA helicases

RNA helicases of the DEAD-box family are a group of highly conserved enzymes and are important players in RNA metabolism. DEAD-box helicases regulate gene expression via the control of processes including transcription and RNA degradation (Linder and Jankowsky, 2011). A secondary, more recent established role for this group of enzymes has been identified, as components of the RNA interference (RNAi) pathway. Canonical RNAi is induced by double stranded RNA (dsRNA) molecules and leads to sequence-specific RNA degradation. Under normal cellular function, this pathway is used to regulate gene expression post-transcriptionally. Many viruses produce dsRNA during replication, triggering the RNAi pathway. As such, RNAi can also function in immunity as a viral restriction factor. The relationship between RNAi and alphaviruses was first investigated in SINV expressing a viral RNAi suppressor protein B2 from the insect-pathogenic Flock House Virus (FHV). Expression of B2 led to an increase in replication and mortality in *Aedes aegypti* mosquitoes (Cirimotich et al., 2009). As RNAi suppression is clearly advantageous in the alphavirus lifecycle, later studies sought to identify the viral proteins responsible for interactions with this pathway. Systematic analysis using a RNAi sensor cell line identified CHIKV nsP2 and nsP3 as viral suppressors of RNAi. Subsequent domain mapping and RNA binding motif analyses identified specific motifs in the nsP3 macrodomain that exert RNAi suppression (Mathur et al., 2016).

Due to the pleiotropic nature of DEAD-box helicases and specific implication of DEAD-box helicase 9 (DHX9) in the lifecycles of multiple viruses, the function of this protein in alphavirus infection of human cells was explored. Nuclear expression of DHX9 was found to be reduced in CHIKV-infected HeLa cells and targeted to foci that also contain nsP3 and dsRNA. Immunoprecipitation then confirmed DHX9 in complex with nsP3 and CHIKV genomic RNA, mediated by the nsP3 HVD. Knockdown of DHX9 with siRNA or CRISP-Cas9 significantly reduced expression of non-structural proteins in early stages of infection but conversely enhanced RNA synthesis during later stages. This suggests a proviral role early on in infection but that it may have some detrimental effects later (Matkovic et al., 2019). Whilst the role of DHX9 in the alphavirus lifecycle appear enigmatic, DEAD-box RNA helicase 56, DDX56, was found to be

exclusively antiviral in SINV and CHIKV infection. DDX56 can bind incoming virus RNA, and negatively impact CHIKV genome stability. Depletion of DDX56 enhanced SINV and CHIKV infection leading to increased expression of non-structural proteins, although the viral protein responsible for interactions with DDX56 is unclear (Taschuk et al., 2020).

In mosquitoes, CHIKV infection also affects the RNAi pathway with specific factors such as Dicer 2 being upregulated, whilst TSN and Ago-2 are downregulated. Co-immunoprecipitation with an emphasis on the RNAi pathway revealed an interaction between nsP3 and the mosquito DEAD-box helicase RM62F. Expression profiling of RM62F transcript showed that this protein was temporally regulated during virus infection (Kumar et al., 2021).

1.2.3.10 nsP3 and RNA interactions

As a member of the replication complex, it is plausible to hypothesise that nsP3 binds RNA and that these interactions are important in the virus lifecycle. However, limited viral and host RNA interactors have been identified. Studies have confirmed the ability of nsP3 to bind RNA oligonucleotides, double stranded RNA (dsRNA) and small interfering RNA (siRNA) (Malet et al., 2009). As discussed in 1.2.3.9, the nsP3 macrodomain binds RNA to function in RNA interference, and the abolishment of RNA binding motifs in this domain reduces its ability to bind dsRNA and therefore, exhibit RNAi activities (Mathur et al., 2016). The nsP3 AUD also has RNA binding properties; a specific double mutation (P247A/V248A) in this domain reduces the ability of this protein to bind the CHIKV negative strand sub-genomic promoter in an *in vitro* RNA filter binding assay (Gao et al., 2019). RNA binding motifs of nsP3 may be more important in determining nsP3-RNA interactions rather than the specific RNA targets as the CHIKV nsP3 macrodomain is able to bind a wide variety of oligonucleotides of unrelated sequences in *in vitro* RNA slot blotting experiments (Malet et al., 2009). Nevertheless, *in vitro* studies do not reflect a true virus infection and whether nsP3 binds RNA specifically or otherwise requires further investigation. Elucidating such interactions may provide crucial information on the role of nsP3 in viral replication and pathogenicity.

1.2.4 *In vivo* identification of protein-RNA interactions

RNA-binding proteins (RBPs) are key regulators of post-transcriptional gene expression. As such it would be advantageous for viruses to possess RNA-binding abilities in order to manipulate host cellular environment. To understand the function and fate of viral and host RNA molecules, it is important to map protein-RNA interactions to unravel their influence in the virus lifecycle.

1.2.4.1 RNA Immunoprecipitation & Cross-Linking Immunoprecipitation

The first approaches to investigate protein-RNA complexes combined RNA affinity purification and immunoprecipitation with various screening techniques including RT-PCR, microarray analysis or library screening (Ule et al., 2005). Unfortunately, these methods produced low-stringency RNA-RNP binding thus were prone to identifying indirect interactions that are not physiologically relevant. Furthermore, these approaches produced low resolution data, leaving the specific RNA-RBP binding sites unresolved. Therefore, developing techniques which can reduce false positives and resolve specific RBP binding sites have since been the primary challenges in experimental ribonomics (König et al., 2012).

The development of *in vivo* UV-crosslinking and immunoprecipitation (CLIP) enabled the identification of precise protein-RNA binding sites with high positional resolution and specificity (Ule et al., 2005). In this method, *in vivo* protein-RNA interactions are preserved by ultraviolet C (UVC) treatment. This induces formation of irreversible covalent cross-linking at sites of direct contact between RNA and proteins. Cells are then lysed and subjected to partial RNase digest to generate fragments of approximately 30- 50 nucleotides (König et al., 2012; Moore et al., 2014). RNA-protein complexes are then immunoprecipitated with an antibody specific for the protein of interest followed by 3' RNA de-phosphorylation using alkaline phosphatase and ligation of a 3' RNA adaptor. Treatment with Polynucleotide kinase K (PNK) is then carried out to phosphorylate and radioactively label the 5' end to allow visualisation of protein-RNA complexes in later steps. SDS-PAGE and transfer to nitrocellulose membrane enable visualisation of RNA-digestion results and isolation of RNA-

tags within an ideal size range. Additionally, SDS-PAGE separates protein-RNA complexes from any non-covalently associated contaminant RNA (König et al., 2012; Moore et al., 2014). Radioactive bands corresponding in size to protein-RNA complexes are then excised from the membrane and treated with Proteinase K to digest the RBP, leaving a polypeptide at the cross-linked residue. An adaptor is ligated to the 5' end and the RNA is reverse transcribed. The resulting cDNAs are amplified by PCR with primers that are complementary to the 5' and 3' adaptor regions, followed by data processing and analysis (König et al., 2012; Moore et al., 2014).
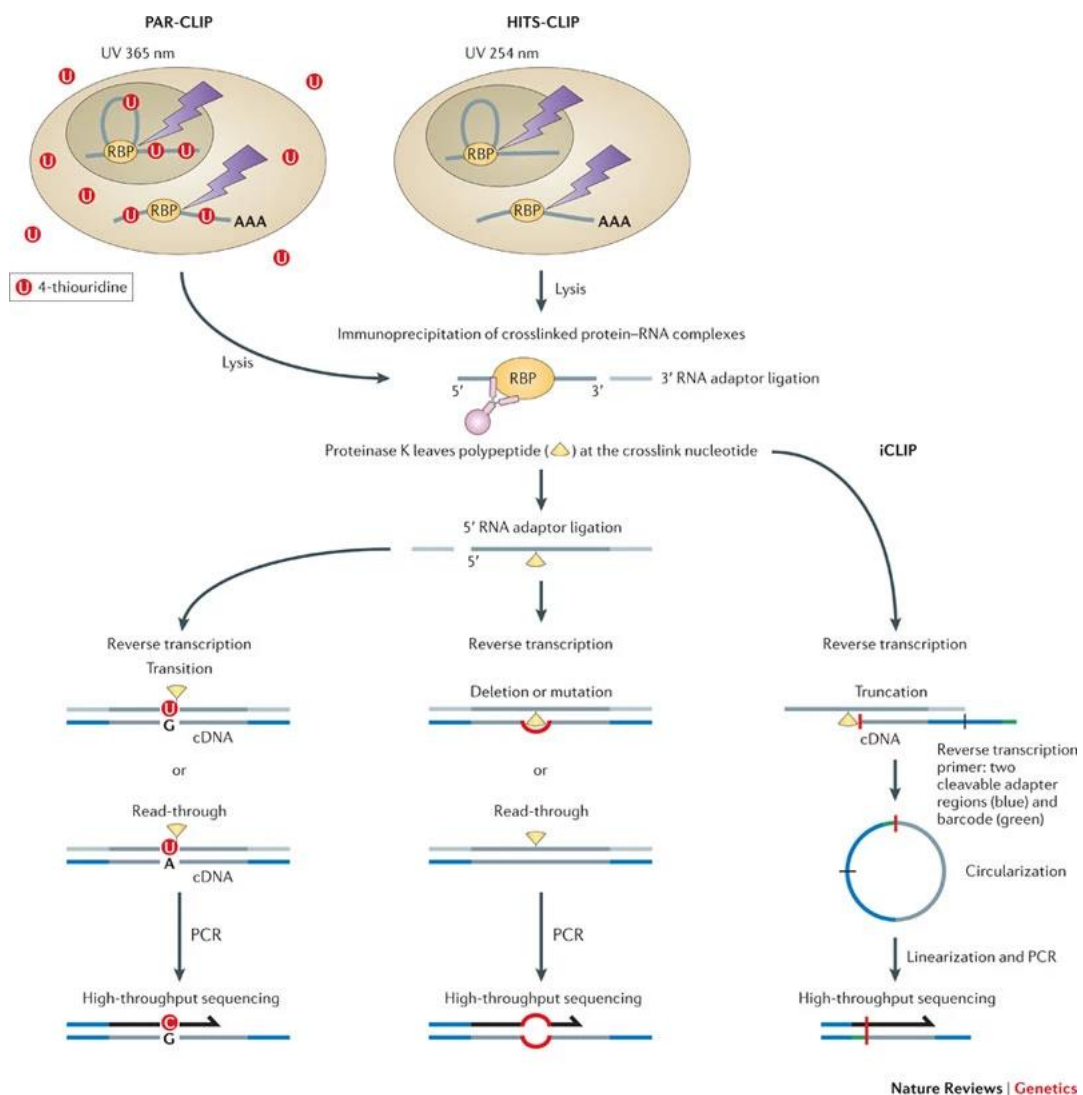
Figure 1.14 Overview of PAR-CLIP, HITS-CLIP and iCLIP. Figure obtained from König et al., 2012.

1.2.4.2 High throughput sequencing-CLIP (or CLIP-Seq)

In high-throughput sequencing of RNA isolated by ultraviolet crosslinking and immunoprecipitation (HITS-CLIP) (Figure 1.14 HITS-CLIP). The cDNA library generated from reverse transcription is subjected to high-throughput next generation sequencing. The first round of PCR product is converted into an Illumina-compatible sequencing library followed by deep sequencing (Lee et al., 2017).

In this method, the exact protein-RNA crosslinked sites can be mapped at single-nucleotide resolution by exploiting crosslink-induced mutation sites (CIMS). Treatment of RNA-protein complexes with Proteinase K removes the bound RBP, however, due to the irreversible nature of UV-crosslinking, the exact crosslinked amino acid is left behind. During reverse transcription, two outcomes can occur: reverse transcriptase can either disregard the polypeptide to generate an unaltered cDNA copy or the residual amino-acid-RNA adduct can poses a physical obstacle for the enzyme to read through during cDNA synthesis, generating a cDNA copy with a deletion or mutation at the site of crosslink. Such altered cDNA copies (CIMS) can be identified by computationally searching and mapping deep sequence reads to the reference genome (Zhang and Darnell, 2011).

HITS-CLIP & CIMS has been exploited to map RNA-interactions of Argonaute, a protein that forms complexes with miRNA to facilitate downstream gene-silencing (Moore et al., 2014). More recently, it has also been used to investigate the association of influenza virus RNA and nucleoprotein (NP) where it was discovered that contrary to interactions described in literature, viral RNA does not bind NP uniformly, but rather exhibit binding profiles that are unique to the specific vRNA segment (Lee et al., 2017).

1.2.4.3 Photoactivatable ribonucleoside-CLIP

Compared to other CLIP approaches, the defining property of photoactivatable-ribonucleoside- enhanced crosslinking and immunoprecipitation (PAR-CLIP) is that photoactivatable nucleoside analogues such as 4-thiouridine (4-SU) or 6-

thioguanosine (6-SG) are used (Figure 1.14 PAR-CLIP). This allows UV cross linking to be performed with UVA/UVB light at a wavelength above 310 nm rather than UVC light at the wavelength of 254 nm. In this method, photoactivatable nucleotide analogues are added directly to cell culture medium around 16 hours prior to cross-linking. The nucleotide analogues are readily taken up by cells and incorporated into newly synthesized transcripts (Danan et al., 2016). Standard downstream CLIP and high-throughput sequencing is then performed.

A key feature of PAR-CLIP is that during reverse transcription, characteristic mutations are generated; T-to-C for 4-SU and G-to-A for 6SG. Such mutations allow resolution of protein-RNA binding sites at single nucleotide. More importantly, it allows computational removal of background RNAs that could generate a false positive signal (Danan et al., 2016). Furthermore, crosslinking of RNA supplemented with photoactivatable nucleoside analogues is comparatively more efficient than unmodified RNA. Another advantage of PAR-CLIP is that crosslinked 4SU favours a U-G pairing instead of the canonical U-A pairing during reverse transcription, so cross-link sites can be clearly labelled (Haecker and Renne, 2014). However, as 4-SU is toxic to cells, effort is required for dose optimisation to determine the maximum dose that can be tolerated. Additionally, the introduction of a single nucleoside analogue can cause nucleotide bias and thus skew the validity of characterised RNA sites (Moore et al., 2014).

1.2.4.4 Individual nucleotide resolution UV-CLIP
The major difference between Individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) and other CLIP approaches is that nucleotide resolution is achieved by circularisation of truncated cDNA (Figure 1.14 iCLIP). In over 80% of CLIP sequencing, any truncated cDNA that are produced from reverse transcriptase stalling lack the 5' adapter required for PCR amplification and would be lost during standard CLIP library preparation. To capture the truncated cDNAs, iCLIP uses alternative adaptor ligation; traditionally, both the 3' and 5' end adaptors are ligated to RNA before reverse transcription whereas

in iCLIP, the 5' adaptor is added post reverse transcription (Huppertz et al., 2014).

For initiation of reverse transcription, a primer containing two cleavable adapter regions and a barcode is used. Target cDNA products are then size selected using gel electrophoresis and the 5' adapter is added followed by circularisation. PCR amplification is then initiated with a primer that anneals to the 5' adapter, yielding a linear product with adapters at both the 5' and 3' end. PCR products are then subjected to high-throughput sequencing, generating reads in which the barcode sequences are immediately followed by the last nucleotide of the truncated cDNA (Huppertz et al., 2014). Compared to HITS-CLIP and PAR-CLIP, iCLIP has enhanced efficiency of target capture and precise mapping of cross-linked sites defined by preservation of truncated cDNA.

1.2.4.5 Redesigning iCLIP

Although current CLIP techniques have facilitated single nucleotide resolution of RNA-RBP interactions, several technical challenges remain. Firstly, for visualisation of UV-crosslinked-protein-RNA complexes, radioisotopes are required. Secondly, labelling of protein-RNA complexes usually occurs on the 5' end of crosslinked RNA, which does not give information about the success of 3' adaptor ligation. Furthermore, radioactive reagents decay, resulting in non-uniform autoradiography signal across experiments (Zarnegar et al., 2016). Further development and redesign of iCLIP have produced two novel variations with 1,000-fold improved efficiency by combining the advantages of iCLIP with optimization at multiple steps (Martin and Zavolan, 2016).
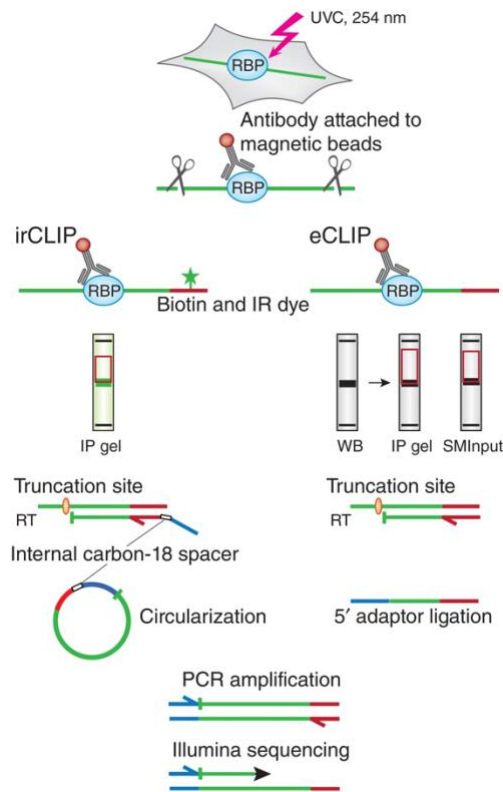
Figure 1.15 Improved CLIP techniques. In infrared-CLIP (irCLIP), an infrared adapter is used to facilitate efficient visualization of correctly sized protein-RNA complexes, downstream workflow mimics that observed in iCLIP. In enhanced-CLIP, size selection is achieved by parallel western blot analysis and capture of truncated cDNA is achieved by ligation of a 5' DNA adapter following reverse transcription. Figure obtained from Martin and Zavolan, 2016.

1.2.4.6 Infrared-CLIP

In infrared-CLIP (irCLIP), a biotinylated and infrared dye conjugated adapter is used to facilitate quantitative and qualitative analysis of *in vivo* captured protein-RNA interactions (Figure 1.15 irCLIP). Compared to the standard adapter, the irCLIP variant demonstrates the same efficiency in ligation reactions, removes the safety issues associated with radioisotope labelling whilst reducing the visualisation time by more than 10-to-100-fold. Additionally, irCLIP uses on-bead nuclease digestion instead of in-lysate digestion as this was found to improve the production of optimal RNA fragments. In iCLIP, RNA-RBP complexes are liberated from nitrocellulose membrane by proteinase K digestion in the presence of urea. However, during optimisation of irCLIP it was found that the use of urea increases the dielectric constant of the surrounding medium. This can inhibit nucleic acid precipitation, resulting in a >80% loss of material.

Replacing urea with SDS significantly improves precipitation of peptide-RNA complexes compared to urea-PK digestion (Zarnegar et al., 2016).

Lastly, reverse transcriptase SuperScript III traditionally used in iCLIP is replaced with thermostable TGIRT-III, allowing reverse transcription to take place at 60°c, followed by cDNA circularization at the same temperature in a one-tube reaction. Using a high temperature removes RNA secondary structures, therefore reducing biases in both reactions. The use of a one-tube reactions also minimises loss of cDNA (Martin and Zavolan, 2016).

1.2.4.7 Enhanced CLIP

Whilst irCLIP exploits cDNA circularization to minimise loss of truncated cDNA, enhanced CLIP (eCLIP) achieves this by adaptor addition in two separate steps: a barcoded 3' RNA adaptor is ligated to crosslinked RNA whist the complex is on immunoprecipitation beads, and a 5' single stranded DNA adapter containing a randomer of 5 or 10 bases is added after reverse transcription (Figure 1.15 eCLIP) (Van Nostrand et al., 2016). Sample-specific barcodes in the 3' adapters allow sample pooling and parallel processing, as samples can be computationally separated following sequencing, greatly increasing the throughput (Martin and Zavolan, 2016). Like irCLIP, eCLIP also omits the need for radioactively labelling of RNA for identification of RNA- protein complexes. Instead, complexes are selected using antibodies on western blots (Van Nostrand et al., 2016).

1.3 Aims and Objectives

The overall goal of this project was to functionally characterise CHIKV nsP3 in the virus lifecycle.

The first aim of this project focused on interrogating the importance of a polyproline motif present in the hypervariable domain of nsP3. For this, mutagenic studies were performed based on sequence alignment with related viruses. Replication specific investigations were carried out using a luciferase-reporter system then further explored in the context of complete virus infection. From this data, further analysis was conducted to identify a critical residue outside of the polyproline motif which was required for replication and production of infectious virus.

Secondly, the RNA binding ability of nsP3 was interrogated. Previous studies have focused on *in vitro* RNA binding properties of nsP3, and little is known about the specific viral and host RNAs targeted by this protein. The first aim of this project was to elucidate nsP3-RNA binding in an *in vivo* model. To achieve this, individual nucleotide UV cross-linking and immunoprecipitation was established and applied to a CHIKV infection model. Subsequent bioinformatics analysis was used to interrogate sequencing product and identify RNA binding sites in both the virus and the host.

Chapter 2 Materials and Methods

2.1 Molecular Biology

2.1.1 Manipulation of nucleic acid

2.1.1.1 Plasmids and virus constructs

DNA constructs of CHIKV sub-genomic replicon with a single luciferase reporter (CHIKV-FLuc-SGR) and full length ICRES ECSA CHIKV were kindly gifted by Professor Andres Merits, University of Tartu. Both replicon and virus constructs are derived from the ECSA genotype LR2006 OPY1 isolate virus. CHIKV-FLuc-SGR contains a firefly luciferase (FLuc) in the subgenomic region of the genome, in place of the structural ORF. The virus construct, termed ICRES-CHIKV throughout this project is a full-length infectious clone of the ECSA genotype LR2006 OPY1 isolate virus. A twin-strep-tag was cloned into nsP3 of both CHIKV-FLuc-SGR and ICRES-CHIKV by Dr Raymond Li (past member of the Harris research group) to generate CHIKV-nsP3-TST-FLuc-SGR and ICRES-CHIKV-nsP3-TST (referred throughout this project as TST-ICRES-CHIKV) for purification purposes.

2.1.1.2 Transformation of bacterial cells with DNA

All plasmids were transformed into chemically competent *Escherichia coli* (DH5α) as follow; briefly, 0.1-100 ng/μl of DNA was incubated with DH5α for 5 minutes on ice then spread out onto Luria-Bertani (LB) agar plates supplemented with 0.1 μg/ml ampicillin (amp). Plates were incubated at 37 °C overnight. Colonies were then picked for incubation in LB broth (0.1 μg/ml amp) overnight at 37 °C. Plasmid DNA was extracted from bacteria using either the ThermoScientific Genejet plasmid midiprep kit or the Monarch plasmid miniprep kit where appropriate, according to the manufacturer's recommended protocol. Resulting DNA concentrations were quantified using Nanodrop (ThermoScientific). DNA was kept at -20 °C for short-term storage. Glycerol preparation was used for long-term storage of plasmid DNA; 2.0 mL of bacteria overnight culture was pelleted and re-suspended with 20% glycerol: LB broth and stored at -80 °C.

2.1.1.3 Endonuclease digestion

For DNA digestion, DNA (5 µg for downstream *in vitro* transcription or 2 µg for cloning purposes) was digested with the appropriate restriction enzyme and incubated at 37 °C for a minimum of 4 hours.

2.1.1.4 DNA TAE agarose gel electrophoresis

DNA agarose gels at 1 or 2% w/v made up with 1X TAE buffer (40 mM Tris, 20 mM Acetate and 1 mM EDTA) and SYBR® safe DNA gel stain (Invitrogen) (1:10,000) were used for quality control and size selection of DNA fragments. DNA samples were electrophoresed at 100 volts for 60 minutes alongside a 1kb DNA size marker (Bioline).

2.1.1.5 Gel extraction of DNA

DNA bands were visualised with blue light (470 nm), and the relevant bands were excised from the gel using a scalpel and placed in a microcentrifuge tube. The excised DNA was then extracted from the gel slices using the Monarch gel extraction kit (New England BioLabs) following the manufacturer's protocol.

2.1.1.6 Ligation of DNA

Following digestion with appropriate restriction enzymes, DNA vector was treated with Calf Intestinal Alkaline Phosphatase (CIP) to prevent vector re-ligation. For this, 10 units of CIP was added to the reaction mix and incubated at 37 °C for 1 hour. Vector and insert were ligated at a ratio of 5:1 using 50 ng of vector DNA. Ligation reactions were assembled with T4 Ligase, buffer, vector DNA, insert DNA and nuclease-free water followed by incubation at 16 °C overnight. The next day, the ligated products were then introduced into competent DH5α cells.

2.1.1.7 Phenol chloroform extraction and ethanol precipitation of DNA

Following digestion, dH_2O was added to DNA to a total volume of 200 µl. 1 volume of UltraPure™ Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v) (Invitrogen) was added and vortexed for 1 min followed by centrifugation at max x g for 5 minutes. The upper phase was then extracted and put into a fresh microcentrifuge tube. 1 volume of chloroform was then added and sample vortexed and centrifuged as described. Following centrifugation, the upper phase was extracted as before. 2 volumes of absolute ethanol and 0.1 volume of 3M Sodium Acetate (pH 5.2) were then added, and the sample was incubated at -20°C for 3 h or -80°C for 1 h. After incubation, the sample was centrifuged at 4°C for 20 minutes at max speed. The supernatant was removed, and pellet washed with 70% ethanol. DNA pellet was then air dried and re-suspended in 20 µL Diethyl pyrocarbonate (DEPC)-dH_2O.

2.1.1.8 *In vitro* transcription

Capped RNA synthesis of CHIKV replicons was achieved using the mMESSAGE mMACHINE SP6 Transcription Kit (ThermoFisher) following the manufacturer's recommended protocol using 1 µg of *Not*I linearised CHIKV replicon DNA (see 2.1.1.3). Resulting RNA was quantified on the Nanodrop using the ratio of absorbance at 260/280 nm.

2.1.1.9 RNA agarose gel electrophoresis

RNA agarose gels at 1% w/v were prepared by adding 0.3 g of agarose powder to 30 mL MOPS buffer (40 mM MOPS, 10mM sodium acetate and 1 mM EDTA) and microwaving until all solids are fully dissolved. 6.5% v/v formaldehyde and SYBR safe DNA gel stain (1:10,000) was then added to the gel and poured into a gel cast. RNA samples were mixed with RNA loading dye (New England BioLabs) and heated to 65 °C for 10 minutes then loaded alongside an ssRNA ladder (New England BioLabs) into the MOPS gel. The gel was then electrophoresed at 80 volts for an hour.

2.1.1.10 Oligonucleotide primers

DNA oligonucleotides were ordered from Integrated DNA Technologies and resuspended with deionised water to 100 µM. All oligonucleotides were stored at -20 °C. All primers used are listed in Appendix Table 6.1.

2.1.1.11 DNA sequencing and analysis

DNA was subjected to Sanger sequencing by the commercial company Genewiz using primers listed in Appendix Table 6.1. Sanger sequencing results were analysed using Benchling, a cloud-based molecular biology software package.

2.1.1.12 RNA sequencing and analysis (Sanger)

RNA purification was performed using TRI Reagent™. Extracted RNA was reverse transcribed with SuperScript II Reverse Transcriptase kit (Life Technologies) using the manufacturer's protocol and random hexamer primers (ThermoScientific). Where appropriate, the resulting cDNA was then PCR amplified for desired regions using specific primers (Appendix Table 6.1). Full length cDNA or PCR products were then sent for Sanger sequencing as described or sequenced in-house using Oxford Nanopore Technologies minION sequencing (Mk1C).

2.1.1.13 RNA sequencing and analysis (Oxford Nanopore)

RNA purification was performed using TRI Reagent™. Oxford nanopore sequencing was performed using the Midnight Sequencing Kit (Oxford Nanopore Technologies C19MAXI) modified with primers designed for the ICRES-CHIKV genome (accession number: DQ443544.2) using primal scheme (Quick et al., 2017). Briefly, extracted RNA was reverse transcribed according to the manufacturer's protocol. PCR reactions were then performed using two separate pools of custom CHIKV specific primers (Appendix Table 6.1) to generate 1200 nt overlapping fragments that cover the entire viral genome. PCR products from both reactions are then combined, barcoded and

purified using AMPure XP bead purification. Purified products are then loaded onto the minION flow cell. Sequencing data was then analysed using a pipeline based on InterARTIC (Ferguson et al., 2022). In brief, sequenced amplicons are PCR tiled to make a consensus sequence which is then aligned to a reference CHIKV sequence.

2.1.2 Mutagenesis

2.1.2.1 Quick-change mutagenesis

Complementary site-directed mutagenesis primers were designed for amplification of the mutants using the NEBaseChanger™ tool (Appendix Table 6.1). Where appropriate, the Q5® Site-Directed Mutagenesis Kit (New England BioLabs) was used. 25 µL reaction mixtures were made up with Q5 Hot Start High-Fidelity Master Mix, 10 µM of forward primer, 10 µM of reverse primer and 25 ng of template DNA. PCR reactions were cycled under the following conditions:

Table 2.1. PCR Programme for Quick-Change mutagenesis

Reaction Step	Temperature	Time
Initial Denaturation	98°C	30 seconds
30 cycles	98°C	10 seconds
	(Annealing temperature dependent on primer)	30 seconds
	72°C	30 seconds/kb (6 minutes and 40 seconds for CHIKV replicons)
Final Extension	72°C	2 minutes

PCR products were then treated with a multi-enzyme mix containing kinase, ligase and *Dpn*I for efficient phosphorylation, ligation of PCR products and removal of input template, followed by transformation into competent DH5α cells as described in 2.1.1.2. DNA was then extracted, purified, and sequenced as described in 2.1.1.2 and 2.1.1.11.

2.1.2.2 PCR mutagenesis

For PCR mutagenesis, two restriction sites, *Spe*I and *Avr*II were chosen for ease of downstream cloning. Briefly, specific forward and reverse primers (Appendix Table 6.1) were used to generate two DNA fragments with each of the restriction site at one end and an overlapping fragment containing the desired mutation on the other. A second PCR reaction was then set up using the DNA fragments as templates to generate one cohesive fragment containing both the *Spe*I restriction site at the 5' end, the *Avr*II site at the 3' end and the desired mutation. This PCR product was then electrophoresed on an DNA TAE agarose gel and extracted as described in 2.1.1.4 and 2.1.1.5. Purified DNA, alongside a vector (CHIKV-FLuc-SGR), was digested using *Spe*I and *Avr*II. The products were electrophoresed on another DNA TAE agarose gel for quality and size control then purified (2.1.1.4 and 2.1.1.5.). Finally, the vector and inserts were ligated, transformed into competent DH5α cells, and left to incubate overnight at 37. The next day, colonies were picked, grown and the DNA was then extracted (2.1.1.6 and 2.1.1.2). To confirm the presence of the desired mutation, colony PCR was performed using a forward and reverse primer covering the fragment of interest (Appendix Table 6.1). PCR products were then run on a DNA TAE agarose gel (2.1.1.4). DNA containing the correct mutation was then sent for sequencing (2.1.1.11). To introduce the desired mutations into ICRES-CHIKV, both CHIKV-FLuc-SGR containing the appropriate mutation and an ICRES-CHIKV construct was digested with *Spe*I and *Age*I. The insert containing the desired mutation from CHIKV-FLuc-SGR was then ligated into the ICRES-CHIKV vector.

2.1.3 SDS PAGE and Western Blotting

2.1.3.1 BCA assay

Protein concentration of cell lysates collected for western blot analysis (see 2.2.7) were determined using the Pierce™ BCA Protein Assay Kit (Thermo Fisher) according to the manufacturer's instructions. Assay standards were made using 1X Glasgow lysis buffer (GLB) (pH7.2) containing 10 mM PIPES, 120 mM KCL, 30 mM NaCl, 5 mM $MgCl_2$, 1% Triton X-100, 10% Glycerol and a cocktail of protease and phosphatase inhibitors (Leupeptin, Pepstatin A,

Aprotinin, Pefabloc, Na_3VO_4 and $Na_4P_2O_7$), and Bovine Serum Albumin (BSA) to concentrations of 2.0 mg/mL, 1.0 mg/mL, 0.5 mg/mL, 0.250 mg/mL, 0.125 mg/mL and 0.0 mg/mL (as a negative control). 5 μL of standard or cell lysates were placed in a single well of a flat-bottom 96-well plate. 200 μL of BCA assay solution was then added into the wells and mixed. The plate was then incubated at 37 °C for 30 minutes in a shaking incubator. The absorbance was read using a BMG plate reader.

2.1.3.2 SDS-PAGE

Equal amounts of protein were resolved using 10% SDS-polyacrylamide gel electrophoresis. The resolving gel was made with 10% (v/v) acrylamide, 375 mM Tris-HCL (pH 8.8), 0.1% (w/v) ammonium persulphate (APS), 0.1% (w/v) sodium dodecyl sulphate and 0.01% (v/v) Tetramethylethylenediamine (TEMED). The stacking gel was made with 6% acrylamide, 376 mM Tris-HCl (pH 6.8), 0.1% APS, 0.1% SDS and 0.01% TEMED. The protein samples were prepared by mixing with SDS-PAGE loading buffer (62.5 mM Tris-HCl, pH 6.8, 10% (v/v) Glycerol, 2% (w/v) SDS, 0.01 % (w/v) Bromophenol blue, 5% (v/v) ß-mercaptoethanol) to a final concentration of 1X then incubated at 95 °C for 5 minutes. Following heat-denaturation, the samples were briefly centrifuged then loaded into the gel alongside a broad range protein marker (New England BioLabs) and electrophoresed for 1 hour at 180 volts.

2.1.3.3 Western blot

Following SDS-PAGE, the gel was transferred onto Hybond PVDF Blotting membrane (Amersham biosciences) using the Hoefer semi-dry transfer unit (TE77X) at 15 volts for 1 hour. The PVDF membrane was then blocked with 50% (v/v) Odyssey blocking buffer (LI-COR) diluted in 1X TBS (20 mM Tris and 150 mM NaCl) for 1 hour, with rocking, at room temperature. The membrane was then probed with primary antibodies (Table 2.2) at the appropriate dilution with 25% (v/v) Odyssey blocking buffer and incubated on a roller at 4 °C overnight. The next morning, the membrane was washed 3X with 1X TBS supplemented with 0.1% (v/v) Tween®-20 (rocking for 5 minutes

for each wash). Blots were then probed with secondary antibodies (Table 2.3) for 1 hour followed by 3 more washes as described. After the final wash, the membrane was left to dry on clean filter paper then scanned using the LI-COR Odyssey® Sa Imaging system (LI-COR).

2.1.3.4 Antibodies

Table 2.2. List of primary antibodies used in experimentation.

Primary Antibody	Species	Origin	Working dilution
Anti-nsP3	Rabbit	Andres Merits	1:1000
Anti-nsP1	Rabbit	GeneTex	1:1000
Anti-capsid	Rabbit	Andres Merits	1:1000
Anti-β actin	Mouse	Sigma A1978	1:10,000
Anti-Strep tag	Mouse	Stratech	1:1000
Anti-HuR	Mouse	SantaCruzBiotechnology	2 µg

Table 2.3. List of secondary antibodies used in experimentation.

Secondary Antibody	Origin	Working dilution
Donkey anti-mouse (680 nm)	LI-COR	1:20,000
Donkey anti-rabbit (680 nm)	LI-COR	1:20,000
Donkey anti-rabbit (800 nm)	LI-COR	1:20,000

2.2 Cell Culture

2.2.1 Mammalian cell lines and maintenance

Mammalian cell lines and their required culture media are shown in Table 2.4. Mammalian cells were maintained at 37 °C in humidified air containing 5% carbon dioxide.

Table 2.4. List of mammalian cells used in this project, their origin/phenotypes and required cell culture.

Cell line	Origin/Phenotype	Media
Huh7 (Nakabayashi et al., 1982)	Human, liver hepatocellular carcinoma	-DMEM (Dulbecco's modified eagle media) (Sigma Aldrich) -10% FBS (Foetal Bovine Serum) (Life Technologies) -1.1% non-essential amino acids (NEAA) (Lonza) -2.8% HEPES (Lonza)
C2C12 (Yaffe and Saxel, 1977)	Mouse, muscle myoblasts (can be differentiated to produce muscle tissue)	-DMEM -20% FBS -1.1% NEAA -2.8% HEPES
BHK-21 (Macpherson and Stoker, 1962)	Baby hamster, kidney fibroblasts	-DMEM -10% FBS -1.1% NEAA -2.8% HEPES

2.2.2 Passaging of mammalian cells

Cells were washed with PBS, and then incubated in 2 mL of trypsin. Following incubation at 37 °C, detached cells were resuspended in complete media and the appropriate number of cells were transferred into a new T175 cell culture flask. Total volume was then made up to 20 mL with media.

2.2.3 Mosquito cell lines and maintenance

Mosquito cell lines and their required culture media are shown in Table 2.5. Mosquito cells were maintained at 28 °C without CO_2.

Table 2.5. Mosquito cells lines used in this project, their origin/phenotypes and required culture media.

Cell line	Origin/Phenotype	Media
C6/36 (Igarashi, 1978)	Mosquito (*Aedes albopictus*), embryonic cells lacking functional RNAi response	-Leibovitz's L-15 medium (Life Technologies) -10% FBS -10% Tryptose phosphate broth (TPB) (Life Technologies)
U4.4 (Condreay and Brown, 1986)	Mosquito (*Aedes albopictus*), embryonic cells	Leibovitz's L-15 medium -10% FBS -10% TBP

2.2.4 Passaging of mosquito cells

Cells were first washed with PBS, then 3 mL of complete mosquito media was added into the flask. Cells were then scraped off using a cell scraper and resuspended in complete media. The appropriate number of cells was transferred into a new T175 cell culture flask. Total volume was then made up to 20 mL with media.

2.2.5 Transfection of cells

Transfection mixes were made by separately mixing an appropriate amount of RNA and Lipofectamine with Opti-MEM (Gibco) at room temperature for 5 minutes. RNA-Opti-MEM was then added to Lipofectamine-Opti-MEM and left

to incubate for 20 minutes at room temperature. The transfection mixture was then delivered to cells and wells were supplemented with 400 µL of Opti-MEM to prevent desiccation followed by incubation at 37 °C. 4 hours post transfection, transfection mixes were removed and replaced with appropriate culture media (see Table 2.4 and Table 2.5), and cells were then left to incubate and harvested when required.

2.2.6 Cell lysis for luciferase assays

For cell lysis, cells were first washed with PBS, then harvested on ice using Passive lysis buffer (Promega) and rocking for 15 mins. Firefly luciferase activity was assessed using LARI reagent (Promega) in a flat-bottom opaque 96-well plate. 30 µL of cell lysate was placed in a well followed by the addition of LARI (50 µL). Firefly luciferase signal was read using the FLUOstar Optima (BMG Labtech).

2.2.7 Cell lysis for western blot analysis

Cells were harvested on ice by the addition of ice-cold Glasgow lysis buffer (GLB) along with rocking for 15 mins.

2.2.8 Electroporation of cells for virus propagation

Cells were first trypsinised and resuspended as described in 2.2.2 then centrifuged at 1000 x g for 3 mins. Following centrifugation, the supernatant was removed, and cells were washed by resuspension in ice cold DEPC (Diethyl pyrocarbonate) PBS. Cells were then counted and resuspended to a density of 3.0×10^6 cells/mL. 400 µL of resuspended cells were then mixed with 1 µg of virus RNA in a cold 4 mm electroporation cuvette (Geneflow), and electroporated with a single pulse at 250 volts for 25 ms. The electroporated cells were then resuspended in 10 mL, placed into a T75 cell culture flask and taken to the BSL3 facility. Media was removed at appropriate time post electroporation (usually 24 or 48 hours post electroporation), aliquoted and stored at -80 °C as virus aliquots for later use.

2.2.9 Virus titration by plaque assay

Plaque assays were performed on BHK cells in 12-well plates with 1.0×10^5 cells. Collected virus was diluted serially 10-fold in serum free media from 10^{-1} to 10^{-8}. 150 µL of the appropriate virus dilutions (determined by expected titres) was then added to the cells seeded 16 hours before and left to rock for 5 mins to facilitate adsorption followed by 1-hour incubation at 37 °C. After incubation, the virus was removed, and cells were washed with PBS then overlaid with 1.6% methyl cellulose (MC) mixed 1:1 with complete media. Cells were incubated at 37 °C for 48-72 hours then fixed with 4% formaldehyde for 30 mins and stained using 0.25% crystal violet staining solution. Plaques were counted and virus titre was calculated using the following equation:

$$PFU/mL = \frac{Number\ of\ plaques\ observed}{dilution\ factor\ (10^{-x}) \times volume\ of\ virus\ added\ (mL)}$$

2.2.10 Infection of cells with CHIKV

Cells used in virus infection were seeded into plates 16 hours before infection. During infection, the culture media is first removed followed by washing with PBS. Virus was diluted in serum-free media to the desired MOI then added to the cells. Plates were then rocked for 5 mins at room temperature then incubated at 37 °C for 1 hour. The media containing the virus was then removed, and cells were washed with PBS followed by the addition of complete media. Infected cells were then left to incubate at 37 °C with 5 % CO_2 until harvest.

2.2.11 TRI Reagent extractions

TRI Reagent™ (FisherScientific) was used to extract RNA from both infected cells and from virus aliquots. For a T-75 flask of infected cells, the culture media was first removed, and then cells were washed with PBS. Cells were scraped into 1 mL PBS, transferred to a microcentrifuge tube and centrifuged twice at 500 x g for 1 minute. The supernatant was then removed, and cells lysed with 1.4 mL of TRI Reagent™. For virus aliquots, 900 µL of TRI Reagent™ was

added to 500 µL of virus to generate a total volume of 1.4 mL. Following TRI Reagent™ addition, samples were incubated at RT for 5 minutes. 100 µL of chloroform was then added to the sample, briefly vortexed and incubated for 5 minutes. The sample was then centrifuged at 12,000 x g for 15 minutes at 4 °C. The aqueous layer was then taken forward into a fresh microcentrifuge tube and 250 µL of iso-propanol was added. The sample was mixed by inversion then incubated for 10 minutes. Following incubation, the sample was centrifuged again at 12,000 x g at 4 °C to pellet the RNA. The supernatant was then removed, and the RNA pellet was washed with 500 µL ice-cold 75% ethanol by centrifugation at 7,500 x g for 5 minutes at 4 °C. The ethanol was then removed, and the pellet was dried for 3 minutes on the bench top. Once dried, the pellet was resuspended in 20 µL of nuclease-free water.

2.3 iCLIP

2.3.1 iCLIP Buffers

Lysis Buffer
50 mM Tris-HCl, pH 7.4
100 mM NaCl
1% NP-40 Alternative (Non-ionic surfactant)
0.1% SDS
0.5% sodium deoxycholate

High-salt Wash Buffer
50 mM Tris-HCl, pH 7.4
1 M NaCl
1 mM EDTA
1% NP-40 Alternative
0.1% SDS
0.5% sodium deoxycholate

PNK Wash Buffer
20 mM Tris-HCl, pH 7.4
10 mM MgCl2
0.2% Tween-20

<u>5x PNK *pH 6.5* Buffer</u>
350 mM Tris-HCl, pH 6.5
50 mM MgCl2
5 mM dithiothreitol
(Aliquots were kept in the freezer and defrosted for single use when required)

<u>10x Ligation Buffer (DTT free)</u>
500mM Tris-HCL 7.5
100mM MgCl2
(Aliquots frozen as above)

<u>PK +SDS Buffer</u>
10 mM Tris-HCl, pH 7.4
100 mM NaCl
1 mM EDTA
0.2% SDS

<u>20X MOPS Running Buffer</u>
50 mM MOPS
50 mM Tris Base
0.1% SDS
1 mM EDTA
pH 7.7

2.3.2 Infection and UV-crosslinking of cells for iCLIP

Cells were seeded in 10 cm^2 dishes 16 hours before infection and either mock infected or infected with ICRES-nsP3-TST-CHIKV the following day as described in 2.2.10. 12 hours post infection, the culture media was removed, and cells washed with PBS. 1 mL of ice-cold PBS was then added, and the plates were placed into the UVP® CL-1000® Ultraviolet Crosslinker and irradiated with 150 mJ/cm^2 at 254 nm. Cells were then harvested manually with a cell scraper and transferred into a microcentrifuge tube for centrifugation twice at 500 x g for 1 min to pellet cells. The supernatant was then removed, and cells were lysed by incubation with 500 µL of iCLIP lysis buffer (see 2.3.1) for 30 mins.

2.3.3 Cell lysate sonication and protein quantification

Cells lysates for sonication are first transferred into polystyrene sonication tubes (Active Motif) then sonicated using the Bioruptor® Pico for 10 cycles of 30 secs on/off. Protein concentration of the lysates is then quantified as described in 2.1.3.1.

2.3.4 In-lysate partial RNA digestion

Where required, lysates were digested with 0.4, 0.8 or 2.5 units/mL lysate by the addition of RNase I (ThermoScientific) or undigested as a negative control. At the same time, 2 µL of Turbo DNase (Invitrogen) was added to the lysates followed by incubation for 3 (0.4 unit/mL RNase), 4 (0.8 unit/mL RNase) or 5 (2.5 unit/mL RNase) mins in a thermomixer (Eppendorf) at 37 °C shaking at 1100 rpm. After incubation, lysates were transferred to ice for a 3-minute incubation.

2.3.5 Bead preparation and lysate incubation

For each experiment, 100 µL of protein G dynabeads (ThermoScientific) was transferred into a LoBind microcentrifuge tube (FisherScientific) and placed on a magnetic rack. The buffer was removed, and beads were washed twice with iCLIP lysis buffer (all washes throughout are performed with 900 µL and a single wash is carried out by reversing the beads on a magnetic rack twice). The beads were then resuspended in 100 µL of iCLIP lysis buffer plus either 5 µg of anti-strep antibody, 2 µg of HuR antibody or no antibody (as a negative control) and tubes were rotated at room temperature for 60 mins. After antibody coupling, the beads were washed thrice with iCLIP lysis buffer. The RNase I digested lysates are then added to the appropriate tubes and left to incubate by rotation at 4 °C overnight.

2.3.6 Immunoprecipitation

Following incubation, the supernatant was discarded, and beads were washed twice with iCLIP high-salt buffer and once with iCLIP PNK buffer. The beads

were then resuspended in 900 µL iCLIP PNK buffer and moved to a fresh LoBind microcentrifuge tube.

2.3.7 3' end RNA dephosphorylation

The PNK wash was then removed, and the beads were resuspended in 40 µL of the following mixture (a master mix was made then 40 µL added to each sample):

8.0 µL	5x PNK *pH 6.5* buffer
1.0 µL	PNK (New England BioLabs)
0.5 µL	FastAP alkaline phosphatase (ThermoFisher)
0.5 µL	RNasin (Promega)
30.0 µL	Nuclease free water

The samples were then incubated for 40 min at 37°C in thermomixer at 1100rpm.

2.3.8 3' end adaptor ligation

Once the incubation finished, the beads were placed on a magnetic rack and the RNA dephosphorylation mix was removed. The beads were then washed once with 1X iCLIP ligation buffer. The wash was then removed, and the beads were resuspended in 25 µL of the following mixture (a master mix was made then 25 µL taken for each sample):

6.3 µL	Nuclease free water
3.0 µL	10X iCLIP ligation buffer
0.8 µL	100% DMSO (New England BioLabs)
2.5 µL	T4 RNA ligase I – high concentration (New England BioLabs)
0.4 µL	RNasin
0.5 µL	PNK (Thermo fisher scientific)
2.5 µL	1 µM pre-adenylated adapter (Kindly gifted by Jernej Ule)
9.0 µL	50% PEG8000 (New England BioLabs)

The samples were then incubated for 75 min at 20 °C in thermomixer at 1100 rpm. After incubation, the beads were washed twice with iCLIP high-salt buffer and once with iCLIP PNK buffer. For the second iCLIP PNK buffer wash, 500 µL was added, samples were resuspended then moved into a new LoBind microcentrifuge tube.

2.3.9 Adaptor removal

The previous PNK wash was then removed by placing beads on the magnetic rack and 20 µL of the following removal mix was added to each sample, with thorough pipetting to ensure complete resuspension:

12.5 µL   Nuclease free water
 2.0 µL   New England BioLabs Buffer 2
 0.5 µL   5' Deadenylase (New England BioLabs)
 0.5 µL   RecJ$_f$ endonuclease (New England BioLabs)
 0.5 µL   RNasin
 4.0 µL   50% PEG8000

Samples were then incubated for 1 hr at 30 °C, then at 37 °C for 30 mins, both shaking at 1100 rpm. The beads were then placed on a magnetic rack and the adaptor removal mix was removed followed by two high-salt buffer washes and one PNK buffer wash. The final wash was then removed, and beads were suspended in 20 µL of 1X NuPAGE® LDS loading buffer (prepared by mixing 4X stock LDS loading buffer (Invitrogen) with nuclease free water and reducing agent DTT (final concentration 100 mM).

2.3.10 SDS-PAGE

Samples resuspended in 1x NuPAGE® LDS loading buffer were placed in a thermomixer preheated to 70 °C and incubated for 2 minutes with shaking at 1100 rpm, and then immediately placed back on a magnetic rack. The supernatants were then collected for loading on a 4-12% NuPAGE® Bis-Tris gel (Invitrogen) in a Novex NuPAGE® buffer system alongside 5 µL of pre-stained broad range protein standard (New England BioLabs), according to the manufacturer's instructions. The gel was then electrophoresed with 0.5 L 1X MOPS running buffer (2.3.1) for 65 minutes at 180V, or until the dye front had migrated to the bottom of the gel. Following electrophoresis, the gel was taken out and the dye front was removed and discarded.

2.3.11 Nitrocellulose transfer and analysis

The protein-RNA-L3 complexes from the gel were transferred to a Amersham™ Protran® nitrocellulose membrane using the Novex wet transfer

apparatus with 1X NuPAGE® Transfer Buffer + 10% methanol for 2 hours at 30 volts, according to the manufacturer's instructions. Once the transfer was complete, the membrane was removed and washed with DEPC-PBS, then wrapped in a plastic pocket and scanned with the Odyssey® LI-COR SA imager using both the 700 nm and 800 nm channels, then stored in the fridge whilst the infrared image was analysed.

2.3.12 Protein digestion and RNA Isolation

The desired protein-RNA complexes were first highlighted with a fine liner then excised using RNase AWAY™ (Thermo Scientific) treated-scissors and placed into a microcentrifuge tube using RNase AWAY™ (Thermo Scientific) treated-tweezers. For protein digestion, 10 µL of proteinase K was mixed with to 390 µL PK+SDS buffer and added to the membrane then incubated at 50 °C for 60 mins with shaking at 1100 rpm. Following incubation, the solution was collected into a fresh tube, and 400 µL of Phenol:Chloroform:Isoamyl Alcohol (ThermoFisher) was added, mixed and left to incubate for 5 mins at 30 °C with shaking at 1100 rpm. The phases were then separated with centrifugation at room temperature for 5 mins at 16000 x g. The aqueous phase was then taken into a separate tube and 800 µL Chloroform (Find Manufacturer) was added, mixed gently, and centrifuged for 5 minutes at room temperature at 16000 x g. After centrifugation, the aqueous phase was then taken into a fresh microcentrifuge tube. The RNA was then precipitated by the addition of 0.75 µL glycogen (Invitrogen) 40 µL 5M NaCl (Fisher Scientific) and 500 µL 100% Ethanol. The samples were then mixed thoroughly and placed at -20 °C overnight. The following morning, the samples were centrifuged for in a 4 °C centrifuge for 20 mins at 16000 x g. The supernatant was then removed, and the pellet was washed with 500 µL 80% nuclease-free ethanol and centrifuged again for 5 minutes. The wash was then removed, and the pellets were left to dry for 3 minutes on the bench and resuspended with 5.5 µL nuclease-free water. The resuspended RNA was then transferred to a fresh PCR tube.

2.3.13 Reverse transcription

For reverse transcription of the purified RNA, the following reagents were added to the RNA:

1.0 µL   primer iCLIP_ddRT_## (Appendix Table 6.1)
0.5 µL   10 mM dNTP mix

The RNA-primer-dNTP mix was then placed in a thermocycler with the following RT thermal program:

65 °C   5 min
25° C   hold until the RT mix is added

Following the initial incubation, an RT mix was then added and mixed thoroughly by pipetting:

2.0 µL   5X SSIV buffer (Invitrogen)
0.5 µL   0.1M DTT
0.25 µL   RNasin
0.25 µL   Superscript IV (Invitrogen)

Reverse transcription was then completed using the following thermal program:

25 °C   5 min
50 °C   5 min
55 °C   5 min
4 °C   hold

After reverse transcription was complete, alkaline hydrolysis of the RNA was carried out by the addition of 1.25 µL of 1M NaOH, followed by mixing and incubation at 85 °C for 15 min. 1.25 µL 1M HCL was then added to neutralise the pH.

2.3.14 cDNA purification using AMPure XP beads capture

For cDNA purification, 3 volumes of Agencourt AMPure XP beads were added to the tube containing the reverse transcription products, mixed and transferred into a microcentrifuge tube. 1.66 volumes of isopropanol were added to the cDNA and mixed by extensive pipetting then incubated for 5 minutes at room temperature. Samples were then placed on a magnetic rack and the beads

66

were allowed to settle for 3 minutes. Following this, the liquid was removed, and the beads were washed twice with 200 µL of 85% ethanol by incubation for 30 seconds. After the last wash was removed, the beads were then centrifuged at 16000 x g and placed back on the magnetic rack to allow for removal of the remaining ethanol. The beads were left to dry on the bench top then eluted with incubation with either 9 or 10 µL of water depending on downstream use (9 µL for circularisation and 10 µL for PCR). The resulting cDNA was collected and placed in a PCR tube.

2.3.15 cDNA circularisation

To the 9 µL of eluted cDNA, the following CircLigase II mix was added:

1.5 µL   10X CircLigase Buffer II (Epicentre)
0.75 µL   CircLigase II (Epicentre)
0.75 µL   50mM $MnCl_2$ (Epicentre)
3.0 µL   5M Betaine (Epicentre)

The samples were incubated at 60 °C for a minimum of 2 hours or overnight. Following circularisation, the cDNA was purified again using AMPure XP beads as described in 2.3.14.

2.3.16 PCR optimisation

For optimisation of PCR amplification, the following mix was prepared:

1.0 µL   cDNA (from cDNA purification for PCR)
0.25 µL   P5Solexa/P3Solexa primer mix (Appendix Table 6.1)
5.0 µL   Phusion HF Master mix (New England BioLabs)
3.75 µL   Nuclease-free water

The samples were then placed in a thermo cycler and ran using the following program:

98°C   40 sec

22, 24, 26 or 28 cycles of
98°C   20 sec
65°C   30 sec
72°C   45 sec

72°C   3 min
25°C   hold

2.3.17 DNA TBE agarose gel electrophoresis

TBE agarose gels at 2% w/v were used for DNA gel electrophoresis of iCLIP PCR products (2% agarose in TBE buffer with 1:10000 SYBR safe (Thermo fisher)). For DNA preparation, 10 µL of PCR product was mixed with 2.27 µL of 6X loading buffer and 1.36 µL 100X Sybr Green DNA stain (Kindly gifted by Prof Eric Blair, University of Leeds). Samples were then electrophoresed alongside a 100bp DNA ladder (New England BioLabs) for 80 mins at 100 volts. The resulting agarose gel was scanned with a Fujifilm FLA-5000 imager.

2.3.18 Preparative PCR for sequencing

From the results gathered in 2.3.16 and 2.3.17, the optimal PCR cycle number was determined and the following PCR mix was prepared:

8.0 µL   cDNA (from cDNA purification for PCR)
30.0 µL   P5Solexa/P3Solexa primer mix (Appendix Table 6.1)
2.0 µL   Phusion HF Master mix (New England BioLabs)
40.0 µL   Nuclease-free water

The PCR products are then separated on a TBE agarose gel with 6X loading dye and 10X SYBR Green for quality control and size selection, then excised and purified as mentioned in 2.1.1.5 with appropriate modifications to the gel dissolving buffer and the elution buffer volumes. Samples with different barcodes that are to be sequenced as one sample are eluted into the same microcentrifuge tube. DNA concentration was then determined using the nanodrop.

2.3.19 Amplicon sequencing

Amplicon sequencing was carried out using the Genewiz® Amplicon-EZ service. The DNA was normalised to 20 ng/ µL and a total of 250 ng was sent for sequencing.

2.3.20 Data analysis

Analysis of the sequencing data was carried out on Galaxy or using Linux command line, as outlined in Appendix Table 6.3 & Table 6.4.

Chapter 3 The role of an nsP3 polyproline motif in the CHIKV lifecycle

3.1 Introduction

The HVD, highlighted by name, is the most variable domain of nsP3 though several features reoccur, suggesting an importance for these in the alphavirus lifecycle. One of these is a proline-rich cluster, hereafter referred to as the polyproline motif, towards the C-terminus of the HVD (Figure 3.1A). Previous studies have demonstrated that large deletions in this region are relatively well-tolerated (Galbraith et al., 2006; Lastarza et al., 1994), however, recent evidence has highlighted the importance of the overall proline-rich region and specific residues within this motif in SFV and SINV, respectively. For SFV, a proline-rich region spanning 34 amino acids was shown to be required for host amphiphysin interaction.
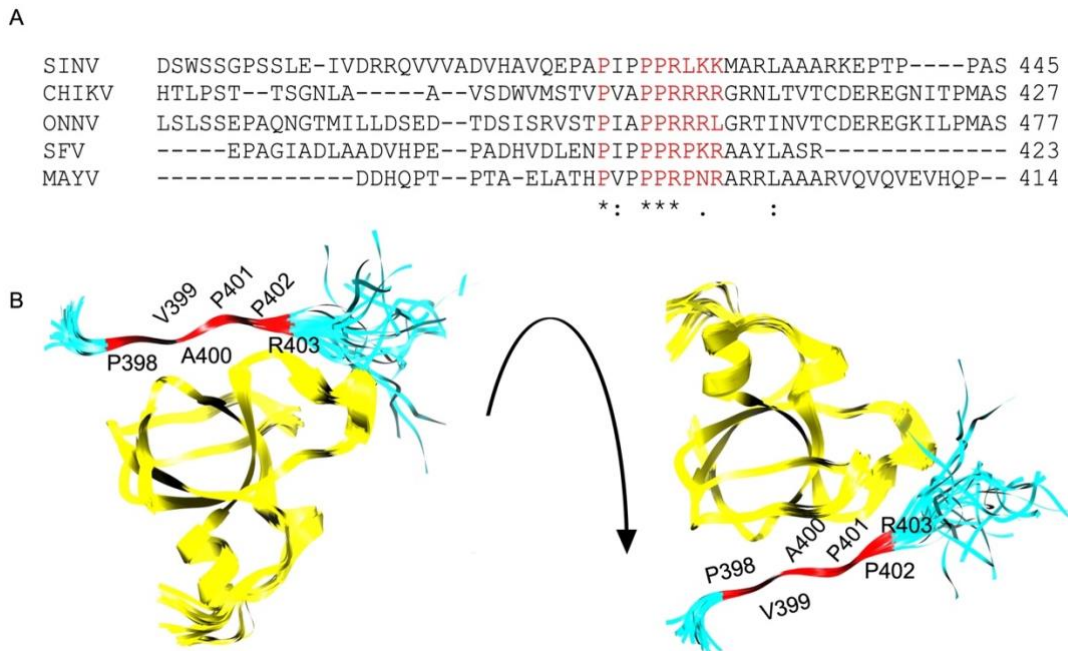
Figure 3.1 Alphavirus nsP3 HVD polyproline motif alignments. A) Amino acid sequence alignment of the alphavirus nsP3 hypervariable domain polyproline motif. Important residues in the canonical polyproline motif are highlighted in red. Alignment is produced using Clustal Omega where * = fully conserved residue, : = residue with strongly similar properties, . = residues with weakly similar properties B) Solution structure of amphiphysin-SH3 domain (yellow) bound to the CHIKV polyproline motif (red = residues directly interacting with amphiphysin, black = residues not directly interacting with amphiphysin). Atomic coordinates and structure factors obtained from PDB (code 5I22), diagram generated with UCSF chimera (Sievers et al., 2011).

In contrast, for SINV, a single positively charged residue, R426, is critical for amphiphysin-1/2 interactions as a charge reversing substitution to glutamic acid completely abolished binding to the cellular protein (Figure 3.1A). The first proline residue was also required as a substitution to alanine compromised amphiphysin-1 binding (Neuvonen et al., 2011). The significance of nsP3-amphiphysin in the context of virus replication was also investigated; RNA synthesis was significantly reduced in both SFV carrying a 34 amino acid deletion, and SINV carrying the R426E mutation (Neuvonen et al., 2011).

Following on from the study conducted by *Neuvonen et al*, a further study interrogated the importance of the nsP3-amphiphysin relationship with a focus on determining the structural and biochemical basis of the interaction. *Tossavainen et al.* found that alphaviruses recruit amphiphysin to the viral replication complex using the nsP3 polyproline motif. Following on from this, the group made several single mutations to the motif, targeting mostly the positively charged arginine residues towards the end of the motif. One mutation, R1739A, caused a substantial decreased in amphiphysin binding. The study also showed that amphiphysin bound nsP3 with higher affinity than dynamin, its cognate binding partner. Using NMR coupled with ITC, it was demonstrated that the strong affinity of amphiphysin for nsP3 is achieved through the positively charged residues within and after the polyproline motif interacting with a large patch of negative electrostatic potential on the surface of amphiphysin. In CHIKV, the five consecutive arginine residues resulted in a nano-molar affinity to amphiphysin (Tossavainen et al., 2016).

Whilst analysis of the alphavirus polyproline motif has been performed in related alphaviruses it has not been investigated in CHIKV. Therefore, the precise function of this motif in the CHIKV lifecycle remains unclear.

3.1.1 Aims

The aim of the work described in this chapter was to characterise the function of the CHIKV nsP3 HVD polyproline motif in the virus lifecycle. To this end, a mutagenic approach was taken targeting the canonical prolines in this motif, in addition to the positively charged arginine residues immediately after the motif. The mutations were then assessed for their replication capacity and infectivity in the context of both sub-genomic replicons and infectious virus, using a range of relevant cell lines.

3.2 Results

3.2.1 Generation of a polyproline motif mutation panel

To determine the function of the polyproline motif, a panel of mutations was generated targeting the specific residues within and following the motif. Firstly, an alignment of the CHIKV nsP3 sequence along with multiple old-world alphaviruses was performed to identify the conserved residues in the polyproline motif (Figure 3.1A). Combining the alignment analysis and solution (Figure 3.1B) structural data for the interactions between nsP3 and amphiphysin produced by *Tossavainen et al*, two proline residues, P398 and P401, along with an arginine residue, R403, were chosen for mutagenesis. This is because the chosen residues are highly conserved in the canonical PxxPxR sequences, and are located at the interface between the nsP3 and amphiphysin SH3-domain interaction. The three arginine residues immediately following the motif were also chosen as the high affinity binding between nsP3 and amphiphysin is mediated by these positively charged residues (Tossavainen et al., 2016). To interrogate whether the proline residues were able to function individually or synergistically, a panel of single, double triple and quadruple alanine or glutamic acid substitutions were made, as shown in Figure 3.2. Furthermore, to understand the function of the entire motif, a deletion of fifteen amino acids was made.

Mutagenesis was performed on CHIKV-FLuc-SGR using PCR mutagenesis as described in 2.1.2. Once confirmed, the region containing the mutations (Figure 3.2) was excised using specific restriction enzymes *Avr*II and *Spe*I then cloned into a freshly prepared CHIKV sub-genomic replicon construct (specific nucleotide changes are shown in Appendix Table 6.2).
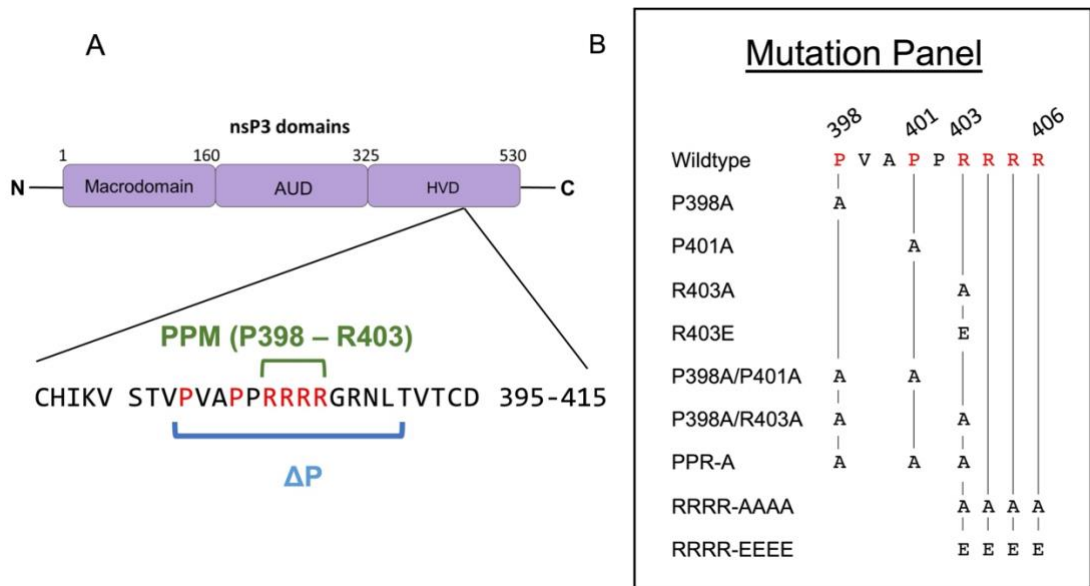
Figure 3.2 Panel of polyproline mutations. A) Schematic of nsP3 with the individual domains highlighted. Green denotes the canonical polyproline motif and important residues are highlighted in red. ΔP denotes the 15 amino acid deletion. B) Mutation panel showing the single, double, triple, or quadruple mutations that were generated in the polyproline motif.

3.2.2 Phenotypic analysis of CHIKV polyproline mutants in human cell lines

The panel of mutants along with a WT CHIKV-FLuc-SGR as a positive control and a replication deficient CHIKV-Fluc-nsP4-GAA-SGR as a negative control, were transfected into multiple relevant cell lines and replication was measured. CHIKV-Fluc-nsP4-GAA-SGR contains two aspartic acid to alanine substitutions in the active site of the RdRp, rendering the polymerase non-functional and therefore unable to replicate. In the subgenomic replicon, the structural genes are replaced with a firefly luciferase reporter, thus the production of firefly luciferase signal is representative of replication levels. Firstly, the mutations were interrogated in a human hepatoma derived cell line (Huh7) as the liver is a natural replication site for the virus and they efficiently support CHIKV replication (Roberts et al., 2017). In Huh7 cells as shown in Figure 3.3A, WT CHIKV-FLuc-SGR produced robust replication with FLuc levels reaching an average of 8×10^5 at 24 h. Peak replication for CHIKV-FLuc-SGR was determined to be at 24 h, thus for all mammalian cells lines, time

points were taken up to 24 h. All the polyproline mutants tested displayed FLuc levels similar to WT, with the exception of the P398A and RRRR-E mutants, which both had a slight (~1 log) but significant reduction in replication. Previous work in this laboratory had used a dual luciferase subgenomic replicon system (Gao et al., 2019; Roberts et al., 2017) where an additional Renilla luciferase cloned into the HVD of nsP3 is used to measure input translation and thus the firefly luciferase can be normalised using this value. However, as the polyproline motif is located in close proximity to the Renilla luciferase insertion site, the decision was taken to use a single luciferase system. This ensures that all observed effects are a result of the mutations in this region, rather than potential interference from Renilla luciferase. Thus, the firefly luciferase data was instead normalised to the 0-h time point as shown in Figure 3.3B. Normalisation of the results showed a similar trend as un-normalised data, which provides confidence that the replication levels are true and not as a result of residual RNA that remain following transfection.
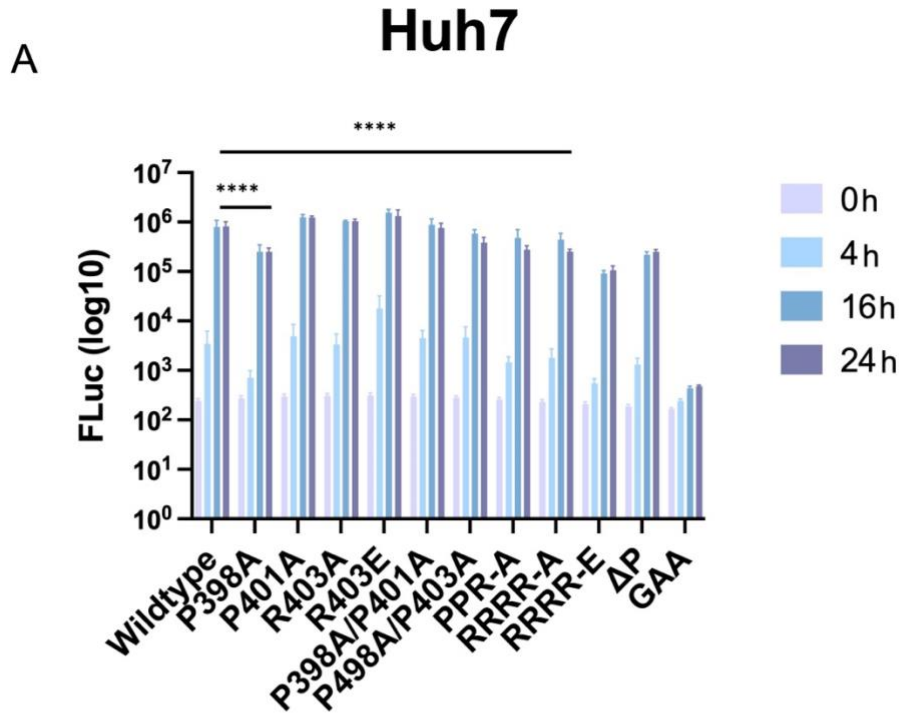
Figure 3.3 CHIKV polyproline motif mutant replication in Huh7 cells. A) Luciferase signal measured for Huh7 cells transfected with WT or mutant CHIKV-FLuc-SGR and harvested at 0, 4, 16 and 24 h. GAA denotes the replication inactive nsP4 mutant. (Data analysed by ordinary Two-way ANOVA with Dunnett's multiple comparison test compared to WT at 24 h, **** = $\leq$ 0.0001. Statistical significance is only shown for mutants that displayed a phenotype, n=2. B) Luciferase values normalised to the 0-h time point.

Secondly, the panel of mutations were transfected into a human rhabdomyosarcoma (RD) cell line derived from muscle tissue. Again, this cell line was chosen as CHIKV naturally infects human muscle cells. Similarly to the results observed in the Huh7 cell line, WT CHIKV-FLuc-SGR can replicate efficiently in RD cells (Figure 3.4A). Compared to Huh7 cells, the levels of replication observed in P398A was not reduced to the same levels. Again, the RRRR-E mutant showed a slight but significant decrease in replication compared to the WT. In accordance with this, the RRRR-A mutant also produced a similar phenotype of reduced replication levels. All other mutations in the panel replicated to levels comparable to WT.

3.2.3 Phenotypic analysis of CHIKV polyproline mutants in non-human mammalian cell lines

Next, the mutant panel was evaluated in two non-human mammalian cell lines. First, the panel was interrogated in baby hamster kidney (BHK) cells. BHK-21 cells are the standard cell line used for alphavirus propagation as they are extremely permissive to infection. Therefore, it was important to screen the replication panel in this cell line to ensure that downstream investigations in the context of virus infection are not affected by cell-line specificity. As observed in the human cell lines, WT CHIKV-FLuc-SGR replicated to a high level in BHK-21 cells. All mutants replicated to high levels in BHK-21 cells, and the phenotypes observed with P398A, RRRR-A and RRRR-E in Huh7/RD cell lines were not displayed in this cell line Figure 3.5A.

The panel was then transfected into a murine myoblast cell line (C2C12) to act as a comparison for BHK-21 cells. As expected, WT CHIKV-FLuc-SGR replicated to high levels in this cell line and in general all other mutants replicated to levels comparable to WT (Figure 3.6). Notably in C2C12 cells, a reduction in replication levels of RRRR-A and RRRR-E which was not observed in BHK-21 cells but observed in RD cells were demonstrated again in this murine cell line.
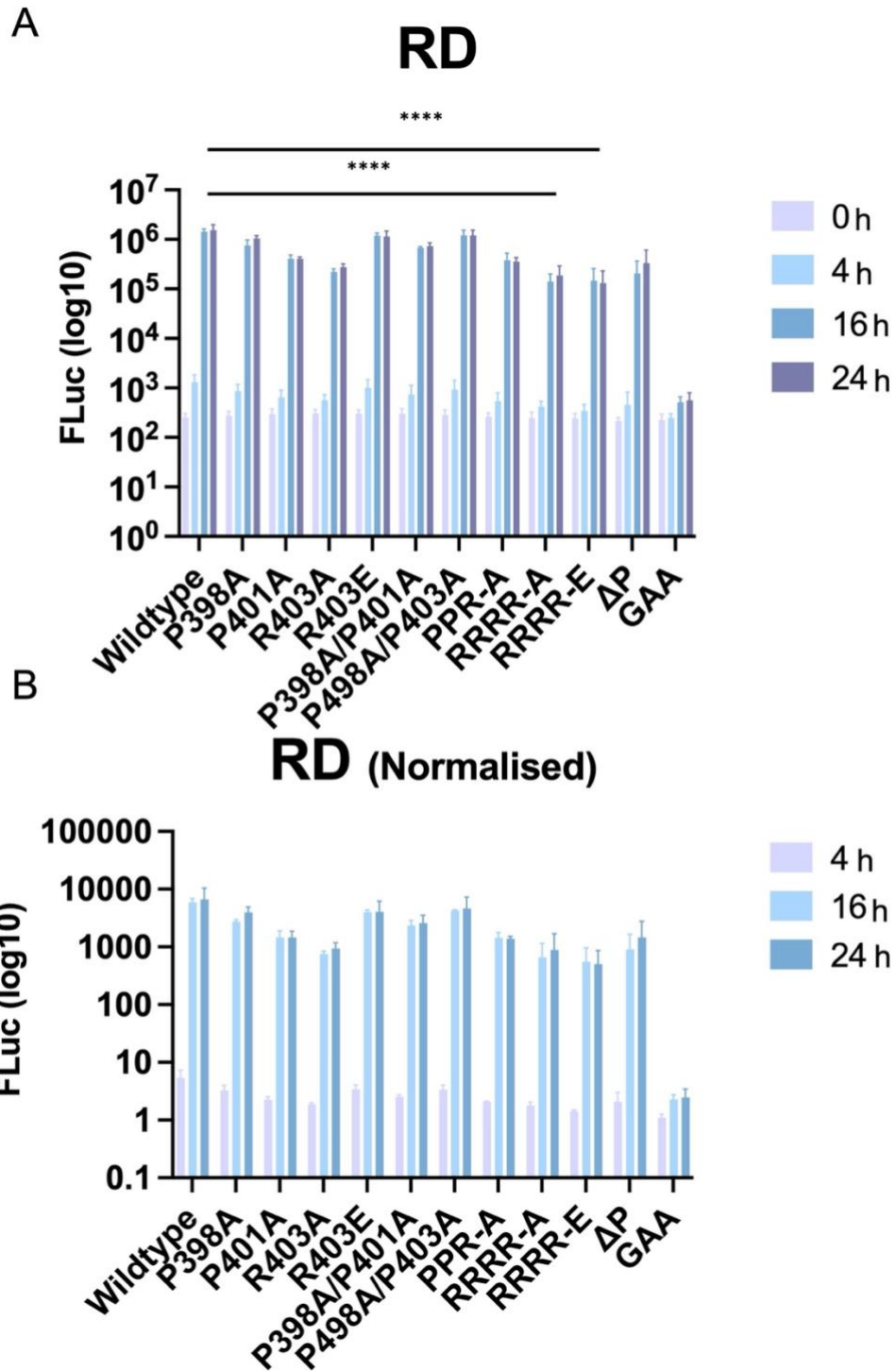
Figure 3.4 CHIKV polyproline motif mutant replication in RD cells. A) Luciferase signal measured for RD cells transfected with WT or mutant CHIKV-FLuc-SGR and harvested at 0, 4, 16 and 24 h. GAA denotes the replication inactive nsP4 mutant. Data analysed by ordinary Two-way ANOVA with Dunnett's multiple comparison test compared to WT at 24 h, **** = ≤ 0.0001. Statistical significance is only shown for mutants that displayed a phenotype, n=2. B) Luciferase values normalised to the 0-h time point.
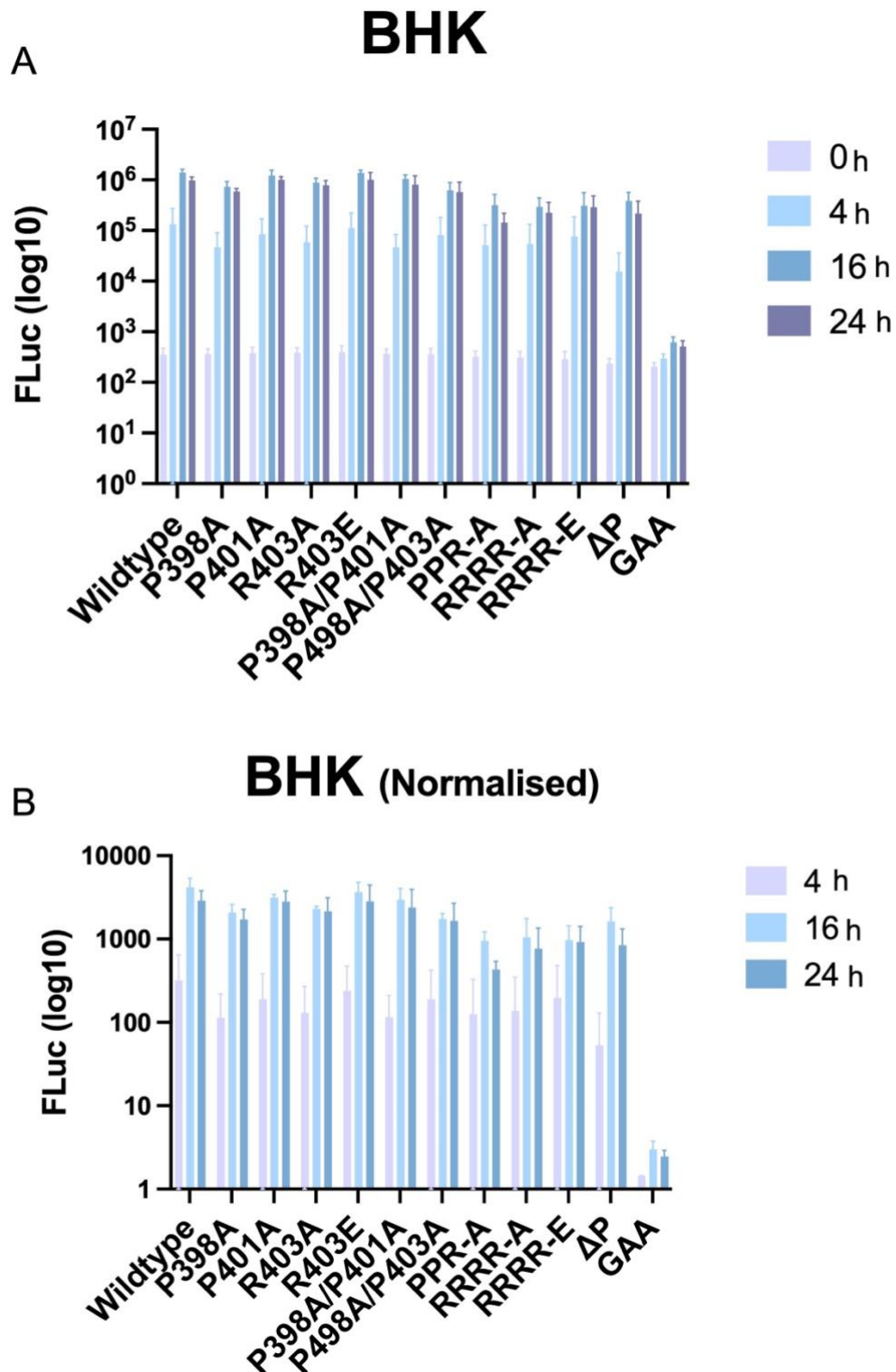
Figure 3.5 CHIKV polyproline motif mutant replication in BHK cells. A) Luciferase signal measured for BHK cells transfected with WT or mutant CHIKV-FLuc-SGR and harvested at 0, 4, 16 and 24 h. GAA denotes the replication inactive nsP4 mutant. Data analysed by ordinary Two-way ANOVA with Dunnett's multiple comparison test compared to WT at 24 h, **** = $\leq$ 0.0001. Statistical significance is only shown for mutants that displayed a phenotype, n=3. B) Luciferase values normalised to the 0-h time point.
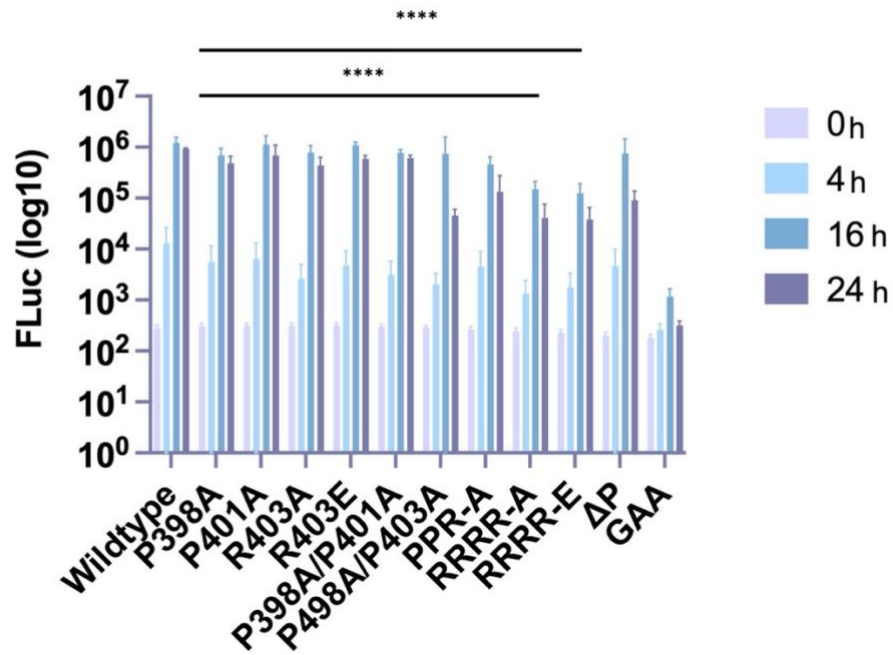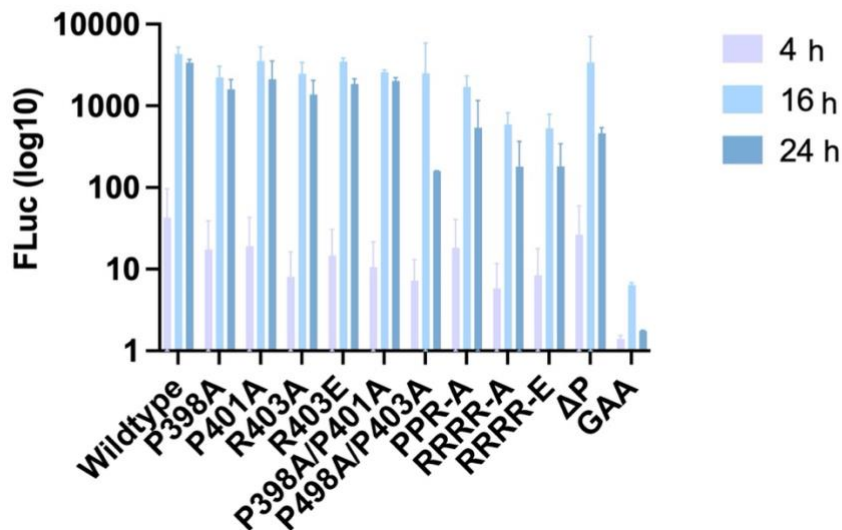
Figure 3.6 CHIKV polyproline motif mutant replication in C2C12 cells. A) Luciferase signal measured for C2C12 cells transfected with WT or mutant CHIKV-FLuc-SGR and harvested at 0, 4, 16 and 24 h. GAA denotes the replication inactive nsP4 mutant. Data analysed by ordinary Two-way ANOVA with Dunnett's multiple comparison test compared to WT at 24 h, **** = ≤ 0.0001. Statistical significance is only shown for mutants that displayed a phenotype, n=2. B) Luciferase values normalised to the 0-h time point.

3.2.4 Phenotypic analysis of CHIKV polyproline mutants in mosquito cell lines

CHIKV is transmitted to animals and humans via an infected *Aedes* mosquito, thus the virus must be able to replicate in both mammalian and mosquito cells. To investigate whether mutations in the polyproline motif has any effects on CHIKV replication in mosquitoes, the panel of mutants was then evaluated in two *Aedes albopictus* cell lines: C6/36 and U4.4 which have both been previously shown to support robust CHIKV replication (Roberts et al., 2017). The main difference between these two cell lines is that C6/36 cells are defective in the RNAi response due to a frameshift mutation in the Dcr2 gene (Morazzani et al., 2012). CHIKV maintains slower replication and reduced cytopathology in mosquitoes compared to mammals, likely due to the need of the vector in transmission. There is also evidence that slower replication kinetics play an important role in determining virus escape from anatomical tissues (Merwaiss et al., 2021). As such, peak replication in mosquito cells is observed at 48 h rather than 24 h and the timepoints taken in the mosquito cell luciferase assays were adjusted accordingly. Nevertheless, in C6/36 cells (Figure 3.7A), WT CHIKV-FLuc-SGR can replicate to similar levels as observed in mammalian cells. Interestingly, whilst all other mutants could replicate well, mutant RRRR-A had a modest but significant reduction (1 log) in replication levels which is further highlighted in the normalised data.

In comparison, as shown in Figure 3.8A, U4.4 cells appear to be less permissive for CHIKV replication and the FLuc levels of WT was one-log lower than all other tested cell lines. Most of the mutants can replicate to WT levels even though some displayed very low luciferase signal at 16 h (R403E, P398A/R403A, PPR-A). Mutant RRRR-A did not produce any replication signal until the 48-h time point, where levels reached a log lower than WT. The ΔP mutant also produced a slight but significant reduction in replication levels. The RRRR-E mutant was observed in multiple other cell lines to have a slight reduction in luciferase signals. However, in U4.4 cells, this mutant appears to be replication deficient and produced luciferase signals like the nsP4 GAA mutant.
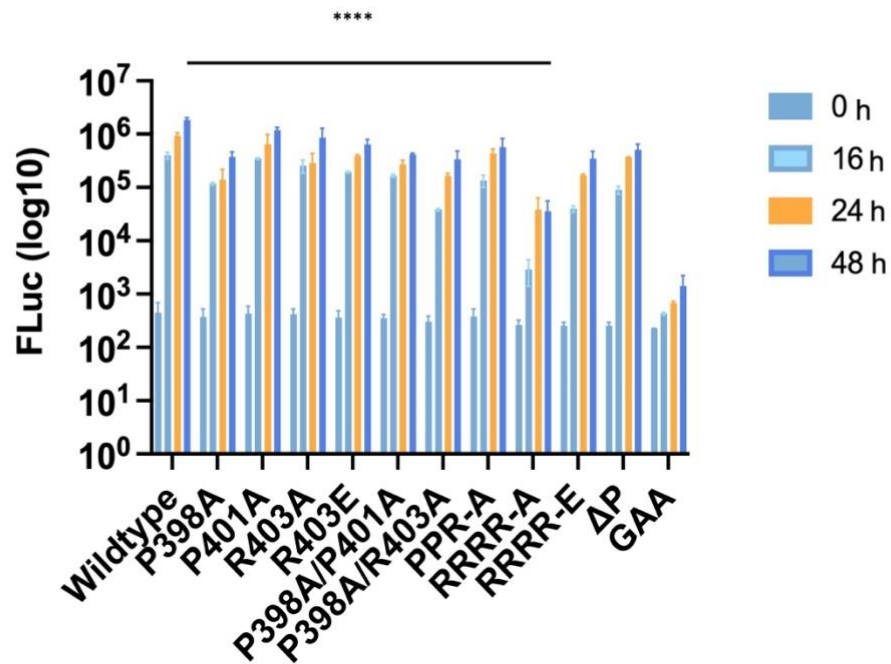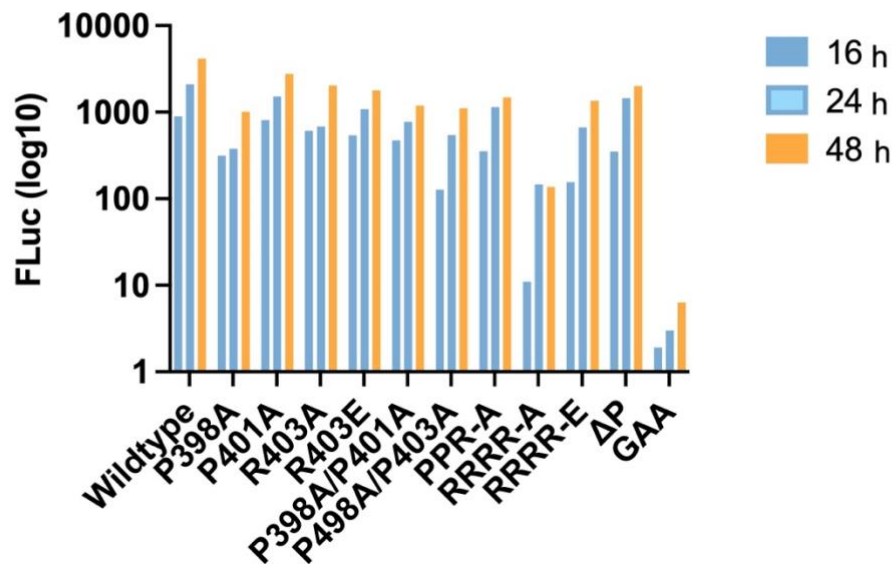
Figure 3.7 CHIKV polyproline motif mutant replication in C3/36 cells. A) Luciferase signal measured for C6/36 cells transfected with WT or mutant CHIKV-FLuc-SGR and harvested at 0, 16, 24 and 48 h. GAA denotes the replication inactive nsP4 mutant. Data analysed by ordinary Two-way ANOVA with Dunnett's multiple comparison test compared to WT at 48 h, **** = ≤ 0.0001. Statistical significance is only shown for mutants that displayed a phenotype. B) Luciferase values normalised to the 0-h time point.
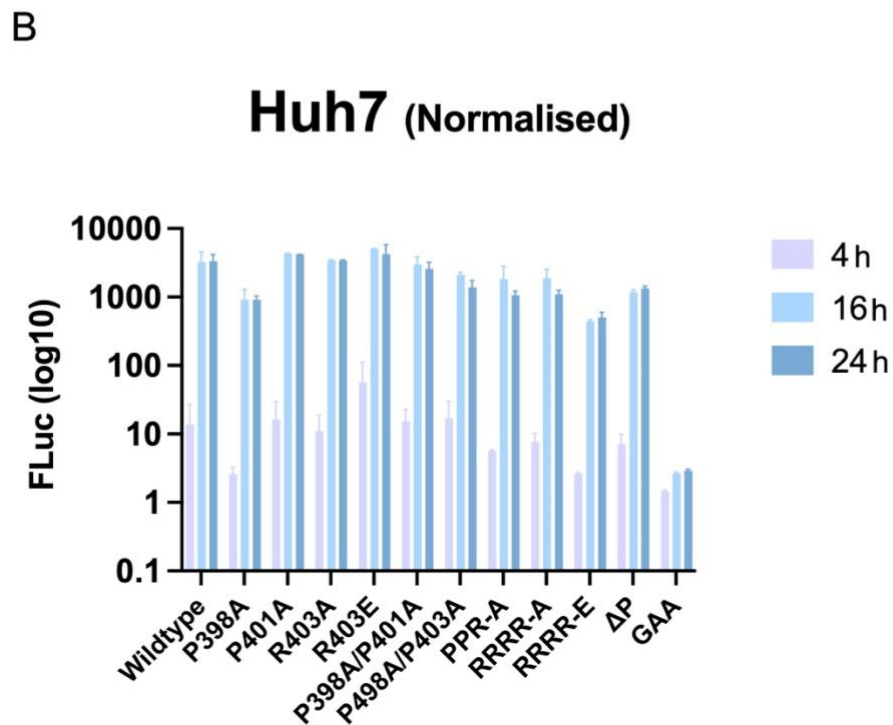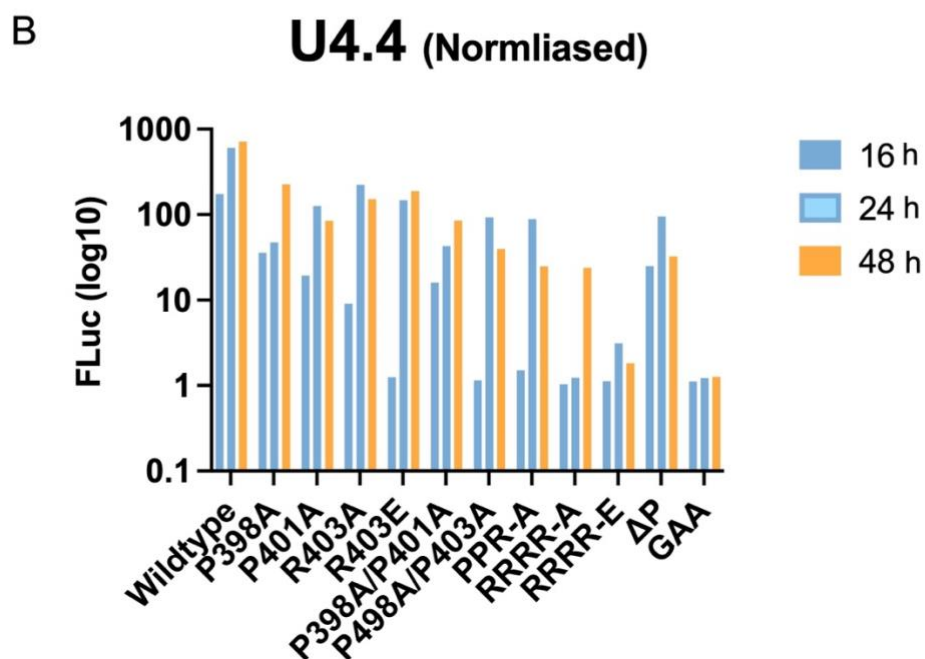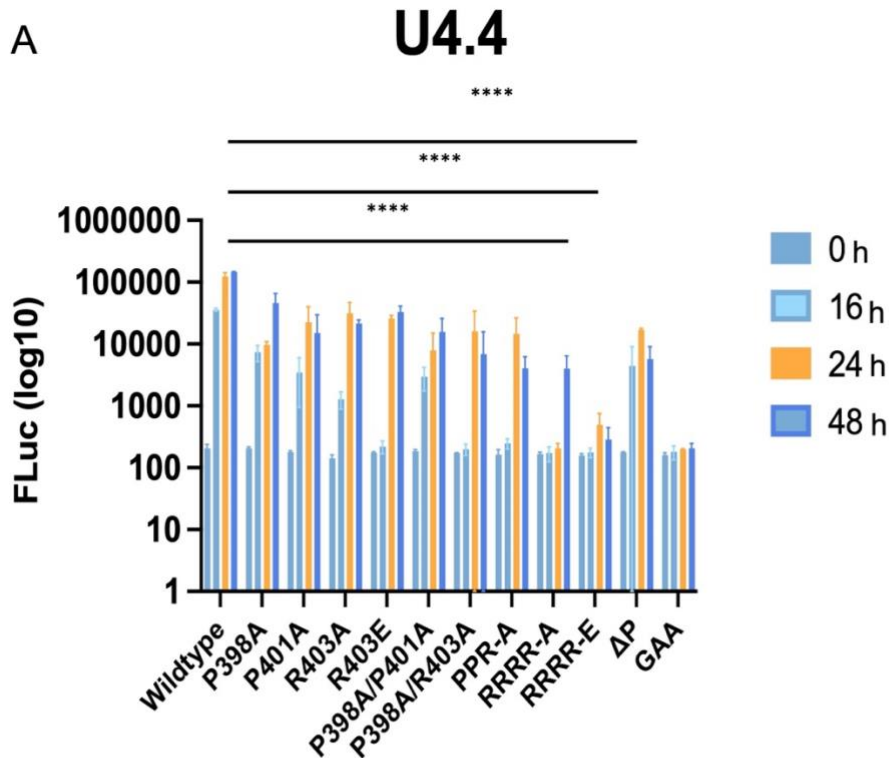
Figure 3.8 CHIKV polyproline motif mutant replication in U4.4 cells. A) Luciferase signal measured for U4.4 cells transfected with WT or mutant CHIKV-FLuc-SGR and harvested at 0, 16, 24 and 48 h. GAA denotes the replication inactive nsP4 mutant. Data analysed by ordinary Two-way ANOVA with Dunnett's multiple comparison test compared to WT at 48 h, **** = ≤ 0.0001. Statistical significance is only shown for mutants that displayed a phenotype. B) Luciferase values normalised to the 0-h time point.

3.2.5 The P398A/P401A mutation prevents production of infectious virus

To determine whether the polyproline motif mutants have impacts on other stages of the virus lifecycle following replication, the panel of mutations were introduced into the full-length infectious virus (ICRES-CHIKV) using an *Age*I restriction site that was previously cloned in by Dr Raymond Li (past member of the Harris research group, University of Leeds). The nsP4 GAA mutant was also introduced into ICRES-CHIKV as a negative control. *In vitro* transcribed RNA was prepared and electroporated into BHK cells. 24 h post electroporation, the supernatant containing virus was collected and quantified by serial dilution in a plaque assay. 48 h following virus infection, the methylcellulose overlay media was removed, and the resulting cells were fixed and stained. As shown by Figure 3.9A, WT CHIKV was able to produce a high titre of infectious virus while the nsP4 GAA negative control failed to produce any infectious virus. Surprisingly, whilst most of the mutants were able to produce virus to the WT level, the P398A/P401A mutant failed to produce infectious virus even though it was able to replicate to WT levels in the luciferase assays in all tested cell lines. The RRRR-E mutant, which showed slightly reduced replication in luciferase assays also demonstrated a slight reduction in production of infectious virus. Taken together, the results show that mutant P398A/P401A can replicate to WT levels but is defective in virus assembly and release of infectious virus particles.
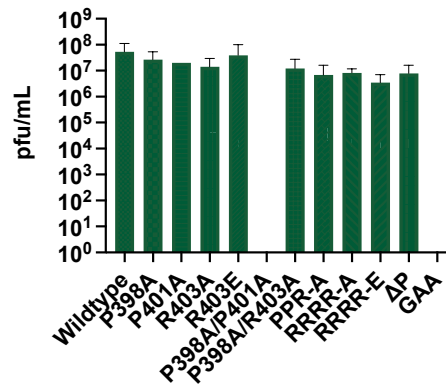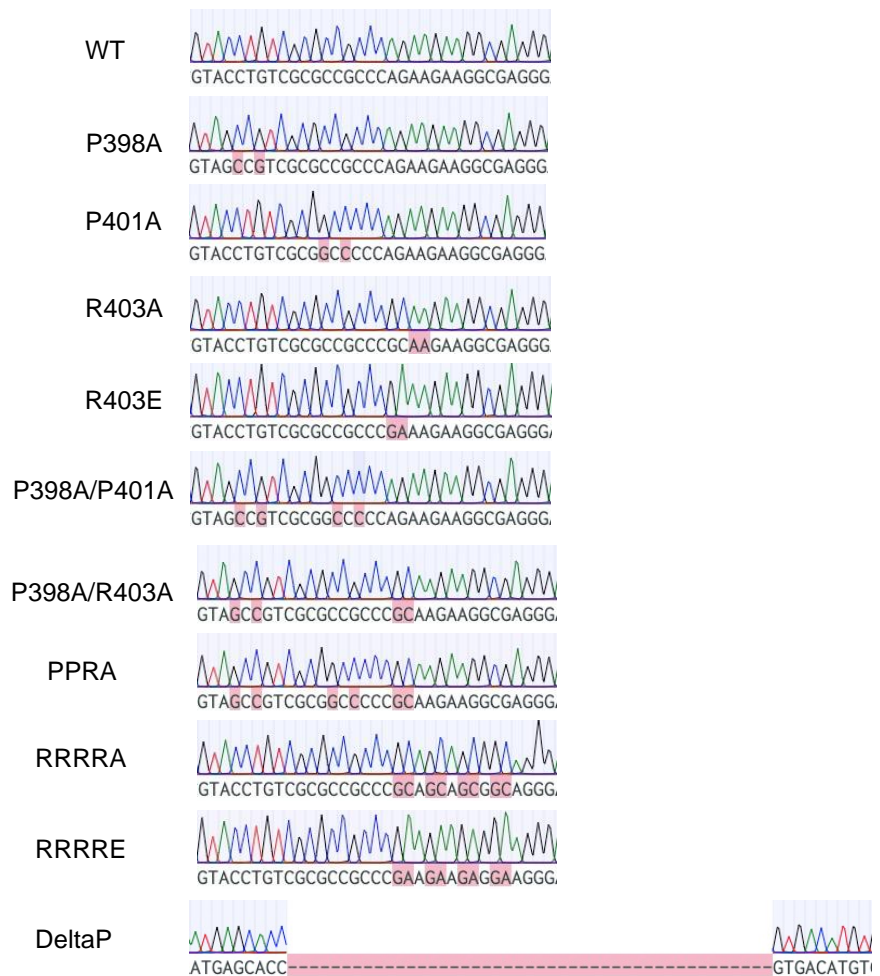
Figure 3.9 Phenotype of CHIKV nsP3 hypervariable domain mutants in BHK cells. A) Plaque assay titration of WT and polyproline motif mutants. Virus harvested from cells electroporated (24 h.p.e) with either WT or mutant ICRES-CHIKV are serially diluted and plated onto fresh cells. 48 h post infection, cells are fixed and stained, n=2. B) Sequencing results of RNA purified from cells electroporated with WT or mutant ICRES-CHIKV.

3.2.6 Sequence analysis of the ICRES-CHIKV-nsP3 RNA following virus production

Interestingly, whilst striking phenotypes were observed for the P398A/P401A mutant in infectious virus production and viral protein expression, none of the other mutants displayed such marked changes. An explanation of this could be that the other mutations had simply reverted to WT following infection, a phenomenon which has been published in literature (Gao et al). To investigate whether this was the case, *in vitro* transcribed RNA was prepared and electroporated into BHK cells. 24h post electroporation, the cells were lysed with TRIzol, and total RNA was purified. The resulting RNA was then reverse transcribed into cDNA and a region containing the polyproline motif mutations was amplified by PCR then subjected to sequence analysis. At the 24-h time point, no reversions were observed for any of the mutants and the expected nucleotide changes remained (Figure 3.9B). This confirms that the phenotypes observed in the infectious virus assays are due to the mutations present in the polyproline motif.

3.2.7 The P398A/P401A mutation reduces viral protein production

As a phenotype was observed for two of the mutants in the virus quantification experiments, viral protein expression levels were next investigated. For this, *in vitro* transcribed RNA was prepared and electroporated into BHK cells. 24-h post electroporation, cell lysates were collected, and western blot analysis was performed using antibodies against two non-structural proteins: nsP1 and nsP3. Relative viral protein levels were normalised using actin as a positive control. Aligned to the results observed in the plaque assays, whilst most mutants expressed nsP1 to WT levels, mutant P398A/P401A displayed a 2-fold decrease (Figure 3.10A). This phenotype was repeated in nsP3 protein levels but at a 3-fold decrease (Figure 3.10B). A decrease in nsP3 protein expression was also observed in the P398A/R403A and PPR-A mutants whilst all other mutants expressed nsP3 to levels comparable to WT. So far, only non-structural proteins have been investigated, therefore, to see whether the polyproline motif mutants has any effects on the expression of structural proteins, western blot analysis was also performed using an antibody against

the capsid protein. Here, the results show that the expression of capsid is again reduced in the P398A/P401A mutant, at a fold decrease (Figure 3.10C). Mutants P398A/R403A and PPR-A also expressed less capsid protein compared to WT (Figure 3.10C). Together, the results indicates that the P398A/P401A mutation has impacts on initial protein translation and replication.
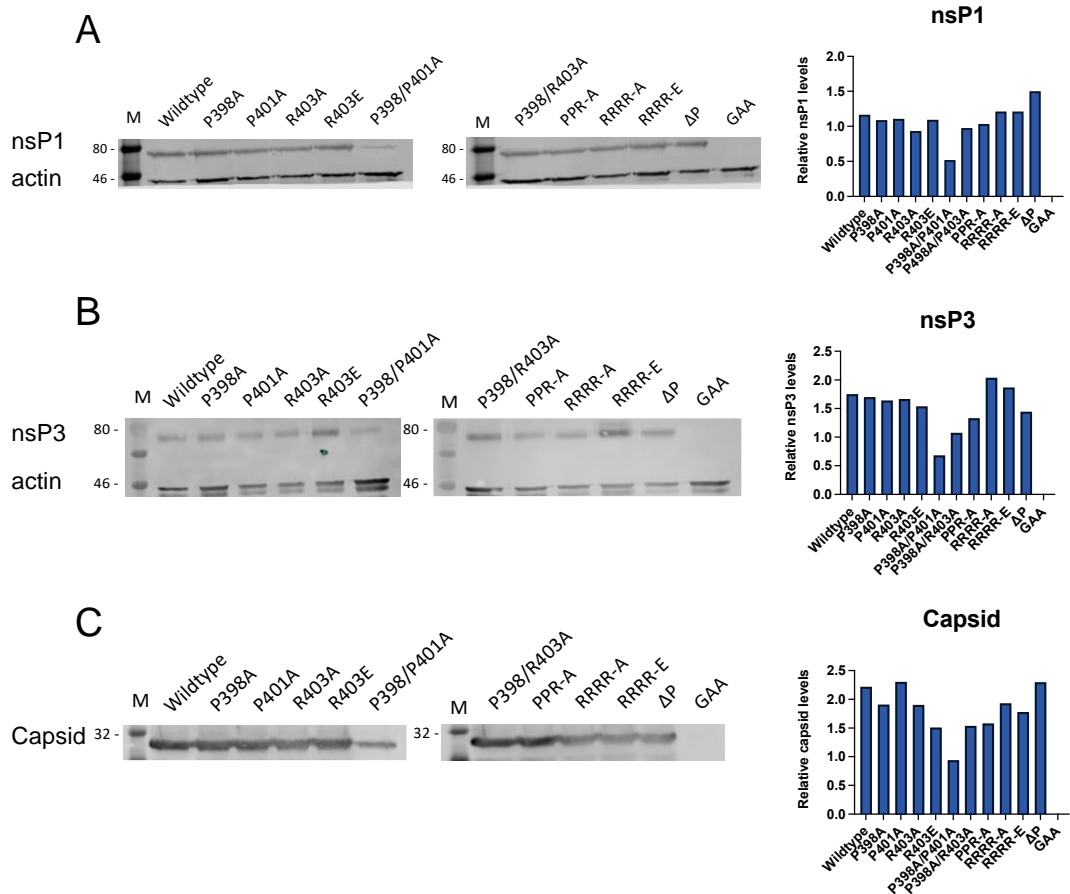
Figure 3.10 Western blot analysis of CHIKV viral proteins in BHK. Cells electroporated with WT or mutant ICRES-CHIKV are harvested 24 h post electroporation. A) Western blot analysis of cells lysates using an antibody against CHIKV nsP1. Graph represents relative nsP1 levels where the protein is normalised to an actin loading control. B) Western blot analysis of cells lysates using an antibody against CHIKV nsP3. Graph represents relative nsP3 levels where the protein is normalised to an actin loading control. C) Western blot analysis of cells lysates using an antibody against CHIKV capsid. Graph represents relative capsid levels where the protein is normalised to the actin loading control shown in A (nsP1 was resolved on the same gel as capsid).

3.2.8 A P398A/P401A mutation prevents virus assembly and release

As the P398A/P401A mutant displayed a clear phenotype, it was taken forward for further analysis. So far, it is not clear which stage of the virus lifecycle is impacted by the mutant. The P398A/P401A mutant cannot produce extracellular infectious virus but whether it can produce infectious virus that is not released into the media is unclear. To dissect this, an intracellular infectious virus production assay was performed. For this, RNA was electroporated into BHK cells as described in section 3.2.6. 24 h post electroporation, the supernatant was collected as the extracellular virus sample. The monolayer was then scraped into fresh media, collected, and subjected to three freeze-thaw cycles to lyse the cells. The resulting sample was collected as the intracellular virus sample. Both intracellular and extracellular virus was then serially diluted and quantified in plaque assays. 48 h later, the plaque assays were fixed and stained as described in section 3.2.6. As shown in Figure 3.11, WT ICRES-CHIKV displayed high viral load both intracellularly and extracellularly. The ICRES-CHIKV-nsP4-GAA negative control failed to produce virus in both samples as expected. The P398A/P401A mutant also failed to produce infectious virus, identical to the negative control. The results indicates that the P398A/P401A mutant is defective in producing intact virus particles that can infect cells.

BHK
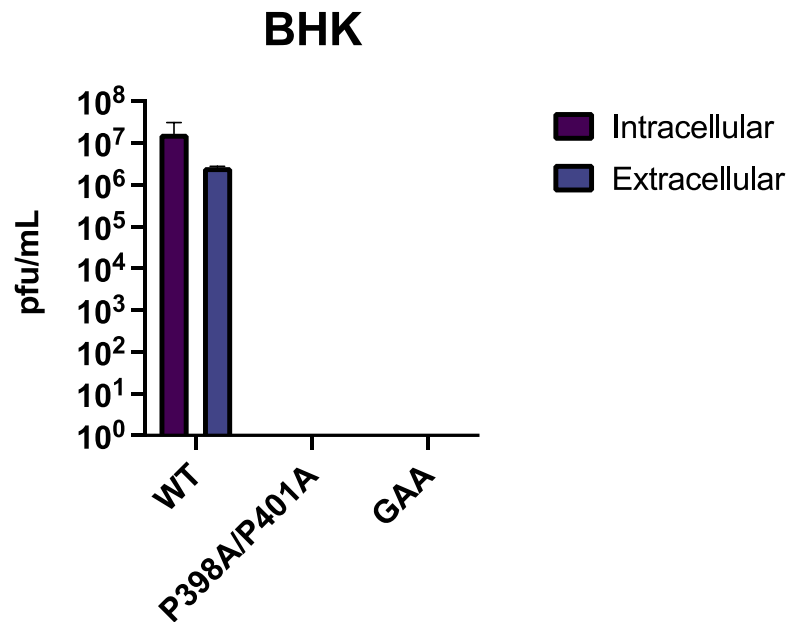
Figure 3.11 Analysis of infectious CHIKV intracellular/extracellular virus. WT or mutant ICRES-CHIKV virus harvested from cells 24 h post-electroporation are divided into two aliquots. One aliquot was then subjected to three freeze-thaw cycle at -80 °C. Plaque assay titration was then performed for both freeze-thawed and unfreeze-thawed viruses. 48 h post infection, cells were fixed and stained.

3.2.9 The P398A/P401A mutant impairs full length genome RNA production

Thus far, the P398A/P401A mutant displayed normal replication capability in the subgenomic replicon assays but defects in protein expression and production of infectious virus in infectious clone analyses. To determine whether the P398A/P401A mutant can replicate in the context of infectious virus, viral genomic RNA was quantified by qRT-PCR using primers against the E1 gene. WT ICRES-CHIKV and ICRES-CHIKV-nsP4-GAA were included as positive and negative controls, respectively. As expected, high levels of genome copies were detected for WT ICRES-CHIKV. ICRES-CHIKV-nsP4-GAA produced genome copies of 1×10^3 which is likely to be background noise rather than replication (Figure 3.12). For the P398A/P401A mutant, both tested input RNA concentrations were able to produce genome copies higher than the negative control suggested there is some level of replication which is aligned

to the results seen in the luciferase assays. Taken together, the results show that the P398A/P401A mutant can replicate but is unable to form infectious virus particles.

BHK
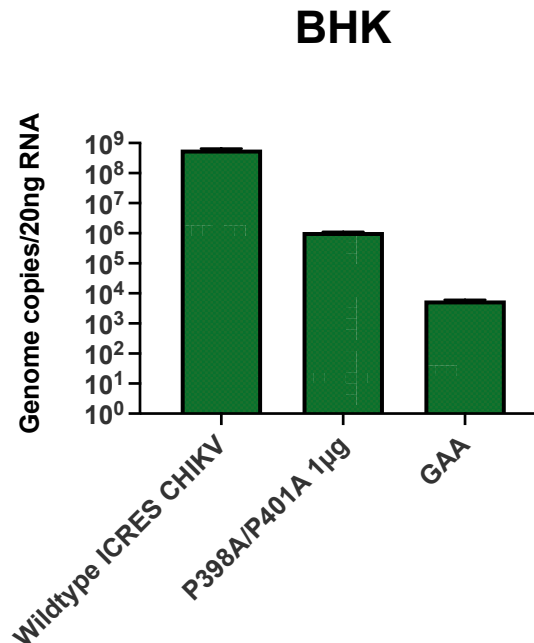
Figure 3.12 qRT-PCR analysis of cells electroporated with multiple ICRES-CHIKV-nsP3-P398A/P401A RNA. BHK cells were electroporated with 1 µg of WT ICRES-CHIKV RNA, 1 µg of ICRES-CHIKV-nsP4-GAA RNA, 1 µg of the ICRES-CHIKV-nsP3-P398A/P401A RNA 24 h post electroporation, the virus was harvested with TRIzol. RNA was then purified and reverse transcribed then analysed using qRT-PCR.

3.2.10 Observed phenotypes are not a result of the nsP3 polyproline motif P398A/P401 mutation

Occasionally, mutations can arise from PCR or restriction digest cloning leading to nucleotide changes outside the desired area. To mitigate this, cloning of the polyproline motif mutations were originally performed in the WT CHIKV-FLuc-SGR. The insert containing the mutations were then digested from the cloning plasmid and re-ligated into a freshly prepared plasmid. This was performed for both the subgenomic replicon and the infectious clone. To confirm that the observed phenotypes with the P398A/P401A mutant were indeed due to these specific mutation, a due diligence check was performed. For this, WT ICRES-CHIKV and CHIKV-nsP3-P398A/P401A plasmids were

digested with *Age*I and *Spe*I restriction enzymes. The resulting backbone and inserts were then purified, and the WT insert was re-ligated to the CHIKV-nsP3-P398A/P401A backbone to generate a CHIKV-nsP3-P398A/P401A revertant clone (Figure 3.13A). The revertant is expected to display the WT phenotype as the P398A/P401A mutation has been replaced with the WT sequence. To investigate whether this was the case, a plaque assay was performed alongside a WT ICRES-CHIKV and ICRES-CHIKV-nsP4-GAA as positive and negative controls, respectively. As seen previously, WT ICRES-CHIKV and ICRES-CHIKV-nsP4-GAA displayed the usual phenotypes. On the other hand, the P398A/P401A revertant did not produce any infectious virus which indicates that the observed phenotypes thus far are not as a result of the polyproline motif mutations (Figure 3.13B).
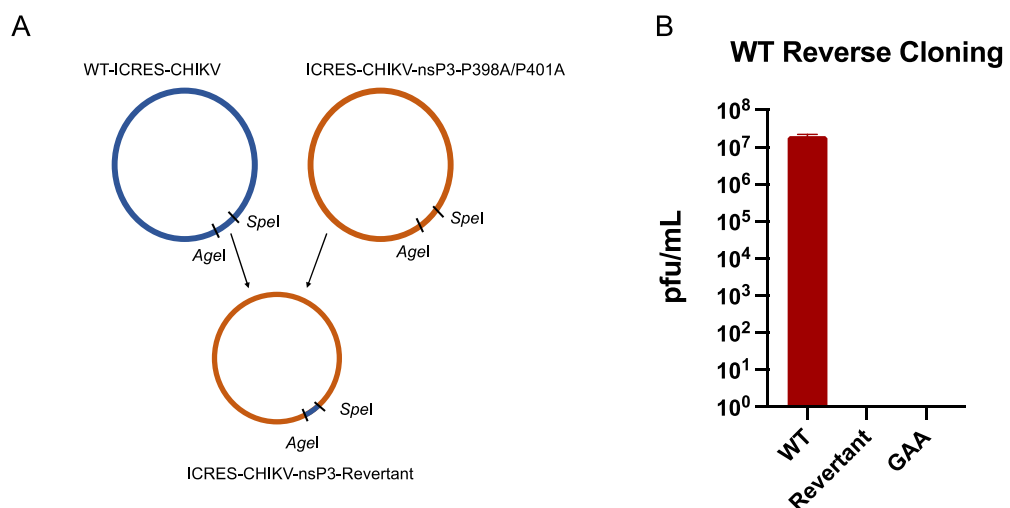
Figure 3.13 Reverse cloning of the ICRES-CHIKV-nsP3-P398A/P401A mutant. A) Schematic of the revertant cloning approach. WT ICRES-CHIKV and ICRES-CHIKV-nsP3-P398A/P401A DNA were digested with restriction enzymes SpeI and AgeI. The purified insert from WT ICRES-CHIKV was then ligated into the backbone of ICRES-CHIKV-nsP3-P398A/P401A to generate a revertant: ICRES-CHIKV-nsP3-Revertant. B) Plaque assay titration of WT ICRES-CHIKV, ICRES-CHIKV-nsP3-Revertant and ICRES-CHIKV-nsP4-GAA.

3.2.11 Identification of the mutation responsible for the observed phenotypes

From the revertant experiments, it was concluded that the phenotypes observed with the P398A/P401A mutant are likely due to alternate mutations outside the investigated region. To confirm that this is the case, sequencing was performed on RNA purified from cells infected with ICRES-CHIKV-nsP3-P398A/P401A. Specifically, 24 h post electroporation, cells were lysed with TRIzol and the RNA was purified then reverse transcribed using random hexamers. The cDNA was then sequenced using primers that cover the full-length virus genome. Unfortunately, it was not possible to sequence further than nsP4 using this sequencing method. Sequencing reactions for the structural genes were either non-specific, of poor quality or produced no priming (Figure 3.14A). This indicated that the mutation causing the observed phenotypes could lie after the gene encoding nsP4 and perhaps could have induced a frameshift event that prevents primer binding from this point onwards. The restriction enzymes used to digest the original insert during cloning span the HVD of nsP3 to the end of nsP4 region, as such it is plausible that the culpable mutation presents following the gene encoding nsP4. To dissect this, nanopore sequencing was used to overcome the issues observed with Sanger sequencing. Purified RNA was converted into two separate pools of 1200 nt DNA fragments using CHIKV specific primers which are then barcoded and sequenced. The sequenced amplicons were then PCR tiled together to build a consensus sequence and aligned to the ICRES-CHIKV reference genome. The results from nanopore sequencing are shown in Figure 3.14B. Both DNA pools were successfully amplified and could be aligned to form a consensus sequence. Compared to sanger sequencing, nanopore sequencing successfully covered most of the structural genes. However, sequencing terminated at nucleotide 11,286 thus it was not possible to get accurate sequencing results for the last 548 nucleotides. Two mutations were identified in nanopore sequencing, the first was the *Spe*I restriction enzyme insertion site and the second was the expected P398A/P401A mutation.

Using the data from Sanger and nanopore sequencing results, a primer was designed targeting the remaining 548 nt region and sequencing was performed again. The first 26 nts of the 548 region spans E1, but sequencing did not identify a mutation in this protein suggesting that the phenotype-causing mutations must lie in the 3′ UTR (Figure 3.15A). Indeed, four mutations were revealed, which were all C-U substitutions (Figure 3.15B & Figure 3.15C). Interestingly, the first two mutations were located close to E1, whereas the last two mutations were presented shortly before the 3′ conserved sequence element (CSE) immediately prior to the poly-A-tail. Taken together, it can be deduced that that the mutation causing phenotypes associated with the P398A/P401A mutant is due to four distinct nucleotide substitutions in the 3′ UTR.
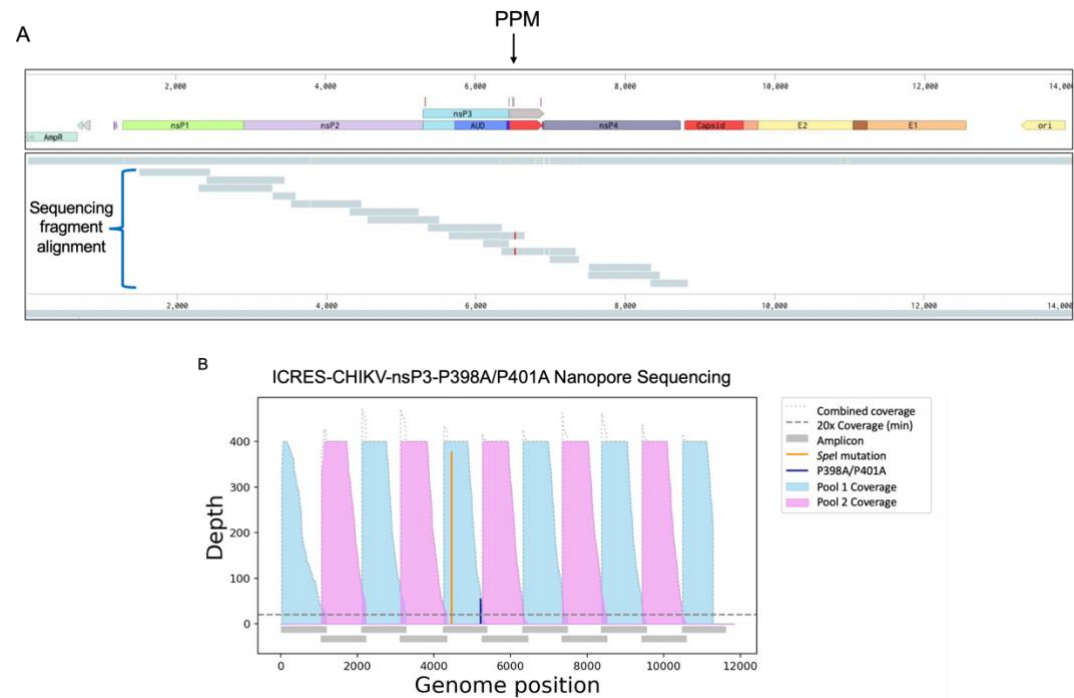
Figure 3.14 Sequence analysis of ICRES-CHIKV-nsP3-P398A/P401A RNA. A) Results of ICRES-CHIKV-nsP3-P398A/P401A sanger sequencing. RNA was purified from cells electroporated with ICRES-CHIKV-nsP3-P398A/P401A, reverse transcribed and sequenced with primers that cover the full-length genome. PPM denotes the location of the polyproline motif. The sequenced fragments are aligned to the reference WT ICRES-CHIKV genome. B) Results of ICRES-CHIKV-nsP3-P398A/P401A nanopore sequencing. RNA was purified from cells electroporated with ICRES-CHIKV-nsP3-P398A/P401A, reverse transcribed and sequenced with primer pairs to generate 1200 bp fragments which are then sequenced. The sequenced amplicons are used to build a consensus sequence which is then aligned to the CHIKV reference genome.

95
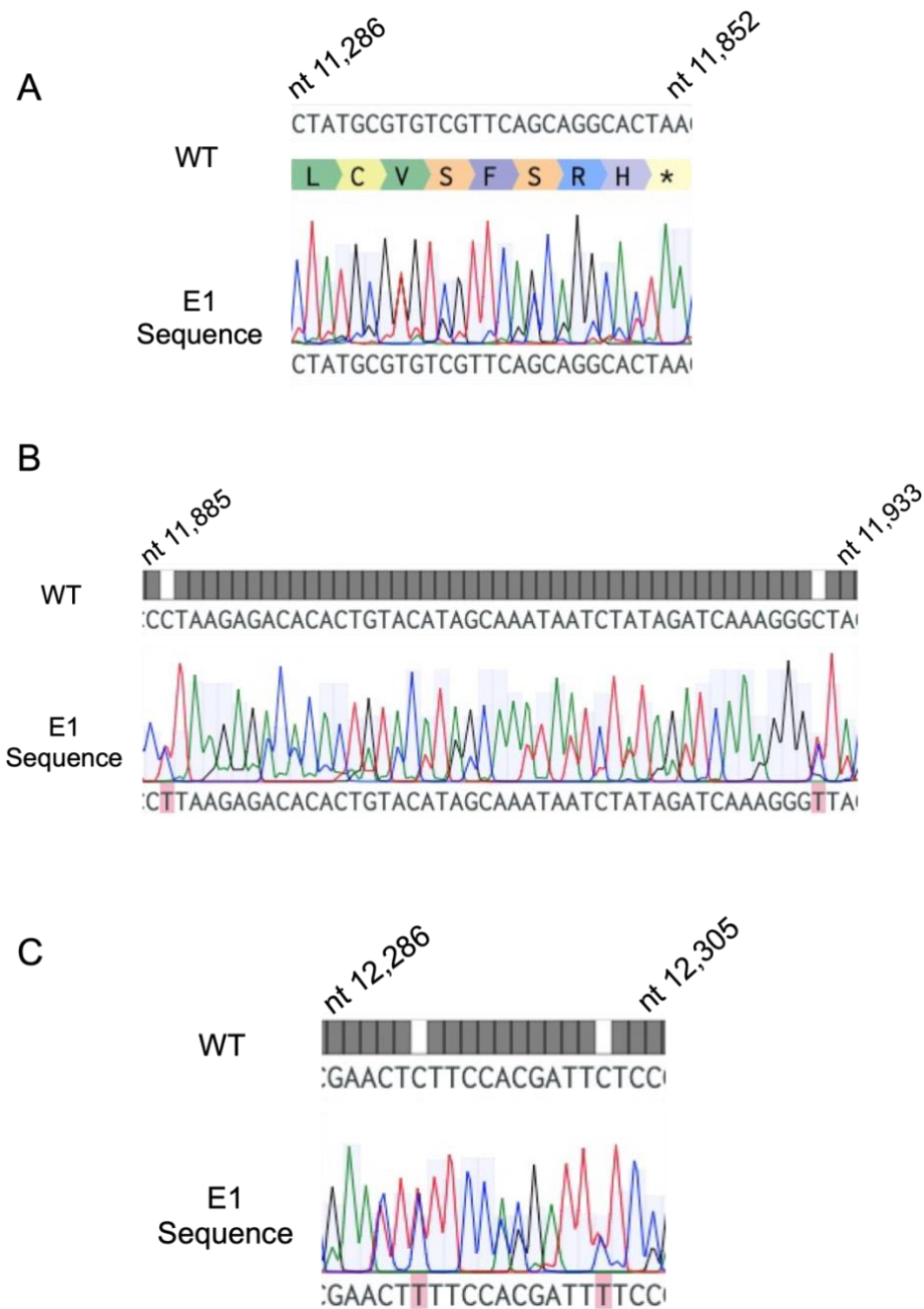
Figure 3.15 Sanger sequencing results of ICRES-CHIKV-nsP3-P398A/P401A E1 and 3′ UTR. A) Sequencing results of the E1 protein. B) Sequencing result of the 3′ UTR, the first two C-to-T identified mutations are highlighted in red (C-to-U in RNA). C) The last two mutations identified in the 3′ UTR, highlighted in red.

3.3 Discussion

3.3.1 The nsP3 HVD polyproline motif is not essential in the CHIKV lifecycle

Although the CHIKV nsP3 HVD polyproline motif is well conserved amongst alphaviruses and individual residues in this region have been shown to be important for amphiphysin binding and RNA replication in other members of this family, the importance of specific motif residues in the context of CHIKV replication has not been explored. This study aimed to perform a comprehensive analysis of individual residues in the CHIKV polyproline motif, to interrogate whether single or combination mutations can impact replication, protein expression or production of infectious virus. The data presented here demonstrates that in all tested cell lines, including mammalian and mosquito, the polyproline motif mutants were capable of replicating to some degree with most able to replicate to WT levels, indicating that this motif is not essential in the CHIKV lifecycle. This has been previously observed in the related hepatitis C virus (HCV); whilst deletion of a proline rich region disrupted interactions with multiple members of the SH3-domain containing Src family of tyrosine kinases, it was determined to be nonessential for RNA replication (Macdonald 2004, Macdonald 2005, Tellinghusen 2008).

In SINV, a single charge reversing mutation R426E completely abolished binding to amphiphysin, reduced RNA synthesis and infectious virus production in BHK cells. The observed phenotypes were especially pronounced in HeLa cells, indicating a difference in infection permissivity (Neuvonen et al., 2011). In CHIKV, a corresponding mutation R403E was generated to investigate whether similar effects can be observed. From the results, this residue clearly does not have the same significance for CHIKV replication, production of infectious virus particles or protein expression as it does for SINV, at least in BHK cells. Whilst a range of cell lines were used to investigate mutant replicative capacity, HeLa cells were not included therefore it would be interesting to perform the same experiments using this cell line to identify whether the R403E mutation displays cell-specific phenotypes.

For SFV, binding to amphiphysin is mediated by a 34 amino acid region spanning across the canonical polyproline motif and an additional polyproline motif further downstream as the deletion of this region reduced interaction with the cellular protein and decreased RNA synthesis (Neuvonen et al., 2011). To recapitulate this mutation, a 15 amino acid deletion (ΔP) was generated as CHIKV only harbours one single polyproline motif, rather than a double seen in SFV. For CHIKV, the targeted polyproline region is not essential for replication in all tested cell lines except U4.4 where the deletion produced a slight but significant reduction in replication levels. However, this could be due to the less permissive nature of this cell line rather than a specific effect caused by the deletion. Regardless, the ΔP mutation was able to express viral proteins and produce infectious virus to levels of WT. During this project, a publication also studying the CHIKV HVD proline rich region became available (Mutso et al., 2018). In this study, two deletions targeting 9 and 16 amino acids from P398 were made and investigated for RNA infectivity and replication capacities. Aligned to the phenotypes observed for the ΔP mutation, the 9 amino acid deletion did not affect virus replication or production of infectious virus. Interestingly, deletion of 16 amino acids was observed to result in ~7-fold decrease in infection. The authors speculated that this could be due to the 7 amino acids following the canonical PxxPxR motif acting as a short linear interaction motif. To compare, the ΔP mutation and the 16 amino acid deletion constructed by *Mutso et al* overlap mostly except for two amino acids: V412 and T413. Combining the results presented here and published by *Mutso et al* suggests that these last two amino acids have some significance for the virus lifecycle.

In CHIKV, previous studies have shown that the last arginine in a string of arginine residues (highlighted in bold PVAPPRRR**R**) following the canonical PxxPxR polyproline motif is essential for binding to amphiphysin. Substitution of this arginine for an alanine residue resulted in a 10-fold decrease in binding affinity. Further to this, ITC analysis revealed that all the CHIKV polyproline motif C-terminal arginine residues are imperative for amphiphysin affinity despite the fact this region is highly disordered and do not present a specific

conformation (Tossavainen et al., 2016). This indicates that it is the positive charge of the residues that are important for amphiphysin binding rather than their structural properties. Furthermore, in HCV a corresponding cluster of quadruple basic residues following the polyproline motif were shown to be essential for production of infectious virus (Zayas et al., 2016). As such, to investigate whether these residues are important in the CHIKV lifecycle, two mutants were generated: One where the four terminal arginines are substituted for alanine residues to abrogate the positive charge (RRRR-A), and the second where they are substituted with glutamic acid to reverse the charge (RRRR-E). For replication, the RRRR-A mutant displayed a slight but significant reduction in RD and C2C12 cells at the 24-h time point, and in C6/36 and U4.4 cells at the 48-h time point. Interestingly, this mutant did not replicate at all in U4.4 cells at earlier time points. Similarly, the RRRR-E mutant also showed significant reduction in luciferase levels in several, but not identical cell lines: Huh7, C2C12 and RD cells at the 24-h time point. In U4.4 cells, the RRRR-E mutant failed to replicate at all. The results indicate that removing the positive charge of the arginine residues has some minor effects to virus replication, which is cell-type dependent. It is plausible to speculate that this could be the result of interrupting interactions with amphiphysin, and further research should focus on the binding capacity of the mutants to this cellular protein following infection. Nevertheless, both mutants can replicate, express viral proteins, and produce infectious virus but subtle phenotypes could be masked by the highly permissive nature of BHK cells. Whilst mutations in the CHIKV proline motif displayed similar phenotypes to corresponding mutations in the HCV polyproline motif, the importance of the basic cluster for HCV was not observed in CHIKV.

3.3.2 Mutation outside of the polyproline motif negatively affect CHIKV protein expression and infectious virus particle assembly

The studies undertaken has demonstrated that the polyproline motif is not essential in the CHIKV lifecycle. However, it has highlighted alternate mutations that is required for CHIKV. Four mutations were identified in the 3′ UTR, which were all C-to-U substitutions. The alphavirus 3′ UTR varies greatly in length between members of this genus and between strains within a single virus with roles described in replication, interaction with cellular proteins and miRNA. In CHIKV, a shorter 3′ UTR in the Asian lineage has been linked to a reduced fitness cost when compared to the ECSA lineage. *In vivo* experiments demonstrated that a chimeric Asian lineage CHIKV carrying the 3′ UTR of the ECSA lineage CHIKV can outcompete the WT Asian lineage CHIKV whilst a chimeric ESCA lineage CHIKV carrying the 3′ UTR of the Asian lineage CHIKV showed lower fitness than WT ESCA CHIKV (Chen et al., 2013). Two of the mutations identified in the 3′ UTR were located near the E1 protein, within the first direct repeat (DR). Direct repeats are short sequence repetitions found in the 3′ UTR which are relatively conserved within CHIKV lineages. This conservation suggests a lineage-specific functional significance (Filomatori et al., 2019). However, the importance of specific nucleotides in DRs are not well understood, therefore it is not immediately clear how these two identified mutations can contribute to the observed phenotypes. Conversely, the last two mutations were found near the highly conserved CHIKV 3′ CSE (the final ~20 nts before the poly-A-tail essential for initiation of minus-strand synthesis and presumed to contain the minus-strand promoter (Rupp et al., 2015). Negative strand synthesis is required to produce both the full-length genome and the 26S RNA genome which is then translated into structural proteins. Therefore, it is plausible that the latter two C-to-U substitutions observed in the P398A/P401A mutant could influence the function of the 3′ CSE, thereby perturbing structural protein synthesis required for virus particle assembly, and lead to loss of infectious virus production.

Chapter 4 Investigating the RNA binding capacity of nsP3

4.1 Optimisation of iCLIP for nsP3 and binding to the CHIKV genome

4.1.1 Introduction

The alphavirus replication complex is composed of four non-structural proteins: nsP1, nsP2, nsP3 and nsP4. As they function to synthesise genomic and subgenomic RNA, these proteins must bind and directly interact with RNA during the alphavirus lifecycle, which been demonstrated for nsP1, nsP2 and nsP4. nsP1 is a methyltransferase and a guanylytransferase and acts as a viral nuclear pore complex to ensure successful capping of newly synthesized viral RNAs (Jones et al., 2021). nsP2 possesses multiple enzymatic activities acting as a helicase, triphosphatase and protease, which all require binding to the CHIKV genome. Furthermore, nsP2 has been proposed to act as a transcription factor, binding to the subgenomic promoter to initiate synthesis of the subgenomic RNA (Rupp et al., 2015). Whilst all the non-structural proteins are required for RNA synthesis, the RNA synthetic property of the replication complex is solely afforded by nsP4, the RNA dependent RNA polymerase.

The inherent functions of the three non-structural proteins described above means that direct viral RNA binding is required for exertion of their enzymatic activities. On the other hand, although it is a member of the replication complex, the specific RNA binding capacity of nsP3 remains relatively unexplored. Individual nsP3 domains have been demonstrated to interact with RNA: the macrodomain has DNA & RNA binding capacity, but the principal function of this domain is to bind and hydrolyse mono-ADP-ribose (Malet et al., 2009). Fragment-based approach and X-ray crystallography studies identified that pyrimidone inhibitors, which are derivatives of nucleobases, localise to the ribose binding site of the macrodomain. This suggests that the nsP3 macrodomain mono-ADP-ribose binding site also has specificity for DNA/RNA molecules (Zhang et al., 2021). The RNA binding activity of the nsP3 AUD was first described by *Shin et al.*, where it was proposed that a patch of basic amino acids close to the zinc coordination site of the AUD has RNA-binding potential. This observation was made as zinc metalloproteins are frequently responsible

for DNA/RNA binding, and gene regulation (Shin et al., 2012). Experimentally, a double mutation in the AUD, P247A/V248A, has been shown to cause defects in synthesis of subgenomic RNA, in parts due to the reduced ability to bind the subgenomic promoter. Furthermore, using a twin-strep-tag system, the double mutation was also demonstrated to have a 100-fold reduction in binding to genomic RNA (Gao et al., 2019).

To date, RNA binding studies of specific nsP3 domains have mostly been performed using bacterially expressed protein rather than in the context of virus infection. Additionally, the experiments were performed in *in vitro* settings. Therefore, to investigate how nsP3 binds to CHIKV and/or host cell RNA *in vivo*, individual nucleotide UV-crosslinking and immunoprecipitation (iCLIP) was employed. Using CLIP techniques to study protein-RNA interactions was first described by *Wang et al.*, where UV irradiation followed by immunoprecipitation and subsequent cDNA library construction allow for high-throughput sequencing of RNA crosslinked to proteins *in vivo* (Wang et al., 2009). However, the nature of reverse transcriptase stalling due to a peptide left behind at the site of cross-linking meant that up to 80% of cDNA can be lost during downstream PCR amplification. Further optimisation by *König et al* led to the establishment of iCLIP, which addresses this problem via a circularisation step to introduce PCR adapters to both ends of the target cDNA (König et al., 2012). Furthermore, the introduction of an infrared dye conjugated adapter removed safety issues surrounding radioisotope labelling of protein-RNA complexes for visualisation. As such, iCLIP was chosen to study the RNA binding capability of CHIKV nsP3.
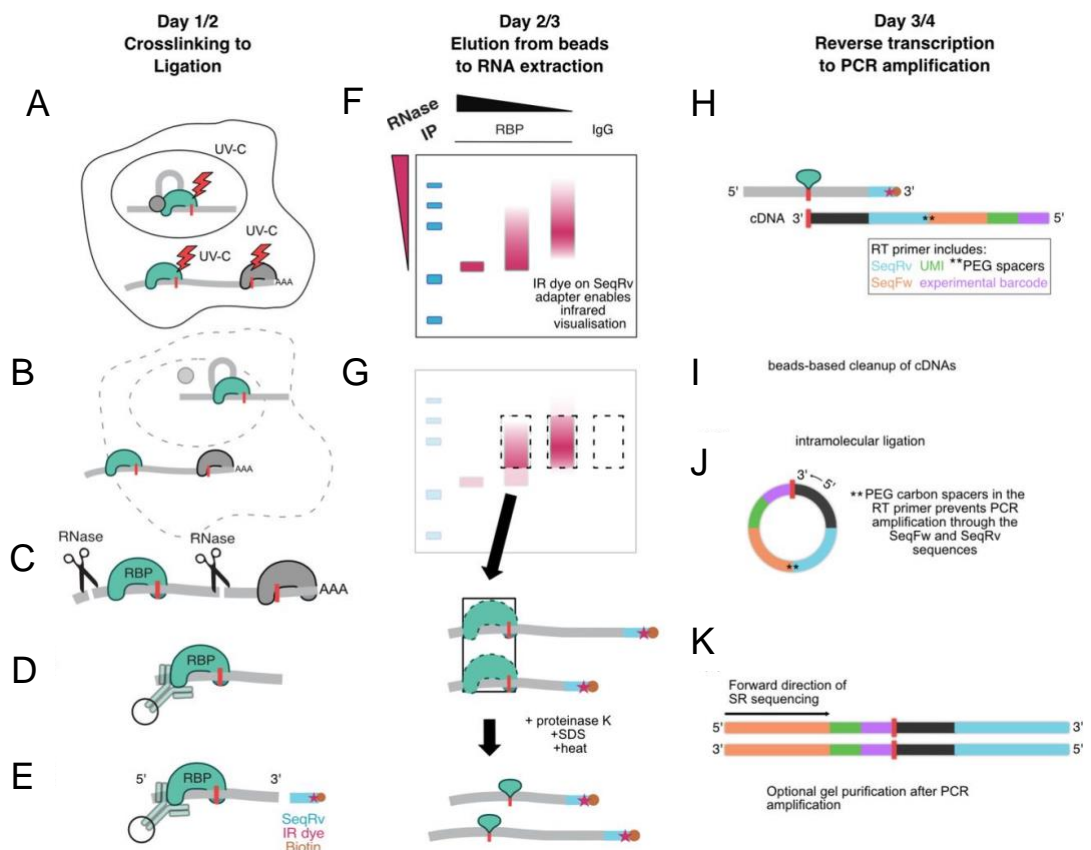
Figure 4.1 iCLIP protocol overview. A) Virus infected cells are first treated with UVC at 254 nm B) Cell were then lysed and sonicated to release protein-RNA complexes. C) Cell lysates were treated with RNase I to fragment RNA. D) Magnetic protein G beads coupled to appropriate antibodies were then used to purify target protein-RNA complexes from the lysates. E) Bound RNA was then dephosphorylated at the 3′ end and an infrared dye conjugated adapter was ligated to allow complex visualisation. F) Protein-RNA complexes were resolved using the NuPAGE system. G) Desired complexes were then excised from the gel and treated with proteinase K to digest away the protein, leaving a small peptide at the site of crosslinking. H) Bound RNA was then reverse transcribed using a reverse transcription primer which contains sequences required for sequencing and a barcode flanked by distinct unique molecular identifiers. I) cDNA libraries were then purified using AMPure XP magnetic beads. J) Purified cDNA products were circularised to introduce sequencing nucleotides to the 5′ end of the cDNA. K) PCR amplification re-linearises the cDNA, generating libraries for sequencing. Figure obtained from Lee et al., 2021.

The iCLIP protocol, as shown in Figure 4.1, begins with virus infection of a monolayer followed by UVC irradiation to crosslink proteins to RNA (Figure 4.1A). The cells are then lysed and sonicated to fragment large fragments of cellular DNA which can trap target protein-RNA complexes (Figure 4.1B). The lysates are then treated with RNase I which cleaves RNA into smaller fragments (Figure 4.1C), important for downstream bioinformatics genome alignment. Magnetic protein G beads coupled with appropriate antibody is then used to immunoprecipitate target protein-RNA complexes from cell lysates (Figure 4.1D). To facilitate visualisation of the complexes, the RNA is then 3′ dephosphorylated and ligated with an infrared dye conjugated adapter (Figure 4.1E). The complexes are then resolved using a NuPAGE gel and transferred onto a nitrocellulose membrane (Figure 4.1F). Following stringent quality control, the desired protein-RNA complexes are then excised and treated with proteinase K to digest most of the protein, leaving a peptide at the site of crosslinking (Figure 4.1G). Reverse transcription of the bound RNA uses a primer which contains the nucleotides required for sequencing and a sample-specific barcode flanked by distinct unique molecular identifiers (Figure 4.1H). The cDNA products generated are then purified using Ampure XP magnetic beads. To introduce the forward sequencing adapter to the 5′ end, the cDNA is then circularised (Figure 4.1J). PCR amplification then proceeds from 5′ to 3′, generating a linear product for sequencing (Figure 4.1K). Appropriately sized cDNA libraries are then sequenced, and bioinformatics analysis is used to identify the protein binding sites.

4.1.1.1 Aims

The overall aim of this chapter is to establish a robust iCLIP experiment for the investigation of nsP3-RNA interactions. Firstly, individual laboratory-based steps of the iCLIP protocol were interrogated to demonstrate whether the technique can be appropriately applied to nsP3. The computational bioinformatics architecture was then subjected to the same examination to ensure that any observed crosslink/binding sites are accurate and precise. Finally, identified nsP3 binding sites on the CHIKV genome are explored with the aim of detecting the determinants of RNA binding.

4.1.2 Results

4.1.2.1 Optimisation of TST-nsP3 expression

An integral part of the iCLIP technique is the immunoprecipitation of the target protein. To facilitate this, a Twin-Strep-Tag (TST) was previously cloned into the hypervariable domain of nsP3 in our research group (Gao et al., 2019). First, it was important to demonstrate that the insertion of the TST does not adversely affect the virus. For this, *in vitro* transcribed RNA was generated for both WT ICRES-CHIKV and ICRES-CHIKV-nsP3-TST (referred to from hereafter as TST ICRES-CHIKV) then electroporated into BHK cells. 24 h post electroporation, the virus was collected and quantified using plaque assay. As shown in Figure 4.2A, the introduction of a TST in the HVD of nsP3 does not negatively impact the virus as there is no significant difference observed between this and WT ICRES-CHIKV. Next, to capture the point where nsP3-RNA binding is at the highest, an infection optimisation was performed. It was hypothesised that maximum nsP3-RNA binding should correlate with the expression level of nsP3 thus the ideal combination of time points and multiplicity of infection (MOI) was determined. From the results, a 12h post infection time point at an MOI of 1 showed peak nsP3 protein expression (Figure 4.2B). Interestingly, from the results shown in Chapter 3, whilst peak replication was determined to be at 24h post electroporation, the 24h time point in virus infection did not produce a higher nsP3 expression.
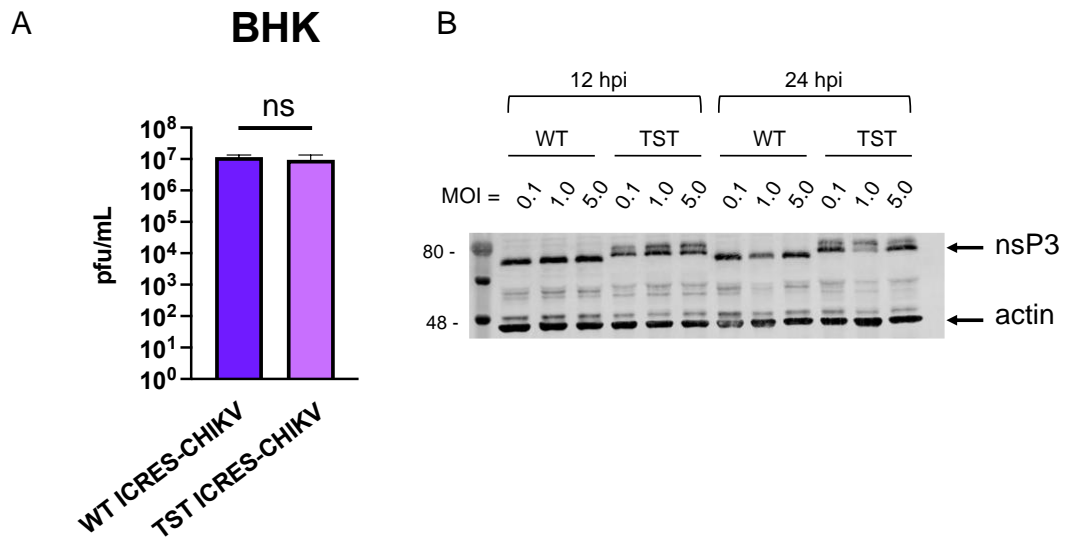
Figure 4.2 Comparison of WT ICRES-CHIKV and TST ICRES-CHIKV. A) Plaque assay titration of WT ICRES-CHIKV and TST ICRES-CHIKV in BHK cells. Virus harvested from cells electroporated (24 h.p.e) with either WT or TST ICRES-CHIKV are serially diluted and plated onto fresh cells. 48 h post infection, cells are fixed and stained. n=2 B) Western blot analysis of cell lysates using an antibody against CHIKV nsP3. Cells were infected with 0.1, 1.0 and 5.0 MOI of WT or TST ICRES-CHIKV and lysed either 12 or 24 h post infection.

4.1.2.2 UVC treatment does not have any adverse effects on nsP3 immunoprecipitation

UVC irradiation can cause damage to proteins, causing disulphide bond breakage and generation of reactive oxygen species which has the potential to affect subsequent downstream experimentation (Durchschlag et al., 1996). Therefore, following the optimisation of nsP3 expression, the next step was to investigate whether UVC treatment can have adverse effects on nsP3 immunoprecipitation as this step forms the basis of protein-RNA preservation in the iCLIP technique. First, 5.0×10^6 BHK cells were seeded into 10 cm^2 dishes then infected with TST ICRES-CHIKV. 12h post infection, the media was removed and replaced with 1 mL PBS and the cells were irradiated once with 150 mJ/cm^2 at 254 nm (Figure 4.3A). After UVC treatment, the cells were lysed then sonicated to fragment cellular DNA that prevent access to the target protein. The lysate was then immunoprecipitated using an antibody against the

TST and analysed by western blot. As shown in Figure 4.3B, UVC treatment does not prevent immunoprecipitation of nsP3, and the protein can be robustly purified using the anti-strep tag system.
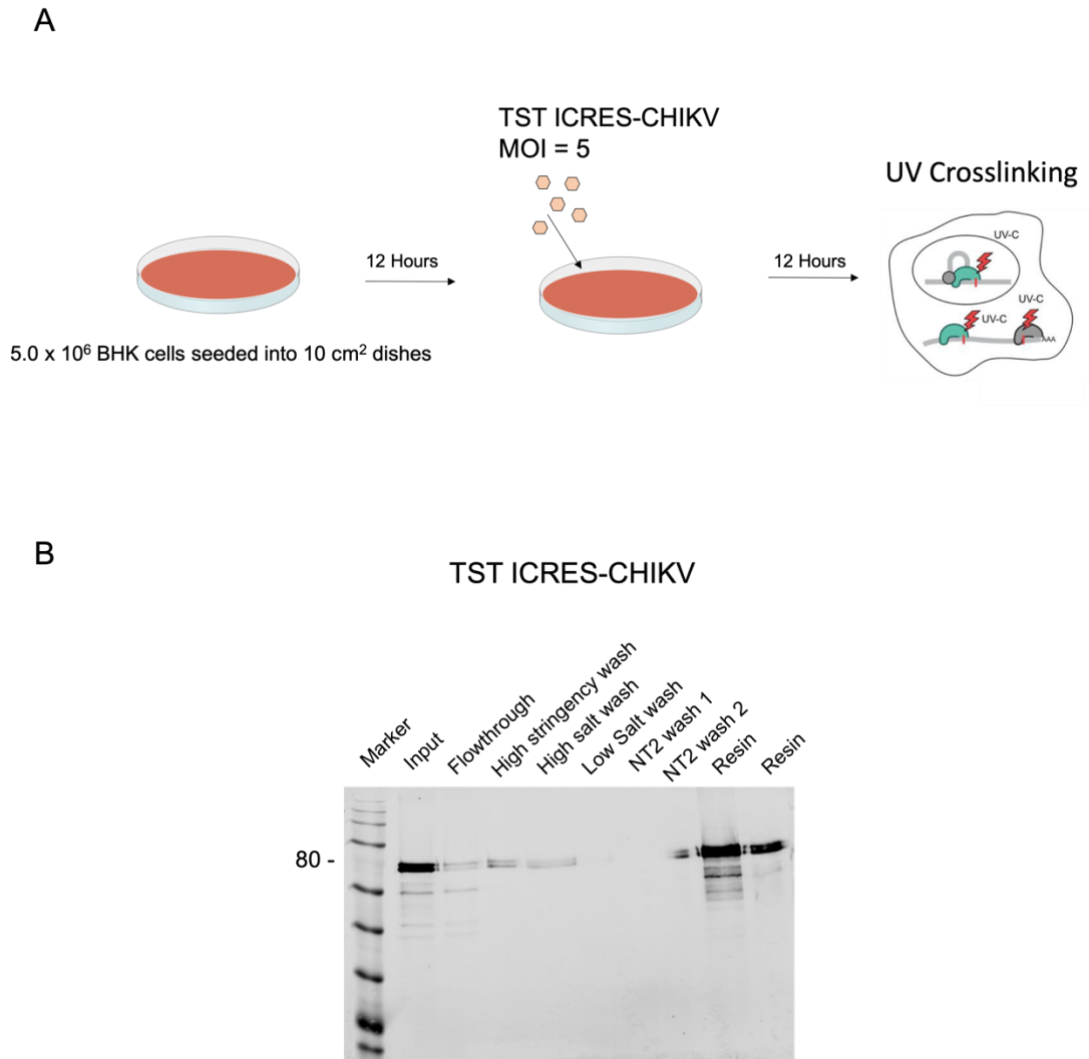
Figure 4.3 nsP3 immunoprecipitation. A) 12 h prior to infection, cells are seeded onto 10 cm2 plates. When the monolayer is 80% confluent, the cells are infected with TST ICRES-CHIKV. 12 h post infection, the cells are irradiated with UVC, and cell lysates are harvested for immunoprecipitation. B) Western blot analysis of immunoprecipitation performed with UVC-irradiated and TST ICRES-CHIKV-infected cell lysate. Immunoprecipitation was performed using buffers listed in Appendix Table 6.7.

4.1.2.3 UV Cross-linked and RNase digested nsP3-RNA complexes can be successfully purified

At this point, steps A, B and D of the iCLIP protocol (Figure 4.1) have been optimised experimentally for CHIKV nsP3. Step 3 in the technique involves RNase digestion of RNAs bound by nsP3. RNase treatment is crucial for determining single nucleotide resolution of protein-RNA binding and therefore must be performed in a controlled manner. This is because the length of digested RNA is important for ensuring that cDNA generated downstream are the ideal size of 50 – 300 nts for genome mapping. Over digestion can introduce sequence bias, and insufficient digestion leads to co-purification of off-target RNA binding proteins and RNAs. Different CLIP techniques utilise different RNA fragmentation enzymes. Many RNases preferentially cleave after specific single or double nucleotides, which can lead to sequence bias and increase sequence constraints at cDNA ends. In iCLIP, RNase I is used as this enzyme can cleave all nucleotides thus have lower sequence specificity (Lee and Ule, 2018). To confirm RNase I treatment is suitable for this project, the experiment described in 4.1.2.2 was repeated up to the sonication of cell lysates. Prior to immunoprecipitation, the lysates were treated with three increasing concentrations of RNase I at 0.4, 0.8 and 2.5 units/mL lysate. Immunoprecipitation was then performed as in Section 4.1.2.2.

RNase digestion success is demonstrated by NuPAGE gel shift analysis of nsP3-RNA complexes. This is driven by the principle that the association of nsP3 with RNA increases the molecular weight of the protein and the resulting complexes therefore migrate slower on an NuPAGE gel. As RNase I digestion is random, this creates RNA fragments of varying lengths causing a diffused signal down the NuPAGE gel (Figure 4.4A). For visualisation of nsP3-RNA complexes, an adapter conjugated to infrared dye was ligated to the 3′ end of the RNAs after immunoprecipitation. The complexes were then separated on an NuPAGE gel and transferred onto nitrocellulose membrane and visualised using the LiCOR Odyssey SA system. The results are shown in Figure 4.4B.
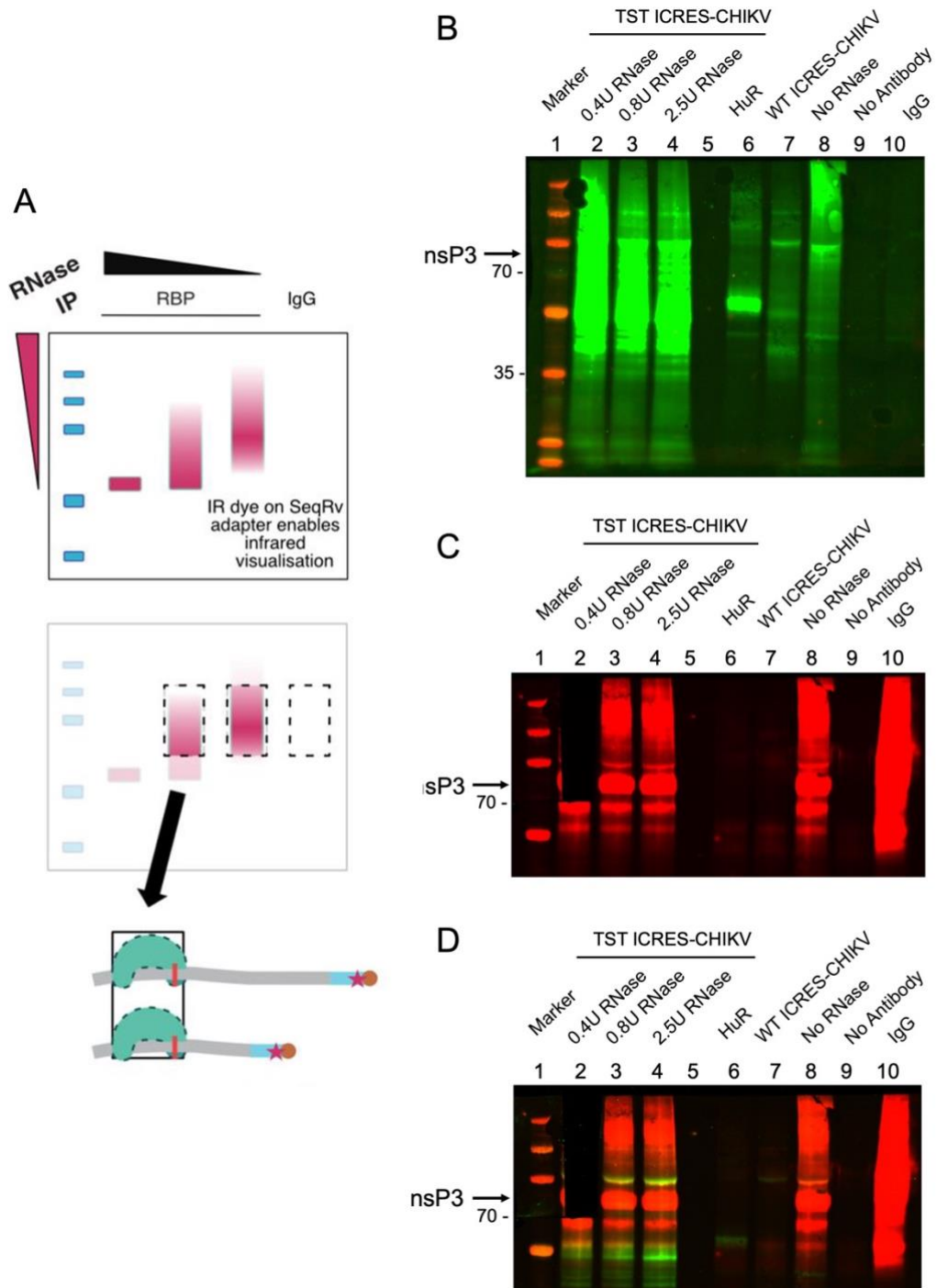
Figure 4.4 Quality control of WT or TST ICRES-CHIKV infected, UVC-irradiated-cell lysate RNase digest, immunoprecipation of nsP3-RNA complexes and 3' adapter ligation. A) Illustration of the quality control step, as concentration of RNase decreases, the 3' adapter signal becomes more diffused on the NuPAGE gel. The appropriately sized complexes were then extracted, highlighted by the dotted rectangles. The excised products of protein (green) bound to RNA (grey) are shown. B) Nitrocellulose membrane image of nsP3-RNA complexes resolved on a NuPAGE gel. C) Western blot analysis of nsP3-RNA complexes on the nitrocellulose membrane using an antibody against nsP3. D) Merge of both nsP3 signal and 3' RNA adapter signal.

111

Quality control (QC) checkpoints are essential during the early steps of iCLIP as determinants of whether the immunoprecipitated nsP3-RNA complexes can be isolated for library preparation. For UVC cross-linking, human antigen R (HuR) was included as a positive control. HuR is a well-characterised, constitutively expressed RNA binding protein responsible for regulating RNA metabolism (Siang et al., 2020). It was also used as a positive control in immunoprecipitation as it can be effectively purified using an anti-huR antibody. For RNase digestion, a 'no RNase' treated control was included as an indication of which complexes should be isolated following NuPAGE analysis. Unfortunately, as shown in Figure 4.4B lane 6, UV cross-linking and immunoprecipitation was not successful for HuR. However, this can be mitigated by the other included controls and observed results. Signal was observed in the TST ICRES-CHIKV + RNase digested samples, and the 'No RNase' (Figure 4.4B lane 8) treated sample, demonstrating positive UV crosslinking.

Three negative controls were included for immunoprecipitation: an anti-Immunoglobulin G (IgG) antibody which does not recognise specific epitope thus is unable to purify any proteins from the cell lysates; a 'no antibody' negative control where immunoprecipitation was performed using beads which were not coupled to the anti-strep antibody; and WT ICRES-CHIKV where nsP3 cannot be purified as it does not contain the twin-strep-tag. As shown in Figure 4.4B lanes 7, 9 and 10, all the negative controls performed as expected.

Once the controls had been analysed and confirmed to work as expected or any observed anomalies mitigated, the nsP3-RNA could be analysed. Firstly, positive signal observed on the gel image confirms that the ligation of 3′ adapter was successful. The results shown in Figure 4.4B lanes 2 – 4 show that increasing the concentration of RNase leads to a higher rate of RNA fragmentation as the signal is more diffused at 0.4 units compared to 0.8 and 2.5 units. Samples treated with 0.8 and 2.5 units of RNase show a sharper band towards the expected molecular weight of nsP3. At 0.4 units, the signal

is uniformly diffused, indicating a proportional spread of increasing nsP3-RNA complexes from the expected molecular weight of ~70 kDa. As the predicted signals were observed on the image, the nsP3-RNA complexes were excised from the membrane for downstream processing. To confirm that the signal observed on the membrane was indeed a result of nsP3-RNA complexes, and that the correct portion of the membrane was excised for downstream processing, the membrane was then probed with an antibody against nsP3 and visualised on a LiCOR imaging system. As seen in Figure 4.4C lane 2-4, red signal indicates the presence of nsP3 and corroborates that the diffused signal was due to nsP3-RNA complexes. It also showed that the excision of nsP3-RNA complexes was successful. Furthermore, it confirmed that immunoprecipitation of nsP3 only occurs in TST ICRES-CHIKV infected samples in combination with beads coupled with the anti-strep tag antibody as nsP3 is not observed in the WT ICRES-CHIKV infected sample or the no antibody sample. A positive signal is seen in the IgG negative control, the reason for this is because the anti-IgG antibody is the same species (rabbit) as the anti-nsP3 antibody, rather than the signal being a true positive for nsP3. Together, the data shows a robust system for RNase digestion of nsP3 bound RNA and immunoprecipitation of nsP3-RNA complexes.

4.1.2.4 cDNA libraries can be successfully prepared from isolated nsP3-RNA complexes

The next step in the protocol is to isolate the nsP3-RNA complexes from the excised membrane fragment, digest nsP3 to leave a peptide at the site of crosslink, purify the RNA then reverse transcribe the bound RNA fragments into cDNA. For this, the membrane was first treated with proteinase K which cleaves peptide bonds to remove most of the protein from the RNA. UVC-crosslinking of nsP3 to RNA protects the part of protein bound to target RNA, so this peptide is maintained following digestion. The peptide-RNA is then subjected to phenol:chloroform:isoamyl alcohol purification. Next, reverse transcription is performed using a primer which is complementary to a portion of the 3′ adapter ligated prior to NuPAGE analysis and contains a barcode. One uniquely barcoded reverse transcription primer is used for one sample

type to allow for multiplexing downstream. The resulting cDNA is then purified using an AMPureXP bead capture system. During reverse transcription, the enzyme proceeds 5′ – 3′ until it reaches the peptide remaining from protein digestion (Figure 4.5A). Around 80% of the time, the peptide causes the reverse transcriptase to disengage due to the physical obstruction. In previous CLIP methods, an adapter is ligated to both the 5′ and 3′ ends of the RNA prior to reverse transcription therefore any prematurely terminated cDNAs are lost as they do not receive an adapter at the 3′ end for downstream PCR amplification (Huppertz et al., 2014). iCLIP addresses this issue by circularising the cDNA following reverse transcription (Figure 4.5A). Therefore, the purified cDNA was circularised using CircLigase™ then purified again using AMPureXP beads.
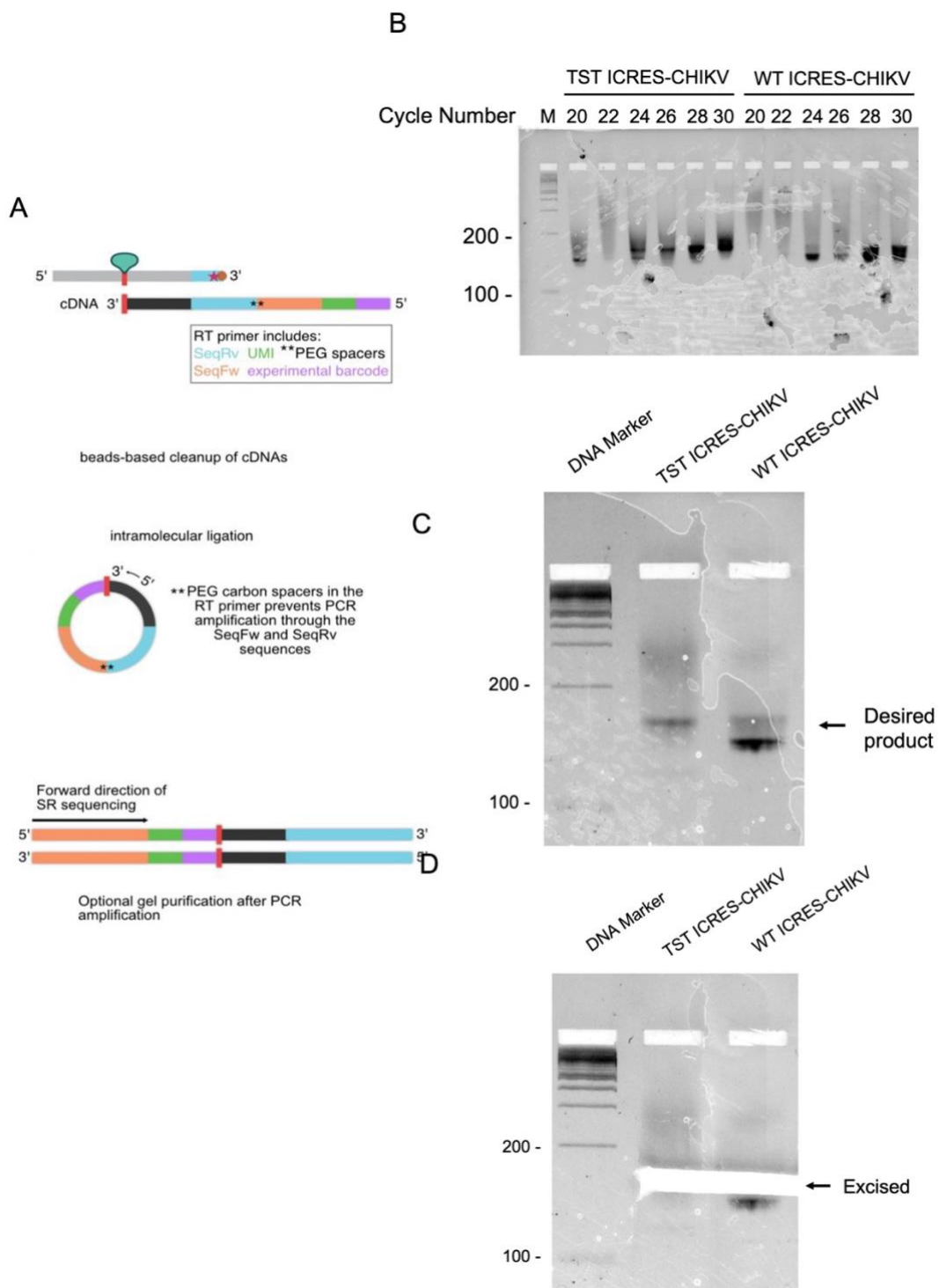
Figure 4.5 Quality control of the reverse transcription, circularisation, and PCR amplification steps. A) Illustration of the reverse transcription, circularisation, and PCR amplification steps. B) PCR cycle optimisation. TBE agarose gel of cDNA products obtained from increasing PCR cycle number (20 – 30). C) PCR amplification of WT and TST ICRES-CHIKV nsP3-bound RNA. Desired product ~150 – 170 nts (Primer 131 nt, insert 20 - 40 nts) D) Gel image following excision of target cDNA products.

The size of cDNA that will be mapped to the reference CHIKV genome should be the size of the insert (20-40 nt) without the two PCR amplification primers and the barcode which equate to 131 nt. Therefore, the expected PCR product for sequencing should be around 150-170 nt. Over-amplification of cDNA is a common problem during the PCR step leading to the generation of contaminating secondary products which migrate as a diffused band of higher size (Huppertz et al., 2014). To avoid this, a test PCR amplification was performed to optimise the PCR cycle number required for the cDNA, where six identical PCR reactions were set up for the WT ICRES-CHIKV cDNA sample and for the TST ICRES-CHIKV cDNA sample. The reactions were then amplified using cycles between 20 – 30 with a stepwise increase of 2 cycles between different reactions. The cDNA products were then electrophoresed on a TBE agarose gel stained with SYBR green (Figure 4.5B). In both the WT and TST ICRES-CHIKV samples, amplification with less than 24 cycles resulted in a weak signal, indicating that the cDNA concentration would not be sufficient for sequencing. Cycles 28 and above produced a thick, diffused band at the expected fragment size, suggesting the presence of secondary products. For the cDNA produced, cycle numbers 24 and 26 appeared ideal for sequencing library preparation as they produced the expected PCR product without highly detectable secondary products. For the TST ICRES-CHIKV sample, this was particularly true at 26 cycles whilst for the WT ICRES-CHIKV samples, 24 cycles appeared more favourable. However, to standardise sample processing, all the following PCR reactions were performed at 26 cycles.

The published iCLIP protocol suggests that a small portion of the PCR amplified cDNA libraries should be visualised on a 6% TBE agarose gel to confirm presence of the designed cDNA population, then the remaining cDNA should be purified using AMPure XP beads. However, when this was performed in this project, the purified cDNA concentration was low thus bead purification was not robust for this specific application. To resolve this, cDNA product was instead electrophoresed on a 2% TAE agarose gel (Figure 4.5C) and desired product was excised. This method also provided assurance that any PCR generated secondary products can be physically excluded. Once the

target fragment was excised, the gel was re-imaged to ensure correct excision (Figure 4.5D). The gel slices were then purified using a gel extraction kit and the cDNA was resuspended in nuclease free water. A separate barcode was used for the WT and the TST ICRES-CHIKV samples during reverse transcription, meaning both samples could be mixed and sent for sequencing together as they can be de-multiplexed bioinformatically later. Sequencing was performed using the commercially available Genewiz Amplicon-EZ service. The results presented here shows that protein digestion and RNA purification was successful and that cDNA libraries can be correctly prepared from nsP3-RNA complexes isolated from a nitrocellulose membrane.

4.1.2.5 Bioinformatics analysis overview

The bioinformatics pipeline used in this project to analyse the sequencing data follows the pipeline outlined by *Busch et al*, and a simplified schematic is shown in Figure 4.6. First, the sequencing data is subjected to a quality control step for filtering out reads that are poor quality or are outside the expected fragment size for alignment. Quality control is important as contaminating sequences can skew alignments to the reference genome and highlight crosslink or binding sites that are untrue. Next, the reads are de-multiplexed into WT or TST ICRES-CHIKV pools so that they can be aligned separately, this is facilitated by the barcoded primer that was used in reverse transcription. Once this has occurred, the libraries can be mapped to the reference CHIKV genome. For iCLIP alignment purposes, the reference genome used is the ECSA strain LR2006 OPY1 modified to include the twin-strep-tag in nsP3. Following alignment, any PCR duplicates are collapsed, which is essential as these can cause over representation of aligned sequences, reducing the accuracy and precision of crosslinked sites. This is again facilitated by the reverse transcription primer, as the five-nucleotide barcode is flanked at the 5′ end by a three-nucleotide wobbler sequence, and at the 3′ end by a four-nucleotide wobbler sequence, known collectively as unique molecular identifiers (UMI), therefore as well as acting as a de-duplication filter, it can also provide information on the orientation of the read (see Figure 4.12 for detailed information of the reverse transcription primer). The specific reverse

transcription primers used in this project are listed in Appendix Table 6.1. Once all the necessary quality control and mapping steps have been performed, PureCLIP is applied to the aligned sequences to mine crosslink and binding sites. PureCLIP is a method to capture specific protein-RNA interaction footprints from iCLIP sequencing date and calls individual crosslink sites based on regions enriched in aligned sequences and truncation patterns (Krakau et al., 2017). Simply, reverse transcription truncates at the site of crosslinking, and the nature of downstream cDNA library preparation means that the site of crosslinking must be a single nucleotide prior to where the sequenced read aligns to the reference genome. Using this information, the crosslink and thus binding site can be deduced.
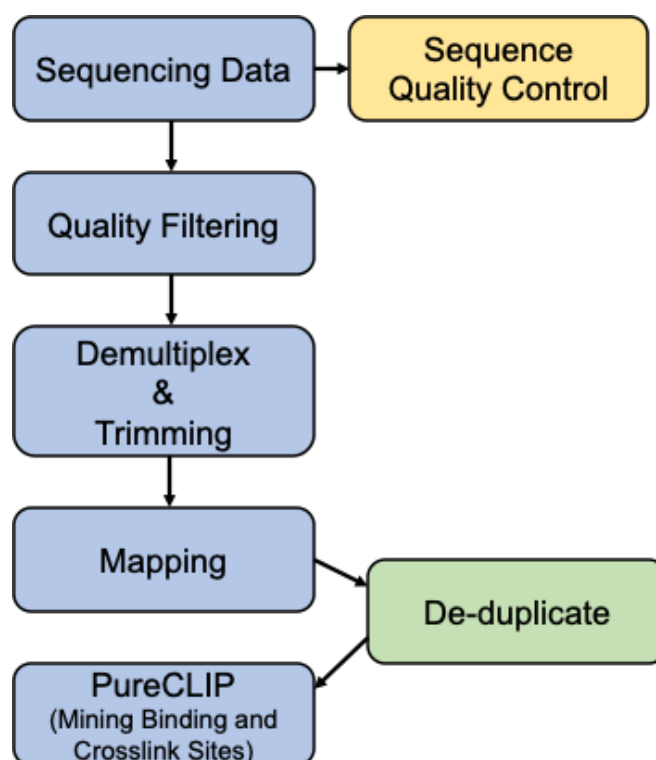
Figure 4.6 Illustration of the bioinformatics analysis pipeline applied to the iCLIP sequencing data. The sequencing data was firstly subjected to quality control. Reads that did not satisfied the defined quality score were removed. Filtered reads were then demultiplexed using the specific barcodes introduced during reverse transcription, the barcodes are then trimmed. Alignment to the reference CHIKV genome was then performed and aligned reads were subjected to PureCLIP analysis which is used to mine crosslinking/binding sites.

4.1.2.6 cDNA library preparation from WT ICRES-CHIKV and TST ICRES-CHIKV infected samples generate good quality sequencing reads

The first step of the bioinformatics analysis involves quality control of the sequenced reads. This was performed for the sequencing data of both the WT and TST ICRES-CHIKV libraries. As shown in the barcode frequency plot in Figure 4.7B, the TST ICRES-CHIKV is the most represented population in the sequenced reads and is around 3-fold higher than the WT ICRES-CHIKV reads. This also corroborates that the sequenced reads from the TST ICRES-CHIKV is truly as a result of nsP3 binding, rather than the result of any non-specific binding. Next, the distribution of read lengths was investigated (Figure 4.7C). Here, the results show that the sequencing reads vary up to 200 nucleotides, but a big proportion lie below 100 nucleotides. Specifically, most reads fell under the 30-34 or 45-49 bp brackets, but small peaks are also observed in the 60-64 and 90-94 regions. Following Genewiz Amplicon-EZ sequencing, the adapters are automatically trimmed from the reads therefore the sequencing data does not contain the primers and the only feature that needs to be considered when filtering read length is the barcode and UMIs (12 nts total). Considering the extra 12 nts, the ideal fragment size would therefore be 30 – 50 nts, which aligns with the peaks present in the sequencing reads providing confidence that the reads can be carried forward for analysis.
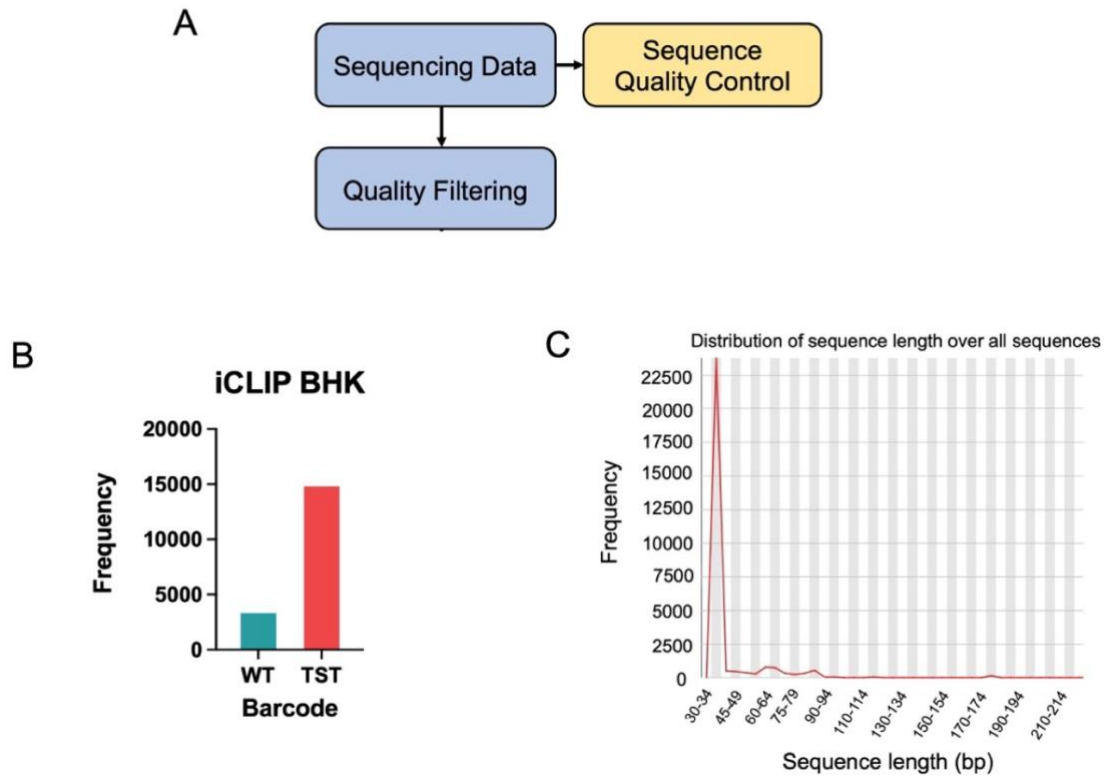
Figure 4.7 Quality control of the sequencing data. A) Schematic outlining the first two steps of the bioinformatics pipeline. B) Number of WT and TST barcodes C) Distribution of read length from the sequencing data.

4.1.2.7 Sequenced reads align across the entire CHIKV genome

Once quality control has been completed, the reads are demultiplexed into two pools for alignment, driven by the unique barcode. The barcode and UMIs are also removed at this point so these sequences do not contaminate the alignment. First, alignment was performed for the WT ICRES-CHIKV sample. Here, out of 3623 sequences that passed all quality control steps, only 24 aligned to the reference genome and the maximum number of reads that aligned to a single nucleotide was 10 (Figure 4.8). In the TST ICRES-CHIKV sequenced pool, 2612 reads, out of 12,495, aligned to the CHIKV genome. Whilst alignments are observed in the WT ICRES-CHIKV sample, the frequency is at a much lower level. The sequences in WT ICRES-CHIKV are not present in the TST ICRES-CHIKV sample therefore they cannot be removed to form a baseline. However, this control serves as an indication that cross-contamination events did not occur during sample preparation.

Furthermore, the TST ICRES-CHIKV reads aligned to reference genome approximately a 100-fold more than the WT ICRES-CHIKV sample, confirming again that the use of WT ICRES-CHIKV as a negative control for iCLIP is suitable, and provides confidence that the nsP3 binding sites from the TST ICRES-CHIKV sample are true positives. The maximum number of reads that aligned to a single nucleotide in the TST ICRES-CHIKV sample was 33 (Figure 4.8). Whilst it is the 5′ end of the sequenced read that is important for deducing the crosslink and binding sites, the number of reads aligned to a given nucleotide highlights the prominence of the genome region. Where higher numbers of alignment to the genome increases the likelihood that a crosslink site and subsequently binding site, is present at that region.
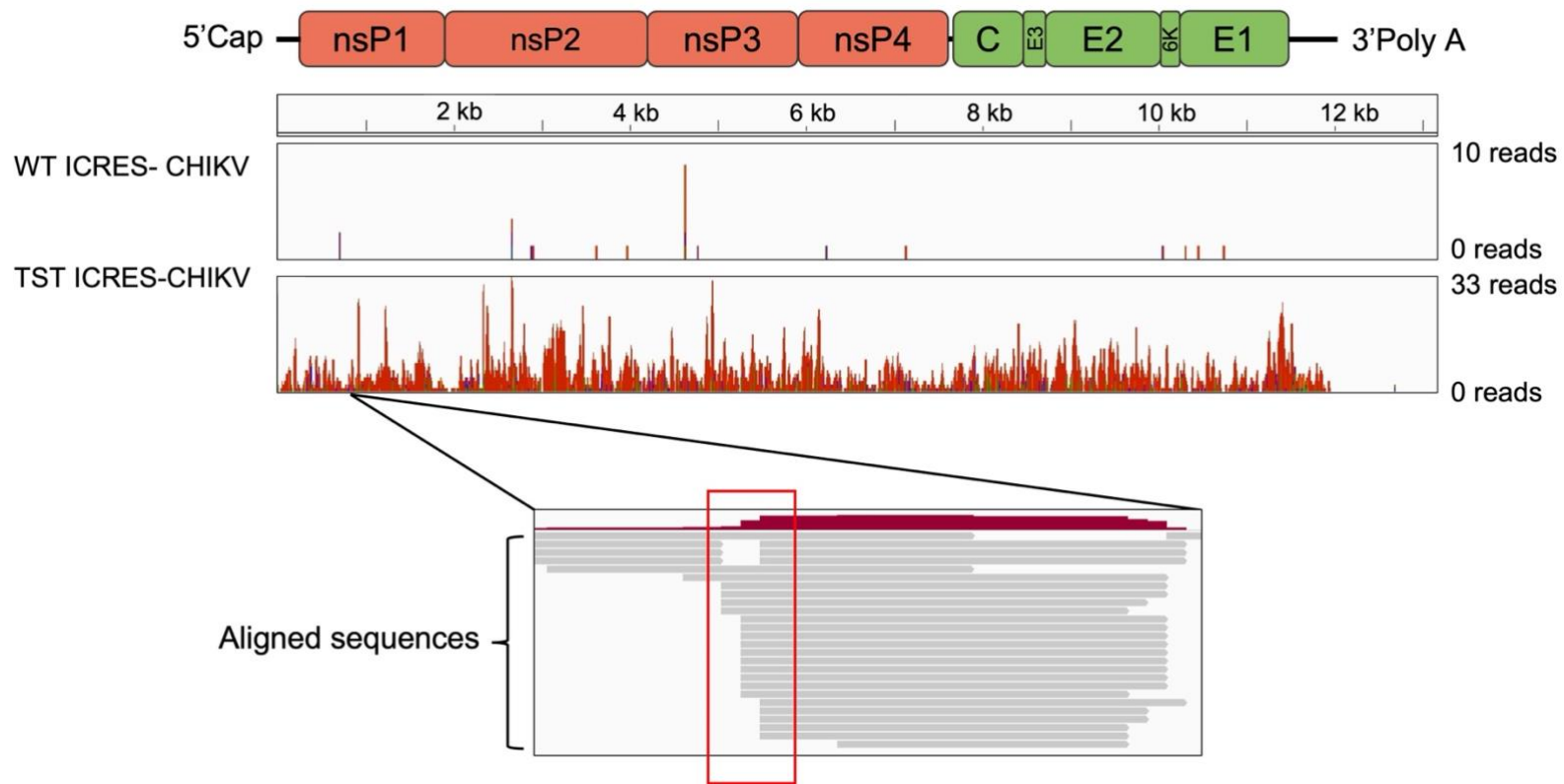
Figure 4.8 Demultiplex, trimming and reference genome mapping. The schematic of the CHIKV genome annotated as encoded genes is shown at the top. Below is the outline of the CHIKV genome by base pairs. Both WT and TST read alignments are shown. The maximum number of alignments at a single nucleotide for the WT sample was 10 reads, and the maximum number for the TST sample was 33 reads. Important for downstream crosslink/binding site identification are the read starts and the frequencies of reads aligned to a single nucleotide position, as highlighted by the red rectangle.

4.1.2.8 nsP3 binds to the CHIKV genome promiscuously

As discussed in 4.1.2.5, to mine the nsP3 crosslink and binding sites, PureCLIP was used. This method detects regions that are enriched in mapped reads, caused by nsP3-RNA binding. Mapped reads are then used to detect where significant fractions of read starts, generated from truncated cDNAs, accumulate (Figure 4.8, red rectangle). Combining all the information, crosslink sites are calculated. Due to the low number of aligned sequences in the WT ICRES-CHIKV sample, it was not possible to process this sample using PureCLIP therefore no crosslink or binding sites could be mined from this pool. On the other hand, the TST ICRES-CHIKV sample was successfully processed by PureCLIP, and 34 crosslink sites were identified (Figure 4.9). Next, PureCLIP computes nsP3 binding regions by merging crosslink sites within a specified distance. The default parameter in the method is to condense crosslink sites within 8 bp into a single binding site. Whilst there is limited evidence in literature on a standard RNA binding sequence length for viral RNA binding proteins (RBP), some evidence exists for both human and viral RBPs exist to suggest that *k*mers can range from 5 bp with no defined upper limit (Schmidt et al., 2021; Van Nostrand et al., 2020). Using the default setting of 8 bp would generate stringent and accurate binding sites, but it can also filter out true, shorter binding sites. Considering the risks and benefits, 5 bp was chosen for the analysis, yielding a total of 17 binding sites (for nucleotide location of binding sites, see appendix Table 6.5) (Schmidt et al., 2021; Van Nostrand et al., 2020). In this experiment, both the crosslink sites and binding sites were observed across the whole CHIKV genome. The binding sites were also all located on the plus strand. The analyses of specific binding sites were first performed as genes encoded by the RNA (Figure 4.9B). Using this method, a higher number of binding sites are observed in the non-structural region, compared to the structural region. In the untranslated regions, no binding site was observed at the 5′ UTR whilst a single binding site was observed at the 3′ UTR. Next, the binding sites were represented as groups of 999 nucleotides (Figure 4.9C). This analysis highlights the spread of binding sites across the genome as there were only 3 regions: 5000-5999 nt, 7000 – 7999 nt and 10000 – 10999 nts where there was no binding.
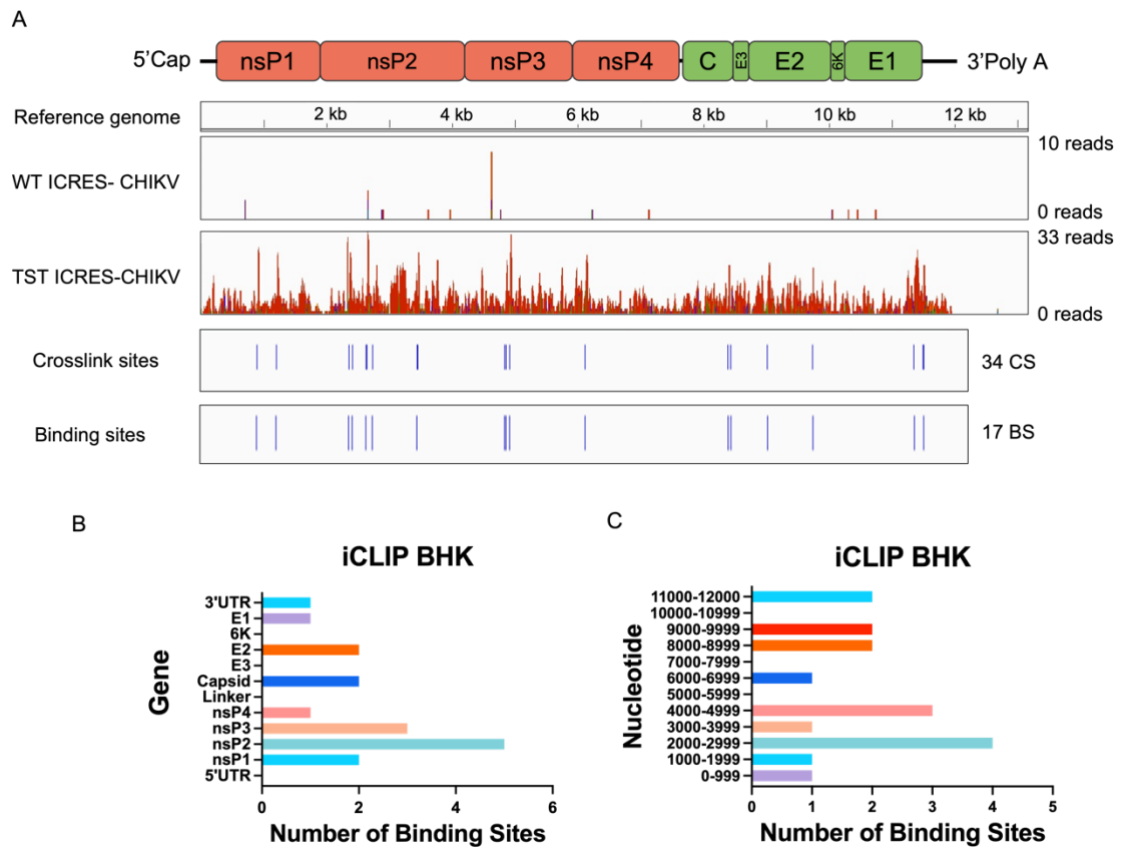
Figure 4.9 PureCLIP analysis of aligned sequences. A) The crosslink and binding sites of aligned sequences from the TST ICRES-CHIKV sample are shown. 34 crosslink sites and 17 binding sites were identified. B) Number of binding sites represented as genes encoded by the CHIKV genome. B) Number of binding sites represented as the CHIKV genome divided into 999 nucleotide-long regions.

4.1.2.9 Exploring binding motifs for nsP3 recognition

Once the RNA binding sites had been determined, further analysis could be performed to extract any binding motifs that may be present in the binding sites. Binding recognition motifs usually precede the binding sites and act as a signal to the RBP for engagement. Therefore, to search for binding motifs, each binding site was flanked at the 5′ end with 10 nucleotides, and at the 3′ end with 20 nucleotides such that the binding site is located directly after the first third of the entire sequence. The sequences can then be subjected to analysis using Multiple EM for Motif Elicitation (MEME) (for sequences submitted to MEME, see Appendix Table 6.6). This tool searches for novel, recurring, fixed-length patterns in biological sequences. The variable length patterns can then be separated into individual motifs (Bailey and Elkan, 1994). In the default parameters, the programme stops searching for additional motifs when a single motif has been identified. The motif that was extracted is shown in Figure 4.10A. However, this was only shared with 6 of the binding sites, and the motif was not present in the remaining 11 binding sites. The sequence conservation is represented in Figure 4.10B and shows that the binding motif is not completely conserved in the 6 binding sites, with only 4 nt positions being complete conserved across the sites. To investigate whether additional binding motifs are present in the binding site sequences, the analysis was repeated but this time the search parameter was increased from 1 to 10 binding motifs. The results are shown in Figure 4.10C. When the search parameter was increased, 10 binding motifs were produced (see Appendix Figure 6.1 for list of binding motifs). However, as observed in the results, the binding motifs are unspecific, demonstrated by the variety of colours on the motif location chart (Figure 4.10C). Furthermore, aside from the motif that was identified when the parameter was set at 1, which appeared 6 times, the maximum number of a single binding motif appearing in the second analysis was 2, suggesting that the additional binding motifs that were found when the parameter was increased to 10 do not represent a universal binding motif. Notably, all the binding motifs were located on the positive strand, depicted by the '+' symbol before the motif location, which aligns with the identified binding sites. Combined, the results presented here, and in section 4.2.8 suggests that nsP3 binds across the entire CHIKV genome and suggests that binding may not be driven by a single consensus recognition motif.
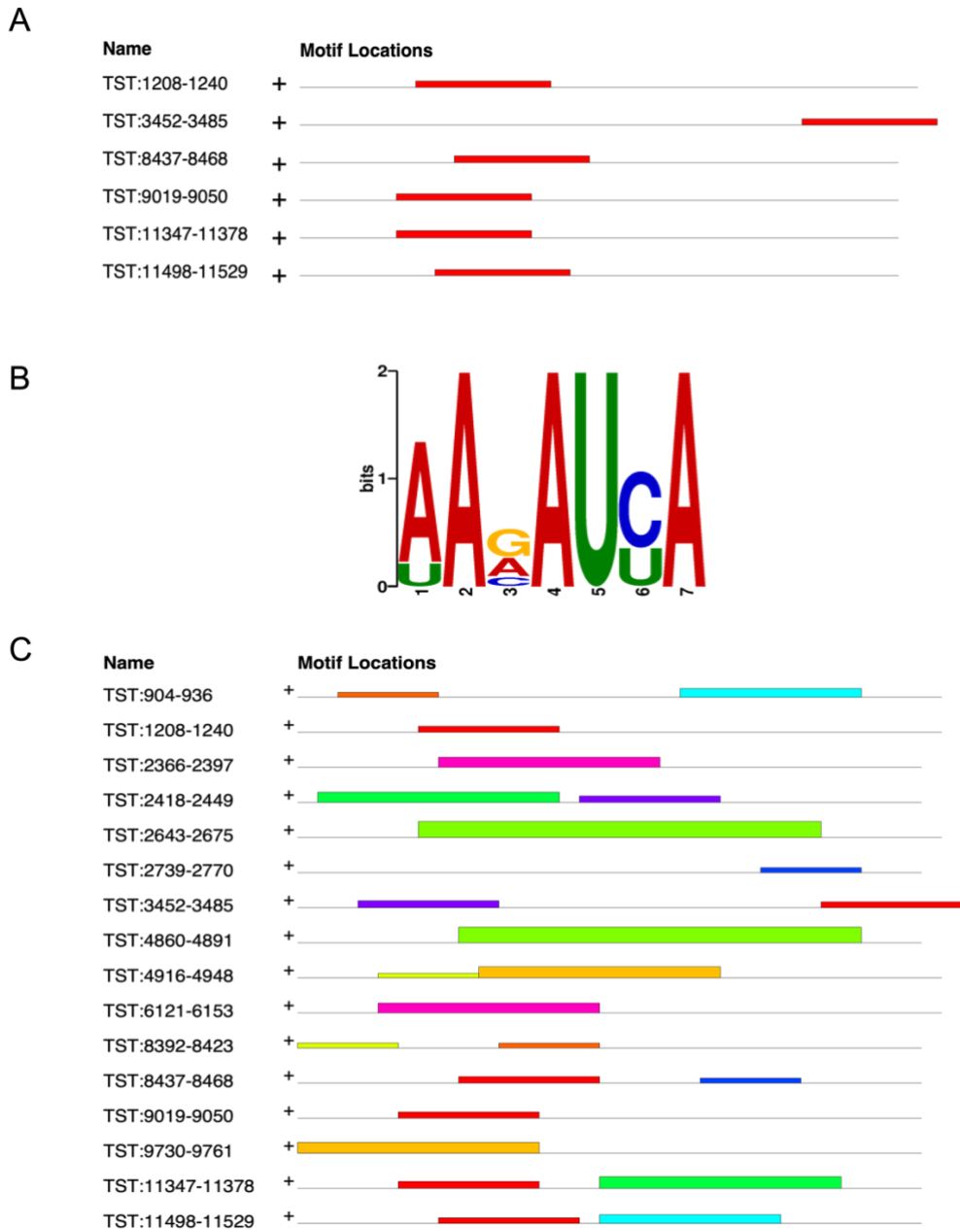
Figure 4.10 Binding motif mining using Multiple EM for Motif Elicitation (MEME)
A) A single binding motif, shared between six of the 17 identified binding sites.
B) Consensus sequence of the binding motif. C) Location of the motifs in the
binding sites identified when MEME parameters was increased from 1 to 10.
Each colours represent a single motif. The height of a colour block indicates the
significance of the motif where a taller block represents a higher significance.

4.1.2.10 Secondary RNA structures influence nsP3 binding to the CHIKV genome

Binding recognition is usually determined by two major features: a binding motif or secondary RNA structures. As a binding motif could not be identified, the next analysis was to investigate whether nsP3 uses secondary structures as a signal for RNA binding. To increase the accuracy of output structures, the binding sites were once again flanked with 50 nucleotides on either side. The sequences were then analysed using the mfold software which simulate single-stranded nucleic acid folding by calculating the optimal arrangement that produces minimum free energy (Churkin et al., 2015; Zuker, 2003). The specific binding site nucleotides were then highlighted and the nature of nsP3 binding to the secondary RNA structures were divided into 5 categories (Figure 4.11A). The first category, 'bulge', is defined by nsP3 binding to a bulge region or mini loop around paired regions where the binding nucleotides are unpaired, an example of this is shown in Figure 4.11C. The second is 'paired region' where every binding nucleotide are paired, as shown in Figure 4.11D. nsP3 binding to completely unpaired nucleotide in a stem loop is the third type and is referred to as 'end of stem loop' binding, shown in Figure 4.11E. The fourth type of binding is called 'overlap', where the some of the binding nucleotides are paired, and some are unpaired, as observed in Figure 4.11F. Lastly, where nsP3 binds to a linear single stranded region, as shown in Figure 4.11G, is termed 'linear unpaired region'. From the results presented here, nsP3 appears to prefer unpaired RNA, especially where the region forms part of a loop. This is likely due to the structural availability of space around the secondary structure as paired regions are less accessible. The length of binding sites was then examined and were found to vary between 2 to 4 nucleotides (Figure 4.11B). Binding sites consisting of 2 nucleotides were the most frequent, which aligns with the nature of iCLIP where the goal is to resolve protein-RNA interactions to a single nucleotide, but bioinformatics restraints mean that binding sites cannot start and end on the same nucleotide.

A

Type of binding in secondary RNA structure	Frequency
C) Bulge	5
D) Paired region	1
E) End of stem loop	5
F) Overlap	4
G) Linear unpaired region	2

B

Binding Site Nucleotide	Frequency
4	1
3	5
2	11

Figure 4.11 Secondary structures identified from binding sites flanked with 50 nucleotides on both sites. Binding nucleotides are circled in green. A) Types of binding in the secondary RNA structures and their frequencies. B) Binding site nucleotide length and their frequencies. C) mFold representation of binding type 'Bulge' is defined as binding nucleotides that are in unpaired region of a stem bulge. D) mFold representation of binding type 'Paired region' is defined as binding where every bound nucleotide is paired in a stem. E) mFold representation of 'End of stem loop' binding defined as binding nucleotides located at the end of a stem loop where all nucleotides are unpaired. F) mFold representation of an 'overlap' binding type described as binding where some of the binding nucleotides are paired, and some are unpaired. G) mFold representation of the 'linear unpaired region' where binding nucleotides lie in a single stranded unpaired region.

4.1.2.11 cDNA library optimisation

The sequencing data presented thus far has provided an indication of how nsP3 interacts with the CHIKV genome during infection. However, whilst conducting data analysis, it was noticed that many of the sequencing reads did not pass quality control. Approximately ~60,000 reads were generated from sequencing, but only ~18000 passed quality control and could be used in alignment to the reference genome meaning almost 70% of sequencing reads were not suitable for bioinformatics analysis. Therefore, a review was conducted to pinpoint the reason behind poor quality sequencing. As multiple quality control steps precede library preparation to ensured that the nsP3-RNA complexes that are taken forward for analysis are of good quality, an error in sequencing due to failed immunoprecipitation was ruled out. The most probable cause of sequencing failure would be during reverse transcription, circularisation, PCR amplification or sequencing. Figures 12 – 14 show detailed depictions of how the nsP3 bound RNA is reverse transcribed to generate a cDNA library and subsequently sequenced. Reverse transcription uses a primer containing a reverse sequencing primer separated from the forward sequencing primer by two carbon spacers and a *BamH*I restriction site, a 4 nucleotide UMI, a specific barcode and a 3 nucleotide UMI (Figure 4.12A and Figure 4.12B). The forward sequencing primer is then introduced into the 5′ end of the cDNA by circularisation (Figure 4.13A). The circularised cDNA must then be PCR amplified to satisfy sequencing concentration requirements. For this, a pair of primers named P3 Solexa and P5 Solexa (kindly gifted from Jernej Ule, The Crick Institute) were used (sequence in Appendix Table 6.1), which are paired end primers that are compatible for illumina sequencing. PCR amplification also linearises the cDNA as DNA polymerase proceeds until it reaches the carbon spacers at the end of the reverse sequencing primer sequence (Figure 4.13B and Figure 4.13C). PCR products were then analysed using TBE/TAE agarose gel electrophoresis where appropriate and the results confirmed that both reverse transcription and circularisation was successful. If either one of these reactions failed, PCR products would not be observed on the gel. As the library preparation steps were successful, it suggests that an error in sequencing could be the cause of poor-quality reads.
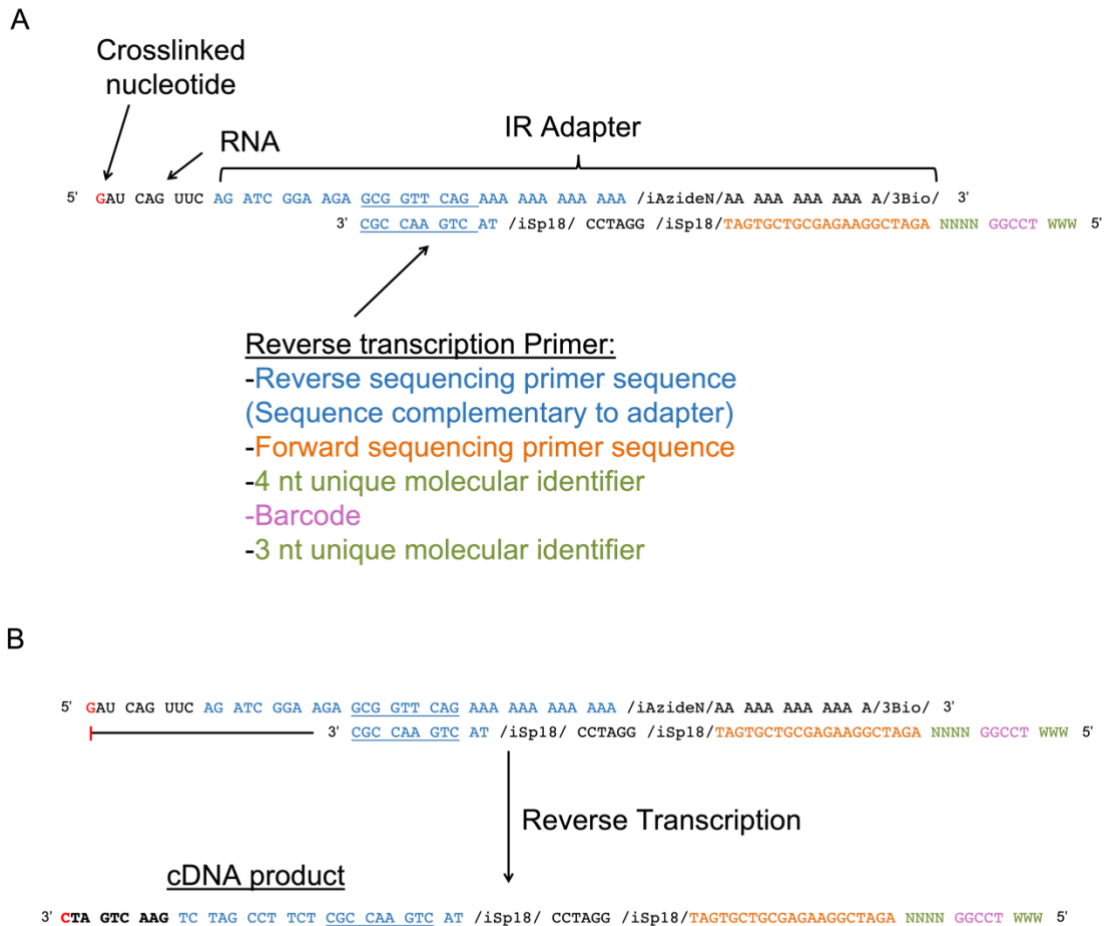
Figure 4.12 Detailed schematic of iCLIP reverse transcription. A) The reverse transcription primer binds to a complementary sequence in the infrared adapter and contains a reverse sequencing primer sequence, a forward sequencing primer sequence, a 4-nucleotide unique molecular identifier, a barcode and a 3-nucleotide unique molecular identifier. B) Reverse transcription product with the sequences required for sequencing introduced into the cDNA.
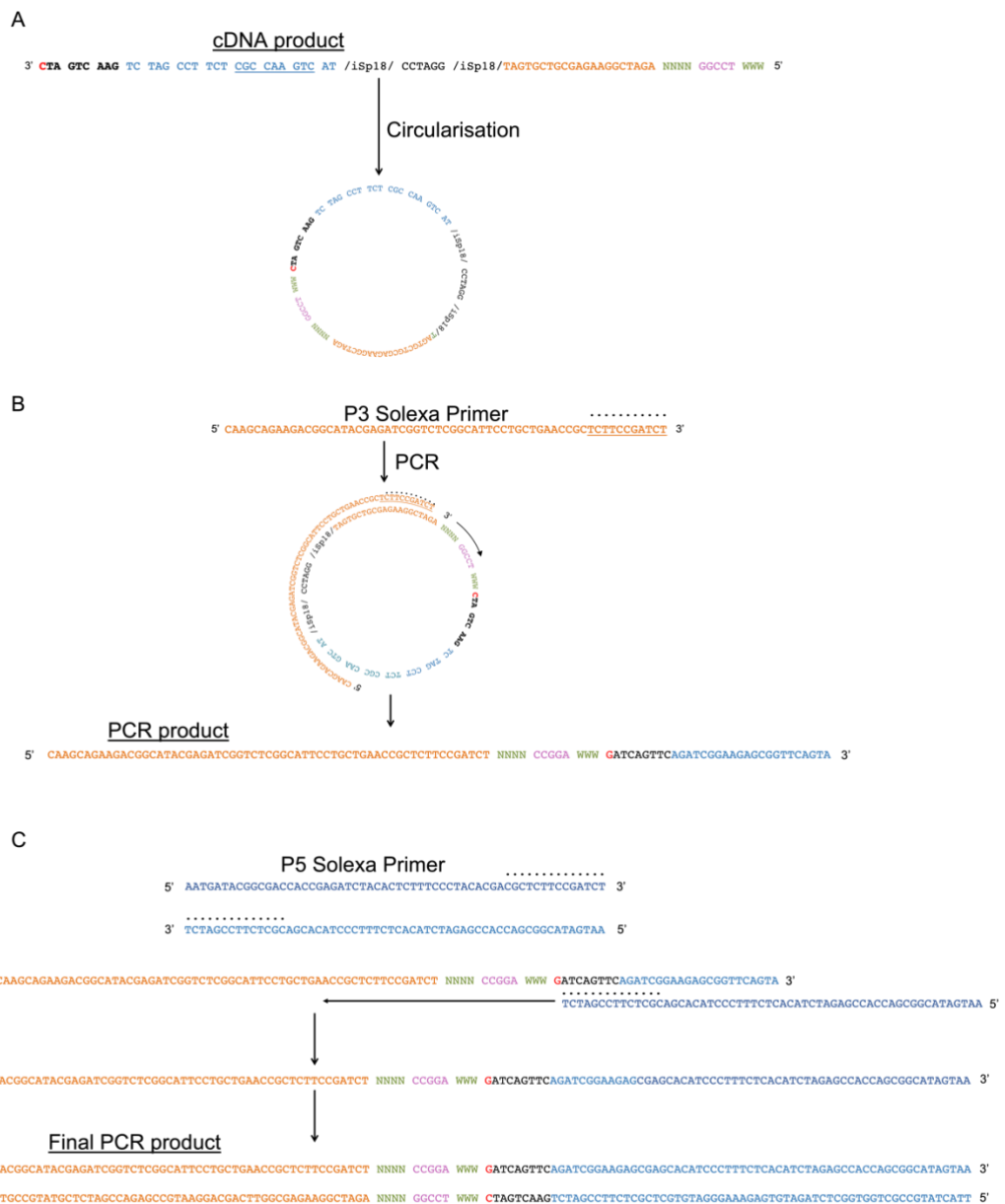
Figure 4.13 Detailed schematic of the circularization and PCR amplification steps of the iCLIP protocol. A) The cDNA product was circularized using circ-ligase which ligates the 5′ end of the cDNA to the 3′ end. B) A P3 Solexa Primer was then used to perform PCR amplification of the cDNA product, which also linearised the cDNA due to the carbon spacers present in the cDNA. C) Second strand synthesis was achieved using a P5 Solexa primer to generate double stranded PCR products.
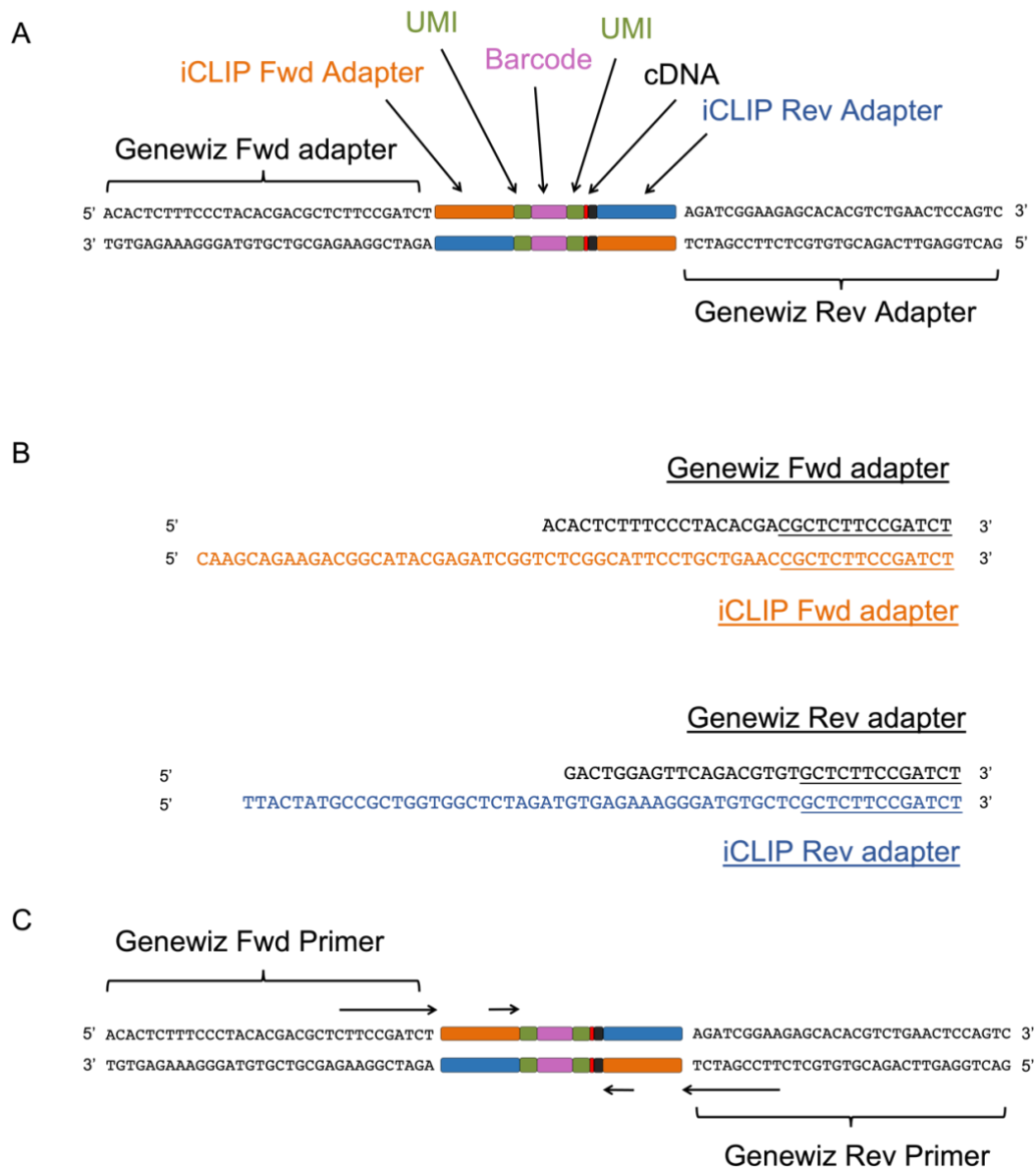
Figure 4.14 Detailed schematic of adapters used in cDNA library sequencing. A) As the adapters introduced during the iCLIP protocol are not compatible with commercial Genewiz EZ sequencing, additional Genewiz specific adapters are ligated prior to the sequencing reactions. B) Sequences of the forward and reverse Genewiz and iCLIP adapters, with identical nucleotides underlined. C) Sequencing of cDNA products can occur from either the Genewiz or iCLIP adapters due to sequence similarity shown in B. Positions where sequencing can initiate are depicted using black arrows.

Genewiz Amplicon-EZ is a commercially available standardised sequencing service. As such, all sequencing performed using this method must follow the listed submission guidelines. Genewiz Amplicon-EZ uses specific sequencing

primers which bind pre-defined adapters ligated to the cDNA. As the libraries produced in this experiment did not contain the Genewiz specified adapters, they were submitted as naked reads and the adapters were subsequently ligated by Genewiz (Figure 4.14A). However, analysis of the P3/P5 Solexa primers showed that they introduced a partial sequence identical to the Genewiz adapters into the cDNA libraries (Figure 4.14B, identical sequences are underlined). Therefore, during sequencing the Genewiz primers can bind at two positions on the cDNA, depicted by arrows on Figure 4.14C, leading to the production of poor-quality reads. Furthermore, the adapter sequences could not be correctly trimmed from the sequencing reads and led to adapter sequence contamination in the results. To prevent a repeat of this sequencing error, PCR amplification reactions were performed with Genewiz specified Illumina adapter sequences (shown in Appendix Table 6.1) instead of P3/P5 Solexa primers.

4.1.2.12 iCLIP with new primers

The iCLIP protocol was repeated and the relevant experimental quality control steps were performed as described in sections 4.1.2.3 and 4.1.2.4. As in the first iCLIP experiment, 3′ adapter ligated and immunoprecipitated nsP3-RNA complexes were resolved on using NuPAGE gel electrophoresis (Figure 4.15A). In this experiment, the HuR positive control (Figure 4.15A, lane 6) performed as expected and confirmed that UVC-irradiation, RNase digest, immunoprecipitation and adapter ligation were all successful. The negative control for RNase digest (Figure 4.15A, lane 7) showed a weaker signal than that observed in the previous iCLIP experiment (Figure 4.4A, lane 8) which could be due to a lower density of nsP3-RNA complexes in that specific sample, rather than an inherent failure in the experiment. The 'no antibody' control performed as expected as no signal was present in this sample. Similarly, the WT ICRES-CHIKV infected sample also had no signal. As shown in the last iCLIP experiment, when the TST ICRES-CHIKV infected samples were digested with increasing concentration of RNase I, the diffusion of signal decreased. The sample treated with 2.5 units of enzyme showed a sharp band at the expected weight of nsP3, whereas treatment with the lowest RNase concentration led to a diffused signal, indicating a shift in molecular weight.
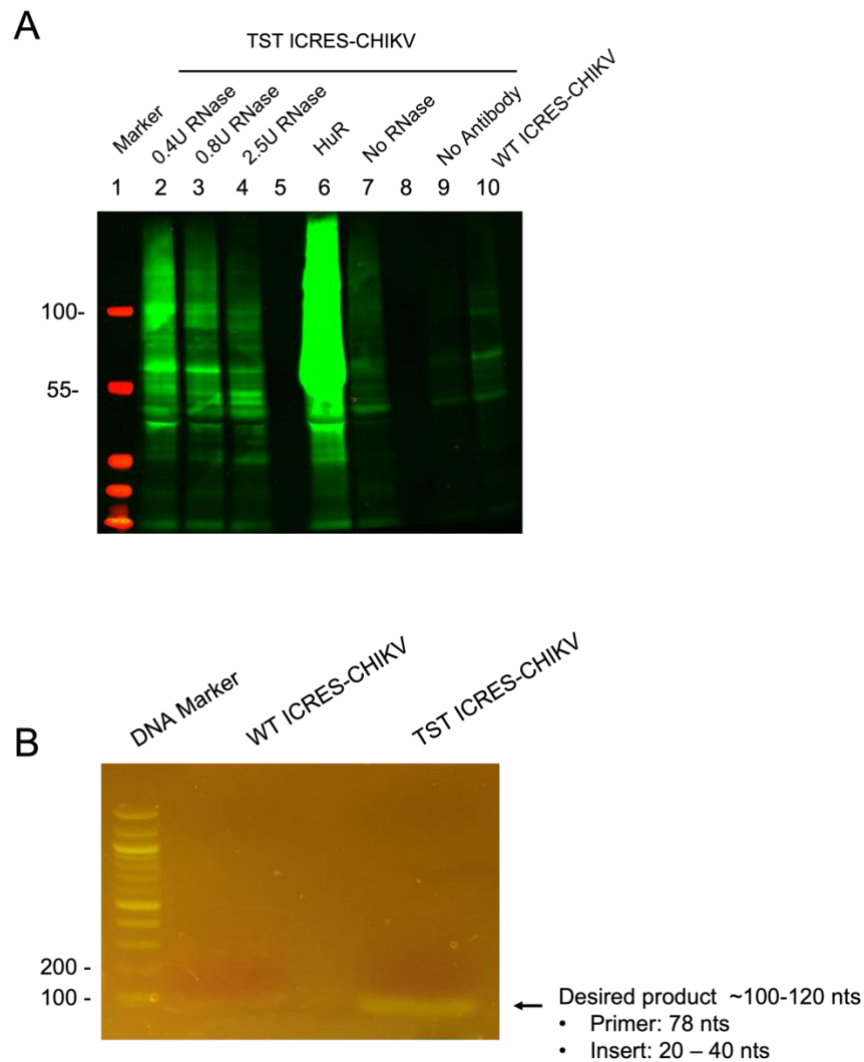
Figure 4.15 Quality control steps of iCLIP using Genewiz sequencing appropriate PCR primers. A) nsP3-RNA complexes from WT or TST ICRES-CHIKV infected cells were resolves on a NuPAGE gel and transferred onto nitrocellulose membrane. B) TAE agarose gel of PCR products obtained with 26 cycles. The desired product is highlighted using a black arrow.

The correct nsP3-RNA complexes were then excised from the membrane, digested with proteinase K, purified and reverse transcribed using the new Genewiz adapters. PCR amplification was then performed for 26 cycles, and the products were electrophoresed on a 2% TAE agarose gel (Figure 4.15B). The previous expected product size was 150-170 nts, made up of 131 nts of primer and a 20-40 nt insert. The Genewiz adapters are shorter in length than the previous primers and add up to 78 nucleotides. Including the insert, the total expected product would be between 100-120 nucleotides which can be observed

in Figure 4.15B. Both the WT and TST ICRES-CHIKV infected samples produced a band under 100 nucleotides, demonstrating that PCR amplification using the new primers was successful. Notably, the WT ICRES-CHIKV infected sample produced a band less concentrated than the TST ICRES-CHIKV sample providing confidence that the TST ICRES-CHIKV sample contains RNA bound by nsP3, rather than non-specific residual RNA. Together, the results show that changing the primers does not have adverse effects on the experimental steps of nsP3-RNA iCLIP.

4.1.2.13 Improved sequencing reads as a result of primer switching

cDNA libraries prepared from the iCLIP experiment were sequenced as previously described and bioinformatics analysis was then performed. To assess whether changing the adapters improved sequencing, a comparison analysis was performed, and the results are shown in Figure 4.16A. The first iCLIP sequencing experiment produced ~60,000 reads whereas sequencing using the new primers yielded ~120,000 reads, resulting in a 2-fold increase. Although the number of reads significantly increased, it was important to investigate whether they were of good quality thus quality control treatment was then applied. In the first iCLIP experiment, 31.5% of reads successfully passed QC. Comparatively in the second experiment, this increased to 66.5%, confirming that changing the primers enhanced the quality of the sequencing reads. Next, the distribution of the WT and TST barcodes was examined. Of the reads that passed QC, the proportion of WT and TST barcodes from the first iCLIP experiment were 19.4% and 80.6%, respectively. On the other hand, in the second iCLIP experiment, the WT barcodes appeared in 5.3% of the read pool, whilst the remaining 94.7% of reads presented a TST barcode (shown numerically in Figure 4.16A, and graphically in Figure 4.16B). The reduction in WT barcoded frequency and the increase in TST barcoded frequency shows that the removal of residual RNA during immunoprecipitation in the second iCLIP experiment was more successful than the preparation performed in the first experiment. In line with the number of reads that passed QC, the number of TST barcoded reads that aligned to the genome also increased 6-fold in the second iCLIP experiment compared to the first iCLIP sequencing results whilst the number of WT reads remained low at

2.9% of all aligned reads. Lastly, the sequence length of reads was inspected (Figure 4.16C). As observed in the previous experiment, most inserts fell under the 30-34 bracket. However, in this experiment, a second distinct peak is shown at 90-94 bps. These reads are unsuitable for alignment as long reads can contaminate crosslink/binding site mining, but they can be easily excluded from further analysis. The results obtained using the new primers display improvements in the sequencing process and increases the confidence of input material for downstream studies.

A

Experiment	Total sequenced reads	Number of reads that passed quality control (WT and TST)	% of reads that passed quality control	Number of reads with WT barcode	Number of reads with TST barcode	Number of WT sequences that aligned to the CHIKV genome	Number of TST sequences that aligned to the CHIKV genome
BHK Exp 1 (Original Adapters)	59,583	18,754	31.5%	3,647 (19.4%)	15,107 (80.6%)	24	2612
BHK Exp 2 (Genewiz Adapters)	119,821	79,682	66.5%	4,240 (5.3%)	75,442 (94.7%)	469	15,479

B

C

Figure 4.16 Sequencing results using Genewiz sequencing appropriate PCR primers. A) Comparative analysis of sequencing reads obtained from iCLIP experiment 1 (old primers) and iCLIP experiment 2 (new primers). The number of total sequenced reads, number of reads that passed quality control, % of reads that passed quality control, number of reads with WT and TST barcodes and the number of reads from both samples that aligned to the reference CHIKV genome are shown. B) WT and TST barcode frequency in iCLIP experiment 2. C) Distribution of read length from the iCLIP experiment 2 sequencing data.

4.1.2.14 Reference genome alignment, crosslink, and binding site mining

As the sequencing reads produced from the new iCLIP experiment were of good quality, genome alignment could be conducted. As observed in the previous alignment analysis, the results in Figure 4.17A confirm that nsP3 binds across the entire CHIKV genome. In this experiment, the number of reads that aligned to a single genome region from the WT ICRES-CHIKV infected sample was increased from 10 in the first experiment, to 61. As expected, the maximum number of reads in the TST ICRES-CHIKV infected sample also increased from 33 reads to 1899. Interestingly, the alignment profile of the WT ICRES-CHIKV infected sample was almost identical to that shown in the TST ICRES-CHIKV sample, which was not the case in the previous experiment, where the coverage was completely different between the two samples. A distinct peak of alignment was located around 8 kb in both samples at the gene encoding nsP4. PureCLIP was then applied to the TST ICRES-CHIKV read pool. 432 crosslinked sites were identified, which were combined into 153 binding sites. In the first experiment, both crosslink and binding sites were spread across the whole genome. Comparatively, the sites found in this experiment binned into three regions. The first group of sites are located around the first 2 kb of the genome. The second group is in the centre of the genome, covering 4kb to 9.9 kb. Lastly, a small group of sites is located at 11 kb (Figure 4.17C). The binding sites are then represented as genes encoded by the genome (Figure 4.17B). Again, no specific trend can be deduced from the binding sites. However, in agreement with the previous experiment, both the 5′ UTR and gene encoding 6K had no binding but this could be due to the fact both regions are relatively small rather than nsP3 selecting to omit interaction. As observed in the previous data set, all the binding sites were again located on the positive strand, indicating a preference over the negative sense genome. To investigate whether the reads aligned in the WT ICRES-CHIKV sample were significant, PureCLIP was also applied to this pool of reads. Crosslink site and binding site mining revealed 2 regions and 1 region, respectively. This suggests that some contamination may have been introduced into the sample. It is unlikely that the binding site is the result of nsP3 immunoprecipitation, as shown in the NuPAGE analysis. The results from both iCLIP experiments confirms the previous result that nsP3 does not bind to the

CHIKV genome with specific preference and suggests that the interactions between nsP3 and the viral genome is stochastic.
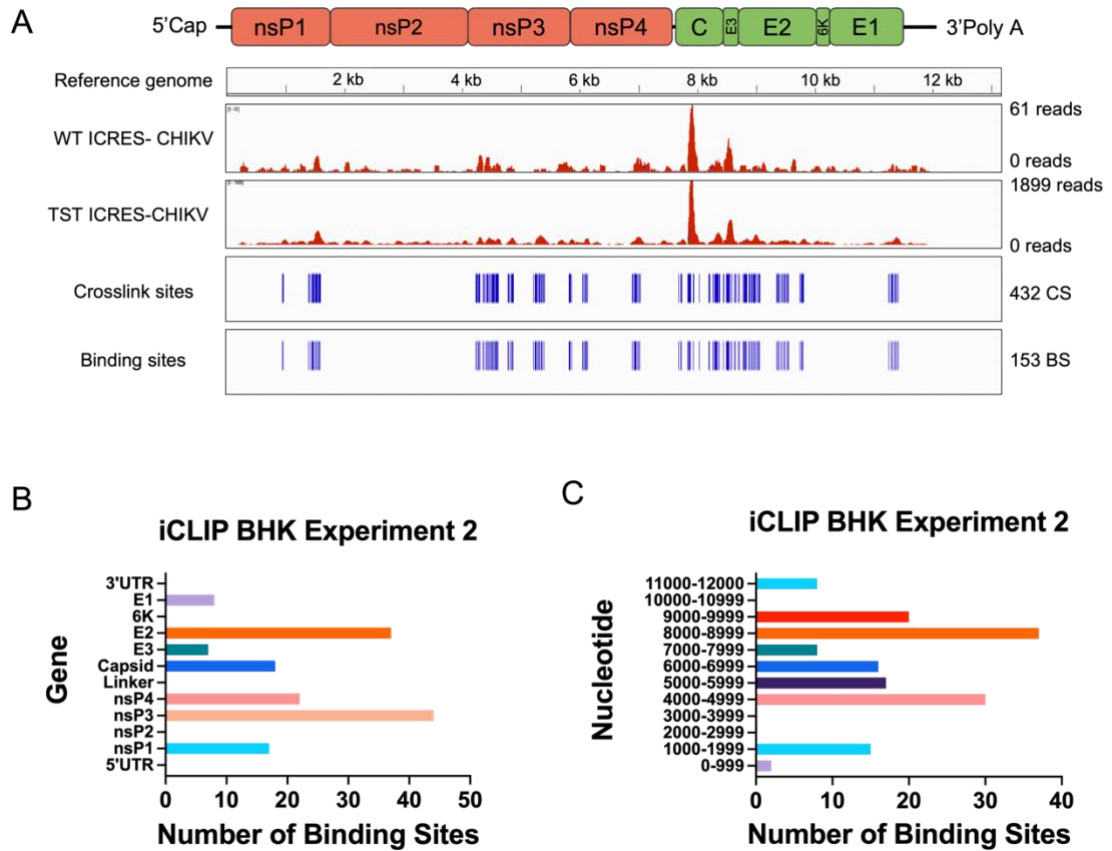
Figure 4.17 PureCLIP analysis of aligned sequences from iCLIP experiment 2. A) The crosslink and binding sites of aligned sequences from the TST ICRES-CHIKV sample are shown. 432 crosslink sites and 153 binding sites were identified. B) Number of binding sites obtained from iCLIP experiment 2 represented as genes encoded by the CHIKV genome. B) Number of binding sites obtained from iCLIP experiment 2 represented as the CHIKV genome divided into 999 nucleotide-long regions.

4.1.2.15 nsP3 does not interact with viral genome using conserved binding motifs

It was not possible to discover a robust conserved binding motif from the previous data, but this could be due to a small number of identified binding sites. The new data set presented a 9-fold increase in binding sites therefore a more comprehensive analysis into binding motif could be conducted. As discussed in section 4.1.2.9, the sites were once again flanked by 10 nucleotides before, 20

nucleotides after, and analysed using the MEME tool. From the 153 binding sites, with the settings limited to searching for a single motif, a 16-nucleotide long motif (Figure 4.18A) appearing 13 times across the binding sites was identified. The first 3 nucleotides of this motif were completely conserved across all 13 binding sites. However, nine positions varied between 2 nucleotides with the remaining four positions varying between 3 nucleotides which demonstrates that the motif was not identical across the 13 instances (Figure 4.18B). In line with the binding sites, the binding motifs were all located in the positive strand. The results from both datasets confirm that nsP3 does not bind to the CHIKV genome with sequence specificity but may have some preference for the positive sense RNA.
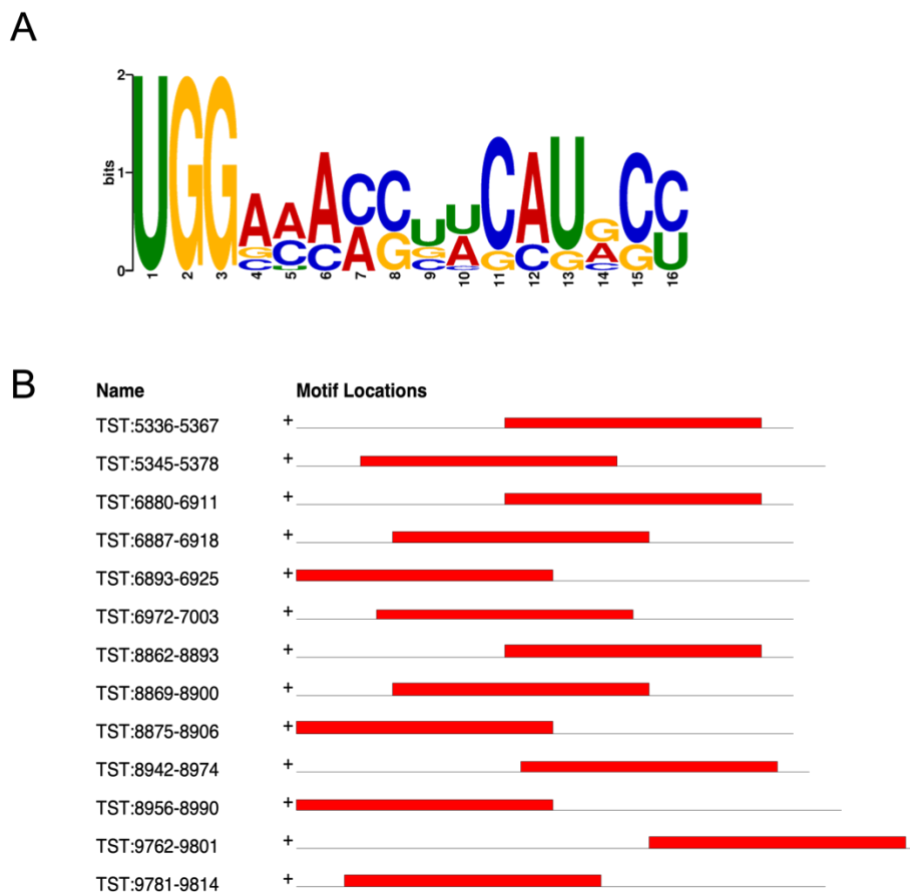
Figure 4.18 Binding motif mining of binding sites identified in iCLIP experiment 2 using Multiple EM for Motif Elicitation (MEME) A) Consensus sequence of a single binding motif, shared between 13 of the 153 identified binding sites. C) Location of the motif in the 13 binding sites.

4.1.2.16 nsP3 binds both paired and unpaired secondary RNA structures

Analyses so far has confirmed that nsP3 does not use a conserved motif for RNA binding and the previous dataset suggested that perhaps secondary RNA structures are more important for binding recognition. To corroborate this, the binding sites obtained from the second iCLIP experiment were subjected to mfold analysis again. Here, the results (Figure 4.19A) show that most of the binding sites lie within 'overlap' regions comprising both paired and unpaired nucleotides. Consistent with the earlier results, linear unpaired regions had the lowest frequency of binding. Previously, the data (section 4.1.2.10) showed that nsP3 binds preferentially to unpaired RNA, but this is not the case here. Interestingly, the binding frequencies of 'bulge', 'paired region' and 'end of stem loop' were almost equal, suggesting that nsP3 binding may not be determined by structural availability of the target RNA. Comparing the length of the binding sites, the data obtained from the second experiment had greater variability than the first iCLIP analysis (Figure 4.19B) where binding sites did not exceed 4 nucleotides. Aligned with the first dataset, most binding sites were between 2 to 7 nucleotides long, with 2 nucleotide binding sites represented most frequently. However, a small proportion of binding sites in the second iCLIP experiment where much longer, reaching up to 29 nucleotides in length. The three longest binding sites are shown in Figure 4.19C (25 nts), Figure 4.19D (29 nts) and Figure 4.19E (16 nts). The 25 and 16 nt binding sites span across paired regions and loops. On the other hand, the 29-nucleotide binding site is located on one side of a stem, where most of the nucleotides are paired. The shortest binding site of 16 nucleotides lies in the gene encoding nsP3, specifically in the AUD. On the other hand, the 25 and 29 nucleotide-long binding sites are both located in the structural protein region and lie in genes encoding capsid protein and E3 protein, respectively.

A

Type of binding in secondary RNA structure	Frequency
C) Bulge	22
D) Paired region	24
E) End of stem loop	28
F) Overlap	66
G) Linear unpaired region	13

B

Binding Site Nucleotide	Frequency
29	1
25	1
16	1
13	2
12	3
11	3
10	3
9	6
8	1
7	15
6	10
5	19
4	18
3	6
2	64

C

D

E
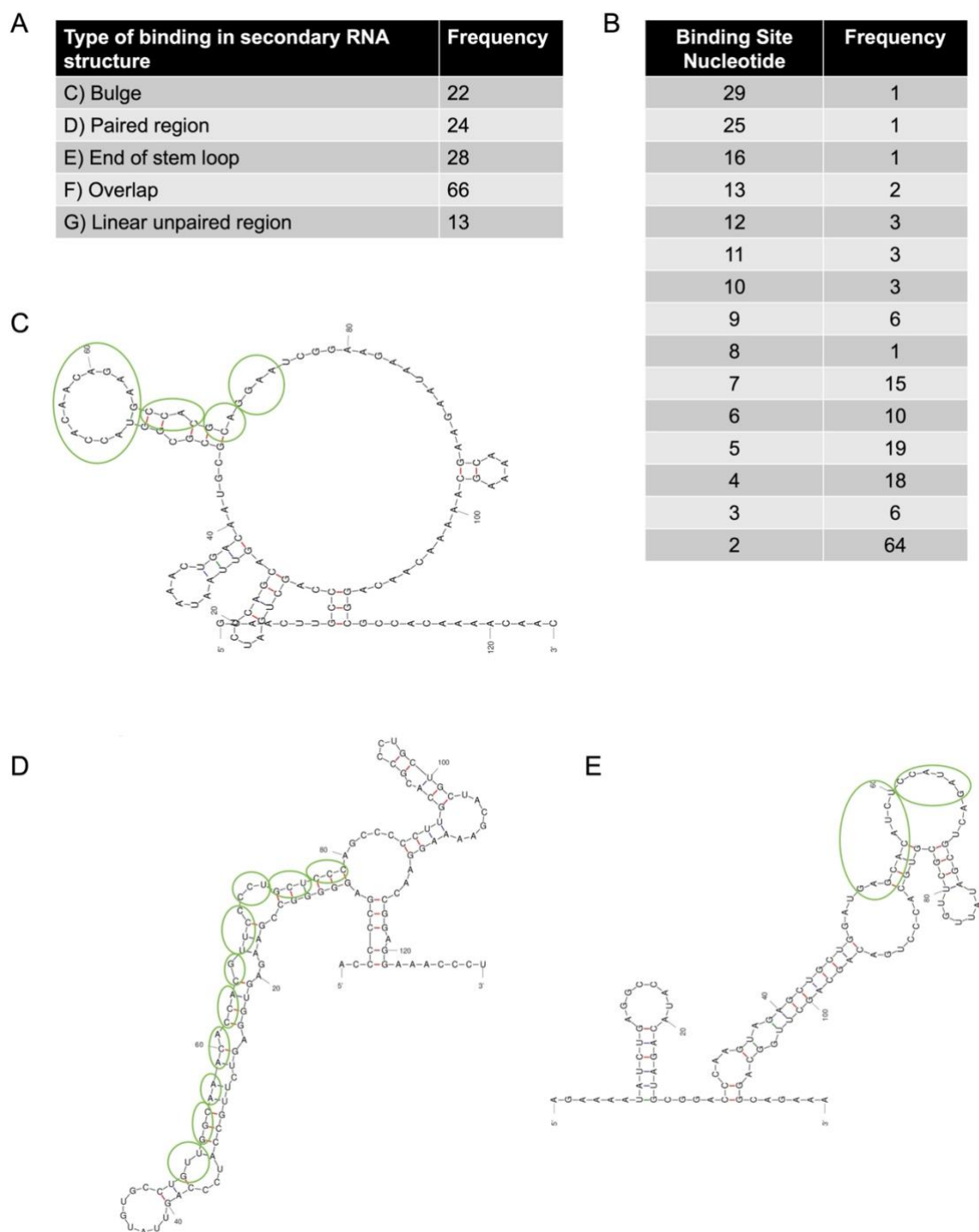
Figure 4.19 Secondary structures identified from iCLIP experiment 2 binding sites flanked with 50 nucleotides on both sites. Binding nucleotides are circled in green. A) Types of binding in the secondary RNA structures and their frequencies. B) Binding site nucleotide length and their frequencies. C) mFold representation of the 25 nucleotide-long binding site. D) mFold representation of the 29 nucleotide-long binding site. E) mFold representation of the 16 nucleotide-long binding site.

4.1.2.17 nsP3 global binding profile

Following the establishment of a robust iCLIP system for investigating the binding of nsP3 to the CHIKV genome RNA, the experiment was repeated to increase the confidence of the downstream bioinformatics analyses. In the second iCLIP experiment, alignments to the CHIKV genome and a single binding site was observed in the WT ICRES-CHIKV infected sample. To rule out that this is the result of nsP3 binding, a mock infection was included. Experimentation, sequencing, and bioinformatics were performed as previously described. No alignments to the CHIKV genome, and thus no crosslink/binding sites were observed in the mock infected sample. As observed before, alignments and crosslink/binding sites, albeit at a low level were present in the WT ICRES-CHIKV infected sample. The inclusion of a mock infected sample confirms that the results seen in the WT ICRES-CHIKV sample is likely due to residual RNA on the immunoprecipitation beads, rather than true binding from nsP3 contamination.

The alignment coverage, crosslink, and binding profiles of all three repeats are shown in Figure 4.20. The read alignment for each experiment appears distinct, without specific conservation (Figure 4.20, A - C). This also applies to the crosslink and binding sites (Figure 4.20, D - I). Whilst the crosslink/binding sites from experiment two could be binned into distinct binding regions, the profile of the third experiment is comparatively more dispersed, and binding is observed across the entire CHIKV genome. The data presented here show that nsP3 biding to the CHIKV genome may not be driven by a single consensus motif, or an obvious secondary RNA structure.
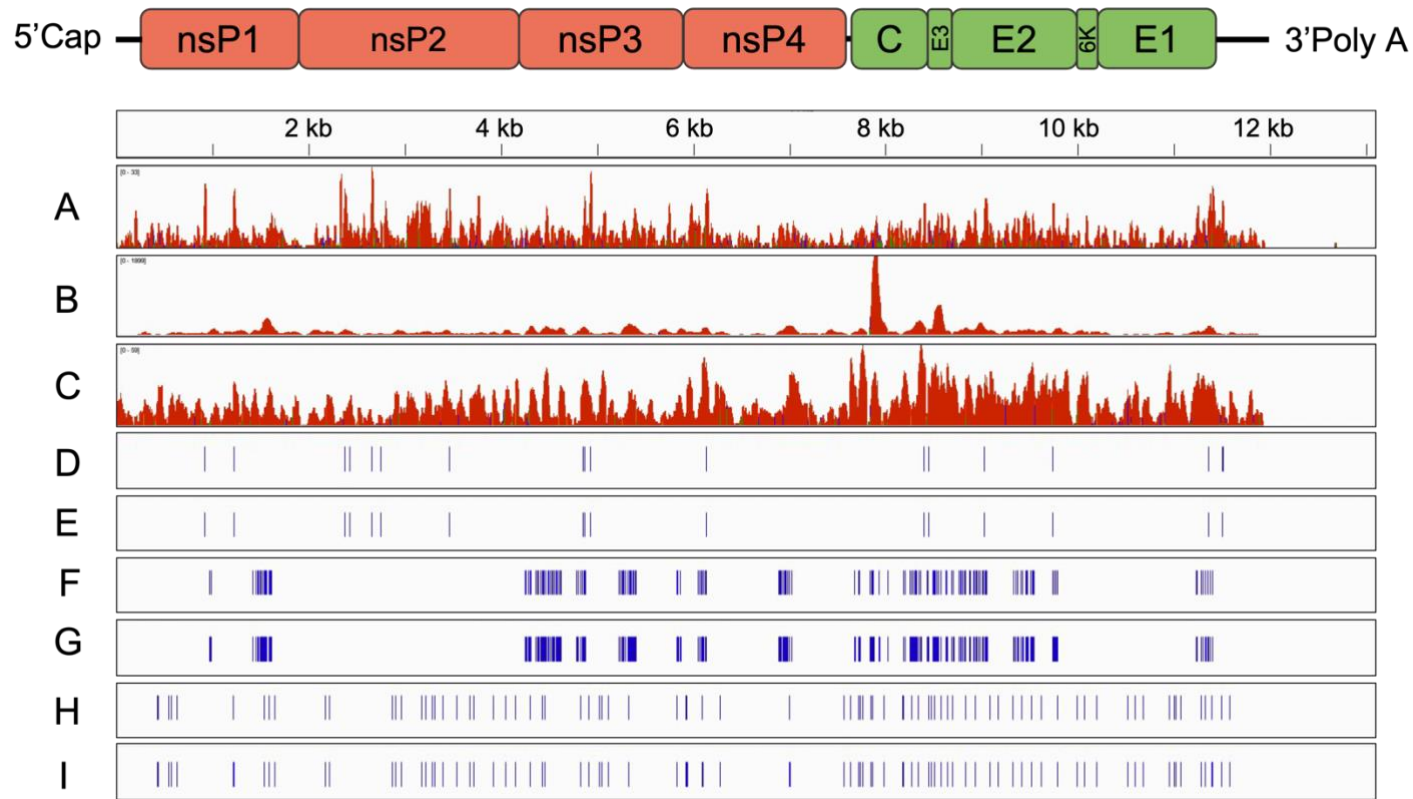
Figure 4.20 Global overview of nsP3 alignment, crosslink and binding profiles. A) Alignment profile of iCLIP experiment 1. b) Alignment profile of iCLIP experiment 2. c) Alignment profile of iCLIP experiment 3. d) Sites of crosslink identified from iCLIP experiment 1. e) Binding sites identified from iCLIP experiment 1. f) Sites of crosslink identified from iCLIP experiment 2. g) Binding sites identified from iCLIP experiment 2. h) Sites of crosslink identified from iCLIP experiment 3. i) Binding sites identified from iCLIP experiment 3.

4.1.3 Discussion

To date, limited information is available on the RNA binding ability of CHIKV nsP3 even though as a member of the replication complex it is predicted to interact with the viral RNA. Furthermore, previous nsP3-RNA binding experiments have been focused on individual domains, performed in an *in vitro* nature (Gao et al., 2019; Malet et al., 2009). To address the gap of knowledge on this, the studies undertaken here describe the establishment of a robust iCLIP technique which can be applied to CHIKV nsP3, to elucidate the genome binding capability of this protein

4.1.3.1 CHIKV nsP3 can be immunoprecipitated from virus infected and UVC-irradiated cell lysates

In general, an intact nsP3 HVD is required for alphavirus replication but targeted modifications in this region can be tolerated. This provides an opportunity to introduce affinity tags for purification of nsP3 and has been demonstrated for SINV (Liang and Li, 2012) and CHIKV (Gao et al., 2019; Remenyi et al., 2017). In this study, the introduction of a twin-strep-tag cloned into HVD (TST ICRES-CHIKV) was compared to WT ICRES-CHIKV and showed no adverse effects on infectious virus production. Therefore, it is an appropriate method of nsP3 purification, consistent with previously published data. Further, whilst UVC irradiation can induce protein unfolding and negatively impact immunoprecipitation (Urdaneta and Beckmann, 2020), the conclusion from this section can be made that nsP3 can be successfully immunoprecipitated from cell lysates following UVC treatment, indicating that the purification system used here is viable for iCLIP application.

4.1.3.2 RNase digested nsP3-RNA complexes can be purified from virus infected and UVC-irradiated cell lysates

NuPAGE gel analysis of purified and RNase-digested nsP3-RNA complexes showed that virus infection, UV crosslinking, RNase digest, adapter ligation and immunoprecipitation can be robustly performed for CHIKV nsP3. Of note, nsP3 has a molecular weight of ~60 kDa but the protein is usually located around 80

kDa on NuPAGE/SDS-PAGE analyses therefore the appropriate protein-RNA complexes should have a minimum weight of 80 kDa which then increase in weight as the length of bound RNA fragment increases. However, signal is observed on Figure 4.5B lanes 2-4 from ~35 kDa upwards. This observation suggests that perhaps a smaller RBP co-immunoprecipitated with nsP3. The most obvious co-purified candidates are the other members of the replication complex; however, this is unlikely as the molecular weight of nsP1, nsP2 and nsP4 are ~60 kDa, ~90 kDa and ~70 kDa, respectively (Rupp et al., 2015). The co-purified protein is an interesting point for further study and could be identified by Coomassie blue staining of the protein-RNA complexes resolved using SDS-PAGE gel electrophoresis. Any distinct bands below the expected molecular weight of nsP3 can then be excised and subjected to mass spectrometry analysis. The interactions of nsP3 with other proteins is an inherent artefact of viral infection (Götte et al., 2018), but contaminating RNA bound by co-purified proteins can be mitigated by RNase digestion. Therefore, given that the RNA fragments taken forward for downstream processing are all above the molecular weight of nsP3, off-target RNA contamination should be limited.

The inclusion of established positive and negative controls are imperative for the success of CLIP experiments. HuR was used as a positive UVC treatment control as this protein is highly susceptible to crosslink formation following irradiation (Urdaneta and Beckmann, 2020). In the first iCLIP experiment, HuR did not produce a positive signal, but a strong positive result was observed in the second experiment. Positive control failure was likely due to problems in immunoprecipitation using the anti-huR antibody or during 3' adapter ligation rather than UVC treatment as positive signal is observed for other samples from the same cell lysate. Nevertheless, the HuR result observed in the second experiment demonstrates that HuR is a robust positive control in iCLIP.

Parallel to the bioinformatics results observed for the TST ICRES-CHIKV, an essential concomitant is to interrogate the data generated from the negative controls (Ule et al., 2018). Four negative controls were included in the iCLIP experiment: WT ICRES-CHIKV infected cell lysate subjected to

immunoprecipitation using the same anti-strep-tag antibody used to purify TST-nsP3, a 'No RNase' control, a 'No antibody control' and an IgG control. The latter three were included to control for the RNase digestion and purification experiments whilst the first control was specifically designed to determine data specificity following sequencing. The absence of detectable signal during NuPAGE analysis demonstrates that the positive signals seen in the TST ICRES-CHIKV samples are true.

4.1.3.3 cDNA libraries can be prepared from excised nsP3-RNA complexes

To sequence the nsP3 bound fragments, the RNA must be reverse transcribed into cDNA then PCR amplified. The result (Figure 4.5) shows that each of the steps addressed here were successful and comparable to published data by *Huppertz et al* therefore cDNA libraries can be obtained from iCLIP performed on CHIKV nsP3. Whilst a desired band was observed at both 24 and 26 PCR cycles, further optimisation could be considered at single PCR cycle increment to increase the concentration of target insert whilst decreasing the presence of secondary products. Final PCR amplification was performed at 26 cycles, and the published iCLIP protocol recommends using AMPure XP beads for purification of PCR products but experimentation performed here found that this was not appropriate. Instead, purification for the CHIKV cDNA libraries was performed using gel extraction following gel electrophoresis of the entire PCR product. This method also allowed any secondary products from overamplification to be physically avoided and is a robust alternative to magnetic bead purification.

4.1.3.4 nsP3 binds stochastically across the CHIKV genome

To date, available studies suggest that CHIKV nsP3 binds to oligonucleotides promiscuously without specificity. In agreement with this, the first iCLIP experiment highlighted 34 crosslink sites which were then computationally combined into 17 binding sites. The location of binding sites was then represented as encoded gene or 999 nucleotide-long regions, but no immediate

binding pattern could be deduced as the identified sites were spread out across the length of the genome and there was no marked difference in the frequency of binding sites at any given gene or region. Interestingly, all the binding sites were observed on the positive strand, suggesting a preference of nsP3 for the plus-sense genome. This could in parts be explained by the intrinsically disproportionate ratio of positive to negative strand present during the virus lifecycle (Kendall et al., 2019). The binding profile of the second iCLIP experiment differed from the first experiment and could be binned into three regions (Figure 4.17A and Figure 4.17C). It is not immediately clear why nsP3 would preferentially bind to the three distinct regions and as every effort was made to limit natural variation between iCLIP experiments, the disparities between the two data sets suggests that nsP3 binding to the CHIKV genome is stochastic. In general, RBP-RNA interaction studies are limited for viruses. However, the alignment of in vitro and in vivo promiscuity of a viral RNA binding protein has been previously demonstrated for influenza virus where the observation that neuraminidase protein binds to viral RNA without sequence specificity was confirmed using PAR-CLIP (Lee et al., 2017).

4.1.3.5 nsP3 binding to the CHIKV genome is not driven by a consensus motif or secondary RNA structures

Limited data exist on the capacity of nsP3 to bind RNA, as such, there is currently no published studies on how nsP3 can recognise target RNA as a signal for binding. RBP-RNA interactions are usually mediated by RBP recognition of a binding motif, or secondary structures present in the target RNA (Corley et al., 2020). Analysis performed in this study using MEME represents the first insight into how nsP3-RNA interactions are achieved and showed that a consensus motif has little influence in binding as all the identified motifs were poorly conserved across the binding sites. In Group A Simian Agent 11 rotavirus, which possess a segmented dsRNA genome, a consensus sequence is required for the binding of nsP3 to viral mRNA to promote translation. However, it was shown here that the recognition of the 3′ consensus sequence is achieved using a network of hydrogen bonds, van der Waals interactions, salt bridges and

stacking interactions suggesting a specific bonding requirement beyond a consensus motif (Deo et al., 2002).

On the other hand, initial mFold prediction showed that nsP3 preferentially binds to unpaired regions, which is the case for many well-known RBPs (Orenstein et al., 2018), due to increased spatial availability around unpaired RNA. However, analysis of the improved dataset showed no specific binding preference to secondary RNA structures. Notably, it is important to consider the robustness of using iCLIP data to identify secondary structures that are targeted for binding. iCLIP is designed for single nucleotide RBP-RNA binding, and short RNA sequences are unlikely to be able to form stable secondary structures. For this study, identified binding sites were flanked with 50 nts on the 5′ and 3′ end to facilitate mFold analysis, but it is possible that bias can be introduced by the selection of this nucleotide length.

Most of the binding sites identified across the iCLIP experiments were at single nucleotide resolution, but a small number of longer binding sites were also observed. The two longest binding sites (Figure 4.19C and Figure 4.19D) were in the RNA encoding structural proteins whilst the 16-nucleotide binding site (Figure 4.19E) lie in the gene encoding nsP3. Information supporting the significance of these binding sites is currently unavailable in literature, therefore whilst the presence of such large binding sites is interesting, no obvious conclusions can be drawn from this data. Further investigations can include the disruption of RNA structures formed by these binding sites to observe effects on the virus lifecycle.

An overview containing the alignment coverage, identified crosslink and binding sites from three separate iCLIP experiments is shown in Figure 4.20. The promiscuous nature of nsP3 interactions with RNA highlighted in this in vivo study agrees with in vitro data published in literature where the protein can bind a variety of RNA oligonucleotides of unrelated sequences (Malet et al., 2009). nsP3 has also been shown to bind promiscuously to both CHIKV 3′ UTR and

Hepatitis C virus 3′ UTR in in vitro RNA filter binding assays even though these two sequences do not share obvious conservation (Dr. Yanni Gao, personal communication, University of Leeds). This could, in part, be explained by patches of positively charged, basic amino acids present in all three domains of nsP3: around the ADP-ribose binding site in the macrodomain (Malet et al., 2009), in proximity to the zinc coordination site of the AUD (Shin et al., 2012), and the string of amino acids immediately following the polyproline motif in the hypervariable domain (Tossavainen et al., 2016).

4.2 nsP3 interactions with the host transcriptome

4.2.1 Introduction

In order to modify the host cell into an optimal environment for infection, viruses often modulate multiple cellular processes including gene expression, transcription, translation and protein folding. Alphaviruses have been shown to induce transcription shut-off in host cells during infection. SINV causes downregulation of cellular mRNA synthesis to almost undetectable levels, which is maintained throughout infection. SINV infection also negatively impacts the synthesis and processing of pre-ribosomal RNA (Gorchakov et al., 2005). Combined, this transcriptional shutoff was theorised to be an efficient immune evasion method (Frolova et al., 2002). Transcriptional shut off capability has been mostly attributed to the non-structural proteins in both SINV and SFV (Garmashova et al., 2007), as sub genomic replicons that do not contain the structural proteins can also exert this effect (Frolov et al., 1999).

RNA-seq analysis has been performed in CHIKV infection of a mouse model, where RNA was extracted from infected tissues over peak viraemia and symptomatic chronic arthritis. Sequence analysis identified up-regulation of genes and pathways involved in the immune response such as those that function in the interactome and granzyme A (Wilson et al., 2017). However, this study focused on the global RNA expression profile during infection rather than how viral proteins can interact with host RNA, thus whether nsP3 can specifically bind to host RNA remains unanswered.

nsP3 binding to CHIKV RNA was addressed in the first part of this chapter, where the results show that this protein binds stochastically to the viral genome. During bioinformatics analysis of the iCLIP sequencing data, it was observed that the maximum number of TST barcoded reads that aligned mapped to the CHIKV genome was 20.5%. Therefore, the remaining unmapped sequences were explored further to see whether they align to the host transcriptome, leading to identification of host binding sites.

4.2.1.1 Aims

This part of the project aims to determine whether CHIKV nsP3 targets host RNA during infection by aligning iCLIP sequencing results to the host transcriptome. Aligned reads are then processed to locate crosslink sites. The frequency of crosslinking is then examined to explore whether significant host transcript interactions can be observed.

4.2.2 Results

4.2.2.1 Bioinformatics overview

The bioinformatics pipeline used to analyse nsP3 binding to the host transcriptome follows the same pathway described in 4.1.2.5 with minor adaptations (Figure 4.21). Specific software and arguments used are described in Appendix Table 6.4. Briefly reads were filtered by quality and length, then de-multiplexed and adapter trimmed. The baby hamster transcriptome files were obtained from Ensembl (Genome assembly: MesAur1.0 GCA_000349665.1), while the viral genome was accessed from the National Centre for Biotechnology Information database (GenBank: MT668625.1). Reads that did not align to the baby hamster kidney transcriptome were extracted and mapped to the CHIKV genome and both alignment files were then merged. Resultant genomic intervals were then subjected to rearrangement and manipulations to account for the cross-linking. Overlapping intervals were then counted.
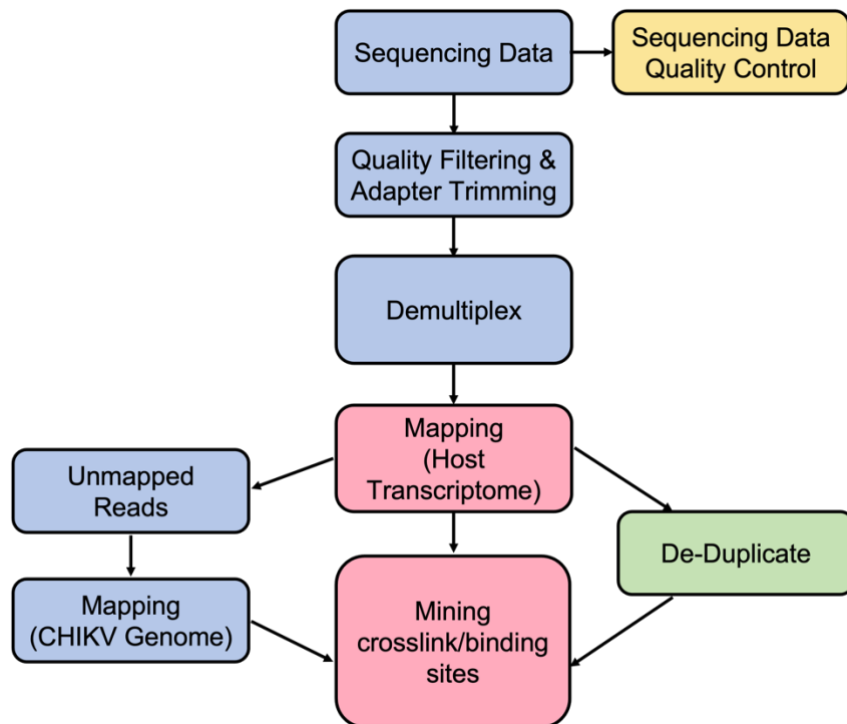
Figure 4.21 Bioinformatics pipeline used to determine nsP3 binding to the BHK transcriptome. Reads were first mapped to the host transcriptome and any remaining reads were aligned to the CHIKV genome.

4.2.2.2 nsP3 targets host transcripts during infection

In the viral genome analysis, PureCLIP was used to determine crosslink and binding sites. However, due to the low number of sequencing reads obtained using the amplicon EZ method, it was not possible to use PureCLIP for the host transcriptome analysis. Instead, crosslink sites were determined by computationally shifting transcriptomic interval entries 1 nucleotide upstream to the location where the crosslinked nucleotide was present. Whilst it was not possible to use PureCLIP to determine binding sites, crosslink sites provide an indication of where the binding sites are located therefore in this part of the project binding describes the direct interaction between nsP3 and host transcript, rather than an identified binding site.

Multiple transcripts can be generated from a single gene, therefore the coverage of crosslink sites across the genome was then summed at both gene (Figure 4.22) and transcript levels (Figure 4.23). In the CHIKV genome analysis two crosslink sites were observed in the WT pool. For host transcriptome crosslink counts in the WT pool, 6 transcripts had a count over 1 with 14 as the highest number (Figure 4.23A). On the other hand, 156 transcripts had a crosslink count over 1 in the TST pool with the maximum count at 116 (Figure 4.23B). Like the binding profiles observed in the CHIKV genome, TST pool had a much higher frequency of crosslinked nucleotides. Comparing the differences between individual transcripts and genes associated with crosslinked transcripts, there were no gene duplicates in the WT ICRES CHIKV pool. In the TST ICRES CHIKV pool, two genes were associated with more than one crosslinked transcript: one gene was associated with three crosslinked transcripts, and the other with two crosslinked transcripts. Interestingly, when the data obtained from CHIKV RNA crosslinking was included alongside this data, it showed that nsP3 targeted CHIKV RNA over 100-fold more than the host transcriptome in the TST read pool, indicating specific preference for the viral genome. Notably, 100% of targeted biotype of transcripts were protein coding which suggests that such interactions function to modulate the transcription or translation of host RNA
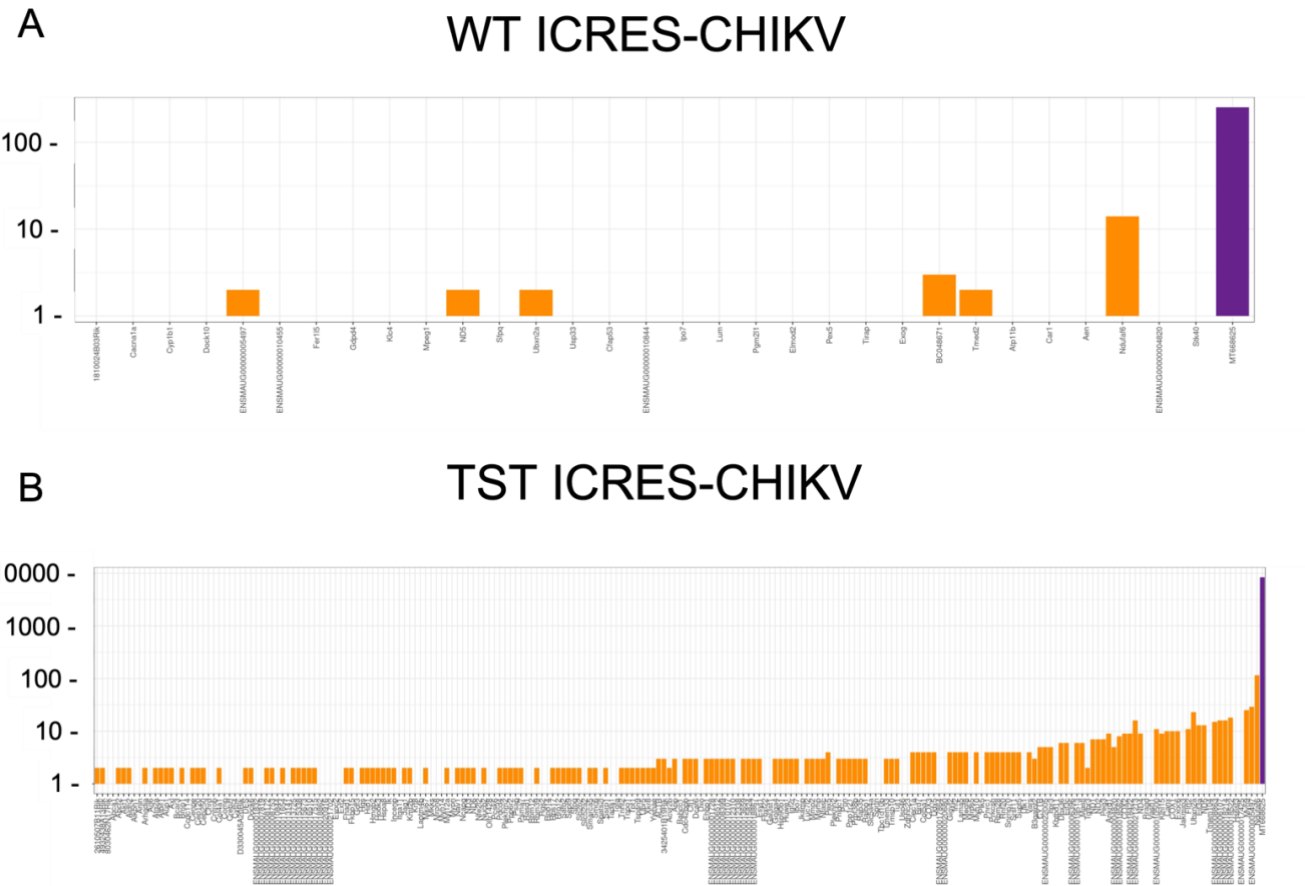
Figure 4.22 Genes targeted by nsP3 (orange). A) Crosslink counts identified in the WT ICRES-CHIKV pool. For the full list of genes on the x-axis, see Appendix Table 6.8. B) Crosslink counts identified in the TST ICRES-CHIKV pool. Purple bar represents CHIKV RNA binding. For the full list of genes on the x-axis, see Appendix Table 6.9.
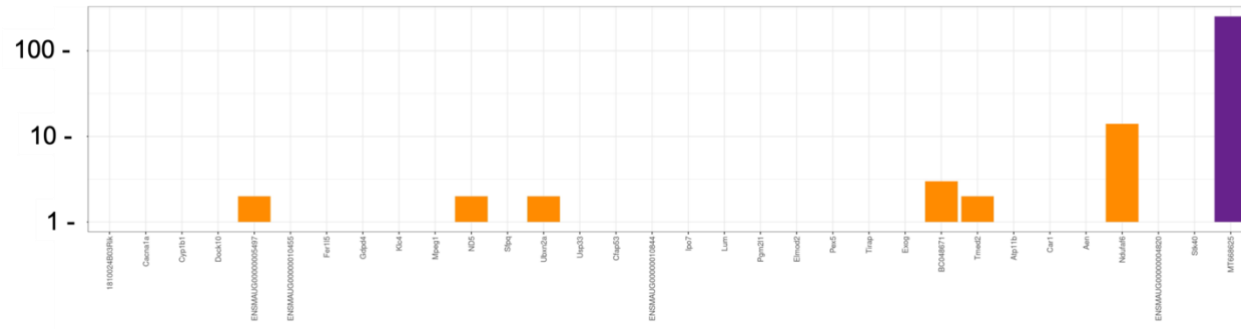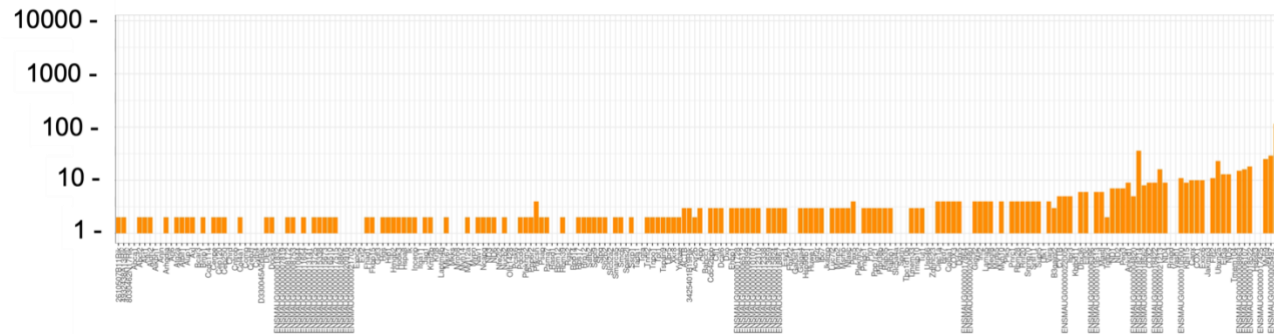
Figure 4.23 Transcripts targeted by nsP3 (orange). A) Crosslink counts identified in the WT ICRES-CHIKV pool. For the full list of genes on the x-axis, see Appendix Table 6.8. B) Crosslink counts identified in the TST ICRES-CHIKV pool. Purple bar represents CHIKV RNA binding. For the full list of genes on the x-axis, see Appendix Table 6.9.

4.2.2.3 Specific host transcripts are targeted by nsP3 during infection

To investigate whether specific host transcripts targeted by nsP3 were reproducible, the same analysis was performed for all 3 iCLIP experiments. This produced 1221 transcripts that were targeted at least once by nsP3 in the TST pool. Significantly targeted transcripts were then extracted by applying a 10-count cut off, yielding 23 distinct transcripts (Figure 4.24A). When the same analysis was applied to the WT sample, 4 distinct transcripts were identified (Figure 4.24B).

4.2.2.4 nsP3 binds to host transcripts with specific functions in mitochondrial complex assembly, vesicle trafficking, membrane assembly & integrity, and cell adhesion/migration

During data analysis, it was noticed that the most frequently targeted RNAs included a high number of transcripts for which a specific function has not been assigned, these are designated with 'ENSMAUG' followed by a string of numbers shown in Figure 4.24A. For this reason, they were eliminated from further analysis. The five transcripts with the highest count and an associated function were identified as NADH:ubiquinone oxidoreductase complex assembly factor 6 (NDUFAF6), myosin heavy chain 9 (MYH9), UBX Domain protein 2A (UBXN2A), heparan sulphate 6-O-sulfotransferase 3 (HS6ST3) and transmembrane protein 104 (TMEM104) (Figure 4.25A). Further to this, transcripts which appeared in more than one iCLIP experiment were also highlighted and shown in Figure 4.25B which were HDUFAF6, Fibronectin leucine rich transmembrane protein 2 (FLRT2) and Exocyst complex component 6 (EXOC6). The seven transcripts identified can be categorised by their functions: NUDFAF6 has functions in the mitochondrial respiratory complex I, MYH9 and EXOC6 play important roles in vesicular trafficking, particularly associated with the Golgi, UBXN2A and TMEM104 are important for membrane assembly/integrity. Finally, HS6ST3 and FLRT2 function in cell adhesion and migration.

TST ICRES-CHIKV

A

Transcript	Count	Experiment
NDUFAF6	116	2
ENSMAUG00000005497	29	2
MYH9	25	2
ENSMAUG00000017425	24	1
UBXN2A	23	2
HS6ST3	19	1
ENSMAUG00000018522	18	2
ENSMAUG00000006107	16	2
ENSMAUG00000021213	16	2
ENSMAUG00000008863	15	2
TMEM104	14	1
EIF3A	13	2
FLRT2	13	3
ND5	13	2
ENSMAUG00000018214	12	2
NDUFAF6	12	3
ENSMAUG00000010500	11	2
FLRT2	11	2
NDUFAF6	11	1
CANX	10	2
COX1	10	2
ENSMAUG00000004820	10	3
EXOC6	10	3
EXOC6	10	2
JAKMIP3	10	1
KLHL15	10	3
TGFBR3	10	1

B

WT ICRES-CHIKV

Transcript	Count	Experiment
NDUFAF6	14	2
STK40	13	3
TMED2	12	1
ENSMAUG00000004820	11	2

Figure 4.24 High frequency transcript counts in A) TST ICRES CHIKV sample and B) WT ICRES CHIKV sample.

A

Transcript	Count	Experiment
NDUFAF6	116	3
MYH9	25	3
UBXN2A	23	3
HS6ST3	19	1
TMEM104	14	1

B

Transcript	Total Count	Number of Experiments
NDUFAF6	139	3
FLRT2	24	2
EXOC6	20	2

Figure 4.25 Specific transcripts targeted by nsP3. A) The 5 most frequently targeted transcripts with characterised function B) Transcripts which appeared more than once over the three iCLIP experiments and the total count.

4.2.3 Discussion

The results presented in this section represents the first *in vivo* analysis dissecting the direct binding of host transcriptome by CHIKV nsP3. It was confirmed here that during infection nsP3 indeed targets host RNAs, with specific transcripts being targeted at a high frequency. Whilst it was not possible to pinpoint specific motifs or secondary structures that direct nsP3 binding to the CHIKV genome, the combined results from both viral and mammalian analyses demonstrates that nsP3 preferentially binds to viral RNA, the frequency of which was 100-fold higher than host RNA. This indicates that nsP3 can specifically distinguish between self and foreign RNA. It could also suggest that whilst nsP3 can target host transcripts, it may not be the primary function of this protein and the modulation of host RNA may be afforded by another viral protein. Simultaneously, the low level of binding to host RNA could be due to a high proportion of viral RNA to host RNA *in vivo* during infection as shown in the RNA-seq analysis performed by *Wilson et al*, therefore leading to lowered availability of RNA for direct binding (Wilson et al., 2017). Another explanation for preference of viral RNA over host RNA could be the subcellular localisation of nsP3 in spherules or replication complexes at the cytoplasmic membrane during early

stages of CHIKV infection (Froshauer et al., 1988). Here, levels of host RNA would be much lower and therefore naturally there would be less binding.

Nevertheless, specific 7 BHK transcripts that were targeted at high frequency and those that were reproducible over multiple iCLIP experiments were identified from this study. The transcript that was targeted the most frequently was NDUFAF6 (Figure 25A and Figure 25B), which encodes a mitochondria-associated protein and functions in the assembly of NADH-ubiquinone oxidoreductase complex I of the mitochondrial respiratory chain. The relationship between nsP3 and mitochondria-associated proteins has been previously shown by mass spectrometry analysis (Dr Yanni Gao, personal communication, University of Leeds). The observed binding shown in the iCLIP experiment corroborates that nsP3 targets mitochondria-associated transcripts and proteins during infection. However, it is important to highlight that binding to the NDUFAF6 transcript was also observed in the WT sample, albeit at 10-fold lower than the TST sample, and that this transcript is intrinsically expressed at a high level in cells with increased energy demands (Mercer et al., 2011). Thus, future experiments exploring the interaction between nsP3 and mitochondrial RNA should take into account the relative abundance of this transcript.

The remaining 6 host transcripts that were identified fell into three distinct functional categories: golgi associated trafficking, membrane assembly/integrity and cell adhesion/migration, cellular processes that are important during virus infection. Interestingly whilst limited information exists on how most of these transcripts are targeted by viruses, FLRT2 has previously been observed to be downregulated during West Nile virus infection (Maximova et al., 2021). iCLIP provides information on protein-RNA binding, but it does not shed light on the downstream effects mediated by such interactions. Therefore, it would be interesting to explore whether FLRT2 downregulation also occurs in CHIKV infection and confirm whether this is due to direct targeting of FLRT2 by nsP3. Clearly, the contribution of these specific transcripts to CHIKV infection requires further investigation, particularly those that were only targeted in a single iCLIP experiment such as UBXN2A, HS6ST3 and TMEM104.

Combined, the bioinformatics analyses of iCLIP results demonstrates that nsP3 preferentially targets the viral genome under the experimental conditions that were performed in this project. The interaction between nsP3 and viral RNA appear to be stochastic, and the data presented suggest that this binding may not be driven by a single consensus binding motif or a specific conformation of secondary RNA structure. Furthermore, the results revealed several specific host transcripts that are targeted by nsP3 during infection which have not been previously described. Continuation of this project could include functionally characterising the importance of the identified binding sites.

Chapter 5 Discussion and future perspectives

Despite decades of research, nsP3 remains the most enigmatic of the CHIKV non-structural proteins. Whilst it plays clear roles in the virus lifecycle, the specific mechanism behind these functions are not well defined. nsP3 consists of three distinct domains: the macrodomain which possesses ADP-ribose binding and hydrolase activities, the AUD which binds zinc and RNA and has roles in subgenomic RNA transcription, and a HVD, important for interactions with host proteins (Gao et al., 2019; Götte et al., 2018; Malet et al., 2009). The HVD, defined by its name, is poorly conserved between other members within the alphavirus genus and even between strains of the same virus. However, specific features such as a polyproline motif and repeated elements containing FGDF motifs are always maintained. The relationship between the CHIKV HVD and specific host proteins including amphiphysin and G3BP1 has been attributed to the polyproline motif and two FGDF repeats, respectively. In the case of amphiphysin, most studies to date have focused on the loss of nsP3 binding due to various mutations in the polyproline motif, but whether specific residues are important within the virus lifecycle is unclear.

This first part of this study set out to characterise the function of the CHIKV nsP3 HVD polyproline motif in the virus lifecycle. Initially, a systematic mutagenesis approach, guided by conservation analysis of the polyproline motif in other alphaviruses, was performed in a subgenomic replicon system. The panel of mutants were able to replicate well in all six tested mammalian and mosquito cell lines, with most reaching wildtype levels, indicating that the targeted residues are not important for CHIKV replication. However, as subgenomic replicon systems only recapitulate the early stages of the virus lifecycle, the mutations were then tested in the context of infectious virus. Here, all of the mutants were able to produce infectious virus to levels comparable to wildtype, except for a double mutation - P398A/P401A where infectious virus production was completely abrogated. This result led to an investigation into the ability of the mutants to express viral proteins. Western blot analysis of non-structural and structural proteins showed similar phenotype: all of the mutants, except for P398A/P401A, were able to express nsP1, nsP3 and capsid protein to high levels. The P398A/P401A mutation had consistently lowered expression of all the tested

CHIKV proteins. qRT-PCT analysis of this mutant also showed impaired genome RNA production. The phenotype of P398A/P401A was further investigated by reverse cloning of a wildtype insert into the backbone of the P398A/P401A mutant. This showed that the observed phenotype associated with this mutant was not because of the targeted polyproline motif nucleotide changes but rather an off-target mutation outside of the polyproline motif. Sequencing revealed four C-to-U mutations located in the 3′ UTR of the P398A/P401A mutant. Two of the mutations were particularly interesting as they were identified in proximity to the 3′ CSE, possible disrupting the requirement of this feature in minus-strand synthesis. The data presented in this project provide a strong argument that neither the CHIKV nsP3 HVD polyproline motif, nor the positively charged residues immediately adjacent are essential for CHIKV replication or production of infectious virus. However, the 3′ UTR mutations present an interesting focus for continued investigations, which should aim to characterise how these nucleotide changes contribute to the observed phenotypes.

As a component of the replication complex, nsP3 must bind to and interact with RNA. All three nsP3 domains have predicted or demonstrated RNA binding capability, but in-depth analyses on how this protein interacts with RNA in vivo have never been performed. The second part of this study focused on characterising nsP3 as an RNA binding protein. To achieve this, iCLIP, which allows determination of protein-RNA interactions down to single nucleotide resolution, was used. The iCLIP experiments performed in this study were based on the published protocol developed in the Ule laboratory which uses specific Illumina adapters for PCR amplification. For the application in this study, it was identified that the adapters were not compatible with the sequencing method that was used and led to poor quality sequencing results. Adaptation of the PCR primers produced a better sequencing outcome, allowing a clear nsP3-RNA profile to be characterised.

Using the optimised iCLIP technique combined with appropriate bioinformatics analysis, it was determined that nsP3 binds across the entire viral genome without obvious specificity. Across multiple repeats of the experiment, nsP3

bound to CHIKV RNA stochastically. Specific binding sites on the genome, and the frequency of binding changed each time. The data from the first iCLIP attempt revealed that nsP3 binding to RNA was not driven by a specific recognition motif, which was confirmed in subsequent repeats. The same data suggested that secondary structures influenced how the protein targets RNA, with preferential binding to unpaired regions. However, this was not the case in other iCLIP experiments where the number of binding sites increased and there was no obvious trend between binding to paired or unpaired regions of RNA.

As viruses routinely remodel the host cellular environment by targeting transcription and translation, the binding of nsP3 to host transcriptome was also interrogated. Whilst an obvious determinant for nsP3 binding to viral RNA could not be identified, nsP3 does favour the virus genome over host RNA. This could be driven by the subcellular localisation of nsP3 into spherules during early stages of infection. Here, levels of host RNA is likely to be significantly lower than viral RNA. The cellular localisation of nsP3 changes during the later stages 1.2.3, and the protein can be found in cytopathic vacuoles or in large aggregates. It is likely that in these environments, the levels of host RNA is altered, therefore it would be interesting to investigate whether the binding profile of nsP3 to host RNA would change accordingly.

Nevertheless, nsP3 did indeed target specific host transcripts at a high frequency. One of these transcripts, NDUFAF6, plays a role in mitochondrial function. Interaction between nsP3 and mitochondrial proteins was first described in mass spectrometry analysis performed in our laboratory, and this observation provides increased confidence that nsP3 may modulate the mitochondria during infection.

The results obtained in this part of the project represents a first insight into how nsP3 targets both viral and host RNA. It is important to highlight that all of the iCLIP experiments described in this project were performed in baby hamster kidney cells, the gold standard cell line used for CHIKV virus experiments, as

they are highly permissive to infection and allow high levels of nsP3 protein expression to maximise RNA binding. However, as this organism and cell type are not natural CHIKV infection targets, the experiments should be performed in more appropriate cells lines such as RD cells for human host, and C6/36 cells for insect host. Work on optimising nsP3 expression, UVC treatment and immunoprecipitation in both these cell lines has been performed and confirms the feasibility of applying iCLIP to alternate cell lines for CHIKV investigations.

Furthermore, the iCLIP experiments performed here were all taken at a 12-hour time point, chosen for the peak expression of nsP3. However, evidence in literature has demonstrated that nsP3 may have different roles over the course of infection reflecting different functions depending on the stage of the virus lifecycle. In the early stages of infection, this is most likely to support genome replication. Therefore, it would be interesting to investigate how the binding profile of the protein to both viral and host RNA changes by performing iCLIP over a time-course of infection.

The work carried out in this project has increased knowledge of both a specific domain of nsP3, and the whole protein. Specifically, it has confirmed that a polyproline motif located in the hypervariable domain is not essential for the virus lifecycle. In addition, this is the first comprehensive analysis of how nsP3 can target RNA during infection.

Chapter 6 Appendix

Table 6.1. List of oligonucleotide primers used in this project

Mutation	Orientation	Site-directed mutagenesis (Quickchange or Q5)
P398A	Forward	GAGCACCGTACGGGTCGCGCCGC
P398A	Reverse	ATTACCCAATCAGACACGGC
P401A	Forward	ACCTGTCGCGGCCCCCAGAAGAA
P401A	Reverse	ACGGTGCTCATTACCCAATC
R403A	Forward	CGCGCCGCCCGCAAGAAGGCGAG
R403A	Reverse	ACAGGTACGGTGCTCATTACC
P398A/R403A	Forward	GCCCCCGCAAGAAGGCGAGGGAGAAAC
P398A/R403A	Reverse	CGCGACGGCTACGGTGCTCATTACCCA
PPR-A	Forward	GCCCCCGCAAGAAGGCGAGGGAGAAAC
PPR-A	Reverse	CGCGACGGCTACGGTGCTCATTACCCA
Sequencing Primer	Forward	CGAGTACAATCACGTCACTGACG
Sequencing Primer	Reverse	GTCCGACGAGAATATATACCCACC

Mutation	Orientation	PCR mutagenesis
R043E	First step Forward	CCCAAAGTACAAAATAGAAGGAGTGC
R043E	First step Reverse	TTCGGGCGGCGCGACA
R043E	Second step Forward	CGCCGCCCGAAAGAAGGCGAGGGAGAAACCTG
R043E	Second step Reverse	CCTAGGCCACCATGGAAGATG
P398A/P401A	First step Forward	CCCAAAGTACAAAATAGAAGGAGTGC
P398A/P401A	First step Reverse	TCTGGGGGCCGCGACGGCTACGGTGCTCATTACCCAATCAGAC
P398A/P401A	Second step Forward	GCCGTCGCGGCCCCCAGAAGAAGGCGAGGG
P398A/P401A	Second step Reverse	CCTAGGCCACCATGGAAGATG
RRRR-A	First step Forward	CCCAAAGTACAAAATAGAAGGAGTGC
RRRR-A	First step Reverse	CCTGCCGCTGCTGCGGGCGGCGCGACA

	Second step Forward	GCAGCAGCGGCAGGGAGAAACCTGACTGTGACATG
	Second step Reverse	CCTAGGCCACCATGGAAGATG
RRRR-E	First step Forward	CCCAAAGTACAAAATAGAAGGAGTGC
	First step Reverse	CCTTCCTCTTCTTCGGGCGGCGCGACA
	Second step Forward	CCCGAAGAAGAGGAAGGGAGAAACCTGACTGTGACATG
	Second step Reverse	CCTAGGCCACCATGGAAGATG
ΔP	First step Forward	CCCAAAGTACAAAATAGAAGGAGTGC
	First step Reverse	CTCTCGTCACAGGTGCTCATTACCCAATCAGAC
	Second step Forward	GAGCACC*TGTGACGAGAGAGAAGGG
	Second step Reverse	CCTAGGCCACCATGGAAGATG
Colony PCR primer	Forward	CCCAAAGTACAAAATAGAAGGAGTGC
	Reverse	GTCCGACGAGAATATATACCCACC

	Primer Name	**CHIKV genome sequencing**
Sequencing for CHIKV whole genome	CF1	AATTGGCAGCTGCCTATCGA
	CF3	GCAGACGGATTCCTGATGTG
	CF5	TACACCATATTGCGATGCAC
	CF7	CAACTATTAAGGAGTGGGAG
	CF9	GCAGCCTTCGTAGGACAGGT
	CF11	TGCTATTTGACCACAACGTG
	CF13	ACGTCCGATTGTCCAATCCC
	CF15	AAACATCACCATCGCCAGCC
	CS1	ATCGATAACGCGGACCTGGC
	CS3	ACCGCAGCACGGTAAAGAGC
	CS5	TAATGAGCGTCGGTGCCCAC
	CS7	GCAAGAAAGGCAAGTGTGCG
	CHR2	GCCCACTTACTGAAGGCTTG
	CHR4	TGCCAGATCCCGGTACTCCG

	CHR6	GCCCGCTGTCTAGATCCACC
	CHR8	TCCCGGTCCCCTTCAGACTC
	CHR10	CAGGTACGGTGCTCATTACC
	CHR12	GCTGCTGCCAGTACATTCTG
	CHR14	TTAGCGGGTCTGCCACTCTG
	CSR2	GCTCCTCCTAAGACTATGGC
	CSR4	GTGACCGCGGCATGACATTG
	CSR6	CGGTGAAGACCTTACAGCTG
	CSR8	CCTCCCGTGATCTTCTGCAC
	CSR10	CATCTCCTACGTCCCTGTGG

	Orientation	Reversion sequencing
Polyproline fragment amplification	Forward	CCCAAAGTACAAAATAGAAGGAGTGC
	Reverse	CATCTTCCATGGTGGCCTAGG
Sequencing	Forward	GGCATGGGTCAGAAGGAT T
	Reverse	GTCCGACGAGAATATATACCCACC

Primer Name	Sequencing Pool	Nanopore Sequencing
CHIKV_ICRES _1_LEFT	Pool 1	AGACACACGTAGCCTACCAGTT
CHIKV_ICRES _1_RIGHT	Pool 1	TGTTCGTATTCCGTTGCGTTCT
CHIKV_ICRES _2_LEFT	Pool 2	AATGTCATTCTCGGTGTGCACA
CHIKV_ICRES _2_RIGHT	Pool 2	TCCGAAGACTCCTATGACTGCA
CHIKV_ICRES _3_LEFT	Pool 1	TGTAAGAAGGAAGAAGCCGCAG

CHIKV_ICRES _3_RIGHT	Pool 1	AATAGCCCGCTGTCTAGATCCA
CHIKV_ICRES _4_LEFT	Pool 2	TTGTTGGGCTAAGAGCTTGGTC
CHIKV_ICRES _4_RIGHT	Pool 2	CTTTGCGACTTCTCGATAGGCA
CHIKV_ICRES _5_LEFT	Pool 1	CGCAAAAACAGTTATGTGCGGT
CHIKV_ICRES _5_RIGHT	Pool 1	TGTTTCTTGTACGACCGGACAC
CHIKV_ICRES _6_LEFT	Pool 2	GGTAATGAGCACCGTACCTGTC
CHIKV_ICRES _6_RIGHT	Pool 2	CTGCTGCTTTTGGCCCTTTTAG
CHIKV_ICRES _7_LEFT	Pool 1	ACCCACTTTGGACTCAGCAGTA
CHIKV_ICRES _7_RIGHT	Pool 1	ACCTATTTAGGACCGCCGTACA
CHIKV_ICRES _8_LEFT	Pool 2	ATGGCAACGAACAGGGCTAATT
CHIKV_ICRES _8_RIGHT	Pool 2	CGGTGGGGAGAACATGTTAAGG
CHIKV_ICRES _9_LEFT	Pool 1	ATCCCAGTTATGTGCCTGTTGG
CHIKV_ICRES _9_RIGHT	Pool 1	CGGCCAATACTTATACGGCTCG
CHIKV_ICRES _10_LEFT	Pool 2	AACACTCCTGTCCTACCGGAAT
CHIKV_ICRES _10_RIGHT	Pool 2	GCCAAATTGTCCTGGTCTTCCT
CHIKV_ICRES _11_LEFT	Pool 1	TTCATTGTGGGGCCAATGTCTT

CHIKV_ICRES_11_RIGHT	Pool 1	TTGTTACTATTCAGGGGTTTTATAGCC

		iCLIP 3′ infrared adaptor
		/5Phos/AG ATC GGA AGA GCG GTT CAG AAA AAA AAA AAA /iAzideN/AA AAA AAA AAA A/3Bio/

Primer Name	Barcode	**iCLIP Reverse Transcription Primers**
irCLIP_ddRT_13	CCGGA	/5Phos/ WWW TCCGG NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC
irCLIP_ddRT_14	AGGCA	/5Phos/ WWW TGCCT NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC
irCLIP_ddRT_16	TTTAA	/5Phos/ WWW TTAAA NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC
irCLIP_ddRT_17	CATTT	/5Phos/ WWW AAATG NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC
irCLIP_ddRT_18	ACCTT	/5Phos/ WWW AAGGT NNNN AGATCGGAAGAGCGTCGTGAT /iSp18/ GGATCC /iSp18/ TACTGAACCGC

Primer Name	Orientation	**iCLIP PCR Primers**
P3 Solexa	Forward	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTG AACCGCTCTTCCGATCT
P5 Solexa	Reverse	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATCT
Genewiz Adapter	Forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Genewiz Adapter	Reverse	GACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Table 6.2. Nucleotide sequence of the polyproline mutants generated in this project

Wildtype/Mutation	Sequence
Wildtype	CCTGTCGCGCCGCCCAGA
P398A	GCCGTCGCGCCGCCCAGA
P401A	CCTGTCGCGGCCCCCAGA
R403A	CCTGTCGCGCCGCCCGCA
R403E	GCCGTCGCGCCGCCCGAA
P398A/P401A	GCCGTCGCGGCCCCCAGA
P398A/R403A	GCCGTCGCGCCGCCCGCA
PPR-A	GCCGTCGCGGCCCCCGCA
Wildtype	AGAAGAAGGCGA
RRRR-A	GCAGCAGCGGCA
RRRR-E	GAAGAAGAGGAA

Table 6.3. Bioinformatics analysis tools, functions and parameters used for the virus genome

Tool	Function	Key Parameters
FastQC	Quality control of sequence reads	-Default parameters of Galaxy version 0.73+galaxy0
UMI-tools extract	Extraction of UMI barcode from a read and adding it to the read ID, leaving sample barcode in place	-Barcode pattern: NNNNXXXXXNNN -Barcode at the 5'end -Phred score threshold = 20
Je-Demultiplex metrics	Demultiplexing the barcoded libraries to separate pools according to supplied barcode file	-Supplied barcodes as text document -Remove barcode from reads -Keep unassigned reads
Map with Bowtie for Illumina	Mapping reads to the CHIKV genome	-Mostly unchanged from default settings (Galaxy version 1.2.0) except 'Suppress alignment reads to 1' which filters out reads that map to more than one place on the genome
BAM filter	Filtering unmapped reads from mapped BAM files	Keep only mapped reads

UMI-tools de-duplicate	Collapse reads with identical UMIs to remove duplicates that may have arisen from PCR amplification	-Extract UMI by ReadID, -Duplicates share the exact same UMI -Distance = 1 -Soft clip = 1
PureCLIP	Mining crosslink/binding sites	-Supplied Target BAM file and Genome reference file -Distance used to merge individual crosslink sites to binding sites = 5
Bedtools GetFastaBed	Generate fasta file for MEME analysis	-Supply a .txt file with flanked binding sites -Supply reference genome
MEME (Web-based MEME version 5.4.1)	Identify binding motifs from supplied sequences	-Classic Mode -Minimum width of motif = 5
Quikfold	Fast folding of multiple RNA sequences	-Energy rules = RNA 2.3 -Sequence type linear -Maximum of 1 folding

Table 6.4 Bioinformatics analysis tools, functions and parameters used for the host genome

Tool	Function	Key Parameters
FastQC	Quality control of sequence reads	-Min read quality = 30 -Min read length = 16
Ultraplex	Demultiplexing of reads	-No 5' or 3' mismatches
STAR Aligner (v2.7.10a)	Index hamster transcriptome and virus genome + align sequence reads	Argument: *--alignEndsType Extend5pOfRead1 --outSAMunmapped Within --outFilterMismatchNmax 999 -- outFilterMultimapNmax 1 -- outFilterMismatchNoverReadLmax 0.04*
Samtools (v1.15.1)	Extract reads	Argument: *Samtools view -b -f 4*
yanBedtools bamtobed	Conversion of .bam files to .bed files	Default arguments used
Samtools merge	Merging alignment files	Default arguments used
Samtools index	Generate alignment file indices (.bai)	-Supply a .txt file with flanked binding sites -Supply reference genome
UMI-tools extract	Deduplication	Argument: *umi_tools –umi-separator=":" -- extract-umi-method read_id –method unique*

Bedtools bamtobed	Conversion to a .bed genomic interval file	Default arguments used
Bedtools shift	Shifting of Genomic/transcriptomic interval entries upstream by 1 nt	Argument: *bedtools shift –m 5 –p -1*
Bedtools genome coverage	Splitting files into positive and negative strand and determining genome coverage	Arguments: *bedtools genomecov –bg –strand + -5* *bedtools genomecov –bg –strand - -5*
Bedtools merge	Merging of 'book-ended' intervals	Argument: *bedtools merge –c 4 –o sum*
Linux command *cat.*	Contatenating positive and negative strand .bedGraph files	N/A
R (v4.1.3)	Summing reads at transcript + gene level and manipulation of data for plotting	N/A
ggplot2 (v3.3.6)	Plotting graphs	N/A

Table 6.5. Identified binding sites from iCLIP experiment 1

Binding Site	Start (nt)	End (nt)
1	914	916
2	1218	1220
3	2376	2377
4	2428	2429
5	2653	2655
6	2749	2750
7	3462	3465
8	4850	4851
9	4870	4871
10	4926	4928
11	6131	6133
12	8402	8403
13	8447	8448

14	9029	9030
15	9740	9741
16	11357	11358
17	11508	11509

Table 6.6. Flanked binding sequences from iCLIP experiment for MEME analysis

Binding Site	Start (nt)	End (nt)	Sequence
1	904	936	GAUACAGUGGUUUCGUGUGAGGGCUACGUCGU
2	1208	1240	CCAUGAAAAAUUAUCUGCUUCCCGUGGUCGCC
3	2366	2397	UAUCUGCACGUACGGUUGACUCGCUGCUCUU
4	2418	2449	CGACGUGUUGUACGUAGACGAGGCGUUUGCG
5	2643	2675	GUCAUCGUUGCAUUACGAAGGCAAAAUGCGCA
6	2739	2770	AGACCUCGUGUUAACGUGCUUCAGAGGGUGG
7	3452	3485	GGAUAGAAGACUUUAACCCUACCACCAACAUCA
8	4840	4871	CCCCCCAAAACUGUCCCGUGCCUUUGCCGUU
9	4860	4891	CCUUUGCCGUUACGCUAUGACUCCAGAACGC
10	4916	4948	UCACAAGCAUAAUUGUGUGUUCUUCGUUUCCC
11	6121	6153	AUCAACGUCCGAUUGUCCAAUCCCGAGUCCGC
12	8392	8423	AAGGAGCCCGUACAGCCCUCUCGGUGGUGAC
13	8437	8468	UUGUCACUAAAAUCACCCCCGAGGGGGCCGA
14	9019	9050	GUAGGAAGAUUAGUCACUCAUGUACGCACCC
15	9730	9761	UAAUUCUGUAUUAUUAUGAGCUGUACCCCAC
16	11347	11378	UGCAGAAGAUCACGGGAGGUGUGGGACUGGU
17	11498	11529	UAAUCUAUAGAUCAAAGGGCUACGCAACCCC

Table 6.7. Buffers used for UVC irradiated nsP3 immunoprecipitation

Buffer	Composition
High stringency wash	20 mM Tris, pH 7.5; 120 mM NaCl; 25 mM KCl; 5 mM EDTA; 1% Trition-X100; 1% Na-deoxycholate
High salt wash	20 mM Tris, pH 7.5; 1 M NaCl; 5 mM EDTA; 1% Trition-X100; 1% Na-deoxycholate; 0.001% SDS
Low salt wash	20 mM Tris, pH 7.5; 5 mM EDTA
NT2 Buffer	50 mM Tris, pH 7.5; 150 mM NaCl; 1 mM MgCl2; 0.0005% Igepal

Table 6.8 List of genes from Figure 4.22 & Figure 4.23 WT samples x-axis

Sample	Gene
WT	1810024B03Rik
WT	Cacna1a
WT	Cyp1b1
WT	Dock10
WT	ENSMAUG00000005497
WT	ENSMAUG00000010455
WT	Fer1l5
WT	Gdpd4
WT	Klc4
WT	Mpeg1
WT	ND5
WT	Sfpq
WT	Ubxn2a
WT	Usp33
WT	Cfap53
WT	ENSMAUG00000010844
WT	Ipo7

WT	Lum
WT	Pgm2l1
WT	Elmod2
WT	Pex5
WT	Tirap
WT	Exog
WT	BC048671
WT	Tmed2
WT	Atp11b
WT	Car1
WT	Aen
WT	Ndufaf6
WT	ENSMAUG00000004820
WT	Stk40
WT	MT668625

Table 6.9 List of genes from Figure 4.22 & Figure 4.23 TST sample x-axis

Sample	Gene
TST	2610507B11Rik
TST	4932438A13Rik
TST	Abca1
TST	Abcf2
TST	Akt1
TST	Aldh2
TST	Alkbh1
TST	Anln
TST	Arhgdia
TST	Arl6

TST	Atg2a
TST	Atpif1
TST	Aup1
TST	Axl
TST	Bcar3
TST	Bmp1
TST	Ccdc114
TST	Cenpe
TST	Cep120
TST	Cep250
TST	Chrd
TST	Cntrob
TST	Col4a1
TST	Comp
TST	Crebzf
TST	Cyth4
TST	D330045A20Rik
TST	DES
TST	Dock6
TST	ENSMAUG00000001832
TST	ENSMAUG00000007819
TST	ENSMAUG00000008172
TST	ENSMAUG00000008753
TST	ENSMAUG00000010844
TST	ENSMAUG00000011654
TST	ENSMAUG00000013141
TST	ENSMAUG00000013725
TST	ENSMAUG00000015338

TST	ENSMAUG00000015673
TST	ENSMAUG00000015824
TST	ENSMAUG00000016510
TST	ENSMAUG00000017515
TST	ENSMAUG00000018862
TST	ENSMAUG00000020916
TST	ENSMAUG00000021702
TST	Exoc2
TST	Ext2
TST	Fhad1
TST	Fkbp15
TST	Gdf3
TST	Hgs
TST	Hip1
TST	Hmgb2
TST	Hook3
TST	Hspa8
TST	Ik
TST	Incenp
TST	Itga11
TST	Jag1
TST	Kmt2b
TST	Krt8
TST	Laptm4b
TST	Mki67
TST	Mkks
TST	Mroh8
TST	Myh14

TST	179	Myl12a
TST		Mypn
TST		Nav1
TST		Ncapg
TST		ND3
TST		ND6
TST		Nfe2l2
TST		Nucb2
TST		Olfr1428
TST		Pcsk6
TST		Pdcd4
TST		Plekhm2
TST		Ppp6r2
TST		Prkcsh
TST		Psap
TST		Psma7
TST		Rasd1
TST		Rbm12
TST		Rbm39
TST		Rgs2
TST		Rpp14
TST		Rps17
TST		Rrp12
TST		Safb2
TST		Sep9
TST		Sfpq
TST		Sh2b2
TST		Slc52a2

TST	Smarca5
TST	Smc6
TST	Spats2l
TST	Ssrp1
TST	Tada1
TST	Tgfa
TST	Tnnt2
TST	Tnpo1
TST	Tpt1
TST	Tspan9
TST	Ubr5
TST	Xkr8
TST	Ywhae
TST	Acvr2b
TST	App
TST	Bahcc1
TST	Cdc42bpb
TST	Dcaf6
TST	Dxo
TST	Ehbp1
TST	ENSMAUG00000001728
TST	ENSMAUG00000006512
TST	ENSMAUG00000008399
TST	ENSMAUG00000010107
TST	ENSMAUG00000012310
TST	ENSMAUG00000012338
TST	ENSMAUG00000013659
TST	ENSMAUG00000015824

TST	ENSMAUG00000018867
TST	Esyt1
TST	Fbxl22
TST	Gal3st1
TST	Golga4
TST	Hsp90b1
TST	Huwe1
TST	Igf2r
TST	Ipo7
TST	Jkamp
TST	Luc7l2
TST	Mmp2
TST	Msmp
TST	Nfasc
TST	Pex5
TST	Plekhh2
TST	Pnldc1
TST	Pnn
TST	Ppp1r21
TST	Prpf38b
TST	Rab35
TST	Ralbp1
TST	Slc5a4a
TST	Sprtn
TST	Tbc1d10b
TST	Tmem33
TST	Tmsb10
TST	Tut1

TST	Usp46
TST	Zdhhc23
TST	Clec1a
TST	Brd7
TST	Col5a1
TST	COX3
TST	Ddx5
TST	ENSMAUG00000017864
TST	ENSMAUG00000020680
TST	Gigyf2
TST	Inf2
TST	Lama5
TST	Megf8
TST	Mrps5
TST	Myh10
TST	Plk5
TST	Pnrc1
TST	Rbm28
TST	Rnf20
TST	Snrnp70
TST	Srsf11
TST	Supt5
TST	Ulk1
TST	Vars
TST	B3galnt2
TST	CYTB
TST	ENSMAUG00000022005
TST	Itpr3

TST	Kbtbd11
TST	Dhx36
TST	Eif3c
TST	ENSMAUG00000006396
TST	ENSMAUG00000019513
TST	Mastl
TST	Tgfbr3
TST	ND1
TST	ND2
TST	Peg3
TST	Ankrd1
TST	ENSMAUG00000004820
TST	ENSMAUG00000018214
TST	ENSMAUG00000018562
TST	COX2
TST	ENSMAUG00000018422
TST	ENSMAUG00000021213
TST	ND4
TST	Rragd
TST	Vash1
TST	ENSMAUG00000010500
TST	Klhl15
TST	Canx
TST	COX1
TST	Exoc6
TST	Jakmip3
TST	Flrt2
TST	Ubxn2a

TST	Eif3a
TST	ND5
TST	Tmem104
TST	ENSMAUG00000008863
TST	ENSMAUG00000006107
TST	ENSMAUG00000018522
TST	Hs6st3
TST	ENSMAUG00000017425
TST	Myh9
TST	ENSMAUG00000005497
TST	Ndufaf6
TST	MT668625

Motif	Symbol	Motif Consensus
1.		AARAUYA
2.		AGGGCUACG
3.		GUURCRYUAYGAMKSCARAA
4.		UAGAMGA
5.		UAAUUSUGURUU
6.		GASGUGUKGKAC
7.		GAGGG
8.		ACGUMCGRUUG
9.		UACAG
10.		AAGSA

Figure 6.1 Binding motif sequences from iCLIP experiment 1

Chapter 7 Bibliography

Abdelnabi, R., Neyts, J., Delang, L., 2015. Towards antivirals against chikungunya virus. Antiviral Res 121, 59–68. https://doi.org/10.1016/j.antiviral.2015.06.017

Abraham, R., Hauer, D., McPherson, R.L., Utt, A., Kirby, I.T., Cohen, M.S., Merits, A., Leung, A.K.L., Griffin, D.E., 2018. ADP-ribosyl–binding and hydrolase activities of the alphavirus nsP3 macrodomain are critical for initiation of virus replication. Proc Natl Acad Sci U S A 115, E10457–E10466. https://doi.org/10.1073/pnas.1812130115

Akahata, W., Yang, Z., Andersen, H., Sun, S., Holdaway, H.A., Kong, W.-P., Lewis, M.G., Higgs, S., Rossmann, M.G., Rao, S., Nabel, G.J., 2010. A VLP vaccine for epidemic Chikungunya virus protects non-human primates against infection. Nat Med 16, 334–338. https://doi.org/10.1038/nm.2105

Aliota, M.T., Walker, E.C., Uribe Yepes, A., Velez, I.D., Christensen, B.M., Osorio, J.E., 2016. The wMel Strain of Wolbachia Reduces Transmission of Chikungunya Virus in Aedes aegypti. PLoS Negl Trop Dis 10, e0004677. https://doi.org/10.1371/journal.pntd.0004677

Atkins, G.J., Sheahan, B.J., 2016. Molecular determinants of alphavirus neuropathogenesis in mice. J Gen Virol 97, 1283–1296. https://doi.org/10.1099/jgv.0.000467

Bailey, T.L., Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2, 28–36.

Basore, K., Kim, A.S., Nelson, C.A., Zhang, R., Smith, B.K., Uranga, C., Vang, L., Cheng, M., Gross, M.L., Smith, J., Diamond, M.S., Fremont, D.H., 2019. Cryo-EM Structure of Chikungunya Virus in Complex with the Mxra8 Receptor. Cell 177, 1725-1737.e16. https://doi.org/10.1016/j.cell.2019.04.006

Bernard, E., Solignat, M., Gay, B., Chazal, N., Higgs, S., Devaux, C., Briant, L., 2010. Endocytosis of Chikungunya Virus into Mammalian Cells: Role of Clathrin and Early Endosomal Compartments. PLOS ONE 5, e11479. https://doi.org/10.1371/journal.pone.0011479

Buchan, J.R., Parker, R., 2009. Eukaryotic Stress Granules: The Ins and Outs of Translation. Molecular Cell 36, 932–941. https://doi.org/10.1016/j.molcel.2009.11.020

Burt, F.J., Chen, W., Miner, J.J., Lenschow, D.J., Merits, A., Schnettler, E., Kohl, A., Rudd, P.A., Taylor, A., Herrero, L.J., Zaid, A., Ng, L.F.P., Mahalingam, S., 2017. Chikungunya virus: an update on the biology and pathogenesis of this emerging pathogen. Lancet Infect Dis 17, e107–e117. https://doi.org/10.1016/S1473-3099(16)30385-1

Burt, F.J., Rolph, M.S., Rulli, N.E., Mahalingam, S., Heise, M.T., 2012. Chikungunya: a re-emerging virus. The Lancet 379, 662–671. https://doi.org/10.1016/S0140-6736(11)60281-X

Busch, A., Brüggemann, M., Ebersberger, S., Zarnack, K., 2020. iCLIP data analysis: A complete pipeline from sequencing reads to RBP binding sites. Methods, High-Throughput Approaches in RNA Biology 178, 49–62. https://doi.org/10.1016/j.ymeth.2019.11.008

Button, J.M., Qazi, S.A., Wang, J.C.-Y., Mukhopadhyay, S., 2020. Revisiting an old friend: new findings in alphavirus structure and assembly. Current Opinion in Virology, 45. Virus structure & Expression (Dec 2020) 45, 25–33. https://doi.org/10.1016/j.coviro.2020.06.005

Caglioti, C., Lalle, E., Castilletti, C., Carletti, F., Capobianchi, M.R., Bordi, L., 2013. Chikungunya virus infection: an overview. New Microbiol 36, 211–227.

Campos, G.S., Albuquerque Bandeira, A.C., Diniz Rocha, V.F., Dias, J.P., Carvalho, R.H., Sardi, S.I., 2017. First Detection of Chikungunya Virus in Breast Milk. Pediatr Infect Dis J 36, 1015–1017. https://doi.org/10.1097/INF.0000000000001658

Cerny, T., Schwarz, M., Schwarz, U., Lemant, J., Gérardin, P., Keller, E., 2017. The Range of Neurological Complications in Chikungunya Fever. Neurocrit Care 27, 447–457. https://doi.org/10.1007/s12028-017-0413-8

Champer, J., Buchman, A., Akbari, O.S., 2016. Cheating evolution: engineering gene drives to manipulate the fate of wild populations. Nat Rev Genet 17, 146–159. https://doi.org/10.1038/nrg.2015.34

Chang, A.Y., Martins, K.A.O., Encinales, L., Reid, P., Acuña, M., Encinales, C., Matranga, C.B., Pacheco, N., Cure, C., Shukla, B., Arteta, T.R., Amdur, R., Cazares, L.H., Gregory, M., Ward, M.D., Porras, A., Mendoza, A.R., Dong, L., Kenny, T., Brueggemann, E., Downey, L.G., Kamalapathy, P., Lichtenberger, P., Falls, O., Simon, G.L., Bethony, J., Firestein, G.S., 2018. Chikungunya Arthritis Mechanisms in the Americas: A cross-sectional analysis of chikungunya arthritis patients 22-months post-infection demonstrate no detectable viral persistence in synovial fluid. Arthritis Rheumatol 70, 585–593. https://doi.org/10.1002/art.40383

Chang, L.-J., Dowd, K.A., Mendoza, F.H., Saunders, J.G., Sitar, S., Plummer, S.H., Yamshchikov, G., Sarwar, U.N., Hu, Z., Enama, M.E., Bailer, R.T., Koup, R.A., Schwartz, R.M., Akahata, W., Nabel, G.J., Mascola, J.R., Pierson, T.C., Graham, B.S., Ledgerwood, J.E., VRC 311 Study Team, 2014. Safety and tolerability of chikungunya virus-like particle vaccine in healthy adults: a phase 1 dose-escalation trial. Lancet 384, 2046–2052. https://doi.org/10.1016/S0140-6736(14)61185-5

Chen, G.L., Coates, E.E., Plummer, S.H., Carter, C.A., Berkowitz, N., Conan-Cibotti, M., Cox, J.H., Beck, A., O'Callahan, M., Andrews, C., Gordon, I.J., Larkin, B., Lampley, R., Kaltovich, F., Gall, J., Carlton, K., Mendy, J., Haney, D., May, J., Bray, A., Bailer, R.T., Dowd, K.A., Brockett, B., Gordon, D., Koup, R.A., Schwartz, R., Mascola, J.R., Graham, B.S., Pierson, T.C., Donastorg, Y., Rosario, N., Pape, J.W., Hoen, B., Cabié, A., Diaz, C., Ledgerwood, J.E., 2020. Effect of a Chikungunya Virus–Like Particle Vaccine on Safety and Tolerability Outcomes. JAMA 323, 1369–1377. https://doi.org/10.1001/jama.2020.2477

Chen, R., Wang, E., Tsetsarkin, K.A., Weaver, S.C., 2013. Chikungunya Virus 3′ Untranslated Region: Adaptation to Mosquitoes and a Population Bottleneck as Major Evolutionary Forces. PLOS Pathogens 9, e1003591. https://doi.org/10.1371/journal.ppat.1003591

Churkin, A., Weinbrand, L., Barash, D., 2015. Free Energy Minimization to Predict RNA Secondary Structures and Computational RNA Design,

in: Picardi, E. (Ed.), RNA Bioinformatics, Methods in Molecular Biology. Springer, New York, NY, pp. 3–16. https://doi.org/10.1007/978-1-4939-2291-8_1

Cirimotich, C.M., Scott, J.C., Phillips, A.T., Geiss, B.J., Olson, K.E., 2009. Suppression of RNA interference increases alphavirus replication and virus-associated mortality in Aedes aegypti mosquitoes. BMC Microbiology 9, 49. https://doi.org/10.1186/1471-2180-9-49

Coffey, L.L., Failloux, A.-B., Weaver, S.C., 2014. Chikungunya Virus–Vector Interactions. Viruses 6, 4628–4663. https://doi.org/10.3390/v6114628

Condreay, L.D., Brown, D.T., 1986. Exclusion of superinfecting homologous virus by Sindbis virus-infected Aedes albopictus (mosquito) cells. J Virol 58, 81–86.

Constant, L.E.C., Rajsfus, B.F., Carneiro, P.H., Sisnande, T., Mohana-Borges, R., Allonso, D., 2021. Overview on Chikungunya Virus Infection: From Epidemiology to State-of-the-Art Experimental Models. Frontiers in Microbiology 12.

Corley, M., Burns, M.C., Yeo, G.W., 2020. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. Molecular Cell 78, 9–29. https://doi.org/10.1016/j.molcel.2020.03.011

Cristea, I.M., Carroll, J.-W.N., Rout, M.P., Rice, C.M., Chait, B.T., MacDonald, M.R., 2006. Tracking and Elucidating Alphavirus-Host Protein Interactions *. Journal of Biological Chemistry 281, 30269–30278. https://doi.org/10.1074/jbc.M603980200

Cristea, I.M., Rozjabek, H., Molloy, K.R., Karki, S., White, L.L., Rice, C.M., Rout, M.P., Chait, B.T., MacDonald, M.R., 2010. Host Factors Associated with the Sindbis Virus RNA-Dependent RNA Polymerase: Role for G3BP1 and G3BP2 in Virus Replication. Journal of Virology 84, 6720–6732. https://doi.org/10.1128/JVI.01983-09

Danan, C., Manickavel, S., Hafner, M., 2016. PAR-CLIP: A Method for Transcriptome-Wide Identification of RNA Binding Protein Interaction Sites. Methods Mol Biol 1358, 153–173. https://doi.org/10.1007/978-1-4939-3067-8_10

Das, I., Basantray, I., Mamidi, P., Nayak, T.K., M, P.B., Chattopadhyay, Subhasis, Chattopadhyay, Soma, 2014. Heat Shock Protein 90 Positively Regulates Chikungunya Virus Replication by Stabilizing Viral Non-Structural Protein nsP2 during Infection. PLOS ONE 9, e100531. https://doi.org/10.1371/journal.pone.0100531

De Caluwé, L., Coppens, S., Vereecken, K., Daled, S., Dhaenens, M., Van Ostade, X., Deforce, D., Ariën, K.K., Bartholomeeusen, K., 2021. The CD147 Protein Complex Is Involved in Entry of Chikungunya Virus and Related Alphaviruses in Human Cells. Frontiers in Microbiology 12.

Della-Flora Nunes, G., Wilson, E.R., Marziali, L.N., Hurley, E., Silvestri, N., He, B., O'Malley, B.W., Beirowski, B., Poitelon, Y., Wrabetz, L., Feltri, M.L., 2021. Prohibitin 1 is essential to preserve mitochondria and myelin integrity in Schwann cells. Nat Commun 12, 3285. https://doi.org/10.1038/s41467-021-23552-8

Deo, R.C., Groft, C.M., Rajashankar, K.R., Burley, S.K., 2002. Recognition of the Rotavirus mRNA 3′ Consensus by an Asymmetric NSP3 Homodimer. Cell 108, 71–81. https://doi.org/10.1016/S0092-8674(01)00632-8

Deperasińska, I., Schulz, P., Siwicki, A.K., 2018. Salmonid Alphavirus (SAV). J Vet Res 62, 1–6. https://doi.org/10.2478/jvetres-2018-0001

Durchschlag, H., Fochler, C., Feser, B., Hausmann, S., Seroneit, T., Swientek, M., Swoboda, E., Winklmair, A., Wlček, C., Zipper, P., 1996. Effects of X- and UV-irradiation on proteins. Radiation Physics and Chemistry, Tihany Symposium on Radiation Chemistry 47, 501–505. https://doi.org/10.1016/0969-806X(95)00138-N

Eckei, L., Krieg, S., Bütepage, M., Lehmann, A., Gross, A., Lippok, B., Grimm, A.R., Kümmerer, B.M., Rossetti, G., Lüscher, B., Verheugd, P., 2017. The conserved macrodomains of the non-structural proteins of Chikungunya virus and other pathogenic positive strand RNA viruses function as mono-ADP-ribosylhydrolases. Sci Rep 7, 41746. https://doi.org/10.1038/srep41746

Edelman, R., Tacket, C.O., Wasserman, S.S., Bodison, S.A., Perry, J.G., Mangiafico, J.A., 2000. Phase II safety and immunogenicity study of live chikungunya virus vaccine TSI-GSD-218. The American Journal of Tropical Medicine and Hygiene 62, 681–685. https://doi.org/10.4269/ajtmh.2000.62.681

Farooq, S., Imran, M., Karim, E., Kanwar, D., Tariq, M.B., 2018. Chikungunya virus associated Guillain-Barre Syndrome with variable presentation: A case series. V O L . 13, 6.

Ferguson, J.M., Gamaarachchi, H., Nguyen, T., Gollon, A., Tong, S., Aquilina-Reid, C., Bowen-James, R., Deveson, I.W., 2022. InterARTIC: an interactive web application for whole-genome nanopore sequencing analysis of SARS-CoV-2 and other viruses. Bioinformatics 38, 1443–1446. https://doi.org/10.1093/bioinformatics/btab846

Filomatori, C.V., Bardossy, E.S., Merwaiss, F., Suzuki, Y., Henrion, A., Saleh, M.C., Alvarez, D.E., 2019. RNA recombination at Chikungunya virus 3'UTR as an evolutionary mechanism that provides adaptability. PLOS Pathogens 15, e1007706. https://doi.org/10.1371/journal.ppat.1007706

Fongsaran, C., Jirakanwisal, K., Kuadkitkan, A., Wikan, N., Wintachai, P., Thepparit, C., Ubol, S., Phaonakrop, N., Roytrakul, S., Smith, D.R., 2014. Involvement of ATP synthase β subunit in chikungunya virus entry into insect cells. Arch Virol 159, 3353–3364. https://doi.org/10.1007/s00705-014-2210-4

Forrester, N.L., Palacios, G., Tesh, R.B., Savji, N., Guzman, H., Sherman, M., Weaver, S.C., Lipkin, W.I., 2012. Genome-Scale Phylogeny of the Alphavirus Genus Suggests a Marine Origin. Journal of Virology.

Frolov, I., Agapov, E., Hoffman, T.A., Prágai, B.M., Lippa, M., Schlesinger, S., Rice, C.M., 1999. Selection of RNA Replicons Capable of Persistent Noncytopathic Replication in Mammalian Cells. J Virol 73, 3854–3865.

Frolova, E., Gorchakov, R., Garmashova, N., Atasheva, S., Vergara, L.A., Frolov, I., 2006. Formation of nsP3-Specific Protein Complexes during Sindbis Virus Replication. Journal of Virology 80, 4122–4134. https://doi.org/10.1128/JVI.80.8.4122-4134.2006

Frolova, E.I., Fayzulin, R.Z., Cook, S.H., Griffin, D.E., Rice, C.M., Frolov, I., 2002. Roles of Nonstructural Protein nsP2 and Alpha/Beta Interferons in Determining the Outcome of Sindbis Virus Infection. J Virol 76, 11254–11264. https://doi.org/10.1128/JVI.76.22.11254-11264.2002

Fros, J.J., Domeradzka, N.E., Baggen, J., Geertsema, C., Flipse, J., Vlak, J.M., Pijlman, G.P., 2012. Chikungunya Virus nsP3 Blocks Stress Granule Assembly by Recruitment of G3BP into Cytoplasmic Foci. J Virol 86, 10873–10879. https://doi.org/10.1128/JVI.01506-12

Fros, J.J., Geertsema, C., Zouache, K., Baggen, J., Domeradzka, N., van Leeuwen, D.M., Flipse, J., Vlak, J.M., Failloux, A.-B., Pijlman, G.P., 2015. Mosquito Rasputin interacts with chikungunya virus nsP3 and determines the infection rate in Aedes albopictus. Parasit Vectors 8, 464. https://doi.org/10.1186/s13071-015-1070-4

Froshauer, S., Kartenbeck, J., Helenius, A., 1988. Alphavirus RNA replicase is located on the cytoplasmic surface of endosomes and lysosomes. J Cell Biol 107, 2075–2086. https://doi.org/10.1083/jcb.107.6.2075

Galbraith, S.E., Sheahan, B.J., Atkins, G.J., 2006. Deletions in the hypervariable domain of the nsP3 gene attenuate Semliki Forest virus virulence. J Gen Virol 87, 937–947. https://doi.org/10.1099/vir.0.81406-0

Ganesan, K., Diwan, A., Shankar, S.K., Desai, S.B., Sainani, G.S., Katrak, S.M., 2008. Chikungunya Encephalomyeloradiculitis: Report of 2 Cases with Neuroimaging and 1 Case with Autopsy Findings. AJNR Am J Neuroradiol 29, 1636–1637. https://doi.org/10.3174/ajnr.A1133

Gao, Y., Goonawardane, N., Ward, J., Tuplin, A., Harris, M., 2019. Multiple roles of the non-structural protein 3 (nsP3) alphavirus unique domain (AUD) during Chikungunya virus genome replication and transcription. PLoS Pathog 15, e1007239. https://doi.org/10.1371/journal.ppat.1007239

Garmashova, N., Gorchakov, R., Volkova, E., Paessler, S., Frolova, E., Frolov, I., 2007. The Old World and New World alphaviruses use different virus-specific proteins for induction of transcriptional shutoff. J Virol 81, 2472–2484. https://doi.org/10.1128/JVI.02073-06

Ghildiyal, R., Gabrani, R., 2021. Deciphering the human cellular interactors of alphavirus unique domain of chikungunya virus. Virus Research 295, 198288. https://doi.org/10.1016/j.virusres.2020.198288

Ghosh, A., Desai, A., Ravi, V., Narayanappa, G., Tyagi, B.K., 2017. Chikungunya Virus Interacts with Heat Shock Cognate 70 Protein to Facilitate Its Entry into Mosquito Cell Line. INT 60, 247–262. https://doi.org/10.1159/000489308

Ghosh, A., Mullapudi, T., Bomanna, S., Tyagi, B.K., Ravi, V., Desai, A., 2019. Understanding the mechanism of Chikungunya virus vector competence in three species of mosquitoes. Medical and Veterinary Entomology 33, 375–387. https://doi.org/10.1111/mve.12376

Göertz, G.P., Lingemann, M., Geertsema, C., Abma-Henkens, M.H.C., Vogels, C.B.F., Koenraadt, C.J.M., Oers, M.M. van, Pijlman, G.P., 2018. Conserved motifs in the hypervariable domain of chikungunya virus nsP3 required for transmission by Aedes aegypti mosquitoes. PLOS Neglected Tropical Diseases 12, e0006958. https://doi.org/10.1371/journal.pntd.0006958

Gorchakov, R., Frolova, E., Frolov, I., 2005. Inhibition of Transcription and Translation in Sindbis Virus-Infected Cells. J Virol 79, 9397–9409. https://doi.org/10.1128/JVI.79.15.9397-9409.2005

Gorchakov, R., Garmashova, N., Frolova, E., Frolov, I., 2008. Different Types of nsP3-Containing Protein Complexes in Sindbis Virus-Infected Cells. Journal of Virology 82, 10088–10101. https://doi.org/10.1128/JVI.01011-08

Gorchakov, R., Wang, E., Leal, G., Forrester, N.L., Plante, K., Rossi, S.L., Partidos, C.D., Adams, A.P., Seymour, R.L., Weger, J., Borland, E.M., Sherman, M.B., Powers, A.M., Osorio, J.E., Weaver, S.C., 2012. Attenuation of Chikungunya virus vaccine strain 181/clone 25 is determined by two amino acid substitutions in the E2 envelope glycoprotein. J Virol 86, 6084–6096. https://doi.org/10.1128/JVI.06449-11

Götte, B., Liu, L., McInerney, G.M., 2018. The Enigmatic Alphavirus Non-Structural Protein 3 (nsP3) Revealing Its Secrets at Last. Viruses 10, 105. https://doi.org/10.3390/v10030105

Götte, B., Panas, M.D., Hellström, K., Liu, L., Samreen, B., Larsson, O., Ahola, T., McInerney, G.M., 2019. Separate domains of G3BP promote efficient clustering of alphavirus replication complexes and recruitment of the translation initiation machinery. PLOS Pathogens 15, e1007842. https://doi.org/10.1371/journal.ppat.1007842

Grivard, P., Le Roux, K., Laurent, P., Fianu, A., Perrau, J., Gigan, J., Hoarau, G., Grondin, N., Staikowsky, F., Favier, F., Michault, A., 2007. Molecular and serological diagnosis of Chikungunya virus infection. Pathologie Biologie, Numéro spécial Bactériologie - Virologie - Mycologie - Parasitologie - Maladies infectieuses - 2e partie 55, 490–494. https://doi.org/10.1016/j.patbio.2007.07.002

Gurer, C., Cimarelli, A., Luban, J., 2002. Specific Incorporation of Heat Shock Protein 70 Family Members into Primate Lentiviral Virions. J Virol 76, 4666–4670. https://doi.org/10.1128/JVI.76.9.4666-4670.2002

Haecker, I., Renne, R., 2014. HITS-CLIP and PAR-CLIP advance viral miRNA targetome analysis. Crit Rev Eukaryot Gene Expr 24, 101–116.

Hallengärd, D., Kakoulidou, M., Lulla, A., Kümmerer, B.M., Johansson, D.X., Mutso, M., Lulla, V., Fazakerley, J.K., Roques, P., Le Grand, R., Merits, A., Liljeström, P., 2014. Novel Attenuated Chikungunya Vaccine Candidates Elicit Protective Immunity in C57BL/6 mice. J Virol 88, 2858–2866. https://doi.org/10.1128/JVI.03453-13

Hameed, S., Khan, S., 2019. Rare variant of Guillain-Barré syndrome after chikungunya viral fever. BMJ Case Reports CP 12, e228845. https://doi.org/10.1136/bcr-2018-228845

Hanley, K.A., 2011. The Double-Edged Sword: How Evolution Can Make or Break a Live-Attenuated Virus Vaccine. Evo Edu Outreach 4, 635–643. https://doi.org/10.1007/s12052-011-0365-y

Hawman, D.W., Stoermer, K.A., Montgomery, S.A., Pal, P., Oko, L., Diamond, M.S., Morrison, T.E., 2013. Chronic joint disease caused by persistent Chikungunya virus infection is controlled by the adaptive immune response. J Virol 87, 13878–13888. https://doi.org/10.1128/JVI.02666-13

Hermanns, K., Marklewitz, M., Zirkel, F., Overheul, G.J., Page, R.A., Loaiza, J.R., Drosten, C., van Rij, R.P., Junglen, S., 2020. Agua Salud alphavirus defines a novel lineage of insect-specific alphaviruses

discovered in the New World. J Gen Virol 101, 96–104.
https://doi.org/10.1099/jgv.0.001344

Holmes, A.C., Basore, K., Fremont, D.H., Diamond, M.S., 2020. A molecular understanding of alphavirus entry. PLOS Pathogens 16, e1008876. https://doi.org/10.1371/journal.ppat.1008876

Hoornweg, T.E., van Duijl-Richter, M.K.S., Ayala Nuñez, N.V., Albulescu, I.C., van Hemert, M.J., Smit, J.M., 2016. Dynamics of Chikungunya Virus Cell Entry Unraveled by Single-Virus Tracking in Living Cells. J Virol 90, 4745–4756. https://doi.org/10.1128/JVI.03184-15

Hucke, F.I.L., Bugert, J.J., 2020. Current and Promising Antivirals Against Chikungunya Virus. Front Public Health 8, 618624. https://doi.org/10.3389/fpubh.2020.618624

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., König, J., Ule, J., 2014. iCLIP: Protein–RNA interactions at nucleotide resolution. Methods 65, 274–287. https://doi.org/10.1016/j.ymeth.2013.10.011

Igarashi, A., 1978. Isolation of a Singh's Aedes albopictus cell clone sensitive to Dengue and Chikungunya viruses. J Gen Virol 40, 531–544. https://doi.org/10.1099/0022-1317-40-3-531

Jayabalan, A.K., Adivarahan, S., Koppula, A., Abraham, R., Batish, M., Zenklusen, D., Griffin, D.E., Leung, A.K.L., 2021. Stress granule formation, disassembly, and composition are regulated by alphavirus ADP-ribosylhydrolase activity. Proceedings of the National Academy of Sciences 118, e2021719118. https://doi.org/10.1073/pnas.2021719118

Jones, J.E., Long, K.M., Whitmore, A.C., Sanders, W., Thurlow, L.R., Brown, J.A., Morrison, C.R., Vincent, H., Peck, K.M., Browning, C., Moorman, N., Lim, J.K., Heise, M.T., 2017. Disruption of the Opal Stop Codon Attenuates Chikungunya Virus-Induced Arthritis and Pathology. mBio 8, e01456-17. https://doi.org/10.1128/mBio.01456-17

Jones, R., Bragagnolo, G., Arranz, R., Reguera, J., 2021. Capping pores of alphavirus nsP1 gate membranous viral replication factories. Nature 589, 615–619. https://doi.org/10.1038/s41586-020-3036-8

Jose, J., Taylor, A.B., Kuhn, R.J., 2017. Spatial and Temporal Analysis of Alphavirus Replication and Assembly in Mammalian and Mosquito Cells. mBio 8, e02294-16. https://doi.org/10.1128/mBio.02294-16

Kay, B.K., Williamson, M.P., Sudol, M., 2000. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. The FASEB Journal 14, 231–241. https://doi.org/10.1096/fasebj.14.2.231

Kendall, C., Khalid, H., Müller, M., Banda, D.H., Kohl, A., Merits, A., Stonehouse, N.J., Tuplin, A., 2019. Structural and phenotypic analysis of Chikungunya virus RNA replication elements. Nucleic Acids Research 47, 9296–9312. https://doi.org/10.1093/nar/gkz640

Kielian, M., Chanel-Vos, C., Liao, M., 2010. Alphavirus Entry and Membrane Fusion. Viruses 2, 796–825. https://doi.org/10.3390/v2040796

König, J., Zarnack, K., Luscombe, N.M., Ule, J., 2012. Protein–RNA interactions: new genomic technologies and perspectives. Nat Rev Genet 13, 77–83. https://doi.org/10.1038/nrg3141

Krakau, S., Richard, H., Marsico, A., 2017. PureCLIP: capturing target-specific protein–RNA interaction footprints from single-nucleotide CLIP-seq data. Genome Biology 18, 240. https://doi.org/10.1186/s13059-017-1364-2

Krieg, S., Pott, F., Eckei, L., Verheirstraeten, M., Bütepage, M., Lippok, B., Goffinet, C., Lüscher, B., Verheugd, P., 2020. Mono-ADP-ribosylation by ARTD10 restricts Chikungunya virus replication by interfering with the proteolytic activity of nsP2. https://doi.org/10.1101/2020.01.07.896977

Kujala, P., Ikäheimonen, A., Ehsani, N., Vihinen, H., Auvinen, P., Kääriäinen, L., 2001. Biogenesis of the Semliki Forest virus RNA replication complex. J Virol 75, 3873–3884. https://doi.org/10.1128/JVI.75.8.3873-3884.2001

Kumar, R., Shrivastava, T., Samal, S., Ahmed, S., Parray, H.A., 2020. Antibody-based therapeutic interventions: possible strategy to counter chikungunya viral infection. Appl Microbiol Biotechnol 104, 3209–3228. https://doi.org/10.1007/s00253-020-10437-x

Kumar, R., Srivastava, P., Mathur, K., Shrinet, J., Dubey, S.K., Chinnappan, M., Kaur, I., Nayak, D., Chattopadhyay, S., Bhatnagar, R.K., Sunil, S., 2021. Chikungunya virus non-structural protein nsP3 interacts with Aedes aegypti DEAD-box helicase RM62F. VirusDis. 32, 657–665. https://doi.org/10.1007/s13337-021-00734-y

Kutchko, K.M., Madden, E.A., Morrison, C., Plante, K.S., Sanders, W., Vincent, H.A., Cruz Cisneros, M.C., Long, K.M., Moorman, N.J., Heise, M.T., Laederach, A., 2018. Structural divergence creates new functional features in alphavirus genomes. Nucleic Acids Res 46, 3657–3670. https://doi.org/10.1093/nar/gky012

Lalitha, P., Rathinam, S., Banushree, K., Maheshkumar, S., Vijayakumar, R., Sathe, P., 2007. Ocular involvement associated with an epidemic outbreak of chikungunya virus infection. Am J Ophthalmol 144, 552–556. https://doi.org/10.1016/j.ajo.2007.06.002

Lanciotti, R.S., Valadere, A.M., 2014. Transcontinental Movement of Asian Genotype Chikungunya Virus. Emerg Infect Dis 20, 1400–1402. https://doi.org/10.3201/eid2008.140268

Langsjoen, R.M., Haller, S.L., Roy, C.J., Vinet-Oliphant, H., Bergren, N.A., Erasmus, J.H., Livengood, J.A., Powell, T.D., Weaver, S.C., Rossi, S.L., 2018. Chikungunya Virus Strains Show Lineage-Specific Variations in Virulence and Cross-Protective Ability in Murine and Nonhuman Primate Models. mBio 9, e02449-17. https://doi.org/10.1128/mBio.02449-17

Lastarza, M.W., Grakoui, A., Rice, C.M., 1994. Deletion and duplication mutations in the C-terminal nonconserved region of Sindbis virus nsP3: effects on phosphorylation and on virus replication in vertebrate and invertebrate cells. Virology 202, 224–232. https://doi.org/10.1006/viro.1994.1338

Lee, F.C.Y., Chakrabarti, A.M., Hänel, H., Monzón-Casanova, E., Hallegger, M., Militti, C., Capraro, F., Sadée, C., Toolan-Kerr, P., Wilkins, O., Turner, M., König, J., Sibley, C.R., Ule, J., 2021. An improved iCLIP protocol. https://doi.org/10.1101/2021.08.27.457890

Lee, F.C.Y., Ule, J., 2018. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. Molecular Cell 69, 354–369. https://doi.org/10.1016/j.molcel.2018.01.005

Lee, N., Le Sage, V., Nanni, A.V., Snyder, D.J., Cooper, V.S., Lakdawala, S.S., 2017. Genome-wide analysis of influenza viral RNA and nucleoprotein association. Nucleic Acids Research 45, 8968–8977. https://doi.org/10.1093/nar/gkx584

Lee, R.C.H., Hapuarachchi, H.C., Chen, K.C., Hussain, K.M., Chen, H., Low, S.L., Ng, L.C., Lin, R., Ng, M.M.-L., Chu, J.J.H., 2013. Mosquito Cellular Factors and Functions in Mediating the Infectious entry of Chikungunya Virus. PLOS Neglected Tropical Diseases 7, e2050. https://doi.org/10.1371/journal.pntd.0002050

Li, M., Bui, M., Yang, T., White, B.J., Akbari, O.S., 2017. Germline Cas9 Expression Yields Highly Efficient Genome Engineering in a Major Worldwide Disease Vector, Aedes aegypti. https://doi.org/10.1101/156778

Liang, Z., Li, G., 2012. Recombinant Sindbis virus expressing functional GFP in the nonstructural protein nsP3, in: In Marketing Science Institute Conference on Big Data.

Linder, P., Jankowsky, E., 2011. From unwinding to clamping — the DEAD box RNA helicase family. Nat Rev Mol Cell Biol 12, 505–516. https://doi.org/10.1038/nrm3154

Lindh, E., Argentini, C., Remoli, M.E., Fortuna, C., Faggioni, G., Benedetti, E., Amendola, A., Marsili, G., Lista, F., Rezza, G., Venturi, G., 2018. The Italian 2017 Outbreak Chikungunya Virus Belongs to an Emerging Aedes albopictus–Adapted Virus Cluster Introduced From the Indian Subcontinent. Open Forum Infect Dis 6, ofy321. https://doi.org/10.1093/ofid/ofy321

Luers, A.J., Adams, S.D., Smalley, J.V., Campanella, J.J., 2005. A phylogenomic study of the genus Alphavirus employing whole genome comparison. Comparative and Functional Genomics 6, 217–227. https://doi.org/10.1002/cfg.478

Lulla, A., Lulla, V., Merits, A., 2012. Macromolecular assembly-driven processing of the 2/3 cleavage site in the alphavirus replicase polyprotein. J Virol 86, 553–565. https://doi.org/10.1128/JVI.05195-11

Macpherson, I., Stoker, M., 1962. Polyoma transformation of hamster cell clones--an investigation of genetic factors affecting cell competence. Virology 16, 147–151. https://doi.org/10.1016/0042-6822(62)90290-8

Madden, E.A., Plante, K.S., Morrison, C.R., Kutchko, K.M., Sanders, W., Long, K.M., Taft-Benz, S., Cruz Cisneros, M.C., White, A.M., Sarkar, S., Reynolds, G., Vincent, H.A., Laederach, A., Moorman, N.J., Heise, M.T., 2020. Using SHAPE-MaP To Model RNA Secondary Structure and Identify 3'UTR Variation in Chikungunya Virus. J Virol 94, e00701-20. https://doi.org/10.1128/JVI.00701-20

Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvonen, M., Ahola, T., Forrester, N., Gould, E.A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A.E., de Lamballerie, X., Canard, B., 2009. The Crystal Structures of Chikungunya and Venezuelan Equine Encephalitis Virus nsP3 Macro Domains Define a Conserved Adenosine Binding Pocket. J Virol 83, 6534–6545. https://doi.org/10.1128/JVI.00189-09

Martin, G., Zavolan, M., 2016. Redesigning CLIP for efficiency, accuracy and speed. Nat Methods 13, 482–483. https://doi.org/10.1038/nmeth.3870

Mathur, K., Anand, A., Dubey, S.K., Sanan-Mishra, N., Bhatnagar, R.K., Sunil, S., 2016. Analysis of chikungunya virus proteins reveals that non-structural proteins nsP2 and nsP3 exhibit RNA interference (RNAi) suppressor activity. Sci Rep 6, 38065. https://doi.org/10.1038/srep38065

Matkovic, R., Bernard, E., Fontanel, S., Eldin, P., Chazal, N., Hassan Hersi, D., Merits, A., Péloponèse, J.-M., Briant, L., 2019. The Host DHX9 DExH-Box Helicase Is Recruited to Chikungunya Virus Replication Complexes for Optimal Genomic RNA Translation. J Virol 93, e01764-18. https://doi.org/10.1128/JVI.01764-18

Maximova, O.A., Sturdevant, D.E., Kash, J.C., Kanakabandi, K., Xiao, Y., Minai, M., Moore, I.N., Taubenberger, J., Martens, C., Cohen, J.I., Pletnev, A.G., 2021. Virus infection of the CNS disrupts the immune-neural-synaptic axis via induction of pleiotropic gene regulation of host responses. Elife 10, e62273. https://doi.org/10.7554/eLife.62273

Mehta, R., Gerardin, P., de Brito, C.A.A., Soares, C.N., Ferreira, M.L.B., Solomon, T., 2018. The neurological complications of chikungunya virus: A systematic review. Rev Med Virol 28, e1978. https://doi.org/10.1002/rmv.1978

Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., Shearwood, A.-M.J., Haugen, E., Bracken, C.P., Rackham, O., Stamatoyannopoulos, J.A., Filipovska, A., Mattick, J.S., 2011. The human mitochondrial transcriptome. Cell 146, 645–658. https://doi.org/10.1016/j.cell.2011.06.051

Merwaiss, F., Filomatori, C.V., Susuki, Y., Bardossy, E.S., Alvarez, D.E., Saleh, M.-C., 2021. Chikungunya Virus Replication Rate Determines the Capacity of Crossing Tissue Barriers in Mosquitoes. J Virol 95, e01956-20. https://doi.org/10.1128/JVI.01956-20

Michel, G., Petrakova, O., Atasheva, S., Frolov, I., 2007. Adaptation of Venezuelan equine encephalitis virus lacking 51-nt conserved sequence element to replication in mammalian and mosquito cells. Virology 362, 475–487. https://doi.org/10.1016/j.virol.2007.01.009

Mittal, A., Mittal, S., Bharati, M.J., Ramakrishnan, R., Saravanan, S., Sathe, P.S., 2007. Optic neuritis associated with chikungunya virus infection in South India. Arch Ophthalmol 125, 1381–1386. https://doi.org/10.1001/archopht.125.10.1381

Moore, M., Zhang, C., Gantman, E.C., Mele, A., Darnell, J.C., Darnell, R.B., 2014. Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nat Protoc 9, 263–293. https://doi.org/10.1038/nprot.2014.012

Morazzani, E.M., Wiley, M.R., Murreddu, M.G., Adelman, Z.N., Myles, K.M., 2012. Production of Virus-Derived Ping-Pong-Dependent piRNA-like Small RNAs in the Mosquito Soma. PLOS Pathogens 8, e1002470. https://doi.org/10.1371/journal.ppat.1002470

Murthy, K.R., Venkataraman, N., Satish, V., Babu, K., 2008. Bilateral retinitis following Chikun- gunya fever. Indian J Ophthalmol 56, 329–331.

Mutso, M., Morro, A.M., Smedberg, C., Kasvandik, S., Aquilimeba, M., Teppor, M., Tarve, L., Lulla, A., Lulla, V., Saul, S., Thaa, B., McInerney, G.M., Merits, A., Varjak, M., 2018. Mutation of CD2AP and SH3KBP1 Binding Motif in Alphavirus nsP3 Hypervariable Domain Results in Attenuated Virus. Viruses 10, 226. https://doi.org/10.3390/v10050226

Nakabayashi, H., Taketa, K., Miyano, K., Yamane, T., Sato, J., 1982. Growth of human hepatoma cells lines with differentiated functions in chemically defined medium. Cancer Res 42, 3858–3863.

Neuvonen, M., Kazlauskas, A., Martikainen, M., Hinkkanen, A., Ahola, T., Saksela, K., 2011. SH3 Domain-Mediated Recruitment of Host Cell Amphiphysins by Alphavirus nsP3 Promotes Viral RNA Replication. PLoS Pathog 7, e1002383. https://doi.org/10.1371/journal.ppat.1002383

Ooi, Y.S., Stiles, K.M., Liu, C.Y., Taylor, G.M., Kielian, M., 2013. Genome-Wide RNAi Screen Identifies Novel Host Proteins Required for Alphavirus Entry. PLOS Pathogens 9, e1003835. https://doi.org/10.1371/journal.ppat.1003835

Orenstein, Y., Ohler, U., Berger, B., 2018. Finding RNA structure in the unstructured RBPome. BMC Genomics 19, 154. https://doi.org/10.1186/s12864-018-4540-1

Panas, M.D., Ahola, T., McInerney, G.M., 2014. The C-Terminal Repeat Domains of nsP3 from the Old World Alphaviruses Bind Directly to G3BP. J Virol 88, 5888–5893. https://doi.org/10.1128/JVI.00439-14

Panas, M.D., Schulte, T., Thaa, B., Sandalova, T., Kedersha, N., Achour, A., McInerney, G.M., 2015. Viral and Cellular Proteins Containing FGDF Motifs Bind G3BP to Block Stress Granule Formation. PLoS Pathog 11, e1004659. https://doi.org/10.1371/journal.ppat.1004659

Panas, M.D., Varjak, M., Lulla, A., Er Eng, K., Merits, A., Karlsson Hedestam, G.B., McInerney, G.M., 2012. Sequestration of G3BP coupled with efficient translation inhibits stress granules in Semliki Forest virus infection. Mol Biol Cell 23, 4701–4712. https://doi.org/10.1091/mbc.E12-08-0619

Parashar, D., Cherian, S., 2014. Antiviral Perspectives for Chikungunya Virus. Biomed Res Int 2014, 631642. https://doi.org/10.1155/2014/631642

Parola, P., de Lamballerie, X., Jourdan, J., Rovery, C., Vaillant, V., Minodier, P., Brouqui, P., Flahault, A., Raoult, D., Charrel, R.N., 2006. Novel Chikungunya Virus Variant in Travelers Returning from Indian Ocean Islands. Emerg Infect Dis 12, 1493–1499. https://doi.org/10.3201/eid1210.060610

Pialoux, G., Gaüzère, B.-A., Jauréguiberry, S., Strobel, M., 2007. Chikungunya, an epidemic arbovirosis. The Lancet Infectious Diseases 7, 319–327. https://doi.org/10.1016/S1473-3099(07)70107-X

Poteryaev, D., Datta, S., Ackema, K., Zerial, M., Spang, A., 2010. Identification of the Switch in Early-to-Late Endosome Transition. Cell 141, 497–508. https://doi.org/10.1016/j.cell.2010.03.011

Pritchard, J., 2010. Guillain-Barré syndrome. Clin Med (Lond) 10, 399–401. https://doi.org/10.7861/clinmedicine.10-4-399

Pujhari, S., Brustolin, M., Macias, V.M., Nissly, R.H., Nomura, M., Kuchipudi, S.V., Rasgon, J.L., 2019. Heat shock protein 70 (Hsp70) mediates Zika virus entry, replication, and egress from host cells. Emerg Microbes Infect 8, 8–16. https://doi.org/10.1080/22221751.2018.1557988

Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., Burton, D.R., Lewis-Ximenez, L.L., de Jesus, J.G., Giovanetti, M., Hill, S.C., Black, A., Bedford, T., Carroll, M.W., Nunes, M., Alcantara, L.C., Sabino, E.C., Baylis, S.A., Faria, N.R., Loose, M., Simpson, J.T., Pybus, O.G., Andersen, K.G., Loman, N.J., 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc 12, 1261–1276. https://doi.org/10.1038/nprot.2017.066

Rack, J.G.M., Perina, D., Ahel, I., 2016. Macrodomains: Structure, Function, Evolution, and Catalytic Activities. Annu Rev Biochem 85, 431–454. https://doi.org/10.1146/annurev-biochem-060815-014935

Ramful, D., Carbonnier, M., Pasquet, M., Bouhmani, B., Ghazouani, J., Noormahomed, T., Beullier, G., Attali, T., Samperiz, S., Fourmaintraux, A., Alessandri, J.-L., 2007. Mother-to-Child Transmission of Chikungunya Virus Infection. The Pediatric Infectious Disease Journal 26, 811–815. https://doi.org/10.1097/INF.0b013e3180616d4f

Ramsey, J., Mukhopadhyay, S., 2017. Disentangling the Frames, the State of Research on the Alphavirus 6K and TF Proteins. Viruses 9, 228. https://doi.org/10.3390/v9080228

Rathore, A.P.S., Haystead, T., Das, P.K., Merits, A., Ng, M.-L., Vasudevan, S.G., 2014. Chikungunya virus nsP3 & nsP4 interacts with HSP-90 to promote virus replication: HSP-90 inhibitors reduce CHIKV infection and inflammation in vivo. Antiviral Research 103, 7–16. https://doi.org/10.1016/j.antiviral.2013.12.010

Reid, W.R., Lin, J., Williams, A.E., Juncu, R., Olson, K.E., Franz, A.W., 2021. Genomic insertion locus and Cas9 expression in the germline affect CRISPR/Cas9-based gene drive performance in the yellow fever mosquito Aedes aegypti. https://doi.org/10.1101/2021.12.08.471839

Remenyi, R., Roberts, G.C., Zothner, C., Merits, A., Harris, M., 2017. SNAP-tagged Chikungunya Virus Replicons Improve Visualisation of Non-Structural Protein 3 by Fluorescence Microscopy. Sci Rep 7, 5682. https://doi.org/10.1038/s41598-017-05820-0

Ritchie, S.A., Johnson, B.J., 2017. Advances in Vector Control Science: Rear-and-Release Strategies Show Promise… but Don't Forget the Basics. J Infect Dis 215, S103–S108. https://doi.org/10.1093/infdis/jiw575

Roberts, G.C., Zothner, C., Remenyi, R., Merits, A., Stonehouse, N.J., Harris, M., 2017. Evaluation of a range of mammalian and mosquito cell lines for use in Chikungunya virus research. Sci Rep 7, 14641. https://doi.org/10.1038/s41598-017-15269-w

Robinson, M.C., 1955. 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, 28–32. https://doi.org/10.1016/0035-9203(55)90080-8

Rupp, J.C., Sokoloski, K.J., Gebhart, N.N., Hardy, R.W., 2015. Alphavirus RNA synthesis and non-structural protein functions. J Gen Virol 96, 2483–2500. https://doi.org/10.1099/jgv.0.000249

Saul, S., Ferguson, M., Cordonin, C., Fragkoudis, R., Ool, M., Tamberg, N., Sherwood, K., Fazakerley, J.K., Merits, A., 2015. Differences in Processing Determinants of Nonstructural Polyprotein and in the Sequence of Nonstructural Protein 3 Affect Neurovirulence of Semliki Forest Virus. J Virol 89, 11030–11045. https://doi.org/10.1128/JVI.01186-15

Saxton-Shaw, K.D., Ledermann, J.P., Borland, E.M., Stovall, J.L., Mossel, E.C., Singh, A.J., Wilusz, J., Powers, A.M., 2013. O'nyong nyong Virus Molecular Determinants of Unique Vector Specificity Reside in Non-Structural Protein 3. PLoS Negl Trop Dis 7, e1931. https://doi.org/10.1371/journal.pntd.0001931

Schmidt, H., Collier, T.C., Hanemaaijer, M.J., Houston, P.D., Lee, Y., Lanzaro, G.C., 2020. Abundance of conserved CRISPR-Cas9 target sites within the highly polymorphic genomes of Anopheles and Aedes mosquitoes. Nat Commun 11, 1425. https://doi.org/10.1038/s41467-020-15204-0

Schmidt, N., Lareau, C.A., Keshishian, H., Ganskih, S., Schneider, C., Hennig, T., Melanson, R., Werner, S., Wei, Y., Zimmer, M., Ade, J., Kirschner, L., Zielinski, S., Dölken, L., Lander, E.S., Caliskan, N., Fischer, U., Vogel, J., Carr, S.A., Bodem, J., Munschauer, M., 2021. The SARS-CoV-2 RNA–protein interactome in infected human cells. Nat Microbiol 6, 339–353. https://doi.org/10.1038/s41564-020-00846-z

Scholte, F.E.M., Tas, A., Albulescu, I.C., Žusinaite, E., Merits, A., Snijder, E.J., van Hemert, M.J., 2015. Stress granule components G3BP1 and G3BP2 play a proviral role early in Chikungunya virus replication. J Virol 89, 4457–4469. https://doi.org/10.1128/JVI.03612-14

Schuffenecker, I., Iteman, I., Michault, A., Murri, S., Frangeul, L., Vaney, M.-C., Lavenir, R., Pardigon, N., Reynes, J.-M., Pettinelli, F., Biscornet, L., Diancourt, L., Michel, S., Duquerroy, S., Guigon, G., Frenkiel, M.-P., Bréhin, A.-C., Cubito, N., Desprès, P., Kunst, F., Rey, F.A., Zeller, H., Brisse, S., 2006. Genome Microevolution of Chikungunya Viruses Causing the Indian Ocean Outbreak. PLOS Medicine 3, e263. https://doi.org/10.1371/journal.pmed.0030263

Schulte, T., Liu, L., Panas, M.D., Thaa, B., Dickson, N., Götte, B., Achour, A., McInerney, G.M., 2016. Combined structural, biochemical and cellular evidence demonstrates that both FGDF motifs in alphavirus nsP3 are required for efficient replication. Open Biol 6, 160078. https://doi.org/10.1098/rsob.160078

Schwartz, O., Albert, M.L., 2010. Biology and pathogenesis of chikungunya virus. Nat Rev Microbiol 8, 491–500. https://doi.org/10.1038/nrmicro2368

Seneviratne, U., 2000. Guillain-Barré syndrome. Postgraduate Medical Journal 76, 774–782. https://doi.org/10.1136/pgmj.76.902.774

Shin, G., Yost, S.A., Miller, M.T., Elrod, E.J., Grakoui, A., Marcotrigiano, J., 2012. Structural and functional insights into alphavirus polyprotein processing and pathogenesis. Proc Natl Acad Sci U S A 109, 16534–16539. https://doi.org/10.1073/pnas.1210418109

Siang, D.T.C., Lim, Y.C., Kyaw, A.M.M., Win, K.N., Chia, S.Y., Degirmenci, U., Hu, X., Tan, B.C., Walet, A.C.E., Sun, L., Xu, D., 2020. The RNA-binding protein HuR is a negative regulator in adipogenesis. Nat Commun 11, 213. https://doi.org/10.1038/s41467-019-14001-8

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7, 539. https://doi.org/10.1038/msb.2011.75

Silva, L.A., Dermody, T.S., 2017. Chikungunya virus: epidemiology, replication, disease mechanisms, and prospective intervention strategies. J Clin Invest 127, 737–749. https://doi.org/10.1172/JCI84417

Singh, A., Kumar, A., N. Uversky, V., Giri, R., 2018. Understanding the interactability of chikungunya virus proteins via molecular recognition feature analysis. RSC Advances 8, 27293–27303. https://doi.org/10.1039/C8RA04760J

Snyder, J.E., Kulcsar, K.A., Schultz, K.L.W., Riley, C.P., Neary, J.T., Marr, S., Jose, J., Griffin, D.E., Kuhn, R.J., 2013. Functional Characterization of the Alphavirus TF Protein. J Virol 87, 8511–8523. https://doi.org/10.1128/JVI.00449-13

Solignat, M., Gay, B., Higgs, S., Briant, L., Devaux, C., 2009. Replication cycle of chikungunya: A re-emerging arbovirus. Virology 393, 183–197. https://doi.org/10.1016/j.virol.2009.07.024

Song, H., Zhao, Z., Chai, Y., Jin, X., Li, C., Yuan, F., Liu, S., Gao, Z., Wang, H., Song, J., Vazquez, L., Zhang, Y., Tan, S., Morel, C.M., Yan, J., Shi, Y., Qi, J., Gao, F., Gao, G.F., 2019. Molecular Basis of Arthritogenic Alphavirus Receptor MXRA8 Binding to Chikungunya Virus Envelope Protein. Cell 177, 1714-1724.e12. https://doi.org/10.1016/j.cell.2019.04.008

Soto-Garita, C., Carrera, J.-P., López-Vergès, S., Corrales-Aguilar, E., 2018. Advances in Clinical Diagnosis and Management of Chikungunya Virus Infection. Curr Treat Options Infect Dis 10, 397–409. https://doi.org/10.1007/s40506-018-0172-x

Strauss, J.H., Strauss, E.G., 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol Rev 58, 491–562. https://doi.org/10.1128/mr.58.3.491-562.1994

Suhrbier, A., Jaffar-Bandjee, M.-C., Gasque, P., 2012. Arthritogenic alphaviruses—an overview. Nat Rev Rheumatol 8, 420–429. https://doi.org/10.1038/nrrheum.2012.64

Suthar, M.S., Shabman, R., Madric, K., Lambeth, C., Heise, M.T., 2005. Identification of Adult Mouse Neurovirulence Determinants of the Sindbis Virus Strain AR86. J Virol 79, 4219–4228. https://doi.org/10.1128/JVI.79.7.4219-4228.2005

Tang, B.L., 2012. The cell biology of Chikungunya virus infection. Cellular Microbiology 14, 1354–1363. https://doi.org/10.1111/j.1462-5822.2012.01825.x

Taschuk, F., Tapescu, I., Moy, R.H., Cherry, S., 2020. DDX56 Binds to Chikungunya Virus RNA To Control Infection. mBio 11, e02623-20. https://doi.org/10.1128/mBio.02623-20

Tossavainen, H., Aitio, O., Hellman, M., Saksela, K., Permi, P., 2016. Structural Basis of the High Affinity Interaction between the Alphavirus Nonstructural Protein-3 (nsP3) and the SH3 Domain of Amphiphysin-2. J Biol Chem 291, 16307–16317. https://doi.org/10.1074/jbc.M116.732412

Tourrière, H., Chebli, K., Zekri, L., Courselaud, B., Blanchard, J.M., Bertrand, E., Tazi, J., 2003. The RasGAP-associated endoribonuclease G3BP assembles stress granules. J Cell Biol 160, 823–831. https://doi.org/10.1083/jcb.200212128

Tuittila, M.T., Santagati, M.G., Röyttä, M., Määttä, J.A., Hinkkanen, A.E., 2000. Replicase Complex Genes of Semliki Forest Virus Confer Lethal Neurovirulence. J Virol 74, 4579–4589.

Ule, J., Hwang, H.-W., Darnell, R.B., 2018. The Future of Cross-Linking and Immunoprecipitation (CLIP). Cold Spring Harb Perspect Biol 10, a032243. https://doi.org/10.1101/cshperspect.a032243

Ule, J., Jensen, K., Mele, A., Darnell, R.B., 2005. CLIP: A method for identifying protein–RNA interaction sites in living cells. Methods, Post-transcriptional Regulation of Gene Expression 37, 376–386. https://doi.org/10.1016/j.ymeth.2005.07.018

Urdaneta, E.C., Beckmann, B.M., 2020. Fast and unbiased purification of RNA-protein complexes after UV cross-linking. Methods, High-Throughput Approaches in RNA Biology 178, 72–82. https://doi.org/10.1016/j.ymeth.2019.09.013

Valneva Successfully Completes Pivotal Phase 3 Trial of Single-Shot Chikungunya Vaccine Candidate – Valneva, n.d. URL https://valneva.com/press-release/valneva-successfully-completes-pivotal-phase-3-trial-of-single-shot-chikungunya-vaccine-candidate/ (accessed 3.25.22).

Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.-Y., Cody, N.A.L., Dominguez, D., Olson, S., Sundararaman, B., Zhan, L., Bazile, C., Bouvrette, L.P.B., Bergalet, J., Duff, M.O., Garcia, K.E., Gelboin-Burkhart, C., Hochman, M., Lambert, N.J., Li, H., McGurk, M.P., Nguyen, T.B., Palden, T., Rabano, I., Sathe, S., Stanton, R., Su, A., Wang, R., Yee, B.A., Zhou, B., Louie, A.L., Aigner, S., Fu, X.-D., Lécuyer, E., Burge, C.B., Graveley, B.R., Yeo, G.W., 2020. A large-scale binding and functional map of human RNA-binding proteins. Nature 583, 711–719. https://doi.org/10.1038/s41586-020-2077-3

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., Yeo, G.W., 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods 13, 508–514. https://doi.org/10.1038/nmeth.3810

Varjak, M., Žusinaite, E., Merits, A., 2010. Novel Functions of the Alphavirus Nonstructural Protein nsP3 C-Terminal Region. Journal of Virology 84, 2352–2364. https://doi.org/10.1128/JVI.01540-09

Verma, S., Goyal, S., Jamal, S., Singh, A., Grover, A., 2016. Hsp90: Friends, clients and natural foes. Biochimie 127, 227–240. https://doi.org/10.1016/j.biochi.2016.05.018

Voss, J.E., Vaney, M.-C., Duquerroy, S., Vonrhein, C., Girard-Blanc, C., Crublet, E., Thompson, A., Bricogne, G., Rey, F.A., 2010. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. Nature 468, 709–712. https://doi.org/10.1038/nature09555

Wahid, B., Ali, A., Rafique, S., Idrees, M., 2017. Global expansion of chikungunya virus: mapping the 64-year history. International Journal of Infectious Diseases 58, 69–76. https://doi.org/10.1016/j.ijid.2017.03.006

Wang, Z., Tollervey, J., Briese, M., Turner, D., Ule, J., 2009. CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo. Methods 48, 287–293. https://doi.org/10.1016/j.ymeth.2009.02.021

Wichit, S., Diop, F., Hamel, R., Talignani, L., Ferraris, P., Cornelie, S., Liegeois, F., Thomas, F., Yssel, H., Missé, D., 2017. Aedes Aegypti saliva enhances chikungunya virus replication in human skin fibroblasts via inhibition of the type I interferon signaling pathway. Infection, Genetics and Evolution 55, 68–70. https://doi.org/10.1016/j.meegid.2017.08.032

Wilson, J.A.C., Prow, N.A., Schroder, W.A., Ellis, J.J., Cumming, H.E., Gearing, L.J., Poo, Y.S., Taylor, A., Hertzog, P.J., Di Giallonardo, F., Hueston, L., Le Grand, R., Tang, B., Le, T.T., Gardner, J., Mahalingam, S., Roques, P., Bird, P.I., Suhrbier, A., 2017. RNA-Seq analysis of chikungunya virus infection and identification of granzyme A as a major promoter of arthritic inflammation. PLoS Pathog 13, e1006155. https://doi.org/10.1371/journal.ppat.1006155

Wintachai, P., Wikan, N., Kuadkitkan, A., Jaimipuk, T., Ubol, S., Pulmanausahakul, R., Auewarakul, P., Kasinrerk, W., Weng, W.-Y., Panyasrivanit, M., Paemanee, A., Kittisenachai, S., Roytrakul, S., Smith, D.R., 2012. Identification of prohibitin as a Chikungunya virus receptor protein. Journal of Medical Virology 84, 1757–1770. https://doi.org/10.1002/jmv.23403

Yaffe, D., Saxel, O., 1977. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature 270, 725–727. https://doi.org/10.1038/270725a0

Zarnegar, B.J., Flynn, R.A., Shen, Y., Do, B.T., Chang, H.Y., Khavari, P.A., 2016. irCLIP platform for efficient characterization of protein–RNA interactions. Nat Methods 13, 489–492. https://doi.org/10.1038/nmeth.3840

Zayas, M., Long, G., Madan, V., Bartenschlager, R., 2016. Coordination of Hepatitis C Virus Assembly by Distinct Regulatory Regions in Nonstructural Protein 5A. PLoS Pathog 12, e1005376. https://doi.org/10.1371/journal.ppat.1005376

Zeller, H., Van Bortel, W., Sudre, B., 2016. Chikungunya: Its History in Africa and Asia and Its Spread to New Regions in 2013–2014. The Journal of Infectious Diseases 214, S436–S440. https://doi.org/10.1093/infdis/jiw391

Zhang, C., Darnell, R.B., 2011. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechnol 29, 607–614. https://doi.org/10.1038/nbt.1873

Zhang, R., Kim, A.S., Fox, J.M., Nair, S., Basore, K., Klimstra, W.B., Rimkunas, R., Fong, R.H., Lin, H., Poddar, S., Crowe, J.E., Doranz, B.J., Fremont, D.H., Diamond, M.S., 2018. Mxra8 is a receptor for multiple arthritogenic alphaviruses. Nature 557, 570–574. https://doi.org/10.1038/s41586-018-0121-3

Zhang, S., Garzan, A., Haese, N., Bostwick, R., Martinez-Gzegozewska, Y., Rasmussen, L., Streblow, D.N., Haise, M.T., Pathak, A.K., Augelli-Szafran, C.E., Wu, M., 2021. Pyrimidone inhibitors targeting Chikungunya Virus nsP3 macrodomain by fragment-based drug design. PLoS One 16, e0245013. https://doi.org/10.1371/journal.pone.0245013

Zhao, H., Lindqvist, B., Garoff, H., von Bonsdorff, C.H., Liljeström, P., 1994. A tyrosine-based motif in the cytoplasmic domain of the alphavirus envelope protein is essential for budding. EMBO J 13, 4204–4211.

Zhao, S.-H., Wang, Y., Wen, L., Zhai, Z.-B., Ai, Z.-H., Yao, N.-L., Wang, L., Liu, W.-C., Chen, B.-L., Li, Y., Yang, H., 2013. Basigin-2 is the predominant basigin isoform that promotes tumor cell migration and invasion and correlates with poor prognosis in epithelial ovarian cancer. Journal of Translational Medicine 11, 92. https://doi.org/10.1186/1479-5876-11-92

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research 31, 3406–3415. https://doi.org/10.1093/nar/gkg595