

1 Mapping RNA-capsid interactions and RNA secondary structure within authentic virus particles  
2 using next-generation sequencing

3

4 Yiyang Zhou<sup>1,2</sup>, Andrew Routh<sup>1,2</sup>

5

6 1) Department of Biochemistry and Molecular Biology, The University of Texas Medical  
7 Branch, Galveston, TX, USA

8 2) Sealy Centre for Structural Biology and Molecular Biophysics, University of Texas  
9 Medical Branch, Galveston, Texas, USA.

10

11 Keywords: Flock House virus, virus assembly, RNA packaging, photo-activatable ribonucleoside  
12 cross-linking, DMS-MaPseq

13 **Abstract**

14 To characterize RNA-capsid binding sites genome-wide within mature RNA virus particles, we  
15 have developed a Next-Generation Sequencing (NGS) platform: Photo-Activatable  
16 Ribonucleoside Cross-Linking (PAR-CL). In PAR-CL, 4-thiouracil is incorporated into the  
17 encapsidated genomes of authentic virus particles and subsequently UV-crosslinked to adjacent  
18 capsid proteins. We demonstrate that PAR-CL can readily and reliably identify capsid binding  
19 sites in genomic viral RNA by detecting crosslink-specific uridine to cytidine transitions in NGS  
20 data. Using Flock House virus (FHV) as a model system, we identified highly consistent and  
21 significant PAR-CL signals across virus RNA genome indicating a clear tropism of the  
22 encapsidated RNA genome. Certain interaction sites correlate to previously identified FHV RNA  
23 motifs. We additionally performed dimethyl sulfate mutational profiling with sequencing (DMS-  
24 MaPseq) to generate a high-resolution profile of single-stranded genomic RNA inside viral  
25 particles. Combining PAR-CL and DMS-MaPseq reveals that the predominant RNA-capsid sites  
26 favor double-stranded RNA regions. We disrupted secondary structures associated with PAR-CL  
27 sites using synonymous mutations, resulting in varied effects to virus replication, propagation,  
28 and packaging. Certain mutations showed substantial deficiency in virus replication, suggesting  
29 these RNA-capsid sites are multifunctional. These provide further evidence to support that FHV  
30 packaging and replication are highly coordinated and inter-dependent events.

31

32 **Importance**

33 Icosahedral RNA viruses must package their genetic cargo into the restrictive and tight confines  
34 of the protected virions. High resolution structures of RNA viruses have been solved by Cryo-EM  
35 and crystallography, but the encapsidated RNA often eluded visualization due to the icosahedral  
36 averaging imposed during image reconstruction. Asymmetrical reconstructions of some  
37 icosahedral RNA virus particles have revealed that the encapsidated RNAs conform to specific  
38 structures, which may be related to programmed assembly pathway or an energy-minima for RNA  
39 folding during or after encapsidation. Despite these advances, determining whether encapsidated  
40 RNA genomes conform to a single structure and determining what regions of the viral RNA  
41 genome interact with the inner surface of the capsid shell remains challenging. Furthermore, it  
42 remains to be determined whether there exists a single RNA structure with conserved topology in  
43 RNA virus particles or an ensemble of genomic RNA structures. This is important as resolving  
44 these features will inform the elusive structures of the asymmetrically encapsidated genomic  
45 material and how virus particles are assembled.

## 46 **Introduction**

47 Flock House virus (FHV) is a non-enveloped, single-stranded positive-sense RNA (+ssRNA) virus  
48 from the family *Nodaviridae*. The small bipartite genome comprising RNA 1 (3.1kb) and RNA 2  
49 (1.4kb) is packaged into a 34 nm non-enveloped T=3 icosahedral virion. Only two non-structural  
50 proteins are produced by FHV: the RNA-dependent RNA polymerase (RdRp) and sub-genomic  
51 RNA encoded protein called B2. The B2 protein was discovered as the virus's approach to evade  
52 the invertebrate anti-viral RNA silencing machinery (1, 2), which thereafter led to the discovery of  
53 similar mechanisms in plant cells (3). FHV is perhaps the best studied *alphanodavirus* and  
54 provides a powerful model system by virtue of its small genome size (4.5kb), genetic tractability  
55 and ability to infect *Drosophila* and mosquito cells in culture and whole flies (reviewed in (4, 5)).  
56 More recently, FHV has been adapted into medical field. FHV-related vaccine developments  
57 utilized either the viral particle as antibody-display system (6), or the viral RNA as trans-  
58 encapsidated chimeric viral vaccine platform (7-9).

59 Both authentic virions of FHV and the related Pariacoto virus have been reconstructed by cryo-  
60 EM and X-ray crystallography to reveal highly ordered dodecahedral cages of RNAs (10, 11). The  
61 X-ray structure of FHV virion showed electron density at the icosahedral 2-fold axis, which was  
62 modelled as an ordered RNA duplex of approximate 20 nucleotides (12). This would account for  
63 1800nts (more than one third) of the viral genome, implicating a highly-ordered and specific set  
64 of interactions between the viral protein capsid and the encapsidated genome. Interestingly, and  
65 recombinantly expressed virus-like particles (VLPs) of FHV also exhibit a similar dodecahedral  
66 RNA cage despite packaging predominantly cellular RNAs indicating that viral capsid may either  
67 impose structure upon the encapsidated RNA or select for natively structured host RNAs such as  
68 ribosomal RNAs (13, 14). However, as these structures are obtained with icosahedral averaging,  
69 we still do not know what regions or sequences of viral genomic RNA comprise the RNA cage.  
70 Furthermore, it remains to be determined whether there exists a single RNA structure with  
71 conserved topology in FHV virions, or rather an ensemble of related genomic RNA structures.

72 The FHV encapsidation process also remains largely unknown. One molecule of each RNA 1 and  
73 2 is specifically encapsidated into virus particles (15), while subgenomic RNA 3 is excluded (16).  
74 Several components of the capsid protein such as the arginine-rich motif and the C-terminal  
75 FEGFGF motif have been demonstrated to be essential determinants of packaging specificity of  
76 RNA 1, RNA 2, or both (17-19). It was also speculated that FHV packaging process may be in  
77 close association with viral replication and/or translational events (20-23). In the virus genome,  
78 one stem-loop structure in RNA 2 proximal to 5' end was demonstrated to be required for RNA 2  
79 packaging (24). However, it remains unclear whether there are similar packaging sites on RNA 1  
80 or 2, and how these sites interact and thus recruit capsid protein to fulfill virus encapsidation.

81 Next-generation sequencing (NGS) in combination with crosslinking techniques provides a high-  
82 throughput approach to study transcriptome-wide RNA-protein interactions (reviewed in (25)). A  
83 number of new technologies have successfully described interactions between RNA-binding  
84 proteins (RBPs) and different types of RNAs, including nascent transcripts, mRNAs, microRNAs  
85 and ribosomal RNAs. Among these, PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced  
86 Crosslinking and Immunoprecipitation) (26) utilizes a 365 nm UVA-activatable ribonucleoside  
87 analog 4-thiouridine (4SU) to effectively crosslink RNA to bound proteins. The enriched  
88 crosslinked RNAs result in a highly specific U to C mutation during NGS library preparation (27-

89 29), granting the ability to rapidly identify RBP and microRNA target sites on a transcriptome scale  
90 (26).

91 In an analogous fashion to PAR-CLIP, here we applied the same principles to study the interaction  
92 of FHV genomic viral RNA in the context of assembled authentic virions. Unlike the complex  
93 cellular micro-environment, authentic virions represent a highly simplified enclosure with few well-  
94 defined components (viral RNA and capsid). Therefore, we are able to screen for specific in virion  
95 RNA-capsid interaction events without interference from other cellular components. Furthermore,  
96 since viruses can be readily separated from other cellular components, we avoided the need of  
97 immunoprecipitation for RNA recovery, and thus largely simplify PAR-CLIP methodology. This  
98 method is hence named '*PAR-CL*'.

99 Using FHV as a model system, PAR-CL methodology was validated by determining that the  
100 increased U to C (U-C) mutation rate was highly specific to crosslink between viral RNA and  
101 capsid. We noticed that the intensity of PAR-CL signals was subjected to the dose of 4SU and  
102 time of incubation. Triplicated FHV PAR-CL experiments revealed significant and highly  
103 consistent PAR-CL signals across genome, which implicated a clear tropism of RNA cage inside  
104 capsid shell. The multiple clusters of PAR-CL sites suggest that FHV encapsidation may require  
105 multiple synergetic packaging sites. DMS-MaPseq (dimethyl sulfate mutational profiling with  
106 sequencing) was used to chemically probe single-stranded FHV genomic RNA in virions. We thus  
107 constructed a whole genome DMS-MaPseq-imposed RNA secondary structure map for FHV. We  
108 noticed RNA-capsid interaction sites favored double stranded RNA regions. Synonymous  
109 mutations were designed to disrupt predicted PAR-CL sties in dsRNA regions, which resulted in  
110 varied effects to virus RNA replication, propagation, and virulence. Mutations over certain PAR-  
111 CL sites showed evidential deficiency in RNA replication, suggesting these sites serve a  
112 multifunctional role in both virus packaging and replication. This provides further evidence to  
113 support that FHV packaging and replication are highly synchronized and inter-dependent events.

## 114 **Results**

### 115 **Photoactivatable-ribonucleoside-enhanced crosslinking (PAR-CL) in virus particles**

116 PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation)  
117 is a well-established method for identification of RNA-protein binding sites and can provide  
118 nucleotide-resolution through analysis of specific uridine to cytosine transitions that occur at the  
119 site of RNA-crosslinking during cDNA synthesis (26, 30, 31). Here, we simplified the technique  
120 by applying a similar approach to purified authentic virions of RNA virus, thereby removing the  
121 necessity of immunoprecipitation, and hence deriving the name “PAR-CL”. A schematic of the  
122 process is illustrated in **Figure 1**. In PAR-CL, photoactivatable nucleotide 4-thiouridine (4SU) was  
123 provided to cells in culture prior to infection with Flock House virus (FHV). 4SU is rapidly taken by  
124 cultured cells without significant cytotoxic effect (26, 32). Upon infection, 4SU is randomly  
125 incorporated into newly synthesized viral RNA, which is subsequently packaged into authentic  
126 virus particles of FHV. Virus particles were isolated using established methods for virus  
127 purification by ultracentrifugation (13). Next, purified virus particles were subjected to UV 365 nm  
128 irradiation, yielding crosslinks between the thio-group of 4SU in the viral genomic RNA and amino  
129 acid residues in the protein capsid shell only if they are in close proximity (32, 33). Virus particles  
130 were then disrupted by proteinase K treatment and a pool of viral RNAs with varied crosslinking  
131 sites was obtained. We then generated canonical random-primed RNAseq libraries using  
132 ClickSeq (34-36) for Illumina SE150 read sequencing on a MiSeq platform.

133 The raw sequencing reads were trimmed and quality filtered using *cutadapt* (37) and the *FASTX*  
134 *toolkit* ([http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)). Reads were aligned to the FHV  
135 genome with *Bowtie* (*v1.0.1*) (38). The read coverage at each genomic position and the frequency  
136 of each nucleotides found in the mapped reads was enumerated using *samtools* and the mutation  
137 rates at each genomic position was calculated.

138 To validate the PAR-CL methodology, we first sought to determine if there was a substantial  
139 increase in U-C mutation rate as a specific consequence of 4SU-capsid crosslink. We performed  
140 a series experiments in which wild-type (wt) FHV without 4SU (4SU-) or 4SU-containing FHV  
141 (4SU+) were treated with (UV+) or without (UV-) UV irradiation. As illustrated for FHV RNA 1 in  
142 **Figure 2a**, we plot the measured U-to-C (U-C) mutation rate across the genome and calculate  
143 the fold change at each U position. A small number of positions, such as nt. 1259 on RNA 1,  
144 showed high U-C mutation rates in both conditions. This possibly reflects the selection of a  
145 minority variant during virus passaging. Other than this, we did not notice an increased U-C  
146 mutation rate for UV-irradiated FHV in the absence of 4SU (4SU-/UV+). This indicated as  
147 expected that UV irradiation alone was not sufficient to induce novel U-C mutations. Similarly, we  
148 measured the influence of 4SU substitution in FHV genomic RNA without UV exposure  
149 (4SU+/UV-) (**Figure 2b**). We also did not notice an increase in U-C mutation rate. We only  
150 observed increased U-C mutations when 4SU and UV irradiation both were present (4SU+/UV+)  
151 (**Figure 2c**). This confirms the elevated U-C mutation rate is indeed a specific result of 4SU-  
152 induced crosslinking. The FHV RNA 2 data of these experiments is shown in **Supplemental**  
153 **Figure S1**.

154 Histograms of the U-C mutation rate frequencies at all genomic U positions are shown in **Figure**  
155 **2d**. This demonstrated that under 4SU+/UV+ condition, more U positions exhibited high U-C  
156 mutation rate ( $\geq 0.3\%$ ) than controls (4SU-/UV- and 4SU+/UV-). Interestingly, we noticed reduced  
157 U-C mutation rates when wt FHV was exposed to UV (4SU-/UV+), for an unknown reason. We

158 also sought to determine if 4SU incorporation and UV exposure would induce any non-specific  
159 (non-U-C) mutations. A histogram of the frequencies of all non-U-C mutations (A,C,G mutations  
160 and U-A, U-G mutations) over all genomic positions is shown in **Figure 2e**. Importantly,  
161 4SU+/UV+ FHV did not show any significant change in non-U-C mutations. We therefore  
162 conclude that the increased U-C mutation rate is specific to 4SU-induced crosslinking.

### 163 **Magnitude of PAR-CL signals was associated with 4SU dose and incubation time.**

165 Our PAR-CL method requires no immunoprecipitation to recover and enrich for crosslinked RNAs.  
166 However, this permits wild type uridine and/or uncrosslinked 4-thiouridine to persist in the RNA  
167 pool which may dilute the PAR-CL signal. To investigate optimum conditions for PAR-CL, we  
168 conducted three parallel experiments (**Figure 3**). S2 cells were infected with FHV and incubated  
169 with 4SU at 100 $\mu$ M (4SU16h) or 150 $\mu$ M (4SU1.5X). Viruses were harvested from infected cells  
170 16 h post-infection. In a third experiment, we extended the incubation time to 40 h, and applied  
171 4SU in a “prime-boost” manner, to reach a final concentration of 200 $\mu$ M (4SU40h). In each  
172 experiment, U-C mutation profile of 4SU+/UV+ virus was compared with correspondent 4SU+/UV-  
173 virus to yield PAR-CL signals (fold change). Results for FHV RNA 1 are shown in **Figure 3a** and  
174 RNA 2 in **Supplemental Figure S2**. We noticed that the concentration of 4SU and the incubation  
175 time of FHV/4SU had an impact on PAR-CL signal intensities over a number of genomic U  
176 positions.

177 The same results were observed when we plotted the frequency of PAR-CL signals for these  
178 three experiments, as well as two controls (**Figure 3b**). This shows that while the mean PAR-CL  
179 signals and the distribution under all three conditions (4SU16h, 4SU1.5X, and 4SU40h) and a  
180 4SU+/UV- control were all comparable, the magnitude of outliers showed correlation to  
181 experimental conditions (4SU40h > 4SU1.5X > 4SU16h). This indicates that only certain 4SU  
182 substitution sites are available for crosslink and therefore sensitive to the varied 4SU  
183 concentrations/incubation times. Again, for an unknown reason, the 4SU-, UV+ control (**Figure**  
184 **3b**) showed a slightly lesser than 1-fold change PAR-CL signal. Importantly, this does not interfere  
185 with our interpretation of PAR-CL signals in crosslinking samples.

186 We sampled the top 5% of PAR-CL signals in each conditioned experiment (**Figure 3c**) and  
187 concluded that, the 4SU40h group showed significantly higher PAR-CL signals than the rest. For  
188 this reason, the 4SU40h experimental condition was applied to all further PAR-CL experiments,  
189 unless otherwise mentioned. Despite the varied PAR-CL signal intensities under different  
190 experimental conditions, the PAR-CL signals presented good Pearson’s correlation coefficient  
191 ( $\geq 0.6$ ) between these parallel experiments (**Supplemental Figure S3**), which indicates high  
192 reproducibility of PAR-CL experiments.

### 193 **Consistent PAR-CL signals indicates structural tropism of encapsidated viral RNAs**

195 We applied the 4SU40h condition in three parallel experiments: three separate S2 cell cultures  
196 were incubated with virus and 4SU, and individually purified viruses were exposed to UV and  
197 thereafter proceeded to sequencing. Similar to before, PAR-CL experiments were conducted in  
198 pairs, with each PAR-CL dataset comprised of one crosslinked sample (4SU+/UV+) and control  
199 with the same sample but without UV irradiation (4SU+/UV-). To ensure reliable mutation rate  
200 calculation, we selected for U positions with coverage of at least 10,000 reads. This allowed us

201 to detect reliable mutation profiles over U34 – U3034 on RNA 1, and U9 – U1337 on RNA 2. The  
202 PAR-CL signals of these three experiments were compared on each U position on viral RNA  
203 genomes (**Figure 4a, b**). We observed good Pearson’s correlation coefficient ( $\geq 0.6$ ) between  
204 these replicates (**Supplemental Figure S4**). To validate the consistent PAR-CL signals, the  
205 signals over every U position were box-plotted over the triplicates (**Supplemental Figure S5**).  
206 This allows us to readily measure the mean signal strength and signal variation over the triplicates.  
207 In order to distinguish reliable crosslinking sites and avoid potential false positives, we removed  
208 any PAR-CL signals in crosslinked sample by applying a conservative background threshold filter,  
209 retaining only the highest 5% of PAR-CL signals in the uncrosslinked control sample. (**Figure 4c,**  
210 **d, Supplemental data 3**). Among the most consistent PAR-CL sites (passed background  
211 threshold in all three replicates), we identified 20 sites in RNA 1 and 8 sites in RNA 2 that showed  
212 the highest average PAR-CL signals (**Figure 4e, Supplemental data 3**). T-test revealed most of  
213 these sites have significantly ( $P < 0.05$ ) higher PAR-CL signal than average. As these same sites  
214 consistently displayed significant PAR-CL signals over parallel replicates, this indicates a set of  
215 consistent RNA-capsid interactions in authentic FHV virions, which further indicates a structural  
216 tropism of FHV RNA in association with the topology of virus capsid shell.

217

### 218 **Probing FHV in virion RNA secondary structures with DMS-MaPseq.**

219 We sought to understand if there is any sequence motif among the PAR-CL sites. Significant ( $> 2\sigma$ )  
220 PAR-CL sites (28 sites from RNA 1 and 15 from RNA 2) and their flanking sequences were  
221 analyzed with Discriminative Regular Expression Motif Elicitation (DREME, (39)) for possible  
222 sequence motif identification (**Supplemental Figure S6**). However, no common motif was  
223 identified. This led us to hypothesize that the mechanism of RNA recognition by FHV capsids may  
224 be related to similar RNA structures rather than sequences. To reliably predict the RNA secondary  
225 structures of PAR-CL signals, we sought to determine the secondary structure of FHV RNA in  
226 authentic virions.

227 Dimethyl sulfate mutational profiling with sequencing (DMS-MaPseq) provides a reliable and high  
228 throughput method to probe RNA secondary structures *in vivo* (40-42). The resulting constraints  
229 provided improvement to thermodynamic map and free energy-based secondary structure  
230 prediction. We performed DMS-MaPseq using the TGIRT<sup>TM</sup>-III enzyme but in combination with  
231 ClickSeq to generate RNAseq libraries (“TGIRT-ClickSeq”) (**Figure 5a**), demonstrating that  
232 TGIRT<sup>TM</sup>-III enzyme is compatible with ClickSeq. DMS-MaPseq induces RNA modifications to  
233 unpaired adenines and cysteines (and guanine to a lesser level (43)) across the viral genome.  
234 Therefore, as expected, in comparison to untreated control virus (DMS-) DMS-treated FHV  
235 (DMS+) has a higher average mutation rate over genomic A/C positions (**Figure 5b**). Similarly,  
236 we plotted the frequency of mutation rates over A/C or G/U positions and only noticed a significant  
237 higher mutation rate frequency over A/C positions (**Figure 5c**). We analyzed A or C positions with  
238 at least 10k read coverage, which corresponds to nt. 14 -3043 on RNA 1 and nt. 11 - 1378 on  
239 RNA 2. Similar to PAR-CL data, DMS-MaPseq signal represents the mutation rate fold change  
240 between DMS-treated virus and untreated control virus, on all genomic A or C positions. Likewise,  
241 we removed potential false-positive signals by applying a background noise threshold, retaining  
242 only the genomic sites with mutation rate higher than this threshold. The resulting DMS-MaPseq  
243 profile of FHV (**Figure 5d and e**) showed clear signals up to 100-fold change over both RNA 1  
244 and 2. The un-refined DMS-MaPseq profiles with background noise, and mutation rate  
245 comparison between DMS-treated and untreated viruses are shown in **Supplemental Figure S7**.

246

247 **DMS-MaPseq resolved FHV RNA secondary structures reveals that PAR-CL sites favor**  
248 **double stranded structures and are highly clustered.**

249 We incorporated the DMS-MaPseq data into free energy based thermodynamic prediction, by  
250 introducing a series of “soft” constraints. Only the most significant ( $>2\sigma$ ) DMS-MaPseq sites (60  
251 sites in RNA 1 and 30 sites in RNA 2) were forced as unpaired constraints in RNAstructure Web  
252 Server(44) with “Fold” algorithm (44, 45). Regardless of their DMS-MaPseq signals, the remaining  
253 genome positions were left without any constraints, to allow maximum prediction flexibility. We  
254 thereby constructed a DMS-MaPseq-imposed secondary structure map of complete FHV RNA  
255 genome (snapshots in **Figure 6**, full-scaled maps of RNA 1 and RNA 2 were also provided in  
256 **Supplemental data 1 and 2**). Despite the low number of introduced constraints, we were able to  
257 greatly improve the thermodynamic mapping of FHV RNAs. With the 60 RNA 1 constraints, 37%  
258 (1145/3107) of nucleotides underwent refolding compared to the unconstrained model, yielding  
259 different paired/unpaired patterns. Similarly, with the 30 RNA 2 constraints, 20% (273/1400)  
260 nucleotides underwent refolding. The dot-bracket maps comparing the differences between  
261 unconstrained and constrained folding can be found in **Supplemental Figure S8**.

262 In combination with PAR-CL data, we noticed that the significant PAR-CL sites heavily favored  
263 double-stranded base-pairing. In RNA 1, among the 28 most significant ( $>2\sigma$ ) PAR-CL sites, 22  
264 are located in double-stranded regions, whereas 3 sites were 1 nt. adjacent to double-stranded  
265 stems. In the much shorter RNA 2, 8/15 of most significant PAR-CL sites are located in dsRNA  
266 stems, whereas 3 are 1 nt. adjacent. In **Table 1**, we illustrate the detailed structures of 16 PAR-  
267 CL sites (11 on RNA 1 and 5 on RNA 2) that presented with highest consistency and average  
268 PAR-CL signals (**Figure 4a-d**).

269 We also noticed that the distribution of PAR-CL signals was uneven and highly clustered.  
270 Numerous PAR-CL stems showed more than one PAR-CL sites with  $>1\sigma$  significance  
271 (**Supplemental data 1, 2**, some examples were listed in **Table 1**). We calculated the average  
272 shortest distance between adjacent PAR-CL sites. On RNA 1 (3107 nt.), among 721 uridine sites,  
273 102 showed  $>1\sigma$  significant PAR-CL signal. The average shortest distance between these PAR-  
274 CL sites is 7.4 nucleotides, which is substantially shorter than the average shortest distance of  
275 102 random uridines (30.47 nucleotides). On RNA 2 (1400 nt. genome with 351 uridines), the  
276 average shortest distance among 45 PAR-CL sites ( $>1\sigma$  significance) was 8.8 nt., which is also  
277 shorter than the average shortest distance of 45 random uridines (31.11 nt.).

278 Notably, by combining PAR-CL data and DMS-MaPseq-imposed RNA structure, we are able to  
279 characterize a stem loop site which is structurally near identical to a previously predicted stem  
280 loop (nts. 168-249) on RNA 2 (24) (**Supplemental Figure S9**). This stem loop site, as well as the  
281 flanking sequence (nt. 210-249) has been determined to be essential for RNA 2 encapsidation.  
282 We identified three PAR-CL signals within this region, consistent with role of this stem loop site in  
283 RNA 2 packaging.

284

285

286 **Structurally-disrupted PAR-CL sites impact FHV lifecycle and fitness.**



287 To determine whether the identified PAR-CL sites have a biological function, we selected 11 PAR-  
288 CL sites from RNA 1 and 5 PAR-CL sites from RNA 2 as our candidate sites (**Table 1**). Referring  
289 to the DMS-MaPseq-corrected FHV RNA structure maps (**Figure 6**), we introduced synonymous  
290 mutations to disrupt the double-stranded RNA regions of the PAR-CL sites (or the nearest stem  
291 of certain PAR-CL sites, i.e. U2515 on RNA1, U534, U903, U968, and U1155 on RNA2). The  
292 predicted structure of these PAR-CL sites, primers, and replaced nucleotides are listed in **Table**  
293 **1**. Plasmids containing these point mutations were transfected into S2 cells. Each transfection  
294 consisted either a mutated RNA1 and wild-type RNA2, or a mutated RNA2 and wild-type RNA1  
295 (**Figure 7a**). After transfection, induction and incubation, cell viability of each transfected mutant  
296 was determined with alamarBlue assay (**Figure 7b**). Almost all mutant virus transfections showed  
297 reduced cytopathic effect compared to transfection with wild-type FHV RNA. Notably RNA 1  
298 mutants U159, and U1233 resulted in little to no detrimental effect to S2 cells.

299 Total cellular RNA was extracted from transfected cells and in-column DNase digestion was  
300 conducted to remove remaining plasmids. From each transfection, 200ng of purified RNA was  
301 used as template for RT-PCR to detect FHV RNA (**Figure 7c**). We noticed that accumulation of  
302 FHV RNA 2 was unaffected by any PAR-CL mutants, while RNA 1 accumulation varied drastically  
303 among RNA 1 mutants. Notably, RNA 1 mutant U1233 yielded undetectable levels of RNA 1 and  
304 RNA 3, while RNA 2 production was less affected. RNA 1 mutant U159 also produced marginal  
305 amount of RNA 1 and RNA 3, and U227 produced substantially less RNA 1 than that of control  
306 or RNA 2 mutants. The replication deficiency of these three mutants agreed with our findings of  
307 their low virulence (**Figure 7b**). Interestingly, these three sites are found within or adjacent to  
308 previously described FHV RNA regulatory regions (46, 47) (**Figure 8**). The importance of these  
309 three sites in both RNA-capsid interaction shown here and RNA replication regulation indicates  
310 that the same motifs in the RNA genome are involved in multiple stages of the viral life-cycle,  
311 consistent with the notion that replication and RNA genome packaging are tightly coupled  
312 processes (20, 21).

313 To confirm capsid production, we separated cells and supernatant from the transfected cells.  
314 Western blots with anti-FHV were used to detect capsid proteins in both cellular components and  
315 supernatants (**Figure 7d**). In the cellular fraction, we readily detected both mature (alpha peptide)  
316 and autoproteolytically cleaved capsid protein (beta peptide) in all mutants. However, reduced  
317 capsid yields were found in U159 and U1233 mutants, possibly due to the observed RNA 1  
318 replication deficiency. In supernatant fractions, the U159 and U1233 mutants resulted in  
319 undetectable level of capsid protein, while U227 resulted in detectable but very marginal amount  
320 of capsid production. This confirmed that the mutations at these three PAR-CL sites have  
321 significant impact on virus production in S2 cells.

322 To expand mutant viruses, we further inoculated naïve S2 cells with equal amount of transfected  
323 p0 cell mix. From the inoculated P1 cell culture, we observed different degrees of cytopathic effect  
324 (CPE) under microscope (**Supplemental Figure S10**), which was correlated to earlier findings.  
325 P1 mutant viruses were nuclease treated, PEG precipitated, and purified with PES membrane  
326 protein concentrator to remove potentially unassembled capsid subunits. The presence of virus  
327 particles was confirmed with SDS-PAGE (**Figure 7e**), and virus yield was calculated by  
328 densitometry. Similar to before, we failed to detect virus production of U159 and U1233 mutants,  
329 while U227 mutant resulted a marginal virus production which can only be detected by western  
330 blot but not with SDS-PAGE. This result also agreed with our western blot analysis (**Figure 7d**).

331 We further tested P1 mutant virus relative virulence by infecting cells with mutants at MOI = 1  
332 **(Supplemental Figure S11)**. Most mutant viruses still resulted in varied but inferior virulence, in  
333 comparison to wild type virus.

## 334 **Discussion**

335 In this study, we demonstrated that PAR-CL can be used as a reliable and convenient method to  
336 screen for capsid-interacting sites on viral RNA genomes. PAR-CL data analysis features low  
337 background noise and thus, highly distinguishable signals. Therefore, PAR-CL signals are highly  
338 specific and representative of consistent crosslinking events between virus RNA and capsid. We  
339 showed that under the optimized experimental condition (4SU40h), FHV PAR-CL replicates  
340 revealed consistent crosslinking sites, indicating clear structural tropism of the asymmetrically  
341 packaged genomic RNA inside particles.

342 We also demonstrated that by combining a functional screening method PAR-CL and a structural  
343 probing method DMS-MaPseq, we can relate virus RNA-capsid interactions with RNA secondary  
344 structures, and *vice versa*. DMS-MaPseq was used to construct the whole genome RNA  
345 secondary structure maps of FHV, with which we observed that PAR-CL signals were highly  
346 clustered and favored double-stranded RNA stems. Synonymous mutations were designed to  
347 disrupt the double stranded structures in candidate PAR-CL sites. These structural mutants  
348 resulted in varied effects with most displaying reduced cytotoxic effect after transfection, while  
349 certain mutants were detrimental to RNA 1 replication. Together, we demonstrated that in  
350 combination with DMS-MaPseq, PAR-CL method can be used *de novo* to identify RNA regions  
351 that are important in virus packaging, virulence, and/or replication.

352

## 353 **PAR-CL methodology**

354 Photoactivatable nucleoside analogs were successfully utilized in the past to enhance  
355 crosslinking efficiency and hence, providing approaches to study RNA-RNA and RNA-protein  
356 interactions (reviewed in (48)). Thionucleobases such as 4-thiouracil (4SU) and 6-thioguanosine  
357 (6SG)) allows for highly effective crosslinking at 330-365 nm excitation spectrum (49), as well as  
358 advantages such as minimum nucleoside structure perturbation (48, 50), lower cytotoxicity (26,  
359 32, 33, 51), and less photochemistry and/or photodamage (48, 50). Importantly, the 4SU/6SG  
360 incorporated RNA can lead to specific base mismatches during reverse transcription (U-C, and  
361 G-T)(27-29), which enables high-throughput screening as indications of crosslinking. This is best  
362 illustrated with PAR-CLIP (PhotoActivatable-Ribonucleoside-enhanced CrossLinking and  
363 ImmunoPrecipitation) technology (26), which allows for pinpointing crosslinking sites at nucleotide  
364 resolution. PAR-CLIP has been successfully applied in the past to identify crosslinking sites of  
365 Argonaute 2, embryonic lethal abnormal vision (ELAV) protein and pumilio homologue 2 (PUM2),  
366 insulin-like growth factor proteins (26, 52).

367 As its primary purpose, PAR-CLIP was designed to screen entire transcriptome for RNA  
368 sequences binding to RBP-of-interest. Typically (26, 30, 31), PAR-CLIP was conducted by  
369 incubating cell cultures with 4SU, followed by 365 nm UVA irradiation, cell lysis, RNase T1  
370 digestion, immunoprecipitation of RBP-of-interest, second RNase T1 digestion, de-  
371 phosphorylation, radiolabeling, SDS-PAGE and electro-elution, proteinase K digestion, and RNA  
372 extraction. The recovered crosslinked RNA then is used as a template for cDNA library  
373 preparation and deep sequencing. A natural prerequisite is large amounts of starting materials  
374 (between 100-400 x 10<sup>6</sup> cells (30)).

375 The unique aspect of our simplified PAR-CL (PhotoActivatable-Ribonucleoside-enhanced  
376 CrossLinking) method is that we applied the similar PAR-CLIP principles to an RNA virus (FHV),

377 which can be easily separated from cellular components. Crosslinks within purified virus particles  
378 allow us to: (1) eliminate the need for immunoprecipitation to recover crosslinked RNA; (2) look  
379 for specific in virion interactions between viral RNA genomes and viral capsid proteins; (3) study  
380 a reductionist and highly controlled microenvironment. The greatly simplified PAR-CL  
381 methodology, in combination with ClickSeq library construction technology (36), granted the ability  
382 to conduct an experiment with as little as 2  $\mu\text{g}$  of purified FHV particles. A single T25 flask of S2  
383 cells can generate ample amount of pure 4SU-containing viruses to conduct multiple PAR-CL  
384 experiments.

385 In our PAR-CL method, the final pool of purified viral RNA can comprise large number of wild type  
386 uridines, or uncrosslinked 4SUs. As a consequence, the signal of any randomly generated, non-  
387 specific crosslinking event will be largely diluted into background level. Only the consistent  
388 crosslinking sites present due to homogeneity in RNA-capsid interactions within a viral population  
389 can readily provide distinguishable PAR-CL signals from background. Therefore, in contrast to  
390 the canonical PAR-CLIP approach where only cross-linked RNA fragments are sequenced, we  
391 are also able to identify regions of the viral genome where there is no reproducible PAR-CL signal,  
392 either due to a lack of RNA-capsid interactions or heterogeneous interactions. This is best  
393 illustrated in **Figure 3b**, where the background noise levels are largely unchanged, with or without  
394 crosslinking.

395 In both PAR-CLIP and PAR-CL, there are intrinsic limitations of 4SU-induced crosslinking. Firstly,  
396 crosslinking is only limited to U positions. Any potential interaction between protein and other  
397 nucleotides is undiscoverable. Next, 4SU crosslinking with protein is affected by reactivity of  
398 amino acid side chains (27, 29), with aromatic amino acids (phenylalanine, tyrosine, and  
399 tryptophan) being predominant targets but also lysine and cysteine (27). Consequentially, not all  
400 RNA-protein interactions can be depicted by PAR-CL or PAR-CLIP, and certain interactions may  
401 not result in crosslinking.

402

### 403 **FHV PAR-CL experiments and data analysis**

404 Several approaches were used to ensure reliable interpretation of PAR-CL signals on FHV: 1) to  
405 ensure reliable interpretation of mutation rate, we limited our analysis to genomic positions with  
406 at least 10k coverage. For this reason, our FHV PAR-CL experiments reliably covered U34 –  
407 U3034 on RNA 1, and U9 – U1337 on RNA 2. However, it is possible that we omitted potential  
408 capsid interaction sites out of our analyzed range. 2) We previously noticed that certain point  
409 mutations may be selected by virus and could be associated with defective interfering RNA  
410 generation (34). In this study, we also noticed substantially increased mutation rates on certain  
411 genomic positions (such as U1259 on RNA1, as illustrated in **Figure 2a-c**). Thus, to eliminate  
412 virus intrinsic mutational events, we avoided to use U-C mutation rate as a measurement. Instead,  
413 we decided to use fold change of U-C mutation rate, between crosslinked virus and uncrosslinked  
414 virus control, as our PAR-CL signals. 3) Because our PAR-CL signal corresponds to the fold  
415 change of U-C mutation rates of two datasets, a substantial PAR-CL signal can be a consequence  
416 of three scenarios: a high U-C rate in crosslinked virus, a much lower U-C rate in uncrosslinked  
417 control, or both. To minimize the possibility of false positives, we introduced a background  
418 threshold. Only the PAR-CL signals above this threshold were taken into our further consideration,  
419 as they represent mutation rates distinguishable from background fluctuation range (illustrated in  
420 **Supplemental data 3**). Together, we believe these three quality control measurements provided

421 stringent analysis to our PAR-CL data to reveal truly biologically relevant FHV RNA-capsid  
422 interaction sites.

423

#### 424 **DMS-MaPseq and FHV secondary structure mappings**

425 Several studies have proposed lowest free energy-based FHV local or whole genome secondary  
426 structure predictions, with the focus on viral RNA intracellular arrangement and replication  
427 regulations (24, 46, 47, 53). *In vivo* RNA chemical probing methods such as DMS and SHAPE  
428 allow for structure-specific chemical modifications to be screened by next generation sequencing  
429 techniques (40, 54, 55). Using DMS-MaPseq in authentic FHV virions, we are able to provide  
430 experimental validation of the RNA structures inside virus particles. With the same rationale of  
431 PAR-CL, we also applied stringent quality control measurements to ensure reliable interpretation  
432 of mutational profiles generated by DMS-MaPseq: A/C error rates were only analyzed over  
433 positions with at least 10k coverage (A14-A3043 on RNA 1, and C11-A1378 on RNA 2); fold  
434 change of A/C mutation rate was regarded as DMS-MaPseq signals instead of actual mutation  
435 rate; similar background noise threshold was also applied to prevent potential false positives.  
436 Canonically, DMS-MaPseq data is imposed upon thermodynamic prediction by enforcing  
437 unpaired constraints on any position with a signal above a given threshold (40). In this study, we  
438 adjusted this approach by only allowing the most significant (top 5%) DMS-MaPseq signals to be  
439 unpaired constraints. However, in this study, we constructed FHV secondary structural maps over  
440 RNA 1 and RNA 2 separately, omitting potential inter-RNA interactions.

441 Combining PAR-CL and DMS-MaPseq, we demonstrated that these two high-throughput  
442 mutational profile technologies can work synergistically to answer basic virology questions. We  
443 observed that the FHV RNA-capsid sites heavily favored double stranded RNA structures. This  
444 finding agreed with earlier predictions that the RNA duplexes scaffold the 2-fold axis of FHV  
445 capsid (12).

446

#### 447 **Flock house virus PAR-CL sites and biological indications**

448 It has been observed previously that the RNAs of FHV, as well as other Nodaviruses, form a  
449 highly ordered dodecahedral cage inside virus particles (10, 56). However, it was not clear  
450 whether the dodecahedral RNA cage had a fixed topology. From our PAR-CL data (**Figure 4a-d**),  
451 we can clearly identify highly consistent RNA-capsid interactions over certain genomic positions  
452 among multiple replicate experiments. This provides evidence that there is well-defined tropism  
453 between FHV RNA cage and capsid shell, at least at these sites identified here. Among the  
454 most consistent and distinguished PAR-CL sites (**Figure 4c, d**), we noticed that they exhibited a  
455 highly clustered pattern. The clustering effect is more pronounced, when taking RNA secondary  
456 structures into consideration (**Figure 6 and Supplemental data 1, 2**).

457 The multiple RNA-capsid interaction sites spanning the whole FHV genome suggest the  
458 possibility that FHV encapsidation may require multiple packaging signals to assemble the entire  
459 virus genome. Flock House virus genome packaging process may be similar to the two-staged  
460 packaging mechanism of MS2 bacteriophage (57). Numerous synergetic high-affinity RNA-capsid  
461 interaction sites are required to recruit capsid subunits. These widely-distributed interaction sites  
462 facilitate capsid-capsid interactions, which reciprocally mediate RNA folding and tertiary

463 compression of RNA genome. Subsequently, this can be followed by continuous recruitment of  
464 capsids on folded RNA to finalize encapsidation process.

465 Several PAR-CL sites also aligned with, or in close adjacent to, known RNA motifs (**Figure 8**).  
466 On RNA 1 (**Figure 8a**), we could not align any candidate PAR-CL signal to subgenomic RNA 3,  
467 which suggests the possibility that the exclusion of RNA 3 during packaging is due to lack of  
468 strong RNA-capsid interactions. Interestingly, two most significant PAR-CL sites on RNA 1, U159  
469 and U1233, aligned with previously discovered replication regulatory elements: a 5' cis element  
470 (nts. 68-205) that is essential for RNA 1 replication and mitochondria-targeting(46), and short  
471 distal subgenomic control cis-element (nts. 1229-1239) which mediates subgenomic RNA 3  
472 replication (47). Furthermore, U2515 and U2576 were located in the subgenomic promoter region  
473 (47, 58) which are also adjacent to a RNA 1 internal cis-acting replication element (intRE, nts  
474 2322-2501) (47). Similarly, on RNA 2 (**Figure 8b**), we noticed PAR-CL site U534 is adjacent to a  
475 RNA 2 cis-acting regulatory site (59), and U1155 which is within a site required for specific  
476 packaging of both RNAs (18). A previously predicted stem loop site (nts 168-249) on RNA 2  
477 serves as a RNA 2 packaging signal (24). This is also the only established FHV RNA packaging  
478 signal to date. Our DMS-MaPseq map did predict near identical stem loop structure as previous  
479 proposed and we noticed three significant ( $>1\sigma$ ) PAR-CL sites were clustered in this critical region  
480 (**Supplemental Figure S9**). Since these RNA-capsid interaction sites are correlated to RNA  
481 cellular replication/mitochondrial targeting sites, we suggest they might be multi-functional in virus  
482 life cycle, and there can be a strong synergy between protein A-mitochondria localization (9, 60,  
483 61), RNA replication, and virus assembly.

484 It was speculated previously that FHV packaging and replication are coordinated events. When  
485 FHV and brome mosaic virus (BMV) were co-expressed in plant cells, assembled virions only  
486 packed their own respective viral RNAs (20). Intracellular protein-protein interactions between  
487 FHV replicase (protein A) and capsid were detected (21). It has also been shown that FHV can  
488 ensure genome assembly specificity only when capsids were translated from replicating viral  
489 RNAs (23). It was hence suggested that FHV encapsidation may be coupled with the RNA  
490 replication. Our FHV PAR-CL experiments directly implicated only one aspect of FHV biology: the  
491 RNA sites interacting with capsid proteins. However, upon further analysis and mutational assays,  
492 a number of PAR-CL sites clearly indicated their significance in FHV replication and regulation:  
493 U159 and U227 mutants showed severe deficiency of RNA 1 replication and virion production,  
494 while U1233 entirely abolished viral replication. This provides further evidence that FHV  
495 replication and packaging are not sequentially separated events, but rather a synchronized, highly  
496 inter-dependent processes. Furthermore, these RNA-capsid interactions are not only important in  
497 post-replicative/translational RNA packaging, but may also be essential for multiple aspects of  
498 virus early stage activities in host cells.

## 499 **Materials and Methods**

### 500 **Cell culture and virus**

501 *D. melanogaster* (S2) cells were regularly maintained and passaged with Schneider's *Drosophila*  
502 Media (Gibco) containing 10% fetal bovine serum, 1 × Antibiotic-Antimycotic (Gibco), 1 × MEM  
503 non-essential amino acids solution, and 1 mM sodium pyruvate.

504 As described previously (34), wt Flock House virus (FHV) was generated by transfecting S2 cells  
505 with pMT plasmid vectors (Invitrogen) containing respective genomes (NC\_004146 for RNA 1,  
506 and NC\_004144 for RNA 2). Copper sulfate was used to induce the promoter 24 h post  
507 transfection, while viruses were allowed to accumulate until 3 days post induction to yield passage  
508 0 (p0) virus/cell mixture. The p0 transfected cells and viruses were then used to inoculate naïve  
509 S2 cells in a T75 flask for 3 days to yield passage 1 viruses, which were purified and used as FHV  
510 inoculum in this study, unless otherwise mentioned.

511 All virus transfections, infections, and passages with S2 cell culture were maintained in 27°C  
512 incubator, unless otherwise mentioned.

513 To purify FHV, 1% Triton X-100 was added to the cell culture containing p1 viruses. Cell culture  
514 underwent one freeze-thaw cycle, and cell debris was removed with 3000 × g centrifugation. FHV  
515 in the supernatant was crudely purified with 4% polyethylene glycol (PEG) 8000 and centrifuged  
516 (6000 × g) to remove debris (8). This was followed by DNase I and RNase A overnight digestion,  
517 to remove any co-precipitated cellular DNA or RNA. Unless otherwise mentioned, viruses were  
518 further purified with a 10-40% sucrose gradient, and ultracentrifuge at 40,000 RPM for 1.5 h.  
519 Viruses were then concentrated with 100K MWCO polyethersulfone (PES) membrane protein  
520 concentrator (Pierce) and washed three times with 10mM Tris pH 7.4.

521

### 522 **PAR-CL and ClickSeq**

523 S2 cells were maintained in T75 flask until 70% - 90% confluency. Cells were infected with purified  
524 Flock House virus (p1) at MOI = 1 (34, 62, 63). As an initial dose, 4-thiouridine (Sigma-Aldrich)  
525 was supplemented to the cell culture to 100µM as 1 × concentration with virus. An optional “boost”  
526 dose of 4-thiouridine can also be supplied 16 h post infection (**Figure 3a**). Cells and viruses were  
527 harvested at 16 or 40 h post infection (**Figure 3a**). Viruses were purified with methods described  
528 above.

529 The nuclease-treated and purified 4SU-containing viruses were placed uncovered over ice and  
530 irradiated with 0.15 J/cm<sup>2</sup> (26, 30) of 365nm UV light (3UV-38, UVP). After crosslink, viruses were  
531 digested with 8U of proteinase K (NEB) at 37 °C for 30 min. Crosslinked RNAs were extracted  
532 and purified with RNA Clean & Concentrator (Zymo Research) to yield RNA template for  
533 4SU+/UV+ sequencing library sample.

534 Unless otherwise mentioned, the same 4SU-containing virus without any UV irradiation were  
535 prepared in the same way to give RNA template for 4SU+/UV- control library.

536 Both the crosslinked and uncrosslinked viral RNA were used to construct the ClickSeq Illumina  
537 libraries per standard ClickSeq method, which is detailed previously (34-36). 250ng of RNA

538 template was used in reverse transcription reaction with 1:35 Azido-NTPs:dNTPs ratio and  
539 SuperScript III reverse transcriptase (Invitrogen).

540 Equal molar of each indexed library was pooled and run on a HiSeq 1500 platform (Illumina), with  
541 single read rapid run flowcell for 1x150 reads and 7 nucleotides for the index.

542

### 543 **DMS-MaPseq and ClickSeq**

544 Dimethyl sulfate (DMS) RNA methylation method was described previously(40, 42). In this study,  
545 nuclease-treated and purified FHV was supplemented with DMS to 5% final concentration. After  
546 5 min incubation at 30°C, reaction was quenched on ice for 5 min with 2 volumes of 10mM Tris  
547 pH 7.4 and 30% 2-Mercaptoethanol (BME). RNA extraction was conducted with Quick-RNA Viral  
548 Kits (Zymo Research) with additional BME in the extraction buffer. The DMS- control sample  
549 comprises the same virus stock with identical treatments as above, but without DMS supplement.

550 Methylated FHV RNA and respective controls were proceeded with similar ClickSeq library  
551 construction method. One exception is the use of a high-fidelity and processive thermostable  
552 group II reverse transcriptase enzyme (TGIRT-III, InGex) during reverse transcription. 100U of  
553 TGIRT-III was mixed with 250ng of RNA template, 0.5mM of AzNTPs/dNTPs mixture  
554 (AzNTPs:dNTPs = 1:35), and the following reaction conditions: 5 mM Dithiothreitol (Invitrogen),  
555 10 U RNaseOUT (Invitrogen), 50 mM Tri-HCl pH 8.0, 75 mM KCl, and 3 mM MgCl<sub>2</sub>. The reaction  
556 mix was incubated in room temperature for 10 min, followed by 57°C incubation for 1.5 hrs, and  
557 75°C termination for 15 min. The terminated reaction was then digested with RNase H to remove  
558 RNA template. The purified cDNA was proceeded with click reaction with Illumina adapters and  
559 final PCR amplification with indexes.

560 Library pooling and Illumina sequencing platform are the same as above.

561

### 562 **Bioinformatics and data analysis**

563 The Illumina sequencing data of both PAR-CL and DMS-MaPseq were subjected to the following  
564 bioinformatic pipelines: first the Illumina sequencing adapter sequence “AGATCGGAAGAGC”  
565 was trimmed with *cutadapt* (37) (command line parameters: -b AGATCGGAAGAGC -m 40); then,  
566 we used *FASTX toolkit* ([http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)) to remove the  
567 remaining random nucleotides from the Illumina adapter sequence and random base-pairing as  
568 a result of azide-alkyne cycloaddition from cDNA fragments (command line parameters:  
569 *fastx\_trimmer* -Q33 -f 7); a further quality filter was applied to remove any reads that contained  
570 more than 4% nucleotides with a PHRED score <20 (command line parameters:  
571 *fastq\_quality\_filter* -Q33 -q 20 -p 96). The remaining reads were aligned to FHV genomes  
572 (NC\_004146, and NC\_004144). Data generated from PAR-CL experiments were aligned for end-  
573 to-end matches with *Bowtie* (v1.0.1) (38) (command line parameters: -v 2 --best). Data generated  
574 from DMS-MaPseq experiment were aligned with *Bowtie2* (64) to allow longer and gapped  
575 alignments (command line parameters: --local). Using *SAMtools* (65), the aligned reads were  
576 binarily converted, merged, indexed, sorted, and mathematically noted.

