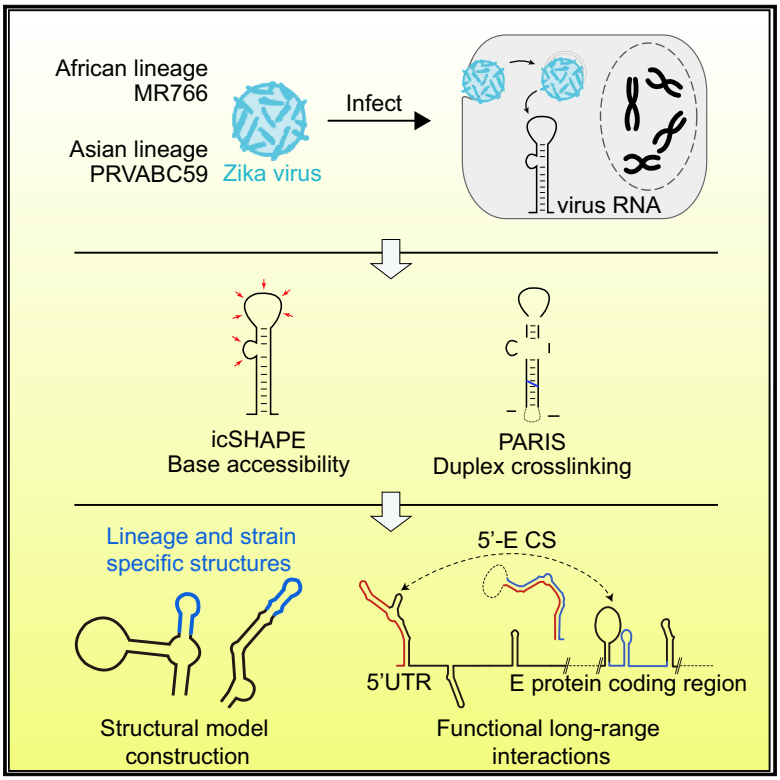
Resource

Cell Host & Microbe

Integrative Analysis of Zika Virus Genome RNA Structure Reveals Critical Determinants of Viral Infectivity

Graphical Abstract



Highlights

- Constructed comprehensive structural maps of ZIKV RNA genomes in infected cells
- Discovered functional RNA structural elements in ZIKV genomes
- Identified an Asian lineage-specific RNA-RNA interaction that affects ZIKV infectivity

Authors

Pan Li, Yifan Wei, Miao Mei, ..., Jianfeng Dai, Xu Tan, Qiangfeng Cliff Zhang

Correspondence

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

In Brief

Li et al. determined structural maps of ZIKV RNA genomes for both Asian and African strains in infected cells. This analysis revealed numerous functional RNA structural elements and, notably, a long-range interaction between the coding and untranslated regions that is specific to epidemic Asian strains and contributes to infectivity.

Li et al., 2018, Cell Host & Microbe *24*, 1–12
December 12, 2018
https://doi.org/10.1016/j.chom.2018.10.011

CellPress

Please cite this article in press as: Li et al., Integrative Analysis of Zika Virus Genome RNA Structure Reveals Critical Determinants of Viral Infectivity, Cell Host & Microbe (2018), https://doi.org/10.1016/j.chom.2018.10.011

Cell Host & Microbe
Resource

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] Wenze Huang,[1] Jianyu Zhou,[3] Chunlin Zou,[4] Shaojun Zhang,[1] Cheng-Feng Qin,[5] Tao Jiang,[3,6] Jianfeng Dai,[4] Xu Tan,[2,*] and 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 215123, 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]These authors contributed equally
[8]Lead Contact
*Correspondence: xutan@tsinghua.edu.cn (X.T.), qczhang@tsinghua.edu.cn (Q.C.Z.)
https://doi.org/10.1016/j.chom.2018.10.011

SUMMARY

Zika virus (ZIKV) strains can be classified into the ancestral African and contemporary Asian lineages, with the latter responsible for recent epidemics associated with neurological conditions. To understand how Asian strains lead to exacerbated disease, a crucial step is identifying genomic variations that affect infectivity and pathogenicity. Here we use two high-throughput sequencing approaches to assess RNA secondary structures and intramolecular RNA-RNA interactions *in vivo* for the RNA genomes of Asian and African ZIKV lineages. Our analysis identified functional RNA structural elements and a functional long-range intramolecular interaction specific for the Asian epidemic strains. Mutants that disrupt this extended RNA interaction between the 5′ UTR and the E protein coding region reduce virus infectivity, which is partially rescued with compensatory mutants, restoring this RNA-RNA interaction. These findings illuminate the structural basis of ZIKV regulation and provide a resource for the discovery of RNA structural elements important for ZIKV infection.

INTRODUCTION

Zika virus (ZIKV) is a mosquito-borne single-stranded RNA virus of the Flaviviridae family (Petersen et al., 2016). 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 (Brasil et al., 2016; Cugola et al., 2016; Mlakar et al., 2016). 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 (Zhu et al., 2016). 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 (Yuan et al., 2017), and another mutation A188V in the NS1 protein promotes transmissibility in mosquito vectors (Liu et al., 2017a). 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 (Figures S1A–S1C). Whether and how these mutations could also contribute to ZIKV infection remains 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 (Rodenhuis-Zybert et al., 2010). 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 that help the virus evade cellular

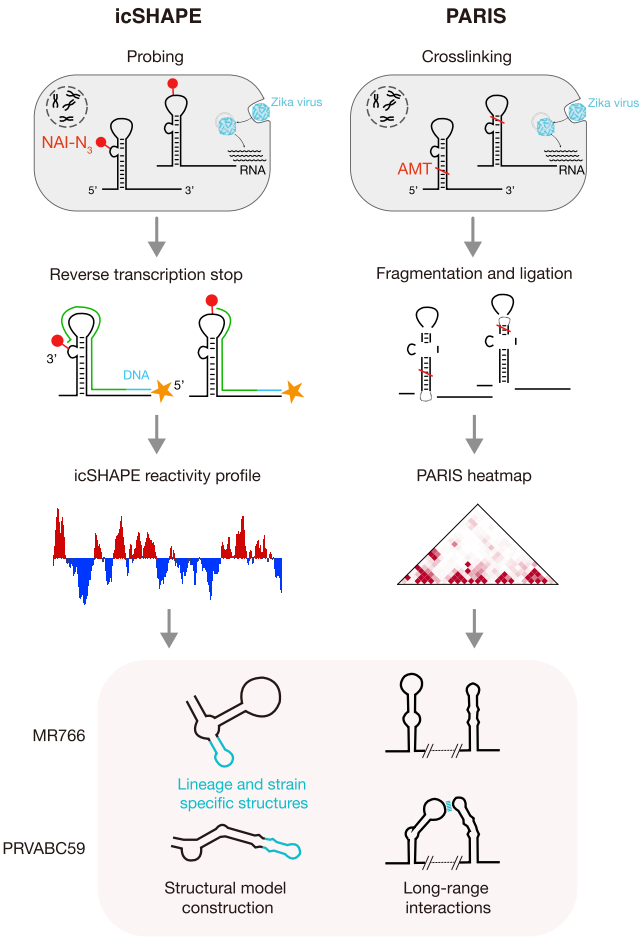


Figure 1. Schematic Presentation of ZIKV RNA Structure *In Vivo* Study by the Combination of icSHAPE and PARIS

For Huh-7 cells with ZIKV PRVABC59 and MR766 72 hours post infection (h.p.i), icSHAPE is performed to detect virus RNA flexibility (paired or unpaired) at nucleotide resolution *in vivo*. PARIS was performed to detect intramolecular interactions of ZIKV RNA genomes *in vivo*. See also STAR Methods, Figure S1.

antiviral processes (Akiyama et al., 2016; Filomatori et al., 2017). 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 role in coordinating virus replication (Villordo and Gamarnik, 2009). To date, our knowledge of ZIKV RNA structure is mainly limited to the UTRs. However, the coding region constitutes more than 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 (Watts et al., 2009). Other studies examined multiple hepatitis C virus (HCV) genome structures and found various structural regulatory elements across the whole genome including coding regions (Mauger et al., 2015; Pirakitikulr et al., 2016). 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 (Spitale et al., 2015) and PARIS (Lu et al., 2016) (Figure 1). We generated maps of the secondary structures of two ZIKV strains representing the Asian and African lineages, with many 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. Therefore, we provide a rich resource for understanding the structure-function relationship of ZIKV genomic RNA.

RESULTS

icSHAPE Defines Nucleotide Flexibility Profiles of ZIKV Genomes

For RNA structural determination, we chose PRVABC59 (Lanciotti et al., 2016) as a representative Asian strain and MR766 (Dick et al., 1952) for the African strains, according to their positions in the ZIKV phylogenetic tree (Figure S1A). We noticed that the PRVABC59 reference genome from GenBank lacks 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′ UTRs, respectively, and 11.3% variation within the polyprotein coding region (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 more than 99.6% of the nucleotides within the two ZIKV genomes (Figures 2A and 2B, Table S1). The icSHAPE scores were reproducible between independent biological replicates ($R \geq 0.89$). The

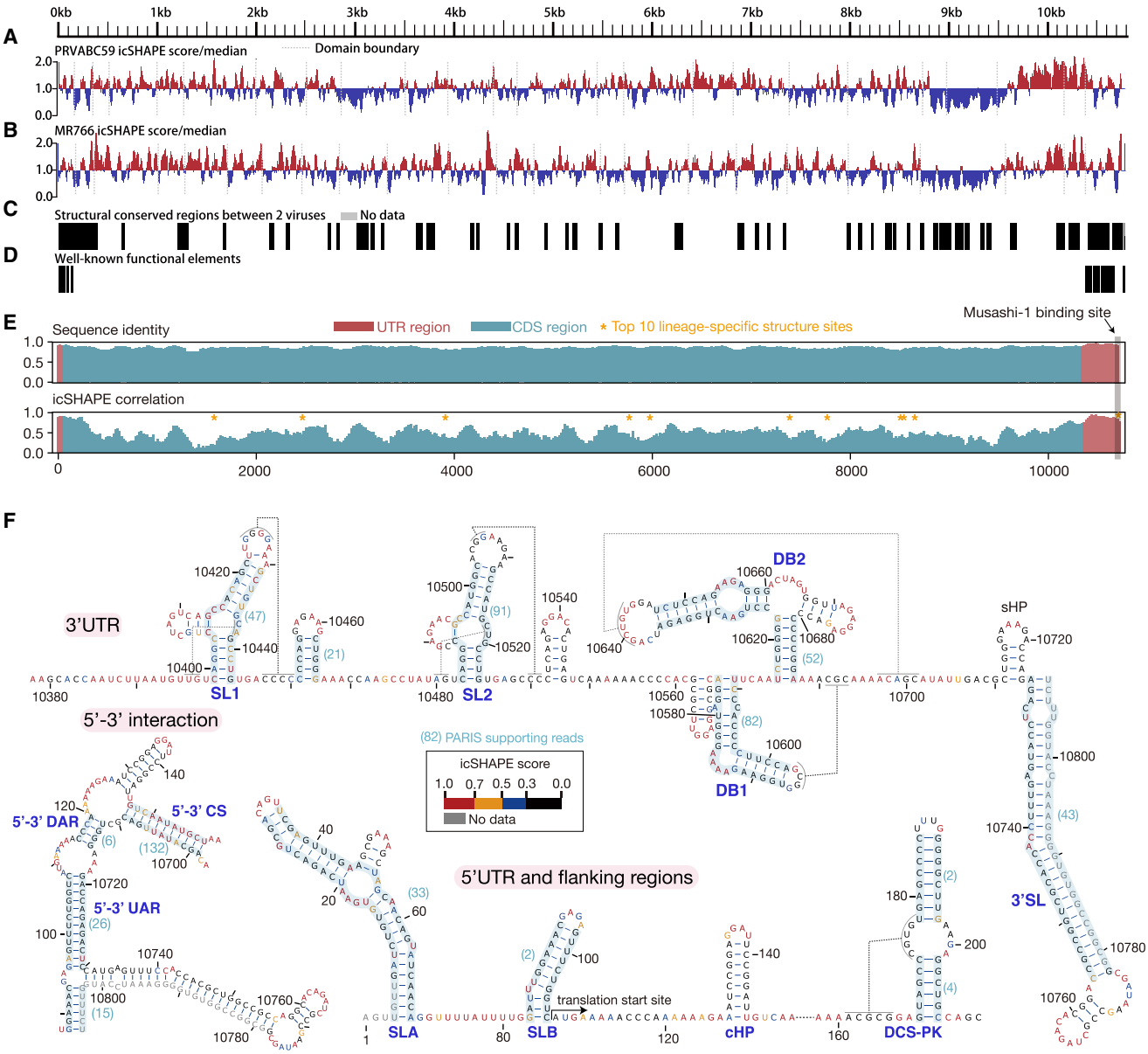


Figure 2. Structural Overview of icSHAPE Profiling in the Genome of ZIKV PRVABC59 and MR766

(A) Normalized *in vivo* icSHAPE reactivity score of PRVABC59 is shown relative to the global median value, with higher values corresponding to more flexible nucleotides. Blue color represents a region more likely to be a pairing state, and red color represents a more likely non-pairing region. Normalized scores are smoothed with a 30-nt window.

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

(C) Structure-conserved regions between two viruses, defined as those regions with icSHAPE difference greater than 0.15 in a 30-nt window.

(D) Well-known functional elements in two viruses as shown in (F).

(E) Sequence identity and correlations of icSHAPE reactivity scores between PRVABC59 and MR766 strains in sliding windows. The CDS region and UTRs are colored with cyan and red, respectively. Sites marked with yellow asterisks 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 read numbers in brackets.

See also STAR Methods, Tables S1 and S3.

structural conservation was significant and contains many well-known functional elements in flaviviruses (Figures 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 sub-genomic flavivirus RNAs 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 (Lodeiro et al., 2009). The DCS-PK element in the 5′ UTR flanking region, is a pseudoknot found in coding region (Liu et al., 2013). 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 (Liu et al., 2013). 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 (Area Under Curve, 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. As one exception, the small stem of 5′-3′ complementary structure (CS) was not reflected in the icSHAPE scores, which might indicate that this structure is dynamic or does not form under these conditions.

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 (Spitale et al., 2015). The remaining steps and data analysis were the same as for 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 (Spitale et al., 2015). 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 versus 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 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 (Lu et al., 2016). 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 RNA-RNA interaction is defined by an alignment of gapped reads that can be mapped to the two stems of the duplex. Although PARIS cannot technically distinguish whether an RNA-RNA interaction is intramolecular or intermolecular, because proximity ligation is a key step of PARIS, and the interactions detected are typically intramolecular.

We discovered a large number of RNA-RNA interactions or pairing structures for both viruses by PARIS (Figures 3A and S2A). The PARIS data were reproducible between independent biological replicates (R = 0.98), clustered into 1,482 duplexes for PRVABC59 and 1,282 duplexes for MR766 (Table S2). Among these duplexes, we defined those with pairing distance of more than 1 kb as long-range interactions, resulting in 230 for PRVABC59 (16% of all duplexes) and 178 for MR766 (14% of all duplexes) (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 (Liu et al., 2016). These interactions, including 5′-3′ upstream of the AUG region, 5′-3′ downstream of the AUG region, and 5′-3′ CS, are well-recovered in our results (Figure 2F).

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 the software suite *RNAstructure* (Deigan et al., 2009) to predict a local minimum free-energy (MFE) structure with icSHAPE scores as a constraint (Figures 3B and S2B). We compared the base pairing in the icSHAPE-assisted predicted structure with those measured by PARIS and found substantial overlap ($p < 0.001$, permutation test, Figures 3B and S2B). 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.

We noticed that short-range interactions are more conserved than long-range ones between the two ZIKV strains (Figure S2C). We found that MR766 and PRVABC59 share more than one-third of short-range interactions, but less than 10% of long-range interactions. PARIS relies on psoralen intercalating in RNA helices, and requires deep sequencing due to limited crosslinking and ligation efficiencies (Lu et al., 2016). 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 (Low and Weeks, 2010; DiChiacchio et al., 2016). Next, we permutated the sequence in both stems 100 times, calculated the minimal fold free energy

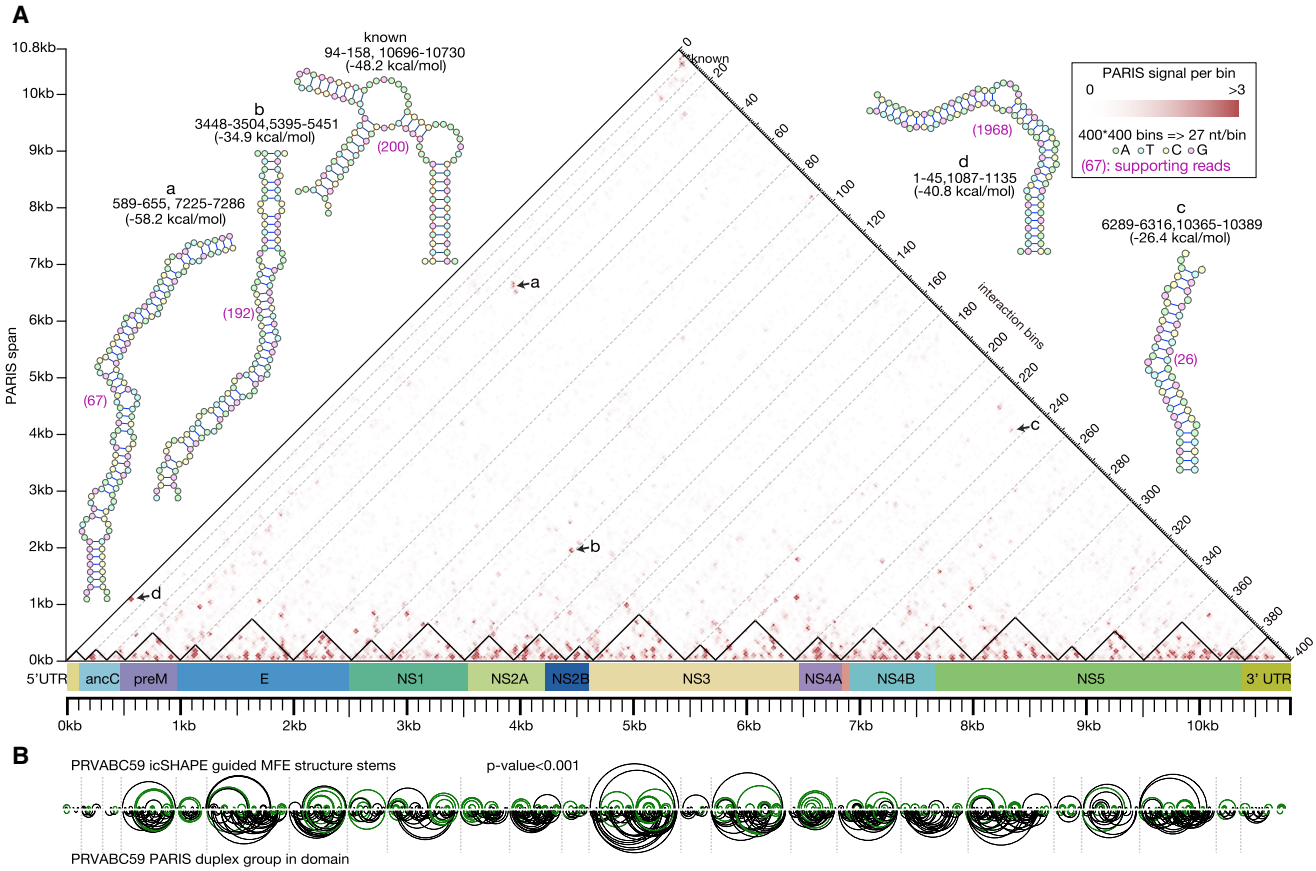


Figure 3. Structural Overview of PARIS Data in PRVABC59

(A) Heatmap of the PARIS connecting reads in the PRVABC59 genomes. Each data point represents an interaction between two regions. The coordinates of the two regions should be read by projecting the data point along the two declining axes. PARIS raw reads are normalized to signal, which is shown as the color depth. 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. p value is calculated by shuffle predicted stems for 1,000 times.

See also STAR Methods, Figures S2–S4, and Tables S1 and S2.

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 = 6.2×10^{-64} for PRVABC59 and p value = 1.5×10^{-45} for MR766, pairwise t test. 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 = 7.9×10^{-24} for PRVABC59-specific long-range interactions, and p value = 2.8×10^{-18} for MR766-specific ones, pairwise t test. Figure S3B), whereas the common interactions displayed similar folding energies (p = 0.64, pairwise t test, 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 Figures 3A and 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.

We also used PARIS to measure ZIKV RNA-RNA interactions that form *in vitro*. We observed a lower fraction of long-range interactions in ZIKV folded *in vitro* compared with *in vivo* (p value = 1.68×10^{-81} for PRVABC59, and p value = 1.50×10^{-23} for MR766, Figures S3C and S3D). This indicates the ZIKV is more extended in cells that form less compacted short-range interactions.

Architecture and Structural Models of ZIKV Genomes

Interestingly, we observed that the intramolecular interactions form clusters in the PARIS data heatmap (Figures 3A and 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 [Yu and Herman, 2005]) between the distribution of interactions in the current domain and that in the two new subdomains. 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 four domains and 28S into six domains, with the domain boundaries accurately demarcated for the two rRNAs (Figures S4A and S4B).

We applied the method to define 23 structural domains for MR766 and 24 domains for PRVABC59 (Figures 3A and 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 (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, $p < 0.001$, permutation test. Figure S4C).

Demarcating ZIKV genomes into small domains helps to accurately build their secondary structural models, as predictions have been successful for small RNAs especially when constrained with experimental probing data (Deigan et al., 2009; Wu et al., 2015). Here we used the software suite *RNAstructure* (Deigan et al., 2009) 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 (Spitale et al., 2015; Rivas et al., 2017).

The method precisely reproduced the known secondary structural elements of the 5′ UTRs and 3′ UTRs within the ZIKV genomes (compare Figures 4, S5, and 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 covariations 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 (Villordo et al., 2016). 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 five 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 (Clyde et al., 2008). 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 structural elements (Figures 4 and 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 five 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 (Table S3); many of them could mediate 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 (Chavali et al., 2017). It has been reported that one base substitution in this site disrupts its binding to MSI1 in African strains (Chavali et al., 2017). 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, Table S4). Among these, we noticed a striking long-range interaction between the 5′ UTR (2–43 nucleotides [nt]) and the E protein coding region (1,089–1,134 nt) 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 (Figures 5B and 5C). Phylogenetic analysis revealed strong conservation of these base pairs in all epidemic strains. In sharp contrast, this conservation is not observed in in pre-epidemic Asian strains or in any African strains. Interestingly, all the unconserved sites within the E protein coding region of pre-epidemic Asian strains and African strains 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 (Liu et al., 2017b). 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-E* or U1116C *mut2-E*). We also designed corresponding mutations in the 5′ UTR (Figure 5C, C36U + G29A *mut1-5′* or A18G *mut2-5′*). Finally, we engineered GZ01 mutants with the compensatory mutations in both the 5′ UTR and the E protein coding region to restore the base pairing (Figure 5C, G1095A + C1101A/C36U + G29A *res1* for *mut1-E* and *mut1-5′*, U1116C/A18G *res2* for *mut2-E* and *mut2-5′*). As a

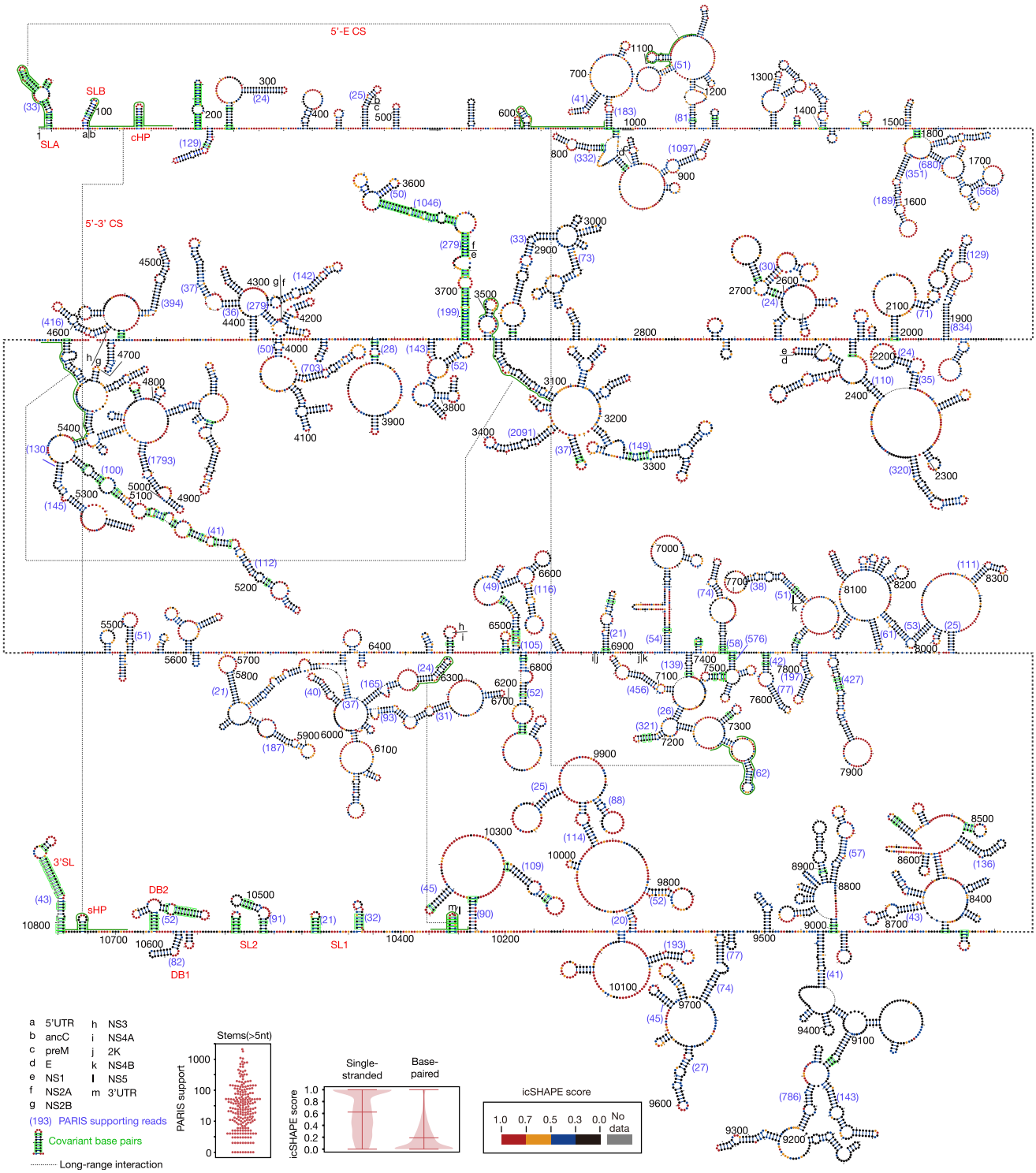
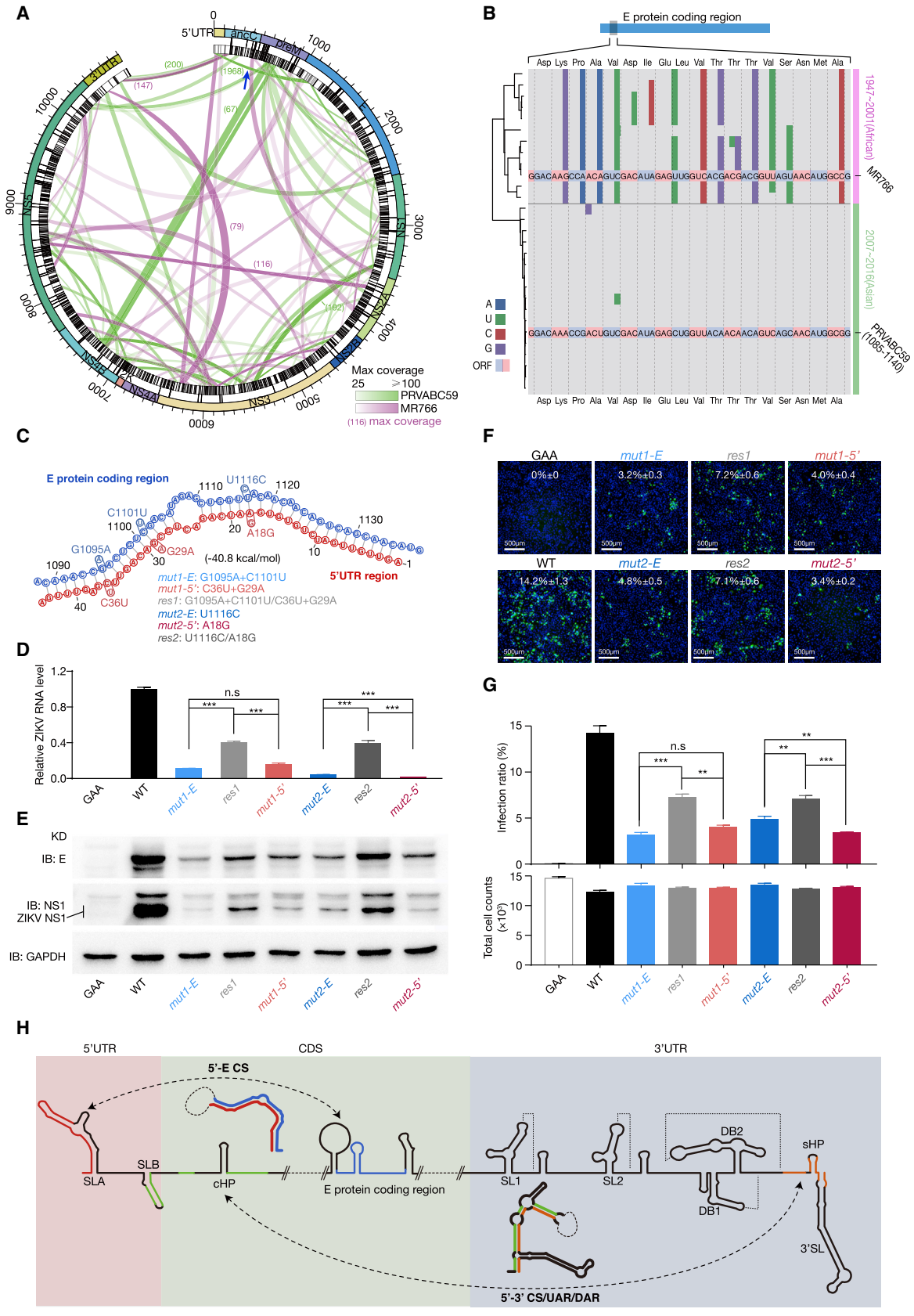


Figure 4. Full-Length Structural Model of the PRVABC59 RNA Genome

Nucleotides are colored with icSHAPE reactivity scores. Base pairs with conserved covariation are boxed with green rectangle. Long-range interactions 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 read numbers for all interactions. See also STAR Methods and Figure S5.



(legend on next page)

control, we included a non-replicating mutant ZIKV that lacks NS5 RNA dependent RNA polymerase function ("GAA") (Liu et al., 2017b). 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 virions were generated from BHK cells transfected with viral genomic RNA, which was produced from *in vitro* transcription. The virions were quantitated with qPCR and equal titer of virions of each strain was used to infect U87MG cell line and infectivity was assessed by qPCR, western blots and immunofluorescence. Similar to the GAA mutation, all of the mutations predicted to disrupt the long-range interaction (*mut1-E*, *mut1-5′*, *mut2-E*, and *mut2-5′*) substantially reduced GZ01 infectivity. Importantly, the compensatory mutations *res1* and *res2* that would restore the 5′ UTR-E interaction partially restored the infectivity in U87MG and Vero cells (Figures 5D–5G). The fact that the combination of two deleterious mutations can rescue either mutant further demonstrate the reductions in infectivity by *mut1-E* and *mut2-E* are unlikely due to codon effects on protein translation. Instead, it strongly argues for the functional importance of the 5′ UTR-E 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′ CS with the 3′ UTR (Ng et al., 2017) (Figure 5H). The switch between the translation and replication conformations is important for ZIKV infection (Liu et al., 2016). Our PARIS experiment identified a third conformation of the ZIKV 5′ UTR, pairing with the E protein coding region to form a 5′-E protein CS (5′-E CS) (Figure 5H), which may play a role in virus replication or translation regulation.

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 (Lichinchi et al., 2016; Gokhale et al., 2016). RNA structure can also reveal molecular properties and functions of viral RNAs (Mauger et al., 2015; Smola et al., 2015). Previous studies have uncovered RNA virus genome structures for HIV (Watts et al., 2009) and HCV (Mauger et al., 2015; Pirakitikulr et al., 2016). 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 UTRs. Here we report an 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 (Mauger et al., 2015). 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 using 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 Figures 2F and 4). However, this is possibly due to the *in vivo* and *in vitro* differences between RNA structures, as our experiments were performed *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.

The structural models revealed secondary structural elements within the ZIKV genomes. In particular, we noticed that some

Figure 5. An Asian Lineage-Specific Long-Range Intramolecular Interaction May Contribute to ZIKV Infection

(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) RT-qPCR quantitation of relative GZ01 viral RNA level from pellets of U87MG cells infected for 48 hr. The GAA mutant is a control with a defective NS5 RdRp domain. n = 3 biological replicates.

(E) Western blotting of the ZIKV E and NS1 proteins from pellets of U87MG cells infected with GZ01 strains in 48 hr. IB, immunoblotting.

(F) Immunofluorescence staining of ZIKV E protein in Vero cells infected for 2 days with GZ01 strains with equal viral titer as normalized by qRT-PCR.

(G) Summarized infection ratio as determined from (F) (up) and total number of cells (down). n = 3 biological replicates.

(H) 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.

Significance was determined by unpaired two-tailed t test; $^{**}p < 0.01$, and $^{***}p < 0.001$; n.s, not significant. Error bars indicate SEM. See also Table S4.

previously unknown long duplexes contain a large number of co-variations, such as duplexes at the positions ~3,600, ~7,600, and ~8,800 in the MR766 strain and at the position ~3,600 in the PRVABC59 strain. In addition, the duplex at the position ~3600 is conserved in both strains, is predicted to be very stable, and has a large number of supporting reads, e.g., 125 PARIS reads in MR766 and 1,046 reads in PRVABC59. The conservation and/or stability of these interactions suggests their possible functional significance.

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 (Yuan et al., 2017; Liu et al., 2017a). 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 strong 5′UTR-E interaction might help ribosome binding and scanning on this 5′ capped RNA genome or influence the SLA element in the 5′ UTR region and affect SLA-mediated recruitment of the Pol to the genome. The delicate balance between translation and replication regulation might alter the infectivity of the PRVABC59 strain. 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 using various techniques, including PARIS, SPLASH, and LIGR-Seq, have started to reveal pervasive and dynamic long-range RNA interactions (Lu et al., 2016; Aw et al., 2016; Sharma et al., 2016). Some of these long-range interactions are conserved among different species (Lu et al., 2016). They can play essential roles in long non-coding RNA modular organization and mRNA translation (Aw et al., 2016). Using PARIS, we systematically revealed many long-range interactions in ZIKV. We hypothesize that, in addition to the well-known 5′-3′ CS and our 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.

High-throughput technologies and system-wide analysis have started to unveil the structural landscape of RNA viruses and have provided a rich resource for the discovery of RNA structural elements that are important for virus infection (Watts et al., 2009; Mauger et al., 2015; Pirakitikulr et al., 2016; Ziv et al., 2018). RNA structures can function in a generic way, for example by affecting translation (Watts et al., 2009). They can also function in a specific way, exemplified by well-known structural elements at UTRs (Rouskin et al., 2014; Gebhard et al., 2011). A recent analysis that synonymously mutated blocks of the HIV-1 genome identified *cis*-acting elements that regulate splicing (Takata et al., 2018). In addition, it is well accepted that RNA structures influence protein and microRNA binding (Beaudoin et al., 2018; Taliaferro et al., 2016). It will be of interest to characterize the specific functions of the candidate structural elements discovered in this study, by large-scale mutational analysis. In sum, our approach and resource open a door to extensively investigate the function of RNA structures in virus infection.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines
  - Viruses
- METHOD DETAILS
  - icSHAPE Experiments in Infected Cells and on Purified RNAs
  - PARIS Experiments in Infected Cells
  - Plaque Assay for Viral Titer Identification
  - ZIKV Recovery with Constructions of Mutant and Rescue Strains
  - Real-Time SYBR Green Quantitative RT-PCR
  - Immunofluorescence Assay
  - Assemble ZIKV Genomes
  - Data Analysis of icSHAPE Experiments
  - Data Analysis of PARIS Experiments
  - Comparison of *In Vitro* and *In Vivo* PARIS Experiments
  - Algorithm to Define Structural Domains
  - Structural Model Construction
  - Phylogenetics Analysis
  - Lineage-Specific Structure and Interactions
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and four tables and can be found with this article online at https://doi.org/10.1016/j.chom.2018.10.011.

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 State Key Research Development Program of China (Grant No. 2016YFC1200300 to X.T.) and the National Natural Science Foundation of China (Grant Nos. 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.

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. J.Z. implemented domain demarcating 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.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 13, 2018
Revised: September 21, 2018
Accepted: October 24, 2018
Published: November 21, 2018

REFERENCES

Akiyama, B.M., Laurence, H.M., Massey, A.R., Costantino, D.A., Xie, X.P., Yang, Y.J., Shi, P.Y., Nix, J.C., Beckham, J.D., and Kieft, J.S. (2016). Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. Science *354*, 1148–1152.

Aw, J.G.A., Shen, Y., Wilm, A., Sun, M., Lim, X.N., Boon, K.L., Tapsin, S., Chan, Y.S., Tan, C.P., Sim, A.Y.L., et al. (2016). In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. Mol. Cell *62*, 603–617.

Beaudoin, J.D., Novoa, E.M., Vejnar, C.E., Yartseva, V., Takacs, C.M., Kellis, M., and Giraldez, A.J. (2018). Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. Nat. Struct. Mol. Biol. *25*, 677–686.

Brasil, P., Pereira, J.P., Jr., Moreira, M.E., Ribeiro Nogueira, R.M., Damasceno, L., Wakimoto, M., Rabello, R.S., Valderramos, S.G., Halai, U.A., Salles, T.S., et al. (2016). Zika virus infection in pregnant women in Rio de Janeiro. N. Engl. J. Med. *375*, 2321–2334.

Chavali, P.L., Stojic, L., Meredith, L.W., Joseph, N., Nahorski, M.S., Sanford, T.J., Sweeney, T.R., Krishna, B.A., Hosmillo, M., Firth, A.E., et al. (2017). Neurodevelopmental protein Musashi-1 interacts with the Zika genome and promotes viral replication. Science *357*, 83–88.

Clyde, K., Barrera, J., and Harris, E. (2008). The capsid-coding region hairpin element (cHP) is a critical determinant of dengue virus and West Nile virus RNA synthesis. Virology *379*, 314–323.

Cugola, F.R., Fernandes, I.R., Russo, F.B., Freitas, B.C., Dias, J.L., Guimaraes, K.P., Benazzato, C., Almeida, N., Pignatari, G.C., Romero, S., et al. (2016). The Brazilian Zika virus strain causes birth defects in experimental models. Nature *534*, 267–271.

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: interactive drawing and editing of the RNA secondary structure. Bioinformatics *25*, 1974–1975.

Deigan, K.E., Li, T.W., Mathews, D.H., and Weeks, K.M. (2009). Accurate SHAPE-directed RNA structure determination. Proc. Natl. Acad. Sci. U S A *106*, 97–102.

Deng, Y.Q., Zhao, H., Li, X.F., Zhang, N.N., Liu, Z.Y., Jiang, T., Gu, D.Y., Shi, L., He, J.A., Wang, H.J., et al. (2016). Isolation, identification and genomic characterization of the Asian lineage Zika virus imported to China. Sci. China Life Sci. *59*, 428–430.

Deng, Y.Q., Zhang, N.N., Li, X.F., Wang, Y.Q., Tian, M., Qiu, Y.F., Fan, J.W., Hao, J.N., Huang, X.Y., Dong, H.L., et al. (2017). Intranasal infection and contact transmission of Zika virus in guinea pigs. Nat. Commun. *8*, 1648.

DiChiacchio, L., Sloma, M.F., and Mathews, D.H. (2016). AccessFold: predicting RNA-RNA interactions with consideration for competing self-structure. Bioinformatics *32*, 1033–1039.

Dick, G.W., Kitchen, S.F., and Haddow, A.J. (1952). Zika virus. I. Isolations and serological specificity. Trans. R. Soc. Trop. Med. Hyg. *46*, 509–520.

Filomatori, C.V., Carballeda, J.M., Villordo, S.M., Aguirre, S., Pallares, H.M., Maestre, A.M., Sanchez-Vargas, I., Blair, C.D., Fabri, C., Morales, M.A., et al. (2017). Dengue virus genomic variation associated with mosquito adaptation defines the pattern of viral non-coding RNAs and fitness in human cells. PLoS Pathog. *13*, e1006265.

Flynn, R.A., Zhang, Q.C., Spitale, R.C., Lee, B., Mumbach, M.R., and Chang, H.Y. (2016). Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. Nat. Protoc. *11*, 273–290.

Gebhard, L.G., Filomatori, C.V., and Gamarnik, A.V. (2011). Functional RNA elements in the dengue virus genome. Viruses *3*, 1739–1756.

Gokhale, N.S., McIntyre, A.B.R., McFadden, M.J., Roder, A.E., Kennedy, E.M., Gandara, J.A., Hopcraft, S.E., Quicke, K.M., Vazquez, C., Willer, J., et al. (2016). N6-methyladenosine in flaviviridae viral RNA genomes regulates infection. Cell Host Microbe *20*, 654–665.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

Lanciotti, R.S., Lambert, A.J., Holodniy, M., Saavedra, S., and Signor Ldel, C. (2016). Phylogeny of Zika virus in Western Hemisphere, 2015. Emerg. Infect. Dis. *22*, 933–935.

Lichinchi, G., Zhao, B.S., Wu, Y., Lu, Z., Qin, Y., He, C., and Rana, T.M. (2016). Dynamics of human and viral RNA methylation during Zika virus infection. Cell Host Microbe *20*, 666–673.

Liu, Z.Y., Li, X.F., Jiang, T., Deng, Y.Q., Zhao, H., Wang, H.J., Ye, Q., Zhu, S.Y., Qiu, Y., Zhou, X., et al. (2013). Novel cis-acting element within the capsid-coding region enhances flavivirus viral-RNA replication by regulating genome cyclization. J. Virol. *87*, 6804–6818.

Liu, Z.Y., Li, X.F., Jiang, T., Deng, Y.Q., Ye, Q., Zhao, H., Yu, J.Y., and Qin, C.F. (2016). Viral RNA switch mediates the dynamic control of flavivirus replicase recruitment by genome cyclization. eLife *5*, https://doi.org/10.7554/eLife.17636.

Liu, Y., Liu, J., Du, S., Shan, C., Nie, K., Zhang, R., Li, X.F., Zhang, R., Wang, T., Qin, C.F., et al. (2017a). Evolutionary enhancement of Zika virus infectivity in *Aedes aegypti* mosquitoes. Nature *545*, 482–486.

Liu, Z.Y., Yu, J.Y., Huang, X.Y., Fan, H., Li, X.F., Deng, Y.Q., Ji, X., Cheng, M.L., Ye, Q., Zhao, H., et al. (2017b). Characterization of cis-acting RNA elements of Zika virus by using a self-splicing ribozyme-dependent infectious clone. J. Virol. *91*, https://doi.org/10.1128/JVI.00484-17.

Lodeiro, M.F., Filomatori, C.V., and Gamarnik, A.V. (2009). Structural and functional studies of the promoter element for dengue virus RNA replication. J. Virol. *83*, 993–1008.

Low, J.T., and Weeks, K.M. (2010). SHAPE-directed RNA secondary structure prediction. Methods *52*, 150–158.

Lu, Z.P., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., et al. (2016). RNA duplex map in living cells reveals higher-order transcriptome structure. Cell *165*, 1267–1279.

Mauger, D.M., Golden, M., Yamane, D., Williford, S., Lemon, S.M., Martin, D.P., and Weeks, K.M. (2015). Functionally conserved architecture of hepatitis C virus RNA genomes. Proc. Natl. Acad. Sci. U S A *112*, 3692–3697.

Mlakar, J., Korva, M., Tul, N., Popovic, M., Poljsak-Prijatelj, M., Mraz, J., Kolenc, M., Resman Rus, K., Vesnaver Vipotnik, T., Fabjan Vodusek, V., et al. (2016). Zika virus associated with microcephaly. N. Engl. J. Med. *374*, 951–958.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933–2935.

Ng, W.C., Soto-Acosta, R., Bradrick, S.S., Garcia-Blanco, M.A., and Ooi, E.E. (2017). The 5' and 3' untranslated regions of the flaviviral genome. Viruses *9*, https://doi.org/10.3390/v9060137.

Petersen, L.R., Jamieson, D.J., and Honein, M.A. (2016). Zika virus. N. Engl. J. Med. *375*, 294–295.

Pickett, B.E., Sadat, E.L., Zhang, Y., Noronha, J.M., Squires, R.B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., et al. (2012). ViPR: an open

bioinformatics database and analysis resource for virology research. Nucleic Acids Res. *40*, D593–D598.

Pirakitikulr, N., Kohlway, A., Lindenbach, B.D., and Pyle, A.M. (2016). The coding region of the HCV genome contains a network of regulatory RNA structures. Mol. Cell *62*, 111–120.

Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics *11*, 129.

Rivas, E., Clements, J., and Eddy, S.R. (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat. Methods *14*, 45–48.

Rodenhuis-Zybert, I.A., Wilschut, J., and Smit, J.M. (2010). Dengue virus life cycle: viral and host factors modulating infectivity. Cell. Mol. Life Sci. *67*, 2773–2786.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature *505*, 701–705.

Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B.J. (2016). Global mapping of human RNA-RNA interactions. Mol. Cell *62*, 618–626.

Smola, M.J., Calabrese, J.M., and Weeks, K.M. (2015). Detection of RNA-protein interactions in living cells with SHAPE. Biochemistry *54*, 6867–6875.

Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., et al. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature *519*, 486–490.

Takata, M.A., Soll, S.J., Emery, A., Blanco-Melo, D., Swanstrom, R., and Bieniasz, P.D. (2018). Global synonymous mutagenesis identifies cis-acting RNA elements that regulate HIV-1 splicing and replication. PLoS Pathog. *14*, e1006824.

Taliaferro, J.M., Lambert, N.J., Sudmant, P.H., Dominguez, D., Merkin, J.J., Alexis, M.S., Bazile, C., and Burge, C.B. (2016). RNA sequence context effects measured in vitro predict in vivo protein binding and regulation. Mol. Cell *64*, 294–306.

Villordo, S.M., and Gamarnik, A.V. (2009). Genome cyclization as strategy for flavivirus RNA replication. Virus Res. *139*, 230–239.

Villordo, S.M., Carballeda, J.M., Filomatori, C.V., and Gamarnik, A.V. (2016). RNA structure duplications and flavivirus host adaptation. Trends Microbiol. *24*, 270–283.

Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Jr., Swanstrom, R., Burch, C.L., and Weeks, K.M. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. Nature *460*, 711–716.

Wu, Y., Shi, B., Ding, X., Liu, T., Hu, X., Yip, K.Y., Yang, Z.R., Mathews, D.H., and Lu, Z.J. (2015). Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. Nucleic Acids Res. *43*, 7247–7259.

Yu, Z.H. and Herman, G. (2005). 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.

Yuan, L., Huang, X.Y., Liu, Z.Y., Zhang, F., Zhu, X.L., Yu, J.Y., Ji, X., Xu, Y.P., Li, G., Li, C., et al. (2017). A single mutation in the prM protein of Zika virus contributes to fetal microcephaly. Science *358*, 933–936.

Zhu, Z., Chan, J.F., Tee, K.M., Choi, G.K., Lau, S.K., Woo, P.C., Tse, H., and Yuen, K.Y. (2016). Comparative genomic analysis of pre-epidemic and epidemic Zika virus strains for virological factors potentially associated with the rapidly expanding epidemic. Emerg. Microbes Infect. *5*, e22.

Ziv, O., Gabryelska, M.M., Lun, A.T.L., Gebert, L.F.R., Sheu-Gruttadauria, J., Meredith, L.W., Liu, Z.Y., Kwok, C.K., Qin, C.F., MacRae, I.J., et al. (2018). COMRADES determines in vivo RNA structures and interactions. Nat. Methods *15*, 785–788.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Flavivirus Group Antigen Antibody, clone D1-4G2-4-15	EMD Millipore	Cat#MAB10216; RRID:AB_827205
Zika virus NS1 protein antibody	GeneTex	Cat#GTX133307; RRID:AB_2688020
Zika virus E antibody	GeneTex	Cat#GTX133314; RRID:AB_2747413
GAPDH antibody	ZSGB-Bio	Cat#TA-08; RRID:AB_2747414
anti-mouse Alexa Fluor 488	Invitrogen	Cat#R37120; RRID:AB_2556548
peroxidase-conjugated AffiniPure antirabbit antibody	ZSGB-Bio	Cat#ZB-2301; RRID:AB_2747412
peroxidase-conjugated AffiniPure anti-mouse antibody	ZSGB-Bio	Cat#ZB-2305; RRID:AB_2747415
Bacterial and Virus Strains		
ZIKV- PRVABC59	Cugola et al., 2016	N/A
ZIKV; MR766	Dick et al., 1952	N/A
ZIKA: GZ01	Liu et al., 2017b	N/A
ZIKV: SZ01	Deng et al., 2016	N/A
Escherichia coli strain Stbl3	TIANDZ	Cat#121112
Chemicals, Peptides, and Recombinant Proteins		
DMEM	Hyclone	Cat#SH30243.01
Fetal Bovine Serum	PAN Biotech	Cat#P30-3306
NAI-N3	Spitale et al., 2015	N/A
4'-aminomethyltrioxsalen	Sigma	Cat#A4330-5mg
Critical Commercial Assays		
RNAiso plus	Takara	Cat#9109
Rneasy Mini kit	QIAGEN	Cat#74106
HiPure Total RNA Mini Kit	Magen	Cat#R4111
Rnase-Free Dnase Set	QIAGEN	Cat#79254
Ribo-Zero Gold rRNA Removal Kit (Human/Mouse/Rat)	Illumina	Cat#MRZG12324
Click-IT biotin DIBO alkyne	ThermoFisher	Cat#C10412
RNA Fragmentation Reagents	ThermoFisher	Cat#AM8740
FastAP Thermosensitive Alkaline Phosphatase	ThermoFisher	Cat#EF0652
T4 Polynucleotide Kinase	NEB	Cat#M0201L
T4 RNA ligase 1 (ssRNA ligase), high concentration	NEB	Cat#M0437M
Single-Stranded DNA Binding Protein	Promega	Cat#M3011
5′ Deadenylase	NEB	Cat#M0331S
RecJf	NEB	Cat#M0264L
S1 Nuclease	ThermoFisher	Cat#EN0321
ShortCut Rnase III	NEB	Cat#M0245L
Proteinase K	Promega	Cat#V3021
Superase-In	ThermoFisher	Cat#AM2694
CircLigase II	Lucigen	Cat#CL9025K
Phusion HF PCR master mix	NEB	Cat#M0531L
EndoFree Maxi Plasmid Kit	TIANGEN	Cat#DP117
Ribomax SP6 large-scale RNA production kit	Promega	Cat#P1280
Ribo m^7G Cap Analog	Promega	Cat#P1711
HiPure Total RNA Mini Kit	Magen	Cat#R4111
5X All-In-One MasterMix	Abm	Cat#G490

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AccuRT Genomic DNA Removal Kit	Abm	Cat#G492
2X ChamQ SYBR Master Mix	Vazyme	Cat#Q321-02
DAPI	Invitrogen	Cat#62247
Deposited Data		
icSHAPE and PARIS sequencing	This paper	European Nucleotide Archive (ENA): PRJEB28648
Experimental Models: Cell Lines		
Huh-7	JCRB	JCRB0403; RRID:CVCL_0336
U87MG	ATCC	HTB-14; RRID:CVCL_0022
C6/36	ATCC	CRL-1660; RRID:CVCL_Z230
Vero	ATCC	CCL-81; RRID:CVCL_0059
Oligonucleotides		
ZIKV-NS5F: GGTCAGCGTCCTCTCTAATAAACG	Ribobio	(Deng et al., 2017)
ZIKV-NS5R: GCACCCTAGTGTCCACTTTTTCC	Ribobio	(Deng et al., 2017)
GAPDH-F: CGGAGTCAACGGATTTGGTCGTAT	Ribobio	N/A
GAPDH-R: AGCCTTCTCCATGGTGGTGAAGAC	Ribobio	N/A
Recombinant DNA		
GZ01-WT	Liu et al., 2017b	N/A
GZ01-GAA	Liu et al., 2017b	N/A
GZ01-mutU1116C	This paper	N/A
GZ01-mutG1095A+**C1101U**	This paper	N/A
GZ01-mut**C36U+G29A**	This paper	N/A
GZ01-mut**G1095A+C1101U-comC36U+G29A**	The paper	N/A
GZ01-mutU1116C-comA18G	This paper	N/A
GZ01-mutA18G	This paper	N/A
Software and Algorithms		
icSHAPE pipeline	Flynn et al., 2016	https://github.com/qczhang/icSHAPE
PARIS pipleine	Lu et al., 2016	https://github.com/zhipenglu/
Trinity	Grabherr et al., 2011	https://github.com/trinityrnaseq/trinityrnaseq/wiki
RNAstructrue	Reuter and Mathews, 2010	https://rna.urmc.rochester.edu/RNAstructure.html
VARNA	Darty et al., 2009	http://varna.lri.fr/
R-scape	Rivas et al., 2017	http://eddylab.org/R-scape/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Qiangfeng Cliff Zhang (qczhang@tsinghua.edu.cn).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

Huh-7 cells (human, sex: male), BHK21 cells (Mesocricetus auratus, sex:male) and Vero cells (Chlorocebus sabaeus, sex: female) were cultured in Hyclone DMEM medium with 10% FBS (PAN Biotech) in an incubator at 37°C with 5% CO_2, and C6/36 cells (mosquito, sex: unspecified) were maintained with the same condition but at 28°C.

Viruses

ZIKV PRVABC59 infections were performed in Huh-7 cells with MOI of 1 for 72 hr. The infections of ZIKV MR766 were performed in the same cell line for 72 hr with MOI of 0.1.

METHOD DETAILS

icSHAPE Experiments in Infected Cells and on Purified RNAs

The icSHAPE experiments were performed in ZIKV-infected Huh7 cells 72 h.p.i as previously described with some modifications (Spitale et al., 2015; Flynn et al., 2016). Briefly, cells were harvested with trypsin and washed with PBS to remove adhered virus particles. Then for *in vivo* modification and control samples, cells were incubated with 0.1M NAI-N_3 diluted in PBS or same proportion of DMSO in PBS at 37°C for 5 min on rotation. The incubation was stopped by centrifugation of 2500g for 1min at 4°C, and supernatant was discarded. Cells were then lysed with RNAiso plus (TaKaRa) and phases were split with adding of chloroform. The aqueous phase was purified with RNeasy Mini kit (Qiagen) to obtain total RNA. On-column DNA digestion was performed by RNase-Free DNase Set (Qiagen) during purification. The rRNA fraction were removed with Ribo-zero Gold kit (Illumina) from the total RNA sample. For *in vitro* NAI-N3 modification, unmodified RNA was extracted from cells, followed by rRNA depletion. Then ribo-zero RNA was refolded in 1 X SHAPE reactivity buffer (100mM HEPES, 100mM NaCl, 6mM $MgCl_2$) and modified with 0.1M NAI-N_3 in 37°C for 10min.

The *in vivo* modified RNA, *in vitro* modified RNA and control unmodified RNA samples were added with 2μL 1.85mM DIBO-biotin (ThermoFisher) solution and 1μL RiboLock (ThermoFisher) and stirred on Thermomixer at 37°C for 2 hr. The biotinylated RNA was fragmented by fragmentation reagent (Ambion). The appending of adapters to 3′ of RNA was started with end repairing by resuspending with 10 μL end repairing mix (70 mM Tris 7.0, 18mM $MgCl_2$, 5mM DTT, 4 U/μL RiboLock, 0.1 U/μL FastAP (Life Technology), 2 U/μL T4 PNK(NEB)) and kept at 37°C for 1 hr. Then 10 μL ligation mix (5mM DTT, 1.25μM 5′ adenylylated and 3′-blocked linker (3′-bio for unmodified, 3′-ddc for modified), 0.66 U/μL T4 RNA ligase (NEB, M0437M), 15% PEG8000, 1X RNA ligase buffer) was added into each sample at 25°C for extra 3 hr. The purified mix of 3′ adaptor-ligated RNAs and additional adaptors was put into 1μL FastAP, 0.2μL SSB (Promega), 0.8μL Ribolock and 1μL 5′ Deadenylase (NEB) in 1X NEB buffer 2 (NEB) and kept at 30°C for 90 min. And then 1μL RecJf (NEB) was added in to remove 3′ adaptors with 5′ exonuclease reaction by incubation at 37°C for 1 hr. The following procedures, including reverse transcription, biotin-streptavadin enrichment, size selection of cDNA, circularization and PCR amplification, were the same as the standard protocol (Flynn et al., 2016).

PARIS Experiments in Infected Cells

The PARIS experiments were performed in ZIKV 72 h.p.i Huh7 cells as previously described with some modifications (Lu et al., 2016). Briefly, ZIKV-infected 72.h.p.i cells were incubated with AMT-DMEM media (40% v/v media, 0.5mg/mL in 1× PBS) for 30min in 37°C, 5% CO2. After the incubation, the media was replaced with 0.5mg/mL AMT in 1× PBS. Cells were crosslinked in a Stratalink 2400 UV crosslinker with 365nm UV on ice bed in a distance of 15cm to bulbs. The crosslinked cells were collected by scraping, washed with PBS. Cell pellets were snap frozen with liquid nitrogen and lysed by urea/SDS buffer (4M urea, 1% SDS) thoroughly. Each tube of cells were incubated with 2μL S1 nuclease (Thermofisher EN0321) in 1x S1 nuclease buffer at room temperature in thermomixer for 10min. The digestion was stopped by adding 10% SDS and Protease K (PK) to a final concentration of 1% SDS and 10μg/mL PK respectively and reacted at 50°C for 30min. RNA was purified by the standard Trizol LS protocol and further digested with ShortCut RNase III (NEB M0245L) as manufacturer's manual description.

AMT crosslinked dsRNA fragments were separated by 2D gel electrophoresis (by 160V, 50min in 12% native PAGE for the first dimension and 50W, 35min in 20% Urea-TBE denatured PAGE for the second dimension). Proximity ligation was performed to purified dsRNA fragments by adding 5μL T4 RNA ligase (NEB M0437M), 0.3μLATP, 1μL Superase-In (Thermofisher) and keeping the mix in a final concentration of 1X T4 RNA ligase buffer overnight. The reaction was stopped by boiling at 95°C for 2 min and RNA was precipitated by ethanol. The RNA was reverse-crosslinked by irradiating with 254UV for 10 min on ice bed. The library preparation for RNA was performed as the icSHAPE protocol (Flynn et al., 2016). Briefly, the RNA was ligated with 3′-end biotin adapters and redundant adapters was removed by RecJf (NEB). After reverse transcription and biotin enrichment, cDNA was circularized with CircLigase II (Lucigen CL9025K) and amplified with Phusion HF PCR master mix (NEB M0531L). Size selected libraries were sequenced on Hiseq 2500 with single end 75bp reads.

Plaque Assay for Viral Titer Identification

Titration (in PFU mL^{-1}) of each C6/36 subculture was obtained by plaque assay to determine the amount of infectious viral particles (PFU). The virus titration was performed in Vero cells, and in DMEM medium with 2% FBS. Briefly, 0.6× 10^6/mL Vero cells were seeded in each well of a 6-well plate for 12h at 37°C to allow Vero cell adherence. Then, a serial dilution of each virus stock from ZIKV SZ01 subculture in DMEM medium (without FBS, 1% penicillin/streptomycin, and 1% glutamine or GlutaMX) was performed from 10^{-1} to 10^{-9}. Then, 400 μL of each dilution was added in each well for virus adsorption 1-2 hr. After this, the virus was discarded and each well was overlaid with 4mL 2% DMEM media containing final 0.8% complete LMP agarose (42°C). After 5 days of incubation at 37°C, the plaque visualization was made using 4% PFA fixation and 0.2% crystal violet in 20% ethanol staining to visualize the plaques. The most appropriate viral dilution was estimated to determine the amount of infected cells visible (PFU mL^{-1}). All the subculture aliquots were stored -80°C.

ZIKV Recovery with Constructions of Mutant and Rescue Strains

The recovery of ZIKV from plasmids was performed as described before (Liu et al., 2017b). We mutated specific site on the GZ01 mRNA with site-directed mutagenesis technique to gain the mutant of wanted, and the mutant clones were amplified in Stbl3 competent cells (TIANDZ, 121112) at 30°C, and maxi-prepared (TIANGEN, DP117) following the manufacture's manual. *In vitro* transcription

of all linearized ZIKV plasmids was performed by using the Ribomax SP6 large-scale RNA production kit (Promega, P1280) with Ribo m^7G Cap Analog (Promega, P1712) as manufacturer's description. The RNA products were spliced in splicing buffer (40 mM Tris-HCl , 1 M NH_4Cl, 10 mM $MgCl_2$, and 0.02% SDS) at 45°C for 1 hr. The transfection of equal molar RNA into BHK21 with Lipofectamine 2000 (Thermo Fisher Scientific) for 3 days. Then the viruses with equal volume supernatant from transfected BHK21 were propagated in C6/36 cells, incubating for ten days. And the virus subculture were centrifuged to discard cell pellets and aliquoted to store at -80°C until use.

Real-Time SYBR Green Quantitative RT-PCR

Total RNA was extracted from Huh7, U87MG, Vero cells or supernatant samples using HiPure Total RNA Mini Kit (Magen, R4111). All RNA pellets were resuspended in 30-50 μL of RNase-free distilled water, quantified using a DeNovix DS-11+ spectrophotometer (DeNovix) and stored at −80°C. The set of primers specific for ZIKV were synthesized by Ribobio:

ZIKV-NS5F: GGTCAGCGTCCTCTCTAATAAACG,
ZIKV-NS5R: GCACCCTAGTGTCCACTTTTTCC.
GAPDH-F: CGGAGTCAACGGATTTGGTCGTAT.
GAPDH-R: AGCCTTCTCCATGGTGGTGAAGAC.

RT reaction was performed with 20μL of each sample to cDNA synthesis using 5x All-In-One MasterMix (Abm, G490) or (with AccuRT Genomic DNA Removal Kit) (abm:G492) by incubation at 25°C for 10min, 42°C for 15min, 85°C for 5min and 10μL of 2xChamQ SYBR Master Mix qPCR reagents (Vazyme, Q321-02). The amplification was done in an Bio-Rad CFX Manager 3.1 real-time PCR system by incubation at 95°C pre-denaturation for 3min, then amplified with 39 cycles (95°C for 10s, 60°C for 30s, and 95°C for 15s), followed by a melt curve process (65°C-95°C with increment 0.5°C for 0.05s). The real-time data were analyzed using Bio-Rad CFX Manager 3.1 system or GraphPad Prism. For the detection and quantification of viral RNA, the real-time PCR of each sample was normalized with a threshold cycle (Ct) value to the internal reference gene GAPDH by the $2^{-\Delta\Delta Ct}$ method:

$$\text{Normalized viral RNA expression level} = 2^{-(\Delta Ct1 - \Delta Ct2)}$$

In the equation, ΔCt1=Ct(viral RNA) - Ct(GAPDH), and ΔCt2=Average of ΔCt1(WT biological replicates)

Immunofluorescence Assay

Seven thousand Vero cells were seeded into each well in a 96-well plate, and infected with equal viral titer inoculum of the ZIKV GZ01 wildtype and mutants for incubation about 48 hr since the next day. And for co-localization analysis, the 24-well cell climbing slices were placed into a 24-well plate, and seeded with $5x10^4$ cells for each, then infected with MOI of 2 SZ01 ZIKV for 2 days since the next day. Then, those cells were fixed by directly adding equal volume of 4% paraformaldehydel for 15 mins at room temperature. After washing, cells were permeabilized with 0.2% Triton X-100 diluted in PBS for 15 mins. The following primary antibodies in antibody dilution buffer (ADB) directed against the following were added: Anti-Flavivirus Group Antigen Antibody, clone D1-4G2-4-15 (polyclonal mouse, Millipore, MAB10216, 1:1000). Cells were then incubated overnight at 4°C. The following secondary antibodies were added for a one-hour incubation at room temperature: anti-mouse Alexa Fluor 488 (Invitrogen, R37120, 1: 1000 in PBS). After washing for three times, the nuclei were stained with DAPI (Invitrogen, 62247, 1:10,000) diluted in a PBS 1× solution for 5 mins. Images were acquired by high content analysis (Thermo: Cellomics ArrayScan VTI) and a Zeiss upright confocal microscope, to obtain the virus infection ratio by Cellomics (Thermo).

Assemble ZIKV Genomes

We used *trinity* (Grabherr et al., 2011) and icSHAPE DMSO library sequence data to assemble the genome sequences of two ZIKV strains with default parameters. We confirmed the assembled genome with the original version from GenBank (GenBank: KU501215.1 for PRVABC59, GenBank: AY632535.2 for MR766). The sequences are basically the same with a few nucleotide substitutions and an extra fragment at the 3′ end.

Data Analysis of icSHAPE Experiments

icSHAPE scores were calculated as previously described (Flynn et al., 2016) from the high-throughput sequencing data. To improve icSHAPE data visualization, we used a sliding window (window size: 30nt, window step: 1nt) to smooth the raw data which is base-resolution. All processed data were divided by their median.

For the analysis of icSHAPE score conservation, we aligned the two ZIKV genomes and then used a sliding window (window size: 100nt, window step: 20nt) to scan the genomes and calculate the sequence identity and the Pearson correlation coefficient of the icSHAPE scores in each window.

Data Analysis of PARIS Experiments

PARIS duplex groups were called as previously described (Lu et al., 2016). Each duplex group was supported by at least 10 reads for the *in vivo* datasets and 3 reads for *in vitro* datasets.

Comparison of *In Vitro* and *In Vivo* PARIS Experiments

We obtained ~22,000 reads for the *in vitro* PARIS experiment of PRVABC59, ~110,000 reads for *in vivo* PRVABC59; ~7,700 reads for *in vitro* MR766, ~98,000 reads for *in vivo* MR766. In the comparison of *in vivo* and *in vitro* interactions, we down-sampled the same number of reads from *in vivo* datasets as that of *in vitro* datasets to balance the size of data. The raw gapped reads were converted to a matrix with each dimension equal to the length of virus. The matrix was then condensed to a 300×300 heatmap. The value of each cell in the condensed heatmap equals to the average for all corresponding cells in the original matrix. Those cells with a value greater than a cutoff in both the *in vivo* and the *in vitro* datasets are filled with blue, and red if only in *in vivo* datasets, green if only in *in vitro* datasets. The cutoff was defined as 1.0 for PRVABC59, and 0.5 for MR766.

Algorithm to Define Structural Domains

We used a two-step algorithm to split a full RNA secondary structure into k separate domains based on PARIS data. In the first step, the full RNA is seen as a single domain, and we try to find a site to split it into two subdomains. Each domain has a distribution of PARIS interaction signals and we used earth movers' distance (Yu and Herman, 2005) to measure the difference between two distributions. To find the best site, we maximized the difference of distribution of PARIS interactions between the domain and two proposed subdomains. We next split all subdomains iteratively and this yields a hierarchical domain partition. In the second step, we try to find a best top-down path in the hierarchy for splitting the RNA into k domains. The quality of each domain is measured by the coefficient of variation (CV, σ/μ). So, we choose the subdomain with a larger CV to split at each iteration. It will result in k domains with the largest coefficient of variation minimized.

Structural Model Construction

To construct RNA secondary structural models for complete ZIKV genomes, we used the *Fold* program in *RNAstructure* software suite (Reuter and Mathews, 2010) to predict secondary structure with icSHAPE score as the pseudo-energy constraint. We first tried all parameters to predict the structure of the 3′ UTR, then chose the parameter with the highest accuracy in the structural model construction for each domain separately.

We visualized RNA structure with *VARNA* (http://varna-gui.software.informer.com/).

Phylogenetics Analysis

For each domain of the PRVABC59 or MR766 RNA genomes, we used *cmbuild* and *cmcalibrate* programs in the Infernal 1.1 software suite (Nawrocki and Eddy, 2013) to build and calibrate a covariance model with the sequence and the predicted structure of the domain . Then we used *cmsearch* to search aligned homologous sequences from 4,256 Flaviviridae virus genomes. All Flaviviridae virus genomes were downloaded from ViPR (Virus Pathogen Database and Analysis Resource, Pickett et al., 2012). Covariance in the resulting alignment was calculated with *R-scape* (Rivas et al., 2017) with default parameters.

Lineage-Specific Structure and Interactions

We first aligned the two ZIKV genomes and then used a sliding window (window size: 5nt; window step: 1nt) to scan the genomes for the regions that contain more than 3 nucleotides with icSHAPE difference more than 0.6. We combined the overlapping windows and obtained a list of structural divergent regions. We downloaded all ZIKV full genomes from ViPR (Pickett et al., 2012), and defined those genomes with less than 700 mutations in the CDS region to the MR766 genome (and the PRVABC59 genome respectively) as the MR766 lineage (and the PRVABC59 lineage respectively). From the above list of structural divergent regions we defined those with less than 25 mutations in a lineage as lineage-specific structures. To search for lineage-specific PARIS interactions, we kept those interactions only in MR766 or PRVABC59 and defined lineage-specific interactions the same as described above.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were carried out using Prism software (GraphPad). Data are presented as mean ± sem. All experiments were biologically repeated for at least three times ($n \geq 3$). Unpaired two tailed t tests were used to calculate p values. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

DATA AND SOFTWARE AVAILABILITY

The accession number for the icSHAPE and PARIS sequencing data reported in this paper is European Nucleotide Archive (ENA): PRJEB28648. The scripts used in this project are available from github (https://github.com/lipan6461188/ZIKV-RNA-Structure).