

Introduction

The emergence of coronavirus disease 2019 (COVID-19) as a pandemic in 2020 spurred many investigations on functional RNA structures in coronaviruses, particularly SARS coronavirus 2 (SARS-CoV-2) [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. Among the more unexpected findings was an RNA:RNA interaction between the frameshifting stimulation element (FSE) and another sequence up to 1,475 nt downstream, which the authors named the FSE-arch [3]. The FSE-arch was detected in infected cells using COMRADES [11] and proposed to comprise three nested long-range RNA:RNA interactions (Figure ??a): an outer 38 bp bulged stem spanning coordinates 13,370-14,842 (which encompasses the FSE); a middle 18 bp bulged stem spanning coordinates 13,533-14,673; and an inner 14 bp bulged stem spanning coordinates 13,580-14,552 [3]. We had discovered that the FSE folds into at least two alternative structures in infected cells, in roughly equal proportions, and that the predicted structure for one of them resembles the FSE-arch [10]. Because computational RNA structure prediction – even guided by chemical probing data – is unreliable for long RNA sequences especially [12], we sought stronger, hypothesis-driven evidence for the existence of the FSE-arch.

Chemical probing followed by mutational profiling is a common strategy for inferring secondary structures of RNA molecules [13, 14].

Here, we present a method to probe RNA–RNA interactions spanning hundreds to thousands of nucleotides, “Structure Ensemble Ablation by Reverse Complement Hybridization with Mutational Profiling” (SEARCH-MaP). To compute, compare, and deconvolute data from mutational profiling experiments (including SEARCH-MaP, DMS-MaPseq, and SHAPE-MaP), we introduce the software “Structure Ensemble Inference by Sequencing, Mutation Identification, and Clustering of RNA” (SEISMIC-RNA).

Results

Strategy of SEARCH-MaP and SEISMIC-RNA

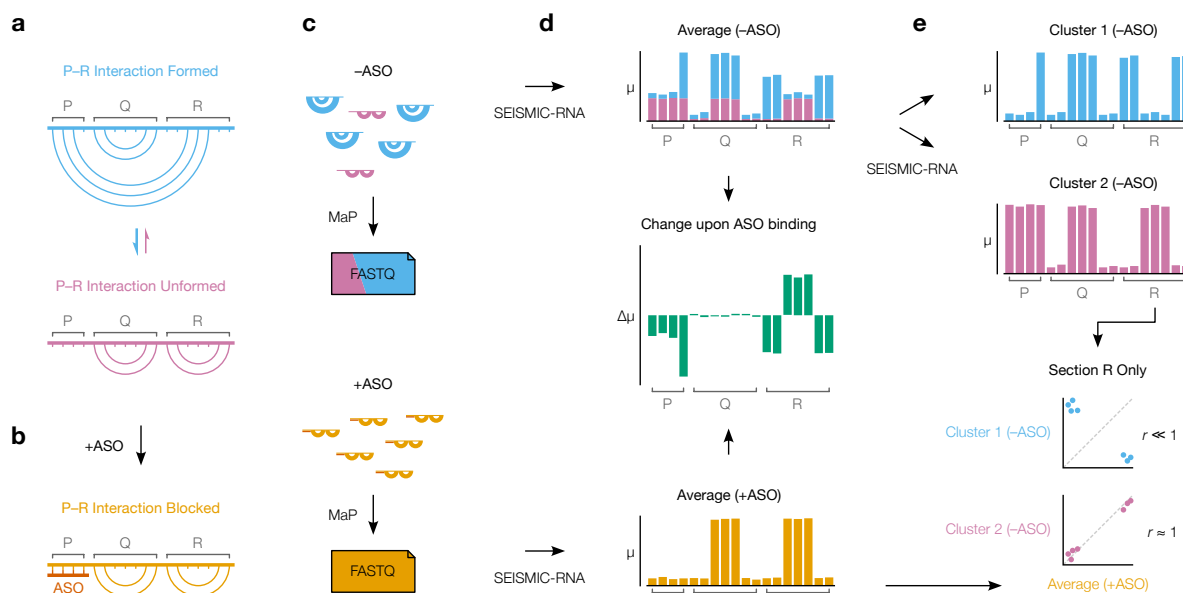


Figure 1: The strategy of SEARCH-MaP and SEISMIC-RNA. (a) This toy RNA is partitioned into three sections (P, Q, and R) whose molecules exist in two structural states: one in which an interaction between P and R forms (blue) and one in which it does not (purple). (b) Hybridizing an ASO (red) to P blocks it from interacting with R and forces all RNA molecules into the state where the P-R interaction is unformed. (c) A SEARCH-MaP experiment entails separate chemical probing and mutational profiling (MaP) with (+ASO) and without (-ASO) the ASO, followed by sequencing to generate FASTQ files. The RNA molecules and FASTQ files use the same color scheme as in (a) and are illustrated/colored in proportion to their abundances in the ensemble. (d) Ensemble average mutational profiles with (+ASO) and without (-ASO) the ASO, computed with SEISMIC-RNA. The x-axis is the position in the RNA sequence; the y-axis is the fraction of mutations (μ) at the position. Each bar in the -ASO profile is drawn in two colors merely to illustrate how much each structural state contributes to each position; in a real experiment, states cannot be distinguished before clustering. The change upon ASO binding (green) indicates the difference in the fraction of mutations ($\Delta\mu$) between the +ASO and -ASO conditions. (e) Mutational profiles of two clusters (top) obtained by clustering the -ASO ensemble in (d) using SEISMIC-RNA, and the scatter plot of the mutation rates of bases in R (bottom) between the +ASO ensemble average (x-axis) and each cluster (y-axis). The expected correlation (r) is shown beside each scatter plot.

We illustrate SEARCH-MaP with an RNA comprising three sections (P, Q, and R) that folds into an ensemble of two structural states: one in which a base-pairing interaction between P and R forms, another in which it does not (Figure 1a). Searching for sections that interact with P begins with hybridizing an antisense oligonucleotide (ASO) to P, which blocks P from base pairing with any other section, ablating the state in which the P–R interaction forms (Figure 1b). The RNA is chemically probed separately with (+ASO) and without (–ASO) the ASO, followed by mutational profiling and sequencing, e.g. using DMS-MaPseq [13] (Figure 1c).

SEISMIC-RNA can detect RNA–RNA interactions by comparing the +ASO and –ASO mutational profiles. Theoretically, each structural state has its own mutational profile [15], but the mutational profile of a single state is not directly observable because all states are physically mixed during the experiment (Figure 1c, top). Instead, the directly observable mutational profile is the “ensemble average” – the average of the states’ (unobserved) mutational profiles, weighted by the states’ (unobserved) proportions (Figure 1d, top). Because the structures – and therefore mutational profiles – of R differ between the interaction-formed and -unformed states, the ensemble averages of R also differ between the +ASO and –ASO conditions (Figure 1d, middle). However, this is not the case for element Q, which has the same secondary structure in both states (Figure 1d, middle). Therefore, one can deduce that P interacts with R – but not with Q – because hybridizing an ASO to P alters the mutational profile of R but not of Q.

