Title
Zika Virus Disrupts the Nonsense-mediated mRNA Decay Pathway feat. A Clinical Application of Next Generation Sequencing

Author
Leon, Kristoffer Edgar

Publication Date
2020

Peer reviewed|Thesis/dissertation

Zika Virus Disrupts the Nonsense-mediated mRNA Decay Pathway feat. A Clinical Application of Next Generation Sequencing

by
Kristoffer Edgar Leon

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Michael Wilson
354F13DACE76493...

Michael Wilson
Chair

DocuSigned by:

Melanie Ott

DocuSigned by:
Stephen Floor
3709C3C6AF93461...

Stephen Floor

Committee Members

Dedication and Acknowledgements

The work in this thesis was done with the support of an F31 Ruth L. Kirschstein National Research Service Award (NRSA) Individual Predoctoral Fellowship to Promote Diversity in Health-Related Research and an Otellini Family UCSF Discovery Fellowship.

This work was made possible by my advisor, Melanie Ott. As an inspiring physician scientist, Melanie was a role model of how to create a career where medicine motivated the science and allowed me to establish a space where I could pursue those types of questions. Her guidance and mentorship during my PhD were invaluable to my career and growth as an independent scientist. I would also like to thank Krystal Fontaine, who provided an indescribable amount of mentorship and guidance. Lastly, all members of the Ott lab provided scientific support, boosts to morale and a wonderful place to do science.

I would also like to thank my thesis committee members, Michael Wilson and Stephen Floor. Their guidance and support helped guide my graduate school path and career.

Lastly, I'd like to thank my friends and family who have supported me during my PhD.

Contributions

The chapters described in this work are reproduced from these publications:

Chapter 1:

Leon, K. E. & Ott, M. An 'Arms Race' between the Nonsense-mediated mRNA Decay Pathway and Viral Infections. *Seminars in cell & developmental biology*, doi:10.1016/j.semcdb.2020.05.018 (2020).

Chapter 2:

Fontaine, K. A. and Leon, K.E. et al. The Cellular NMD Pathway Restricts Zika Virus Infection and Is Targeted by the Viral Capsid Protein. mBio 9, doi:10.1128/mBio.02126-18 (2018).

Chapter 3:

Leon, K. E. et al. Zika Virus Infection Prevents Host mRNA Nuclear Export by Disrupting UPF1 Function. bioRxiv, 2020.2012.2003.410837, doi:10.1101/2020.12.03.410837 (2020).

Chapter 4:

Leon, K. E. et al. Genomic and serologic characterization of enterovirus A71 brainstem encephalitis. Neurol Neuroimmunol Neuroinflamm 7, doi:10.1212/nxi.0000000000000703 (2020).

Zika Virus Disrupts the Nonsense-mediated mRNA Decay Pathway

feat. A Clinical Application of Next Generation Sequencing

Kristoffer Edgar Leon

Abstract

Zika virus (ZIKV) is a mosquito-borne illness responsible for an outbreak of microcephaly in Brazil in 2015. To better understand the molecular mechanisms behind the neurodevelopmental delay seen, we explored the pathways perturbed by ZIKV to establish an infection and cause the neurological sequelae seen in patients. We focused particularly on the Nonsense-mediated mRNA Decay (NMD) pathway: a cellular RNA quality control pathway responsible for degrading mRNAs with premature termination codons that also regulates a significant subset of normal mRNAs. To best study the impacts of ZIKV, we used neural progenitor cells (NPCs) derived from induced pluripotent stem cells. We found that during ZIKV infection, the viral Capsid protein interacted with and degraded UPF1, the master regulator of the NMD pathway, in the nucleus. UPF1 was a viral restriction factor, and removal of UPF1 from the cell leads to increased permissivity to ZIKV infection. We next found that ZIKV-mediated UPF1 degradation led to a loss of UPF1 occupancy on host transcripts, which caused a shift in cellular localization and trapped specific transcripts in the nucleus. An extracellular matrix protein involved in fetal development, FREM2, had mRNA retained in the nucleus, leading to decreased protein levels. Depletion of FREM2 in NPCs led to premature neuronal differentiation. Overall, it leads to a model system where ZIKV, using its capsid protein, causes destruction of UPF1 to promote a productive infection As we find UPF1 linked to many neurodevelopment pathways, we propose that the disruption of the NMD and the lack of host mRNA export contributes to virally induced neurodevelopmental disorders in ZIKV infection. Lastly, we show an example of how laboratory techniques used in the study of viral infections *in vitro* can be used in a clinical application. An outbreak of EV-A71 in Catalonia, Spain caused widespread brainstem encephalitis in a pediatric

population. To better understand the outbreak, we compare metagenomic next generation sequencing (mNGS) to quantitative PCR (qPCR) as a diagnostic tool in the CSF of patients. We show that mNGS is capable of identifying virus in samples that qPCR can't. Using mNGS data, we obtained several full-length genomes, allowing for phylogenetic analyses and the identification of a single mutation potentially responsible for the enhanced neuroinvasive characteristics. We then used VirScan, a phage display library expressing a wide variety of human viral protein peptides, to identify patients exposed to virus and characterize which viral epitopes were the most immunoreactive, complementing the mNGS data. Overall, we show that this standard laboratory technique can play an important role in patient care, especially for viral detection in typically hard-to-diagnose body sites. This also applies to surveillance of current and future viral outbreaks. Given the circumstances—writing a dissertation regarding the molecular pathogenesis of a virus responsible for an epidemic concurrent with the COVID-19 pandemic highlights the need for translation of cutting-edge laboratory techniques to improve virus detection and comprehension for more personalized therapeutics.

Table of Contents

Chapter 4

List of Figures

List of Tables

List of Abbreviations

ZIKV – Zika Virus

DENV- Dengue Virus

HCV – Hepatitis C Virus

NMD – Nonsene-mediated mRNA Decay

UPF1 – upframeshift protein 1

qPCR – quantitative polymerase chain eaction

mNGS – metagenomic next generation sequencing

CLIP – crosslinking and immunoprecipitation

EJC – exon-junction complex

PTC – premature termination codon

EV-A71 – enterovirus A71

UTR – untranslated eegion

NPC – neural progenitor cell

iPSC- induced pluripotent stem cell

HIV – Human Immunodeficiency Virus

Chapter 1

Introduction

An 'Arms Race' between the Nonsense-mediated mRNA Decay Pathway and Viral Infections

A version of this chapter was published as a review:

Leon, K. & Ott, M. An 'Arms Race' between the Nonsense-mediated mRNA Decay Pathway and Viral Infections. *Seminars in cell & developmental biology*, doi:10.1016/j.semcdb.2020.05.018 (2020).

1. Introduction

Transcripts with premature termination codons (PTCs) lead to the production of truncated proteins, which is potentially harmful for the cell (1, 2). To address this problem, the cell employs the NMD pathway, which is a translation-dependent process that targets PTC - containing transcripts with a set of factors conserved from yeast to man (3). However, recent research has revealed that NMD plays a broader role in biology. This includes surveillance of non-PTC containing messenger RNAs (mRNAs) implicated in a wide swath of biological functions, including development, stem cell differentiation, stress response, and protection from viral infections (4-8).

While our understanding of the NMD pathway is still evolving, it can be divided into at least two branches: one which is dependent on the Exon-Junction Complex (EJC-dependent or EJC-enhanced) and one which operates in an EJC-independent manner. The current understanding of these models is summarized below. After transcription occurs, mRNAs are spliced and components of the exon-junction complex (EJC) are loaded onto mRNAs. The EJC consists of several proteins, including MAGOH, RBM8A, BTZ and eIF4A3 (7). Transcripts are then

transported out of the nucleus, where translation begins, displacing EJCs positioned along the transcript (9). If the transcript contains a PTC, translation will terminate with one or more downstream EJCs non-displaced, which triggers initiation of “EJC-dependent” NMD (reviewed in (3)) (**Fig. 1.1A**).

The ribosome and retained EJC recruit core NMD proteins: the Up-Frameshift (UPF) proteins, which recognize and bind to the transcript, and Suppressors with Morphological Effects on Genitalia (SMG) proteins, which induce decay (6). The translation termination complex, which contains the helicase UPF1, the kinase SMG1, and the eukaryotic peptide chain release factor eRF1-3, provides the first signal for the NMD pathway. The remaining downstream EJC provides the second signal by binding to the UPF3 (or the partially redundant UPF3X) and UPF2 complex. The UPF2/3 complex acts as a bridge linking the EJC to the UPF1/SMG1 complex, which then activates SMG1 to phosphorylate UPF1 (1). Phosphorylation of UPF1 activates the Decay-Inducing Complex (DECID) pathway, which consists of decapping and deadenylation of the transcript by the deadenylase CCR4-NOT via the SMG5/7 heterodimer (10, 11), as well as endonucleolytic cleavage by SMG6 (12, 13). The resulting cleavage products are then further degraded, likely by the exoribonuclease XRN1 (14).

Recent systems-wide approaches uncovered that NMD not only targets “faulty” transcripts, but also a large number of “normal” transcripts, potentially 10-20% of the human transcriptome (15-20). Certain transcripts, especially those with long 3’ untranslated regions (UTRs) and GC-rich regions, are particularly susceptible to NMD degradation (21-23). In addition, NMD triggered by events other than EJCs deposited downstream of stop codons (EJC-independent) is particularly relevant in interactions of NMD with positive-sense RNA viruses.

This target range and preference for “unusual” RNAs form the basis for the ability of the NMD pathway to also degrade incoming viral RNAs or viral transcripts (24). In turn, viruses have evolved mechanisms to circumvent this innate antiviral defense mechanism, either to protect the viral RNA from degradation or to turn members of the NMD pathway in enabling factors to enhance viral replication (5). Below, we summarize examples of how many different kinds of viruses interact with the NMD system, and the implications that these mechanisms may have for the development of anti-viral therapies.

2. Positive Sense RNA Viruses

2.1 Alphavirus

Alphaviruses are small, enveloped viruses with a single strand positive-sense RNA genome. One of the first reports indicating a role for the NMD pathway in RNA virus infection came from the study of two alphaviruses, Semliki Forest Virus (SFV) and Sindbis Virus (SINV) (25). In a genome-wide screen in HeLa cells, UPF1 was identified as a restriction factor of SFV infection, and this function was confirmed in a model of SINV infection. UPF1 was found to act early in infection, and knockdown of UPF1 led to increased viral RNA synthesis and production of viral progeny. Downstream effectors of the NMD pathway, SMG5 and SMG7, were also found to restrict alphavirus infection. The study’s conclusion was that UPF1, SMG5, and SMG7 target the incoming viral RNA, leading to a decrease of all downstream viral events. Interestingly, the NMD pathway is known to target host transcripts with long 3’ UTRs, but deletion of the viral 3 UTR did not appear to impact UPF1-mediated viral RNA degradation (25). However, alphavirus proteins are translated from two different open reading frames: the nonstructural proteins from the genomic RNA and the structural proteins from shorter so-called subgenomic RNAs. Thus, when nonstructural proteins are translated, the remaining sequence encoding the structural

proteins forms a long 3□ sequence including the template for the structural proteins, potentially allowing for recognition by the NMD despite the deletion of the shorter genomic 3□ UTR (26). UPF1 also acted as a viral restriction factor of SINV in a fly model system (27). These data indicate that the NMD pathway, relying on UPF1 and the SMG proteins, can target incoming viral RNA, serving as an intrinsic immune system against RNA viruses.

2.2 Flavivirus

Flaviviridae also contain a positive-sense single-stranded RNA genome as a possible target for NMD. Flaviviridae were first linked to the antiviral activity of NMD in a screen for protein-protein interactions between host proteins and proteins encoded by the Hepatitis C Virus (HCV)(28). HCV chronically infects liver cells, in most cases leading to life-long progressive, and ultimately terminal, liver disease (29). Our group found that one NMD-associated protein, WIBG (also called partner of Y14/RBM8A and Magoh, PYM), interacted robustly with the HCV capsid protein called core. A parallel RNAi screen found that knockdown of WIBG led to decreased HCV replication. WIBG is known to interact with two members of the exon-junction complex, RBM8A and MAGOH (9, 30, 31), and is thought to recycle these proteins back into the nucleus (9) (**Fig. 1.1B, Fig. 1.2A**). Binding of HCV core to WIBG disrupts this interaction, leading to an overall decrease in cellular NMD activity as shown by the enhanced RNA levels of select NMD targets (28). Thus, HCV evolved to disrupt the antiviral activity of NMD. However, unlike with alphaviruses, the knockdown of UPF1 did not enhance viral replication. In contrast, our data support a model wherein HCV has disrupts WIBG and the NMD pathway to support its own life cycle.

Zika virus (ZIKV) is a mosquito-borne flavivirus that causes microcephaly in fetuses and a dengue-like fever in adults (32). Our group found that ZIKV has also evolved mechanisms to

disrupt the NMD pathway, one of them being that the ZIKV capsid interacts with and degrades UPF1 in the nucleus (33) (**Fig. 1.1B, Fig. 1.2A**). Depletion of UPF1 results in a marked increase in the permissiveness of neural progenitor cells to viral infection, indicating that NMD is active in tissue-specific progenitor cells, possibly as an added protection of these cells from viral infection (33).

WIBG, MAGOH, and UPF1 were all identified as restriction factors for Dengue Virus (DENV), ZIKV and West Nile Virus (WNV) infections in human cell lines (34), emphasizing that NMD is a common host response to many flaviviruses. Furthermore, MAGOH and RBM8A proteins re-localize from the cytoplasm to host membranes when WNV capsid protein is overexpressed (34), which may explain why NMD is disrupted in WNV-infected cells (**Fig. 1.1B, Fig 1.2A**). RBM8A was found to interact with viral RNA, dependent on the expression of WIBG or MAGOH proteins (34). Interaction of RBM8A with viral RNA is surprising because EJC deposition typically occurs in the nucleus, whereas flavivirus RNA replication occurs in the cytoplasm. Relocalization of EJC members from the nucleus to the cytoplasm during WNV infection could expose not only the viral RNA but also other cytoplasmic RNAs to members of the EJC. In addition, other members of the NMD pathway, such as UPF1, which is found in the nucleus and cytoplasm, could potentially recruit EJC family members to viral RNA. Flaviviruses are vector-borne, and the interactions between NMD proteins and viral proteins likely occur in both the insect vectors and human hosts. Collectively, these data uncover multiple interactions of flaviviruses with the NMD pathway, from acute suppression of NMD in self-limiting diseases like ZIKV, DENV and WNV infections to chronic disruption of individual NMD-associated proteins, like WIBG in HCV infection.

2.3 Coronavirus

Coronaviruses (CoV) comprise a group of large positive-stranded RNA viruses (~30 kb) that include SARS-CoV, MERS-CoV and the new SARS-CoV-2 causing the COVID-19 outbreak (35). Mouse Hepatitis Virus (MHV), a mammalian coronavirus, has been shown to be the target of the NMD pathway and replicate more efficiently when UPF1, UPF2, SMG5 or SMG6 proteins are individually knocked down (36). Furthermore, viral RNA transfected into cells that were rendered NMD-incompetent gained a significantly increased half-life, arguing that CoV RNA, like flavivirus and alphavirus RNA, may be a substrate for NMD-mediated degradation. Interestingly, the nucleocapsid of MHV, the N protein, which protects viral RNA from degradation, impairs the NMD pathway when co-transfected with an NMD reporter system (36), indicating an adaptation to actively antagonize the host response. Most recently, the SARS-CoV-2 nucleoprotein has also been shown to interact with UPF1 (37), highlighting that interaction of coronaviruses with the NMD pathway is conserved in both mice and humans.

2.4 Alphaflexiviridae and Tombusviridae

NMD is highly conserved among eukaryotes including plants (38). Potato virus X (PVX) is a positive-sense single-stranded RNA virus in the Alphaflexiviridae family that typically infects solanaceous crops, which include potatoes, tomatoes and peppers. PVX is also capable of infecting Nicotiana benthamiana, a permissive laboratory model and close relative of tobacco indigenous to Australia (39, 40). Using a standard Arabidopsis plant model system artificially expressing PVX, a screen was performed to identify restricting and enabling host factors. This screen identified UPF1, SMG7, and UPF3 as restriction factors, that when inactivated led to enhanced viral replication (41). In N. benthamiana, PVX replicated more efficiently when plants expressed a dominant negative mutant of UPF1. Furthermore, removal of the long 3’UTR from

the PVX genome also conferred protection from NMD-mediated degradation (**Fig. 1.2B**). Interestingly, PVX infection also impaired NMD activity and stabilized known NMD substrates in the Arabidposis model, although this function could not be confirmed in the Nicotiana system due to the lack of known NMD targets in these plants (41). This indicates that PVX may actively hamper NMD, as seen in other viral species.

Similar enhancement of viral replication by dominant negative UPF1 was observed when N. benthamiana was infected with Turnip crinkle virus (TCV). This virus belongs to the family of tombusviridae, and unlike PVX lacks both a 5' cap and a 3' polyA tail (41). Interestingly, tombusviridae showed a spectrum of susceptibility to the antiviral activity of NMD, with Pea Enation Mosaic Virus 2 (PEMV2) being more sensitive than TCV (**Fig. 1.2B**). While the 3'UTR of TCV conferred protection against NMD degradation, the 3'UTR of PEMV2 did not appear to provide the same protection (42) (**Fig. 1.2C**). Furthermore, the PEMV2 long distance movement protein p26 protects viral and host mRNAs (**Fig. 1.1B, 1.2B**) against NMD in a N. benthamiana system. The authors found that PEMV2 infection can protect nearly half of the NMD-sensitive target RNAs, which consist of host transcripts with long, GC-rich 3'UTRs. P26 expression alone increased transcript abundance for about a third of the NMD-sensitive target RNAs (**Fig. 1.1B**) (43). This indicates that PEMV2 likely has multiples ways of perturbing the NMD pathway.

3. Retroviruses

3.1 HTLV (Human T lymphotropic virus)

The human T lymphotropic virus (HTLV) is a human oncogenic retrovirus that contains two single-stranded RNA copies in its virion (44). Similar to the other viruses discussed above, HTLV has an RNA genome, but fundamental differences in the HTLV lifecycle, including reverse

transcription of the viral genome and subsequent integration into the host, provide different avenues by which NMD proteins could target and interact with the virus. HTLV-1 infected cells exhibit suppressed NMD function (45), and viral RNA co-immunoprecipitates with UPF1 (46). The central role of UPF1 in virus-targeted NMD was demonstrated by the finding that HTLV-1 mRNAs are sensitive to NMD degradation, which can be rescued by UPF1 knockdown (45). In contrast, ectopic UPF1 expression decreases the presence of viral RNA, whereas inhibition of NMD either with wortmannin (a PI3K inhibitor that targets SMG1 and is known to disrupt UPF1 function (47)) or siRNA knockdown leads to enhanced levels of viral RNA (46). In addition, two HTLV-1 proteins, the transactivator protein Tax and the viral RNA binding protein Rex, directly influence the NMD pathway (45, 46, 48, 49)(**Fig. 1.2D**).

Functional analysis showed that a reporter construct containing a PTC was significantly enriched in HTLV-1 infected cells, but only if those cell lines expressed Tax (45). Tax alone could stabilize PTC-containing reporter mRNAs, but also stabilized both NMD-susceptible host and HTLV-1 mRNAs (45). HTLV Tax protein interacts with three NMD proteins: UPF1, UPF2, and INT6 (45) (**Fig. 1.1B, Fig. 1.2D**). Tax interaction with UPF1 reduces the ability of UPF1 to interact with nucleic acids, which is thought to occur by steric hindrance of the RNA binding site. Tax also prevents translocation of UPF1 across nucleic acid substrates inducing substrate dissociation (49). Tax expression also reduces the interaction between INT6, a member of the eIF3 translation initiation factor, and UPF1. While UPF1 is known to interact with eIF3 to inhibit translation (50), the role of the INT6-UPF1 interaction is not fully understood. However, INT6 appears to play an important role in the NMD, as Tax mutants unable to bind to INT6 do not disrupt NMD (45). Interestingly, Tax expression induces UPF1 re-localization to P-bodies, cytoplasmic foci of untranslated mRNAs and proteins involved with mRNA decay and translation inhibition (51), and causes P-body enlargement, indicating inhibition of mRNA decay.

Similar to Tax, the HTLV Rex protein, responsible for stabilizing and exporting viral mRNAs from the nucleus (52) (**Fig. 1.1B**), can significantly hamper normal cellular NMD function. This effect was seen both for the PTC reporter system and for endogenous substrates of the NMD pathway (46). Rex overexpression stabilizes viral RNA, possibly due to NMD inhibition, which leads to increased viral RNA and consequently protein production (46).

Collectively, these data on both Tax and Rex indicate multiple pathways inhibit NMD, both on viral RNAs and on host RNAs. This leads to the hypothesis that not only are these viral proteins protecting viral RNA, but that disruption of the NMD pathway may result in a favorable environment for HTLV to replicate.

3.2 HIV

Like HTLV discussed above, HIV is a retrovirus with a single-stranded RNA genome that gets incorporated into the host genome. The first indication that UPF1 interacts with HIV was based on known links between HIV and the Staufen protein (53, 54). Staufen is a dsRNA-binding protein known to recruit UPF1 to specific RNAs (55). Staufen is also incorporated into HIV virions (53, 54), indicating that HIV could interact with UPF1. Indeed, UPF1 is incorporated into the HIV virion, and HIV viral RNA is destabilized when UPF1 is knocked down (56) (**Fig. 1.2E**). Furthermore, UPF1 overexpression enhances HIV replication and protein production, although interestingly, this effect is dependent on UPF1 ATPase activity, and independent of its role in NMD (56). Thus UPF1, separate from its role in the NMD pathway, appears to play a uniquely enabling role in the HIV life cycle. The necessity of UPF1 for HIV replication was independently confirmed by a group of researchers who found that HIV virions produced in UPF1-depleted cells showed reduced infectivity. Furthermore, the UPF1 ATPase region was also found to be responsible for proper virion infectivity (57).

UPF1 controls HIV replication at multiple levels, including nucleo-cytoplasmic shuttling of HIV RNA (**Fig. 1.2E**). Indeed, in the absence of the viral Rev protein, which exports unspliced viral RNAs from the nucleus (58), overexpression of UPF1 is sufficient to support viral export, and this requires the UPF1 nuclear localization sequence and nuclear export signal (59). In the same study, UPF2 and UPF3 were shown to be negative regulators of HIV, and UPF2 was excluded from HIV RNPs (59). This indicates that the canonical, NMD-related function of UPF1, including its interaction with UPF2 and UPF3, actually inhibits HIV replication by interfering with UPF1 recruitment to the HIV Genome. Blocking UPF2 interaction with UPF1 is thought to be particularly important for the production of infectious HIV virions.

Multiple NMD factors also control HIV latency, wherein viral gene expression is transcriptionally silenced. Specifically, UPF1 can reverse latency by enhancing viral RNA levels and viral gene expression, whereas UPF2 and SMG6 promote latency by binding to UPF1 and hindering its proviral activities. UPF2 sequesters UPF1 and prevents association with viral RNAs, while SMG6, an endonuclease, is recruited to RNA bound by UPF1 and destabilizes the viral RNA (60). The same study showed that SMG6 and UPF2 negatively impact HIV reactivation from latency in both primary CD4+ T-cells and macrophages (60, 61). In turn, SMG6 and UPF2 protein levels were decreased in infected macrophages. This underscores that the effects of NMD factors on HIV replication are complex, and point to UPF1 as a unique enabling factor for HIV independent of its role in NMD.

3.3 Rous Sarcoma Virus (RSV)

Rous Sarcoma virus (RSV), which causes sarcoma in fowl, was the first virus discovered to have oncogenic properties (62). While early studies on RSV found that certain RNA elements of the genome determined stability of the viral RNA (63-65), it was only after NMD was better understood that it became clear that these elements were linked to UPF1 and the NMD pathway (66-69). It was found that RSV RNAs containing PTCs were degraded at a higher rate than wild type, and co-transfection of a dominant negative UPF1 prevented degradation of the PTC-containing RSV RNAs, highlighting their status as NMD targets (66). Subsequent studies found that a 155-nucleotide element, known as the RNA Stability Element (RSE), in the 3'UTR of RSV RNAs conferred protection from NMD (67, 68) (**Fig. 1.2C**). The RSE contains polypyrimidine tracts that bind PTBP1, which excludes UPF1 from the target RNA (70). This evasion strategy is also relevant for several NMD-resistant host mRNAs (70). Furthermore, the RSE inhibited deadenylation and decay by XRN1 when inserted into 3'UTR of canonical NMD transcripts (69). This indicates that RSV has evolved a mechanism to escape the restriction imposed by the cellular NMD pathway and to successfully protect its genome and mRNAs from NMD attacks.

3.4 Pararetroviruses

Cauliflower Mosaic Virus (CaMV), a plant pararetrovirus in the family Caulimoviridae that infects the Cruciferae family, including Arabidopsis thaliana, also disrupts cellular NMD processes (71, 72). Pararetroviruses are similar to mammalian retroviruses, in that reverse transcription of an RNA intermediate is required to replicate, but pararetroviruses have a DNA genome that does not integrate into the host genome but instead remains in nuclear episomes (73). The CaMV viral transactivator protein (TAV) binds and destabilizes the RNA decapping complex VARICOSE (VCS), which then leads to the accumulation of NMD substrates in the cell (**Fig.**

1.1A). Specifically, transcripts with PTCs, normally targeted and eliminated by the NMD pathway, are stabilized by TAV overexpression. The proposed model is that by inhibiting NMD, TAV increases viral RNA levels present within the cell (72). Even though the pararetrovirus has a DNA genome, the production of a significant number of RNA intermediates for replication produces NMD substrates. Thus, protection of viral RNA from the NMD pathway remains an important aspect of infection.

4. Conclusions

In the last several years, our understanding of the importance of NMD as a cell-intrinsic antiviral restriction pathway in animals and plants has expanded rapidly. With this, we have also come to recognize that viruses have evolved multiple mechanisms to evade the antiviral activity of NMD (**Fig. 1.1, Fig. 1.2**). These range from hijacking UPF1 to inhibiting the deadenylation step of NMD. Remarkably, viruses such as HIV and HCV have co-opted components of the NMD pathway, making the NMD factors required for viral replication. Understanding NMD and its nuances in response to viral infections can inform diverse fields. For example, Cas9 activity generates nonsense mutations that are subsequently depleted as transcripts by the NMD pathway (74), which is an important consideration for the therapeutic applications of the CRISPR/Cas9 technology.

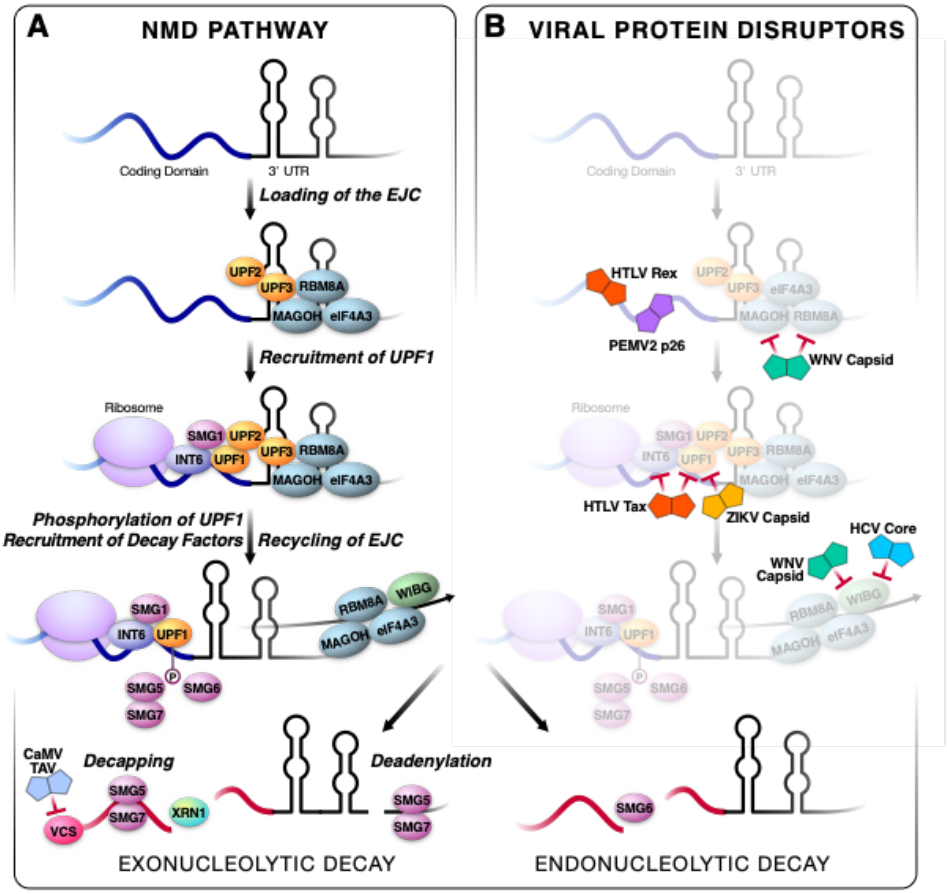
Understanding how viruses perturb NMD could also have implications for nonviral diseases. For example, in Duchenne's Muscular Dystrophy, mutations in the dystrophin gene lead to truncated transcripts that are subject to NMD. Preservation of those transcripts could result in partially functional proteins, thus taking lessons from viral pathogens on how to inhibit NMD could lead to novel therapeutic approaches (75).

In other cases, there may be therapeutic benefit to amplifying NMD. For example, microcephaly is caused by haploinsufficiency of members of the EJC, which are required for EJC-dependent NMD (76, 77). In the context of ZIKV infection, the capsid protein interferes with NMD by degrading UPF1 in the nucleus (33) or binding WIBG (34). Thus, small-molecule therapeutics that could correct the UPF1 downregulation and prevent capsid:WIBG interactions could exert anti-viral and anti-pathogenesis effects, as active NMD would be restored to prevent the establishment of viral infection and maintenance of transcript homeostasis.

As with many emerging fields, plenty of open questions remain. For example, is the interaction restricted to RNA viruses, or do DNA viruses such as herpesviruses also interact with NMD? In addition, little is known about the recognition of viral RNAs by NMD. What aspects of the NMD pathway are required to recognize “foreign” RNA, and what defines specificity towards aberrant RNA molecules? Spatial aspects are also important to clarify: many of the upstream factors of the canonical NMD pathway act in the nucleus, but many of the viruses described above, such as the flaviviruses, replicate in virally-induced compartments in the cytoplasm. Where does NMD of viral RNAs occur, and how do nuclear NMD factors interact with viral RNA?

Much more work is required to fully understand the impact of viruses on NMD and vice versa. Until then, the findings that many viruses have evolved elaborate mechanisms to impair NMD or to subvert select NMD members into enabling factors promoting viral infection remain the strongest argument for a critical role of the pathway in the innate anti-viral response.

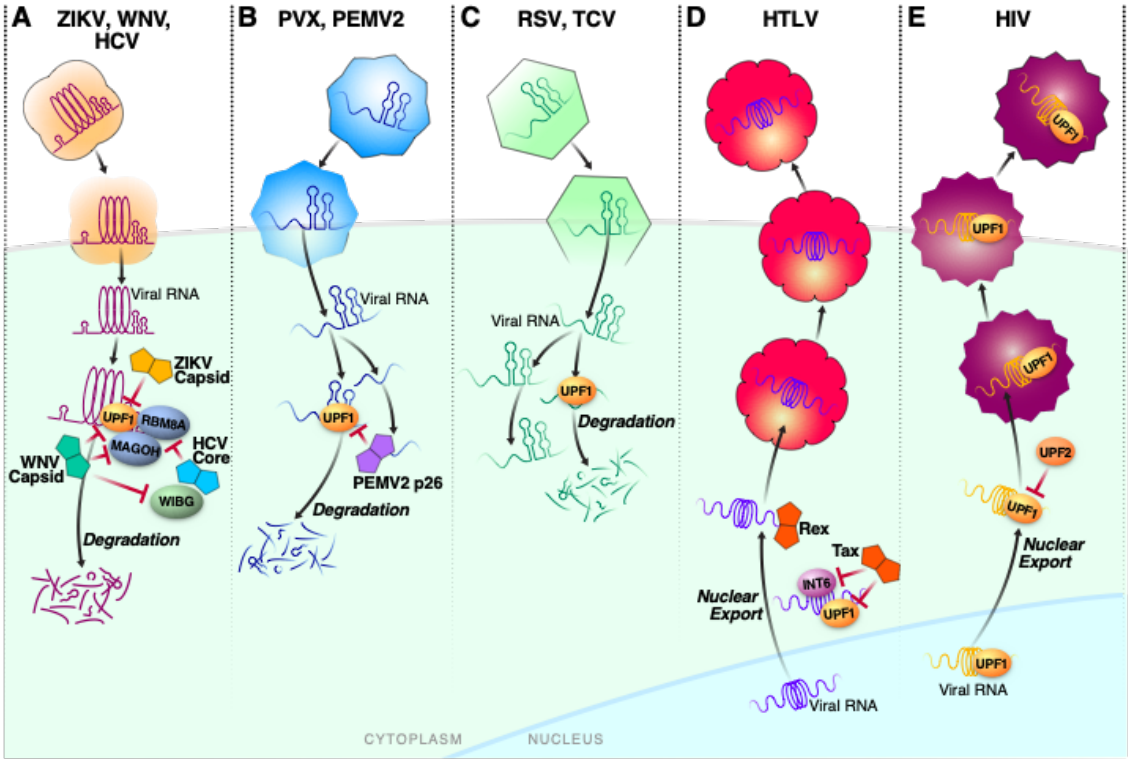
Figure 1.1



A) Schematic of the NMD pathway: Initially, a spliced mRNA has the exon junction complex (EJC) loaded on to the mRNA. This includes the core members: RBM8A, eIF4A3, and MAGOH. When there is a PTC upstream of a retained EJC, UPF2 and UPF3 act as a bridge to recruit the UPF1/SMG1 complex and initiate the NMD pathway. UPF1 is then phosphorylated by SMG1, leading to the recruitment of the decapping and deadenylating complex created by SMG5/SMG7, resulting in exonucleolytic decay. SMG6 is recruited for a separate endonucleolytic decay pathway. In plants, the complex VARICOSE (VCS) is responsible for decapping mRNAs.

B) Viral interactors overlaid on the NMD pathway: HTLV Rex and PEMVp62 protects mRNA from degradation by blocking NMD initiation on transcripts. WNV capsid interacts and interferes with RBM8A and MAGOH. HTLV Tax interferes with INT6 and UPF1 function, whereas ZIKV capsid blocks UPF1 function. WNV capsid and HCV core interfere with WIBG recycling of EJC factors. The CaMV TAV protein blocks decapping by interfering with the complex VCS.

Figure 1.2



A) Zika Virus (ZIKV), West Nile Virus (WNV) and Hepatitis C Virus (HCV): Flavivirus structural proteins protect viral mRNA. ZIKV and WNV capsid degrade UPF1. HCV core and WNV capsid interact with WIBG to block interaction between WIBG and RBM8A/MAGOH. WNV capsid also relocalizes RBM8A and MAGOH.
B) Potato Virus X (PVX) and Pea Enation Mosaic Virus 2 (PEMV2): Elements of the viral mRNA lead to NMD degradation. The 3'UTR contains elements responsible for NMD degradation, where removal of the 3'UTR protects viral RNA from the NMD. PEMV2 p62 interacts with RNA to protect it from degradation.
C) Rous Sarcoma Virus (RSV) and Turnip Crinkle Virus (TCV): Elements of the viral mRNA protect from NMD degradation. A segment of the 3' UTR confers protection to viral mRNAs from the NMD and protect from RNA decay machinery.
D) Human T-cell Lymphotropic Virus (HTLV): Two different HTLV proteins protect viral RNA from the NMD pathway. While Rex interaction with viral RNA confers protection, Tax plays a more active role by both inhibiting the ability of UPF1 to interact with RNA and properly translocate across bound RNA molecules. Furthermore, Tax inhibits INT6 interactions with UPF1.
E) Human Immunodeficiency Virus (HIV): UPF1 is involved in the RNA export of HIV genomes from the nucleus into the cytoplasm and then loaded into infectious virions. The host protein UPF2 inhibits UPF1 interaction with viral RNA and inhibits viral replication.

References

1. T. Kurosaki, L. E. Maquat, Nonsense-mediated mRNA decay in humans at a glance. *J Cell Sci* **129**, 461-467 (2016).
2. N. Hug, D. Longman, J. F. Cáceres, Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research* **44**, 1483-1495 (2016).
3. T. Kurosaki, M. W. Popp, L. E. Maquat, Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol* **20**, 406-420 (2019).
4. X. Han *et al.*, Nonsense-mediated mRNA decay: a 'nonsense' pathway makes sense in stem cell biology. *Nucleic Acids Res* **46**, 1038-1051 (2018).
5. G. Balistreri, C. Bognanni, O. Muhlemann, Virus Escape and Manipulation of Cellular Nonsense-Mediated mRNA Decay. *Viruses* **9**, (2017).
6. S. Lykke-Andersen, T. H. Jensen, Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol* **16**, 665-677 (2015).
7. M. W. Popp, L. E. Maquat, Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu Rev Genet* **47**, 139-165 (2013).
8. J. E. Smith, K. E. Baker, Nonsense-mediated RNA decay--a switch and dial for regulating gene expression. *Bioessays* **37**, 612-623 (2015).
9. N. H. Gehring, S. Lamprinaki, A. E. Kulozik, M. W. Hentze, Disassembly of exon junction complexes by PYM. *Cell* **137**, 536-548 (2009).
10. L. Unterholzner, E. Izaurralde, SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. *Mol Cell* **16**, 587-596 (2004).
11. B. Loh, S. Jonas, E. Izaurralde, The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev* **27**, 2125-2138 (2013).
12. S. Lykke-Andersen *et al.*, Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev* **28**, 2498-2517 (2014).

13. T. M. Franks, G. Singh, J. Lykke-Andersen, Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense- mediated mRNA decay. *Cell* **143**, 938-950 (2010).

14. N. Hug, D. Longman, J. F. Cáceres, Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res* **44**, 1483-1495 (2016).

15. J. T. Mendell, N. A. Sharifi, J. L. Meyers, F. Martinez-Murillo, H. C. Dietz, Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* **36**, 1073-1078 (2004).

16. H. Tani *et al.*, Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol* **9**, 1370-1379 (2012).

17. J. Weischenfeldt *et al.*, NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev* **22**, 1381-1396 (2008).

18. J. Wittmann, E. M. Hol, H. M. Jack, hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol Cell Biol* **26**, 1272-1287 (2006).

19. H. Yepiskoposyan, F. Aeschimann, D. Nilsson, M. Okoniewski, O. Mühlemann, Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* **17**, 2108-2118 (2011).

20. J. A. Hurt, A. D. Robertson, C. B. Burge, Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome Res* **23**, 1636-1650 (2013).

21. K. G. Toma, I. Rebbapragada, S. Durand, J. Lykke-Andersen, Identification of elements in human long 3' UTRs that inhibit nonsense-mediated decay. *Rna* **21**, 887-897 (2015).

22. B. W. Kebaara, A. L. Atkin, in *Nucleic Acids Res.* (2009), vol. 37, pp. 2771-2778.

23. N. Imamachi, K. A. Salam, Y. Suzuki, N. Akimitsu, A GC-rich sequence feature in the 3′ UTR directs UPF1-dependent mRNA decay in mammalian cells. *Genome Res* **27**, 407-418 (2017).

24. L. Contu, S. Steiner, V. Thiel, O. Muhlemann, The Role of Stress Granules and the Nonsense-mediated mRNA Decay Pathway in Antiviral Defence. *Chimia (Aarau)* **73**, 374-379 (2019).

25. G. Balistreri *et al.*, The host nonsense-mediated mRNA decay pathway restricts Mammalian RNA virus replication. *Cell Host Microbe* **16**, 403-411 (2014).

26. J. L. Hyde *et al.*, The 5' and 3' ends of alphavirus RNAs--Non-coding is not non-functional. *Virus Res* **206**, 99-107 (2015).

27. M. F. Wernet, M. Klovstad, T. R. Clandinin, Generation of infectious virus particles from inducible transgenic genomes. *Curr Biol* **24**, R107-108 (2014).

28. H. R. Ramage *et al.*, A Combined Proteomics/Genomics Approach Links Hepatitis C Virus Infection with Nonsense-Mediated mRNA Decay. *Mol Cell* **57**, 329-340 (2015).

29. M. Houghton, Hepatitis C Virus: 30 Years after Its Discovery. *Cold Spring Harb Perspect Med* **9**, (2019).

30. M. D. Diem, C. C. Chan, I. Younis, G. Dreyfuss, PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat Struct Mol Biol* **14**, 1173-1179 (2007).

31. F. Bono *et al.*, Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex. *EMBO Rep* **5**, 304-310 (2004).

32. T. C. Pierson, M. S. Diamond, The emergence of Zika virus and its new clinical syndromes. *Nature* **560**, 573-581 (2018).

33. K. A. Fontaine *et al.*, The Cellular NMD Pathway Restricts Zika Virus Infection and Is Targeted by the Viral Capsid Protein. *MBio* **9**, (2018).

34. M. Li *et al.*, Identification of antiviral roles for the exon-junction complex and nonsense-mediated decay in flaviviral infection. *Nat Microbiol* **4**, 985-995 (2019).

35. N. Zhu *et al.*, A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* **382**, 727-733 (2020).

36. M. Wada, K. G. Lokugamage, K. Nakagawa, K. Narayanan, S. Makino, Interplay between coronavirus, a cytoplasmic RNA virus, and nonsense-mediated mRNA decay pathway. *Proc Natl Acad Sci U S A* **115**, E10157-10166 (2018).

37. D. E. Gordon *et al.*, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, (2020).

38. J. P. B. Lloyd, The evolution and diversity of the nonsense-mediated mRNA decay pathway. *F1000Res* **7**, 1299 (2018).

39. C. Lico, E. Benvenuto, S. Baschieri, The Two-Faced Potato Virus X: From Plant Pathogen to Smart Nanoparticle. *Front Plant Sci* **6**, 1009 (2015).

40. M. M. Goodin, D. Zaitlin, R. A. Naidu, S. A. Lommel, Nicotiana benthamiana: its history and future as a model for plant-pathogen interactions. *Mol Plant Microbe Interact* **21**, 1015-1026 (2008).

41. D. Garcia, S. Garcia, O. Voinnet, Nonsense-mediated decay serves as a general viral restriction mechanism in plants. *Cell Host Microbe* **16**, 391-402 (2014).

42. J. P. May, X. Yuan, E. Sawicki, A. E. Simon, RNA virus evasion of nonsense-mediated decay. *PLoS Pathog* **14**, e1007459 (2018).

43. J. P. May, P. Z. Johnson, M. Ilyas, F. Gao, A. E. Simon, The Multifunctional Long-Distance Movement Protein of Pea Enation Mosaic Virus 2 Protects Viral and Host Transcripts from Nonsense-Mediated Decay. *mBio* **11**, (2020).

44. N. Futsch, R. Mahieux, H. Dutartre, HTLV-1, the Other Pathogenic Yet Neglected Human Retrovirus: From Transmission to Therapeutic Treatment. *Viruses* **10**, (2017).

45. V. Mocquet *et al.*, The human T-lymphotropic virus type 1 tax protein inhibits nonsense-mediated mRNA decay by interacting with INT6/EIF3E and UPF1. *J Virol* **86**, 7530-7543 (2012).

46. K. Nakano *et al.*, Viral interference with host mRNA surveillance, the nonsense-mediated mRNA decay (NMD) pathway, through a new function of HTLV-1 Rex: implications for retroviral replication. *Microbes Infect* **15**, 491-505 (2013).

47. M. Pal, Y. Ishigaki, E. Nagy, L. E. Maquat, Evidence that phosphorylation of human Upfl protein varies with intracellular location and is mediated by a wortmannin-sensitive and rapamycin-sensitive PI 3-kinase-related kinase signaling pathway. *Rna* **7**, 5-15 (2001).

48. V. Mocquet *et al.*, in *Retrovirology*. (2014), vol. 11, pp. O61.

49. F. Fiorini *et al.*, HTLV-1 Tax plugs and freezes UPF1 helicase leading to nonsense-mediated mRNA decay inhibition. *Nat Commun* **9**, 431 (2018).

50. O. Isken *et al.*, Upf1 Phosphorylation Triggers Translational Repression during Nonsense-Mediated mRNA Decay. *Cell* **133**, 314-327 (2008).

51. C. J. Decker, R. Parker, P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harb Perspect Biol* **4**, a012286 (2012).

52. K. Nakano, T. Watanabe, HTLV-1 Rex Tunes the Cellular Environment Favorable for Viral Replication. *Viruses* **8**, 58 (2016).

53. L. Chatel-Chaix *et al.*, Identification of Staufen in the human immunodeficiency virus type 1 Gag ribonucleoprotein complex and a role in generating infectious viral particles. *Mol Cell Biol* **24**, 2637-2648 (2004).

54. A. J. Mouland *et al.*, The double-stranded RNA-binding protein Staufen is incorporated in human immunodeficiency virus type 1: evidence for a role in genomic RNA encapsidation. *J Virol* **74**, 5441-5451 (2000).

55. Y. K. Kim, L. Furic, L. Desgroseillers, L. E. Maquat, Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. *Cell* **120**, 195-208 (2005).

56. L. Ajamian *et al.*, Unexpected roles for UPF1 in HIV-1 RNA metabolism and translation. *RNA* **14**, 914-927 (2008).

57. A. K. Serquina *et al.*, UPF1 is crucial for the infectivity of human immunodeficiency virus type 1 progeny virions. *J Virol* **87**, 8853-8861 (2013).

58. M. Blissenbach, B. Grewe, B. Hoffmann, S. Brandt, K. Uberla, Nuclear RNA export and packaging functions of HIV-1 Rev revisited. *J Virol* **84**, 6598-6604 (2010).

59. L. Ajamian *et al.*, HIV-1 Recruits UPF1 but Excludes UPF2 to Promote Nucleocytoplasmic Export of the Genomic RNA. *Biomolecules* **5**, 2808-2839 (2015).

60. S. Rao, R. Amorim, M. Niu, A. Temzi, A. J. Mouland, The RNA surveillance proteins UPF1, UPF2 and SMG6 affect HIV-1 reactivation at a post-transcriptional level. *Retrovirology* **15**, 42 (2018).

61. S. Rao *et al.*, Host mRNA decay proteins influence HIV-1 replication and viral gene expression in primary monocyte-derived macrophages. *Retrovirology* **16**, 3 (2019).

62. R. A. Weiss, P. K. Vogt, 100 years of Rous sarcoma virus. *J Exp Med* **208**, 2351-2355 (2011).

63. G. F. Barker, K. Beemon, Rous sarcoma virus RNA stability requires an open reading frame in the gag gene and sequences downstream of the gag-pol junction. *Mol Cell Biol* **14**, 1986-1996 (1994).

64. G. F. Barker, K. Beemon, Nonsense codons within the Rous sarcoma virus gag gene decrease the stability of unspliced viral RNA. *Mol Cell Biol* **11**, 2760-2768 (1991).

65. S. Arrigo, K. Beemon, Regulation of Rous sarcoma virus RNA splicing and stability. *Mol Cell Biol* **8**, 4858-4867 (1988).

66. J. J. LeBlanc, K. L. Beemon, Unspliced Rous Sarcoma Virus Genomic RNAs Are Translated and Subjected to Nonsense-Mediated mRNA Decay before Packaging. *J Virol* **78**, 5139-5146 (2004).

67. J. E. Weil, K. L. Beemon, A 3′ UTR sequence stabilizes termination codons in the unspliced RNA of Rous sarcoma virus. *RNA* **12**, 102-110 (2006).

68. J. B. Withers, K. L. Beemon, Structural features in the Rous sarcoma virus RNA stability element are necessary for sensing the correct termination codon. *Retrovirology* **7**, 65 (2010).

69. V. Balagopal, K. L. Beemon, Rous Sarcoma Virus RNA Stability Element Inhibits Deadenylation of mRNAs with Long 3′UTRs. *Viruses* **9**, (2017).

70. Z. Ge, B. L. Quek, K. L. Beemon, J. R. Hogg, Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. *Elife* **5**, (2016).

71. M. Haas, M. Bureau, A. Geldreich, P. Yot, M. Keller, Cauliflower mosaic virus: still in the news. *Mol Plant Pathol* **3**, 419-429 (2002).

72. N. Lukhovitskaya, L. A. Ryabova, Cauliflower mosaic virus transactivator protein (TAV) can suppress nonsense-mediated decay by targeting VARICOSE, a scaffold protein of the decapping complex. *Sci Rep* **9**, 7042 (2019).

73. T. Hohn, H. Rothnie, Plant pararetroviruses: replication and expression. *Curr Opin Virol* **3**, 621-628 (2013).

74. M. W. Popp, L. E. Maquat, Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. *Cell* **165**, 1319-1322 (2016).

75. J. N. Miller, D. A. Pearce, Nonsense-Mediated Decay in Genetic Disease: Friend or Foe? *Mutat Res Rev Mutat Res* **0**, 52-64 (2014).

76. J. J. McMahon, E. E. Miller, D. L. Silver, The exon junction complex in neural development and neurodevelopmental disease. *Int J Dev Neurosci* **55**, 117-123 (2016).

77. H. Mao, J. J. McMahon, Y. H. Tsai, Z. Wang, D. L. Silver, Haploinsufficiency for Core Exon Junction Complex Components Disrupts Embryonic Neurogenesis and Causes p53-Mediated Microcephaly. *PLoS Genet* **12**, e1006282 (2016).

Chapter 2

The cellular NMD pathway restricts Zika virus infection and is targeted by the viral capsid protein

A version of this chapter was published:
Fontaine, K. A. et al. The Cellular NMD Pathway Restricts Zika Virus Infection and Is Targeted by the Viral Capsid Protein. mBio 9, doi:10.1128/mBio.02126-18 (2018).

Introduction

ZIKV is a mosquito-borne RNA virus that belongs to the *Flaviviridae* family. First isolated in Uganda in 1947, ZIKV remained relatively obscure for decades following its discovery because infection was associated with only mild disease. However, more severe clinical manifestations, including microcephaly, have been observed during the recent spread of ZIKV through the Americas (1). ZIKV infection induces cell cycle growth arrest and apoptosis in neural progenitor cells (NPCs) in *in vitro* studies and *in vivo* mouse models, with the latter resulting in cortical thinning and microcephaly (2-6). While it is now established that ZIKV infection during pregnancy is a causative agent of microcephaly (7), the molecular mechanisms underlying ZIKV-induced neuropathogenesis remain largely unknown.

Similar to other flaviviruses, ZIKV contains a single-stranded, positive-sense RNA genome of ~11 kb in size. The genome encodes a single polyprotein that is post-translationally processed by both host and viral proteases to produce 3 structural and 7 nonstructural proteins (8, 9). The flavivirus capsid, which is the first protein encoded in the genome, is a major structural element required for the encapsidation of the RNA genome during virion assembly (10). While flavivirus replication is known to occur in the cytoplasm, a significant portion of the viral capsid protein

localizes to the nucleus during infection (10, 11). Although the role of nuclear capsid during infection is less clear, several functions have been suggested. The capsid protein from dengue virus, a close relative of ZIKV, binds to core histones and inhibits nucleosome formation, thus implicating the protein in altering host gene expression (12). Furthermore, several flavivirus capsid proteins, including ZIKV capsid, localize to the nucleolus, with many interacting with nucleolar proteins to promote viral particle production (13-16).

The nonsense-mediated mRNA decay (NMD) pathway was initially discovered as a highly conserved quality control system that destroys transcripts containing premature termination codons (PTCs)(17). Following splicing of pre-mRNAs, a multi-subunit protein complex called the exon-junction complex (EJC) is deposited onto mRNAs near the sites of exon-exon junctions. If a PTC is found ~50–55 nucleotides upstream of an EJC, the mRNA will be subjected to NMD-mediated degradation initiated by the recruitment of the RNA helicase up-frameshift protein 1 (UPF1). UPF1 plays a central role in the NMD pathway by linking the translation termination event to the assembly of a surveillance complex, resulting in NMD activation (18). Interestingly, microcephaly has been associated with genetic mutations that result in the impairment of the NMD pathway. While knockout of Upf1 and other NMD factors is embryonic lethal in mice (19), mice haploinsufficient for the EJC components Magoh, Rbm8a, and Eif4a3 exhibit aberrant neurogenesis and microcephaly (20-22).

In addition to PTC-containing transcripts, it is now known that the NMD pathway recognizes a broader range of RNA substrates. Notably, the NMD controls the “normal” expression of ~10% of the cellular transcriptome and is regarded as a post-transcriptional mechanism of gene regulation (23). Furthermore, the NMD pathway also regulates viral infections. While it was first reported that UPF1 promotes the infectivity of HIV-1 progeny virions (24), replication of several human RNA viruses, including human T-cell lymphotropic virus type 1 (HTLV-1), Semliki Forest

virus and Sindbis virus, is enhanced following UPF1 knockdown, implicating UPF1 and the NMD pathway, either directly or indirectly, in the host antiviral response (25-28). As ZIKV infection and NMD impairment both promote microcephaly development, and we previously described disruption of the NMD pathway in cells infected with a related flavivirus, the hepatitis C virus (HCV) (29), we hypothesized that ZIKV infection manipulates the cellular NMD pathway, a process contributing to ZIKV-induced neuropathology.

Results

The NMD pathway is impaired during ZIKV infection. To determine if ZIKV infection affects NMD, we infected human hepatoma cells (Huh7) and human induced pluripotent stem cell (iPSC)-derived NPCs with ZIKV for 48 h. We isolated total RNA from infected cells and measured mRNA levels of three canonical NMD substrates: asparagine synthetase (ASNS), cysteinyl-tRNA synthetase (CARS), and SR protein SC35 (29). ASNS, CARS, and SC35 transcripts were significantly elevated in Huh7 cells and NPCs following infection with Asian lineage ZIKV strain P6-740 (**Fig. 2.1a**). Levels of NMD substrates were also elevated in Huh7 cells infected with the contemporary ZIKV clinical isolate PRVABC59 (Puerto Rico, 2015)(**Fig. 2.1a**). We found that ZIKV-induced increase in NMD transcripts did not reflect a global increase in transcription, as mRNA levels of housekeeping genes, including glyceraldehyde 3-phosphate dehydrogenase (GAPDH), were not altered in infected cells (**Fig. 2.1a**). Together, these results indicate that the NMD pathway is impaired in ZIKV-infected cells.

NMD substrates are regulated through the ATP-dependent RNA helicase activity of UPF1, the central regulator of NMD (18). To determine if ZIKV infection broadly affects NMD, we utilized two publicly available RNA sequencing (RNA-Seq) datasets to compare genome-wide transcriptional alterations found during ZIKV infection (6) to those found following UPF1

knockdown (30). As shown in **Figure 2.1b**, there is a significant overlap in upregulated genes between these two datasets. Interestingly, several of the overlapping genes are canonical NMD substrates (31-35) involved in cell cycle arrest and induction of apoptosis, two conditions linked to ZIKV-associated neuropathology (5). These genes include DNA damage-inducible transcript 3 (DDIT3)(36) and growth arrest and DNA damage-inducible protein 45 alpha and beta (GADD45A and GADD45B, respectively)(37). Via quantitative real-time RT-PCR, we confirmed that transcripts of each were upregulated following infection of Huh7 cells with ZIKV PRVABC59, while the mRNA levels of the housekeeping genes GAPDH, hypoxanthine phosphoribosyltransferase 1 (HPRT1), and lactate dehydrogenase A (LDHA) were not elevated (**Fig. 2.1c**). Combined, these data show that ZIKV infection is associated with dysregulated expression of NMD substrates relevant to ZIKV-mediated neuropathogenesis.

ZIKV capsid interacts with the NMD pathway. We previously showed that the core protein of HCV and capsid protein of the related flaviviruses dengue virus and West Nile virus interact with within bgcn homolog (WIBG/PYM1), an EJC disassembly factor associated with NMD (29, 38). To examine potential interactions between ZIKV and the NMD pathway, we separately analyzed data generated from an affinity purification-mass spectrometry (AP-MS) screen to specifically query whether the capsid protein of ZIKV interacts with NMD-associated host factors (Shah et al., submitted). ZIKV-host protein-protein interaction (PPI) maps were generated in HEK293T cells using ZIKV proteins from the Ugandan 1947 strain MR 766 or the French Polynesian 2013 strain H/PF/2013 as bait proteins.

From this analysis, we found that ZIKV capsid proteins interacted with several factors of the NMD pathway, including multiple members of the EJC complex, as well as UPF1 and UPF3B, the latter an NMD effector that stimulates UPF1 helicase activity (**Fig. 2.2a**)(39). Importantly, the NMD host factors that interacted with each of the two different capsid proteins greatly

overlapped, revealing that the interaction between capsid and the NMD pathway is conserved across the Asian and African lineages of ZIKV (**Fig. 2.2a**). Next, we validated the binding of ZIKV capsid to select NMD host factors by co-immunoprecipitating FLAG-tagged capsid protein with endogenous UPF3B or UPF1 in HEK293T cells. Both UPF3B and UPF1 proteins co-immunoprecipitated with ZIKV capsid, thus confirming the AP-MS results (**Fig. 2.2b,c**, respectively). UPF1 interacted with the viral capsid protein independently from its RNA-binding and ATPase/helicase capacities, as ZIKV capsid co-immunoprecipitated with UPF1 mutants deficient in these functions following overexpression (**Fig 2.2d**).

ZIKV capsid degrades UPF1, the master regulator of NMD. Surprisingly, we consistently observed a decrease in UPF1, but not UPF3B, protein levels in the input lysate of ZIKV capsid-transfected cells, pointing to a specific perturbation of endogenous UPF1 expression by ZIKV capsid (**Fig. 2.2c**). To confirm that UPF1 protein levels are dysregulated during ZIKV infection, we performed western blot analysis of infected Huh7 cells and NPCs. Cellular UPF1 protein levels were consistently downregulated by ~50% in ZIKV-infected Huh7 cells, whereas a ~25% reduction was observed in ZIKV-infected NPCs (**Fig. 2.2e**), mirroring the difference in infection efficiencies achieved in these two cell systems. UPF1 transcript levels were not decreased in ZIKV-infected cells or following ZIKV capsid overexpression, indicating that UPF1 is post-transcriptionally downregulated during ZIKV infection (**Supplemental Fig. 2.1a,b**, respectively).

Because ZIKV capsid and UPF1 both localize to the nucleus and the cytoplasm (13, 40) we performed fractionation studies in ZIKV capsid-transfected HEK293T cells to determine if UPF1 is downregulated within a specific cellular compartment. Capsid expression markedly decreased nuclear UPF1 levels, whereas cytoplasmic levels were unchanged (**Fig. 2.3a**). We next examined a potential role for the autophagic and proteasomal pathways, both of which are known to mediate nuclear protein degradation, in ZIKV capsid-induced UPF1 downregulation.

As shown in **Supplemental Figure 2.2**, nuclear UPF1 levels in ZIKV capsid-transfected cells were not rescued by inhibition of cellular autophagy via bafilomycin A1 treatment. Instead, nuclear UPF1 levels were restored in a dose-dependent manner following treatment with the proteasome inhibitor bortezomib (**Fig. 2.3b**), indicating enhanced proteasomal degradation of nuclear UPF1 in the presence of ZIKV capsid. Although ZIKV capsid colocalized with endogenous UPF1 in the cytoplasm of transfected Huh7-Lunet cells (Mander's colocalization coefficient of ~57%), we detected very little colocalization within the nucleus (~7%) (**Supplemental Figure 2.3**). We hypothesized that this was due to specific degradation of colocalizing UPF1 by ZIKV capsid in the nucleus. Indeed, when cells were treated with bortezemib, the fraction of nuclear UPF1 colocalizing with ZIKV capsid increased ~8-fold, while the fraction of nuclear capsid interacting with UPF1 remained unchanged (**Figure 2.3c**). These results demonstrate that ZIKV capsid interacts with UPF1 both in the cytoplasm and nucleus, but specifically targets nuclear UPF1 for proteasomal degradation.

UPF1 is a restriction factor of ZIKV. To test the effect of lowered UPF1 levels on ZIKV infection, we further decreased UPF1 expression prior to ZIKV infection by transfecting NPCs with either non-targeting siRNA or a pool of UPF1-specific siRNAs. We then infected the transfected cells with ZIKV and measured viral RNA levels, as well as infectious titers, 48 h post-infection (hpi). UPF1 knockdown was successful in siRNA-treated cells, as confirmed by western blot analysis (**Fig. 2.4a**). The depletion of UPF1 in NPCs prior to infection resulted in a significant increase in both ZIKV RNA levels and infectious virus production (**Fig. 2.4b,c,** respectively), indicating that expression of UPF1 restricts ZIKV infection at or before the RNA replication stage. To differentiate between these two stages, we analyzed double-stranded RNA (dsRNA) intermediates representing presumed viral RNA replication centers in infected NPCs (**Fig. 2.4d**)(41). Using confocal microscopy and 3D reconstruction analyses, we observed no significant difference in the number and size of dsRNA foci per cell when comparing ZIKV-

infected, UPF1-depleted NPCs to ZIKV-infected cultures expressing UPF1 (**Fig. 2.4e,f**). Instead, we found a significant increase in the number of infected cells in NPC cultures when UPF1 was depleted, indicating that UPF1 regulates permissivity of NPCs to ZIKV infection at an early stage prior to viral RNA replication (**Fig. 2.4g**).

Discussion

In summary, we identified the NMD pathway as a restriction mechanism for ZIKV infection in human NPCs. NMD was partially inactivated in ZIKV-infected NPCs through expression of the viral capsid protein and the resulting degradation of host nuclear UPF1. As further weakening NMD by depleting UPF1 resulted in a marked increase in the number of infected cells, we propose a model in which an evolutionary "arms race" between cellular NMD and ZIKV determines whether a cell is successfully infected (**Fig. 2.4h**).

Downregulation of UPF1 by ZIKV capsid is not complete and is likely limited by the damaging effects of NMD disruption, as illustrated by the upregulation of genes involved in cell growth arrest and apoptosis. Indeed, knockout of UPF1 and other members of the NMD pathway is embryonic lethal in mice (19). However, mice haploinsufficent for NMD factors upstream of UPF1, including Magoh, Rbm8a, and Eif4a3, develop microcephaly (20-22). Thus, the reduction in nuclear UPF1 we observe in ZIKV-infected NPCs could contribute to the microcephaly phenotype caused by ZIKV infection in the fetal brain. While fetal and adult NPCs appear to be transcriptionally distinct (42), it has been shown that adult NPCs are also permissive to ZIKV infection (43). As the NMD pathway is a ubiquitous cellular surveillance mechanism, it is likely that ZIKV capsid targets UPF1 for degradation in any cell type that is susceptible to ZIKV infection. Accordingly, we have found that UPF1 is degraded following infection of both NPCs and hepatic Huh7 cells (**Figure 2.2e**). Therefore, we believe that the interaction between ZIKV

capsid and UPF1 occurs during infection of both fetal and adult NPCs, but that the loss of this stem cell population is far more deleterious in the developing brain.

Why ZIKV capsid specifically downregulates nuclear UPF1, and how nuclear UPF1 contributes to ZIKV restriction remains unanswered. Several studies suggest that NMD is associated with the nucleus, although this issue remains controversial. Multiple transcripts, such as those encoding T cell receptor beta, triosephosphate isomerase and mouse major urinary protein, have been shown to be specifically degraded in purified nuclei or reduced in nuclear fractions (44). These data support the model that reducing nuclear UPF1 levels selectively disrupts NMD function in ZIKV-infected cells. In addition, UPF1 is involved in several other processes within the nucleus, including nuclear-associated RNA metabolism, cell cycle progression, and DNA replication, suggesting that ZIKV capsid could target nuclear UPF1 to disrupt these processes and program target cells for viral replication (40). Notably, viral RNA replication is thought to occur solely within the cytoplasmic compartment (45, 46). Using confocal microscopy and 3D reconstruction, we did not detect dsRNA foci localized within the nuclei of ZIKV-infected cells (data not shown), supporting our finding that UPF1 does not restrict viral RNA replication. While our results suggest a role for nuclear UPF1 in ZIKV restriction, it is possible that UPF1 also serves as a restriction factor of ZIKV within the cytoplasm. Previously, it was shown that UPF1 suppresses alphavirus replication by degrading the incoming viral RNA following uncoating in the cytosol (28). Thus, ZIKV may possess an additional mechanism to prevent cytoplasmic UPF1 from targeting its incoming RNA genome for destruction.

Our data reveal that nuclear UPF1 is degraded by ZIKV capsid in a proteasome-dependent manner. While the nuclear proteasome has not been specifically linked to microcephaly, it plays critical roles in the regulation of chromatin structure, gene expression, DNA repair, and protein quality control (47). Thus, the co-opting of the nuclear proteasome by ZIKV capsid to degrade

UPF1 could disrupt its normal proteasomal activity and further contribute to the cytopathic effects associated with ZIKV infection. Furthermore, given that the flavivirus capsid protein can translocate cell membranes directly (48), it is possible that capsid released from apoptotic, ZIKV-infected cells can enter neighboring, uninfected cells to degrade UPF1 and increase the permissivity of the cell to ZIKV infection. Studies are ongoing to determine the precise molecular mechanism of ZIKV capsid-mediated UPF1 degradation and how UPF1 depletion enhances ZIKV replication, directly or indirectly. Ultimately, these data may help inform new therapeutic approaches, as reinforcement of the antiviral properties of the NMD pathway is expected to enhance resistance of NPCs to ZIKV infection and to promote normal neurodevelopment in infected fetuses.

Materials and Methods

Viruses and cells. Two Asian lineage strains of ZIKV, P6-740 (ATCC VR-1845) and PRVABC59 (ATCC VR-1843), were used for all experiments. ZIKV stocks were propagated in Vero cells (ATCC) and titers were determined by plaque assays on Vero cells. Huh7 cells (ATCC), Huh7-Lunet cells (Ralf Bartenschlager, Heidelberg University) and Vero cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum (FBS), 2 mM L-glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin. HEK293T cells (ATCC) and the GFPu-1 cell line (ATCC CRL-2794) were maintained in DMEM/H21 medium supplemented with 10% FBS, 100 U/mL penicillin, 100 µg/mL streptomycin, and 1 mM sodium pyruvate or DMEM with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin. Human iPSC-derived NPCs were generated and maintained as described previously (49). All of the human fibroblast cell lines used to generate iPSCs came from the Coriell Institute for Medical Research and Yale Stem Cell Center. The iPSCs used in these studies were the CTRL2493nXX, CS2518nXX, and Cs71iCTR-20nXX lines. CTRL2493nXX was derived from the parental fibroblast line ND31845 that was biopsied from a healthy female at 71

years of age. CS2518nXX was derived from the parental fibroblast line ND30625 that was biopsied from a healthy male at 76 years of age. CS71iCTR-20nXX was derived from the parental fibroblast line ND29971 that was biopsied from a female at 61 years of age. For virus infections, NPCs plated on Matrigel-coated (Corning) multi-well plates or Huh7 cells were infected with ZIKV at a multiplicity of infection (MOI) of 0.1 or 1 for 2 h at 37°C. Infected cells were harvested at 48 hpi for all analyses.

Affinity purification, mass spectrometry, and AP-MS scoring. The ZIKV capsid open reading frames (ORFs) from the Ugandan 1947 strain MR 766 or the French Polynesian 2013 strain H/PF/2013 were cloned into pCDNA4_TO with a C-terminal 2xStrep II affinity tag for expression in human cells. The viral capsid proteins (three biological replicates), as well as GFP (two biological replicates) and empty vector (ten biological replicates) as negative controls, were expressed in HEK293T cells and affinity purifications were performed as previously described (50). Briefly, clarified lysates were incubated with Strep-Tactin Superflow (IBA) overnight at 4°C. Proteins were eluted with 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA containing 2.5 mM Desthiobiotin (IBA) for 30 min at 4°C. Lysates and affinity purified eluates were analyzed by western blot and silver stain PAGE to confirm expression and purification. Purified protein eluates were digested with trypsin for LC-MS/MS analysis. Samples were denatured and reduced in 2M urea, 10 mM NH_4HCO_3, 2 mM DTT for 30 min at 60°C, then alkylated with 2 mM iodoacetamide for 45 min at room temperature. Trypsin (Promega) was added at a 1:100 enzyme:substrate ratio and digested overnight at 37°C. Following digestion, samples were concentrated using C18 ZipTips (Millipore) according to the manufacturer's specifications. Peptides were resuspended in 15 µL of 4% formic acid and 3% ACN, and 1-2 µL of sample was loaded onto a 75 µm ID column packed with 25 cm of Reprosil C18 1.9 µm, 120Å particles (Dr. Maisch GmbH). Peptides were eluted into a Q-Exactive Plus (Thermo Fisher) mass

spectrometer by gradient elution delivered by an Easy1200 nLC system (Thermo Fisher). The gradient was from 4.5% to 32% acetonitrile over 53 min. All MS spectra were collected with oribitrap detection, while the 20 most abundant ions were fragmented by higher energy collisional dissociation (HCD) and detected in the orbitrap. All data was searched against the SwissProt Human protein sequences, combined with ZIKV sequences and GFP. Peptide and protein identification searches, as well as label-free quantitation, were performed using the MaxQuant data analysis algorithm and all peptide and protein identifications were filtered to a 1% false-discovery rate (51, 52). SAINTq (53) was used to calculate the probability of bait-prey interactions for both Ugandan ZIKV capsid and French Polynesian ZIKV capsid against the negative controls, including GFP and empty vector, with protein intensities as input values. We applied a combined threshold of probability of interaction (AvgP) greater than 0.90 and Bayesian False Discovery Rate of less than 0.05.

Quantitative real-time reverse transcription-PCR (qRT-PCR). Total cellular RNA was isolated from Huh7 cells and NPCs using the RNeasy Mini Kit (Qiagen). cDNA was synthesized with oligo$(dT)_{18}$ (ThermoFisher Scientific) primers, random hexamer (Life Technologies) primers, and AMV reverse transcriptase (Promega). The cDNA was then used in SYBR Green PCR Master Mix (ThermoFisher Scientific) according to manufacturer's instructions and analyzed by qPCR (Bio-Rad ABI 7900). The primers used for ASNS, CARS, SC35 1.7, GAPDH, HPRT1, LDHA, and 18S rRNA have been described previously(29). The additional primers used were ZIKV PRVABC59 forward primer 5'- GAG ACG AGA TGC GGT ACA GG -3', ZIKV PRVABC59 reverse primer 5'- CGA CCG TCA GTT GAA CTC CA -3', UPF1 forward primer 5'- CTG CAA CGG ACG TGG AAA TAC -3', UPF1 reverse primer 5'- ACA GCC GCA GTT GTA GCA C -3', DDIT3 forward primer 5'- CTG CTT CTC TGG CTT GGC TG -3', DDIT3 reverse primer 5'- GCT CTG GGA GGT GCT TGT GA -3', GADD45A forward primer 5'- GAG CTC CTG CTC TTG GAG AC -3', GADD45A reverse primer 5'- GCA GGA TCC TTC CAT TGA GA -3',

GADD45B forward primer 5'- TGA CAA CGA CAT CAA CAT C -3', and GADD45B reverse primer 5'- GTG ACC AGA GAC AAT GCA G -3'. Relative levels of each transcript were normalized by the delta threshold cycle method to the abundance of 18S rRNA or GAPDH, with mock-infected cells or vector-transfected cells set to 1.

Western blot analysis. Cells were lysed in RIPA lysis buffer (50 mM Tris-HCl, pH 8, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, supplemented with Halt™ protease inhibitor cocktail (ThermoFisher Scientific)) to obtain whole cell lysates or lysed using the NE-PER nuclear and cytoplasmic extraction kit (ThermoFisher Scientific) to obtain cytoplasmic and nuclear fractions. Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes (Bio-Rad). Blots were incubated with the indicated primary antibody: anti-UPF3B (ab134566, Abcam), anti-UPF1 (12040, Cell Signaling Technology, Inc.), anti-ZIKV Capsid (C) (GTX133304, GeneTex), anti-Flag (F7425, Sigma-Aldrich), anti-β-actin (A5316, Sigma-Aldrich), anti-ZIKV Envelope (E) (GTX133314, GeneTex), anti-SP1 (sc-14027, Santa Cruz Biotechnology), anti-GAPDH (5174, Cell Signaling Technology, Inc.), anti-Myc tag (ab9106, Abcam), anti-Strep tag (ab18422, Abcam), and anti-p62 (ab56416, Abcam). Proteins were visualized by chemiluminescent detection with ECL and ECL Hyperfilm (Amersham). Differences in band intensity were quantified by densitometry using ImageJ.

Immunoprecipitations. Cells were lysed in either RIPA lysis buffer or IP lysis buffer (150mM NaCl, 50mM Tris pH 7.4, 1mM EDTA, 0.5% NP-40 substitute, supplemented with Halt™ protease inhibitor cocktail (ThermoFisher Scientific)) at 4°C and passed through a G23 needle. Clarified lysates were immunoprecipitated with Flag M2 agarose (Sigma), anti-Myc tag (ab9106, Abcam), or normal rabbit IgG (sc-2027, Santa Cruz Biotechnology) overnight, washed in lysis buffer, and resuspended in Laemmli buffer for SDS-PAGE. Western blot analysis of immunoprecipitated proteins was performed as described above.

Immunofluorescence. Transfected Huh7-Lunet cells or infected NPCs were collected at 48 h and plated onto 22 × 22 mm #1.5 coverslips. Cells were fixed in 4% paraformaldehyde, permeabilized with 0.1% Triton X-100, and blocked in 3% bovine serum albumin. Cells were then immunostained with the indicated antibodies: anti-Strep Tag (Abcam, ab184224), anti-UPF1 (Abcam, ab109363), human anti-DENV mAb 1.6D (Sharon Isern and Scott Michael, Florida Gulf Coast University), which recognizes the ZIKV envelope protein, anti-dsRNA mAb J2 (SCICONS), and the appropriate fluorophore-conjugated secondary antibodies. Coverslips were mounted onto glass slides using Vectashield® Mounting Medium with DAPI (Vector Laboratories) and analyzed by fluorescence microscopy (Zeiss Axio Observer ZI) or confocal microscopy (Zeiss LSM 880). For acquiring high-resolution images, cells were imaged on the Zeiss LSM 880 with Airyscan using a 20x/0.8 or 63x/1.4 M27 oil immersion objective. A total of 15-20 (20x objective) or 60-80 (63x objective) Z-slices were acquired every 0.88 µm or 0.15 µm, respectively. The resulting Z-stack was reconstructed and rendered in 3D using Imaris software (Bitplane). Viral dsRNA foci were reconstructed via the Imaris spot detection function, which provided an analysis of total number and mean volume of foci within a cell, for images acquired using the 20x objective. Strep-tagged ZIKV capsid, UPF1, and dsRNA channels acquired using the 63x objective were reconstructed using the Imaris surfaces package. The Imaris colocalization function was used to determine overlap of fluorescence. Thresholding for background fluorescence was determined by the Imaris automatic thresholding tool that utilizes the Costes approach (54). The thresholded Mander's correlation coefficient (MCC) measures the fraction of voxels with fluorescence positive for one channel that also contain fluorescence from another channel. The MCC is typically more appropriate for analysis of three-dimensional colocalization (33).

Statistical analysis. Statistical differences between groups were analyzed using either a two-tailed unpaired Student's *t*-test or a two-tailed ratio paired Student's *t*-test, as stated in the figure legends. Hypergeometrical tests were used to calculate the probability of an overlap in gene dysregulation between ZIKV-infected NPCs and UPF1-depleted cells and to calculate the probability of ZIKV capsid bait-prey interactions. Data are represented as mean ± s.e.m. Statistical significance was defined as $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, and $****P \leq 0.0001$.

Figures

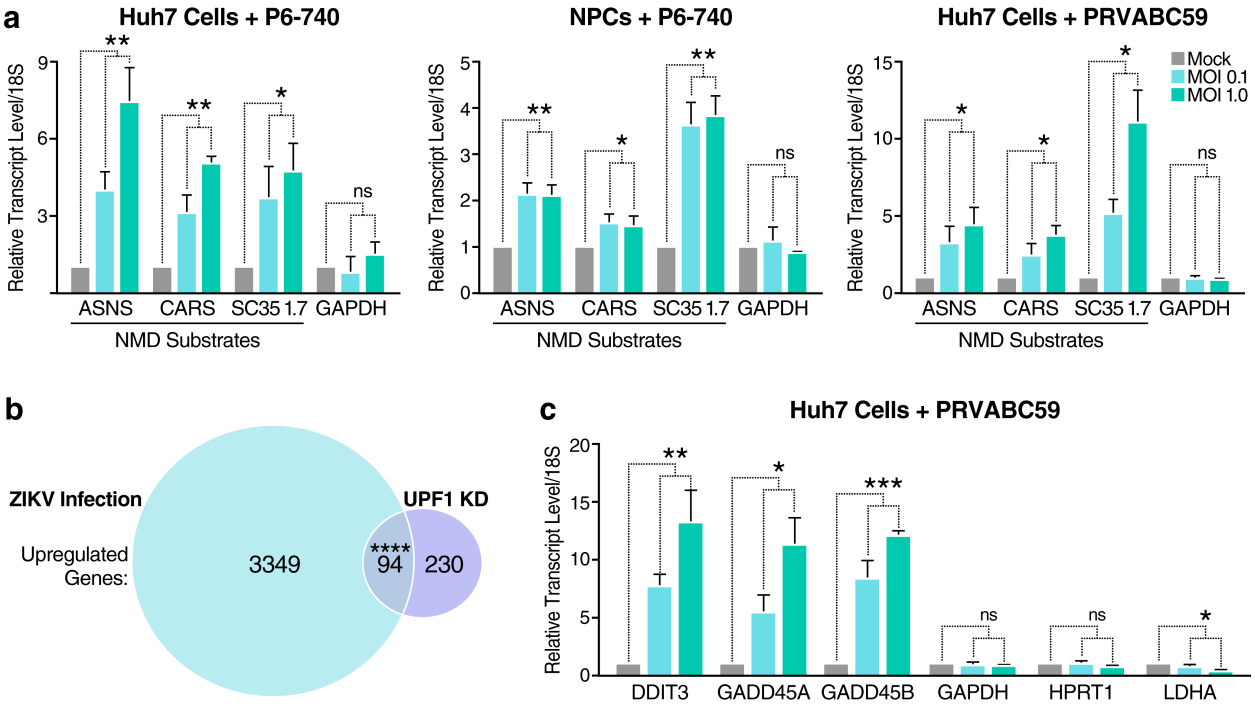


Figure 2.1. ZIKV infection disrupts the NMD pathway.

(a) Transcript levels of NMD substrates and housekeeping genes from Huh7 cells or NPCs mock-infected or infected with ZIKV strain P6-740 or the contemporary clinical isolate PRVABC59. Cells were infected at a multiplicity of infection (MOI) of 0.1 or 1 and harvested at 48 hpi. Data are represented as mean ± s.e.m. *P* values were calculated by unpaired Student's *t*-test. **P* ≤ 0.05; ***P* ≤ 0.01; ns, not significant. n= 3 independent experiments.
(b) Venn diagram showing overlap of significantly upregulated genes associated with ZIKV infection of NPCs and UPF1 knockdown in HeLa cells. RNA-Seq analyses of mock-infected or ZIKV-infected NPCs harvested at 56 hpi and control siRNA-treated or UPF1 siRNA-treated HeLa TO cells harvested at 72 h post-transfection (hpt). The GeneProf hypergeometric probability calculator was then used to generate a hypergeometric *P* value. *****P* ≤ 0.0001.
(c) Transcript levels of housekeeping genes and select genes involved in cell cycle growth arrest and apoptosis that were identified in (b). Huh7 cells were mock-infected or infected with ZIKV PRVABC59 at an MOI of 0.1 or 1 and harvested at 48 hpi. Data are represented as mean ± s.e.m. *P* values were calculated by unpaired Student's *t*-test. **P* ≤ 0.05; ***P* ≤ 0.01; ****P* ≤ 0.001; ns, not significant. n= 3 independent experiments.

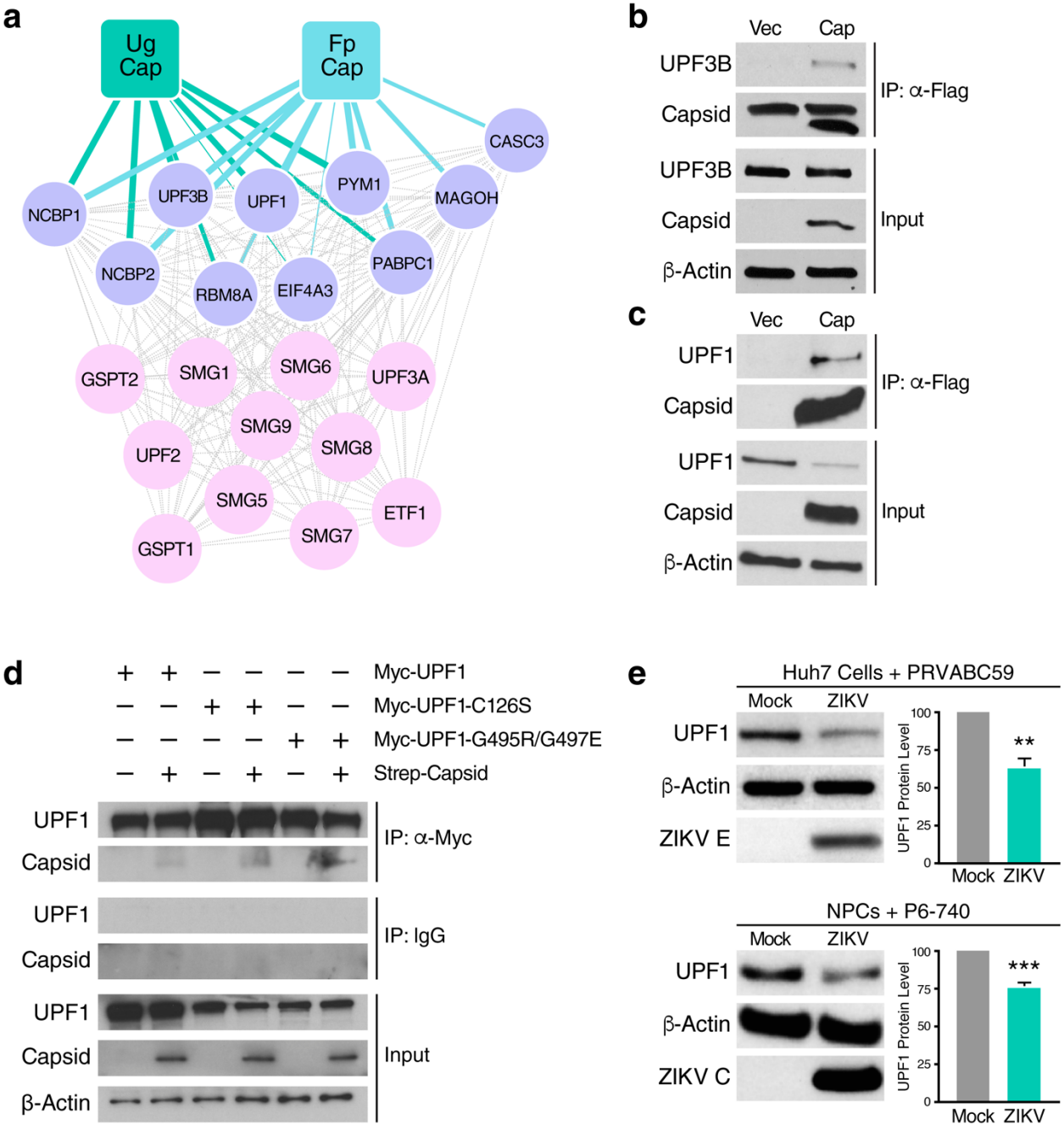


Figure 2.2. The capsid protein of ZIKV interacts with the NMD pathway.

(a) Ugandan ZIKV capsid (Ug Cap, MR 766) and French Polynesian ZIKV capsid (Fp Cap, H/PF/2013) PPI maps that show significant enrichment for host NMD-associated factors (purple), as identified by AP-MS (SAINTq probability score > 0.9 and FDR < 0.05). Ten interactions between Fp Cap and host NMD factors (hypergeometrical test, *P* value = 7.16×10^{-10}) and eight interactions between Ug Cap and host NMD factors (*P* value = 3.45×10^{-7}) were identified.
(b) Co-immunoprecipitation (co-IP) and western blot analysis of HEK293T cells transfected with vector or Flag-tagged ZIKV capsid (H/PF/2013, Asian lineage) and harvested at 48 hpt to

immunoprecipitate endogenous UPF3B. The upper band detected in the IP Capsid blot represents a non-specific artifact.
(c) Co-IP and western blot analysis of HEK293T cells transfected with vector or Flag-tagged ZIKV capsid and harvested at 48 hpt to immunoprecipitate endogenous UPF1.
(d) Myc-tag co-IP and western blot analysis of HEK293T cells transfected with Strep-tagged ZIKV capsid and Myc-UPF1 (WT), Myc-UPF1-C126S (RNA-binding mutant), or Myc-UPF1-G495R/G497E (ATPase/helicase mutant) and harvested at 48 hpt to immunoprecipitate ZIKV capsid.
(e) Western blot analysis of UPF1 levels in mock-infected and ZIKV-infected (PRVABC59, MOI of 1) Huh7 cells or mock-infected and ZIKV-infected (P6-740, MOI of 1) NPCs harvested at 48 hpi, with β-actin and ZIKV envelope (ZIKV E) or ZIKV capsid (ZIKV C) protein serving as loading and infection controls, respectively. Densitometric analyses were performed using ImageJ to quantify relative band intensities. Data are represented as mean ± s.e.m. *P* values were calculated by unpaired Student's *t*-test. $^{**}P \leq 0.01$; $^{***}P \leq 0.001$. n= 3 independent experiments.

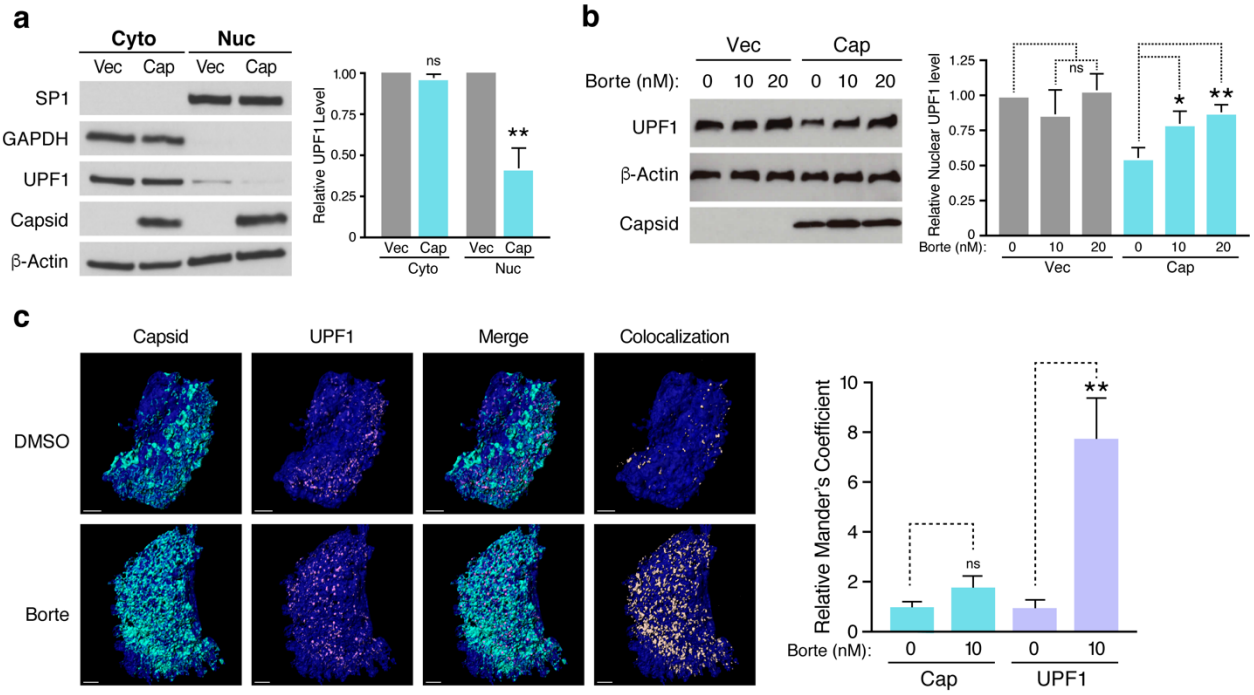


Figure 2.3. ZIKV capsid degrades UPF1, the master regulator of NMD, via a proteasome-dependent mechanism.

(a) Western blot analysis of UPF1 levels in subcellular fractionated HEK293T cells transfected with vector or Flag-tagged ZIKV capsid (H/PF/2013, Asian lineage) for 48 h. GAPDH was used as a cytoplasmic marker and SP1 as a nuclear marker to ensure optimal fractionation. Densitometric analyses were performed using ImageJ to quantify relative band intensities. Data are represented as mean ± s.e.m. *P* values were calculated by unpaired Student's *t*-test. $^{**}P \leq 0.01$; ns, not significant. n= 3 independent experiments.
(b) Western blot analysis of nuclear UPF1 levels in fractionated HEK293T cells transfected with vector or Flag-tagged ZIKV capsid for 48 h. Cells were treated with DMSO or increasing concentrations of the proteasome inhibitor bortezomib for 24 h before harvest. Densitometric analyses were performed using ImageJ to quantify relative band intensities. Data are represented as mean ± s.e.m. *P* values were calculated by one-way ANOVA with multiple comparisons. $^{*}P \leq 0.05$; ns, not significant. n= 3 independent experiments.
(c) Representative 3D confocal microscopy images of the nuclei of Huh7-Lunet cells transfected with Strep-tagged ZIKV capsid. Cells were treated at 24 hpt with DMSO or 10 nM bortezomib and processed for immunostaining at 48 hpt with antibodies against Strep-tag (turquoise) and endogenous UPF1 (purple). DAPI (blue) was used to stain and define the nuclei. Each channel was reconstructed digitally for visualization of the 3D colocalization. The thresholded Mander's correlation coefficients were determined and *P* value was calculated by unpaired Student's *t*-test. $^{**}P \leq 0.01$. n = 8 cells per condition. Scale bar represents 3 μm.

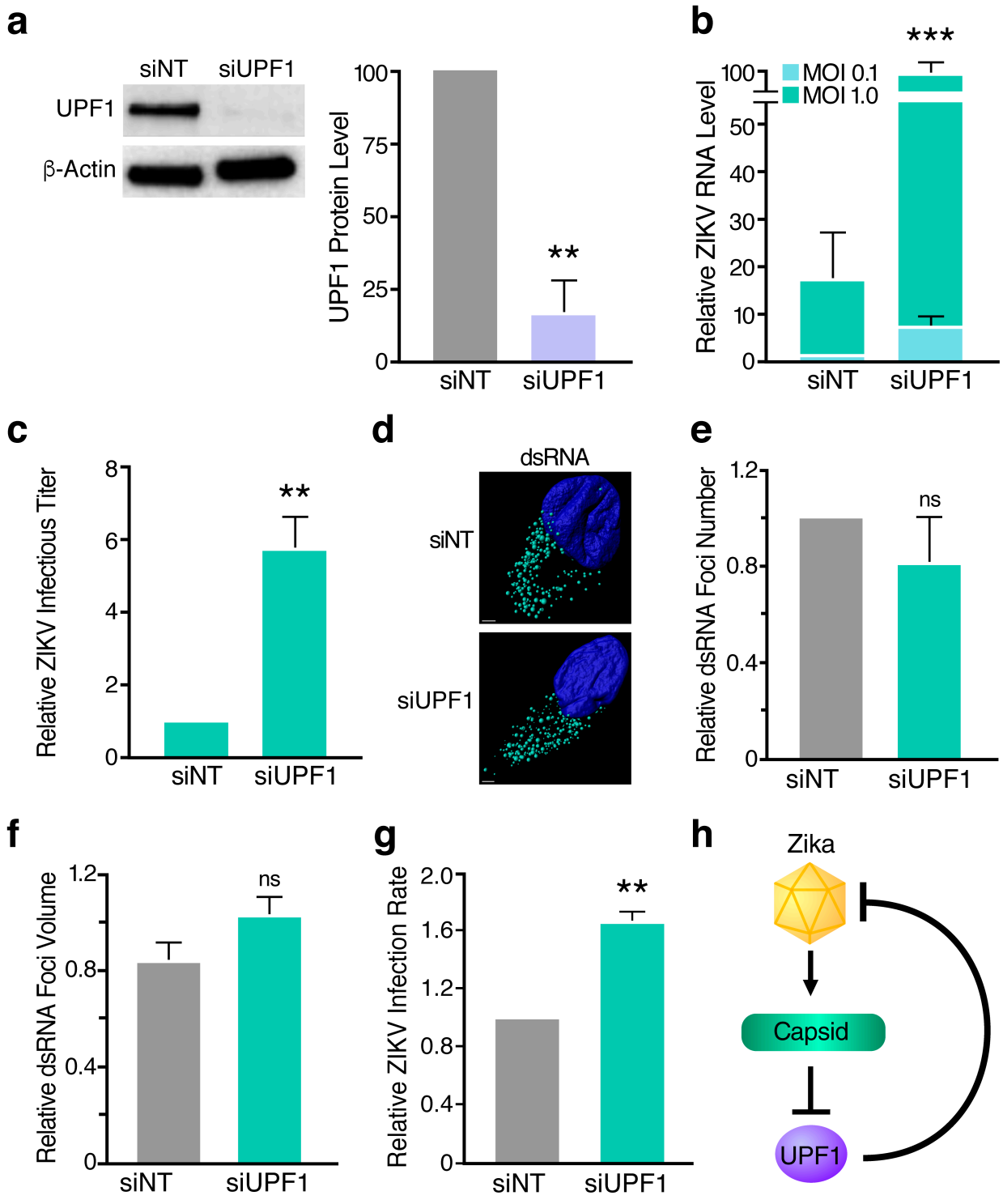
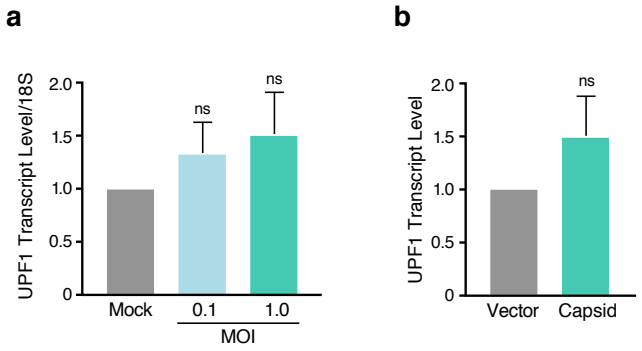


Figure 2.4. UPF1 knockdown enhances the permissivity of NPCs to ZIKV infection.

(a) Western blot analysis of UPF1 levels in NPCs transfected with non-targeting siRNA (siNT) or a pool of UPF1-specific siRNAs (siUPF1) at 96 hpt. Densitometric analyses were performed using ImageJ to quantify relative band intensities. Data are represented as mean ± s.e.m. *P*

value was calculated by unpaired Student's *t*-test. $^{**}P \leq 0.01$. n= 3 independent experiments using one NPC line.
(b) ZIKV RNA levels in siNT-treated or siUPF1-treated NPCs infected with ZIKV strain PRVABC59 at an MOI of 0.1 or 1 and harvested at 48 hpi. Data are represented as mean ± s.e.m. *P* value was calculated by two-tailed ratio paired Student's *t*-test. $^{***}P \leq 0.001$. n= 3 independent experiments using one NPC line.
(c) Released infectious virus from siNT-treated or siUPF1-treated, ZIKV-infected (MOI of 1) NPCs harvested at 48 hpi. Data are represented as mean ± s.e.m. *P* value was calculated by unpaired Student's *t*-test. $^{**}P \leq 0.01$. n= 3 independent experiments using one NPC line.
(d) Representative confocal microscopy images of a ZIKV-infected, siNT-treated NPC or a ZIKV-infected, siUPF1-treated NPC with the nuclei stained with DAPI (blue) and ZIKV dsRNA foci stained with the anti-dsRNA mAb J2 (teal). 3D image rendering and reconstructed dsRNA foci were produced using the Imaris spot detection function. Scale bar represents 2 μm.
(e) Number of dsRNA foci were averaged for each cell. Data are represented as mean ± s.e.m. *P* value was calculated by two-tailed ratio paired Student's *t*-test. ns, not significant. n = 3 independent experiments using two NPC lines, with 3-10 cells analyzed per condition for each experiment.
(f) Measurement of dsRNA foci volume were averaged for each cell. Data are represented as mean ± s.e.m. *P* value was calculated by two-tailed ratio paired Student's *t*-test. ns, not significant. n = 3 independent experiments using two NPC lines, with 3-10 cells analyzed per condition for each experiment.
(g) Infection rates of siNT-treated or siUPF1-treated, ZIKV-infected (MOI of 1) NPCs measured at 48 hpi. Fixed cells were subjected to the anti-DENV mAb 1.6D, which also recognizes the ZIKV envelope protein (55). Data are represented as mean ± s.e.m. *P* value was calculated by two-tailed ratio paired Student's *t*-test. $^{**}P \leq 0.01$. n = 3 independent experiments using two NPC lines.
(h) Model of the interaction between the capsid protein of ZIKV and UPF1 of the NMD pathway.

Supplementary Figure 1

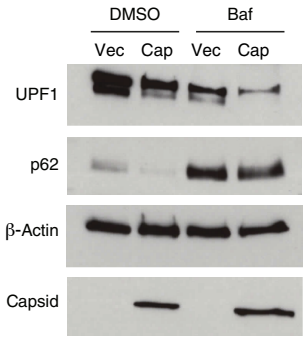


18.010 D4 M.Ott

Supplemental Figure 2.1. UPF1 is not transcriptionally downregulated during ZIKV infection or following ZIKV capsid overexpression.

(a) UPF1 transcript levels from Huh7 cells mock-infected or infected with ZIKV strain PRVABC59 at an MOI of 0.1 or 1 and harvested at 48 hpi. Data are represented as mean ± s.e.m. *P* values were calculated by unpaired Student's *t*-test. ns, not significant. n= 3 independent experiments.
(b) UPF1 transcript levels from HEK293T cells transfected with vector or Strep-tagged ZIKV capsid (H/PF/2013, Asian lineage) and harvested at 48 hpt. Data are represented as mean ± s.e.m. *P* values were calculated by unpaired Student's *t*-test. ns, not significant. n= 3 independent experiments.

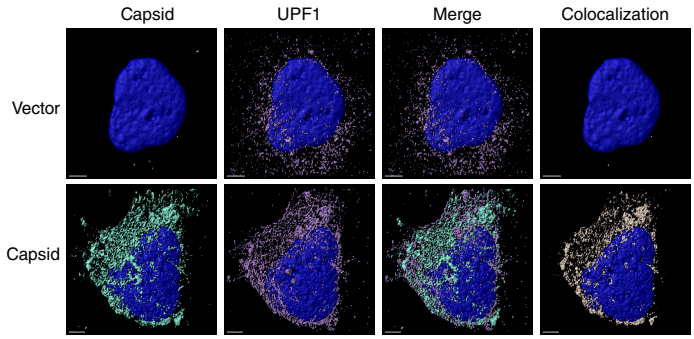
Supplementary Figure 2



Supplemental Figure 2.2. ZIKV capsid-induced UPF1 degradation is not dependent on autophagy.

Western blot analysis of nuclear UPF1 levels in fractionated HEK293T cells transfected with vector or Flag-tagged ZIKV capsid for 48 h. Cells were treated with DMSO or the autophagy inhibitor bafilomycin A1 (Baf) (10 nM) for 24 h before harvest. Levels of p62, which is degraded by autophagy, were monitored to confirm autophagic inhibition following bafilomycin A1 treatment.

Supplementary Figure 3



Supplemental Figure 2.3. ZIKV capsid colocalizes with endogenous UPF1.

Representative 3D confocal microscopy images of Huh7-Lunet cells transfected with vector or Strep-tagged ZIKV capsid. Cells were processed for immunostaining at 48 hpt and probed with antibodies against Strep-tag (turquoise) and endogenous UPF1 (purple). DAPI (blue) was used to stain the nuclei. Each channel was reconstructed digitally for visualization of the 3D colocalization. The thresholded Mander's correlation coefficient for ZIKV capsid was 0.57 (n = 17), indicating that approximately 57% of the voxels positive for capsid fluorescence were also positive for UPF1 fluorescence. Scale bar represents 5 μm.

References

1. Fauci AS, Morens DM. 2016. Zika Virus in the Americas--Yet Another Arbovirus Threat. N Engl J Med 374:601-604. 10.1056/NEJMp1600297.
2. Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JLM, Guimarães KP, Benazzato C, Almeida N, Pignatari GC, Romero S, Polonio CM, Cunha I, Freitas CL, Brandão WN, Rossato C, Andrade DG, Faria DdP, Garcez AT, Buchpigel CA, Braconi CT, Mendes E, Sall AA, Zanotto, Paolo M. de A., Peron JPS, Muotri AR, Beltrão-Braga PCB. 2016. The Brazilian Zika virus strain causes birth defects in experimental models. Nature 534:267-271. 10.1038/nature18296.
3. Shao Q, Herrlinger S, Yang S, Lai F, Moore JM, Brindley MA, Chen J. 2016. Zika virus infection disrupts neurovascular development and results in postnatal microcephaly with brain damage. Development 143:4127-4136. 10.1242/dev.143768.
4. Souza BSF, Sampaio GLA, Pereira CS, Campos GS, Sardi SI, Freitas LAR, Figueira CP, Paredes BD, Nonaka CKV, Azevedo CM, Rocha VPC, Bandeira AC, Mendez-Otero R, Dos Santos RR, Soares MBP. 2016. Zika virus infection induces mitosis abnormalities and apoptotic cell death of human neural progenitor cells. Scientific reports 6:39775. 10.1038/srep39775.
5. Li C, Xu D, Xu Z, Ye Q, Hong S, Jiang Y, Liu X, Zhang N, Shi L, Qin C. 2016. Zika Virus Disrupts Neural Progenitor Development and Leads to Microcephaly in Mice. Cell Stem Cell 19:120-126. 10.1016/j.stem.2016.04.017.
6. Tang H, Hammack C, Ogden S, Wen Z, Qian X, Li Y, Yao B, Shin J, Zhang F, Lee E, Christian K, Didier R, Jin P, Song H, Ming G. 2016. Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth. Cell Stem Cell 18:587-590. 10.1016/j.stem.2016.02.016.

7. Sonja A Rasmussen, Denise J Jamieson, Margaret A Honein, Lyle R Petersen. 2016. Zika Virus and Birth Defects--Reviewing the Evidence for Causality. The New England journal of medicine 374:1981-1987. 10.1056/NEJMsr1604338.
8. Harris E, Holden KL, Edgil D, Polacek C, Clyde K. 2006. Molecular biology of flaviviruses. Novartis Found Symp 277:253.
9. Lindenbach BD, Rice CM. 2003. Molecular biology of flaviviruses. Adv Virus Res 59:23-61.
10. Oliveira ERA, Mohana-Borges R, de Alencastro RB, Horta BAC. 2017. The flavivirus capsid protein: Structure, function and perspectives towards drug design. Virus Res 227:115-123. 10.1016/j.virusres.2016.10.005.
11. den Boon JA, Diaz A, Ahlquist P. 2010. Cytoplasmic Viral Replication Complexes. Cell Host Microbe 8:77-85. 10.1016/j.chom.2010.06.010.
12. Colpitts TM, Barthel S, Wang P, Fikrig E. 2011. Dengue virus capsid protein binds core histones and inhibits nucleosome formation in human liver cells. PLoS ONE 6:e24365. 10.1371/journal.pone.0024365.
13. Lukasz P Slomnicki, Dong-Hoon Chung, Austin Parker, Taylor Hermann, Nolan L Boyd, Michal Hetman. 2017. Ribosomal stress and Tp53-mediated neuronal apoptosis in response to capsid protein of the Zika virus. Scientific Reports (Nature Publisher Group) 7:1-15. 10.1038/s41598-017-16952-8.
14. Tsuda Y, Mori Y, Abe T, Yamashita T, Okamoto T, Ichimura T, Moriishi K, Matsuura Y. 2006. Nucleolar protein B23 interacts with Japanese encephalitis virus core protein and participates in viral replication. Microbiol Immunol 50:225-234.
15. Xu Z, Hobman TC. 2012. The helicase activity of DDX56 is required for its role in assembly of infectious West Nile virus particles. Virology 433:226-235. 10.1016/j.virol.2012.08.011.
16. Rawlinson SM, Moseley GW. 2015. The nucleolar interface of RNA viruses. Cell Microbiol 17:1108-1120. 10.1111/cmi.12465.

17. Peccarelli M, Kebaara BW. 2014. Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. Eukaryotic cell 13:1126-1135. 10.1128/EC.00090-14.

18. Hug N, Longman D, Cáceres JF. 2016. Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic Acids Res 44:1483-1495. 10.1093/nar/gkw010.

19. Han X, Wei Y, Wang H, Wang F, Ju Z, Li T. 2018. Nonsense-mediated mRNA decay: a 'nonsense' pathway makes sense in stem cell biology. Nucleic acids research 46:1038-1051. 10.1093/nar/gkx1272.

20. Silver DL, Watkins-Chow DE, Schreck KC, Pierfelice TJ, Larson DM, Burnetti AJ, Liaw H, Myung K, Walsh CA, Gaiano N, Pavan WJ. 2010. The exon junction complex component Magoh controls brain size by regulating neural stem cell division. Nat Neurosci 13:551-558. 10.1038/nn.2527.

21. Mao H, Pilaz L, McMahon JJ, Golzio C, Wu D, Shi L, Katsanis N, Silver DL. 2015. Rbm8a haploinsufficiency disrupts embryonic cortical development resulting in microcephaly. J Neurosci 35:7003-7018. 10.1523/JNEUROSCI.0018-15.2015.

22. Mao H, McMahon JJ, Tsai Y, Wang Z, Silver DL. 2016. Haploinsufficiency for Core Exon Junction Complex Components Disrupts Embryonic Neurogenesis and Causes p53-Mediated Microcephaly. PLoS Genetics 12:e1006282. 10.1371/journal.pgen.1006282.

23. Kurosaki T, Maquat LE. 2016. Nonsense-mediated mRNA decay in humans at a glance. J Cell Sci 129:461-467. 10.1242/jcs.181008.

24. Serquiña AKP, Das SR, Popova E, Ojelabi OA, Roy CK, Göttlinger HG. 2013. UPF1 is crucial for the infectivity of human immunodeficiency virus type 1 progeny virions. J Virol 87:8853-8861. 10.1128/JVI.00925-13.

25. Jerome M Molleston, Sara Cherry. 2017. Attacked from All Sides: RNA Decay in Antiviral Defense. Viruses 9:2. 10.3390/v9010002.

26. Balistreri G, Bognanni C, Mühlemann O. 2017. Virus Escape and Manipulation of Cellular Nonsense-Mediated mRNA Decay. Viruses 9:. 10.3390/v9010024.

27. Rigby RE, Rehwinkel J. 2015. RNA degradation in antiviral immunity and autoimmunity. Trends in Immunology 36:179-188. 10.1016/j.it.2015.02.001.

28. Balistreri G, Horvath P, Schweingruber C, Zünd D, McInerney G, Merits A, Mühlemann O, Azzalin C, Helenius A. 2014. The host nonsense-mediated mRNA decay pathway restricts Mammalian RNA virus replication. Cell Host Microbe 16:403-411. 10.1016/j.chom.2014.08.007.

29. Ramage HR, Kumar GR, Verschueren E, Johnson JR, Von Dollen J, Johnson T, Newton B, Shah P, Horner J, Krogan NJ, Ott M. 2015. A combined proteomics/genomics approach links hepatitis C virus infection with nonsense-mediated mRNA decay. Mol Cell 57:329-340. 10.1016/j.molcel.2014.12.028.

30. Tani H, Imamachi N, Salam KA, Mizutani R, Ijiri K, Irie T, Yada T, Suzuki Y, Akimitsu N. 2012. Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. RNA Biology 9:1370-1379. 10.4161/rna.22360.

31. Weischenfeldt J, Damgaard I, Bryder D, Theilgaard-Mönch K, Thoren LA, Nielsen FC, Jacobsen SEW, Nerlov C, Porse BT. 2008. NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. Genes Dev 22:1381-1396. 10.1101/gad.468808.

32. Li T, Shi Y, Wang P, Guachalla LM, Sun B, Joerss T, Chen Y, Groth M, Krueger A, Platzer M, Yang Y, Rudolph KL, Wang Z. 2015. Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. EMBO J 34:1630-1647. 10.15252/embj.201489947.

33. Brazão TF, Demmers J, van IJcken W, Strouboulis J, Fornerod M, Romão L, Grosveld FG. 2012. A new function of ROD1 in nonsense-mediated mRNA decay. FEBS Lett 586:1101-1110. 10.1016/j.febslet.2012.03.015.

34. Chan W, Huang L, Gudikote JP, Chang Y, Imam JS, MacLean JA, Wilkinson MF. 2007. An alternative branch of the nonsense-mediated decay pathway. EMBO J 26:1820-1830. 10.1038/sj.emboj.7601628.

35. Nelson JO, Moore KA, Chapin A, Hollien J, Metzstein MM. 2016. Degradation of Gadd45 mRNA by nonsense-mediated decay is essential for viability. Elife 5:. 10.7554/eLife.12876.

36. Jauhiainen A, Thomsen C, Strömbom L, Grundevik P, Andersson C, Danielsson A, Andersson MK, Nerman O, Rörkvist L, Ståhlberg A, Åman P. 2012. Distinct cytoplasmic and nuclear functions of the stress induced protein DDIT3/CHOP/GADD153. PLoS ONE 7:e33208. 10.1371/journal.pone.0033208.

37. Salvador JM, Brown-Clay JD, Fornace AJ. 2013. Gadd45 in stress signaling, cell cycle control, and apoptosis. Adv Exp Med Biol 793:1-19. 10.1007/978-1-4614-8289-5_1.

38. Gehring NH, Lamprinaki S, Kulozik AE, Hentze MW. 2009. Disassembly of exon junction complexes by PYM. Cell 137:536-548. 10.1016/j.cell.2009.02.042.

39. Chamieh H, Ballut L, Bonneau F, Le Hir H. 2008. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nat Struct Mol Biol 15:85-93. 10.1038/nsmb1330.

40. Varsally W, Brogna S. 2012. UPF1 involvement in nuclear functions. Biochem Soc Trans 40:778-783. 10.1042/BST20120052.

41. Klema VJ, Padmanabhan R, Choi KH. 2015. Flaviviral Replication Complex: Coordination between RNA Synthesis and 5′-RNA Capping. Viruses 7:4640-4656. 10.3390/v7082837.

42. Maisel M, Herr A, Milosevic J, Hermann A, Habisch H, Schwarz S, Kirsch M, Antoniadis G, Brenner R, Hallmeyer-Elgner S, Lerche H, Schwarz J, Storch A. 2007. Transcription profiling of adult and fetal human neuroprogenitors identifies divergent paths to maintain the neuroprogenitor cell state. Stem Cells 25:1231-1240. 10.1634/stemcells.2006-0617.

43. Li H, Saucedo-Cuevas L, Regla-Nava J, Chai G, Sheets N, Tang W, Terskikh A, Shresta S, Gleeson J. 2016. Zika Virus Infects Neural Progenitors in the Adult Mouse Brain and Alters Proliferation. Cell Stem Cell 19:593-598. 10.1016/j.stem.2016.08.005.

44. Nickless A, Bailis JM, You Z. 2017. Control of gene expression through the nonsense-mediated RNA decay pathway. Cell & Bioscience 7:1-12. 10.1186/s13578-017-0153-7.

45. Cortese M, Goellner S, Acosta EG, Neufeldt CJ, Oleksiuk O, Lampe M, Haselmann U, Funaya C, Schieber N, Ronchi P, Schorb M, Pruunsild P, Schwab Y, Chatel-Chaix L, Ruggieri A, Bartenschlager R. 2017. Ultrastructural Characterization of Zika Virus Replication Factories. Cell Rep 18:2113-2123. 10.1016/j.celrep.2017.02.014.

46. Grant A, Ponia SS, Tripathi S, Balasubramaniam V, Miorin L, Sourisseau M, Schwarz MC, Sánchez-Seco MP, Evans MJ, Best SM, García-Sastre A. 2016. Zika Virus Targets Human STAT2 to Inhibit Type I Interferon Signaling. Cell Host Microbe 19:882-890. 10.1016/j.chom.2016.05.009.

47. von Mikecz A. 2006. The nuclear ubiquitin-proteasome system. J Cell Sci 119:1977-1984. 10.1242/jcs.03008.

48. Freire JM, Veiga AS, Conceição TM, Kowalczyk W, Mohana-Borges R, Andreu D, Santos NC, Da Poian AT, Castanho, Miguel A. R. B. 2013. Intracellular nucleic acid delivery by the supercharged dengue virus capsid protein. PLoS ONE 8:e81450. 10.1371/journal.pone.0081450.

49. HD iPSC Consortium. 2017. Developmental alterations in Huntington's disease neural cells and pharmacological rescue in cells and mice. Nat Neurosci 20:648-660. 10.1038/nn.4532.

50. Jäger S, Gulbahce N, Cimermancic P, Kane J, He N, Chou S, D'Orso I, Fernandes J, Jang G, Frankel AD, Alber T, Zhou Q, Krogan NJ. 2011. Purification and characterization of HIV-human protein complexes. Methods 53:13-19. 10.1016/j.ymeth.2010.08.007.

51. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. 2014. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13:2513-2526. 10.1074/mcp.M113.031591.

52. Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367-1372. 10.1038/nbt.1511.

53. Teo G, Koh H, Fermin D, Lambert J, Knight JDR, Gingras A, Choi H. 2016. SAINTq: Scoring protein-protein interactions in affinity purification - mass spectrometry experiments with fragment or peptide intensity data. Proteomics 16:2238-2245. 10.1002/pmic.201500499.

54. Costes SV, Daelemans D, Cho EH, Dobbin Z, Pavlakis G, Lockett S. 2004. Automatic and quantitative measurement of protein-protein colocalization in live cells. Biophys J 86:3993-4003. 10.1529/biophysj.103.038422.

55. Paul LM, Carlin ER, Jenkins MM, Tan AL, Barcellona CM, Nicholson CO, Michael SF, Isern S. 2016. Dengue virus antibodies enhance Zika virus infection. Clinical & Translational Immunology 5:n/a. 10.1038/cti.2016.72.

Chapter 3

ZIKV prevents host mRNA export from the nucleus by disrupting UPF1 function

A version of this paper is published:

Leon, K. E. et al. Zika Virus Infection Prevents Host mRNA Nuclear Export by Disrupting UPF1 Function. bioRxiv, 2020.2012.2003.410837, doi:10.1101/2020.12.03.410837 (2020).

Introduction

Zika virus (ZIKV) is a mosquito-borne, enveloped virus with a positive sense, single stranded RNA genome (*1*). ZIKV is a member of the *Flaviviridae* family that includes Dengue virus (DENV), West Nile virus (WNV) and Hepatitis C virus (HCV). In 2015, an outbreak of ZIKV in Brazil was linked to a dramatic increase in the number of infants born with microcephaly (*2, 3*). It was subsequently shown that *in utero* ZIKV infects neural progenitor cells (NPCs), resulting in neurodevelopmental delays that ultimately cause a range of birth defects including microcephaly, ocular damage, and contractures, collectively known as congenital Zika syndrome (*4-8*). NPCs are critical for brain development as they differentiate into the glial and neuronal cells that compose the majority of the brain parenchyma (*9*). Proposed molecular mechanisms by which ZIKV disrupts NPC function and differentiation include perturbation of the ANKLE2/VRK1, centrosomal organization, autophagy, apoptosis and unfolded protein response pathways (*10-14*). However, it remains unknown how ZIKV manipulates multiple cellular pathways at once to cause such widespread developmental reprogramming.

We and others previously showed that ZIKV infection suppressed the host nonsense-mediated mRNA decay (NMD) pathway (*15, 16*). NMD is an RNA quality control mechanism that targets faulty host transcripts for degradation and acts as an antiviral pathway on many viral species, particularly single-stranded RNA viruses (*17, 18*). Different models of the NMD pathway have

been described, but the exon-junction complex (EJC)-mediated NMD pathway is the best defined (*19, 20*). This pathway is classically triggered by premature termination codons in mRNAs, but can be also be initiated by normal 3'UTRs (*21*).

EJCs are normally loaded onto mRNAs in the nucleus (*22*); the RNAs are then exported to the cytoplasm to undergo translation, leading to proper displacement of the EJCs (*23*). Non-displaced EJCs recruit the master regulator and helicase up-frameshift protein 1 (UPF1), which induces a recruitment cascade activating NMD-mediated degradation of the faulty transcript (*24, 25*). However, the NMD pathway and UPF1 also target non-faulty mRNAs, especially those with long and GC-rich 3' untranslated regions (UTRs), a feature found in many viral RNAs (*26-29*). Because knockdown of UPF1 enhances replication of many RNA viruses including ZIKV, WNV, DENV, Rous Sarcoma virus, Potato virus X, Pea Enation Mosaic virus 2, and Turnip Crinkle virus, NMD is recognized as a *bone fide* antiviral restriction pathway, often targeted by viruses including ZIKV for inactivation (*17*).

ZIKV inactivation of the antiviral activity of NMD through interactions between the ZIKV capsid protein and UPF1 has been previously described (*15, 16*). Similar interactions of capsid proteins with the NMD pathway were subsequently confirmed in other flaviviruses (*16*). However, ZIKV is unique in that infection or ZIKV capsid expression selectively downregulated UPF1 expression in the host nucleus (*15*), which is unusual as viral RNA replication occurs exclusively in the host cytoplasm. Thus, it remains unclear why nuclear UPF1 is targeted during ZIKV infection.

Here we show that UPF1 occupancy of host transcripts is globally decreased during infection with many of these transcripts downregulated in the cytoplasm. ZIKV infection or UPF1 knockdown resulted in polyadenylated transcript accumulation in the nucleus, causing decreased protein levels of FREM2, a UPF1 target, and consequent perturbation of NPC differentiation. We propose

that by targeting nuclear UPF1 and trapping host mRNAs in the nucleus, ZIKV has evolved a mechanism to “shut off” host mRNA function while promoting translation of its own proteins. This mechanism describes a new central role for UPF1 in mRNA export connected to many cellular pathways associated with neurodevelopment.

Results

ZIKV infection decreases UPF1 interaction with the 3’UTR of host transcripts

While ZIKV inhibits NMD by capsid-mediated degradation of UPF1, the functional impact of UPF1 loss is still unknown. We first aimed to define the UPF1-RNA binding landscape by performing infrared crosslinking immunoprecipitation and RNA sequencing (irCLIP-Seq) in human induced pluripotent stem cell-derived NPCs (*30*). UPF1 is an RNA helicase and ATPase found associated with many transcripts (*31*). To identify differences in UPF1 binding during infection, NPCs were mock-infected or infected with ZIKV (isolate PRVABC59) at an MOI of 1 for 48 hours, followed by UV-crosslinking of transcripts with proteins and immunoprecipitation of UPF1 (**Figure 3.1A**). RNA-protein complexes were separated by SDS-PAGE, and mass spectrometry was performed on the excised bands to confirm UPF1 enrichment (**Figure 3.S1A**) before RNA was extracted and submitted to next-generation sequencing (**Figure 3.1A**).

In mock-infected NPCs, UPF1 bound 6778 transcripts, predominantly at the 3’UTR, as previously reported (*27, 29*). In ZIKV-infected NPCs, this number decreased to 4557 transcripts, with a marked decrease in binding at the 3’UTR (**Figure 3.1B**). Overall, a greater than 4-fold decrease in binding was observed in 3% of UPF1 target transcripts. Notably, known NMD-resistant transcripts such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH) maintained UPF1 interaction, highlighting that ZIKV infection did not impact all transcripts uniformly (**Figure 3.1C**).

In infected samples, UPF1 also interacted with ZIKV RNA, again mostly within the 3'UTR, supporting the model that UPF1 targets the viral RNA to restrict infection (**Figure 3.1D**). To determine if viral RNA sequestered UPF1 away from host RNA, leading to decreased UPF1 occupancy during infection, we quantified the percent of UPF1-bound reads mapping to host vs. viral RNAs. The majority of sequences isolated with UPF1 pulldown mapped to the host transcriptome; viral reads only accounted for 0.33% of all mapped reads, excluding a "sponge" effect of viral RNA for UPF1 **(Figure 3.1E).**

Selective reduction of UPF1 target transcripts in the cytoplasm of infected NPCs

During ZIKV infection, we saw a widespread decrease in UPF1 occupancy of host transcripts. To understand the consequences of the loss of UPF1, we sought to define the changes in the abundance of transcripts identified in the irCLIP-Seq experiment. We performed whole transcriptome sequencing of ZIKV-infected NPCs followed by differential expression analysis. As expected, genes associated with the interferon response were predominantly upregulated upon infection, consistent with previous RNA-seq studies of ZIKV-infected NPCs (*32, 33*) (**Figure 3.2A, Supplemental Figure 3.2C**). The NMD pathway and UPF1 are associated with RNA degradation, and we previously found that ZIKV infection upregulated select NMD target mRNAs (*15*). We thus expected that transcripts that had lost UPF1 interaction upon infection would be selectively upregulated. However, the abundance of transcripts which lost UPF1 occupancy upon infection was not significantly changed, indicating that the majority of UPF1 targets were not automatic subjects of NMD (**Figure 3.2B**). Metascape analysis of transcripts with stable abundance but loss of UPF1 binding showed a characteristic enrichment for neural and neurodevelopmental functions, underscoring the relevance of UPF1 binding for proper neurodevelopment. Axon guidance, which is essential for appropriate establishment of neural circuits in the brain, represented the top pathway identified (**Figure 3.2C**) (*34*).

Since transcript abundance was not affected by UPF1 occupancy, we next hypothesized the loss of UPF1 binding may instead impact the subcellular localization of cellular mRNAs. This was based on the knowledge that UPF1 translocates between the nucleus and cytoplasm (*35*), and our previous observation that UPF1 is selectively downregulated in the cell nucleus upon infection (*15*). To test this hypothesis, we fractionated ZIKV-infected NPCs into cytoplasmic and nuclear compartments and performed RNA sequencing and differential expression analysis with each fraction. We confirmed successful fractionation by comparing the $\log_2$ fold change of known cytoplasmic (*GAPDH*) and nuclear (*ANRIL*) transcripts, which showed 3.2-fold enrichment of *GAPDH* in the cytoplasmic fraction and a 2.3-fold increase of *ANRIL* in the nuclear fraction, as expected (**Supplemental Figure 3.2A**). Upon ZIKV infection, we identified 585 and 312 significantly upregulated mRNAs in the cytoplasm and nucleus, respectively (**Figure 3.2D**). These upregulated transcripts were predominantly interferon response genes (**Supplemental Figure 3.2D**). In contrast, downregulated genes were overwhelmingly cytoplasmic, with 912 genes downregulated in the cytoplasm compared to 22 in the nuclear fraction (**Figure 3.2E**). A significant overlap was found between the 912 transcripts with reduced cytoplasmic expression and the 331 transcripts that lost UPF1 occupancy without changing overall abundance **(Figure 3.2F)**. In contrast, considerably less overlap existed with transcripts that were upregulated in the cytoplasm and had lost UPF1 binding (**Supplemental Figure 3.2B**). Moreover, of the top 53 genes with the most significant decrease in UPF1 interaction upon infection, all but 6 were exclusively downregulated in the cytoplasm (**Figure 3.2G**). These results show a selective loss of UPF1 target mRNAs in the cytoplasm of infected cells. As total abundance of these transcripts was unchanged, we considered the possibility that they accumulate in the nucleus upon ZIKV infection.

ZIKV infection, ZIKV capsid expression or UPF1 knockdown results in accumulation of host transcripts in the nucleus

To test the hypothesis that transcripts were selectively accumulating in the nucleus in ZIKV-infected cells, we analyzed global transcript localization using RNAscope analysis with a polyA tail probe. We first optimized the technology in Huh7-Lunet cells, a hepatoma cell line frequently used to study flavivirus infection (*36*), and NPCs, both infected with ZIKV at an MOI of 1. In both cell types, polyadenylated RNA fluorescence was significantly increased in the nucleus upon infection, supporting the model that mRNA location rather than abundance is perturbed in ZIKV-infected cells (**Figure 3.3A,B**). To understand if this effect was UPF1-mediated, we leveraged the previously demonstrated ability of ZIKV capsid to specifically degrade nuclear UPF1 (*15*), and used Huh7-Lunet cells expressing ZIKV capsid protein in a tetracycline-inducible manner. Induction of capsid expression by addition of doxycycline led to reduced levels of nuclear UPF1 (**Figure 3.3C**). This also led to a similar accumulation of polyadenylated RNAs in the cell nucleus, mirroring the effect observed in ZIKV-infected cells (62% nuclear fluorescence in vector vs 78% in capsid overexpressing cells) (**Figure 3.3D**). Furthermore, siRNA-mediated knockdown of UPF1 in NPCs also resulted in retention of polyadenylated RNAs in the nucleus (36% nuclear fluorescence in control vs 40% in siUPF1 NPCs) (**Figure 3.3E**, **Supplemental Figure 3.3**). Collectively, these data indicate that UPF1 regulates nuclear mRNA export, a function perturbed upon nuclear UPF1 degradation during ZIKV infection.

Depletion of UPF1 leads to nuclear retention of *FREM2* mRNA, depletion of FREM2 protein, and shifts in NPC differentiation

Next, we focused on *FREM2*, the transcript with the largest fold decrease in UPF1 interaction in the irCLIP studies (**Figure 3.1B**). FREM2 is an extracellular matrix protein involved in cell-cell

interactions that is important in many developmental pathways including tissue and vascular morphogenesis (*37, 38*). The *FREM2* transcript was also specifically downregulated in the cytoplasm of infected NPCs (**Figure 3.2G**). We confirmed select cytoplasmic downregulation of *FREM2* mRNA expression upon infection by qPCR (**Figure 3.4A**). Using RNAscope, we found enrichment of the *FREM2* transcript in the nucleus of cells treated with UPF1-targeting siRNAs as compared to cells treated with non-targeting control siRNAs, where it was found in both the cytoplasm and the nucleus (**Figure 3.4B**). FREM2 protein levels were significantly decreased in UPF1 siRNA-treated cells (42% decrease in siUPF1 compared to the siNT control), supporting the model that nuclear mRNA retention leads to reduced translation of UPF1 target transcripts during ZIKV infection (**Figure 3.4C**).

To determine the functional consequences of reduced FREM2 expression, we used siRNAs to knockdown FREM2 in NPCs (**Supplemental Figure 3.4).** FREM2 knockdown increased the percentage of cells expressing the neural marker βIII-Tubulin by 44%, and decreased the percentage of cells expressing the pluripotency marker Sox2 by 51% (**Figure 3.4D**). These results indicate that by preventing FREM2 nuclear export and protein expression, ZIKV infection may promote premature neuronal differentiation of NPCs, thus reducing the number of NPCs that determine brain and ultimately head size.

Discussion

In this study, we identified a novel mechanism by which ZIKV infection prevents RNA export from the nucleus, a phenotype described in other viral infections, but not seen previously with flaviviruses (*39-42*). Many transcripts lost UPF1 occupancy during ZIKV infection, especially in the 3'UTR, consistent with the finding that UPF1 expression is downregulated in ZIKV infection. Loss of UPF1 binding altered transcript localization with only a minor effect on transcript abundance, underscoring the functional significance of the role of UPF1 in mRNA export. The

accumulation of mRNA in the nucleus induced by ZIKV infection could be recapitulated both by ZIKV capsid expression, which is known to degrade nuclear UPF1, and by UPF1 knockdown, linking the new function of UPF1 with its nuclear localization. Lastly, we show that when UPF1 is unable to export *FREM2* mRNA from the nucleus, FREM2 protein abundance decreases, which may disrupt neural differentiation. The majority of these experiments were performed in human NPCs, the natural target cells of ZIKV in fetuses, providing a physiologically relevant view into pathways and transcripts disrupted by ZIKV infection and how that might lead to congenital Zika syndrome.

UPF1 is an RNA helicase involved with a plethora of functions within the cell. Besides its function in NMD, UPF1 also plays important roles in several other RNA decay pathways (*43*), as well as telomere maintenance (*44*), and ubiquitination (*45*). While this study was underway, a study in fruit flies connected nuclear UPF1 function with co-transcriptional binding and export of mRNAs from the nucleus (*46*). UPF1 is also required for HIV RNA export from the nucleus (*47*).

Here we show a role of nuclear UPF1 in mRNA export in mammalian cells, which is disrupted by ZIKV infection. While the precise mechanisms of UPF1-mediated nuclear export are not yet clear, two findings underscore its significance: 1) most transcripts were not upregulated in NPCs after loss of UPF1 binding, which points to another major function of UPF1 in cells other than NMD, and 2) many polyadenylated transcripts were retained in the nucleus upon UPF1 knockdown, demonstrating that the effect of UPF1 on mRNA export is widespread. Additional studies are needed to determine whether specific features of these mRNAs, such as the 3'UTR sequence or structure, underlie their targeting for export by UPF1 (*48*). Similarly, additional studies are needed to determine the features that lead to UPF1 enrichment in the ZIKV 3'UTR. Interestingly, a non-coding subgenomic flavivirus RNA (SfRNA) is derived from the 3' UTR of ZIKV and is known to be resistant to XRN1 degradation (*49*). The production of an SfRNA is shared among flaviviruses and antagonizes the interferon response. UPF1 potentially interacts with ZIKV SfRNA, which

could play an important role in NMD evasion. These avenues of research would be important to explore in future studies.

The host shutoff effect is a common viral strategy to reduce host gene expression, potentially for the purposes of competing with cellular resources and evading immune responses. Flaviviruses have been reported to cause host shutoff by translational repression (*50*). The host shutoff described here is achieved by retention of host transcripts in the nucleus. We confirm that a single viral protein, the capsid protein, is sufficient to specifically degrade nuclear UPF1 and cause nuclear mRNA retention. This underscores the importance of the localization of ZIKV capsid to the cell nucleus (*51*), which is independent from its function as a structural component of the virion as the virus replicates exclusively outside of the nucleus (*52*). Our study also highlights the significance of nuclear UPF1, which is less studied than cytoplasmic UPF1, although it has been associated with cell cycle progression, DNA replication, telomere maintenance, and mRNA release (*35, 46*).

Nuclear retention of host mRNAs caused by the disruption of UPF1 function could contribute to the neurodevelopmental defects seen during fetal ZIKV infections. Since UPF1 depletion impacts many mRNAs, it may represent a central mechanism explaining pleotropic effects of ZIKV infection on diverse cellular pathways. We pursued *FREM2* as a top UPF1 target, and show that it is downregulated at the protein level when UPF1 is unable to perform its role as a nuclear mRNA export regulator. FREM2 is both a member of the FRAS/FREM complex and regulates its formation (*37*). The FRAS/FREM complex is found in cellular basement membranes and is expressed differentially during development (*38*). FREM2 itself is important for proper development of the eye and FREM2 mutation is associated with Fraser syndrome, in which cryptopthalamos (a congenital defect where the eyes are covered completely by skin and often associated with small or missing eyeballs) is commonly seen (*53*). As developmental defects in

the eyes are also seen in children with congenital Zika syndrome, the retention of the *FREM2* transcript in the nucleus could be directly involved in pathogenesis (*54, 55*).

Surprisingly, our studies in NPCs and Huh7 hepatoma cells indicate that at least in these cell types the role of UPF1 in nuclear export affects more mRNAs than its known function in NMD. How this function is regulated and whether it involves known UPF1 interacting proteins such as the EJC or the known RNA export pathways such as CRM-1, remains to be determined. The finding that ZIKV has evolved a mechanism to selectively target nuclear UPF1 and the mRNA export function of the protein underscores its central yet so far underappreciated role in biology.

Materials and Methods

Cell Culture and Viruses

Human iPSC-derived NPCs were generated and maintained as described previously (*56*). The human fibroblast cell line used to generate iPSCs came from the Coriell Institute for Medical Research and Yale Stem Cell Center. The iPSCs used in these studies was the CTRL2493n17 line. CTRL2493n17 was derived from the parental fibroblast line ND31845 that was biopsied from a healthy female at 71 years of age. iPSCs were cultured and maintained in complete mTESR (StemCell Techologies). NPCs were differentiated and maintained using EFH media (Stemline Neural Stem Cell Media [Sigma Aldrich], EGF [R&D Biosystems], rhFGF basic [R&D Biosystems]. Heparin Sulfate [Sigma Aldrich]). NPCs were dissociated and plated onto Matrigel-coated plates (Corning) prior to infecting with ZIKV (MOI of 1) or treating with siRNAs. Experiments were harvested 48 hours post infection and up to 7 days after siRNA treatment. Huh7-Lunet cells (Ralf Bartenschlager, Heidelberg University), and Vero cells were maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS), 2 mM l-glutamine, 100 U/ml penicillin, and 100 µg/ml streptomycin.

N-terminally Strep TagII tagged ZIKV Capsid was cloned to pLVX-TetOne-Puro Vector (Clontech, Cat: 631849) using BamHI & EcoRI cut sites. We used a 2nd generation lentiviral system to generate a stably expressing ZIKV Capsid Huh7-Lunet cell line. After transduction, Huh7-Lunet cells were selected with puromycin (2ug/ml) for one weeks and later clonally expanded. Clones with the highest and specific expression were used for the experiment.

The strain PRVABC59 of ZIKV (ATCC VR-1843) was used for all experiments. ZIKV stocks were propagated in Vero cells (ATCC), and titers were determined by plaque assays on Vero cells. ZIKV infections were performed by adding viral inoculum to DMEM with 2% FBS or EFH followed by a two-hour incubation at 37C with a rock every 15 minutes. After infection was completed, inoculum was aspirated and then fresh DMEM with 10% FBS or EFH was added to the cells. Infected cells were cultured for 48 hours prior to harvesting for all sequencing and IF experiments.

Antibodies and other reagents

Primary antibodies used were anti-UPF1 (Bethyl laboratories, A300-38A, CST, 12040S and Abcam, ab109363), anti-FREM2 (Invitrogen, PA5-20982), anti-FLAG (Abcam, ab18230), anti-Actin (CST, 4967S), anti-GFAP (Abcam, ab53554), anti-Nestin (Abcam, ab22035) and anti-Beta III tubulin (Abcam, ab18207). Secondary antibodies used include goat anti-rabbit Alexa 488 (Invitrogen, A-11008), goat anti-mouse Alexa 488 (Invitrogen, A-11001), goat anti-rabbit Alexa 594 (Invitrogen, A-11012), goat anti-mouse Alexa 594 (Invitrogen, A-11005), donkey anti-goat Alexa 647 (Invitrogen, A-21447), donkey anti-rabbit Alexa 488 (Invitrogen, A-21206), donkey anti-mouse Alexa 594 (Invitrogen, A-21203). The RNAScope Multiplex Fluorescent V2 Assay (ACD, 323100) was used with RNAscope Probes include polyA RNA (ACD, 318631) and FREM2 (ACD, 482841). Opal570 (Akoya Biosciences, FP1488001KT) was used for visualization of RNAscope probes. Accell siRNA was used for knockdown of FREM2 (Dharmacon, E-021693-00-0010) and UPF1 (Dharmacon, E-011763-00-0010) according to manufacturer's instructions. Non-targeting

siRNAs used were cat. D-001910-10-20 (Dharmacon). Huh7-Lunets were treated with 60 ng/mL of Leptomycin B (Cayman Chemical) for 16 hours.

Infrared Crosslinking and Immunoprecipitation

irCLIP was performed as in Zarnegar et al. 2016 (*57*). HeLa cells grown as described above and UV crosslinked to a total of 0.35 J/cm^2. Whole-cell lysates were generated in CLIP lysis buffer (50 mM HEPES, 200 mM NaCl, 1 mM EDTA, 10% glycerol, 0.1% NP-40, 0.2% Triton X-100, 0.5% N-lauroylsarcosine) and briefly sonicated using a probe-tip Branson sonicator to solubilize chromatin. Each experiment was normalized for total protein amount, typically 1 mg, and partially digested with RNase A (ThermoFisher Scientific, EN0531) for 10 minutes at 37°C and quenched on ice. UPF1 (Bethyl laboratories, A300-38A) IP's were performed using 15 μg of each antibody with 50 μL Protein G Dynabeads (ThermoFisher Scientific), for 8 hours at 4°C on rotation. Samples were washed sequentially in 1 mL for 1 minute each at 25°C: 1× high stringency buffer (15 mM Tris-HCl, pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 120 mM NaCl, 25 mM KCl), 1× high salt buffer (15 mM Tris-HCl pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% sodium deoxycholate, 1 M NaCl), 2× NT2 buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP-40). After the NT2 wash, RNA-protein complexes were dephosphorylated with T4 PNK (NEB) for 45 minutes in an Eppendorf Thermomixer at 37°C, 15 seconds 1400rpm, 90 seconds of rest in a 30 μL reaction, pH 6.5, containing 10 units of T4 PNK, 0.1 μL SUPERase-IN (ThermoFisher Scientific), and 6 μL of PEG-400 (16.7% final). Dephosphorylated RNA-protein complexes were then rinsed once with NT2 buffer and 3'-end ligated with T4 RNA Ligase 1 (NEB) overnight in an Eppendorf Thermomixer at 16°C, 15 seconds 1400rpm, 90 seconds of rest in a 60 μL reaction containing 10 units T4 RNA Ligase, 1.5 pmol pre-adenylated-IR800-3'biotin DNA-adapter, 0.1 μL SUPERase-IN, and 6 μL of PEG400 (16.7% final). The following day, samples were again rinsed once with 500 μL NT2 buffer

and resuspended in 30μL of 20 mM DTT, 1x LDS (ThermoFisher Scientific) in NT2 buffer. Samples were heated to 75°C for 10 min, and released RNA-protein complexes were separated on 4-12% Bis-Tris SDS-PAGE (1.0mm X 12 well) at 200V for 45 min. Resolved RNP complexes were wet-transferred to nitrocellulose at 550 mA for 45 minutes at 4°C.

Nitrocellulose membranes were imaged using an Odyssey CLx scanner (LiCor), RBP-RNA complexes were excised using scalpels, and RNA was recovered by adding 0.1 mL of Proteinase K reaction buffer (100 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.2% SDS) and 5 μL of 20mg/mL Proteinase K (ThermoFisher Scientific). Proteins were digested for 60 minutes at 50°C in an Eppendorf Thermomixer. Next, 200 μL of saturated-phenol-chloroform, pH, 6.7 was added to each tube and incubated for 10 minutes at 37°C in an Eppendorf Thermomixer, 1400 rpm. Tubes were briefly centrifuged and the entire contents transferred to a 2 mL Heavy Phase Lock Gel (5Prime, 2302830). Samples were centrifuged for 2 minutes at >13000 rpm. The aqueous layer was re-extracted with 1 mL of chloroform (inverting 10 times to mix; no vortexing) in the same 2 mL Phase Lock Gel tube and centrifuged for 2 minutes at >13000 rpm. The aqueous layer was then transferred to a new 2 mL Heavy Phase Lock Gel tube and extracted again with an additional 1 mL of chloroform. After 2 minutes centrifugation at >13000 rpm, the aqueous layer was transferred to a siliconized 1.5 mL tube and precipitated overnight at -20°C by addition of 10 μL 5M NaCl, 3 μL Linear Polyacrylamide (ThermoFisher Scientific) and 0.8 mL 100% ethanol. RNA fragments were pelleted at >13000 rpm for 45 minutes at 4°C, washed once with 1 mL of ice cold 75% ethanol and air dried.

RNA pellets were resuspended in 12 μL water 1 μL of 3 μM cDNA and 1 μL of 10mM dNTPs and heated to 70°C for 5 minutes then rapidly cooled to 4°C. cDNA Master Mix (4 μL 5x Super Script IV (SSIV) Buffer, 1 μL 100mM DTT, 1 μL SSIV, 6 μL total) was added to the annealed RNA and

incubated for 30minutes at 55°C. cDNA:RNA hybrids were captured by addition of 5 μL of MyOne Streptavidin C1 Dynabeads (ThermoFisher Scientific) that had been rinsed and suspended in 50 μL of Biotin-IP buffer (100mM Tris, pH 7.5, 1M NaCl, 1mM EDTA, 0.1% Tween), and end over end rotation for 45 minutes at room temperature. Beads were placed on a 96-well magnet and washed sequentially twice with 100 μL of Biotin IP buffer and 100 μL ice-cold 1xPBS. Beads were resuspended in 10 μL of cDNA elution buffer (8.25 μL water, 1 μL of 1 μM P3 short oligo, and 0.75 μl of 50 mM $MnCl_2$) and heated to 95°C for 10 minutes, ramp 0.1 degree/second to 60°C forever. Next 5 μL of circularization reaction buffer was added (3.3 μL water, 1.5 μL 10x Circligase-II buffer, and 0.5 μL of Circligase-II (Epicentre)). cDNA was circularized for 2 hours at 60°C. cDNA was purified with 30 μL of AMPure XP beads (Beckman Coulter) and 75 μL of isopropanol. Samples were incubated for 20 minutes at 25°C, washed twice with 100 μL 80% ethanol, air dried for 5 minutes, and eluted in 14 μL of water. Elution took place at 95°C for 3 minutes and the eluent was immediately transferred to a 96-well magnet. Eluted cDNA was transferred to a new PCR tube containing 15 μL of 2X Phusion HF-PCR Master Mix (NEB), 0.5 μL of 30 μM P3/P6 PCR1 oligo mix and 0.5 μl of 15x SYBR Green I (ThermoFisher Scientific). Real-time quantitative PCR was performed: 98°C 2 min, 15 cycles of 98°C 15 seconds, 65°C 30 seconds, 72°C, 30 seconds, with data acquisition set to the 72°C extension. PCR1 reactions were cleaned up by adding of 4.5 μL of isopropanol, 54 μL of AMPure XP beads and incubation for 10 min. Beads were washed once with 80% ethanol, dried for 5 min, and eluted in 15 μl of water. Illumina flow cell adaptors were added by adding 15 μL 2X Phusion HF-PCR Master Mix and 0.4 μL P3solexa/P6solexa oligo mix and amplified: 98°C 2 min, 3 cycles of 98°C 15 seconds, 65°C 30 seconds, 72°C, 30s seconds. Final libraries were purified by addition of 48 μL of AMPure XP beads and incubation for 5 min. Beads were washed twice with 70% ethanol, dried for 5 min, and eluted in 20 μL of

water. 1-2μL of libraries were quantitated by HS-DNA Bioanalyzer. Samples were deep sequenced on the Illumina NextSeq machine: single-end, no index, high-output, 75-bp cycle run.

Whole transcriptome RNA sequencing was performed using the methods described above for RNA extraction, library preparation and sequencing.

Nuclear/cytoplasmic RNA fractionation and sequencing

Fractionation was performed using the Cytoplasmic and Nuclear RNA Purification Kit (Cat. # 2100, Norgen Biotek Corp). Purified RNA was treated with DNase I followed by library preparation using the NuGEN V2 RNA-Sequencing Library Preparation kit (Tecan Genomics). Both RNA and library quality were analyzed via a Bioanalyzer (Agilent). Sequencing was performed on a NextSeq 500 (Illumina): single-end, no index, high-output, 75-bp cycle run.

IF and RNAscope protocol

For immunofluorescence and RNAscope, infected NPCs were collected at 48 hr and plated onto 22- by 22-mm no. 1.5 coverslips. Cells were then fixed with 4% PFA in PBS for 15 minutes. For the RNAscope protocol, we followed manufacturer's instructions for adherent cell lines. Briefly, we first dehydrated the cells using 50%, 70% and then 100% ethanol in PBS. This was followed by a rehydration of the cells using 70% and then 50% of ethanol in PBS. Lastly, cells were fully rehydrated in PBS. Cells were then permeabilized by hydrogen peroxide, followed by protease 3 treatment. Next, we hybridized the RNAscope probes to the cells for 2 hours or O/N at 40C. Probe amplification was then performed, followed by labelling with Opal570 (Akoya Biosciences). Nuclei were stained using Hoechst 33258 (Thermofisher).

For immunofluorescence, cells were fixed, permeabilized by 0.1% Triton X-100 (Sigma Aldrich), blocked with 3% bovine serum albumin (Sigma Aldrich) in PBS. Cells were then immunostained

with the indicated antibody, followed by the appropriate secondary. Lastly, nuclei were stained with Hoechst 33258.

Microscopy was performed on an LSM880 with Airyscan (Zeiss) or an Olympus FV3000RS. On the LSM880, imaging of Huh7-Lunet cells was performed with 20x magnification objective, while NPCs were imaged with a 63x oil objective. On the FV3000RS, the 20x objective was used for imaging NPCs. All images were taken as a Z-stack.

Western blot analysis

Cells were lysed in RIPA lysis buffer (50 mM Tris-HCl [pH 8], 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, supplemented with Halt protease inhibitor cocktail [Thermo Fisher Scientific]) to obtain whole-cell lysates or lysed using the NE-PER nuclear and cytoplasmic extraction kit (Thermo Fisher Scientific) to obtain cytoplasmic and nuclear fractions. Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes (Bio-Rad). Proteins were visualized by chemiluminescent detection with ECL and visualized on a ChemiDoc MP Imaging System (Bio-Rad).

Computational and Statistical Analyses

For western blot analysis, differences in band intensity were quantified by densitometry using ImageJ (*58*). Student's t-test was used for statistical analysis of Western Blots. Imaris (Oxford Instruments) was used for analysis of confocal images, using the surface function for polyA mRNA analysis and dot function identifying specific transcripts by RNA-scope. Nuclei were also bounded and identified by the surface function of Imaris. PolyA RNA-scope experiments were statistically analyzed using a linear mixed model to account for individual cell values across multiple biological replicates. Data are represented as means plus standard errors of the means (SEM). Gene set overlap statistics were performed using a hypergeometric test. Statistical significance was defined

as follows: *, P ≤ 0.05; **, P ≤ 0.01; ***, P ≤ 0.001; ****, P ≤ 0.0001. Biological replicates are defined as the same experimental design but performed sequentially, with a different cell passage number and on different days.

Analysis of RNA sequencing data

PCR duplicates were removed using unique molecular identifiers in the RT primer region. The adaptor and barcode sequences were trimmed and reads were mapped step-wise to viral (ZIKV), repetitive and finally non-repetitive (GRCh38) genomes. Bowtie2 indexes were generated using the 'bowtie2-build' command in Bowtie2 for the ZIKV (KU501215.1) RNA genome sequences. The specific parameters used for the FAST-iCLIP pipeline were as follows: -f 18 (trims 17 nt from the 5′ end of the read), -l 16 (includes all reads longer than 16 nt), –bm 29 (minimum MAPQ score from bowtie2 of 29 is required for mapping; unique mapping only), –tr 2,3 (repetitive genome) and –tn 2,3 (non-repetitive genome) RT stop intersection (n,m; where n = replicate number and m = number of unique RT stops required per n replicates). Using the –tr/tn 2,3 parameters, a minimum of six RT stops are required to support any single nucleotide identified as a crosslinking site.

Analysis of the sequencing data was performed using a custom analysis pipeline, with the peak finding software uploaded to Github (https://github.com/ChangLab/FAST-iCLIP/tree/lite). Other analyses were performed by aligning the reads to the human genome using STAR, followed by gene counts using Bedtools (*59*). Only reads in the canonical 3'UTR of the human transcripts were counted. The count distribution across the metagene of the CDS and 3'UTRs was visualized using deeptools (*60*). Log_2 fold changes were calculated by comparing RPKM values between Mock and ZIKV infected cells. Genes with fewer than 10 total read counts were excluded from analysis.

Heat maps and XY plots were produced using Prism 8 (GraphPad).

For total RNA and cytoplasmic/nuclear fractionated sequencing, reads were aligned to the human genome using the STAR aligner (*61*) (version 2.7.5) followed by HTSeq (*62*) (version 0.12.3) to obtain counts and then using the DeSeq2 (*63*) (version 1.28.1) pipeline to determine $\log_2$ fold changes in transcripts. Volcano plots were produced using the R-package, Enhanced Volcano (https://github.com/kevinblighe/EnhancedVolcano).

Figures

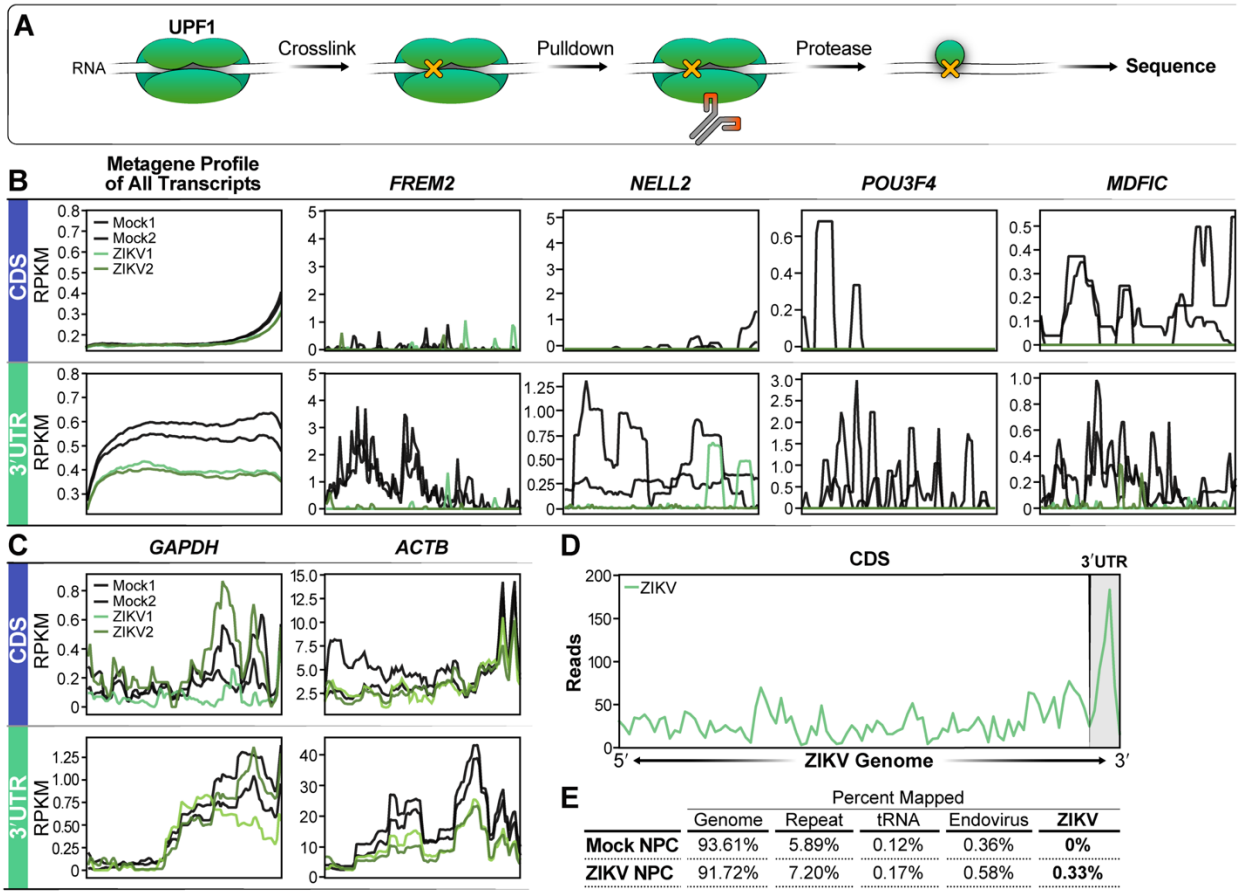


Figure 3.1: UPF1 interactions with many host transcripts are lost during ZIKV infection of NPCs

A) irCLIP schematic describing the workflow to obtain sequencing data. After 48 hrs of ZIKV infection, UPF1 and RNA are crosslinked using UV light, followed by UPF1 pulldown and protease degradation to expose the UPF1-bound RNA for sequencing. **B)** Metagene profile of all transcripts sequenced from the irCLIP experiment. The graphs show Reads Per Kilobase of transcript, per Million mapped reads (RPKM) values for positions in the CDS and the 3'UTR. Experiment was produced from 2 biological replicates. Representative metagene plots (FREM2, NELL2, POU3F4, MDFIC) of loci found to have a loss of UPF1 interaction. **C)** A metagene plot of GAPDH, a transcript resistant to UPF1 degradation, and ACTB are shown as negative controls. **D)** Metagene plot of reads mapping to the ZIKV genome, with the 3'UTR marked. **E)** Tabular breakdown of read map percentages from the UPF1-CLIP experiment.

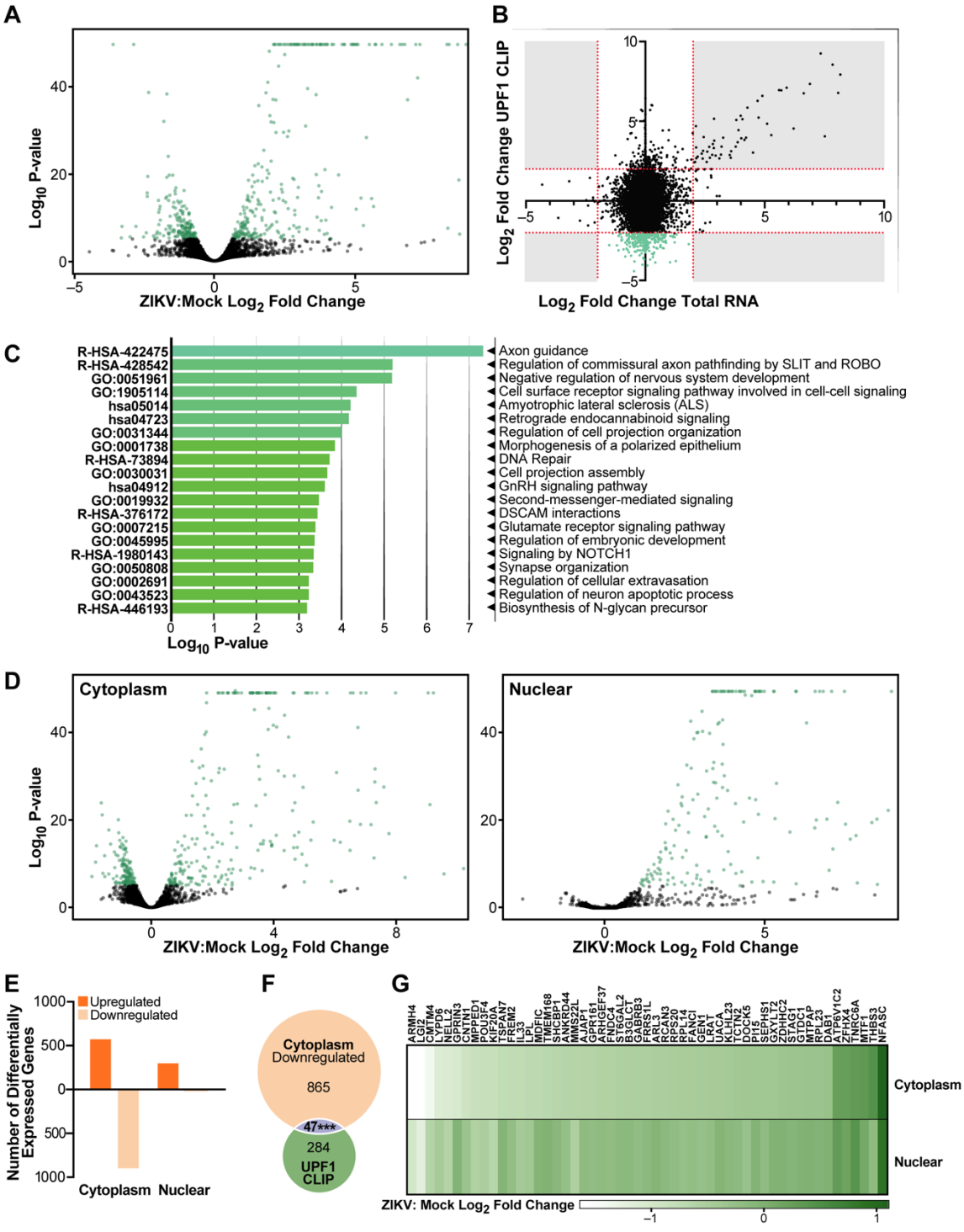

A
Log10 P-value
ZIKV:Mock Log2 Fold Change
B
Log2 Fold Change UPF1 CLIP
Log2 Fold Change Total RNA
C
R-HSA-422475
R-HSA-428542
GO:0051961
GO:1905114
hsa05014
hsa04723
GO:0031344
GO:0001738
R-HSA-73894
GO:0030031
hsa04912
GO:0019932
R-HSA-376172
GO:0007215
GO:0045995
R-HSA-1980143
GO:0050808
GO:0002691
GO:0043523
R-HSA-446193
Axon guidance
Regulation of commissural axon pathfinding by SLIT and ROBO
Negative regulation of nervous system development
Cell surface receptor signaling pathway involved in cell-cell signaling
Amyotrophic lateral sclerosis (ALS)
Retrograde endocannabinoid signaling
Regulation of cell projection organization
Morphogenesis of a polarized epithelium
DNA Repair
Cell projection assembly
GnRH signaling pathway
Second-messenger-mediated signaling
DSCAM interactions
Glutamate receptor signaling pathway
Regulation of embryonic development
Signaling by NOTCH1
Synapse organization
Regulation of cellular extravasation
Regulation of neuron apoptotic process
Biosynthesis of N-glycan precursor
Log10 P-value
D
Cytoplasm
Nuclear
Log10 P-value
ZIKV:Mock Log2 Fold Change
ZIKV:Mock Log2 Fold Change
E
Upregulated
Downregulated
Number of Differentially Expressed Genes
Cytoplasm Nuclear
F
Cytoplasm
Downregulated
865
47***
284
UPF1 CLIP
G
Cytoplasm
Nuclear
ZIKV: Mock Log2 Fold Change

Figure 3.2: **Loss of UPF1 interaction during ZIKV infection causes mRNA downregulation in the cytoplasm**

A) Volcano plot of whole transcriptome RNA sequencing of ZIKV infected NPCs compared to Mock, n=2 biological replicates. **B)** X-axis shows ZIKV: Mock log_2 fold change in whole transcriptome sequencing, Y-Axis shows ZIKV:Mock log_2 fold change from the UPF1 CLIP experiment. The green transcripts are >2 log_2 fold change in the UPF1-CLIP and <2 log2 fold change in total RNA. **C)** Metascape analysis of transcripts identified to have a significant loss of UPF1 interaction in ZIKV infected cells compared to Mock. **D)** Volcano plot of sequencing from ZIKV-infected NPCs compared to Mock fractionated into cytoplasmic and nuclear fractions, n=2 biological replicates. **E)** Number of differentially expressed genes from sequencing of ZIKV-infected NPCs compared to Mock fractionated into cytoplasmic and nuclear fractions. **F)** Overlap of transcripts found to be significantly downregulated in the cytoplasm and those with greater than a 2 log_2 fold decrease in the UPF1 CLIP. Statistics performed by hypergeometric probability test, ***, $P \leq 0.001$. **G)** A heatmap showing log_2 fold changes of transcripts identified in the CLIP experiment in cytoplasmic and nuclear fractions.

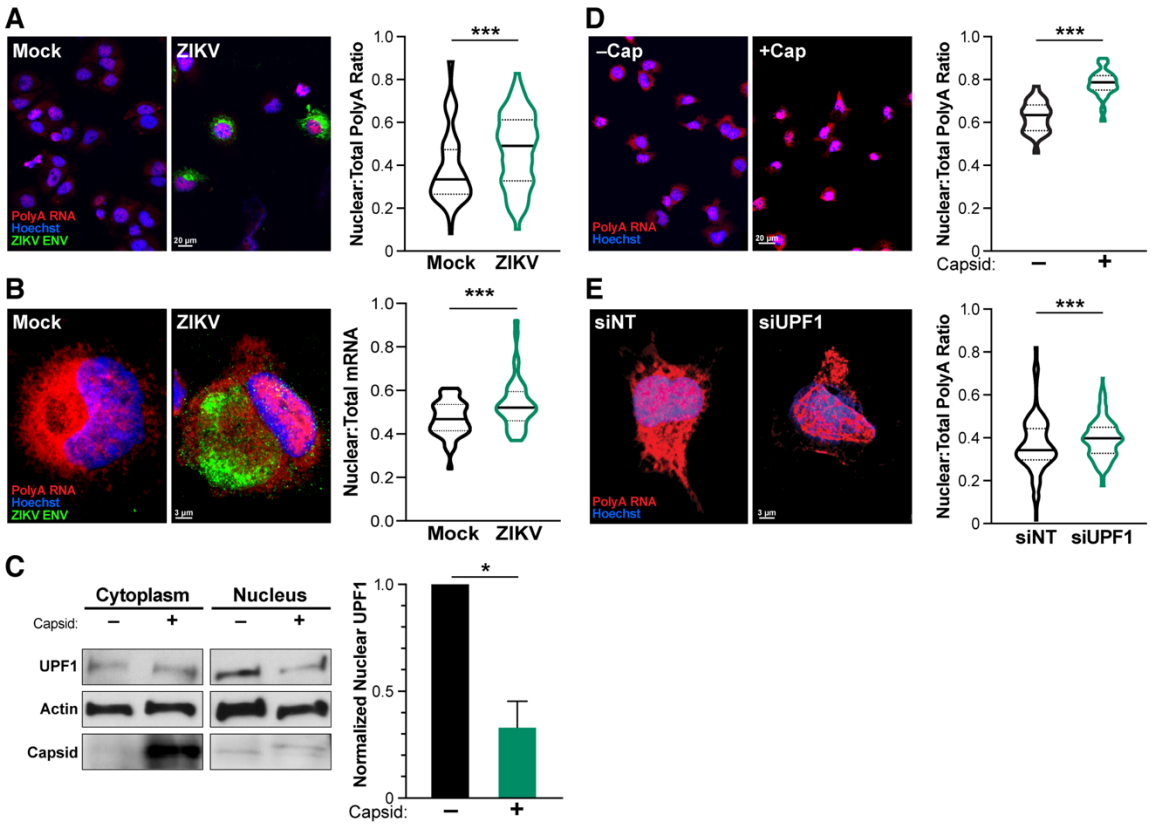


Figure 3.3: ZIKV-mediated degradation of UPF1 leads to mRNA retention in the nucleus. **A and B)** During ZIKV infection (ZIKV ENV, green), polyA RNA (red) is shifted toward the nucleus (Hoechst, blue) in Huh7-Lunet cells. Statistics produced by a Linear Mixed Model. 3 biological replicates, n=25 cells (A) or 15 cells (B) per condition per replicate. **C)** Tetracycline-inducible Capsid expression in Huh7-Lunet cells was used to degrade nuclear UPF1. Leptomycin B (LMB) was used to increase UPF1 degradation. Statistics performed by Student's t-test, n=3, and representative western blot shown. **D)** Capsid overexpression in Huh7-Lunet cells results in an increased ratio of polyA RNAs (red) in the nucleus. N=3 biological replicates, 25 cells per condition per replicate. **E)** Knockdown of UPF1 in NPCs results in an increased ratio of polyA RNAs (red) in the nucleus. NPCs were treated with siNT and siUPF1 for 96 hours. Statistics produced by a Linear Mixed Model. n=3 biological replicates, 15 cells per condition per replicate. *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$. Error bars are SEM.

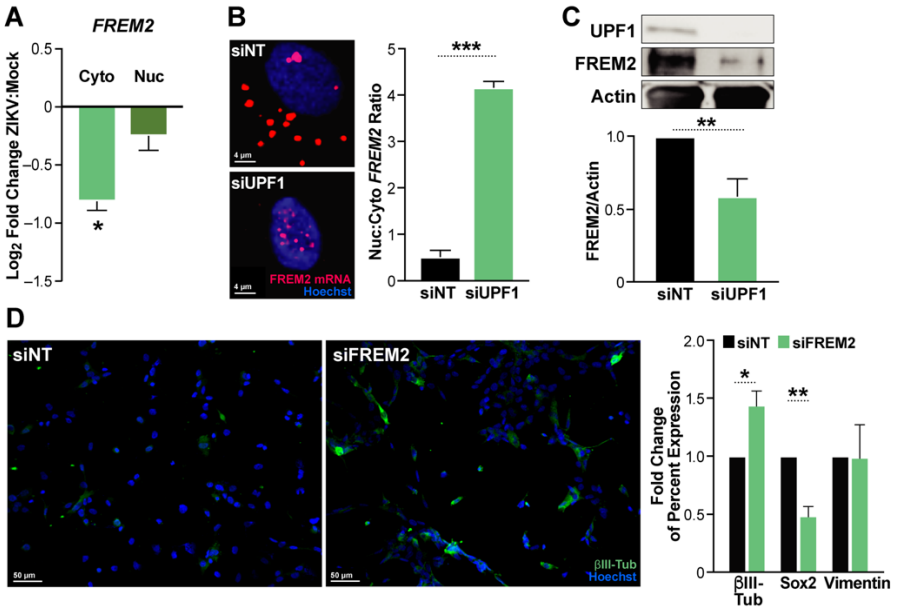
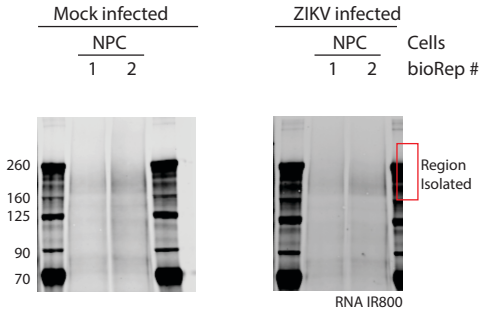
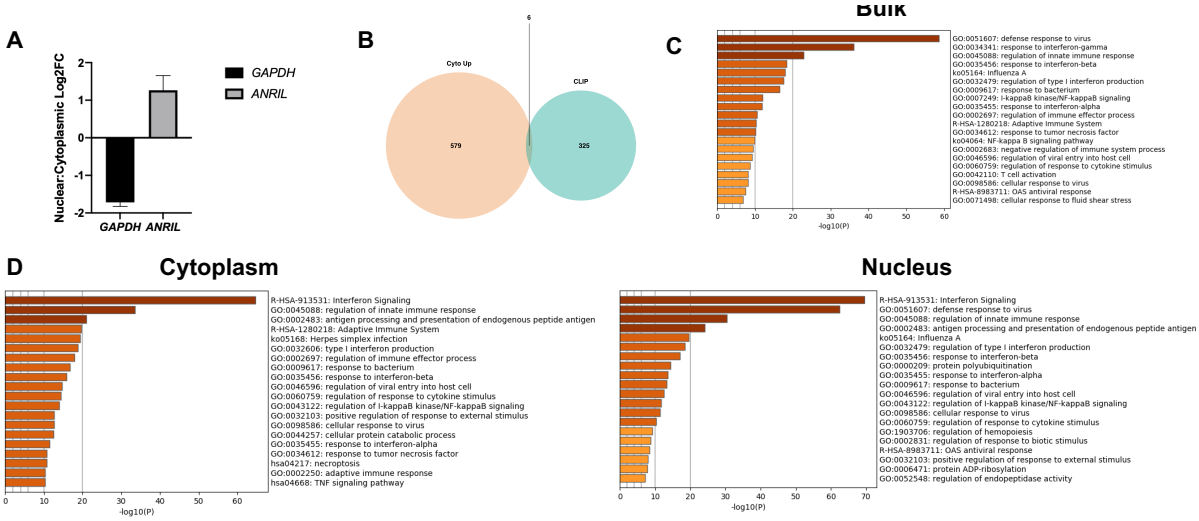


Figure 3.4: UPF1 knockdown leads to retention of *FREM2* mRNA and decreased protein production, which causes premature differentiation of NPCs.
A) qPCR of *FREM2* in cytoplasmic and nuclear fractions in ZIKV infected NPCs. **B)** RNAscope of *FREM2* in UPF1 knockdown NPCs. NPCs were treated with siRNAs for 96 hours prior to harvesting for microscopy. Number of nuclear puncta compared to cytoplasmic as calculated by Imaris. 3 biological replicates, 10 cells per biological replicate per condition averaged. Statistics produced by Student's t-test. **C)** Western blot of siNT and siUPF1 treated NPCs. NPCs were treated with siRNAs for 96 hours prior to harvesting for microscopy. Densitometric analyses of FREM2 were performed using ImageJ to quantify relative band intensities. **D)** siNT and siFREM2 treated NPCs stained for βIII-Tubulin, Soxs2 and Vimentin. NPCs were treated with siRNAs for 7 days prior to analysis. Statistics produced by Student's t-test. *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$. Error bars are SEM.



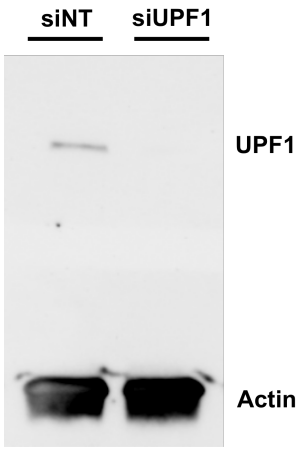
Supplemental Figure 3.1

Electrophoresis of RNA pulled down with UPF1, visualized using the IR handle ligated to the RNAs. The region bounded by the red box indicates the part of the gel excised and then analyzed by irCLIP and AP-MS. AP-MS analysis is indicated in Supplemental File 1.



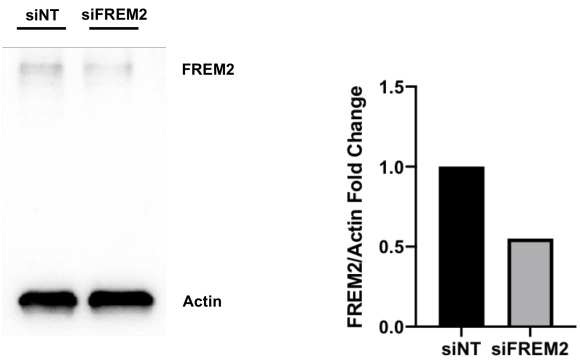
Supplemental Figure 3.2

A) Log_2 fold change from the fractionated RNA-sequencing between the nucleus and cytoplasm for markers of successful fractionation in the mock infected samples: GAPDH for cytoplasm and ANRIL for nucleus.
B) Comparison of transcripts increased in the cytoplasm and transcripts that lost interaction with UPF1 as determined by CLIP. Only 6 transcripts overlapped between the two groups.
C) Metascape Analysis of significantly upregulated transcripts found in the bulk RNA sequencing of Figure 2A. The top 400 upregulated genes were used to produce this GO clustering.
D) Metascape Analysis of significantly upregulated transcripts found in the cytoplasmic and nuclear fractionated RNA sequencing of Figure 2D. The top 400 and 300 upregulated genes were used to produce this GO clustering.



Supplemental Figure 3.3
Western blot for UPF1 in siNT and siUPF1 treated cells. Actin is shown as a loading control.

Supplemental Figure 4



Supplemental Figure 3.4
Western blot for FREM2 in siNT and siFREM2 treated NPCs after 7 days. Densitometric analyses of FREM2 were performed using ImageJ to quantify relative band intensities

References

1. D. Musso, D. J. Gubler, Zika Virus. *Clinical Microbiology Reviews* **29**, 487-524 (2016).
2. T. C. Pierson, M. S. Diamond, The emergence of Zika virus and its new clinical syndromes. *Nature* **560**, 573-581 (2018).
3. A. S. Fauci, D. M. Morens, Zika Virus in the Americas--Yet Another Arbovirus Threat. *The New England Journal of Medicine* **374**, 601-604 (2016).
4. F. R. Cugola *et al.*, The Brazilian Zika virus strain causes birth defects in experimental models. *Nature* **534**, 267-271 (2016).
5. H. Tang *et al.*, Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth. *Cell Stem Cell* **18**, 587-590 (2016).
6. Q. Shao *et al.*, Zika virus infection disrupts neurovascular development and results in postnatal microcephaly with brain damage. *Development (Cambridge, England)* **143**, 4127-4136 (2016).
7. B. S. F. Souza *et al.*, Zika virus infection induces mitosis abnormalities and apoptotic cell death of human neural progenitor cells. *Scientific reports* **6**, 39775 (2016).
8. F. Costa *et al.*, Emergence of Congenital Zika Syndrome: Viewpoint From the Front Lines. *Ann Intern Med* **164**, 689-691 (2016).
9. V. Martínez-Cerdeño, S. C. Noctor, Neural Progenitor Cell Terminology. *Front Neuroanat* **12**, 104 (2018).
10. C. R. Ojha *et al.*, Complementary Mechanisms Potentially Involved in the Pathology of Zika Virus. *Front Immunol* **9**, 2340-2340 (2018).
11. P. S. Shah *et al.*, Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. *Cell* **175**, 1931-1945.e1918 (2018).

12. N. Link *et al.*, Mutations in ANKLE2, a ZIKA Virus Target, Disrupt an Asymmetric Cell Division Pathway in Drosophila Neuroblasts to Cause Microcephaly. *Dev Cell* **51**, 713-729.e716 (2019).

13. F. Wen *et al.*, Zika virus increases mind bomb 1 levels, causing degradation of pericentriolar material 1 (PCM1) and dispersion of PCM1-containing granules from the centrosome. *J Biol Chem* **294**, 18742-18755 (2019).

14. M. Saade *et al.*, Multimerization of Zika Virus-NS5 Causes Ciliopathy and Forces Premature Neurogenesis. *Cell Stem Cell* **27**, 920-936.e928 (2020).

15. K. A. Fontaine *et al.*, The Cellular NMD Pathway Restricts Zika Virus Infection and Is Targeted by the Viral Capsid Protein. *mBio* **9**, (2018).

16. M. Li *et al.*, Identification of antiviral roles for the exon-junction complex and nonsense-mediated decay in flaviviral infection. *Nat Microbiol* **4**, 985-995 (2019).

17. K. Leon, M. Ott, An 'Arms Race' between the Nonsense-mediated mRNA Decay Pathway and Viral Infections. *Seminars in cell & developmental biology*, (2020).

18. G. Balistreri, C. Bognanni, O. Mühlemann, Virus Escape and Manipulation of Cellular Nonsense-Mediated mRNA Decay. *Viruses* **9**, (2017).

19. T. Kurosaki, L. E. Maquat, Nonsense-mediated mRNA decay in humans at a glance. *J Cell Sci* **129**, 461-467 (2016).

20. T. Kurosaki, M. W. Popp, L. E. Maquat, Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol* **20**, 406-420 (2019).

21. N. H. Gehring, G. Neu-Yilik, T. Schell, M. W. Hentze, A. E. Kulozik, Y14 and hUpf3b Form an NMD-Activating Complex. *Molecular Cell* **11**, 939-949 (2003).

22. M. W. Popp, L. E. Maquat, Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu Rev Genet* **47**, 139-165 (2013).

23. N. H. Gehring, S. Lamprinaki, A. E. Kulozik, M. W. Hentze, Disassembly of exon junction complexes by PYM. *Cell* **137**, 536-548 (2009).

24. B. Loh, S. Jonas, E. Izaurralde, The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev* **27**, 2125-2138 (2013).

25. S. Lykke-Andersen *et al.*, Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev* **28**, 2498-2517 (2014).

26. N. Imamachi, K. A. Salam, Y. Suzuki, N. Akimitsu, A GC-rich sequence feature in the 3′ UTR directs UPF1-dependent mRNA decay in mammalian cells. *Genome Res* **27**, 407-418 (2017).

27. B. W. Kebaara, A. L. Atkin, in *Nucleic Acids Res*. (2009), vol. 37, pp. 2771-2778.

28. M. Peccarelli, B. W. Kebaara, Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. *Eukaryot Cell* **13**, 1126-1135 (2014).

29. K. G. Toma, I. Rebbapragada, S. Durand, J. Lykke-Andersen, Identification of elements in human long 3' UTRs that inhibit nonsense-mediated decay. *Rna* **21**, 887-897 (2015).

30. B. J. Zarnegar *et al.*, irCLIP platform for efficient characterization of protein-RNA interactions. *Nat Methods* **13**, 489-492 (2016).

31. F. Fiorini, D. Bagchi, H. Le Hir, V. Croquette, Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. *Nat Commun* **6**, 7581 (2015).

32. C. Li *et al.*, Zika Virus Disrupts Neural Progenitor Development and Leads to Microcephaly in Mice. *Cell Stem Cell* **19**, 120-126 (2016).

33. L. Liu *et al.*, Protection of ZIKV infection-induced neuropathy by abrogation of acute antiviral response in human neural progenitors. *Cell Death Differ* **26**, 2607-2621 (2019).

34. E. T. Stoeckli, Understanding axon guidance: are we nearly there yet? *Development* **145**, (2018).

35. W. Varsally, S. Brogna, UPF1 involvement in nuclear functions. *Biochem Soc Trans* **40**, 778-783 (2012).

36. I. Vicenti *et al.*, Comparative analysis of different cell systems for Zika virus (ZIKV) propagation and evaluation of anti-ZIKV compounds in vitro. *Virus Res* **244**, 64-70 (2018).

37. E. Pavlakis, R. Chiotaki, G. Chalepakis, The role of Fras1/Frem proteins in the structure and function of basement membrane. *Int J Biochem Cell Biol* **43**, 487-495 (2011).

38. J. R. Timmer, T. W. Mak, K. Manova, K. V. Anderson, L. Niswander, Tissue morphogenesis and vascular stability require the Frem2 protein, product of the mouse myelencephalic blebs gene. *Proc Natl Acad Sci U S A* **102**, 11746-11750 (2005).

39. T. Sakuma *et al.*, Murine leukemia virus uses NXF1 for nuclear export of spliced and unspliced viral transcripts. *J Virol* **88**, 4069-4082 (2014).

40. S. K. Kuss, M. A. Mata, L. Zhang, B. M. Fontoura, Nuclear imprisonment: viral strategies to arrest host mRNA nuclear export. *Viruses* **5**, 1824-1849 (2013).

41. G. R. Kumar, B. A. Glaunsinger, Nuclear import of cytoplasmic poly(A) binding protein restricts gene expression via hyperadenylation and nuclear retention of mRNA. *Mol Cell Biol* **30**, 4996-5008 (2010).

42. P. Fortes, A. Beloso, J. Ortín, Influenza virus NS1 protein inhibits pre-mRNA splicing and blocks mRNA nucleocytoplasmic transport. *Embo j* **13**, 704-712 (1994).

43. Y. K. Kim, L. E. Maquat, UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond. *Rna* **25**, 407-422 (2019).

44. R. Chawla *et al.*, Human UPF1 interacts with TPP1 and telomerase and sustains telomere leading-strand replication. *The EMBO journal* **30**, 4047-4058 (2011).

45. Q. Feng, S. Jagannathan, R. K. Bradley, The RNA Surveillance Factor UPF1 Represses Myogenesis via Its E3 Ubiquitin Ligase Activity. *Molecular Cell* **67**, 23-251.e256 (2017).

46. A. K. Singh *et al.*, The RNA helicase UPF1 associates with mRNAs co-transcriptionally and is required for the release of mRNAs from gene loci. *Elife* **8**, (2019).

47. L. Ajamian *et al.*, HIV-1 Recruits UPF1 but Excludes UPF2 to Promote Nucleocytoplasmic Export of the Genomic RNA. *Biomolecules* **5**, 2808-2839 (2015).

48. S. R. Carmody, S. R. Wente, mRNA nuclear export at a glance. *J Cell Sci* **122**, 1933-1937 (2009).

49. G. P. Göertz, S. R. Abbo, J. J. Fros, G. P. Pijlman, Functional RNA during Zika virus infection. *Virus Res* **254**, 41-53 (2018).

50. H. Roth *et al.*, Flavivirus Infection Uncouples Translation Suppression from Cellular Stress Responses. *mBio* **8**, (2017).

51. E. R. A. Oliveira, R. Mohana-Borges, R. B. de Alencastro, B. A. C. Horta, The flavivirus capsid protein: Structure, function and perspectives towards drug design. *Virus Res* **227**, 115-123 (2017).

52. M. I. Mohd Ropidi, A. S. Khazali, N. Nor Rashid, R. Yusof, Endoplasmic reticulum: a focal point of Zika virus infection. *J Biomed Sci* **27**, 27 (2020).

53. A. M. Slavotinek, C. J. Tifft, Fraser syndrome and cryptophthalmos: review of the diagnostic criteria and evidence for phenotypic modules in complex malformation syndromes. *J Med Genet* **39**, 623-633 (2002).

54. C. V. Ventura, L. O. Ventura, Ophthalmologic Manifestations Associated With Zika Virus Infection. *Pediatrics* **141**, S161-s166 (2018).

55. B. de Paula Freitas, C. V. Ventura, M. Maia, R. Belfort, Jr., Zika virus and the eye. *Curr Opin Ophthalmol* **28**, 595-599 (2017).

56. H. i. Consortium, Developmental alterations in Huntington's disease neural cells and pharmacological rescue in cells and mice. *Nat Neurosci* **20**, 648-660 (2017).

57. Z. BJ *et al.*, irCLIP platform for efficient characterization of protein-RNA interactions. *Nature methods* **13**, (2016).

58. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675 (2012).

59. Q. AR, H. IM, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**, (2010).

60. R. F, D. F, D. S, G. BA, M. T, deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **42**, (2014).

61. D. A *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, (2013).

62. A. S, P. PT, H. W, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)* **31**, (2015).

63. L. MI, H. W, A. S, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, (2014).

Chapter 4

Genomic and Serologic Characterization of Enterovirus A71 Brainstem Encephalitis

A version of this chapter was published:

Leon, K. E. et al. Genomic and serologic characterization of enterovirus A71 brainstem encephalitis. Neurol Neuroimmunol Neuroinflamm 7, doi:10.1212/nxi.0000000000000703 (2020).

Foreword

While the majority of my dissertation has been focused on ZIKV, another project I wanted to highlight used many of the same tools, especially RNA sequencing, as a clinical diagnostic for patients suffering from severe neurological disease. This chapter is included to highlight how traditional laboratory techniques can be used to advance the field of medicine and how traditional virology can inform clinical practice.

Introduction

In early 2016, an outbreak of enterovirus A71 (EV-A71) in Catalonia caused more than 100 pediatric cases of neurological disease, ranging from aseptic meningitis to brainstem encephalitis with or without myelitis (*1*). qRT-PCR for EV with genotyping of peripheral blood, respiratory, and gastrointestinal samples, but not cerebrospinal fluid (CSF), identified EV-A71 in 40 out of 57 subjects. Other EVs were also found in 7 out of 57 subjects, including echovirus 30 (E-30), Coxsackievirus (CV) B1, and CV A10. While unable to type the EVs, the BioFire FilmArray Meningitis/Encephalitis panel detected EV in the CSF of 4 out of 20 subjects with brainstem encephalitis (*2*). The EV-A71 strain detected in this study was subtyped as subgenogroup C1, and phylogenetic analyses showed it was closely related to an EV-A71 strain associated with a 2015 German case of brainstem encephalitis (*3-5*). EV-A71 and other EV-related neurologic disease outbreaks were also reported in France and Denmark in 2016 (*6-7*). While EV-A71 was detected in peripheral body sites of many of these cases, EV-A71 was only

identified in the CSF of 0.02% of subjects in the German study, 3% in the Danish study, and 14% in the French study.

Despite the presence of EV in peripheral samples, questions remained about the Catalonia outbreak. Since pan-EV qRT-PCR was negative in the CSF of subjects with brainstem encephalitis (*1,8*), concern persisted that neurological symptoms may have been caused by a para-infectious mechanism or an unidentified co-infection. In addition, standard qRT-PCR assays on peripheral samples recovered small segments of the EV-A71 genome, thereby limiting the ability to assess for viral mutations that may have modulated neurovirulence.

To address these concerns, we deployed metagenomic next-generation sequencing (mNGS) on samples from the 2016 Catalonia outbreak. mNGS is an unbiased assay that can simultaneously identify nucleic acid from viruses, fungi, bacteria and parasites in subjects with neurological symptoms. As opposed to traditional PCR assays that amplify limited and usually highly conserved regions of a microbe's genome, the entire genome of a pathogen can often be recovered with mNGS. This makes it possible to identify genomic changes in the virus that may correlate with increased neurovirulence or reveal strain divergence (*9*). Here, we deployed mNGS on CSF, nasopharyngeal (NP), plasma, and stool samples from children affected by neurologic disease during the 2016 Catalonia outbreak. This allowed us to screen for all EVs and for possible co-infections while simultaneously comparing EV genomes for any differences between subjects with brainstem encephalitis and other manifestations of neurologic disease (i.e. meningitis with or without encephalitis with self-limited and short-lasting symptoms). We supplemented our investigations with an enhanced version of a previously published assay that comprehensively assesses for anti-viral antibodies using phage display (VirScan) and performed orthogonal confirmation with EV ELISA (*10-12*).

Methods

Cohort

The cohort consisted of 20 pediatric cases of EV-related neurological diseases. Specifically, we selected 10 children diagnosed with brainstem encephalitis or encephalomyelitis and 10 children with aseptic meningitis or uncomplicated encephalitis. Cases were recruited from the Hospital Sant Joan de Deu, University of Barcelona between April and June 2016. The Hospital Sant Joan de Deu is a 300-bed tertiary care hospital for high-complexity patients across a catchment area with a pediatric population of ~300,000 and has participated in a Spanish EV molecular surveillance network since 2010. EV-related neurological disease was defined as the detection of EV by a pan-EV qRT-PCR assay or BioFire FilmArray Meningitis/Encephalitis panel (*8,13-14*) in at least one sample from a subject suffering acute neurological disease in the absence of another clear cause (*1*). Case definitions of the World Health Organization's Guide to Clinical Management and Public Health Response for Hand-Foot-and-Mouth Disease (HFMD), detailed in Table e-1, were followed by a group of trained pediatricians to classify cases.

Case demographics and clinical syndromes are described in Table 1. De-identified samples from encephalomyelitis/brainstem encephalitis (n=10) and meningitis/encephalitis (n=10) were transferred to UCSF on dry ice and stored at -80^{o}C. Specimens from cases included CSF, serum, NP, and stool samples (**Figure 1 and Table e-2**). Among the 60 samples, 35 were EV-negative by clinical pan-EV qRT-PCR (Table e-2). Subjects 1-10 were negative by the BioFire FilmArray Meningitis/Encephalitis panel (*2*). All EV-positive subjects (by clinical PCR) were genotyped at the Enterovirus Unit of the Spanish National Centre for Microbiology according to a previously described procedure (*15*). For VirScan controls, we used CSF from pediatric subjects with other neurologic diseases (ONDs). VirScan and available clinical data has been previously published on this control cohort (*10*).

Standard Protocol Approvals, Registrations, and Patient Consents

The institutional ethics board approved the study and informed consent was obtained from parents or caretakers (*1*).

Metagenomic Sequencing Library Preparation

After samples were thawed, RNA was immediately extracted. The plasma, stool and NP samples were homogenized with OMNI-International's 2.8 mm ceramic bead kit and the TissueLyser II (Qiagen) for 5 min at 15 Hz. RNA isolation from the CSF samples (250 uL/sample), the plasma samples (250 uL/sample), the NP swab samples and two water controls was performed using the Direct-zol RNA MicroPrep with TRI reagent (Zymo Research) into 20 uL of nuclease-free water (ThermoFisher Scientific). Homogenized stool samples (250 uL/sample) were extracted using the RNeasy PowerMicrobiome Kit (Qiagen) on a Qiacube (Qiagen) into 100 uL of RNase-free water along with a water control. Sequencing libraries were prepared as described previously (*18*). The libraries were subjected to Depletion of Abundant Sequences by Hybridization (DASH), described previously, to remove human mitochondrial cDNA (*17*). The pooled library was size-selected using Ampure beads (Beckman Coulter), and concentration and quality were determined using a Fragment Analyzer (Advanced Analytical Technologies). Samples were sequenced on an Illumina HiSeq 4000 instrument using 140/140 base pair (bp) paired-end sequencing.

Sequencing libraries produced from the CSF samples were prepared together (including a water control). Sequencing libraries from the plasma, stool and NP swab samples were prepared together with their respective water controls. The water controls for CSF and NP samples contained no EV reads (i.e., sequences). The stool water control, which had been pooled for sequencing with stool samples containing EV reads, had 0.7 EV reads per million (rpM). To

differentiate whether the EV reads present in the stool water control stemmed from sample contamination during library preparation or from sequencer contamination due to barcode hopping (*18*), we re-sequenced the same stool water control library independent of the stool samples. When sequenced separately, we found no EV reads in the stool water control. This suggested that the EV reads originally detected in the stool water control resulted from barcode hopping and not from sample contamination during library preparation. To remove any potential that physical cross contamination had occurred between libraries, we re-sequenced a subset of the original libraries with pathogen abundance below 0.7 rpM and only considered a case positive if the pathogen was detected in both sequencing runs. If a pathogen was only detected in a sample in one of the two sequencing runs, it was considered to have 0 EV reads in that sample.

mNGS Bioinformatics

Sample sequences were analyzed for pathogens using a custom pathogen identification pipeline as previously described (*19*). Sequences that mapped to the EV genus were assembled *de novo* using the Geneious version 10.2.3 and St. Petersburg genome Assembler (SPAdes) version 3.10.0 (*20*). Phylogenetic trees were created in Geneious using a MUltiple Sequence Comparison by Log-Expectation (MUSCLE) or Multiple Alignment using Fast Fourier Transform (MAFFT) alignment algorithm followed by the Geneious Tree Builder tool with Neighbor-Joining build method (3, 2*1*). Bootstrapping was performed with 100 replicates. Statistics comparing the degree of concordance between research-based mNGS results and standard clinical diagnostic testing and comparing direct and indirect testing methods results were performed using Mann-Whitney and McNemar's statistical tests.

Pan-viral CSF Serologic Testing with VirScan

We constructed a T7 bacteriophage display library comprising 481,966 sixty-two amino acid peptides with 14 amino acid overlap tiled across full-length vertebrate, mosquito-borne, and tick-borne viral genomes downloaded from UniProt and RefSeq databases in February 2017, as previously described (VirScan) (*10-11*). VirScan libraries were incubated with 2 uL of CSF overnight, immunoprecipitated with protein A/G beads (ThermoFisher Scientific) for two rounds and sequenced as previously described (*10*). We normalized individual peptide counts and expressed them as reads per 25,000 reads sequenced (rpqK). A sample was considered EV positive if the total EV rpqK value was greater than the mean signal in the OND controls plus one standard deviation.

Enterovirus ELISA Validation

To validate our VirScan results, we generated recombinant EV viral protein 1 (VP1) from EV-A71 and EV-D68, given that complex and cross-reactive serologic responses to EVs are known to occur in subjects, and performed two independent ELISAs on each sample and considered the higher value as previously described (*10*). Signal was measured as the optical density (OD) at 450 nm and reported after background subtraction (background OD = 0.05). A sample was considered positive if the OD was greater than three times the background.

Data Availability

Assembled viral sequences have been deposited in Genbank, and non-human reads have been deposited in the BioProject Database.

Results

Traditional Clinical Testing versus Research mNGS

We obtained an average of 21.8 million (range 4.53-61.1 million) 140 bp paired-end reads per sample (Table e-2). The non-human sequence reads have been deposited at the National Center for Biotechnology Information (NCBI) Sequence Read Archive, BioProject (PRJNA504776).

As expected, EV-A71 was the most common virus detected across all samples (**Fig. 1A**). In the CSF, however, EV-A71 detection remained low relative to NP and fecal testing. By pan-EV qRT-PCR, 0/20 cases were positive in the CSF versus 10/16 in NP and 9/15 in fecal samples (p-value < 0.001 for both comparisons by Fisher's exact test). With mNGS, 3/20 cases were positive in the CSF versus 13/16 in NP and 10/15 in fecal samples (p-value < 0.001 and p = .003 Fisher's exact test). There was no statistically significant difference observed in the number of samples testing positive by NP or fecal testing (p = 1.0 for qRT-PCR and p = 0.43 for mNGS by Fisher's exact test). When EV-A71 was present in the CSF, the number of reads detected was lower than for the other two EVs we found, highlighting the difficulty of detecting EV-A71 in CSF (**Figure 1B**). The mean read count for E-30 was greater in the CSF than that for EV-A71 (mean rpM 107.28, range 16.24-337.8 versus 0.23, range 0.06-0.55, p-value = 0.057 by Mann-Whitney test). We detected CVB in 1 case which, like E-30, was at a much higher abundance than EV-A71 (32.25 rpM). We did uncover co-infection with more than one EV in five subjects (**Figure 1C**). Overall, we tested 60 samples by both mNGS and qRT-PCR. mNGS found EV reads in 100% of qRT-PCR positive samples (n = 25). mNGS detected EV in an additional 22% (n = 8) of samples negative by qRT-PCR, a statistically significant improvement in detection rate for mNGS versus qRT-PCR (**Figure 1D**, p = 0.01 by McNemar's test).

Phylogenetics

Full-length EV-A71 (n = 7), E-30 (n = 4), and CV-B virus (n = 1) genomes (coverage depth range 22-2,296x, Genbank MH484066-MH484076, MN515037) were assembled as described in the Methods. All seven EV-A71 genomes were essentially identical to the German neuroinvasive EV-A71 strain (Genbank KX139462.1, 99.3—99.4% nucleotide similarity and 99.7-99.8% amino acid similarity, **Figure 2A**). In addition, we reproduced a previous but more limited phylogenetic analysis showing greater similarity in the VP1 sequence between the neurovirulent EV-A71 in Catalonia and other neurovirulent EV-A71 strains detected in Germany, China, and West Africa as compared with pre-outbreak HFMD EV-A71 in Spain (**Figure 2B**) (*3, 22-24*).

To determine the degree of divergence of neurovirulent EV-A71, E-30, and CV-B from their respective viral species, we built a phylogenetic tree using the Catalonia genomes and 8,841 full length EV genomes from the NCBI's GenBank database (Accessed February 2018) clustered at 95% similarity. Since we detected more than one EV species in five subjects and because EVs are known to recombine, we used this phylogenetic analysis to determine whether any obvious interspecies recombination events had occurred (*25-26*). However, all 3 viruses clustered within their own species without any major deviation (**Figure 2C**).

Lastly, an EV-A71 genome assembled from a CSF sample was compared to EV-A71 genomes derived from the same subject's stool and NP samples (subject 16). A single amino acid substitution (S241P) in the VP1 region was found only in the CSF, not the two peripheral sites. Indeed, all of the EV-A71 genomes identified in peripheral body sites from this study had a serine in position 241 (**Figure 2D**).

VirScan Increases CSF Detection of Enterovirus

While mNGS improved the detection of EVs across our 20 subjects compared with traditional qRT-PCR, uncovered EV co-infections in 5 instances, and enabled phylogenetic and mutational analyses, detection of EV-A71 in the CSF was more difficult than in the NP or fecal samples, as described above. Therefore, we deployed VirScan for comprehensive anti-viral antibody detection in the CSF. After performing VirScan on 12 cases and 54 pediatric controls with ONDs, the only significantly enriched viral family in our cases was *Picornaviridae* (**Supplemental Figure 4.1A**, mean rpqK was 2,366 in cases versus 224 in controls, p-value < 0.001 by Mann-Whitney with Bonferroni correction). Within *Picornaviridae*, genus *Enterovirus* was the most enriched (**Supplemental Figure 4.1B-C**, median proportion of reads 0.09 in cases versus 0.002 in controls, p-value < 0.001 by Mann-Whitney with Bonferroni correction). To determine the genomic location of enriched EV antigens in VirScan, we used BLASTp to map all genus *Enterovirus* peptides against a model EV-A71 genome (Genbank Accession AXK59213.1), recognizing that not all EV subject antibodies are likely to be truly EV-A71 specific. The location and strength of the EV antigen response detected in cases by VirScan is shown normalized against controls, with immunodominant regions visible in VP1 and 3D (**Figure 4.3A**). The pattern of EV antigen targeting appears to be highly conserved and similar to a group of previously published pediatric acute flaccid myelitis cases predominantly from the United States (**Figure 4.3A**, red overlay) (*10*).

We next confirmed the EV VirScan results in a subset of cases with remaining CSF using an EV VP1 ELISA. The EV ELISA was confirmatory in all of the cases detected by VirScan (6/6). Thus, we considered the other EV cases detected by VirScan for which there was insufficient CSF to do confirmatory ELISA positive as well (n = 2 cases and n = 1 OND control were positive by VirScan but had insufficient CSF remaining for confirmatory ELISA). In addition, EV ELISA was positive for 3 cases in whom the amount of enterovirus signal by VirScan was below the

threshold we set to consider a sample positive. In total, 9/54 (17%) of the pediatric OND controls were positive for CSF EV antibodies. Of note, EV antibody detection by VirScan was not confirmed by EV ELISA in one OND control. This subject had limited clinical information available, and the EV-A71 ELISA had detectable signal above the background but was not above the conservative threshold we used to determine whether a sample was positive.

We found that we were able to detect CSF EV antibodies using VirScan or EV VP1 ELISA regardless of whether EV had been previously detected in the CSF by mNGS or qRT-PCR (**Figure 4.3B**, p-value = ns by Mann-Whitney). EV was especially difficult to detect in the CSF of subjects with brainstem encephalitis using mNGS or qRT-PCR (**Figure 4.3C**). Therefore, we asked whether EV detection in brainstem encephalitis could be improved with EV VirScan or ELISA. Indeed, serologic testing with CSF VirScan or ELISA was able to improve detection of EV in brainstem encephalitis (**Figure 4.3D**, p-value = 0.07 Fisher's Exact Test).

Discussion

The original description of the 2016 pediatric brainstem encephalitis outbreak in Catalonia identified EV-A71 as the likely etiologic agent. However, this conclusion was tempered because 1) there were multiple co-circulating EVs present during the outbreak, and 2) EV-A71 was not identified in the CSF of the majority of children with brainstem encephalitis. Identifying EVs in peripheral body sites of children with severe neurologic disease but failing to find it in the CSF mirrors both the recent North American outbreak of acute flaccid myelitis associated with EV-D68[21] and outbreaks of EV-A71 neurological disease (*27*-*28*). Here, we utilized mNGS and serologic testing with VirScan and confirmatory EV ELISA to further investigate samples from the Catalonia outbreak to enhance detection of pathogens, especially in the CSF.

mNGS identified EV-A71 in the CSF of 3 subjects not detected by qRT-PCR. EV-A71 abundance in the CSF was very low (range 0.06-0.55 rpM) compared to other EVs detected in the CSF in this study (i.e., E-30 and CVB). In addition, only 2.4% (4/166) of the EV-A71 reads mapped to the 5' untranslated region (UTR) targeted by the primers used in the clinical qRT-PCR assay. Since mNGS can identify EV sequences from any part of the viral genome, this may help explain its improved sensitivity over the more targeted clinical qRT-PCR and the BioFire FilmArray Meningitis/Encephalitis panel assays (*1,8*).

In addition, mNGS also identified five subjects co-infected by two different EVs, and two of these subjects had brainstem encephalitis. Due to the small number of subjects in our study, we are not able to determine whether co-infection contributed to disease severity. These findings highlight that in the midst of an outbreak, unbiased mNGS can detect co-infections, and future studies may uncover ways in which co-infections impact clinical presentations and outcomes.

A limitation of this paper is that, due to sample availability, we did not perform orthogonal confirmation of the mNGS-only virus identifications. As a result, our evaluation of the performance of the mNGS assay is vulnerable to incorporation bias because the gold standard by which we are evaluating its performance includes the mNGS results[29]. However, since we tested low abundance samples on two independent sequencing runs, we are confident that even the low levels of EV detected by mNGS in some samples did not result from contamination during sample preparation or from barcode hopping.

Through mNGS, we added significant new knowledge about neurovirulent EV-A71, including 7 full-length genomes. The assembled neurovirulent EV-A71 genomes were related to other neurovirulent EV-A71 strains detected in Germany, West Africa, and China rather than pre-outbreak Spanish EV-A71 strains associated with HFMD. Interestingly, we found that the EV-

A71 strain circulating in Catalonia appeared to be most closely related to the neurovirulent German strain, followed by the West African and then Chinese strains (*22*). The German strain reported in 2015 clusters within the C1 subgenogroup when comparing VP1 regions. However, the 5' UTR clusters more closely with B3 and C2-like strains, and the P2 and P3 regions cluster with C4 strains related to strains reported in China (*30*).

In a single subject, we discovered a single amino acid substitution (S241P) that was present only in EV-A71 sequences derived from the CSF, and not from viral sequences in the periphery of that same subject. This substitution was not seen in any of the other EV-A71 viruses sequenced from the same outbreak, nor was it found in any of the EV-A71 VP1 sequences compared in **Figure 4.2B**. An amino acid substitution (S241L) at this location has been previously reported to be associated with increased virulence (3*1*). However, those authors also identified coincident, additional mutations that increased the virus's ability to use the P-selectin glycoprotein ligand-1 (PSGL-1) cell receptor (*31-34*). Here, the S241P mutation we observed occurs without the mutation at position 145 that is required for PSGL-1 binding, suggesting that this variant of unknown significance at position 241 may impact virulence in a PSGL-1 independent manner. Although we were able to obtain EV-A71 sequences in the CSF from subjects 12 and 17, they did not span position 241 of VP1. No CSF remains from these subjects to obtain additional sequence information. While the clinical and functional significance of the S241P mutation remains uncharacterized, this study highlights the importance of acquiring viral nucleic acid directly from the CNS as opposed to the periphery alone.

Due to the limitations of direct detection of viral pathogens using nucleic acid-based methods, we utilized VirScan to comprehensively profile CSF anti-viral antibodies. This technique has previously shown utility in the diagnosis of encephalitis without a clear etiology (*12*). We show that antibodies directed towards the genus *Enterovirus*, but not towards other viruses, were

present in the CSF of 8/12 cases. When we further examined a subset of these cases with an EV ELISA and combined the results with VirScan, we found evidence for EV antibodies in 11/12 (92%) cases from the Catalonia outbreak versus just 9/54 (17%) pediatric OND controls with significantly improved detection of brainstem encephalitis due to EV. While we are able to detect antibodies to EVs broadly, VirScan and ELISA detected antibodies targeting highly conserved segments of EVs and thus, did not allow for consistent viral typing. In addition to detecting CSF antibodies to EV, our analysis uncovered immunodominant epitopes at VP1 and 3D, similar to what we found previously in EV-associated acute flaccid myelitis (*10,35*). However, evaluation of linear peptides from other regions of the EV genome may be more fruitful for subtyping EV infections (*35*), and future work with natively folded EV proteins may also better serologically type EVs.

While mNGS can detect a variety of pathogens in the CSF (*19,37-38*), there have been few reports on the actual diagnostic yield of mNGS of CSF compared to traditional assays such as pathogen-specific PCR and the BioFire FilmArray Meningitis/Encephalitis panel. In our study, mNGS was more sensitive in cases with very low EV read abundance (*39*). Furthermore, identifying an EV-A71 closely related to strains previously associated with neurologic disease in Germany, West Africa, and China in the CSF of 3 additional subjects provides strong evidence that the Catalonia outbreak was due to neuroinvasive EV. While the experiments described herein utilized a research-based CSF mNGS assay, a clinically validated CSF mNGS assay with a 3- to 7-day turnaround time was recently evaluated in a multi-center prospective study and is now clinically available (*39-40*). Thus, the role of mNGS in individual patient care and public health outbreak investigations will likely expand.

Similarly, the detection of CSV EV antibodies in many of these subjects supports our findings of EV nucleic acid in the CSF while also providing more information about the immune response

against EV. A limitation of the VirScan method is that we are unable to distinguish between antibodies that have been synthesized intrathecally versus peripheral synthesis followed by transudation across the blood brain barrier. Furthermore, it is difficult to definitively say whether these anti-EV IgG antibodies developed in response to this particular outbreak. However, the markedly lower levels of CSF EV antibodies in our pediatric OND control population, many of whom also had blood brain barrier compromise, supports our contention that the high levels of CSF EV antibodies in the cases results from a CNS EV infection and not simply from anti-EV antibodies formed in the periphery in response to the many peripheral EV infections children contract.

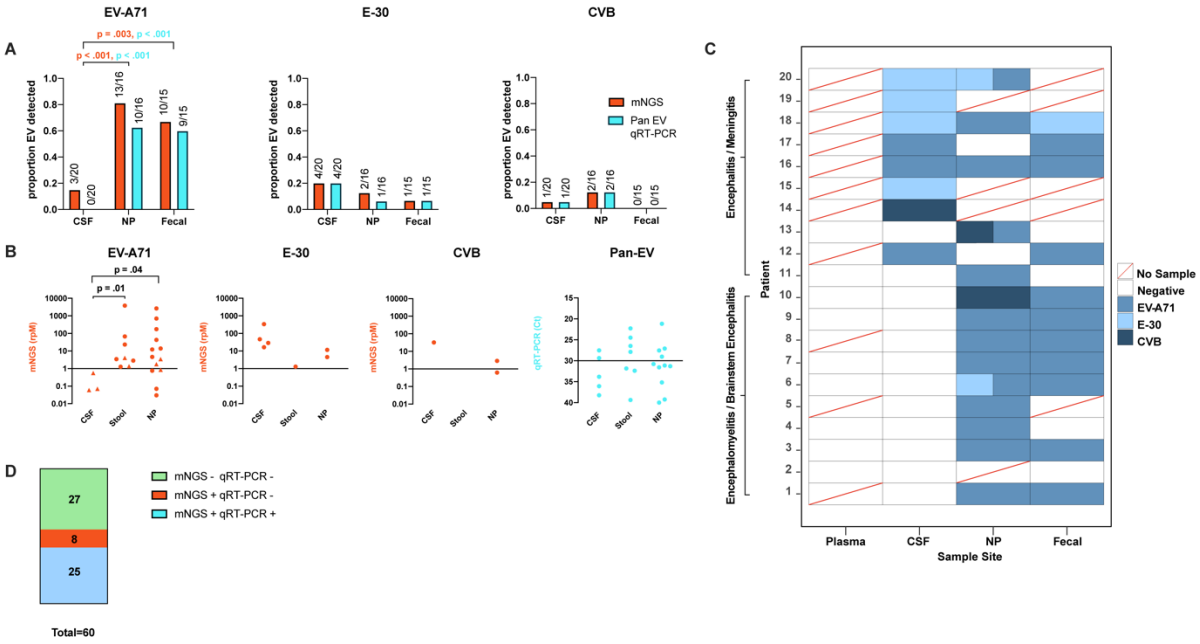


Figure 4.1: Summary of mNGS diagnostics: improvement over traditional clinical testing. **A)** Comparison of mNGS and qRT-PCR detection of EV-A71, E-30 and CVB in CSF, NP and fecal samples. Statistics performed using Fisher's exact test, with orange and blue p-values corresponding to mNGS and qRT-PCR results, respectively. **B)** Comparison of detection levels for each different experiment, with rpM representing mNGS and Ct values for qRT-PCR. Triangles denote samples identified by mNGS but not qRT-PCR. Statistics performed using a Mann-Whitney test. **C)** Heatmap of each individual subject with each body site represented. Boxes with two colors represent co-detections of different viral species. **D)** Comparison of mNGS detection rates to qRT-PCR. Statistics performed using McNemar's test. CSF = cerebrospinal fluid, NP = nasopharyngeal, EV-A71 = enterovirus A71, E-30 = echovirus 30, CVB = Coxsackievirus B

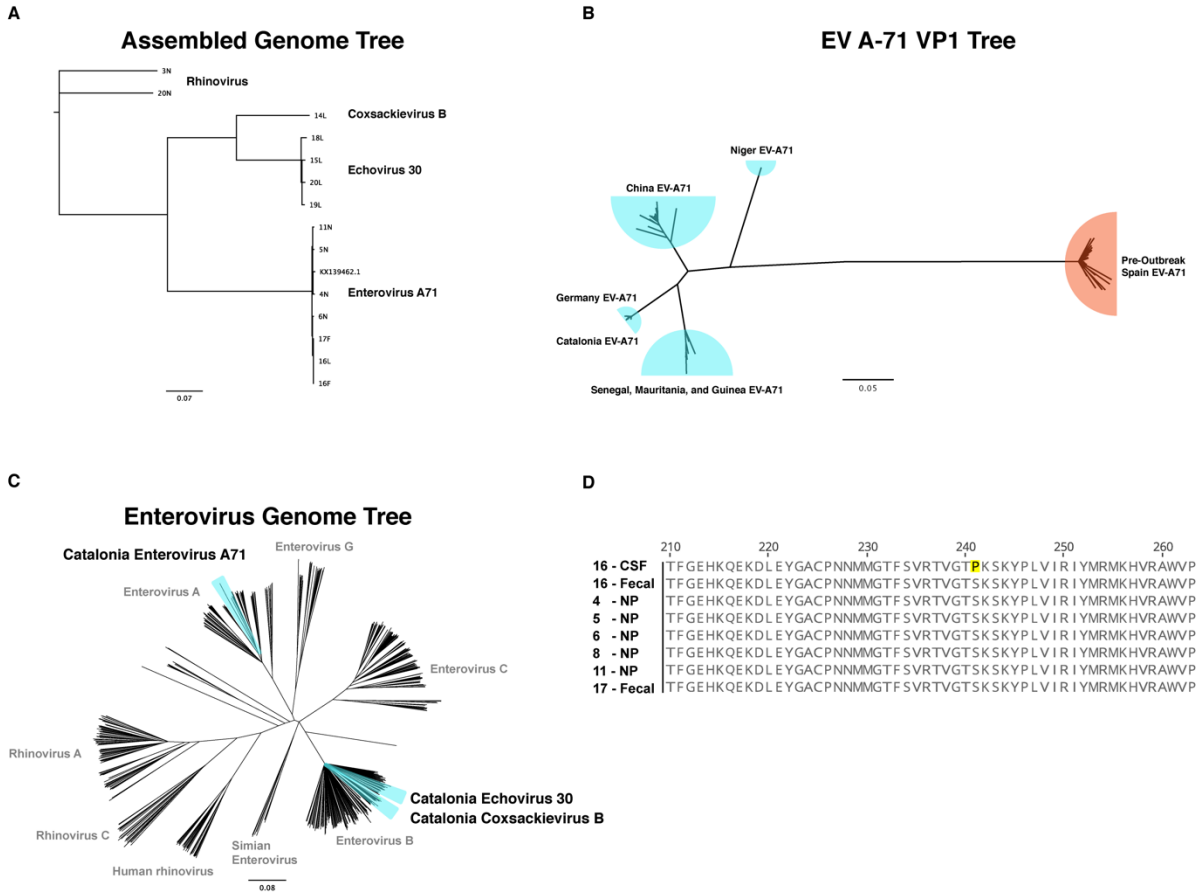


Figure 4.2: Phylogenetic and Genomic Analysis of EV-A71. A) Phylogenetic tree of full-length viral genomes for enterovirus A71, echovirus 30, Coxsackievirus B and rhinovirus isolated from CSF (L), stool (F) and NP (N) compared to the German neuroinvasive strain (KX139462.1). Rhinovirus obtained from subjects acts as the root. The number refers to the subject. **B)** Confirmation of clinical VP1 testing that the Catalonian EV-A71 Viral Protein 1 (VP1) gene is most closely related to a neuroinfectious German strain. Blue = neuroinvasive EV-A71, Orange = HFMD EV-A71. **C)** Phylogenetic tree of 545 EV genomes from every species highlighting the relatedness between the EV strains discovered in this outbreak. **D)** Protein alignment of the VP1 gene highlighting the S241P mutation found exclusively in the CSF of subject 16. Scale bars indicate nucleotide substitution rate per position. CSF = cerebrospinal fluid, NP = nasopharyngeal, F = fecal, EV-A71 = enterovirus A71

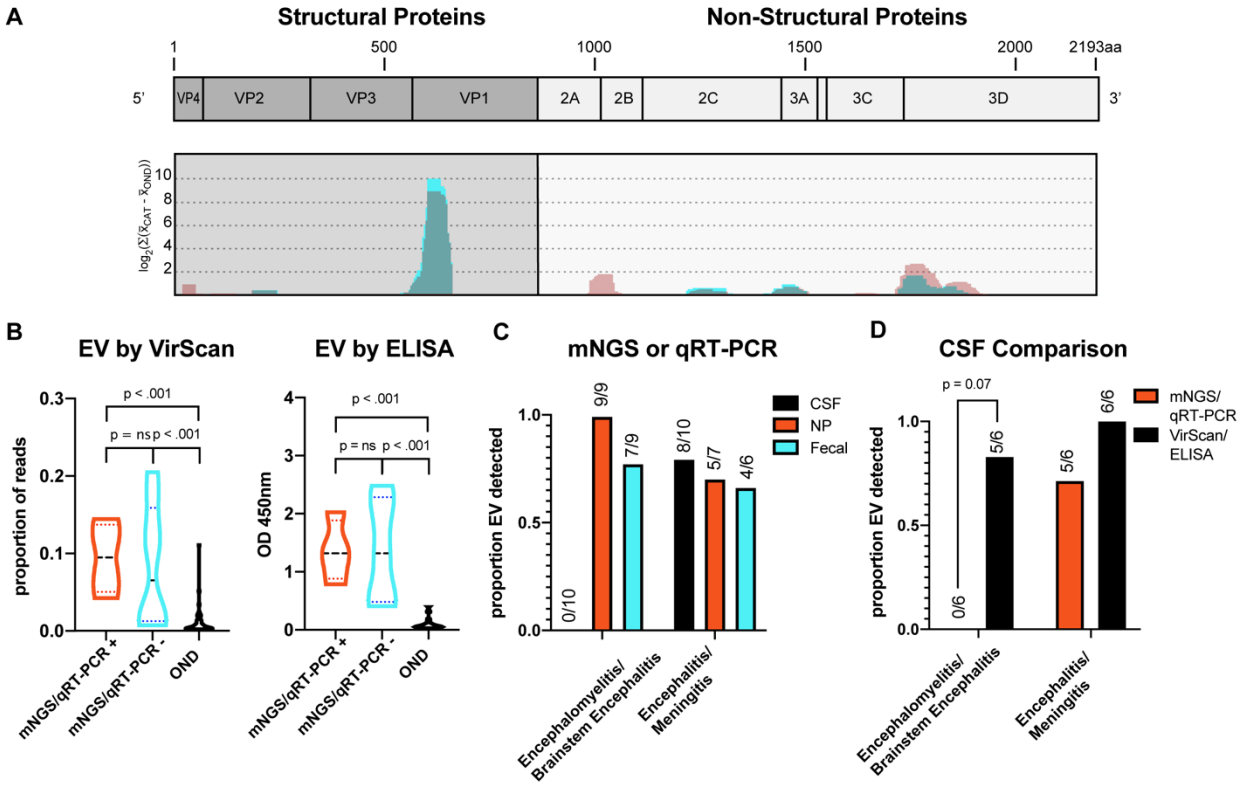
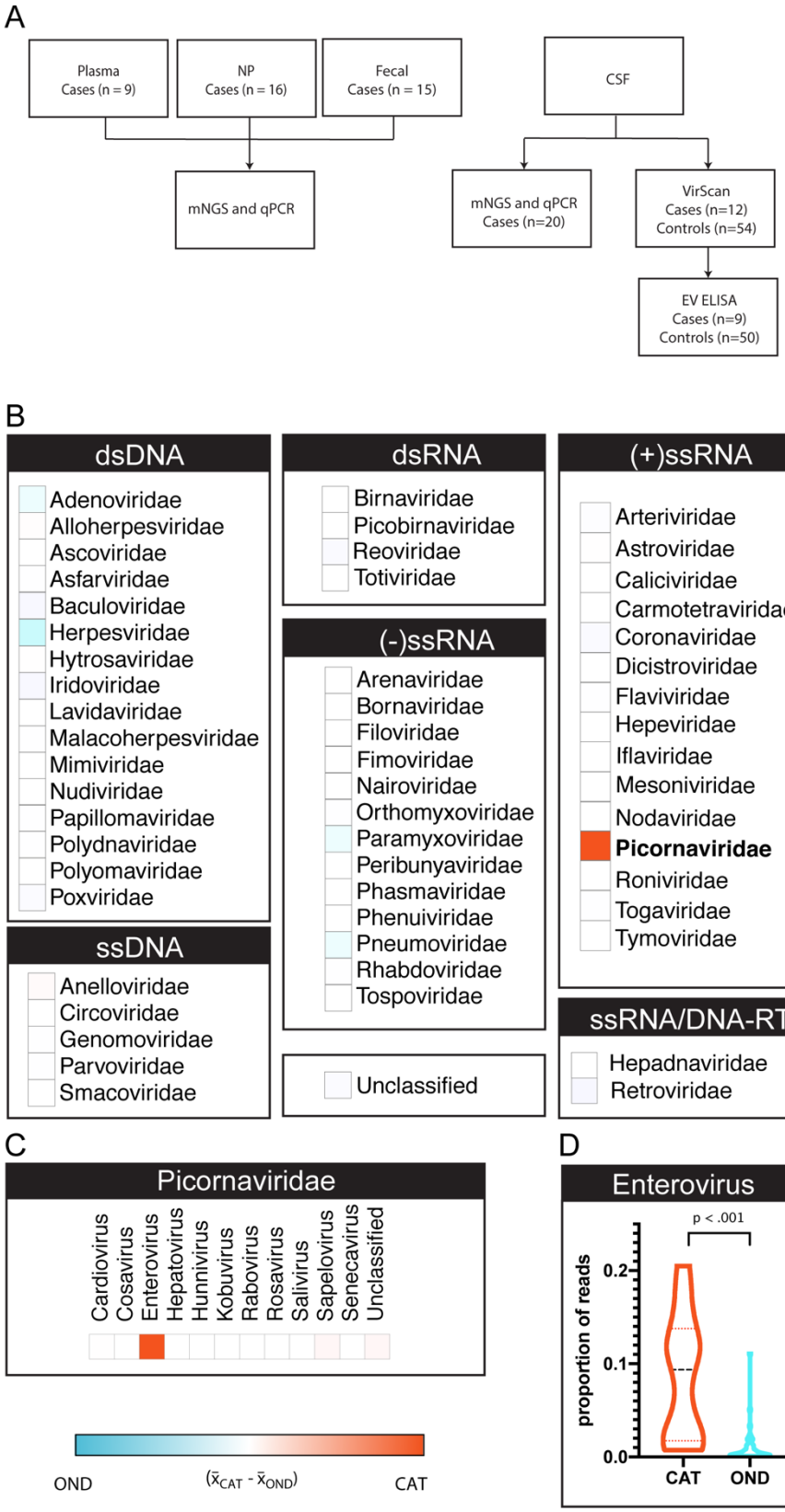


Figure 4.3: VirScan Identifies Immunodominant Enterovirus Antigens and improves the detection of CSF enterovirus in Encephalomyelitis / Brainstem encephalitis A) VirScan identified 136 unique, enriched viral antigens with taxonomies linking them to genus *Enterovirus*. We mapped 123 of these 136 peptides with BLASTP to a reference EV-A71 genome (Genbank Accession AXK59213.1) as described previously[9] (coding genes in light blue, non-coding genes in orange). Mapping revealed the relative locations of the EV antigens identified in the Catalonia cases (graphed blue shading) across the EV genome, which appeared remarkably conserved as seen previously in pediatric acute flaccid myelitis (graphed light red shading with overlap appearing grey)[10]. **B)** Violin plot revealing enrichment for EV antigens by VirScan and EV VP1 antigen by ELISA in Catalonia cases regardless of whether an EV had been previously identified by mNGS or qRT-PCR (p = ns for both comparisons). In both groups, EV detection was significantly greater than in the pediatric OND controls (p < .001 for all comparisons as indicated). The Mann-Whitney test was used, with Bonferroni correction for VirScan. **C)** Differences in virus detection for subjects with encephalomyelitis/brainstem encephalitis or encephalitis/meningitis. The detection of CSF EV by mNGS or qRT-PCR was low (0/10). **D)** VirScan or ELISA improved EV detection in encephalomyelitis/brainstem

encephalitis (0/6 versus 5/6, p = 0.07 by McNemar's test). EV = enterovirus, CSF = cerebral spinal fluid, NP = nasopharyngeal


A
Plasma
Cases (n = 9)
NP
Cases (n = 16)
Fecal
Cases (n = 15)
mNGS and qPCR
CSF
mNGS and qPCR
Cases (n=20)
VirScan
Cases (n=12)
Controls (n=54)
EV ELISA
Cases (n=9)
Controls (n=50)
B
dsDNA
Adenoviridae
Alloherpesviridae
Ascoviridae
Asfarviridae
Baculoviridae
Herpesviridae
Hytrosaviridae
Iridoviridae
Lavidaviridae
Malacoherpesviridae
Mimiviridae
Nudiviridae
Papillomaviridae
Polydnaviridae
Polyomaviridae
Poxviridae
ssDNA
Anelloviridae
Circoviridae
Genomoviridae
Parvoviridae
Smacoviridae
dsRNA
Birnaviridae
Picobirnaviridae
Reoviridae
Totiviridae
(-)ssRNA
Arenaviridae
Bornaviridae
Filoviridae
Fimoviridae
Nairoviridae
Orthomyxoviridae
Paramyxoviridae
Peribunyaviridae
Phasmaviridae
Phenuiviridae
Pneumoviridae
Rhabdoviridae
Tospoviridae
Unclassified
(+)ssRNA
Arteriviridae
Astroviridae
Caliciviridae
Carmotetraviridae
Coronaviridae
Dicistroviridae
Flaviviridae
Hepeviridae
Iflaviridae
Mesoniviridae
Nodaviridae
Picornaviridae
Roniviridae
Togaviridae
Tymoviridae
ssRNA/DNA-RT
Hepadnaviridae
Retroviridae
C
Picornaviridae
Cardiovirus
Cosavirus
Enterovirus
Hepatovirus
Hunnivirus
Kobuvirus
Rabovirus
Rosavirus
Salivirus
Sapelovirus
Senecavirus
Unclassified
OND
$(\bar{x}_{CAT} - \bar{x}_{OND})$
CAT
D
Enterovirus
p < .001
proportion of reads
0.2
0.1
0.0
CAT
OND

Supplemental Figure 4.1: VirScan Detects CSF Enterovirus Antibodies. A) Workflow for sample processing and analysis. **B**) Baltimore classification of viral families detected by VirScan. Heatmap color intensity was calculated by subtracting the mean reads per twenty-five thousand reads sequenced (rpqK) in the Catalonia cases from those in the pediatric other neurologic disease (OND) control CSF. The maximum (orange) and minimum (blue) color intensities represent +2,500 and -2,500 rpqK, respectively. **C**) Genus *Enterovirus* is enriched the strongest in the Catalonia samples. **D**) Focused comparison of the fraction of reads mapping to *Enterovirus* in the Catalonia cases and the OND control CSF. The mean, first quartile, and third quartile are indicated by horizontal lines; Mann-Whitney test corrected for multiple comparisons with a Bonferroni correction.

Table 4.1. Summary of subject group demographics and clinical data.

	Cases Total=10	Controls Total=10
Mean age (months)[a]	22.7 (18.1-31.2)	10.8 (0.9-37.5)
Sex (male)	4	6
Systemic symptoms		
Fever	10	10
Vomiting	6	4
Diarrhea	2	1
Exanthema	5	3
Enanthema	8	2
Neurologic symptoms		
Meningismus	2	3
Irritability	2	2
Lethargy	8	4
Headache	1	2
Myoclonic jerks	6	0
Tremor	5	0
Ataxia	9	0
Paresis	3	0
Nystagmus and/or strabismus	1	0
Bulbar palsy	2	0
Medullary symptoms	3	0
WHO clinical classification		
Meningitis	0	7
Encephalitis	0	3
Brainstem encephalitis	8	0
Encephalomyelitis	2	0
EV results by Clinical Pan-EV qRT-PCR (positive/total)		
CSF	0/10	5/10
Plasma	0/7	0/2
Nasopharyngeal sample	8/9	3/7
Stool	5/9	4/6

[a] Median (interquartile range).

References

1. Casas-Alba D, de Sevilla MF, Valero-Rello A, et al. Outbreak of brainstem encephalitis associated with enterovirus-A71 in Catalonia, Spain (2016): A clinical observational study in a children's reference centre in Catalonia. Clin Microbiol Infect 2017;23:874-881.
2. Leber AL, Everhart K, Balada-Llasat J, et al. Multicenter evaluation of BioFire FilmArray meningitis/encephalitis panel for detection of bacteria, viruses, and yeast in cerebrospinal fluid specimens. J Clin Microbiol 2016;54:2251-2261.
3. Böttcher, S., Obermeier, P. E., Neubauer, K., & Diedrich, S. Recombinant enterovirus A71 subgenogroup C1 strains, Germany, 2015. Emerging Infectious Diseases 2016.
4. González-Sanz R, Casas-Alba D, Launes C, et al. Molecular epidemiology of an enterovirus A71 outbreak associated with severe neurological disease, Spain, 2016. Euro Surveillance: European communicable disease bulletin 2019;24.
5. Karrasch M, Fischer E, Scholten M, et al. A severe pediatric infection with a novel enterovirus A71 strain, Thuringia, Germany. Journal of Clinical Virology 2016;84:90-95.
6. Midgley SE, Nielsen AG, Trebbien R, Poulsen MW, Andersen PH, Fischer TK. Co-circulation of multiple subtypes of enterovirus A71 (EV- A71) genotype C, including novel recombinants characterised by use of whole genome sequencing (WGS), Denmark 2016. Eurosurveillance 2017;22:30565.
7. Antona D, Kossorotoff M, Schuffenecker I, et al. Severe paediatric conditions linked with EV-A71 and EV-D68, France, May to October 2016. Euro Surveill 2016;21.
8. Launes C, Casas-Alba D, Fortuny C, Valero-Rello A, Cabrerizo M, Muñoz-Almagro C. Utility of FilmArray meningitis/encephalitis panel during outbreak of brainstem encephalitis caused by enterovirus in Catalonia in 2016. J Clin Microbiol 2017;55:336-338.
9. Dudas G, Carvalho LM, Bedford T, et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature 2017;544:309-315.

10. Schubert RD, Hawes I, Ramachandran PS, et al. Pan-viral serology implicates enteroviruses in acute flaccid myelitis. Nature Medicine 2019;25:1748-1752.
11. Xu GJ, Kula T, Xu Q, et al. Comprehensive serological profiling of human populations using a synthetic human virome. Science 2015;348:aaa0698.
12. Johnson TP, Larman HB, Lee M, et al. Chronic dengue virus panencephalitis in a patient with progressive dementia with extrapyramidal features. Ann Neurol 2019.
13. Selva L, Martinez-Planas A, García-García J-, Casadevall R, Luaces C, Muñoz-Almagro C. Comparison of an in-house real-time RT-PCR assay with a commercial assay for detection of enterovirus RNA in clinical samples. Eur J Clin Microbiol Infect Dis 2012;31:715-719.
14. Verstrepen WA, Kuhn S, Kockx MM, Van De Vyvere, M. E., Mertens AH. Rapid detection of enterovirus RNA in cerebrospinal fluid specimens with a novel single-tube real-time reverse transcription-PCR assay. J Clin Microbiol 2001;39:4093-4096.
15. Cabrerizo M, Echevarria JE, González I, de Miguel T, Trallero G. Molecular epidemiological study of HEV-B enteroviruses involved in the increase in meningitis cases occurred in Spain during 2006. J Med Virol 2008;80:1018-1024.
16. Wilson MR, Suan D, Duggins A, et al. A novel cause of chronic viral meningoencephalitis: Cache valley virus. Ann Neurol 2017;82:105-114.
17. Gu W, Crawford ED, O'Donovan BD, Wilson MR, Chow ED, Retallack H, DeRisi JL. Depletion of abundant sequences by hybridization (DASH): Using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biol 2016;17:41.
18. Wilson MR, Fedewa G, Stenglein MD, et al. Multiplexed metagenomic deep sequencing to analyze the composition of high-priority pathogen reagents. mSystems 2016;1.
19. Wilson MR, O'Donovan BD, Gelfand JM, et al. Chronic meningitis investigated via metagenomic next-generation sequencing. JAMA Neurol 2018.

20. Bankevich A, Nurk S, Antipov D, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 2012;19:455-477.

21. Greninger AL, Naccache SN, Messacar K, et al. A novel outbreak enterovirus D68 strain associated with acute flaccid myelitis cases in the USA (2012-14): A retrospective cohort study. Lancet Infect Dis 2015;15:671-682.

22. Fernandez-Garcia MD, Volle R, Joffret M, et al. Genetic characterization of enterovirus A71 circulating in Africa. Emerging Infect Dis 2018;24:754-757.

23. Cabrerizo M, Tarragó D, Muñoz-Almagro C, et al. Molecular epidemiology of enterovirus 71, coxsackievirus A16 and A6 associated with hand, foot and mouth disease in Spain. Clinical Microbiology and Infection 2014;20:O150-O156.

24. Sun L, Zheng H, Zheng H, et al. An enterovirus 71 epidemic in Guangdong province of China, 2008: Epidemiological, clinical, and virogenic manifestations. Jpn J Infect Dis 2011;64:13-18.

25. McWilliam Leitch EC, Cabrerizo M, Cardosa J, et al. The association of recombination events in the founding and emergence of subgenogroup evolutionary lineages of human enterovirus 71. J Virol 2012;86:2676-2685.

26. Simmonds P, Welch J. Frequency and dynamics of recombination within different species of human enteroviruses. J Virol 2006;80:483-493.

27. Pérez-Vélez CM, Anderson MS, Robinson CC, et al. Outbreak of neurologic enterovirus type 71 disease: A diagnostic challenge. Clin Infect Dis 2007;45:950-957.

28. Messacar K, Spence-Davizon E, Osborne C, et al. Clinical characteristics of enterovirus A71 neurological disease during an outbreak in children in Colorado, USA, in 2018: An observational cohort study. Lancet Infect Dis 2019.

29. Kohn MA, Carpenter CR, Newman TB. Understanding the direction of bias in studies of diagnostic test accuracy. Acad Emerg Med 2013;20:1194-1206.

30. Böttcher S, Obermeier PE, Neubauer K, Diedrich S. Recombinant enterovirus A71 subgenogroup C1 strains, Germany, 2015. Emerging Infectious Diseases.

31. Chang C, Wu S, Chen Y, et al. Mutations in VP1 and 5'-UTR affect enterovirus 71 virulence. Scientific reports 2018;8:8744-1.

32. Nishimura Y, Lee H, Hafenstein S, Kataoka C, Wakita T, Bergelson JM, Shimizu H. Enterovirus 71 binding to PSGL-1 on leukocytes: VP1-145 acts as a molecular switch to control receptor interaction. PLoS pathogens 2013;9:e1003511.

33. Lin J, Shih S. Cell and tissue tropism of enterovirus 71 and other enteroviruses infections. J Biomed Sci 2014;21:18.

34. Nishimura Y, Shimojima M, Tano Y, Miyamura T, Wakita T, Shimizu H. Human P-selectin glycoprotein ligand-1 is a functional receptor for enterovirus 71. Nat Med 2009;15:794-797.

35. Mishra N, Ng TFF, Marine RL, et al. Antibodies to enteroviruses in cerebrospinal fluid of patients with acute flaccid myelitis. MBio 2019;10.

36. Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med 2014;370:2408.

37. Naccache SN, Peggs KS, Mattes FM, et al. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. Clin Infect Dis 2015;60:919-923.

38. Miller S, Naccache SN, Samayoa E, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. Genome Res 2019;29:831-842.

39. Wilson MR, Sample HA, Zorn KC, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. N Engl J Med 2019;380:2327-2340.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
kristoffer leon
688AE23B744E473...
Author Signature

12/17/2020
Date