577 For both PAR-CL and DMS-MaPseq, we excluded any nucleotide location with less than 10k  
578 coverage to ensure reliable mutation rate calculation. For PAR-CL, we calculated the mutation



579 frequencies of each of the four nucleotides, as well as the U-C mutation rate at each genomic U  
580 position. For DMS-MaPseq, similar analysis was conducted but we focused on the overall  
581 mutation rates of A and C genomic positions. Between test group and respective control  
582 (4SU+/UV+ and 4SU+/UV- for PAR-CL, DMS+ and DMS- for DMS-MaPseq), we compared the  
583 mutation rate at the same genomic position, to yield the fold change map, as presentations of  
584 PAR-CL or DMS-MaPseq signals.

585 A background filter was applied to both PAR-CL triplicates (**Figure 4a-d**) and DMS-MaPseq data  
586 (**Figure 5d, e**), to ensure reliable data analysis and avoid potential false positives. For PAR-CL,  
587 the background threshold is determined by bottom 95% of U-C mutation rate in the uncrosslinked  
588 control group (4SU+,UV-). In the correspondent crosslinked group (4SU+,UV+), we removed any  
589 datapoint with U-C mutation rate below this threshold, as it is indistinguishable from background  
590 fluctuation. An example of applying background threshold for PAR-CL data can be found in  
591 **Supplemental data 3**. For DMS-MaPseq, similar background threshold was determined as the  
592 bottom 95% of A/C mutation rate, in the DMS-untreated control group (DMS-). Only the datapoints  
593 passed the background threshold were used to compile the fold change maps of mutation rate  
594 changes.

595 The raw sequencing data for both PAR-CL and DMS-MaPseq experiments are available in the  
596 NCBI sequence read archive (SRA) with accession number: PRJNA554838.

597

## 598 **RNA secondary structure prediction**

599 RNA secondary structure prediction was conducted with RNAstructure (44) with 310.15 K  
600 temperature and maximum loop size = 30. “Fold” (44, 45) and “Partition” (66) were used to  
601 prediction the structure of RNA and calculate the base pairing probability, respectively. The most  
602 significant DMS-MaPseq signal sites were applied as unpaired constraints in structure prediction.  
603 No other constraints applied to the rest genomic sites, regardless of the DMS-MaPseq signals, to  
604 ensure the flexibility of algorithm. The predicted structure file was then re-organized and certain  
605 nucleotides were highlighted for graphical purposes with StructureEditor (v.1.0) which is also  
606 provided by RNAstructure suite.

607

608

## 609 **Mutated virus with disrupted PAR-CL sites**

610 Plasmids containing FHV genomes were used as PCR templates. Universal upstream primer  
611 (TGCATAATTCTCTTACTGTCATGCCATCCGTAAG) and downstream primer  
612 (TAAGAGAATTATGCAGTGCTGCCATAACCATG) were used in combination with mutation  
613 primers (**Table 1**) to generate overlapped PCR fragments (Phusion High-Fidelity DNA  
614 Polymerase, NEB), with disrupted RNA structure at each selected PAR-CL sites. These  
615 overlapped fragments were then cloned into competent cells with standard In-Fusion HD Cloning  
616 (TaKaRa) techniques. The plasmids containing mutated FHV RNA 1 or RNA 2 sequences were  
617 then sanger-sequenced and mutation sites were confirmed.

618 To generate mutated viruses, the plasmids containing PAR-CL site mutations were used to  
619 transfect S2 cells with above-stated methods. Each mutant transfection consisted of equal  
620 amount of one mutated RNA genome with disrupted PAR-CL site, and wt genome of the other  
621 RNA (**Figure 7a**). These p0 mutant viruses were allowed to propagate in cell culture until 3 days

622 post induction. Similar to before, p1 mutant viruses were generated by inoculating naïve S2 cells  
623 with p0 cell culture/virus mix.

624

### 625 **Relative virulence of mutant viruses**

626 The virulence of p0 PAR-CL mutant viruses was measured via transfecting S2 cells with plasmids  
627 containing mutant viral genomes. For this experiment, transfection was conducted in black 96-  
628 well plate, with each well seeded with 25k S2 cells. The transfection reagents and methods were  
629 similar to above, except for scaling down to adapt for 96 well plate. For each mutant virus, 100ng  
630 plasmids of each mutant genome and the other wild type genome were used. alamarBlue was  
631 incubated with cell culture for 4h, before detecting fluorescence with EnSpire plate reader  
632 (PerkinElmer) at 560 nm excitation and 590 nm emission. The relative fluorescence then  
633 normalized reverse-ratiometrically with mock transfection = 0% and FHV wt RNA transfection =  
634 100%.

635 The relative virulence of p1 PAR-CL mutant viruses was measured via infecting S2 cells with  
636 purified p1 mutant viruses. The p1 mutant viruses were purified through sucrose cushion (30%  
637 sucrose 10mM Tris pH 7.4, ultracentrifuge 80k rpm for 1.5hrs), PEG precipitation (4% v/v  
638 PEG8000), DNase I and RNase A treatment, buffer exchanged and concentrated with PES  
639 membrane protein concentrator (100K MWCO). The concentration of purified p1 viruses was  
640 determined with SDS-PAGE and densitometry. 25k S2 cells were seeded in black 96-well plate  
641 and 0.12 ng (approximately equivalent to MOI = 1 (34, 62, 63)) of serial diluted p1 viruses was  
642 used to infect each well. Standard alamarBlue assay was conducted as before, at 24 h post  
643 infection, to measure cell viability. The relative fluorescence then normalized reverse-  
644 ratiometrically with mock infection = 0% and purified wt FHV infection = 100%.

645

### 646 **RT-PCR**

647 Transfected p0 PAR-CL mutants were sampled for RT-PCR to detect RNA production. Total RNA  
648 was extracted from transfected cells and media with Direct-zol RNA kit (Zymo Research), and  
649 DNase I in-column digestion was conducted to remove plasmids. For each mutants and controls,  
650 200 ng of total RNA was used as template for RT-PCR. RT reaction was conducted with  
651 SuperScript III reverse transcriptase (Invitrogen), per manufacture's protocol. PCR was  
652 conducted with OneTaq® (NEB), per manufacture's protocol. The entire RT-PCR reaction was  
653 loaded on agarose gel for electrophoresis.

654

### 655 **SDS-PAGE and western blot**

656 After collecting p0 transfections, the cell/virus/supernatant mix was centrifuged at 1000 × g for 10  
657 min. Supernatant fraction was removed and collected thereafter. The cell pellet was washed once  
658 with 1 × PBS and centrifuged as before. The washed cellular fraction was then resuspended in 1  
659 × PBS and 1 × cComplete (Roche). 150 µL of supernatant (of each sample) was supplemented  
660 with 1 × cComplete and then reduced with vacuum centrifuge prior to SDS-PAGE.

661 All SDS-PAGE assays were conducted with Bolt 4-12% Bis-Tris Plus Gels (Invitrogen).  
662 Membrane transfer was conducted with iBlot 2 Dry Blotting System (Invitrogen) with standard

663 protocol. Western blot was conducted with iBind Western Device (Invitrogen) with standard  
664 protocol. Rabbit Anti-FHV polyclonal antibody was given as a gift from Dr. Vijay Reddy from  
665 Scripps Research, which was labelled with Alexa Fluor 488 goat anti-rabbit IgG (Invitrogen). Prior  
666 to membrane transfer, part of SDS-PAGE gel was cut and stained with Coomassie brilliant blue  
667 R-250 to highlight  $\alpha$ -tubulin (55 kDa) as a loading control.

668

669

## 670 **Acknowledgements**

671 A.R. is supported by start-up funds from the University of Texas Medical Branch. We thank Dr.  
672 Vijay Reddy from Scripps Research for providing samples of the anti-FHV capsid antibody.

673

674

## 675 **Reference**

- 676 1. Aliyari R, Wu Q, Li HW, Wang XH, Li F, Green LD, Han CS, Li WX, Ding SW. 2008. Mechanism of  
677 induction and suppression of antiviral immunity directed by virus-derived small RNAs in  
678 *Drosophila*. *Cell Host Microbe* 4:387-97.
- 679 2. Li H, Li WX, Ding SW. 2002. Induction and suppression of RNA silencing by an animal virus.  
680 *Science* 296:1319-21.
- 681 3. Seo JK, Kwon SJ, Rao AL. 2012. Molecular dissection of Flock house virus protein B2 reveals that  
682 electrostatic interactions between N-terminal domains of B2 monomers are critical for  
683 dimerization. *Virology* 432:296-305.
- 684 4. Venter PA, Schneemann A. 2008. Recent insights into the biology and biomedical applications of  
685 Flock House virus. *Cell Mol Life Sci* 65:2675-87.
- 686 5. Odegard A, Banerjee M, Johnson JE. 2010. Flock house virus: a model system for understanding  
687 non-enveloped virus entry and membrane penetration. *Curr Top Microbiol Immunol* 343:1-22.
- 688 6. Manayani DJ, Thomas D, Dryden KA, Reddy V, Siladi ME, Marlett JM, Rainey GJ, Pique ME,  
689 Scobie HM, Yeager M, Young JA, Manchester M, Schneemann A. 2007. A viral nanoparticle with  
690 dual function as an anthrax antitoxin and vaccine. *PLoS Pathog* 3:1422-31.
- 691 7. Maharaj PD, Mallajosyula JK, Lee G, Thi P, Zhou Y, Kearney CM, McCormick AA. 2014.  
692 Nanoparticle encapsidation of flock house virus by auto assembly of tobacco mosaic virus coat  
693 protein. *Int J Mol Sci* 15:18540-56.
- 694 8. Zhou Y, McCormick AA, Kearney CM. 2017. Plant Expression of Trans-Encapsidated Viral  
695 Nanoparticle Vaccines with Animal RNA Replicons. *Methods Mol Biol* 1499:77-86.
- 696 9. Zhou Y, Kearney CM. 2017. Chimeric Flock House virus protein A with endoplasmic reticulum-  
697 targeting domain enhances viral replication and virus-like particle trans-encapsidation in plants.  
698 *Virology* 507:151-160.
- 699 10. Tang L, Johnson KN, Ball LA, Lin T, Yeager M, Johnson JE. 2001. The structure of pariacoto virus  
700 reveals a dodecahedral cage of duplex RNA. *Nat Struct Biol* 8:77-83.
- 701 11. Johnson KN, Tang L, Johnson JE, Ball LA. 2004. Heterologous RNA encapsidated in Pariacoto  
702 virus-like particles forms a dodecahedral cage similar to genomic RNA in wild-type virions. *J Virol*  
703 78:11371-8.

- 704 12. Fisher AJ, Johnson JE. 1993. Ordered duplex RNA controls capsid architecture in an icosahedral  
705 animal virus. *Nature* 361:176-9.
- 706 13. Routh A, Domitrovic T, Johnson JE. 2012. Host RNAs, including transposons, are encapsidated by  
707 a eukaryotic single-stranded RNA virus. *Proc Natl Acad Sci U S A* 109:1907-12.
- 708 14. Routh A, Domitrovic T, Johnson JE. 2012. Packaging host RNAs in small RNA viruses: An  
709 inevitable consequence of an error-prone polymerase? *Cell Cycle* 11.
- 710 15. van de Waterbeemd M, Fort KL, Boll D, Reinhardt-Szyba M, Routh A, Makarov A, Heck AJ. 2017.  
711 High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. *Nat Methods*  
712 14:283-286.
- 713 16. Scotti PD, Dearing S, Mossop DW. 1983. Flock house virus: A Nodavirus isolated from *Costelytra*  
714 *zealandica* (White) (Coleoptera: Scarabaeida). *Archives of Virology* 75:181-189.
- 715 17. Marshall D, Schneemann A. 2001. Specific packaging of nodaviral RNA2 requires the N-terminus  
716 of the capsid protein. *Virology* 285:165-75.
- 717 18. Schneemann A, Marshall D. 1998. Specific encapsidation of nodavirus RNAs is mediated through  
718 the C terminus of capsid precursor protein alpha. *Journal of Virology* 72:8738-8746.
- 719 19. Dong XF, Natarajan P, Tihova M, Johnson JE, Schneemann A. 1998. Particle Polymorphism  
720 Caused by Deletion of a Peptide Molecular Switch in a Quasiequivalent Icosahedral Virus.  
721 *Journal of Virology* 72:6024.
- 722 20. Annamalai P, Rofail F, Demason DA, Rao AL. 2008. Replication-coupled packaging mechanism in  
723 positive-strand RNA viruses: synchronized coexpression of functional multigenome RNA  
724 components of an animal and a plant virus in *Nicotiana benthamiana* cells by agroinfiltration. *J*  
725 *Virol* 82:1484-95.
- 726 21. Seo JK, Kwon SJ, Rao AL. 2012. A physical interaction between viral replicase and capsid protein  
727 is required for genome-packaging specificity in an RNA virus. *J Virol* 86:6210-21.
- 728 22. Venter PA, Krishna NK, Schneemann A. 2005. Capsid protein synthesis from replicating RNA  
729 directs specific packaging of the genome of a multipartite, positive-strand RNA virus. *J Virol*  
730 79:6239-48.
- 731 23. Venter PA, Schneemann A. 2007. Assembly of two independent populations of flock house virus  
732 particles with distinct RNA packaging characteristics in the same cell. *J Virol* 81:613-9.
- 733 24. Zhong WD, Dasgupta R, Rueckert R. 1992. Evidence That the Packaging Signal for Nodaviral Rna2  
734 Is a Bulged Stem Loop. *Proceedings of the National Academy of Sciences of the United States of*  
735 *America* 89:11146-11150.
- 736 25. König J, Zarnack K, Luscombe NM, Ule J. 2012. Protein–RNA interactions: new genomic  
737 technologies and perspectives. *Nat Rev Genet* 13:77-83.
- 738 26. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M,  
739 Jr., Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T. 2010.  
740 Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-  
741 CLIP. *Cell* 141:129-41.
- 742 27. Meisenheimer KM, Meisenheimer PL, Koch TH. 2000. Nucleoprotein photo-cross-linking using  
743 halopyrimidine-substituted RNAs. *METHODS IN ENZYMOLOG* 318:88-104.
- 744 28. Testa SM, Disney MD, Turner DH, Kierzek R. 1999. Thermodynamics of RNA–RNA Duplexes with  
745 2- or 4-Thiouridines: Implications for Antisense Design and Targeting a Group I Intron†.  
746 *Biochemistry* 38:16655-16662.
- 747 29. Ascano M, Hafner M, Cekan P, Gerstberger S, Tuschl T. 2012. Identification of RNA-protein  
748 interaction networks using PAR-CLIP. *Wiley Interdiscip Rev RNA* 3:159-77.
- 749 30. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M,  
750 Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T. 2010. PAR-

- 751 CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J Vis*  
752 *Exp* doi:10.3791/2034.
- 753 31. Spitzer J, Hafner M, Landthaler M, Ascano M, Farazi T, Wardle G, Nusbaum J, Khorshid M, Burger  
754 L, Zavolan M, Tuschl T. 2014. PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking  
755 and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of  
756 binding sites of RNA-binding proteins. *Methods Enzymol* 539:113-61.
- 757 32. Favre A, Moreno G, Blondel MO, Kliber J, Vinzens F, Salet C. 1986. 4-thiouridine photosensitized  
758 RNA-protein crosslinking in mammalian cells. *Biochemical and Biophysical Research*  
759 *Communications* 141:847-854.
- 760 33. Favre A, Saintome C, Fourrey JL, Clivio P, Laugaa P. 1998. Thionucleobases as intrinsic  
761 photoaffinity probes of nucleic acid structure and nucleic acid protein interactions. *Journal of*  
762 *Photochemistry and Photobiology B-Biology* 42:109-124.
- 763 34. Jaworski E, Routh A. 2017. Parallel ClickSeq and Nanopore sequencing elucidates the rapid  
764 evolution of defective-interfering RNAs in Flock House virus. *PLoS Pathog* 13:e1006365.
- 765 35. Jaworski E, Routh A. 2018. ClickSeq: Replacing Fragmentation and Enzymatic Ligation with Click-  
766 Chemistry to Prevent Sequence Chimeras, p 71-85. *In* Head SR, Ordoukhanian P, Salomon DR  
767 (ed), *Next Generation Sequencing: Methods and Protocols* doi:10.1007/978-1-4939-7514-3\_6.  
768 Springer New York, New York, NY.
- 769 36. Routh A, Head SR, Ordoukhanian P, Johnson JE. 2015. ClickSeq: Fragmentation-Free Next-  
770 Generation Sequencing via Click Ligation of Adaptors to Stochastically Terminated 3'-Azido  
771 cDNAs. *Journal of molecular biology* 427:2610-2616.
- 772 37. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.  
773 *EMBnetjournal*; Vol 17, No 1: Next Generation Sequencing Data AnalysisDO -  
774 1014806/ej171200.
- 775 38. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of  
776 short DNA sequences to the human genome. *Genome biology* 10:R25-R25.
- 777 39. Bailey TL. 2011. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics*  
778 27:1653-9.
- 779 40. Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2017. DMS-MaPseq for  
780 genome-wide or targeted RNA structure probing in vivo. *Nat Methods* 14:75-82.
- 781 41. Zubradt M, Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2016. Target-  
782 specific DMS-MaPseq for in vivo RNA structure determination. *Protocol Exchange*  
783 doi:10.1038/protex.2016.069.
- 784 42. Zubradt M, Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2016.  
785 Genome-wide DMS-MaPseq for in vivo RNA structure determination. *Protocol Exchange*  
786 doi:10.1038/protex.2016.068.
- 787 43. Tijerina P, Mohr S, Russell R. 2007. DMS footprinting of structured RNAs and RNA-protein  
788 complexes. *Nat Protoc* 2:2608-23.
- 789 44. Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction  
790 and analysis. *BMC Bioinformatics* 11:129.
- 791 45. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating  
792 chemical modification constraints into a dynamic programming algorithm for prediction of RNA  
793 secondary structure. *Proc Natl Acad Sci U S A* 101:7287-92.
- 794 46. Van Wynsberghe PM, Ahlquist P. 2009. 5' cis elements direct nodavirus RNA1 recruitment to  
795 mitochondrial sites of replication complex formation. *J Virol* 83:2976-88.
- 796 47. Lindenbach BD, Sgro JY, Ahlquist P. 2002. Long-distance base pairing in flock house virus RNA1  
797 regulates subgenomic RNA3 synthesis and RNA2 replication. *J Virol* 76:3905-19.

- 798 48. Meisenheimer KM, Koch TH. 1997. Photocross-linking of nucleic acids to associated proteins.  
799 Crit Rev Biochem Mol Biol 32:101-40.
- 800 49. Tanner NK, Hanna MM, Abelson J. 1988. Binding interactions between yeast tRNA ligase and a  
801 precursor transfer ribonucleic acid containing two photoreactive uridine analogs. Biochemistry  
802 27:8852-8861.
- 803 50. Favre A. 1990. 4-thiouridine as an intrinsic photoaffinity probe of nucleic acid structure and  
804 interactions. Bioorganic photochemistry 1:379-425.
- 805 51. Melvin WT, Milne HB, Slater AA, Allen HJ, Keir HM. 1978. Incorporation of 6-Thioguanosine and  
806 4-Thiouridine into Rna - Application to Isolation of Newly Synthesized Rna by Affinity  
807 Chromatography. European Journal of Biochemistry 92:373-379.
- 808 52. Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M. 2011. A quantitative analysis  
809 of CLIP methods for identifying binding sites of RNA-binding proteins. Nat Methods 8:559-64.
- 810 53. Li Y, Ball LA. 1993. Nonhomologous RNA recombination during negative-strand synthesis of flock  
811 house virus RNA. J Virol 67:3854-60.
- 812 54. Smola MJ, Rice GM, Busan S, Siegfried NA, Weeks KM. 2015. Selective 2'-hydroxyl acylation  
813 analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and  
814 accurate RNA structure analysis. Nat Protoc 10:1643-69.
- 815 55. Mortimer SA, Weeks KM. 2007. A fast-acting reagent for accurate analysis of RNA secondary and  
816 tertiary structure by SHAPE chemistry. J Am Chem Soc 129:4144-5.
- 817 56. Tihova M, Dryden KA, Le TvL, Harvey SC, Johnson JE, Yeager M, Schneemann A. 2004. Nodavirus  
818 Coat Protein Imposes Dodecahedral RNA Structure Independent of Nucleotide Sequence and  
819 Length. Journal of Virology 78:2897-2905.
- 820 57. Borodavka A, Tuma R, Stockley PG. 2013. A two-stage mechanism of viral RNA compaction  
821 revealed by single molecule fluorescence. RNA Biol 10:481-9.
- 822 58. Zhou Y, Maharaj PD, Mallajosyula JK, McCormick AA, Kearney CM. 2015. In planta Production of  
823 Flock House Virus Transencapsidated RNA and Its Potential Use as a Vaccine. Mol Biotechnol  
824 57:325-36.
- 825 59. Ball LA, Li Y. 1993. cis-acting requirements for the replication of flock house virus RNA 2. J Virol  
826 67:3544-51.
- 827 60. Miller DJ, Ahlquist P. 2002. Flock House Virus RNA Polymerase Is a Transmembrane Protein with  
828 Amino-Terminal Sequences Sufficient for Mitochondrial Localization and Membrane Insertion. J  
829 Virol 76:9856-9867.
- 830 61. Miller DJ, Schwartz MD, Dye BT, Ahlquist P. 2003. Engineered Retargeting of Viral RNA  
831 Replication Complexes to an Alternative Intracellular Membrane. J Virol 77:12193-12202.
- 832 62. Dasgupta R, Free HM, Zietlow SL, Paskewitz SM, Aksoy S, Shi L, Fuchs J, Hu C, Christensen BM.  
833 2007. Replication of flock house virus in three genera of medically important insects. J Med  
834 Entomol 44:102-10.
- 835 63. Selling Bh Fau - Rueckert RR, Rueckert RR. Plaque assay for black beetle virus.
- 836 64. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-  
837 9.
- 838 65. Li H, Handsaker B Fau - Wysoker A, Wysoker A Fau - Fennell T, Fennell T Fau - Ruan J, Ruan J Fau  
839 - Homer N, Homer N Fau - Marth G, Marth G Fau - Abecasis G, Abecasis G Fau - Durbin R, Durbin  
840 R. The Sequence Alignment/Map format and SAMtools.
- 841 66. Mathews DH. 2004. Using an RNA secondary structure partition function to determine  
842 confidence in base pairs predicted by free energy minimization. RNA (New York, NY) 10:1178-  
843 1190.
- 844 67. Albariño CG, Eckerle LD, Ball LA. 2003. The cis-acting replication signal at the 3' end of Flock  
845 House virus RNA2 is RNA3-dependent. Virology 311:181-191.



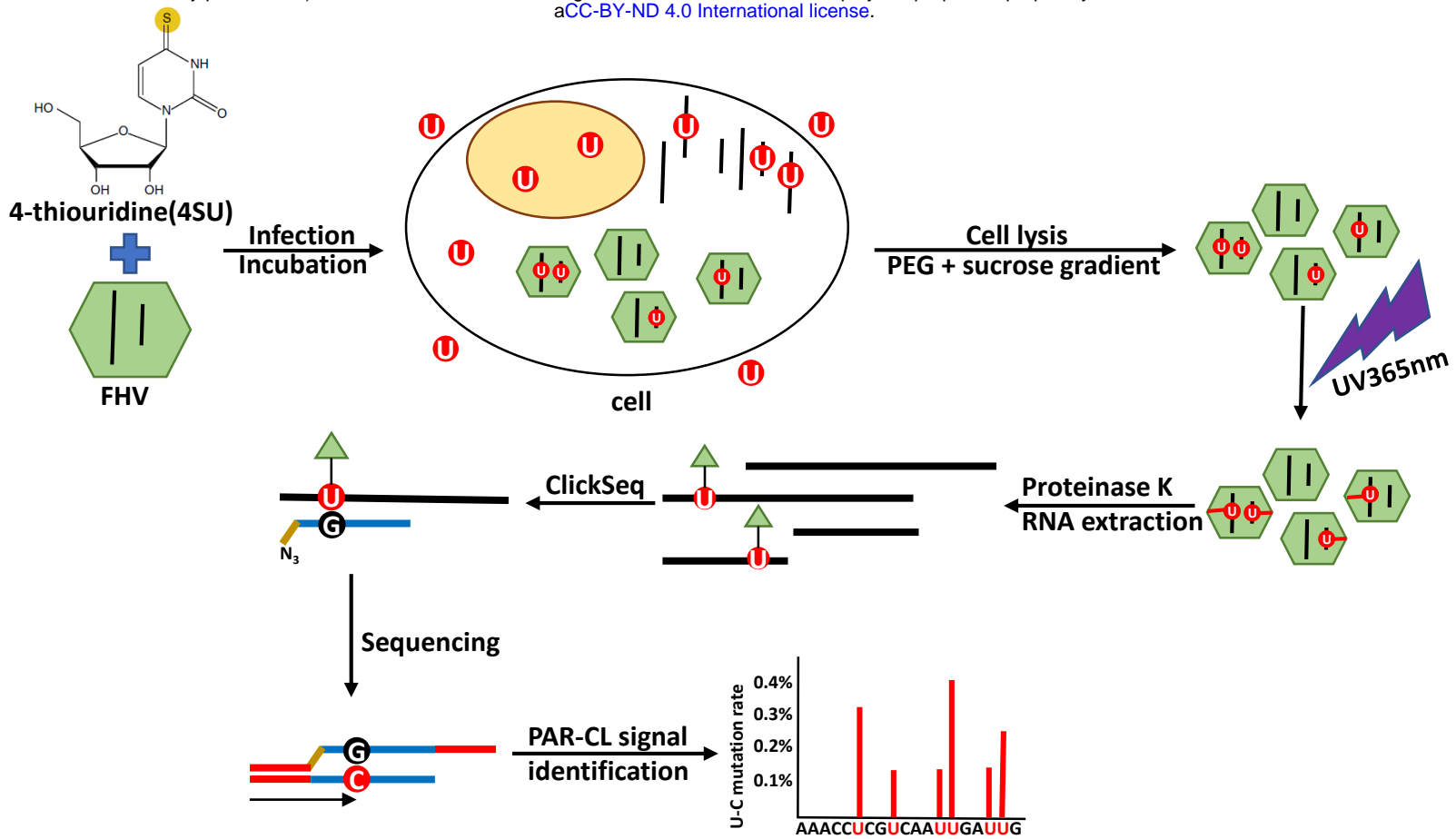


Figure 1



## Figure 1. PAR-CLIP methodology

4-thiouridine (4SU) were supplemented to S2 cells, during Flock House virus (FHV) infection. After incubation, purified viruses are irradiated with 365nm UV to induce crosslink. After proteinase K digestion, crosslinked RNAs are purified and subjected to ClickSeq with Azido-NTPs. In viral genome, the crosslinked sites are characterized with elevated U to C mutation rates.

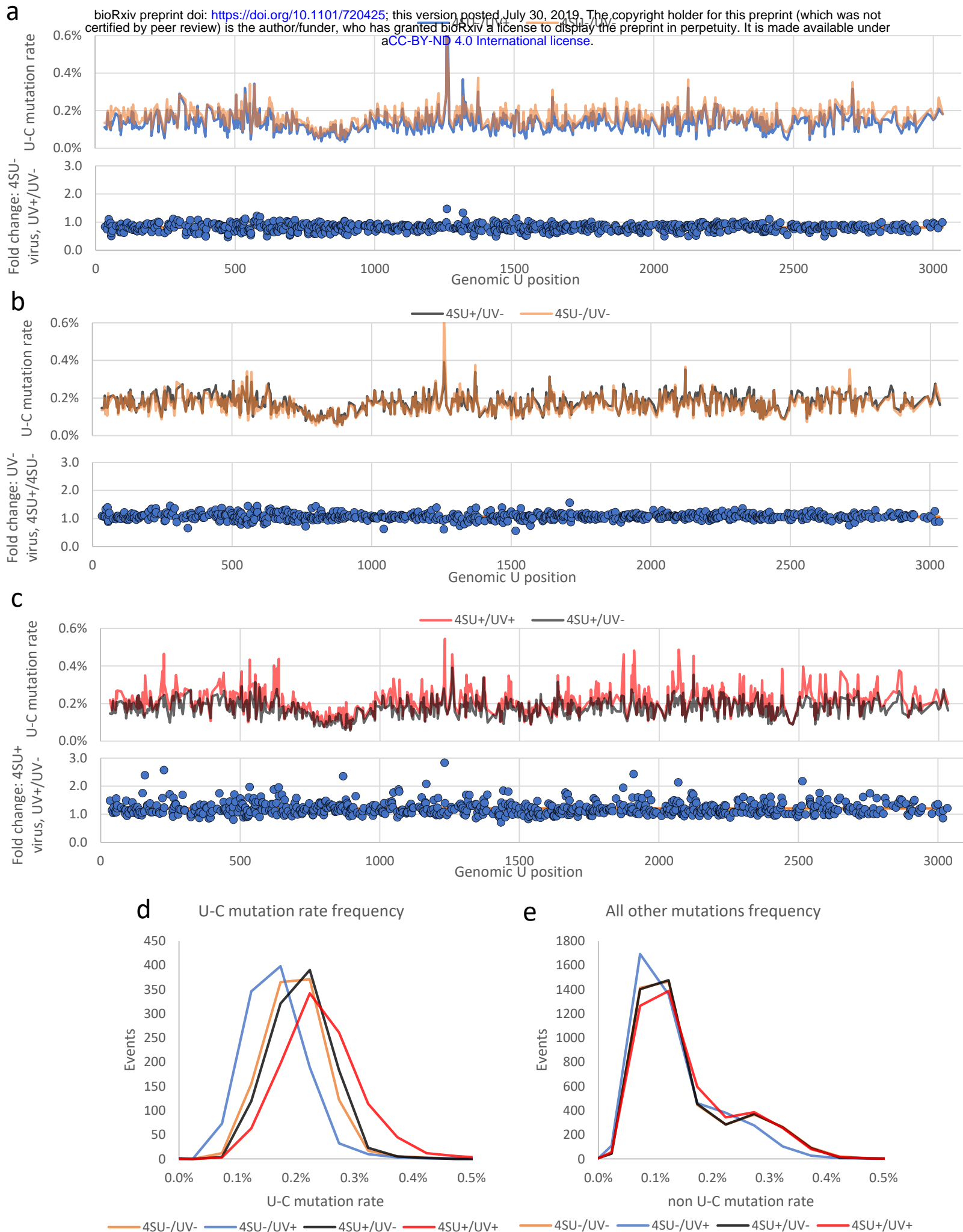


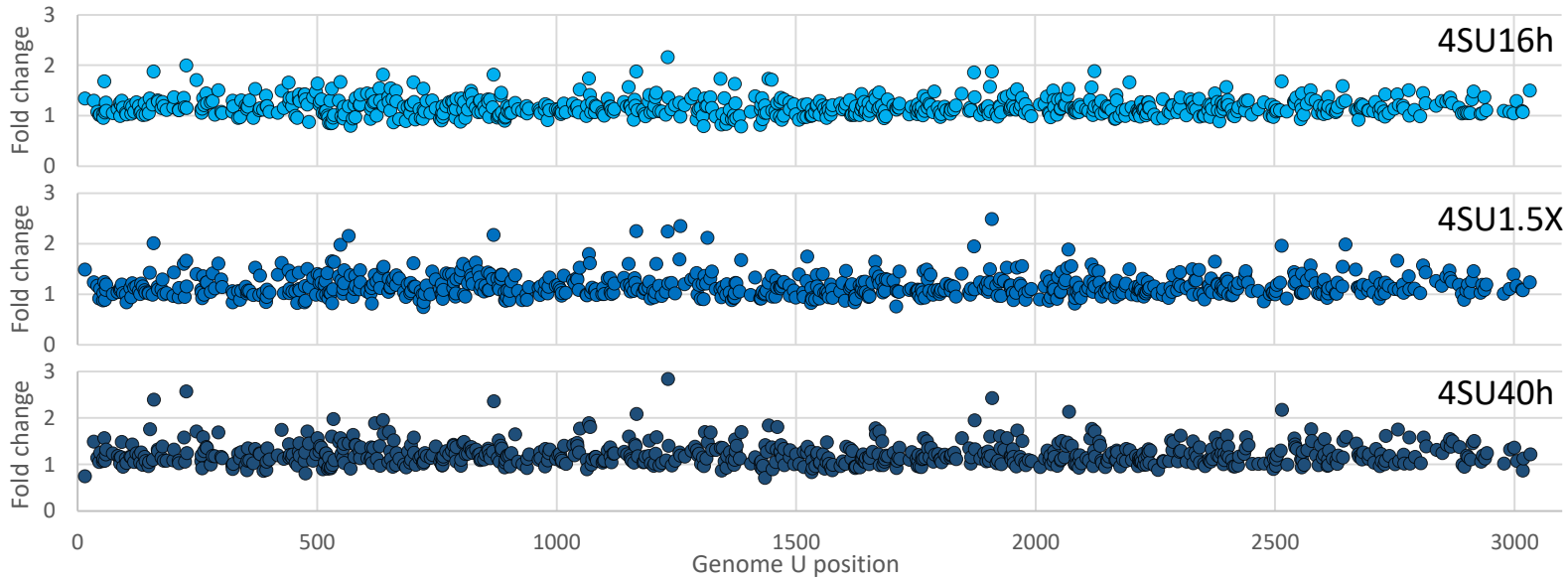
Figure 2

**Figure 2 U-C mutation rate elevation is specific to 4SU crosslink**

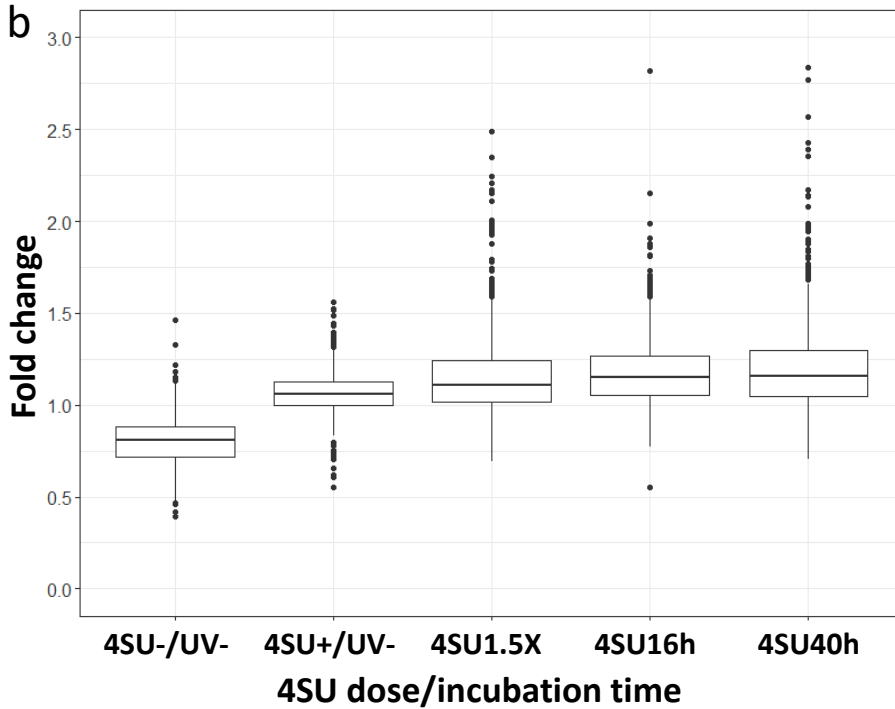
Using FHV RNA 1 as an example, several control experiments were conducted to ensure specificity of PAR-CL signals. **(a)** effects of UV exposure to wt FHV (4SU-) were compared. We did not observe significant U-C mutation rate elevation. **(b)** effects of 4SU incorporation were compared to FHV without UV exposure (UV-). We did not observe significant U-C mutation rate elevation. **(c)** only when 4SU-containing FHV was irradiated with UV, we observed a significant increase in U-C mutation rates. **(a-c)** orange line represents the average fold change in each experiment. **(d)** the distribution of U-C mutation frequency: only when both 4SU and UV both presented, did we notice a shift towards higher U-C mutation rates. **(e)** crosslink did not impact the rates of other mutations (A,G,C-mutations, or U-A/U-G mutations).

**a**

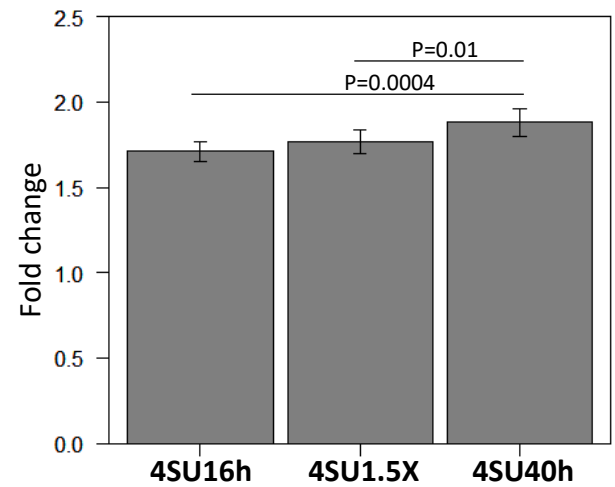
	Priming 4SU cct.	Boosting 4SU cct.	Final cct.	Incubation time
<b>4SU16h</b>	100uM	-	100uM	16hr
<b>4SU1.5X</b>	150uM	-	150uM	16hr
<b>4SU40h</b>	100uM	100uM (@16hpi)	200uM	40hr



**b**



**c**



**b**

**Figure 3**

**Figure 3. PAR-CL signal intensities correlated with 4SU dose and incubation time.** Using FHV RNA 1 as an example, **(a)** three experimental conditions were tested for impact to PAR-CL signals. We observed that the intensities of PAR-CL signals (i.e. fold change of U-C mutation rate) were related to the concentration of 4SU in cell culture and the time of incubation. **(b)** with or without crosslink, the average PAR-CL signals among all 4SU-containing FHVs is similar. However, the outliers of crosslinking groups showed a significant higher PAR-CL signals than that of control (4SU+,UV-), and the magnitude of outliers correlated with 4SU concentration and incubation time. **(c)** We sampled top 5% PAR-CL signals from three experiment groups and determined that optimal PAR-CL signal was achieved under 4SU40h condition, which was applied to all further PAR-CL experiments.

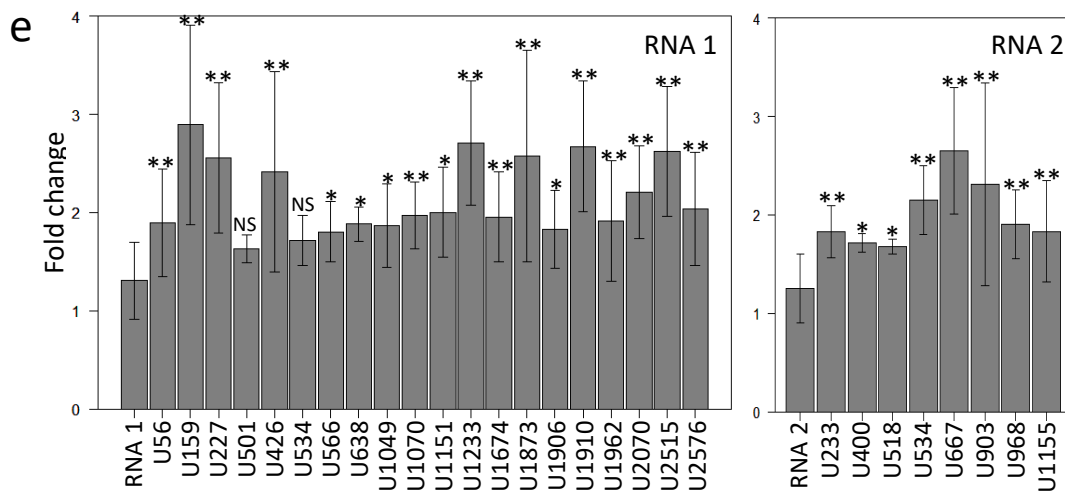
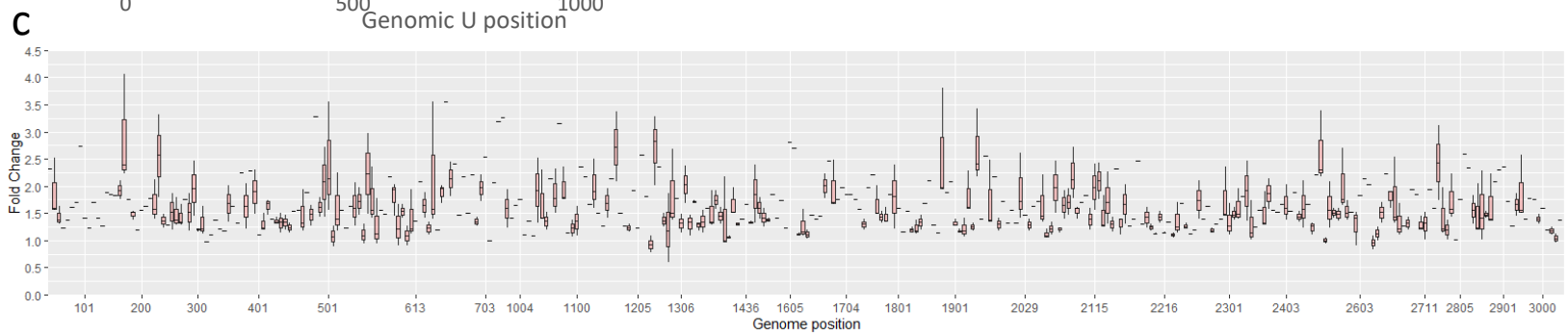
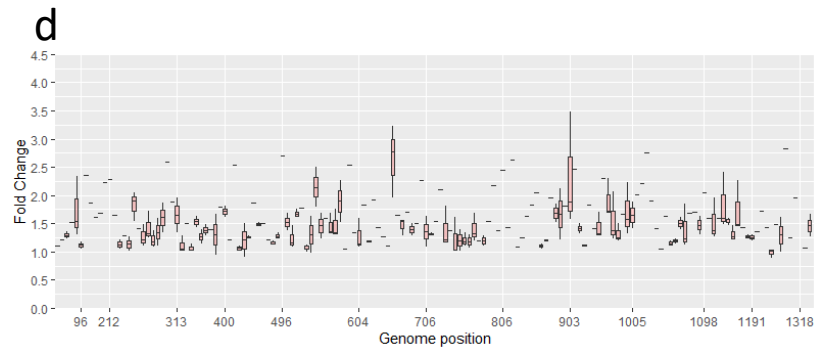
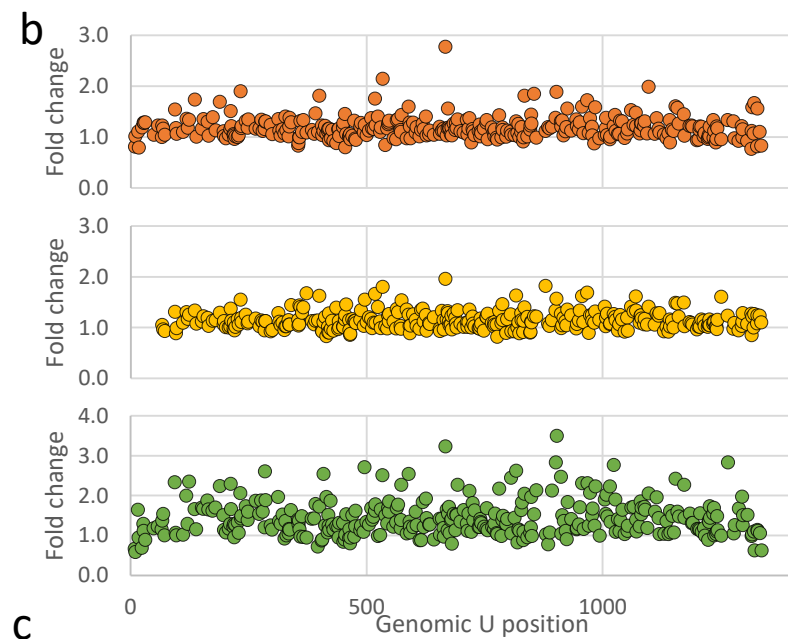
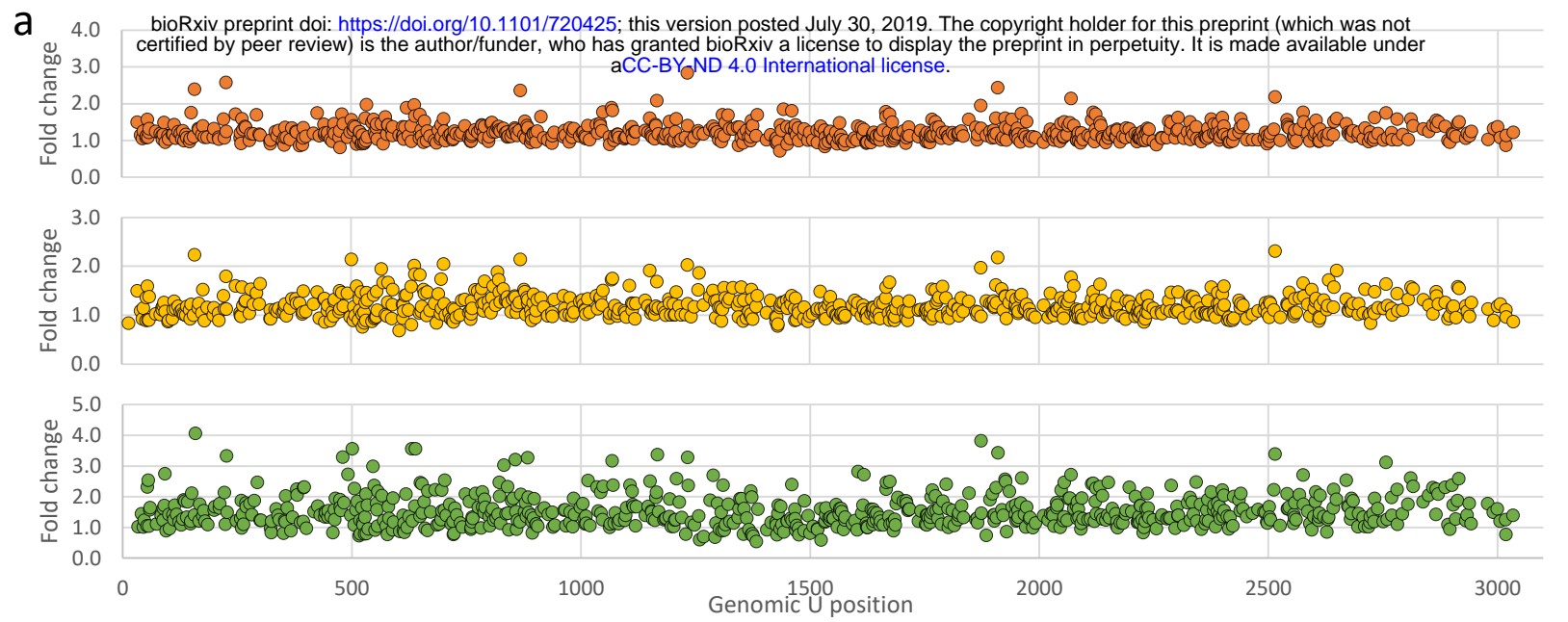


Figure 4

## Figure 4. Consistent PAR-CL sites revealed clear RNA tropism. (a, b) PAR-CL signals of FHV RNA 1 and 2 respectively, in triplicates. (c, d)

Triplicated PAR-CL signals of FHV RNA 1 and 2 were box-plotted. We removed any PAR-CL signal failed to pass the background threshold. A number of sites on both RNA 1 and RNA 2 showed consistently significant PAR-CL signals, indicating reliable crosslinking sites between RNA and protein. These consistent PAR-CL sites also suggest a strong tropism of FHV RNA cage inside virion. X-axis is not continuous. (e) Among these consistent PAR-CL sites, most of them showed significantly higher PAR-CL signals than the average. \*P<0.05; \*\*P<0.01; NS=not significant.

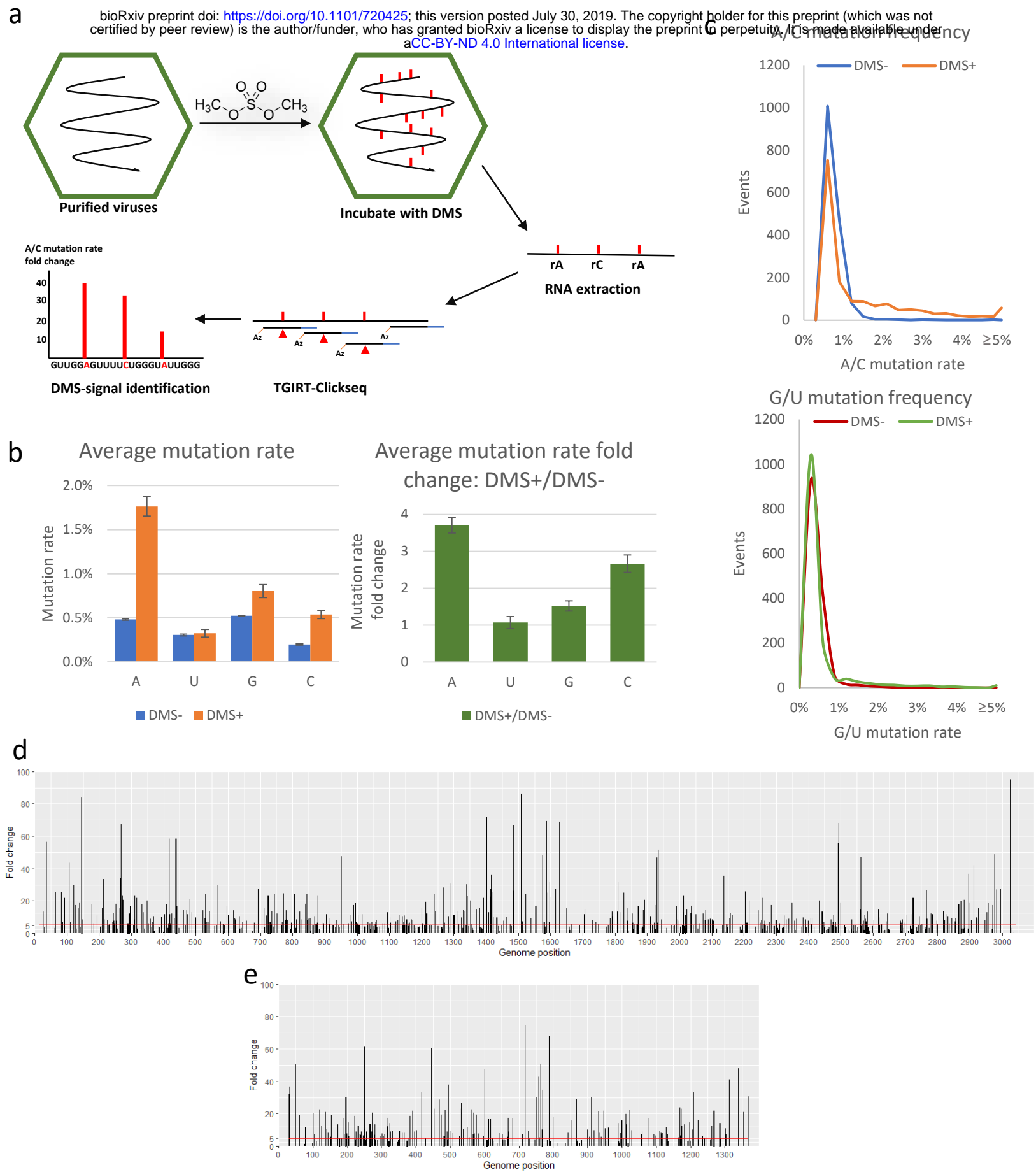
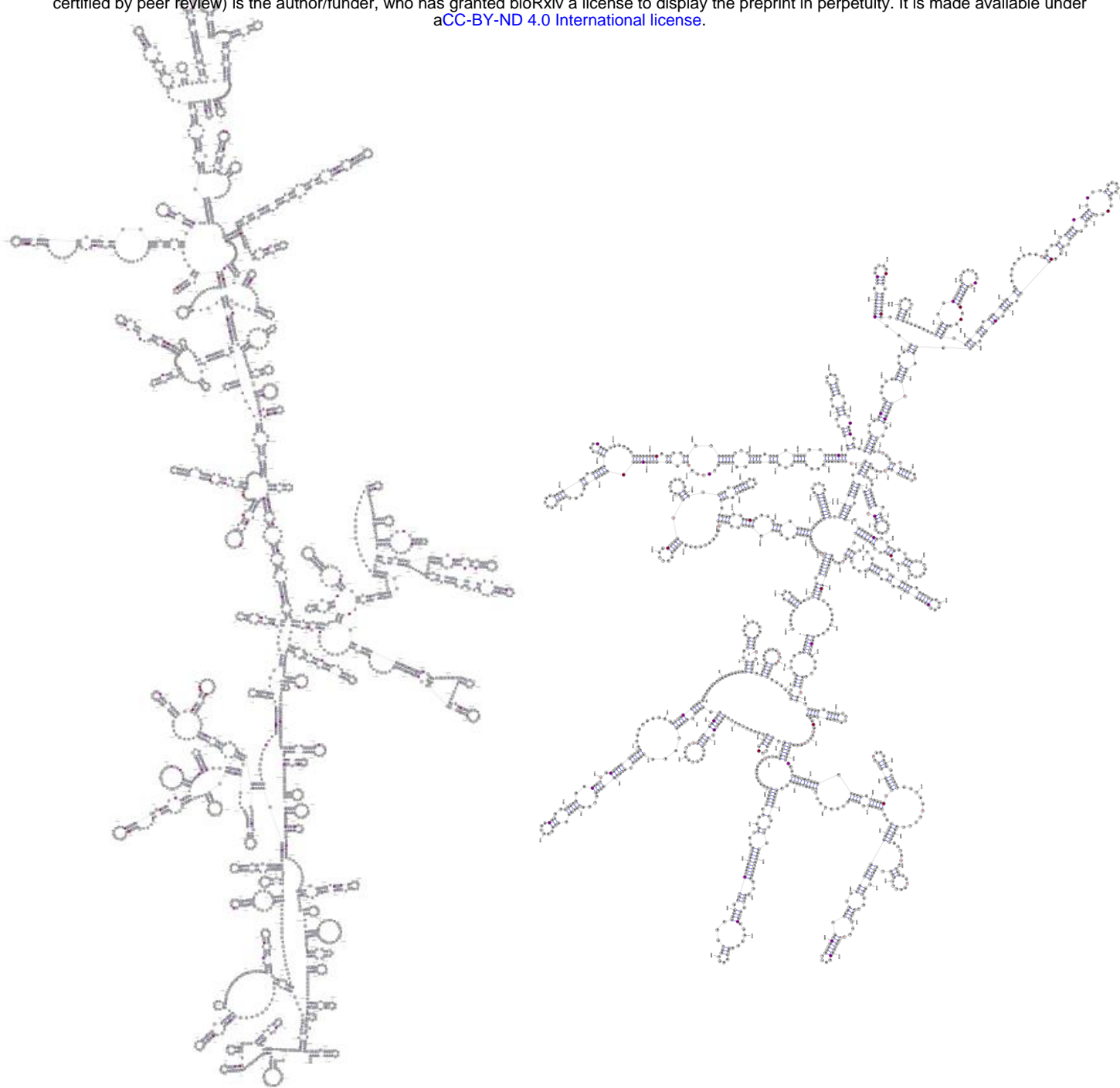


Figure 5



## Figure 5: FHV DMS-MaPseq

**(a)** In order to precisely predict the secondary structure of FHV RNA, we used DMS (dimethyl sulfide) to induce in virion methylation of unpaired adenines and cystines. The extracted RNAs were subjected to ClickSeq library preparation with TGIRT<sup>TM</sup>-III enzyme, which invokes mutations over methylated bases. DMS-MaPseq signal represents the mutation rate fold change over A/C positions. Red markers on RNA represent methylated ribonucleotides; red triangles on cDNA represent DMS-induced mutations. **(b)** of A/C mutations were detected as a result of DMS treatment. **(c)** DMS treated virus exhibited higher mutation rate frequencies over A/C positions. **(e,f)** DMS-MaPseq map of FHV RNA 1 and RNA 2, respectively. Background noise was removed. Red line represents the average DMS-MaPseq signal.



**Color scheme of base pairing probability:**

Probability  $\geq$  99%  
99% > Probability  $\geq$  95%  
95% > Probability  $\geq$  90%  
90% > Probability  $\geq$  80%  
80% > Probability  $\geq$  70%  
70% > Probability  $\geq$  60%  
60% > Probability  $\geq$  50%

**Nucleotide annotations:**

**U** : PAR-CL sites ( $>2\sigma$  significance)

**U** : PAR-CL sites ( $>1\sigma$  significance)

**c** **a** : DMS-MaPseq constraints

bioRxiv preprint doi: <https://doi.org/10.1101/720425>; this version posted July 30, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

**Figure 6. DMS-MaPseq-corrected secondary structure map of complete FHV RNA genome.** Snapshots of RNA 1 (left) and RNA 2 (right) are shown. Full scaled maps can be found in Supplementary data 1 and 2. PAR-CL signal sites of different significance were color annotated. The introduced DMS-MaPseq constraints were highlighted by lower case “a” or “c” in red color. The base pairing probabilities of representative PAR-CL sites are color schemed on the base pairing bonds. Some examples are shown in **Table 1**.

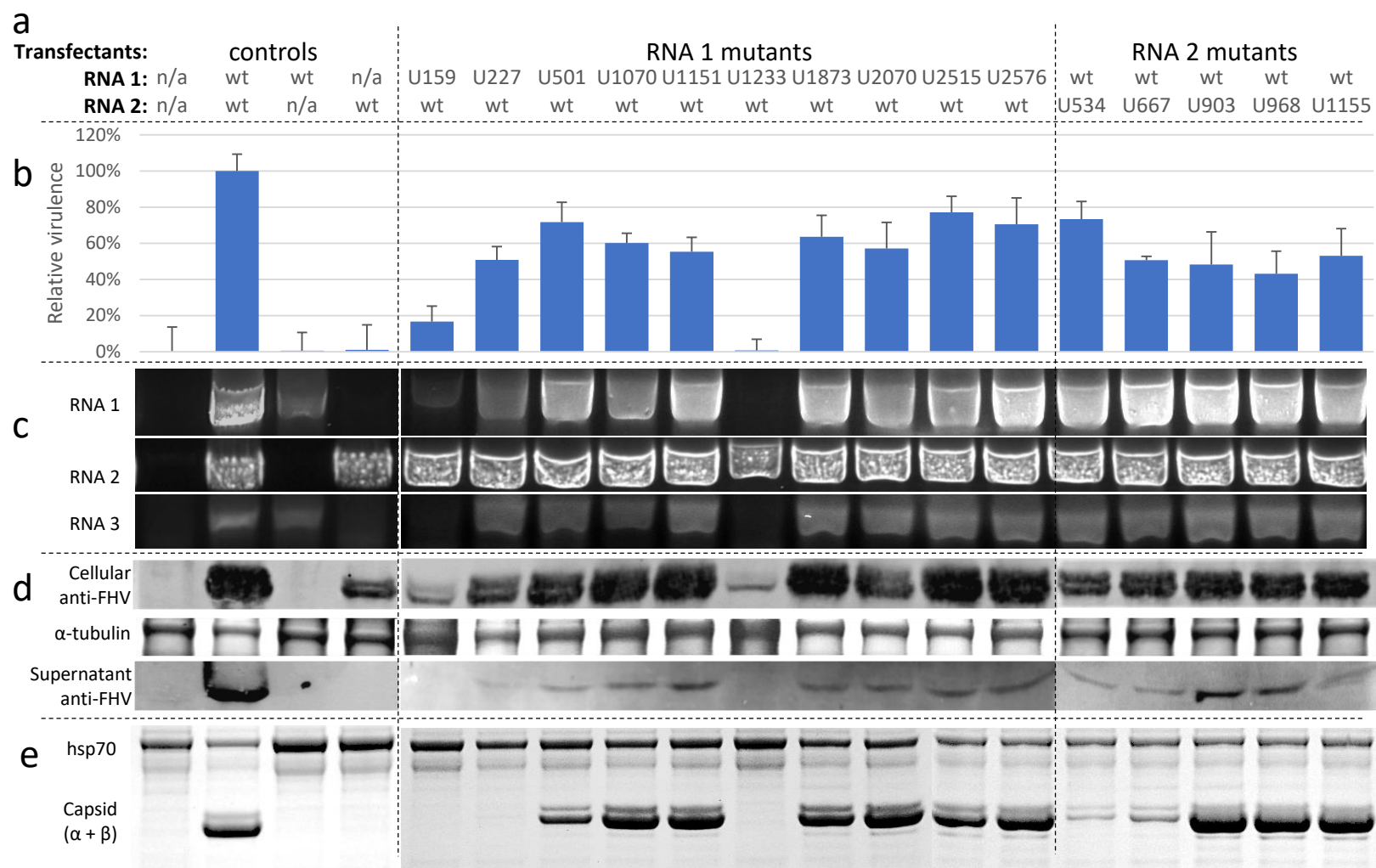
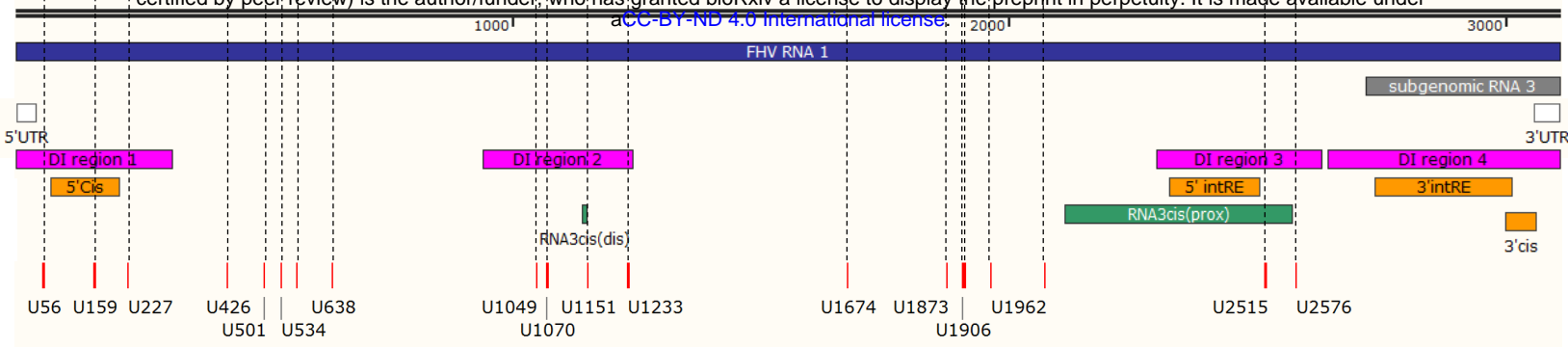
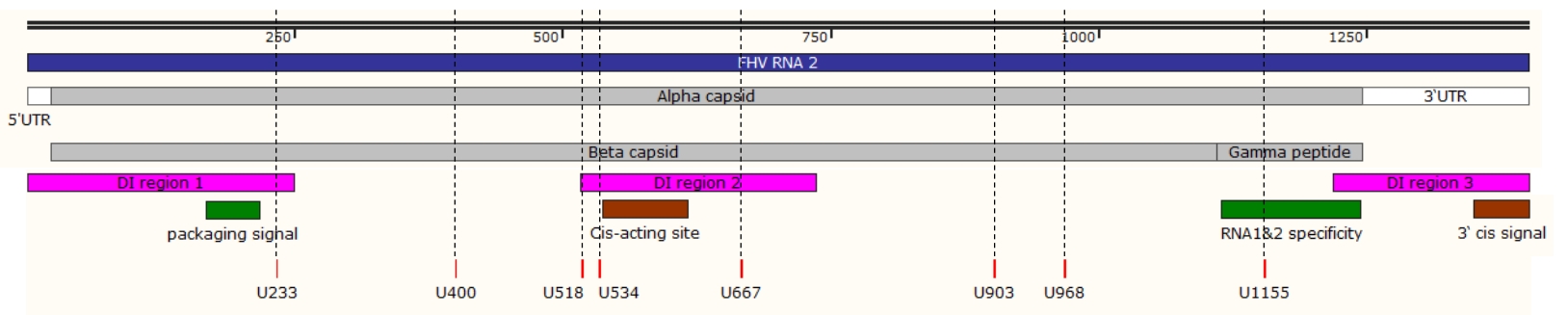


Figure 7

**Figure 7. FHV PAR-CL mutants characterization.** (a) Plasmids containing FHV RNA 1 or RNA 2 with mutated PAR-CL structures were transfected to S2 cells to yield p0 mutant viruses. (b) Relative virulence of p0 mutant viruses was determined with alamarBlue assay to measure cell viability after transfection. (c) 200ng of total cellular RNA of each transfection was analyzed with RT-PCR to measure the accumulation of FHV RNAs. (d) Cellular and supernatant FHV capsid productions were detected with anti-FHV antibodies. Coomassie-stained  $\alpha$ -tubulin as loading control for cellular assay. (e) p1 viruses were purified and filtered with 100 K molecular weight filter, mutant virus production was verified with SDS-PAGE gel. Heat shock protein 70 (hsp70) shown as a loading control.



a



b

**Figure 8** FHV significant PAR-CL signals and known motifs

**(a)** On RNA 1, four highly conserved defective interfering RNA regions (DI regions 1-4)(34) were shown in pink; four RNA 1 replication regulatory elements were shown in orange (5'Cis (46), 5'intRE(47), 3'intRE(47), and (67)); two RNA 3 regulatory elements and a putative RNA 3 subgenomic promoter region were shown in green (RNA3cis(dis) (47), RNA3cis(prox)(47)); the candidate PAR-CL sites (listed in Figure 4e) were shown as red bars on bottom. **(b)** On RNA 2, three conserved defection interfering RNA regions were shown in pink (DI regions 1-3 (34)); a mid-genome cis-acting replicational element(59) and a 3' cis-acting regulatory element(67) were shown in brown; a RNA 2 packaging signal (24) and a capsid site essential for RNA 1&2 specificity (18) are shown in green.

Table 1

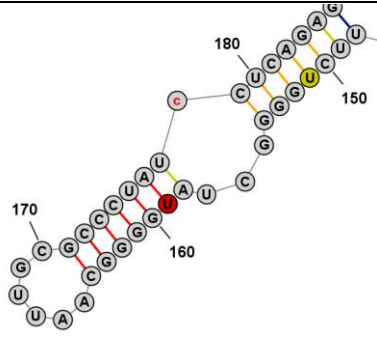
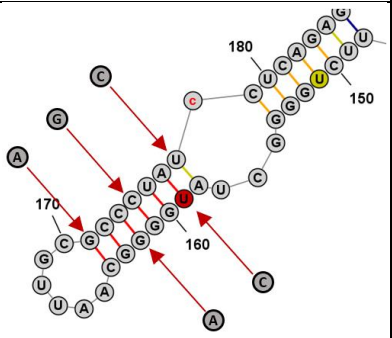
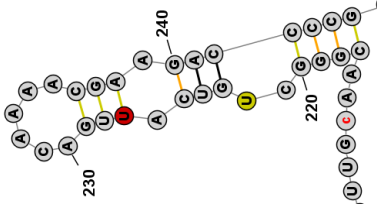
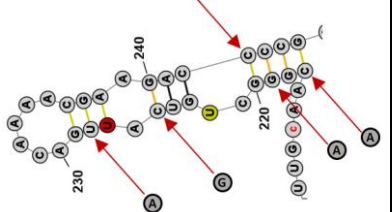
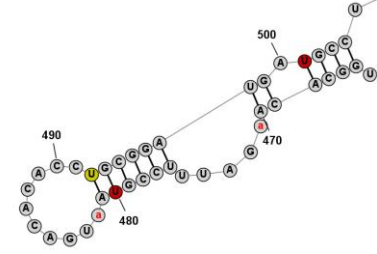
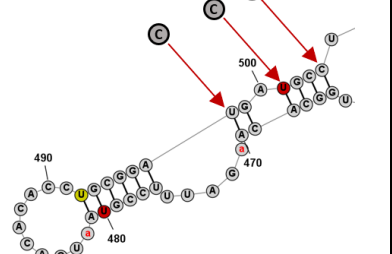
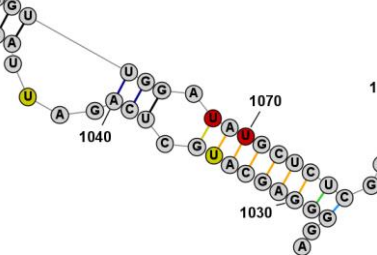
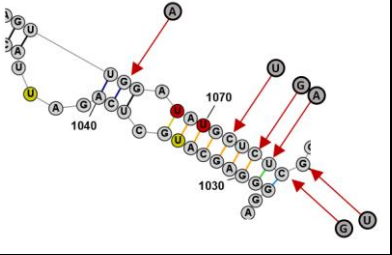
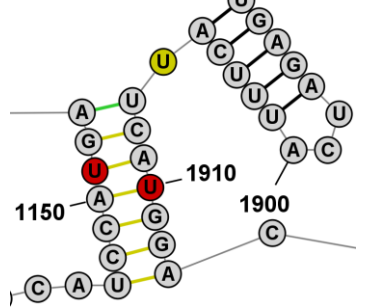
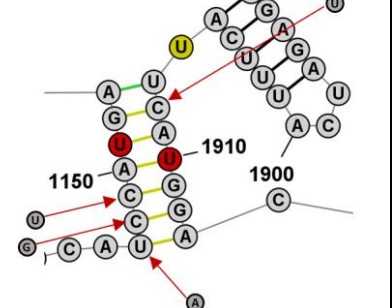
PAR-CL sites	Predicted structure	PRIMERS	Mutated nucleotides
U159(RNA1)		<p>F: TACGGAGCAATTGCACCGTACCCTC AGAGTGGAGGGAACCG</p> <p>R: TGCAATTGCTCCGTAGCCCCAGAAC TTGGATATGCAGTAC</p>	
U227(RNA1)		<p>F: AGAGCTGTGATAGACAAAACGAAG ACGCCGATAGAGACAGTTTCTATC CGCT</p> <p>R: GTCTATCACAGCTCTTTGCAATGCG CGTGTAACTC</p>	
U501(RNA1)		<p>F: GACGACGCGTTCATCGTCGGTGTT GATGTTGAT</p> <p>R: GATGAACGCGTCGTCGCCAGGTGT GTCATTACGGAAATC</p>	
U1070(RNA1)		<p>F: TTAGATATGTTGAGTGATTATCCG CCACCAATCTG</p> <p>R: ACTCAACATATCTAACTTTTCTTTCT CGATAGTAATCTGAGCATG</p>	
U1151 & U1910(RNA1)		<p>F: AGTATGATTGTCCAGTATTATACTG GCAAGAAGGTA</p> <p>R: CTGGACAATCATACTTGTGTATTGC GGTCCTTGTG</p>	



Table 1

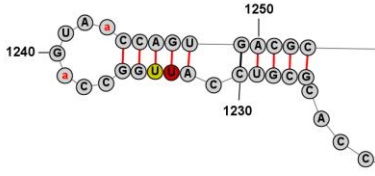
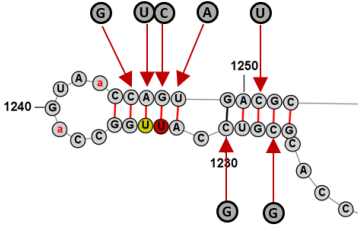
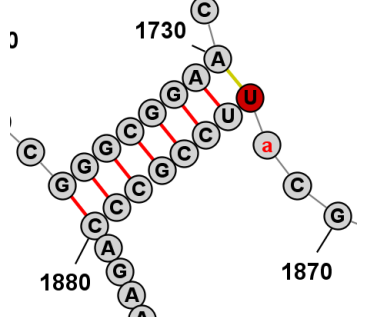
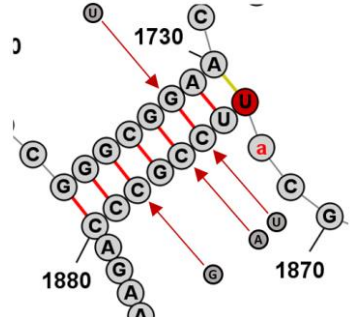
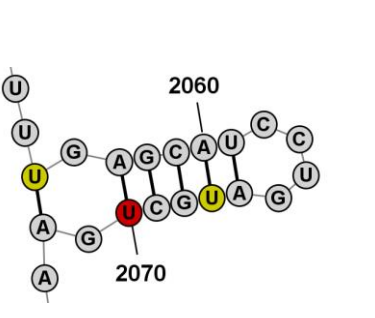
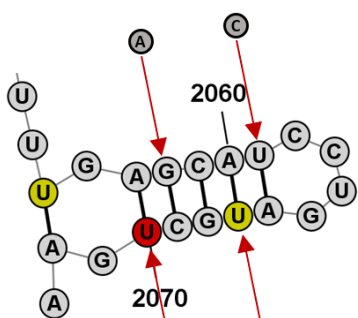
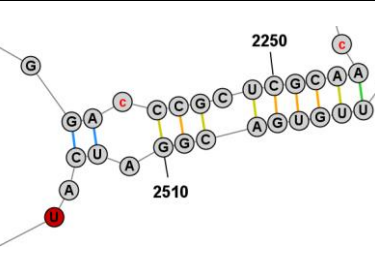
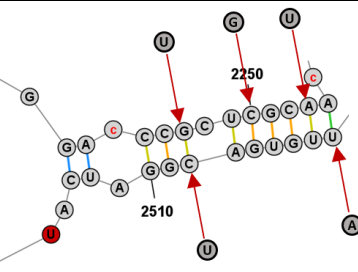
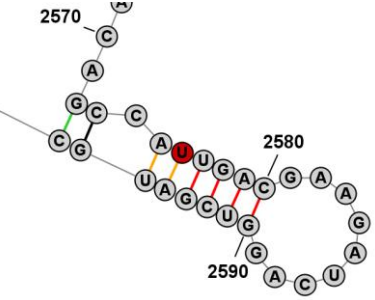
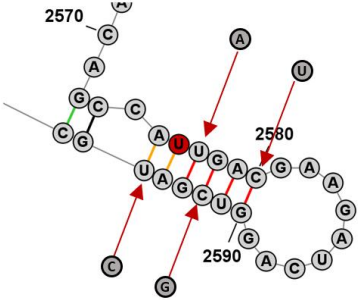
<p>U1233(RNA1)</p>		<p><b>F:</b> CGGGTGCATTGGCCAGTAACGTCA GATGCAGATGTACCAGAAGTGAGC</p> <p><b>R:</b> TGCCCAATGCACCCGTGGCATTGT AGGTTTATACACAGTTG</p>	
<p>U1873(RNA1)</p>		<p><b>F1:</b> GGGCGTAACCCGACTGAGATCGCC GAC</p> <p><b>R1:</b> AGTCGGGTTACGCCGGGTAATA CCAATGTTTCATTGTGTTTC</p> <p><b>F2:</b> GCATTTAGCCAGAATACAGAGAT GAGATCATTTTC</p> <p><b>R2:</b> TTCTGGCCTAAATGCTTGAACCATG GCCTTTTGGGCGA</p>	
<p>U2070(RNA1)</p>		<p><b>F:</b> GAACACCCTGACGCAGAACCTGAA GATTTGTCCGTTTAATCGGAC</p> <p><b>R:</b> TGCGTCAGGGTGTCAAAGGTCAG AGCTGTAATTCGACAC</p>	
<p>U2515(RNA1)</p>		<p><b>F1:</b> GACCCTGGCTACTACGACCACAA TTCAAGACCCAC</p> <p><b>R1:</b> AGTAGCCAGAGGGTCCACAAATAC ACGAGACAGGAAACAAAG</p> <p><b>F2:</b> ACATGTGATGGAAGTTGCCACAG CATCCGC</p> <p><b>R2:</b> ACTTCCATCACATGTCAACCAAGTAG GGCTTCTCTTTATTC</p>	
<p>U2576(RNA1)</p>		<p><b>F:</b> ATAGATGAAGATCAGGTGGACGCA CTCATTGGGCGTTTTG</p> <p><b>R:</b> CTGATCTTCACTATGGCTGTACGT TTGATTAAAACTGCT</p>	

Table 1

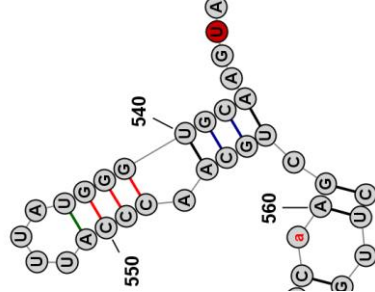
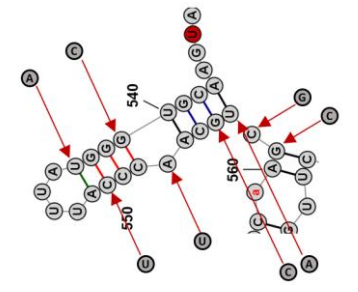
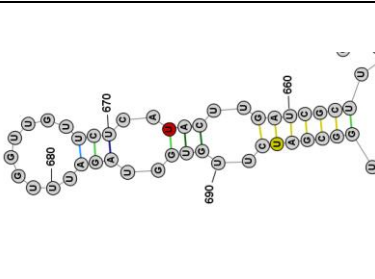
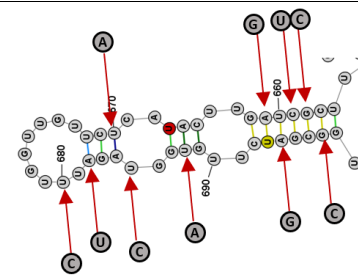
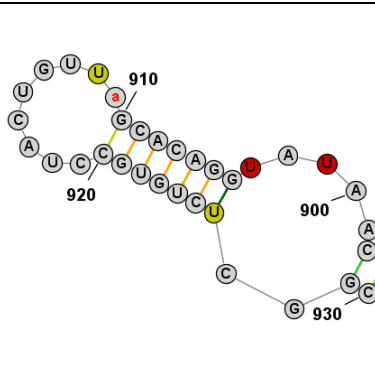
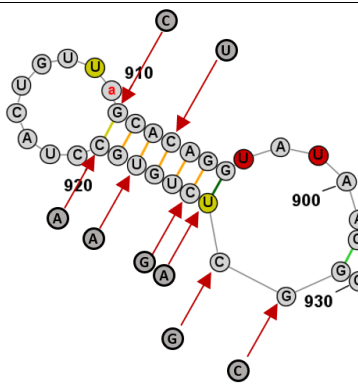
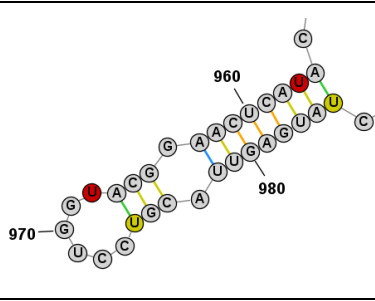
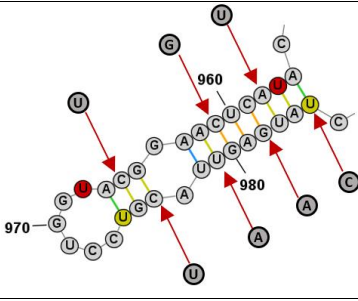
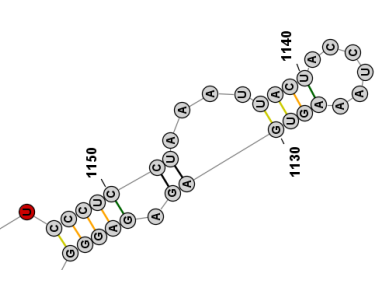
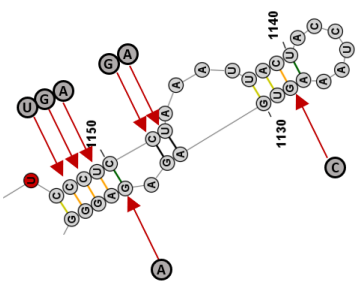
U534(RNA1)		<p><b>F:</b> GTCGGAAITTTATCCTACCAGCAACT TGATGACAGTTTGCCGGAAG</p> <p><b>R:</b> AGGATAAATTCGACGTTTCATGGA AGCGTACTCTGAATGAG</p>	
U667(RNA2)		<p><b>F:</b> TCCTGGTTCATACACTTGTGGTC TTGACGGAGTTCTGCCGTGGGGC CTGACAACCTTCTC</p> <p><b>R:</b> TGTATGAACCAAGGAACTGGTGGC TGGATCTGTTG</p>	
U903(RNA2)		<p><b>F:</b> GATACCATTGTCATCAGAGTGAGC GCCCTGAGGGCGCAGTTAA</p> <p><b>R:</b> GATGACAATGGTATCCATATTGCC CATCCGACTACTCCA</p>	
U968(RNA2)		<p><b>F:</b> ATTCTGAAGGCTTGGTCCTGTATAG AATACCGACCAAATCCAAACGCCA TG</p> <p><b>R:</b> CCAAGCCTTCAGAATGGCAGAGTT AACTGCGCCCT</p>	
U1155(RNA2)		<p><b>F:</b> GAAAGAGTCAAATCCATCATTAAA AGCAGTCTGGCTGCTGCAAGCAAC ATTC</p> <p><b>R:</b> GGATTTGACTCTTCCACATTGAT GCATTTTG</p>	

Table 1

Probability  $\geq$  99%  
99% > Probability  $\geq$  95%  
95% > Probability  $\geq$  90%  
90% > Probability  $\geq$  80%  
80% > Probability  $\geq$  70%  
70% > Probability  $\geq$  60%  
60% > Probability  $\geq$  50%

Ⓢ :  $\geq$  2sigma significant PAR-CL sites

Ⓤ :  $\geq$  1sigma < 2sigma significant PAR-CL sites

ⓐ ⓐ (lower cased red a/c): DMS-MaPseq unpaired constrains