After identifying RNA–RNA interactions, SEISMIC-RNA can also determine the mutational profiles of the states where the P–R interaction is formed and unformed – even if their secondary structures are unknown. Inferring mutational profiles for the interaction-formed and -unformed states requires clustering the –ASO ensemble into two clusters of RNA molecules (Figure 1e, top). Each cluster has its own mutational profile and corresponds to one structural state, but which cluster corresponds to the interaction-formed (or -unformed) state is not yet known. The interaction-unformed state has a mutational

profile similar to that of the +ASO ensemble average, since the ASO blocks the interaction and forces the RNA into the interaction-unformed state. Therefore, a cluster that correlates well ($r \approx 1$) with the +ASO ensemble average (here, Cluster 2) corresponds to the interaction-unformed state; while a cluster that correlates weakly ($r \ll 1$) corresponds to the interaction-formed state (Figure 1e, bottom).

(I hope) SEARCH-MaP detects long-range base-pairing in ribosomal RNA

We first validated SEARCH-MaP using 16S and 23S ribosomal RNA (rRNA) from *E. coli*. For each rRNA, we selected two RNA–RNA interactions spanning \geq [HOW MANY] nt that had been detected in a cell-free system [16]. For each interaction, we hypothesized that binding an ASO to either side would break the interaction and perturb the structure of the other side (distant from the ASO binding site) and designed two ASOs, one targeting each side. As a negative control, we also designed one ASO targeting a stem loop in each rRNA, which we hypothesized would perturb only the structure near the ASO binding site.

We folded the 16S and 23S rRNAs with each ASO, performed DMS-MaPseq over the entire transcripts, and compared ensemble average mutational profiles with and without ASOs using SEISMIC-RNA. [DESCRIBE THE RESULTS]

Figure 2:

SARS-CoV-2 tiling

Long-range RNA–RNA interactions in many species of virus regulate core processes such as viral protein synthesis [17].

In SARS coronavirus 2 (SARS-CoV-2), the frameshift stimulating element (FSE) was shown to base pair with another genomic element over 1,000 nt downstream, a struc-

ture the authors named the "FSE-arch" [3]. We had found that about 45% of the genomic RNA molecules within infected cells have DMS mutational profiles consistent with the FSE-arch [10], which had surprised us given the length of this RNA–RNA interaction. Therefore, we sought to investigate and potentially improve the model of the FSE-arch using SEARCH-MaP.

We *in vitro* transcribed a 2,924 nt RNA segment of the SARS-CoV-2 genome centered on the long-range interaction (Figure 3a). We added groups of DNA ASOs; each group targeted a different section of the RNA (Figure 3a). Groups 9 and 10 targeted the 3' side of the FSE-arch; we expected that if this structure exists, then adding either group should block part of the FSE-arch and change the structure near the FSE. We confirmed the ASOs bound using DMS-MaPseq (SFIG). We then assessed the structure near the FSE via DMS-MaPseq with RT-PCR primers flanking the FSE, including the entire 5' side of the FSE-arch.

The mutational profiles of the FSE region with no ASOs were highly reproducible: the Pearson correlation coefficient (PCC) between two replicates was 0.98 (Figure 3b, light gray). Binding ASO group F (targeting the FSE itself) plunged the correlation with the no-ASO control to 0.55 (Figure 3b, dark gray), confirming that we could detect ASO-induced structural changes around the FSE. Of the other ASO groups, only the addition of group 9 dropped the correlation with the no-ASO control below 0.90 (Figure 3b). This drop in correlation localized to the stems predicted to be part of the FSE-arch (SFIG, on rolling correlation). This result supports the inner two stems of the FSE-arch. That adding ASO group 10 had no effect on the FSE (PCC = 0.97) suggests that the outer stem either does not exist or forms less often or under more specific conditions than do the inner two stems.

We next sought to determine in what fraction of molecules the inner two stems of the FSE-arch fold. We clustered the reads for the no-ASO control using SEISMIC-RNA and found that they form at least two distinct clusters. These clusters were consistent with our previous data in Vero and Huh-7 cells [10] (SFIG), showing that this RNA segment adequately models of the RNA structure in the full-length virus. We compared the mutational

profile of each cluster to that of the ensemble average after adding ASO group 9 (Figure 3c, top). Cluster 2 (57% of the ensemble) was very similar (PCC = 0.95), suggesting that this cluster corresponds to the FSE-arch unformed. Cluster 1 (43% of the ensemble) was distinct (PCC = 0.64), suggesting that it corresponds to the the FSE-arch formed.

To gain further support for these assignments, we took advantage of having a preexisting model of the FSE-arch [3]. If these assignments were true, then the mutational profile of Cluster 1 should agree well with the structure of the FSE-arch (i.e. paired and unpaired bases should have low and high mutation rates, respectively); and Cluster 2 should agree less. We assessed the agreement by constructing a receiver operating characteristic (ROC) curve with respect to the inner two stems of the preexisting model (Figure 3c, bottom). The area under the curve (AUC) for Cluster 1 was 1.0, indicating perfect agreement with the inner two stems of the FSE-arch, while that of Cluster 2 (AUC = 0.57) was only marginally better than the null expectation of 0.50. This result supports that Cluster 1 is the mutational profile when the inner two stems of the FSE-arch form, and Cluster 2 when they do not.

If the FSE-arch exists, then blocking the 5' side of the FSE-arch should also alter the structure of the 3' side. We investigated by RT-PCR the region surrounding the inner two stems of the FSE-arch on the 3' side with and without adding ASO group F, which targets the FSE. Similar to the previous result, we found that the 3' side of the FSE-arch also forms two clusters of roughly even proportions (Figure 3d). The mutational profile (ignoring one outlier) of Cluster 2 resembled blocking the FSE-arch (PCC = 0.95), while Cluster 1 did not (PCC = 0.80); and Cluster 1 agreed with the FSE-arch model (AUC = 1.0), while Cluster 2 did not (AUC = 0.57). Thus, the 3' side of the FSE-arch also appears to form two structural states, one of which corresponds to the long-range interaction forming.

We conclude that this 2,924 nt segment mimics the FSE-arch in cells and exists as a structure ensemble in which the inner two stems of the FSE-arch fold in $47\% \pm 4\%$ of the molecules (Figure 3e). As this structure folds *in vitro*, it depends on the RNA sequence itself,

not on proteins or other cellular/viral factors. Moreover, we generated a mutational profile of the formed and the unformed states on both sides of the FSE-arch, which could assist with structure modeling.

