Frameshift stimulating elements of multiple coronaviruses form 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 **?** contained sixty-two 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 (including SARS-CoV-2) for further study – at least one from each genus (SFIG). 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 moorhen 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 though chemical probing. For each of the ten coronaviruses that passed the computational screen, we *in vitro* transcribed and performed DMS-MaPseq **?** on both a 239 nt segment comprising the FSE and minimal flanking sequences and a 1.8 kb 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 1.8 kb segments (SFIG), suggesting long-range interactions between the FSE and another element within the 1.8 kb segment.

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

To determine whether the FSE interacts with another RNA element – and if so, which – in each coronavirus, we performed SEARCH-MaP on the 1.8 kb RNA segment using ASOs targeting the bases that changed the most between the 239 nt and 1.8 kb segments. Using SEISMIC-RNA, we computed the mutational profiles with and without ASOs and calculated the Spearman correlation coefficient (SCC) between them via a sliding window (Figure 1). In every coronavirus, the ASO target site (light green) showed a large dip in SCC, confirming that the ASOs bound and altered the structure.

The long-range interaction in SARS-CoV-2 was suggested to comprise three stems **?**. SEARCH-MaP found two (dips in SCC around positions 1,500 and 1,615); the third stem (around position 1,375) may have been missed because it is not only the shortest (15 nt) but also has just one unpaired base on each side. Similar long-range interactions appeared in the other SARS-related viruses: SARS-CoV-1 and bat coronavirus BM48-31. Like SARS-

2

CoV-2, both had a dip in SCC near position 1,500 and another dip downstream, between positions 1,575 and 1,675. The SCC also dipped around position 1,375 for SARS-CoV-1, matching the location of the missed stem in SARS-CoV-2. Therefore, this three-stemmed long-range interaction involving the FSE appears to be conserved among three strains of SARS-related viruses, suggesting it is functional.

The SCC dipped prominently below 0.9 at least 200 nt from the ASO binding site in all other coronaviruses except common moorhen coronavirus.

To predict the specific base pairs, we modeled secondary structures using RNAstructure Fold **?** with the no-ASO mutational profiles as DMS constraints **?**. Ideally, Fold would predict the ASO target site pairs with all – and only – dips in SCC outside the target site. We classified predicted base pairs with the ASO target site as true positives (blue) if the base outside the ASO target site had an SCC below 0.9, otherwise false positives (orange).
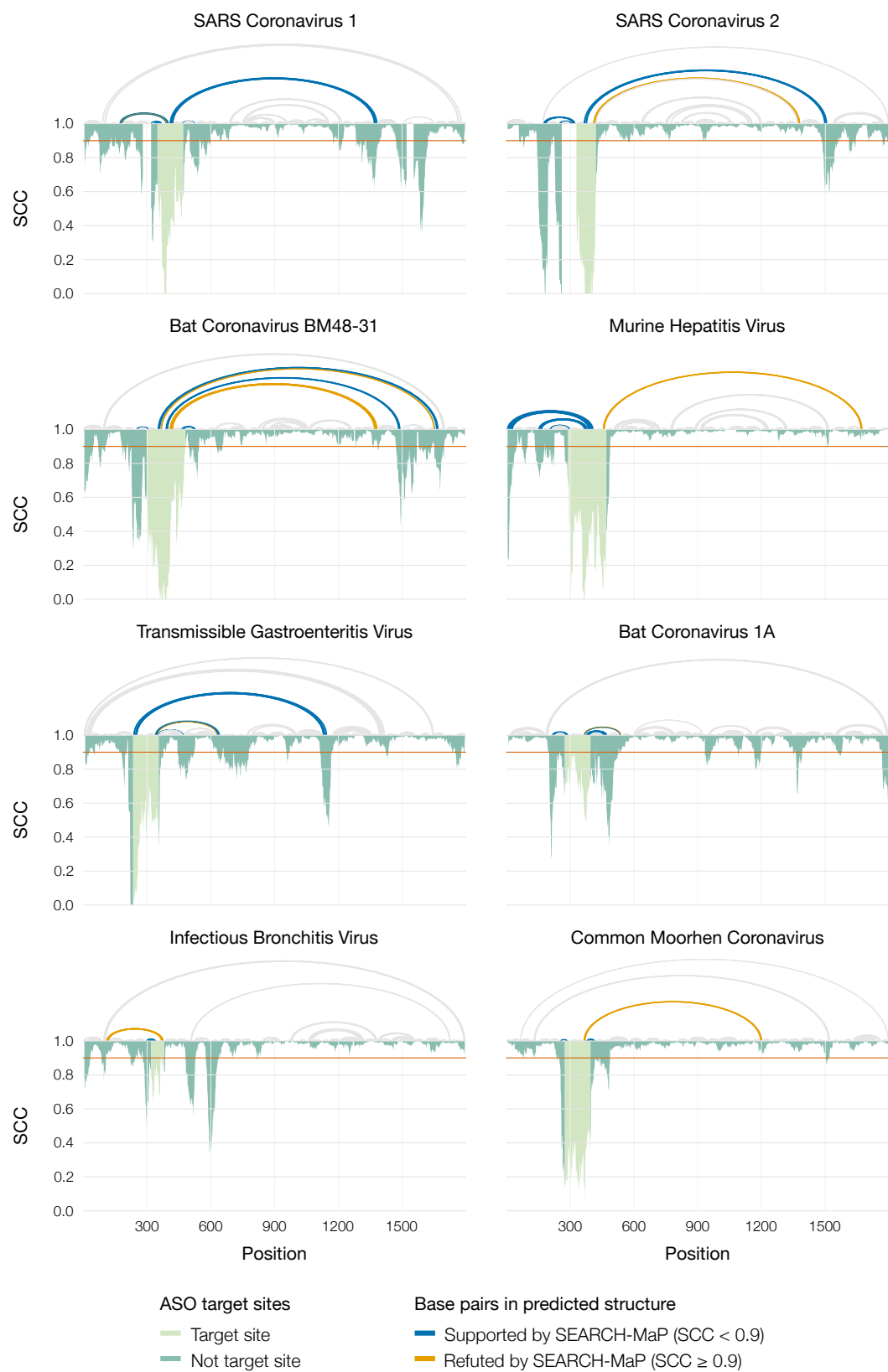
Figure 1: