Integrative analysis of Zika virus genome RNA structure reveals critical determinants of viral infectivity

Pan Li[1,7], Yifan Wei[1,7], Miao Mei[2,7], Lei Tang[1,7], Lei Sun[1,7], Wenze Huang[1], Jianyu Zhou[3], Chunlin Zou[4], Shaojun Zhang[1], Cheng-feng Qin[5], Tao Jiang[6,3], Jianfeng Dai[4], Xu Tan[2,*], Qiangfeng Cliff Zhang[1,8,*]

[1] MOE Key Laboratory of Bioinformatics, Beijing Advanced Innovation Center for Structural Biology, Center for Synthetic and Systems Biology, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

[2] School of Pharmaceutical Sciences, Center for Infectious Disease Research, School of Medicine, Tsinghua University, Tsinghua-Peking Center for Life Sciences, Beijing 100084, China

[3] Bioinformatics Division, BNRIST; Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China.

[4] Institutes of Biology and Medical Sciences, Jiangsu Key Laboratory of Infection and Immunity, Soochow University, Suzhou, China.

[5] Department of Virology, State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China.

[6] Department of Computer Science and Engineering, University of California, Riverside, CA 92521, USA.

[7] Co-first author

[8] Lead Contact

*Correspondence: xutan@tsinghua.edu.cn (X.T.), qczhang@tsinghua.edu.cn (Q.C.Z.)

Running title: RNA secondary structure regulates Zika virus infection

Keywords: RNA secondary structure, Zika virus

SUMMARY

Since its outbreak in 2007, Zika virus (ZIKV) has become a global health threat that causes severe neurological conditions. Here we perform a comparative *in vivo* structural analysis of the RNA genomes of two ZIKV strains to decipher the regulation of their infection at the RNA level. Our analysis identified both known and novel functional RNA structural elements. We discovered a functional long-range intramolecular interaction specific for the Asian epidemic strains, which contributes to their infectivity. Further, we developed a computational screening strategy that uses our RNA structural information to accurately predict host factors of ZIKV. We validated the predicted binding of the host protein ELAVL1 to ZIKV and showed that it promotes infection. Our findings illuminate the structural basis of ZIKV regulation and provide a rich resource for the discovery of RNA structural elements and host factors that are important for ZIKV infection.

HIGHLIGHTS

- Integrative probing *in vivo* generates comprehensive structural maps of ZIKV RNA genomes in infected cells
- Comparative analysis revealed known and novel functional RNA structural elements in ZIKV genomes
- A long-range intramolecular interaction specific for the epidemic Asian ZIKV strains affects infectivity in cell lines
- Structural-based computational screening accurately predicts the RBP host factors of ZIKV in mammalian cells
- The host protein ELAVL1 binds to ZIKV RNA and promotes viral infection

INTRODUCTION

Zika virus (ZIKV) is a mosquito-borne single-stranded RNA virus of the *flaviviridae* family (1). The virus was first isolated in Uganda in 1947 and had remained obscure until its outbreak in Micronesia in 2007. Since then it has caused epidemics in Pacific islands and later in America and Asia. In response to its rapid spread and association with serious brain disorders, including microcephaly in newborns and Guillain-Barre syndrome in adults, the World Health Organization has declared it a public health emergency and tremendous efforts have been devoted to understanding its pathogenesis and to developing effective treatments (2,3,4). However, there remains no specific therapy or approved vaccine for ZIKV infection.

The ZIKV genome is an approximately 10.8 kb positive-sense RNA in an mRNA-like pattern (5). ZIKV strains can be classified into the ancestral African lineage and the contemporary Asian lineage based on genome sequences, with the latter responsible for the current epidemics. To understand how the Asian strains lead to epidemics and illness, a crucial step is to identify genomic variations that affect their infectivity and pathogenicity. Comparative studies have focused on mutations that change protein sequences. For example, it has been proposed that a key amino acid substitution S139N in the prM protein contributes to fetal microcephaly (6), and another mutation A188V in the NS1 protein promotes transmissibility in mosquito vectors (7). However, most genomic variations between the two lineages are synonymous or non-coding, and thus are not expected to affect the encoded proteins. We analyzed a set of representative ZIKV strains and found on average 1, 166 mutations between the Asian strains and the African strains. Among them, about 88%, i.e., 1,020 mutations, are synonymous or in non-coding UTR regions (Supplementary Figure S1A, S1B and S1C). Whether and how these mutations could also contribute to ZIKV infection remain to be answered.

It is well-known that the genomic RNA of flaviviruses participates in viral processes including translation, replication, packaging and evasion of host cell antiviral responses (8). RNA structural elements are functionally involved in many of these processes. For example, conserved multi-pseudoknot structures in the 3' UTR of ZIKV and other flaviviruses can stall the RNA exonuclease Xrn1, thereby giving rise to sub-genomic flavivirus RNAs (sfRNAs) that help the virus evade cellular antiviral processes (9, 10). In addition, an intramolecular RNA-RNA pairing between the 5'- and the 3'-UTRs facilitates transformation between the linear and circular conformations of the genomic RNA, and thus plays an important in role coordinating virus replication (11). To date, our knowledge of ZIKV RNA structure is mainly

limited to the untranslated regions. However, the coding region constitutes over 95% of the ZIKV genome and likely contains a wealth of functional structural elements yet to be discovered. Uncovering these structural elements and their differences between the two lineages may reveal a molecular rationale for the outbreak of epidemics.

Next-generation sequencing-based technologies enable the profiling of the complete genome structures of RNA viruses. A pioneer study of the HIV RNA genome structure discovered known and unrecognized regulatory motifs, as well as a higher-order organizational principle that RNA structures directly encode protein domain junctions (12). Other studies examined multiple HCV genome structures and found various structural regulatory elements across the whole genome including coding regions (13, 14). The studies suggested that the viral RNA structures may have evolved into a sophisticated complex network that protects the genome from both RNase L and double-stranded RNA-induced innate immune sensors. And conformational changes within these structural motifs can influence viral replication or immune evasion. However, all these studies are based on *in vitro* experiments.

Here we investigated RNA secondary structures and intramolecular RNA-RNA interactions of two ZIKV RNA genomes *in vivo* by combing two orthogonal high throughput sequencing based technologies, icSHAPE (15) and PARIS (16) (Figure 1). We generated maps of the secondary structures of two ZIKV strains representing the Asian and African lineages, with many novel long-range intramolecular RNA-RNA interactions including common and lineage-specific structural elements. We identified a functionally important long-range intramolecular interaction that is specific for the Asian strains after 2007 and regulates their infection in certain cell lines. Finally, we predicted many potential ZIKV host protein factors, and showed that one, the RNA binding protein (RBP) ELAVL1, binds ZIKV RNA and promotes viral replication. Therefore, we provide a rich resource for understanding the structure-function relationship of ZIKV genomic RNA and the interactions between ZIKV RNA and human RBPs.

RESULTS

icSHAPE defines nucleotide flexibility profiles of ZIKV genomes

For the RNA structural determination, we chose PRVABC59 (17) as a representative of the Asian strains and MR766 (18) for the African strains according to their positions in the ZIKV phylogenetic tree (Supplementary Figure S1A). We noticed that the PRVABC59 reference genome from GenBank misses a conserved functional small hairpin in the 3'UTR,

so we re-assembled the genomes with our sequencing data and completed the sequences. The two ZIKV genomes displayed 6.5% and 3.9% sequence variation within the 5'- and 3'-UTR regions, respectively, and 11.3% variation within the polyprotein coding region (Supplementary Figure S1D). The elevated sequence conservation within the UTRs relative to the coding regions reflects the importance of known functional RNA elements in flavivirus UTRs.

First, we performed *in vivo* click selective 2-hydroxyl acylation and profiling experiments (icSHAPE) to measure the structural flexibility of every nucleotide within the PRVABC59 and MR766 RNA genomes in infected cells (Figure 2). We treated ZIKV-infected Huh7 cells with the icSHAPE reagent NAI-N3, which preferentially reacts with unstructured and flexible nucleotides. We then purified the modified RNA and performed reverse transcription. NAI-N3-modified bases block reverse transcriptase, yielding cDNA fragments that end at the modified site. We performed deep sequencing and computational analyses to map the reverse transcription termination sites, generating a flexibility or icSHAPE reactivity score for each nucleotide. Flexibility negatively correlates with the likelihood of secondary structure, thus providing a measure of the pairing probability of each nucleotide

We obtained icSHAPE reactivity scores for over 99.6% of the nucleotides within the two ZIKV genomes (Figure 2A and 2B, Supplementary Table S1). The icSHAPE scores were reproducible between independent biological replicates (R≥0.89). The structural conservation was significant and contains many well-known functional elements in flaviviruses (Figure 2C and 2D). Consistent with previous findings, we noticed that the structural conservation by icSHAPE scores was lower than the degree of the sequence conservation between the two strains (R=0.52; Figure 2E). Five RNA elements in the 3'-UTR, including Stem-loop1 (SL1), Stem-loop2 (SL2), Dumbbell1 or pseudo-Dumbbell (DB1), Dumbbell2 (DB2) and the small hairpin 3' Stem-loop (sHP-3'SL), are conserved structures and established benchmarks for RNA folding prediction (10). All these elements, except sHP-3'SL, are at the 5' end of sfRNAs and may function as Xrn1 RNase resistant components in ZIKV (9). At the 5'-end of the viruses, three stem loops, including stem loop A (SLA), stem loop B (SLB) and capsid hairpin (cHP), are defined from previous SHAPE and RNase digestion studies in Dengue virus (19). The DCS-PK element in the 5' UTR flanking region, is a novel pseudoknot found in coding region (20). SLA has been proposed to function in viral RNA synthesis, whereas SLB and cHP form base pairs with the 3' end for genome cyclization and the DCS-PK element helps enhance genome cyclization during replication (20). Generally, unpaired nucleotides within these

structural elements had higher icSHAPE scores, whereas paired nucleotides had lower icSHAPE scores, demonstrating that icSHAPE accurately measured ZIKV genomic RNA structures in infected cells (AUC=0.83, Figure 2F). The known long-range interaction between the 5'UTR and 3'UTR was also generally consistent with the icSHAPE scores. However, the small stem of 5'-3' CS was not reflected in the icSHAPE scores, with one exception, pointing to some limitation of icSHAPE in detecting long-range RNA-RNA interactions.

We also used icSHAPE to measure *in vitro* ZIKV RNA structures. RNA was extracted from ZIKV-infected Huh7 cells, refolded in an icSHAPE modification buffer for RNA structure stability, and then modified with NAI-N3 (15). The remaining steps and data analysis were the same as the *in vivo* icSHAPE measurement. We observed moderate correlations between *in vivo* and *in vitro* viral RNA structures (R=0.75 for both PRVABC59 and MR766), similar to that of eukaryote RNAs in previous studies (15). Our data confirmed that RNA structures *in vivo* are generally more open than *in vitro*. Our *in vitro* icSHAPE data displayed lower agreement with the canonical 3'UTR RNA structure models than the *in vivo* data (AUC=0.74 vs. AUC=0.83), indicating that previous studies have recovered the *in vivo* conformations of many functional structural elements. Importantly, the difference between our *in vitro* and *in vivo* results implies that viral RNAs adopt distinct conformations in infected cells, highlighting the importance of studying RNA structures in their cellular context to uncover biologically relevant conformations.

PARIS uncovers RNA-RNA interactions within ZIKV genomes

The RNA structural flexibility measured by icSHAPE scores represents the probability that a nucleotide is in a non-pairing, or single-stranded form. To directly map the intramolecular RNA-RNA parings in ZIKV genomes, we used the newly developed PARIS method (Psoralen Analysis of RNA Interactions and Structures, Figure 1), which globally determines RNA duplex structures in live cells via reversible psoralen crosslinking (16). ZIKV-infected Huh7 cells were treated with the psoralen derivative AMT, and crosslinked RNA duplexes were purified by 2-D gel separation, proximity ligated and resolved by sequencing and bioinformatics analysis. An intramolecular RNA-RNA interaction is defined by an alignment of gapped reads that can be mapped to the two stems of the duplex.

We discovered a large number of RNA-RNA interactions or pairing structures for both viruses by PARIS (Figure 3A and Supplementary Figure S2A). The PARIS data were reproducible between independent biological replicates (R=0.98), clustered into 1482 duplexes

for PRVABC59 and 1282 duplexes for MR766 (Supplementary Table S2). Among these duplexes, we defined those with pairing distance of more than 1kb as long-range interactions, resulting in 230 for PRVABC59 (16% of all duplexes) and 178 for MR766 (14% of all duplexes) (Supplementary Table S2). Among the 15 short-range interactions of known local structures in the 3'UTR and 5'UTR with flanking regions, 10 were captured by PARIS, with the number of supporting reads ranging from 2 to 91 (data from the PRVABC59 experiment, Figure 2F). Only a few long-range interactions have been previously reported; among them is the interaction between the 5'UTR and the 3'UTR, which is highly conserved in flaviviruses and essential for their genome cyclization and replication (21). These interactions, including 5'-3'UAR, 5'-3'DAR and 5'-3'CS, are well-recovered in our results (Figure 2F).

We noticed that short-range interactions are more conserved than long-range ones between the two ZIKV strains (Supplementary Figure S2B). We found that MR766 and PRVABC59 share over one third of short-range interactions, but less than 10% of long-range interactions. PARIS relies on psoralen intercalating in RNA helices and needs to sequence into very deep because its limited crosslinking and ligation efficiency (16). Our approach might underestimate the conservation of long-range interactions due to the limited coverage of PARIS, but this substantial difference still suggests lower evolutionary constraints for long-range than for short-range interactions.

To evaluate the plausibility of the detected long-range interactions, we analyzed their pairing energy. We first carried out a permutation test to determine the likelihood of an RNA pairing structure. For every PARIS long-range interaction, we calculated the pairing energy, using the *bifold* program from software suite *RNAstructure* with default parameters (22,23). Then we permutated the sequence in both stems 100 times, calculated the minimal fold free energy each time and compared the true pairing energy with the averaged energy of 100 permutated ones. We found that most of the PARIS interactions are more stable than the permuted ones (p-value=6.2E-64 for PRVABC59 and p-value=1.5E-45 for MR766, pairwise t-test. Supplementary Figure S3A). This analysis demonstrates the specificity of PARIS experiments. We also compared the folding energies of the long-range interactions discovered by PARIS with putative interactions between homologous regions of the other virus strain. Briefly, for each stem of a duplex defined by PARIS in one strain, we located their homologous regions in the other strain, and defined a putative homologous interaction by pairing the two homologous regions. We calculated the corresponding folding energies and found that most of the strain-specific PARIS interactions are much more stable than putative homologous ones

(p-value=7.9E-24 for PRVABC59-specific long-range interactions, and p-value=2.8E-18 for MR766-specific ones, pairwise t-test. Supplementary Figure S3B), whereas the common interactions displayed similar folding energies (p-value=0.64, pairwise t-test, Supplementary Figure S3B). Together, our analyses suggest that many of the PARIS long-range interactions are energetically favorable. We highlighted some of the most stable long-range interactions in Figure 3A and Supplementary Figure S2A (labeled as **a**~**d** and **known**), including the well-known 5'-3'CS interaction. We hypothesize that some of the other long-range interactions are also important in virus infection.

Architecture and structural models of ZIKV genomes

Interestingly, we observed that the intra-molecular interactions form clusters in the PARIS data heatmap (Figure 3A and Supplementary Figure S2A), similar to clusters found in genome connectivity maps from Hi-C data. This inspired us to define structural domains in each of the two ZIKV genomes. We implemented an algorithm that identifies clusters with dense intra-domain interactions. The algorithm searches for an optimal position to recursively split an RNA sequence into two disjointed domains. In each iteration, the position is chosen to maximize the difference (measured by the earth mover's distance (24)) between the distribution of interactions in the current domain and that in the two new sub-domains. This yields a hierarchical domain partition, and the top *k* domains in the hierarchy that minimize the maximal coefficient of variation of intra-domain interactions are selected as the output. This algorithm split the structure of the rRNA 18S into 4 domains and 28S into 6 domains, with the domain boundaries accurately demarcated for the two rRNAs (Supplementary Figure S4A and S4B).

We applied the method to define 23 structural domains for MR766 and 24 domains for PRVABC59 (Figure 3A and Supplementary Figure S2A). As expected, most PARIS interactions are clustered into intra-domain pairing structures. The enrichment ratio of intra-domain PARIS signal density over the background is about 32.3 for MR766 and 28.4 for PRVABC59 (Supplementary Figure S4C). The overall architectures of the two ZIKV genomic RNAs agree well, measured by the conservation of domain boundaries (13 boundaries are in common, p-value<0.001, permutation test. Supplementary Figure S4C).

The single-strandedness measured by icSHAPE and the intramolecular interactions measured by PARIS provide two complementary sets of information on RNA structures. We noted that icSHAPE scores are usually very low for PARIS local structures, suggesting good agreement between the data. In addition, we used software suite *RNAstructure* (25) *to* predict

a local minimun free-energy (MFE) structure with icSHAPE scores as a constraint (Figure 3B and Supplementary Figure S2C). We compared the base-pairing in the icSHAPE-assisted predicted structure with those measured by PARIS and found substantial overlap (p-value<0.001, permutation test, Figure 3B and Supplementary Figure S2C). However, it is worth noting that the agreement between icSHAPE and PARIS is lower for long-range interactions, which may reflect that many long-range interactions are transient, switching their conformations to open states under different conditions.

By demarcating ZIKV genomes into small domains, we expect to accurately build their secondary structural models, as predictions have been successful for small RNAs especially when constrained with experimental probing data (25,26). Here we used the software suite *RNAstructure* (25) to predict the secondary structure for each domain separately, with icSHAPE scores as a constraint. We used the parameter that predicts the structure of 3'UTR most accurately for other domains (AUC=0.83 for PRVABC59 and AUC=0.80 for MR766). We verified the validity of our method by assessing the predictive performance on the 5'UTR (AUC=0.95 for PRVABC59 and AUC=0.92 for MR766). We also used a similar pipeline from a previous study of the HCV structures combined with a statistical tool R-scape to call covariant base pairs with 4,256 Flaviviridae virus genomes (15, 27).

The method precisely reproduced the known secondary structural elements of the 5'-UTRs and 3'-UTRs within the ZIKV genomes (compare Figure 4, Figure S5 and Figure 2F). Most stems in these structural models demonstrate some covariation, suggesting that they are functionally conserved. In total, one RNA element in the 5'UTR domain (SLA) and four elements in the 3'UTR domain (SL1, SL2, DB2 and 3'SL) contain 71 co-variations in about 420 nucleotides (Figure 4). We did not find any covariation for one structural element, DB1 in the 3'UTR domain. Indeed, some flaviviruses, like YFV, lack DB1 (28). The coding region, as expected, contains fewer evolutionary covariations (170 in about 10k nucleotides). However, the structural models do reveal some conserved structural elements in the coding region with covariations. For example, the capsid hairpin (cHP) in the coding region of the protein capsid contains 5 covariations. The cHP is involved in the translation and replication of flaviviruses, including DENV and WNV. Although the cHP hairpin structure is conserved, its sequence is quite flexible (29). The functions of other structural elements remain to be elucidated (Figure 4). Overall, our structural analysis with icSHAPE and PARIS provided reliable structural models of two ZIKV genomes, and identified known and novel structural elements (Figure 4, Figure S5).

A lineage-specific long-range interaction in epidemic ZIKV

The structural landscapes of the ZIKV genomic RNAs allowed us to investigate whether and how RNA structural elements may affect the infectivity of ZIKV. We hypothesized that lineage-specific structural elements might underlie outbreaks of epidemic strains. To search for lineage-specific structural elements, we first compared the structural flexibilities of the MR766 and the PRVABC59 strains from the icSHAPE data. We found 132 short regions that contain at least three nucleotides with substantially different icSHAPE scores (larger than 0.6, in a sliding window of 5 nucleotides). For each structurally distinct region, we defined it to be lineage-specific if the encoding sequence is conserved within each lineage but not in both. The analysis revealed a set of 68 elements (Supplementary Table S3); many of them could be the basis or readout of lineage-specific RBP binding, RNA modification, etc. For example, the list contains a previously identified Musashi1(MSI1) protein binding site at the 3'-SL in the 3'UTR (30). It has been reported that one base substitution in this site disrupts its binding to MSI1 in African strains (30). Indeed, this substitution causes a structural change in MR766 and may render the binding with MSI1 energetically unfavorable.

Similarly, we defined lineage-specific intramolecular interactions from the PARIS results (271 short-range interactions for PRVABC59, 165 for MR766; and 127 long-range interactions for PRVABC59, 73 for MR766, Supplementary Table S4). Among these, we noticed a striking long-range interaction between the 5' UTR (2-43nt) and the E protein coding region (1089-1134nt) in the PRVABC59 strain but not in the MR766 strain (Figure 5A). Although the 5' UTR is well-conserved between the two strains, ten nucleotides in the E protein coding region are mutated to form this 5' UTR-E interaction in PRVABC59, consisting of 37 base-pairs (Figure 5B and 5C). Phylogenetic analysis revealed a sharp separation of strong conservation of these base pairs in all epidemic strains, but not in pre-epidemic Asian strains nor any African strains. Interestingly, all ten mutations within the E protein coding region are synonymous to maintain the same amino acid sequence at this region in the E protein of all strains (Figure 5B).

To investigate the function of this 5'UTR-E interaction in epidemic ZIKV, we used a full-length Asian strain GZ01, for which we have an infectious clone available for mutagenic analysis (31). The sequence of GZ01 is identical to that of PRVABC59 in the 5'UTR-E interaction region. We constructed mutants with synonymous mutations that disrupt base pairing without changing the E protein sequence in GZ01 (Figure 5C, G1095A+C1101A *mut1* or U1116C *mut2*). We also engineered GZ01 mutants with the compensatory mutation in the

5’ UTR to restore the base pairing (Figure 5C, C36U+G29A *res1* for *mut1*, A18G *res2* for *mut2*). As a control, we included a non-replicating mutant ZIKV that lacks NS5 RNA dependent RNA polymerase function (“GAA”) (31). We transfected BHK21 cells with *in vitro* transcribed wild-type or mutant ZIKV RNA and harvested ZIKV particles in supernatant. To assess the full life cycle of ZIKV, the viral particles were used to infect C6/36 cell line the infectivity was assessed by qPCR and immunofluorescence. Similar to the GAA mutation, *mut1* and *mut2* substantially reduced GZ01 infectivity. Importantly, the compensatory mutations *res1* and *res2* that restore the 5’UTR-E interaction partially restored the infectivity in C6/36 cells (Figure 5D-F). The reductions in infectivity are unlikely due to codon effects on protein translation, since we avoided rare codons. What’s more, since the effect can be rescued by distant mutations in the primary sequences of the 5’ UTR region, strongly suggesting the functional importance of the long-range interaction.

The 5’ UTR of ZIKV RNA participates in the regulation of virus translation and replication through different pathways. It can initiate virus translation via a cap-dependent mechanism. For virus replication, the local structural element SLA functions as a promoter for the viral polymerase NS5. Another element SLB also facilitates replication by cyclizing the RNA genome via formation of a 5'-3' complementary structure (5'-3' CS) with the 3' UTR (32) (Figure 5G). The switch between the translation and replication conformations is important for ZIKV infection (21). Our PARIS experiment identified a third conformation of the ZIKV 5’UTR, paring with the E protein coding region to form a 5'-E protein complementary structure (5'-E CS) (Figure 5G) , which may play a role in virus replication or translation regulation.

Prediction of host factors that bind to ZIKV RNA

In parallel to structural differences across strains that can be used to elucidate strain or lineage-specific functions, *in vivo* versus *in vitro* structural differences can facilitate the discovery of host regulators of ZIKV (15). We designed a machine learning model that uses a random forest algorithm and structural information for RBP binding site prediction (Supplementary Table S5). First, we collected binding sites of 16 RBPs from published CLIP-seq experiments conducted in two cell lines, 293T and K562. We obtained binding sequence motifs for the RBPs from the existing databases ATtRACT if available, and otherwise called motifs with the MEME software tool (33). Finally, we used the 16 sequence motifs to search for matches on ZIKV genomes and defined a set of binding sites, each containing a motif match. We also obtained icSHAPE structure profiles in the 293T cell line to train a model to learn the structural pattern of the binding sites for each RBP separately, with both *in vivo* and *in vitro*

icSHAPE data as input (15). We validated the prediction accuracy of the machine learning model by cross-validation and found that the algorithm offers fairly accurate prediction for all RBPs evaluated on testing datasets (All AUCs > 0.75) (Supplementary Figure S6B).

We then applied the models to predict RBP binding on ZIKV genomes. We first used the 16 sequence motifs to scan the ZIKV genomes and obtained 274 matches for 16 proteins. We then used the machine learning models and ZIKV icSHAPE scores to distinguish the true and the false bindings within these sequence matches. The strategy identified 59 potential binding sites for 11 RBPs (Figure 6A, Supplementary Figure S6C, Supplementary Table S5). Among the predicted ZIKV interacting proteins, SND1 and DDX3X interact with dengue virus RNA directly (34, 35), FMRP expression is upregulated by the invasion of tick-borne encephalitis virus (36), ELAVL1 and IGF2BP1 are critical factors in HCV replication (37, 38), and CPSF6 and TARDBP are involved in the life cycle of HIV (38,39). The list also includes a few stress granule related proteins, IGF2BP1, IGF2BP3 and ELAVL1 (Figure 6A) (40,41). The stress granule is often hijacked by flaviviruses including ZIKV; the discovery of multiple stress granule proteins corroborates the functional relevance of the predictions (42).

We focused on the single-stranded RBP ELAVL1 as a potential novel host factor for ZIKV infection to validate our prediction. To validate binding regions, we designed 17 RNA oligos of 33-39 nt for ELAVL1 pull-down (P1-Pn, predicted positive interactors, and N1-Nn, predicted negative interactors in Supplementary Figure S6D, Supplementary Table S6). In the selection of pull-down sequences, we carefully chose ZIKV genomic regions for which the full-length RNA and the short RNA fragments had the same predicted structure. The pull-down assay showed significantly higher binding intensity with the sites that are predicted to be bound by ELAVL1 in our model (Figure 6B, 6C and 6D). The fairly high accuracy of the prediction algorithm highlights the importance of RNA structures in RBP binding.

To further investigate the functional modulation of ZIKV infection by ELAVL1, we utilized the U87MG glioblastoma cell line, a model of neural progenitor cells widely used for ZIKV infection. ZIKV SZ01 genomic RNA co-localized and co-immunoprecipitated with ELAVL1 in U87MG cells (Figure 6E, 6F). We found that knockdown of ELAVL1 resulted in decreased viral replication (Figure 6G-I), whereas overexpression of HA-ELAVL1 increased viral replication (Figure 6J-L) in U87MG cells. Thus, ELAVL1 is a host factor that promotes ZIKV infection, likely via a direct interaction with ZIKV RNA.

DISCUSSION

As an exogenous RNA in host cells, ZIKV genomes participate in many infection processes, presenting multiple levels of gene regulation. Studies have now started to reveal the complexity and significance of viral RNA function and regulation in RNA virus infection. For example, it has been shown that m^6A modifications on the genomes of ZIKV and other flaviviruses modulate viral RNA metabolism in cells (43, 44). RNA structure can also reveal molecular properties and functions of viral RNAs (13,45). Previous studies have uncovered RNA virus genome structures for HIV (12) and HCV (13, 14). These studies identified many structural elements with known or uncharacterized regulatory functions, as well as global organizational principles that are important for viral processes. However, the studies are based on *in vitro* or *in viron* chemical probing, which do not completely reflect the complex regulation of RNA structures in infected cells. For ZIKV, due to the lack of global structural information, most studies have been focused on a limited set of RNA structures in the UTR regions. Here we report the first analysis that comprehensively revealed *in vivo* RNA structural regulation of ZIKV.

We combined two *in vivo* methods and obtained two complementary sets of ZIKV genome structural information with a high degree of agreement. Similar to the organization patterns of protein and DNA genome structures, we discovered structural domains in the ZIKV RNA genomes, with a large number of intra-domain interactions and a few inter-domain interactions. The domain organization was conserved between the two ZIKV genomes. Our data and previous studies showed that RNA secondary structures display more variability than RNA sequences (13). Whether the domain organization is more conserved than simple RNA secondary structures, and its relevance to virus regulation, are interesting open questions.

We built structural models for each domain utilizing the *in vivo* probing data as constraints. The good agreement with existing, well-accepted models demonstrates the accuracy and reliability of our modeling algorithm. We did notice a few differences between our models and existing ones. For example, the SL1 structure at the 3' UTR is slightly different in the two models (compare Figure 2F and Figure 4). However, this is possibly due to the *in vivo* and *in vitro* differences between RNA structures, as our experiments are done *in vivo* and previous models combined information from *in silico* predictions and experiments including *in vitro* data. Our models of ZIKV viral genomes *in vivo* are more relevant to the virus life cycle in cells.

We set out to use these RNA structural models to understand the differences in virus infectivity and pathogenicity between the African strains and the epidemic Asian strains. As an RNA virus with a high mutation rate, ZIKV RNA genomes have accumulated many nucleotide variations, some of which account for these functional differences. Conventional approaches evaluate protein differences (6,7). Here, we focused on synonymous and non-coding mutations that change RNA structures without affecting protein sequences. We developed an analysis to identify lineage-specific RNA structural elements. We hypothesized that some of these lineage-specific structural elements could contribute to virus infectivity distinctions in different lineages and strains. We discovered an Asian lineage-specific long-range interaction between the 5' UTR and E protein coding region, and confirmed its significance in ZIKV infection through mutagenesis and rescue experiments. The detailed molecular mechanism of how this interaction contributes to ZIKV infection remains to be investigated. Nevertheless, the substantial changes in infectivity arising from a single nucleotide change exemplifies the important function of many structural elements in ZIKV genomes. Further characterization of other lineage-specific RNA structural elements may reveal additional RNA features that contribute to the viral infectivity and pathogenicity of epidemic strains.

Due to the lack of analytic tools, RNA structural studies have been mainly focusing on local structural elements. However, studies utilizing new techniques, including PARIS, SPLASH and LIGR-Seq, have started to reveal pervasive and dynamic long-range RNA interactions (16, 46, 47). Some of these long-range interactions are conserved among different species (16). They can play essential roles in lncRNA modular organization and mRNA translation (46). Using PARIS, we for the first time systematically revealed many long-range interactions in ZIKV. We hypothesize that, in addition to the well-known 5'-3' CS and our newly confirmed 5'-E CS, many of the other interactions we identified could be functional. Comparative analysis of the two ZIKV strains revealed that most long-range interactions in cells are not shared by the two species. In addition to the possibility of limitations in coverage, it is also possible that many of these are specific interactions that encode important functions for specific lineages and strains.

We also showed that structural analysis, combined with machine learning algorithms, can predict RBPs involved in ZIKV infection with testable structural details. The difference between *in vivo* and *in vitro* RNA structures has been exploited for the functional study of RNA, including the prediction of RBP binding (15). Here we combined *in vivo* and *in vitro*

structural profiles to predict host proteins that bind to ZIKV RNA. We verified a predicted interaction between the protein ELAVL1 and ZIKV RNA in two cell lines, Huh7 and U87MG. We also validated the binding sites with *in vitro* assays. We further confirmed ELAVL1 as a host factor for ZIKV infection in U87MG cells. ELAVL1 is known to bind to poly-U elements and AU-rich elements (AREs) in the 3'-UTR of target mRNAs, thereby increasing their stability (48). ELAVL1 is known to interact with other viruses, including HCV (37), swine fever virus (49), and sindbis virus (50). These interactions are all related to stability of the genomic RNA or mRNA of the viruses. Whether ELAVL1 promotes ZIKV infection also by stabilizing the viral RNA remains to be tested. A further system-wide study of the interactions between the candidate structural elements and the host factors is warranted. In sum, our approach and resource open a door to extensively investigate the function of RNA structures in virus infection.

AUTHOR CONTRIBUTIONS

Q.C.Z. and X.T. conceived this project. P.L. analyzed all the results. Y.W. performed the icSHAPE experiments. L.T. performed the PARIS experiments. M.M. performed the virus mutagenesis and rescue studies with the help from L.S.. L.S. validated ELAVL1 binding sites. M.M and L.S validated the ELAVL1 function in ZIKV infection. J.Z implemented domain demarcating algorithm. W.H. implemented host RBP factor prediction algorithm. Q.C.Z. and X.T. supervised the project. Q.C.Z., Y.W. and X.T. wrote the manuscript with inputs from all authors.

ACKNOWLEDGMENTS

We thank Xiaohua Shen (Tsinghua University) and Barry Honig (Columbia University) for helpful comments on the manuscript. This project is supported by the National the State Key Research Development Program of China (Grant No. 2016YFC1200300 to X.T.) and the National Natural Science Foundation of China (Grants No. 31671355, 91740204, and 31761163007 to Q.C.Z. and No. 31722030 to X.T.), the Beijing Advanced Innovation Center for Structural Biology to Q.C.Z, the Tsinghua-Peking Joint Center for Life Sciences and the National Thousand Young Talents Program of China to Q.C.Z. and X.T..

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Petersen, L. R., Jamieson, D. J., Powers, A. M. & Honein, M. A. Zika Virus. *The New England journal of medicine* **374**, 1552-1563, doi:10.1056/NEJMra1602113 (2016).
2. Brasil, P. *et al.* Zika Virus Infection in Pregnant Women in Rio de Janeiro. *The New England journal of medicine* **375**, 2321-2334, doi:10.1056/NEJMoa1602412 (2016).
3. Cugola, F. R. *et al.* The Brazilian Zika virus strain causes birth defects in experimental models. *Nature* **534**, 267-271, doi:10.1038/nature18296 (2016).
4. Mlakar, J. *et al.* Zika Virus Associated with Microcephaly. *The New England journal of medicine* **374**, 951-958, doi:10.1056/NEJMoa1600651 (2016).
5. Zhu, Z. *et al.* Comparative genomic analysis of pre-epidemic and epidemic Zika virus strains for virological factors potentially associated with the rapidly expanding epidemic. *Emerging microbes & infections* **5**, e22, doi:10.1038/emi.2016.48 (2016).
6. Yuan, L. *et al.* A single mutation in the prM protein of Zika virus contributes to fetal microcephaly. *Science* **358**, 933-936, doi:10.1126/science.aam7120 (2017).
7. Liu, Y. *et al.* Evolutionary enhancement of Zika virus infectivity in Aedes aegypti mosquitoes. *Nature* **545**, 482-486, doi:10.1038/nature22365 (2017).
8. Rodenhuis-Zybert, I. A., Wilschut, J. & Smit, J. M. Dengue virus life cycle: viral and host factors modulating infectivity. *Cellular and molecular life sciences : CMLS* **67**, 2773-2786, doi:10.1007/s00018-010-0357-z (2010).
9. Akiyama, B. M. *et al.* Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science* **354**, 1148-1152, doi:10.1126/science.aah3963 (2016).
10. Filomatori, C. V. *et al.* Dengue virus genomic variation associated with mosquito adaptation defines the pattern of viral non-coding RNAs and fitness in human cells. *PLoS pathogens* **13**, e1006265, doi:10.1371/journal.ppat.1006265 (2017).
11. Villordo, S. M. & Gamarnik, A. V. Genome cyclization as strategy for flavivirus RNA replication. *Virus research* **139**, 230-239, doi:10.1016/j.virusres.2008.07.016 (2009).

12. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711-716, doi:10.1038/nature08237 (2009).

13. Mauger, D. M. *et al.* Functionally conserved architecture of hepatitis C virus RNA genomes. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 3692-3697, doi:10.1073/pnas.1416266112 (2015).

14. Pirakitikulr, N., Kohlway, A., Lindenbach, B. D. & Pyle, A. M. The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures. *Molecular cell* **62**, 111-120, doi:10.1016/j.molcel.2016.01.024 (2016).

15. Spitale, R. C. *et al.* Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486-490, doi:10.1038/nature14263 (2015).

16. Lu, Z. P. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**, 1267-1279, doi:10.1016/j.cell.2016.04.028 (2016).

17. Lanciotti, R. S., Lambert, A. J., Holodniy, M., Saavedra, S. & Signor Ldel, C. Phylogeny of Zika Virus in Western Hemisphere, 2015. *Emerging infectious diseases* **22**, 933-935, doi:10.3201/eid2205.160065 (2016).

18. Dick, G. W., Kitchen, S. F. & Haddow, A. J. Zika virus. I. Isolations and serological specificity. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **46**, 509-520 (1952).

19. Lodeiro, M. F., Filomatori, C. V. & Gamarnik, A. V. Structural and functional studies of the promoter element for dengue virus RNA replication. *Journal of virology* **83**, 993-1008, doi:10.1128/JVI.01647-08 (2009).

20. Liu, Z. Y. *et al*. Novel cis-acting element within the capsid-coding region enhances flavivirus viral-RNA replication by regulating genome cyclization. *Journal of virology* **87**, 6804-6818, doi:10.1128/JVI.00243-13 (2013).

21. Liu, Z. Y. *et al.* Viral RNA switch mediates the dynamic control of flavivirus replicase recruitment by genome cyclization. *eLife* **5**, doi:10.7554/eLife.17636 (2016).

22. Low, J. T. & Weeks, K. M. SHAPE-directed RNA secondary structure prediction. *Methods* **52**, 150-158, doi:10.1016/j.ymeth.2010.06.007 (2010).

23. DiChiacchio, L., Sloma, M. F. & Mathews, D. H. AccessFold: predicting RNA-RNA interactions with consideration for competing self-structure. *Bioinformatics* **32**, 1033-1039, doi:10.1093/bioinformatics/btv682 (2016).

24. Yu, Z. H. & Herman, G. On the earth mover's distance as a histogram similarity metric for image retrieval. *2005 IEEE International Conference on Multimedia and Expo (ICME), Vols 1 and 2*, 686-689 (2005).

25. Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure determination. Proceedings of the National Academy of Sciences of the United States of America 106, 97-102, doi:10.1073/pnas.0806929106 (2009).

26. Wu, Y. *et al.* Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. *Nucleic acids research* **43**, 7247-7259, doi:10.1093/nar/gkv706 (2015)

27. Rivas, E., Clements, J. & Eddy, S. R. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nature methods* **14**, 45-48, doi:10.1038/NMETH.4066 (2017).

28. Villordo, S. M., Carballeda, J. M., Filomatori, C. V. & Gamarnik, A. V. RNA Structure Duplications and Flavivirus Host Adaptation. *Trends in microbiology* **24**, 270-283, doi:10.1016/j.tim.2016.01.002 (2016).

29. Clyde, K., Barrera, J. & Harris, E. The capsid-coding region hairpin element (cHP) is a critical determinant of dengue virus and West Nile virus RNA synthesis. *Virology* **379**, 314-323, doi:10.1016/j.virol.2008.06.034 (2008).

30. Chavali, P. L. *et al.* Neurodevelopmental protein Musashi-1 interacts with the Zika genome and promotes viral replication. *Science* **357**, 83-88, doi:10.1126/science.aam9243 (2017).

31. Liu, Z. Y. *et al.* Characterization of cis-Acting RNA Elements of Zika Virus by Using a Self-Splicing Ribozyme-Dependent Infectious Clone. *Journal of virology* **91**, doi:UNSP e00484-17 10.1128/JVI.00484-17 (2017).

32. Ng, W. C., Soto-Acosta, R., Bradrick, S. S., Garcia-Blanco, M. A. & Ooi, E. E. The 5' and 3' Untranslated Regions of the Flaviviral Genome. *Viruses* **9**, doi:10.3390/v9060137 (2017).

33. Giudice, G., Sanchez-Cabo, F., Torroja, C. & Lara-Pezzi, E. ATtRACT-a database of RNA-binding proteins and associated motifs. *Database : the journal of biological databases and curation* **2016**, doi:10.1093/database/baw035 (2016).

34. Lei, Y. F. *et al.* Functional interaction between cellular p100 and the dengue virus 3 ' UTR. *Journal of General Virology* **92**, 796-806, doi:10.1099/vir.0.028597-0 (2011).

35. Li, G. H., Feng, T. T., Pan, W., Shi, X. H. & Dai, J. F. DEAD-box RNA helicase DDX3X inhibits DENV replication via regulating type one interferon pathway. *Biochemical and biophysical research communications* **456**, 327-332, doi:10.1016/j.bbrc.2014.11.080 (2015).

36. Weinlich, S. *et al.* IGF2BP1 enhances HCV IRES-mediated translation initiation via the 3' UTR. *Rna-a Publication of the Rna Society* **15**, 1528-1542, doi:10.1261/rna.1578409 (2009).

37. Shwetha, S. *et al.* HuR Displaces Polypyrimidine Tract Binding Protein To Facilitate La Binding to the 3 ' Untranslated Region and Enhances Hepatitis C Virus Replication. *Journal of virology* **89**, 11356-11371, doi:10.1128/JVI.01714-15 (2015).

38. Gilmartin, G. M., Fleming, E. S., Oetjen, J. & Graveley, B. R. CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes & development* **9**, 72-83 (1995).

39. Ou, S. H., Wu, F., Harrich, D., Garcia-Martinez, L. F. & Gaynor, R. B. Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *Journal of virology* **69**, 3584-3596 (1995).

40. Bley, N. *et al.* Stress granules are dispensable for mRNA stabilization during cellular stress. *Nucleic acids research* **43**, e26, doi:10.1093/nar/gku1275 (2015).

41. Kobayashi, T., Winslow, S., Sunesson, L., Hellman, U. & Larsson, C. PKCalpha binds G3BP2 and regulates stress granule formation following cellular stress. *PloS one* **7**, e35820, doi:10.1371/journal.pone.0035820 (2012).

42. Hou, S. *et al.* Zika virus hijacks stress granule proteins and modulates the host stress response. *Journal of virology*, doi:10.1128/JVI.00474-17 (2017).

43. Lichinchi, G. *et al.* Dynamics of Human and Viral RNA Methylation during Zika Virus Infection. *Cell host & microbe* **20**, 666-673, doi:10.1016/j.chom.2016.10.002 (2016).

44. Gokhale, N. S. *et al.* N6-Methyladenosine in Flaviviridae Viral RNA Genomes Regulates Infection. *Cell host & microbe* **20**, 654-665, doi:10.1016/j.chom.2016.09.015 (2016).

45. Smola, M. J., Calabrese, J. M. & Weeks, K. M. Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry* **54**, 6867-6875, doi:10.1021/acs.biochem.5b00977 (2015).
46. Aw, J. G. *et al.* In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Molecular cell* **62**, 603-617, doi:10.1016/j.molcel.2016.04.028 (2016).
47. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B. J. Global Mapping of Human RNA-RNA Interactions. *Molecular cell* **62**, 618-626, doi:10.1016/j.molcel.2016.04.030 (2016).
48. Doller, A. *et al.* Posttranslational modification of the AU-rich element binding protein HuR by protein kinase Cdelta elicits angiotensin II-induced stabilization and nuclear export of cyclooxygenase 2 mRNA. *Molecular and cellular biology* **28**, 2608-2625, doi:10.1128/MCB.01530-07 (2008).
49. Nadar, M. *et al.* HuR binding to AU-rich elements present in the 3' untranslated region of Classical swine fever virus. *Virology journal* **8**, 340, doi:10.1186/1743-422X-8-340 (2011).
50. Sokoloski, K. J. *et al.* Sindbis virus usurps the cellular HuR protein to stabilize its transcripts and promote productive infections in mammalian and mosquito cells. *Cell host & microbe* **8**, 196-207, doi:10.1016/j.chom.2010.07.003 (2010).

FIGURE LEGENDS

Figure 1. Schematic presentation of ZIKV RNA structure *in vivo* study by the combination of icSHAPE and PARIS. See also STAR Methods, Figure S1.

For ZIKV PRVABC59 and MR766 h.p.i Huh-7 cells, icSHAPE and PARIS were performed to detect RNA structural profiles and intra-molecular interactions of ZIKV RNA genomes *in vivo*.

Figure 2. Structural overview of icSHAPE profiling in the genome of ZIKV PRVABC59 and MR766. See also STAR Methods, Tables S1 and S3.

A. Normalized *in vivo* icSHAPE reactivity score of PRVABC59 shown relative to the global median value, with higher values correspond to more flexible nucleotides. Blue

color thus represents a region more likely to be pairing state, and red color represents more likely a non-pairing region.

B. Normalized *in vivo* icSHAPE reactivity score of MR766.

C. Structure-conserved regions between two viruses.

D. Well-known functional elements in two viruses.

E. Sequence identity and correlations of icSHAPE reactivity scores between PRVABC59 and MR766 strains in sliding windows. The CDS region and UTRs are respectively colored with cyan and red. Sites marked with yellow asterisk are top-10 lineage-specific structure sites between PRVABC59 lineage and MR766 lineage.

F. Known structural models of the 3'UTR (top), the 5'UTR and flanking regions (bottom middle) and the 5'-3' interaction (bottom left), colored with icSHAPE reactivity scores. Dashed lines represent pseudoknots. RNA-RNA interactions with PARIS data supporting are shadowed in light blue with gapped reads numbers in brackets.

Figure 3. Structural overview of PARIS data in PRVABC59. See also STAR Methods, Figures S2, S3 and S4, and Tables S1 and S2.

A. Distribution of the PARIS connecting reads in the PRVABC59 genomes. Pairing structural models of some most stable long-range interactions including the known 5'-3' interaction are highlighted with coordinates and folding energies (**a**~**d** and **known**). Small triangles flanking the middle bar represent RNA structural domains of PRVABC59.

B. Comparison between icSHAPE-guided predicted base-pairing interactions by MFE (up) and PARIS interactions (below) for each domain of PRVABC59. Green arcs represent the common interactions while gray arcs are interactions only inferred from one method.

Figure 4. Full-Length structural model of the PRVABC59 RNA genome. See also STAR Methods and Figure S5.

Nucleotides are colored with icSHAPE reactivity scores. Base pairs with conserved covariation are boxed with green rectangle. Long range interaction in Figure 1D are indicated by dotted lines. The two boxplot insets at the bottom are the distributions of icSHAPE reactivity scores and PARIS supporting reads number for all interactions.

Figure 5. An Asian lineage-specific long-range intramolecular interaction may contribute to ZIKV infection. See also Table S4.

A. Circos plot of the long-range interaction in the PRVABC59 genome and the MR766 genome. The outermost ring shows the genomic coordinates with protein annotation. The middle and the inner rings indicate amino acids and nucleotide sequence diversity, respectively. The color ribbons represent the long-range interactions of PRVABC59 (green) of MR766 (pink). The blue arrow points to the 5'-E CS interaction that only exists in PRVABC59.
B. Nucleotide sequence diversity between African strains and Asian strains in the E protein coding region of the 5'-E CS. The two viral genomes are highlighted with PRVABC59 as the reference sequence. Mutations to the PRVABC59 genome are colored per type. Strains are organized according to the phylogenetic tree on the left.
C. Predicted secondary structure model of the 5'-E CS interaction in PRVABC59, annotated with genomic coordinates. Blue circles represent designs of mutations and red circles represent rescues in the infection study.
D. Time course of qPCR quantitation of relative GZ01 viral RNA copies from pellets of infected C6/36 cells. The GAA mutant is a control with a defected NS5 RdRp domain.
E. Summarized infection ratio in Vero cells infected with WT, mutant and rescue virus amplified for 1,3 and 5 days in C6/36 cells for 48h by immunostaining analysis with 4G2.
F. E protein immunofluorescence assay of 48 h.p.i Vero cells infected with WT, mutant and rescue strains amplified for 5 days in C6/36 cells. Numbers indicate the ratio of infected cells.
G. Model of alternative ZIKV RNA genomic conformations including 5'UTR, E protein and 3'UTR local structures, long-range 5'-E CS and 5'-3' CS interactions (dashed lines). Regions are marked to help identify pairing patterns.

Figure 6. Prediction and validation of host RBPs that bind to Asian strain ZIKA RNA genomes. See also STAR Methods, Figure S6, Tables S5 and S6.

A. Predicted binding sites of different RBPs on the PRVABC59 RNA genome, indicated by blue bars. All bars are from motif scanning; those with a box are positive and the others are negative sites by the random forest predictions based on structural information.

B. RNA pull-down assays with RNA oligos of predicted ELAVL1-binding sites in the PRVABC59 genome.
C. RNA pull-down assays with RNA oligos of predicted ELAVL1-binding sites in the MR766 genome.
D. Scatter plot of RNA-pull down assay by average icSHAPE scores and ELAVL1 affinity values.
E. ELAVL1 co-localizes with ZIKV RNA in U87MG cells by immunofluorescence assay.
F. RNA-immunoprecipitation of ZIKV RNA in infected U87MG ELAVL1 overexpression cells and U87MG luciferase negative contrl cells.
G. Normalized ELAVL1 expression from qRT-PCR in ELAVL1-knockdown U87MG cells
H. Normalized viral RNA copy numbers from qRT-PCR in ELAVL1-knockdown GZ01 48.h.p.i U87MG cells
I. Western blottings of ELAVL1-knockdown 48.h.p.i U87MG cells
J. Normalized ELAVL1 expression from qRT-PCR in ELAVL1-overexpressed 48.h.p.i U87MG cells.
K. Normalized viral RNA copy numbers from qRT-PCR in ELAVL1-overexpressed 48.h.p.i U87MG cells
L. Western blotting results of ELAVL1-overexpressed 48.h.p.i U87MG cells.

* means p-value ≤ 0.01, **** means p-value ≤ 0.0001 (t-test)

Figure 1

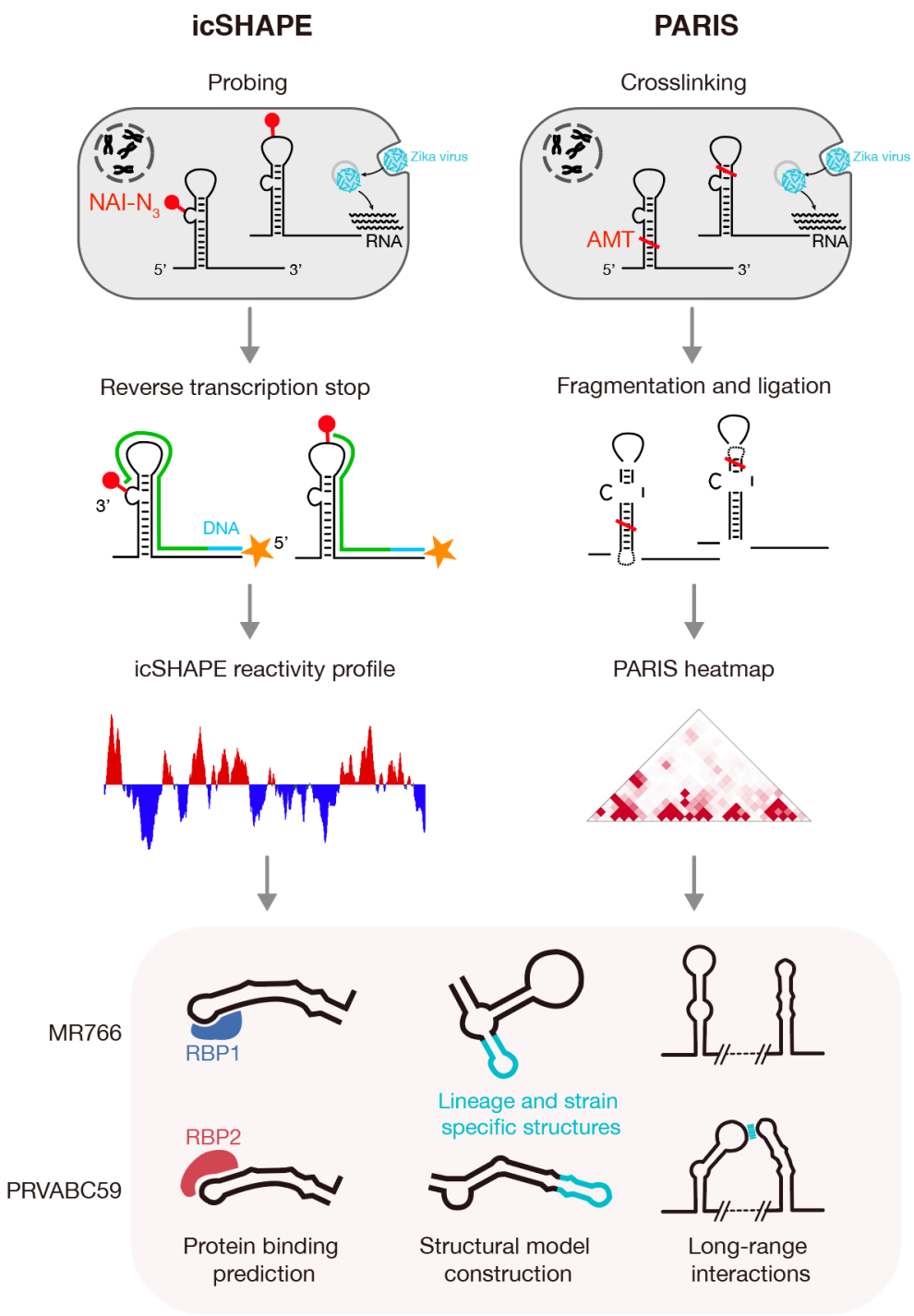

icSHAPE
PARIS
Probing
Crosslinking
Zika virus
NAI-N3
AMT
RNA
5'
3'
Reverse transcription stop
Fragmentation and ligation
DNA
icSHAPE reactivity profile
PARIS heatmap
MR766
PRVABC59
RBP1
RBP2
Lineage and strain specific structures
Protein binding prediction
Structural model construction
Long-range interactions

Figure 2

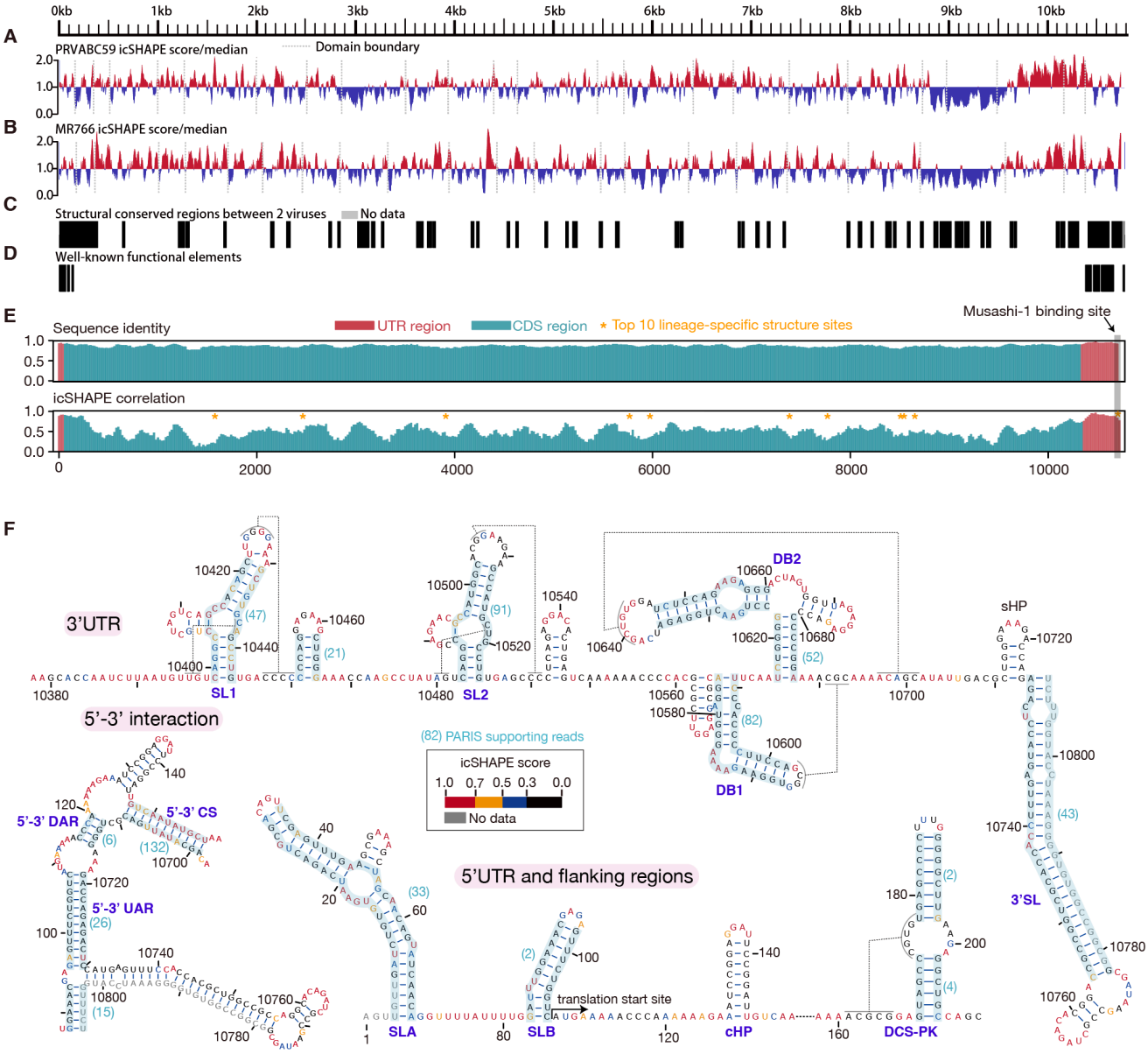
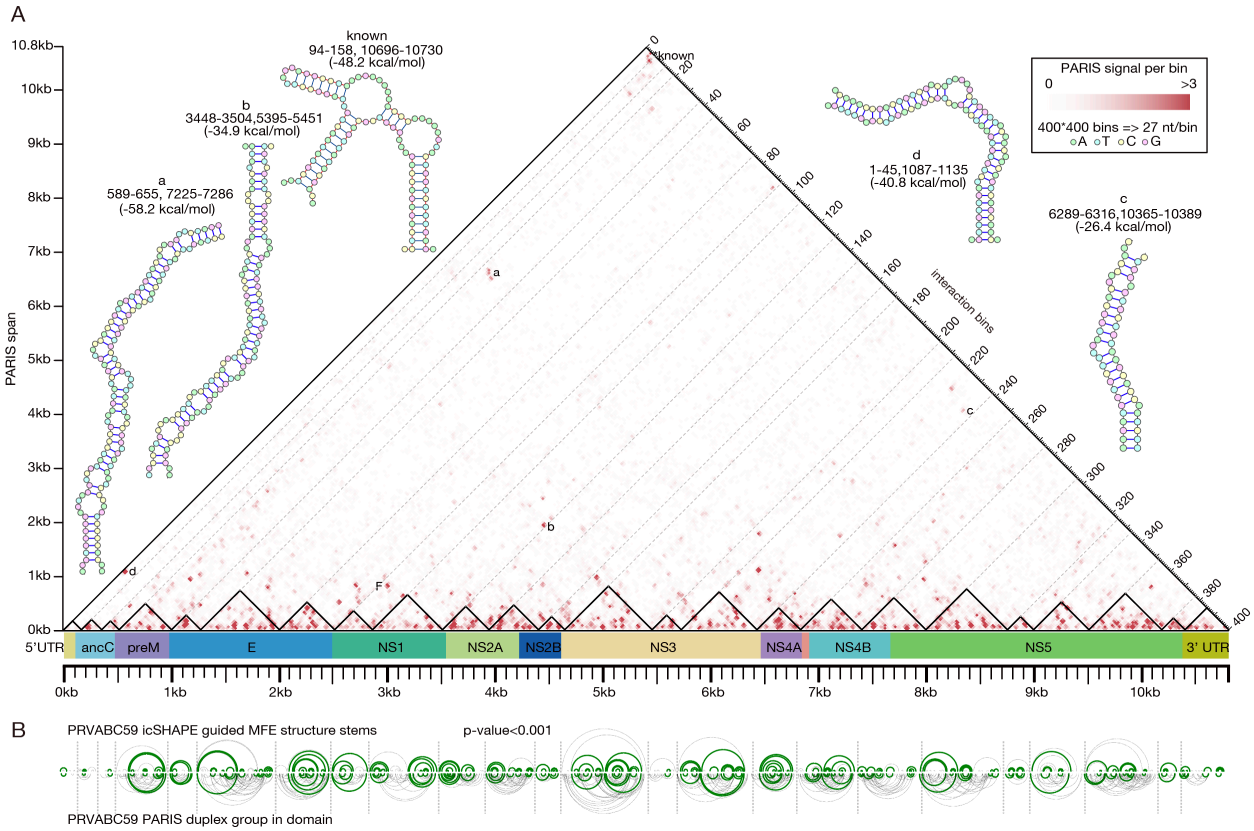

A
PRVABC59 icSHAPE score/median
Domain boundary
B
MR766 icSHAPE score/median
C
Structural conserved regions between 2 viruses
No data
D
Well-known functional elements
E
Sequence identity
UTR region
CDS region
Top 10 lineage-specific structure sites
Musashi-1 binding site
icSHAPE correlation
F
3'UTR
5'-3' interaction
5'UTR and flanking regions
SL1
SL2
DB1
DB2
sHP
5'-3' DAR
5'-3' CS
5'-3' UAR
SLA
SLB
translation start site
cHP
DCS-PK
3'SL
(82) PARIS supporting reads
icSHAPE score
No data

Figure 3


A
known
94-158, 10696-10730
(-48.2 kcal/mol)
b
3448-3504,5395-5451
(-34.9 kcal/mol)
a
589-655, 7225-7286
(-58.2 kcal/mol)
d
1-45,1087-1135
(-40.8 kcal/mol)
c
6289-6316,10365-10389
(-26.4 kcal/mol)
PARIS signal per bin
0
>3
400*400 bins => 27 nt/bin
A T C G
PARIS span
interaction bins
5'UTR
ancC
preM
E
NS1
NS2A
NS2B
NS3
NS4A
NS4B
NS5
3' UTR

B
PRVABC59 icSHAPE guided MFE structure stems
p-value<0.001
PRVABC59 PARIS duplex group in domain

Figure 4

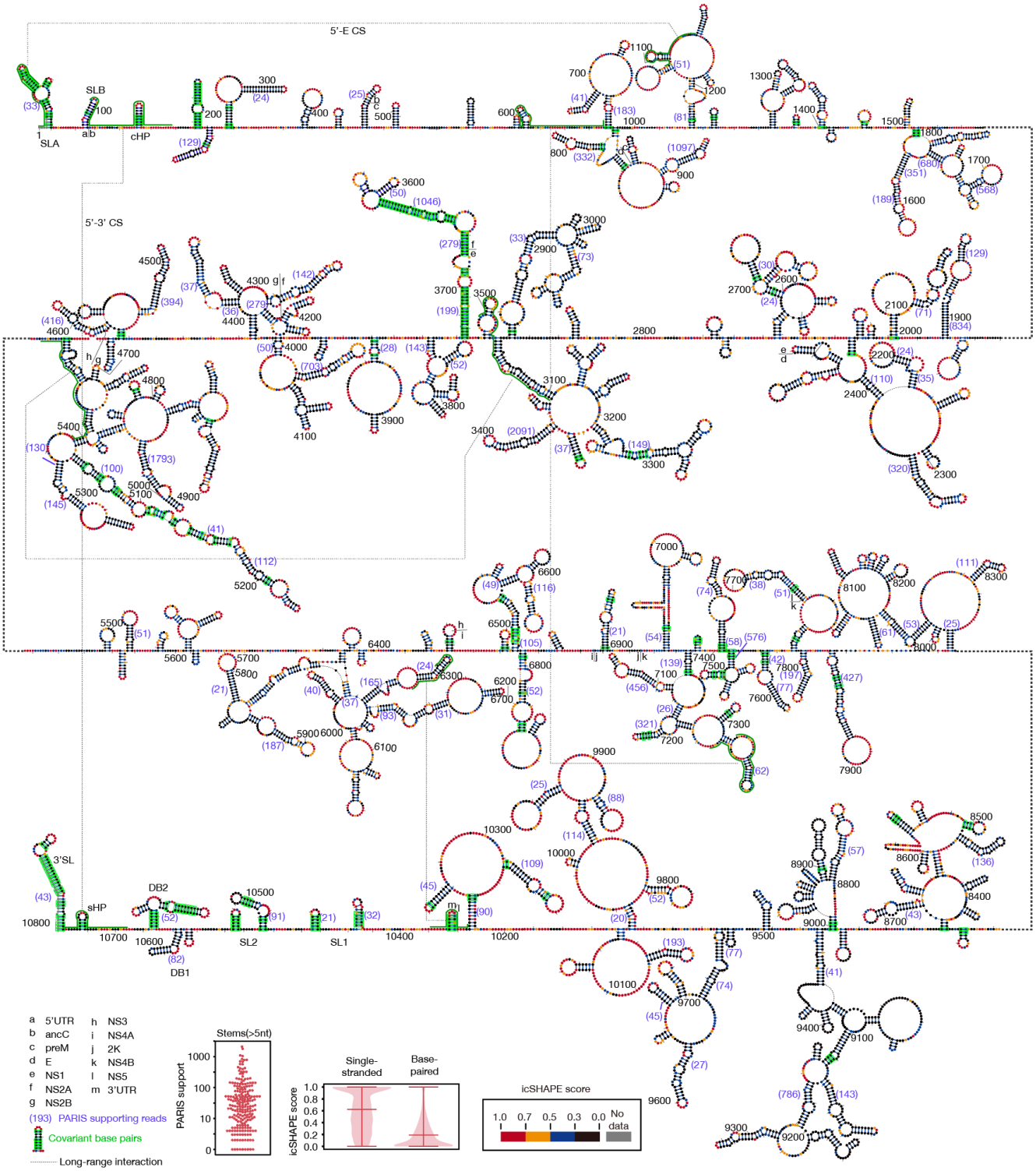

5'-E CS
5'-3' CS
SLB
SLA
cHP
3'SL
sHP
DB2
DB1
SL2
SL1
a 5'UTR
b ancC
c preM
d E
e NS1
f NS2A
g NS2B
h NS3
i NS4A
j 2K
k NS4B
l NS5
m 3'UTR
(193) PARIS supporting reads
Covariant base pairs
Long-range interaction
Stems(>5nt)
PARIS support
Single-stranded
Base-paired
icSHAPE score
icSHAPE score
1.0 0.7 0.5 0.3 0.0 No data

Figure 5

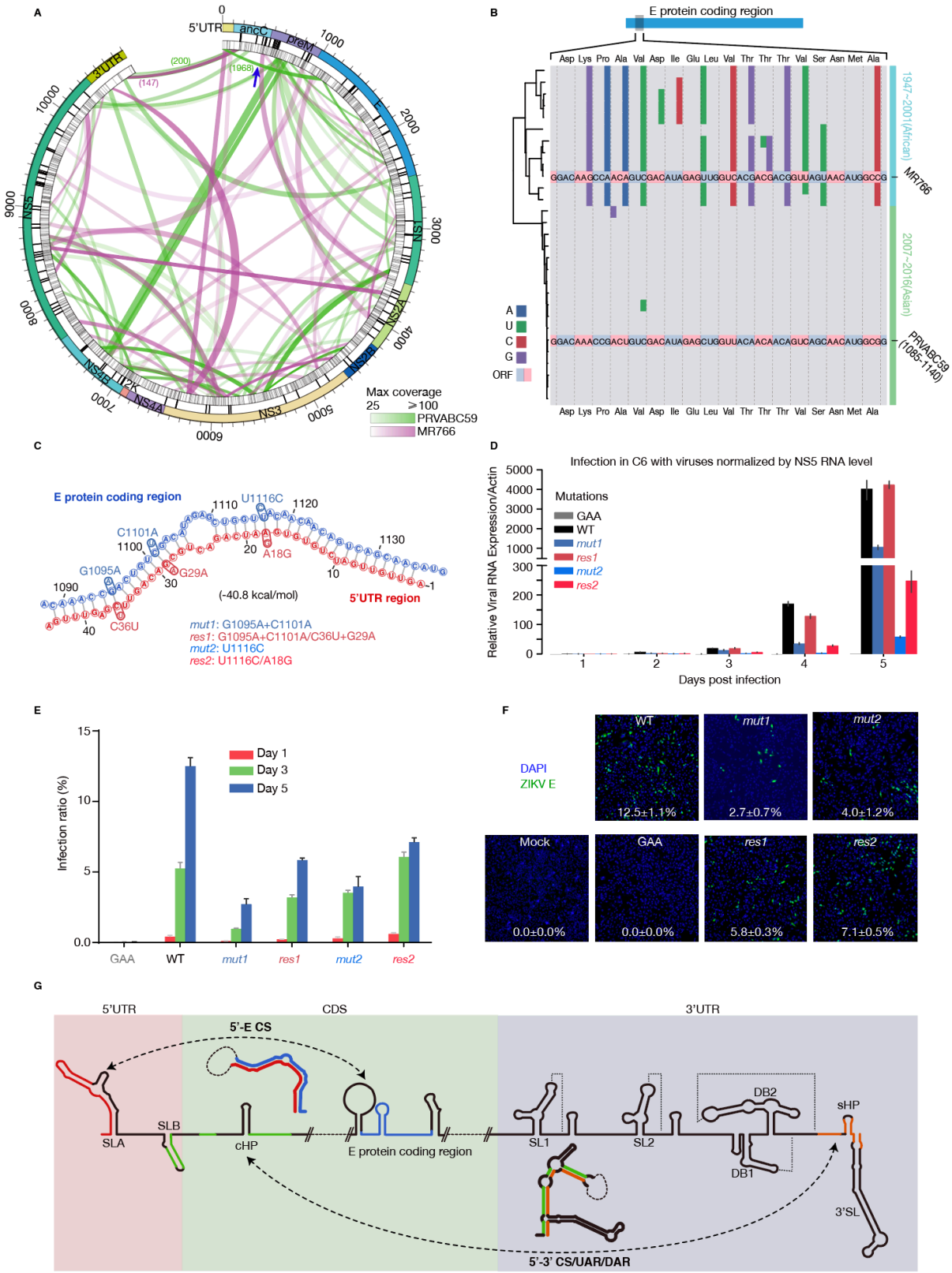
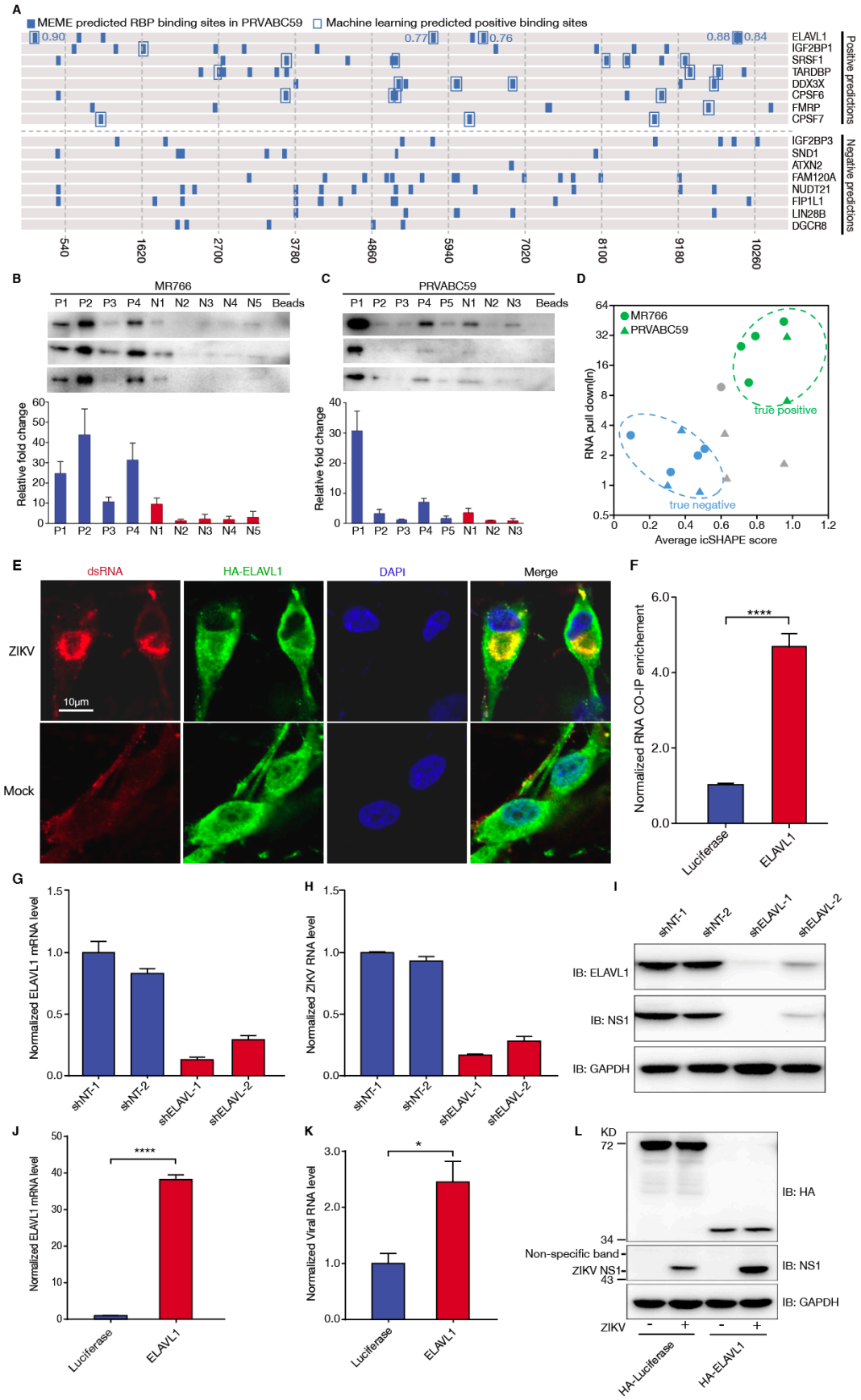

A
5'UTR
ancC
prM
E
NS1
NS2A
NS2B
NS3
NS4A
NS4B
NS5
3'UTR
(200)
(147)
(1968)
Max coverage
25 ≥100
PRVABC59
MR766
B
E protein coding region
Asp Lys Pro Ala Val Asp Ile Glu Leu Val Thr Thr Thr Val Ser Asn Met Ala
1947~2001(African)
2007~2016(Asian)
MR766
PRVABC59 (1085-1140)
A
U
C
G
ORF
C
E protein coding region
U1116C
C1101A
G1095A
G29A
A18G
C36U
(-40.8 kcal/mol)
5'UTR region
mut1: G1095A+C1101A
res1: G1095A+C1101A/C36U+G29A
mut2: U1116C
res2: U1116C/A18G
D
Infection in C6 with viruses normalized by NS5 RNA level
Relative Viral RNA Expression/Actin
Mutations
GAA
WT
mut1
res1
mut2
res2
Days post infection
E
Infection ratio (%)
Day 1
Day 3
Day 5
GAA
WT
mut1
res1
mut2
res2
F
DAPI
ZIKV E
WT
mut1
mut2
12.5±1.1%
2.7±0.7%
4.0±1.2%
Mock
GAA
res1
res2
0.0±0.0%
0.0±0.0%
5.8±0.3%
7.1±0.5%
G
5'UTR
CDS
3'UTR
5'-E CS
SLA
SLB
cHP
E protein coding region
SL1
SL2
DB2
DB1
sHP
3'SL
5'-3' CS/UAR/DAR

Figure 6


A
MEME predicted RBP binding sites in PRVABC59
Machine learning predicted positive binding sites
0.90
0.77
0.76
0.88
0.84
ELAVL1
IGF2BP1
SRSF1
TARDBP
DDX3X
CPSF6
FMRP
CPSF7
Positive predictions
IGF2BP3
SND1
ATXN2
FAM120A
NUDT21
FIP1L1
LIN28B
DGCR8
Negative predictions
540
1620
2700
3780
4860
5940
7020
8100
9180
10260
B
MR766
P1 P2 P3 P4 N1 N2 N3 N4 N5 Beads
Relative fold change
C
PRVABC59
P1 P2 P3 P4 P5 N1 N2 N3 Beads
D
MR766
PRVABC59
true positive
true negative
RNA pull down(ln)
Average icSHAPE score
E
dsRNA
HA-ELAVL1
DAPI
Merge
ZIKV
Mock
10μm
F
Normalized RNA CO-IP enrichement
****
Luciferase
ELAVL1
G
Normalized ELAVL1 mRNA level
shNT-1
shNT-2
shELAVL-1
shELAVL-2
H
Normalized ZIKV RNA level
I
IB: ELAVL1
IB: NS1
IB: GAPDH
J
Normalized ELAVL1 mRNA level
****
K
Normalized Viral RNA level
*
L
KD
72
34
43
IB: HA
Non-specific band
ZIKV NS1
IB: NS1
IB: GAPDH
ZIKV
HA-Luciferase
HA-ELAVL1