The frameshifting pseudoknot competes with the long-range interaction

We refined the model of the long-range interaction using the mutational profiling data from both clusters that corresponded to this interaction forming (Figure 3c, d). We predicted the structure of a 1,799 nt segment of the genome centered on the long-range interaction using RNAstructure Fold [18]. The minimum free energy structure (Figure 4) contained not only the two inner stems of the FSE-arch (LS1 and LS2a/b) but also two additional stems that were not part of the original FSE-arch model [3] (LS3a/b and LS4). The structure also contained the alternative stem 1 (AS1) that we had previously discovered [10], which encircles the attenuator hairpin (AH) [19]. Because we had collected DMS-MaPseq data on both ends of LS3a/b, but only the 5' end of LS4, we focused on LS3a/b.

To our surprise, LS2b, LS3a/b, and LS4 of the new model penetrate into structures within the FSE pseudoknot that is widely thought to stimulate frameshifting – LS2b overlaps with PS2, LS3 with PS3, and LS4 with PS1. We had previously shown that AS1 also overlaps with and seems to outcompete PS1 [10]. If these stems existed, then they would be mutually exclusive with the FSE pseudoknot, which suggests that the long-range interaction could inhibit the formation of the pseudoknot and possibly regulate the rate of frameshifting.

Thus, we verified these structures of the long-range interaction using SEARCH-MaP. We performed SEARCH-MaP on the same 1,799 nt segment for which we had predicted the structure, this time with shorter LNA/DNA mixmer ASOs (15-20 nt) to reach single-stem precision. Each ASO targeted a single stem in the downstream portion of the interaction,

and we measured the change in DMS reactivities of the FSE. ASOs targeting the 3' sides of LS1 and LS2a perturbed the DMS reactivities in exactly the expected locations on the 5' sides (Figure 4b). Binding an ASO to the 3' side of LS2b caused a larger perturbation with more off-target effects, likely because this stem overlaps with stem 2 of the pseudoknot (PS2), so blocking it with an ASO could promote pseudoknot formation. Blocking LS3b also resulted in a main effect around the intended location, with one off-target effect upstream, suggesting that there may be another RNA–RNA interaction with the pseudoknot and this upstream region. These results show that stems LS1, LS2a/b, and LS3b in the refined model of the long-range interaction do exist and can be detected with SEARCH-MaP – cleanly if the stem does not interact with anything else, otherwise with off-target effects.

Assuming that all stems in Figure 4a do exist, we found that there are six possible structure models resulting from all possible combinations of non-overlapping stems (Figure 4c). We ordered these models from most long-range character (A) to most pseudoknot-like character (F). We then determined whether each model actually exists in the ensemble and estimated its proportion.

We found that our no-ASO control clustered reproducibly up to 6 clusters (SFIG). For each cluster, we found how well it agreed with each structure model by calculating the area under the receiver operating characteristic curve (AUC-ROC). We graph the results as a heatmap with clusters on the x-axis, in order of and their widths indicating their proportions in the ensemble, and the models on the y-axis. We consider a cluster and model to be consistent with each other if the AUC-ROC is at least 0.90, and explicitly print all such AUC-ROC values. We found that model C – in which AS1 folds along with stems PS2 and PS3 of the pseudoknot – was consistent with three clusters representing 52% of the ensemble (Figure 4d, top). Model A – where the full long-range interaction forms – was consistent with one cluster (20%). No clusters were consistent with the pseudoknot (Model F); the least-abundant cluster (7%) came close with an AUC-ROC of 0.88. The remaining cluster

(21%) was not even close to being consistent with any model, suggesting that there are still other structures in the ensemble besides those in Figure 4c.

These results suggested to us that the primary competitors of the pseudoknot are AS1 and LS2b, since both are present in the most abundant model, C (52%), while LS3 and LS4 fold in model A, which is only 20% of the ensemble. We thus reasoned that if the long-range interaction does actually compete with the pseudoknot, then blocking AS1 and LS2b simultaneously should allow the pseudoknot to fold, while blocking either would have a much smaller effect on the pseudoknot. To test this hypothesis, we repeated the above experiment while adding an ASO targeting either AS1 or just the part of LS2b that overlaps with the pseudoknot, or both ASOs together. Blocking AS1 (Figure 4d, left) reduced the proportion of clusters consistent with AS1 (Models A, B, and C) from 72% to 16%, as expected; it also resulted in two clusters consistent with the pseudoknot (56% total) and one cluster (20%) consistent with Model D, which includes PS1. Blocking the part of LS2b that seems to overlap with PS2 (Figure 4d, right) eliminated Model A but not Model C (as expected, since Model A includes LS2b, while Model C does not), and also produced one cluster (13%) that was consistent with the pseudoknot. Blocking both AS1 and LS2b together (Figure 4d, bottom) forced the entire ensemble into the pseudoknot state, with 87% of the ensemble highly consistent with Model F (AUC-ROC = 0.97) and the remaining 13% still somewhat consistent (AUC-ROC = 0.87). Thus, we conclude that the long-range interaction – particularly LS2b – along with AS1, does compete with the pseudoknot.

Frameshift stimulating elements of multiple coronaviruses participate in long-range RNA–RNA interactions

We hypothesized that similar long-range interactions could exist in other coronaviruses – particularly other SARS-related viruses. To test this hypothesis, we performed SEARCH-MaP with FSE-targeted ASOs on 1,799 nt segments from eight selected coronaviruses.

Computational and experimental screening identifies eight coronaviruses with potential long-range interactions

As of December 2021, the NCBI Reference Sequence Database [20] contained 62 complete genomes of coronaviruses. To focus on those likely to have long-range interactions involving the FSE, we predicted the likelihood that each base in a 2,000 nt section surrounding the FSE would pair with a base in the FSE (SFIG). Based on these predicted interactions, we selected ten coronaviruses – at least one from each genus (SFIG) – including SARS-CoV-2 as a positive control. Within the genus *Betacoronavirus*, we included all three of the SARS-related viruses – SARS coronaviruses 1 (NC_004718.3) and 2 (NC_045512.2) and bat coronavirus BM48-31 (NC_014470.1) – because they clustered into their own structural outgroup, distinct from all other coronaviruses. The other three strains of *Betacoronavirus* that we selected were MERS coronavirus (NC_019843.3) with a predicted interaction at positions 510-530; and human coronavirus OC43 (NC_006213.1) and murine hepatitis virus strain A59 (NC_048217.1), both with a predicted upstream interaction at positions 10-20. We selected two strains of *Alphacoronavirus*: transmissible gastroenteritis virus (NC_038861.1) and bat coronavirus 1A (NC_010437.1), predicted to have interactions at positions 440-460 and 350-360, respectively. Avian infectious bronchitis virus strain Beaudette (NC_001451.1) – a strain of *Gammacoronavirus* – was predicted to have a strong interaction at positions 330-350, while common mormon coronavirus HKU21 (NC_016996.1) was the species of *Deltacoronavirus* with the most promising FSE interactions.

We reasoned that if an FSE does interact with a distant RNA element, then removing that element by truncating the RNA would break the interaction, causing a structural change in the FSE that could be detected through chemical probing. For each of the ten coronaviruses that passed the computational screen, we *in vitro* transcribed and performed DMS-MaPseq [13] on both a 239 nt segment comprising the FSE and minimal flanking

sequences and a 1,799 nt segment encompassing the FSE and all sites with which it was predicted to interact. All coronaviruses except for human coronavirus OC43 and MERS coronavirus showed differences in their DMS reactivity profiles between the 239 nt and 1,799 nt segments (SFIG), suggesting long-range interactions involving the FSE.

SEARCH-MaP reveals long-range interactions involving the FSE in four other coronaviruses

To determine which RNA elements the FSE base-pairs with in each coronavirus, we performed SEARCH-MaP on the 1,799 nt RNA segment using DNA ASOs targeting the vicinity of the FSE (Figure 5). The rolling Spearman correlation coefficient (SCC) between the +ASO and no-ASO mutational profiles dipped below 0.9 at the ASO target site in every coronavirus segment, confirming the ASOs bound and altered the structure.

To confirm we could detect long-range interactions, we compared the rolling SCC for the SARS-CoV-2 segment to our model of the long-range interaction (Figure 4a, green). The SCC dipped below 0.9 at positions 1,483-1,560 and at 1,611-1,642, which coincide with stems LS2a-LS3b (positions 1,476-1,550 within the 1,799 nt segment) and stem LS4 (positions 1,600-1,622) of the long-range interaction. These dips were the two largest downstream of the FSE; although others (corresponding to no known base pairs) existed, they were barely below 0.9 and could have resulted from base pairing between these regions and other (non-FSE) regions. Near LS1 (positions 1,367-1,381), the SCC dipped only slightly to a minimum of 0.95, presumably because LS1 is the smallest (15 nt) and most isolated long-range stem. Therefore, this method was sensitive enough to detect all but the smallest long-range stem, and specific enough that the two largest dips corresponded to validated base pairs.

We found similar long-range interactions in SARS-CoV-1 and another SARS-related virus, Bat coronavirus BM48-31. Both viruses showed dips in SCC at roughly the same positions as LS2a-LS4 in SARS-CoV-2, indicating that they have homologous structures.

SARS-CoV-1 also had a wide dip below 0.9 at positions 1,284-1,394, corresponding to a homologous LS1. Thus, three SARS-related viruses share this multi-stemmed long-range interaction involving the FSE, suggesting this structure is functional.

In every other species except common moorhen coronavirus, we found prominent dips in SCC at least 200 nt from the ASO target site. To model potential base pairing between these dip positions and the FSE, we used the Fold program from RNAstructure [18] with the no-ASO mutational profiles as DMS constraints [21]. We obtained models consistent with the SEARCH-MaP data for both transmissible gastroenteritis virus (TGEV) and murine hepatitis virus (Figure 5, orange). We subsequently focused on transmissible gastroenteritis virus because its long-range interaction lay downstream of the FSE, as in the SARS-related viruses.

To verify that the long-range interaction also forms in live transmissible gastroenteritis virus (TGEV), we infected ST cells with TGEV (two biological replicates) and performed DMS-MaPseq (two technical replicates per biological replicate). The DMS reactivities were highly reproducible over the whole TGEV genome ($r = 0.96-0.97$, SFIG). As expected, they differed from those of the 1.8 kb segment *in vitro* ($r = 0.82$, SFIG), showing why it is necessary to verify the long-range interaction in TGEV-infected cells.

First, to determine whether the FSE and the region with which it may interact form alternative structures, we amplified and deeply sequenced these two regions from each sample. Clustering the reads using SEISMIC-RNA revealed that both regions adopt at least two alternative structures. The two clusters of the downstream region differed most around positions 1,120-1,140 – the site of the 3' end of the predicted long-range interaction. In cluster 1 (63% of the ensemble), bases 1,129-1,136 (all part of the predicted interaction) had DMS reactivities less than 0.01; while in cluster 2, the DMS reactivities were all greater

than 0.01. This result suggests that cluster 1 corresponds to the state in which the long-range interaction forms.

Discussion

In this work, we developed SEARCH-MaP and SEISMIC and applied them jointly to detect structural ensembles involving long-range RNA:RNA interactions in SARS-CoV-2 and other coronaviruses. This study is certainly not the first to perturb RNA structure with ASOs, nor even the first to use DMS-MaPseq to quantify the structural changes upon binding ASOs to SARS-CoV-2 RNA [22]. But while this previous study examined local structural perturbations caused by binding an ASO, we show that we can detect changes in the structure at more distant locations in an RNA molecule that interact with the nucleotides bound by an ASO.

SEARCH-MaP bears conceptual similarity to another method, mutate-and-map read out through next-generation sequencing (M2-seq) [23]. Both involve perturbing one region of an RNA molecule (in the case of M2-seq, by pre-installing mutations through error-prone PCR) and measuring the effects on other bases in the RNA using chemical probing. The major differences are the precision and scale of the interactions identified, as well as the throughput. M2-seq can pinpoint interactions down to the resolution of a single base pair, and is thus more precise than SEARCH-MaP. However, DMS-guided RNA structure prediction can propose structure models at single-base-pair resolution, which SEARCH-MaP can validate, and in this way achieve single-base-pair resolution. SEARCH-MaP is also capable of finding interactions over a much longer range because M2-seq requires the interacting bases to be in the same Illumina sequencing read. Within this length limit, one M2-seq experiment can theoretically find all pairwise interactions between bases, while one SEARCH-MaP experiment can find only interactions that involve the region to which the ASOs were hybridized. M2-seq is also limited by the formation of alternative structures.

Some methods, such as [CITE something by Rhiju, maybe REEFIT] and DANCE-MaP [CITE], have been designed to work around this limitation SEARCH-MaP; however, [something by Rhiju] has [this problem], and DANCE-MaP requires extremely high sequencing depth of several million reads [MORE PRECISE]. SEARCH-MaP, by contrast, assumes from the start that the RNA may form alternative structures; for simply detecting long-range interactions, even a 5,000 read depth is sufficient coverage; and for clustering, we have found [SOME LIMIT].

Another limitation of SEARCH-MaP as presented here is that it cannot distinguish between direct and indirect interactions. If RNA segment A interacts with segment B, while B interacts with both segment A and C, then hybridizing an ASO to segment A would perturb the structure of B, which could consequentially perturb the structure of C. Hence, C would appear to interact with A, even though this interaction is indirect, through B. One possible workaround (not shown in this study) would be to mutate or hybridize an ASO to segment B, and then repeat the experiment with hybridizing an ASO to segment A. If the interaction between A and C is direct, then C should still be perturbed even when segment B is incapable of interacting with A or C. But if B mediates an indirect interaction between A and C, then disrupting B should eliminate the apparent interaction between A and C.

Functional long-range interactions up to four kilobases involving an FSE have been found previously in two plant viruses [24, 25]. In both cases, frameshifting required the long-range interaction, suggesting that this interaction enables negative feedback on synthesis of viral RNA polymerase [24]. When polymerase levels are low, the interaction would form and stimulate frameshifting, which is needed to synthesize RNA polymerase. Once the polymerase had accumulated, it would begin to replicate the genomic RNA; in its passage from the genomic 3' end to the 5' end, it would disrupt the 3' side of the long-range interaction, attenuating frameshifting and reducing synthesis of more polymerase.

However, this strategy cannot be the role, if any, of the long-range interactions in coronaviruses. Unlike in the two plant viruses, a long-range interaction is not required to stim-

ulate frameshifting in coronaviruses: numerous studies have shown that even the isolated FSE can cause 15 - 40% of ribosomes to frameshift [26, 27, 28, 10, 29, 30, 31]. In coronaviruses, the long-range interaction is not only unnecessary for frameshifting but also may even attenuate it, given that in SARS-CoV-2, the FSE-arch and the frameshift-stimulating pseudoknot seem to be mutually exclusive. Moreover, coronaviruses partition translation and RNA synthesis into two different cellular compartments (the cytosol and the double-membrane vesicles, respectively) [32], so structural changes induced by RNA polymerases would not be seen by ribosomes.

The functions of these long-range interactions involving the FSE in coronaviruses remain mysterious. However, given that they occur in multiple coronaviruses across at least two genera, it seems reasonable that they could play a role in the viral life cycle, possibly by affecting the rate of frameshifting. Further research may reveal new mechanisms of translational regulation in coronaviruses via long-range RNA:RNA interactions.

Methods

Screening coronavirus long-range interactions computationally

All coronaviruses with reference genomes in the NCBI Reference Sequence Database [20] were searched for using the following query:

```
refseq[filter] AND ("Alphacoronavirus"[Organism] OR  
                    "Betacoronavirus"[Organism] OR  
                    "Gammacoronavirus"[Organism] OR  
                    "Deltacoronavirus"[Organism])
```

The complete record of every reference genome was downloaded both in FASTA format (for the reference sequence) and in Feature Table format (for feature locations). The location of the frameshift stimulating element (FSE) in each genome was estimated from the feature table, and the nearest instance of TTTAAAC was used as the slippery site, using a custom Python script. The 2,000 nt segment beginning 100 nt upstream of and ending 1,893 nt downstream of the slippery site was used for predicting long-range interactions involving the FSE. Genomes with ambiguous nucleotides (e.g. N) in this segment were discarded. For each coronavirus genome, up to 100 secondary structure models of the 2,000 nt segment were generated using Fold version 6.3 from RNAstructure [18] with `-M 100` and otherwise default parameters. Then, for each position, the fraction of models for the coronavirus in which the base at the position paired with any other base between positions 101 (the first base of the slippery sequence) and 250 was calculated using a custom Python script. The coronaviruses were clustered by their fraction vectors using the unweighted pair group method with arithmetic mean (UPGMA) and a euclidean distance metric, implemented in Seaborn version 0.11 [33] and SciPy version 1.7 [34]. The resulting hierarchically-clustered heatmap was examined manually to select coronaviruses based on the prominence of potential long-range interactions with the FSE (relatively large fractions far from positions 101-250).

References

- [1] Ramya Rangan, Ivan N. Zheludev, Rachel J. Hagey, Edward A. Pham, Hannah K. Wayment-Steele, Jeffrey S. Glenn, and Rhiju Das. Rna genome conservation and secondary structure in sars-cov-2 and sars-related viruses: a first look. *RNA*, 26(8):937–959, 2020.
- [2] Ilaria Manfredonia, Chandran Nithin, Almudena Ponce-Salvatierra, Pritha Ghosh, Tomasz K. Wirecki, Tycho Marinus, Natacha S. Ogando, Eric J. Snijder, Martijn J. van Hemert, Janusz M. Bujnicki, and Danny Incarnato. Genome-wide mapping of sars-cov-2 rna structures identifies therapeutically-relevant elements. *Nucleic Acids Research*, 48:12436–12452, 2020.

- [3] Omer Ziv, Jonathan Price, Lyudmila Shalamova, Tsveta Kamenova, Ian Goodfellow, Friedemann Weber, and Eric A. Miska. The short- and long-range rna-rna interactome of sars-cov-2. *Molecular Cell*, 80:1067–1077.e5, 12 2020.
- [4] Lei Sun, Pan Li, Xiaohui Ju, Jian Rao, Wenzhe Huang, Lili Ren, Shaojun Zhang, Tuanlin Xiong, Kui Xu, Xiaolin Zhou, Mingli Gong, Eric Miska, Qiang Ding, Jianwei Wang, and Qiangfeng Cliff Zhang. In vivo structural characterization of the sars-cov-2 rna genome identifies host proteins vulnerable to repurposed drugs. *Cell*, 184:1865–1883.e20, 2021.
- [5] Yan Zhang, Kun Huang, Dejian Xie, Jian You Lau, Wenlong Shen, Ping Li, Dong Wang, Zhong Zou, Shu Shi, Hongguang Ren, Youliang Wang, Youzhi Mao, Meilin Jin, Grzegorz Kudla, and Zhihu Zhao. In vivo structure and dynamics of the sars-cov-2 rna genome. *Nature Communications*, 12:5695, 9 2021.
- [6] Nicholas C. Huston, Han Wan, Madison S. Strine, Rafael de Cesaris Araujo Tavares, Craig B. Wilen, and Anna Marie Pyle. Comprehensive in vivo secondary structure of the sars-cov-2 genome reveals novel regulatory motifs and mechanisms. *Molecular Cell*, 81, 2021.
- [7] Ramya Rangan, Andrew M. Watkins, Jose Chacon, Rachael Kretsch, Wipapat Kladwang, Ivan N. Zheludev, Jill Townley, Mats Rynge, Gregory Thain, and Rhiju Das. De novo 3d models of sars-cov-2 rna elements from consensus experimental secondary structures. *Nucleic Acids Research*, 49:3092–3108, 4 2021.
- [8] Edoardo Morandi, Ilaria Manfredonia, Lisa M. Simon, Francesca Anselmi, Martijn J. van Hemert, Salvatore Oliviero, and Danny Incarnato. Genome-scale deconvolution of rna structure ensembles. *Nature Methods*, 18:249–252, 2 2021.
- [9] Siwy Ling Yang, Louis DeFalco, Danielle E. Anderson, Yu Zhang, Jong Ghut Ashley Aw, Su Ying Lim, Xin Ni Lim, Kiat Yee Tan, Tong Zhang, Tanu Chawla, Yan Su, Alexander Lezhava, Andres Merits, Lin Fa Wang, Roland G. Huber, and Yue Wan. Comprehensive mapping of sars-cov-2 interactions in vivo reveals functional virus-host interactions. *Nature Communications*, 12, 2021.
- [10] Tammy C.T. Lan, Matty F. Allan, Lauren E. Malsick, Jia Z. Woo, Chi Zhu, Fengrui Zhang, Stuti Khandwala, Sherry S.Y. Nyeo, Yu Sun, Junjie U. Guo, Mark Bathe, Anders Näär, Anthony Griffiths, and Silvi Rouskin. Secondary structural ensembles of the sars-cov-2 rna genome in infected cells. *Nature Communications*, 13:1128, 3 2022.
- [11] Omer Ziv, Marta M. Gabryelska, Aaron T.L. Lun, Luca F.R. Gebert, Jessica Sheu-Gruttadauria, Luke W. Meredith, Zhong Yu Liu, Chun Kit Kwok, Cheng Feng Qin, Ian J. MacRae, Ian Goodfellow, John C. Marioni, Grzegorz Kudla, and Eric A. Miska. Comrades determines in vivo rna structures and interactions. *Nature Methods*, 15:785–788, 9 2018.

- [12] Sharon Aviran and Danny Incarnato. Computational approaches for rna structure ensemble deconvolution from structure probing data. *Journal of Molecular Biology*, 434:167635, 9 2022.
- [13] Meghan Zubradt, Paromita Gupta, Sitara Persad, Alan M. Lambowitz, Jonathan S. Weissman, and Silvi Rouskin. Dms-mapseq for genome-wide or targeted rna structure probing in vivo. *Nature Methods*, 2254:219–238, 2016.
- [14] Nathan A. Siegfried, Steven Busan, Gregory M. Rice, Julie A.E. Nelson, and Kevin M. Weeks. Rna motif discovery by shape and mutational profiling (shape-map). *Nature methods*, 2014.
- [15] Chringma Sherpa, Jason W. Rausch, Stuart F.J. Le Grice, Marie Louise Hammarskjold, and David Rekosh. The hiv-1 rev response element (rre) adopts alternative conformations that promote different rates of virus replication. *Nucleic Acids Research*, 43:4676–4686, 3 2015.
- [16] Anthony M. Mustoe, Nicole N. Lama, Patrick S. Irving, Samuel W. Olson, and Kevin M. Weeks. Rna base-pairing complexity in living cells visualized by correlated chemical probing. *Proceedings of the National Academy of Sciences of the United States of America*, 116:24574–24582, 11 2019.
- [17] Beth L. Nicholson and K. Andrew White. Functional long-range rna–rna interactions in positive-strand rna viruses. *Nature Reviews Microbiology*, 12:493–504, 6 2014.
- [18] David H. Mathews, Matthew D. Disney, Jessica L. Childs, Susan J. Schroeder, Michael Zuker, and Douglas H. Turner. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of rna secondary structure. *Proceedings of the National Academy of Sciences*, 101:7287–7292, 5 2004.
- [19] Mei Chi Su, Chung Te Chang, Chiu Hui Chu, Ching Hsiu Tsai, and Kung Yao Chang. An atypical rna pseudoknot stimulator and an upstream attenuation signal for -1 ribosomal frameshifting of sars coronavirus. *Nucleic Acids Research*, 2005.
- [20] Nuala A. O’Leary, Mathew W. Wright, J. Rodney Brister, Stacy Ciufu, Diana Haddad, Rich McVeigh, Bhanu Rajput, Barbara Robbertse, Brian Smith-White, Danso Ako-Adjei, Alexander Astashyn, Azat Badretdin, Yiming Bao, Olga Blinkova, Vyacheslav Brover, Vyacheslav Chetvernin, Jinna Choi, Eric Cox, Olga Ermolaeva, Catherine M. Farrell, Tamara Goldfarb, Tripti Gupta, Daniel Haft, Eneida Hatcher, Wratkan Hlavina, Vinita S. Joardar, Vamsi K. Kodali, Wenjun Li, Donna Maglott, Patrick Master-son, Kelly M. McGarvey, Michael R. Murphy, Kathleen O’Neill, Shashikant Pujar, Sanjida H. Rangwala, Daniel Rausch, Lillian D. Riddick, Conrad Schoch, Andrei Shkeda, Susan S. Storz, Hanzhen Sun, Francoise Thibaud-Nissen, Igor Tolstoy, Raymond E. Tully, Anjana R. Vatsan, Craig Wallin, David Webb, Wendy Wu, Melissa J. Landrum, Avi Kimchi, Tatiana Tatusova, Michael DiCuccio, Paul Kitts, Terence D. Murphy, Kim D. Pruitt, O’Leary NA, Wright MW, Brister JR, Ciufu S, Haddad Haft D, McVeigh R, Robbertse Rajput B, Robbertse Rajput B, Smith-White B, Ako-Adjei D, Astashyn A, Badret-din A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell

- CM, Goldfarb Gupta T, Goldfarb Gupta T, Haddad Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, and Pruitt KD. Reference sequence (refseq) database at ncbi: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44:D733–D745, 2016.
- [21] Pablo Cordero, Wipapat Kladwang, Christopher C. Vanlang, and Rhiju Das. Quantitative dimethyl sulfate mapping for automated rna secondary structure inference. *Biochemistry*, 51:7037–7039, 9 2012.
- [22] Chi Zhu, Justin Y. Lee, Jia Z. Woo, Lei Xu, Xammy Nguyenla, Livia H. Yamashiro, Fei Ji, Scott B. Biering, Erik Van Dis, Federico Gonzalez, Douglas Fox, Eddie Wehri, Arjun Rustagi, Benjamin A. Pinsky, Julia Schaletzky, Catherine A. Blish, Charles Chiu, Eva Harris, Ruslan I. Sadreyev, Sarah Stanley, Sakari Kauppinen, Silvi Rouskin, and Anders M. Näär. An intranasal aso therapeutic targeting sars-cov-2. *Nature Communications*, 13:4503, 12 2022.
- [23] Clarence Y. Cheng, Wipapat Kladwang, Joseph D. Yesselman, and Rhiju Das. Rna structure inference through chemical mapping after accidental or intentional mutations. *Proceedings of the National Academy of Sciences of the United States of America*, 114:9876–9881, 9 2017.
- [24] Jennifer K. Barry and W. Allen Miller. A -1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral rna. *Proceedings of the National Academy of Sciences of the United States of America*, 99:11133–11138, 8 2002.
- [25] Yuri Tajima, Hiro oki Iwakawa, Masanori Kaido, Kazuyuki Mise, and Tetsuro Okuno. A long-distance rna–rna interaction plays an important role in programmed -1 ribosomal frameshifting in the translation of p88 replicase protein of red clover necrotic mosaic virus. *Virology*, 417:169–178, 8 2011.
- [26] Pramod R. Bhatt, Alain Scaiola, Gary Loughran, Marc Leibundgut, Annika Kratzel, Romane Meurs, René Dreos, Kate M. O'Connor, Angus McMillan, Jeffrey W. Bode, Volker Thiel, David Gatfield, John F. Atkins, and Nenad Ban. Structural basis of ribosomal frameshifting during translation of the sars-cov-2 rna genome. *Science*, 372:1306–1313, 5 2021.
- [27] Hafeez S. Haniff, Yuquan Tong, Xiaohui Liu, Jonathan L. Chen, Blessy M. Suresh, Ryan J. Andrews, Jake M. Peterson, Collin A. O'Leary, Raphael I. Benhamou, Walter N. Moss, and Matthew D. Disney. Targeting the sars-cov-2 rna genome with small molecule binders and ribonuclease targeting chimera (ribotac) degraders. *ACS Central Science*, 6:1713–1721, 2020.

- [28] Jamie A. Kelly, Alexandra N. Olson, Krishna Neupane, Sneha Munshi, Josue San Emeterio, Lois Pollack, Michael T. Woodside, and Jonathan D. Dinman. Structural and functional conservation of the programmed -1 ribosomal frameshift signal of sars coronavirus 2 (sars-cov-2). *Journal of Biological Chemistry*, 295:10741–10748, 7 2020.
- [29] Ewan P. Plant, Rasa Rakauskaitė, Deborah R. Taylor, and Jonathan D. Dinman. Achieving a golden mean: Mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *Journal of Virology*, 84:4330–4340, 2010.
- [30] Yu Sun, Laura Abriola, Rachel O. Niederer, Savannah F. Pedersen, Mia M. Alfajaro, Valter Silva Monteiro, Craig B. Wilen, Ya-Chi Ho, Wendy V. Gilbert, Yulia V. Surovtseva, Brett D. Lindenbach, and Junjie U. Guo. Restriction of sars-cov-2 replication by targeting programmed -1 ribosomal frameshifting. *Proceedings of the National Academy of Sciences of the United States of America*, 118:e2023051118, 6 2021.
- [31] Kaiming Zhang, Ivan N. Zheludev, Rachel J. Hagey, Raphael Haslecker, Yixuan J. Hou, Rachael Kretsch, Grigore D. Pintilie, Ramya Rangan, Wipapat Kladwang, Shanshan Li, Marie Teng Pei Wu, Edward A. Pham, Claire Bernardin-Souibgui, Ralph S. Baric, Timothy P. Sheahan, Victoria D'Souza, Jeffrey S. Glenn, Wah Chiu, and Rhiju Das. Cryo-em and antisense targeting of the 28-kda frameshift stimulation element from the sars-cov-2 rna genome. *Nature Structural & Molecular Biology*, 28:747–754, 8 2021.
- [32] Georg Wolff, Charlotte E. Melia, Eric J. Snijder, and Montserrat Bárcena. Double-membrane vesicles as platforms for viral replication. *Trends in Microbiology*, 28:1022–1033, 12 2020.
- [33] Michael Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6, 2021.
- [34] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

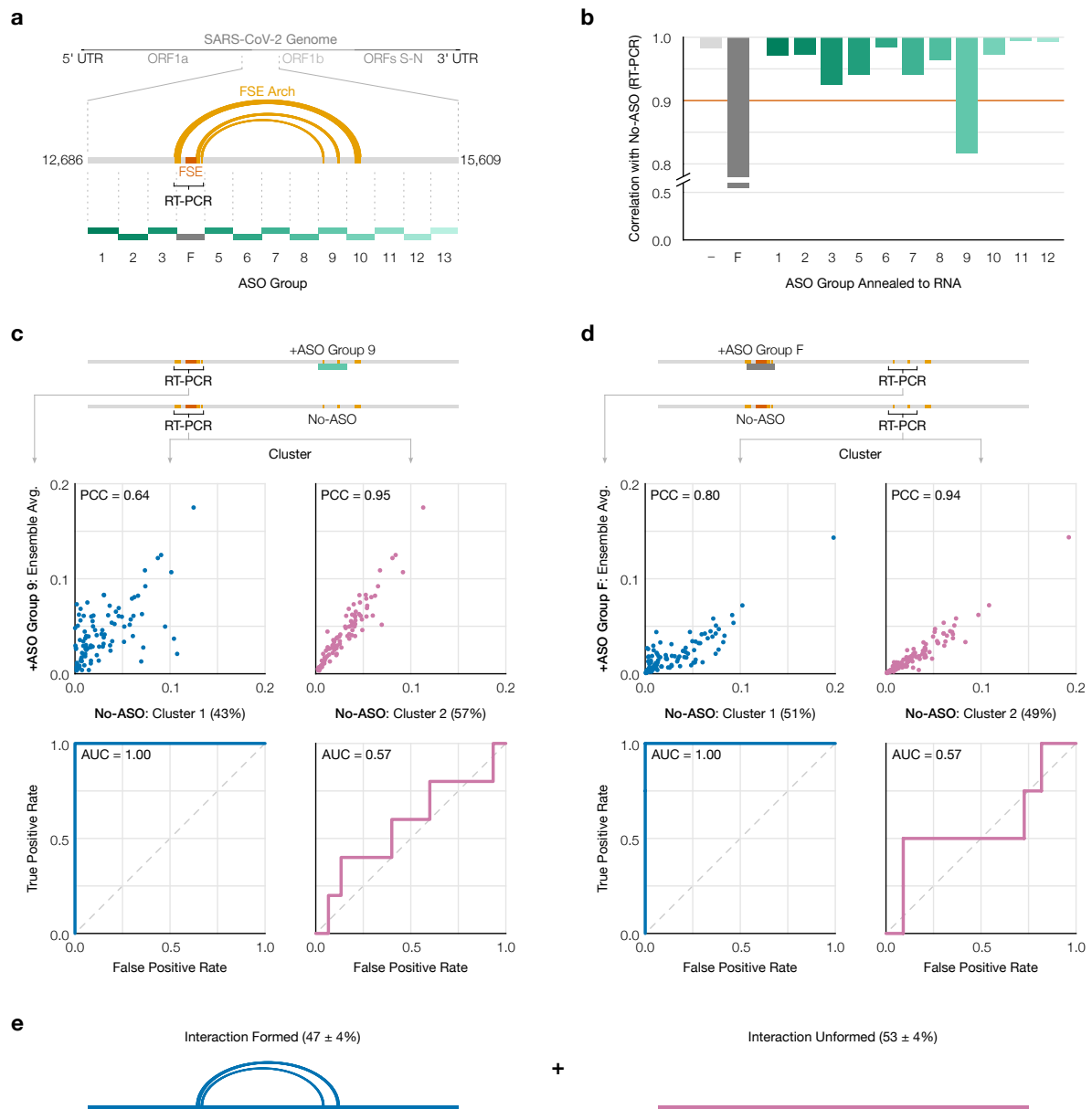


Figure 3:

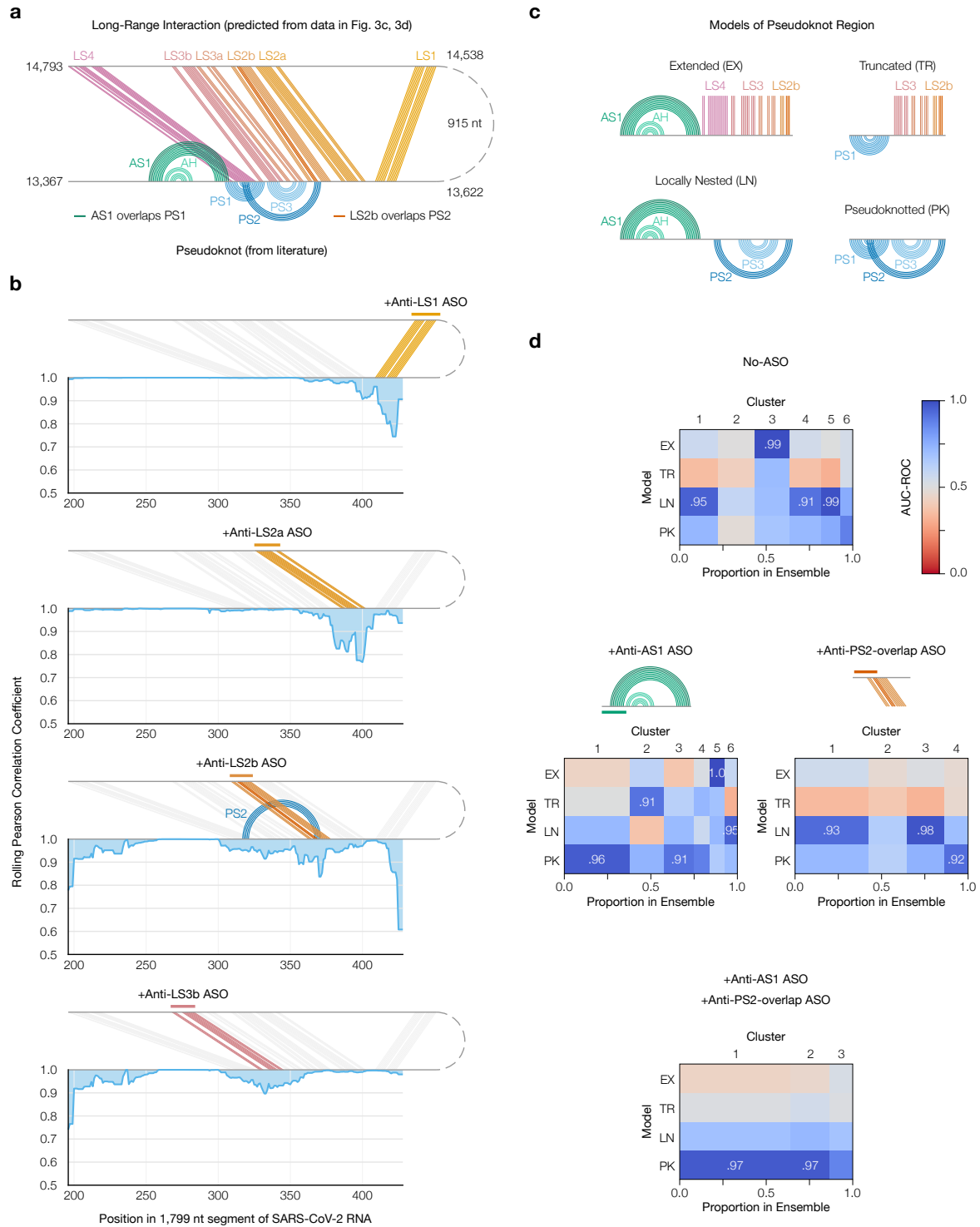


Figure 4:

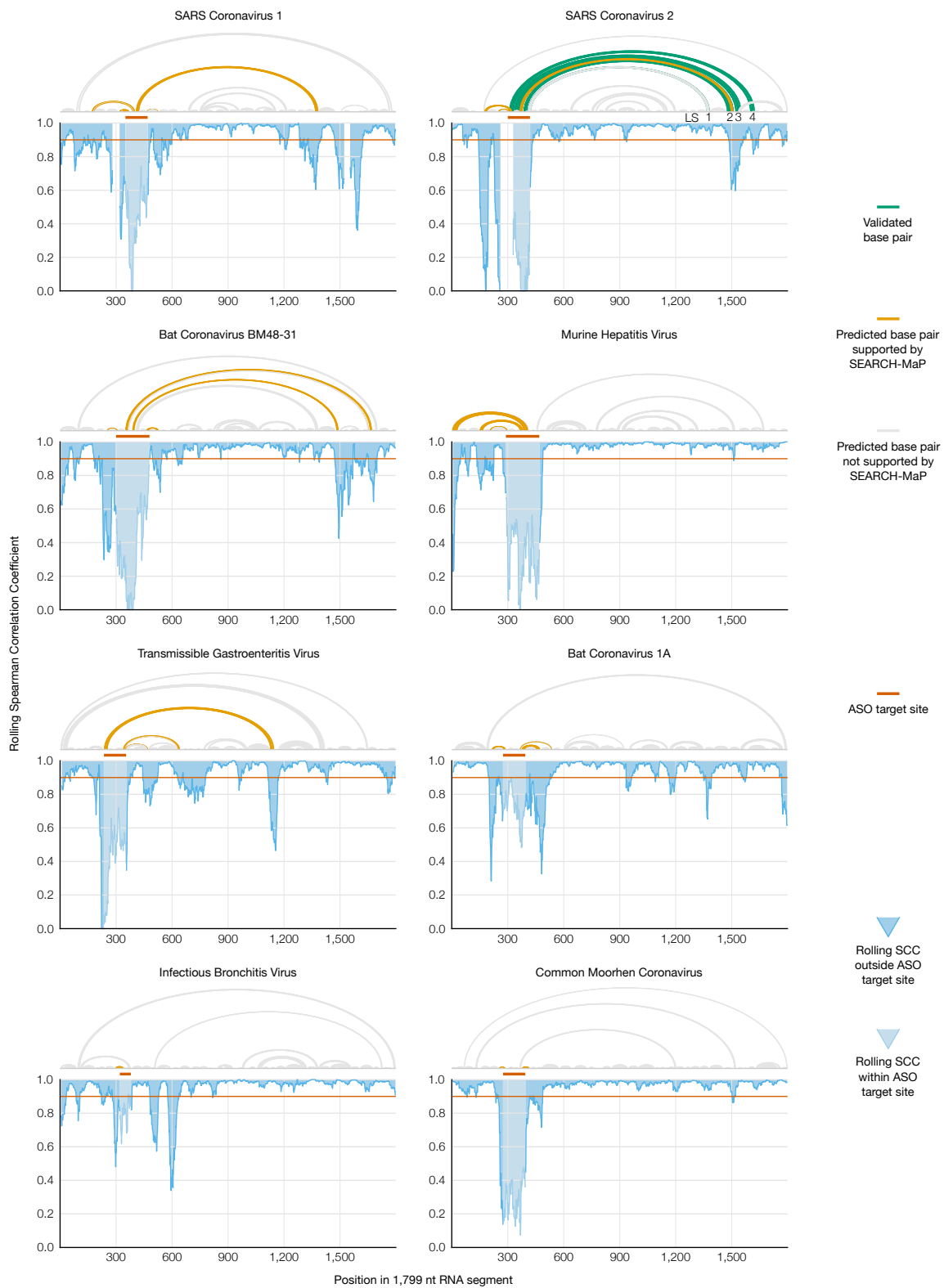


Figure 5: