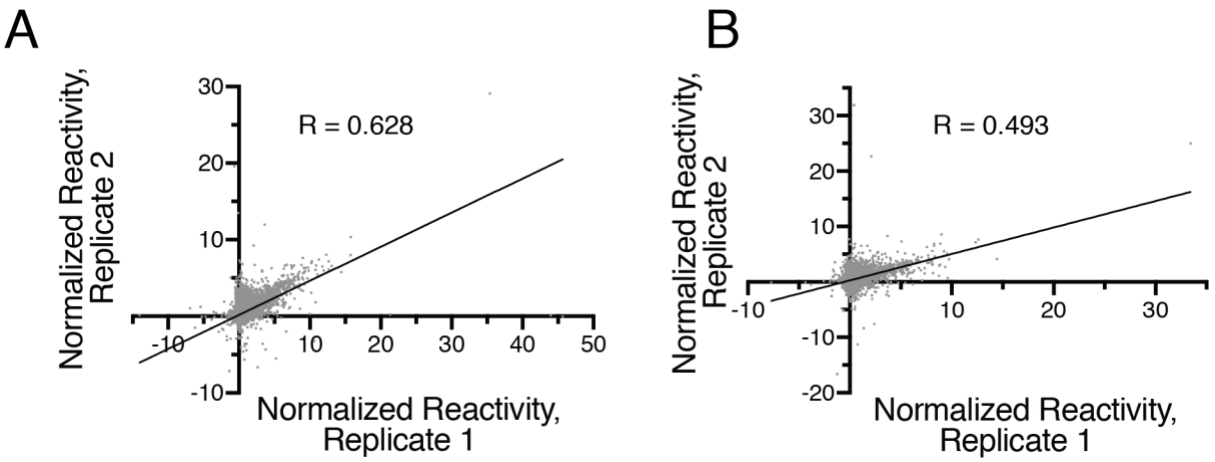
Molecular Cell, Volume *81*

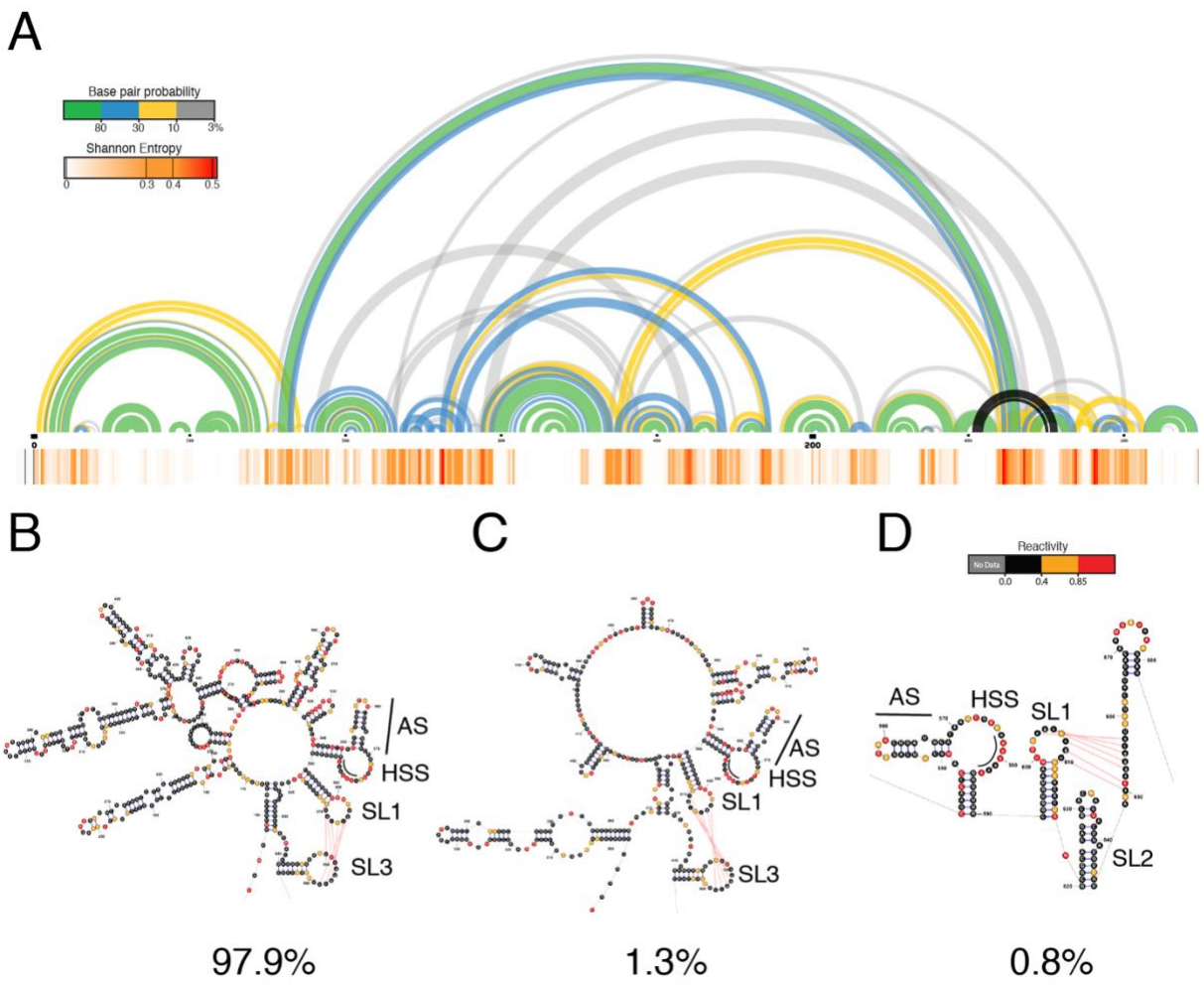
Supplemental Information

Comprehensive *in vivo* secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms

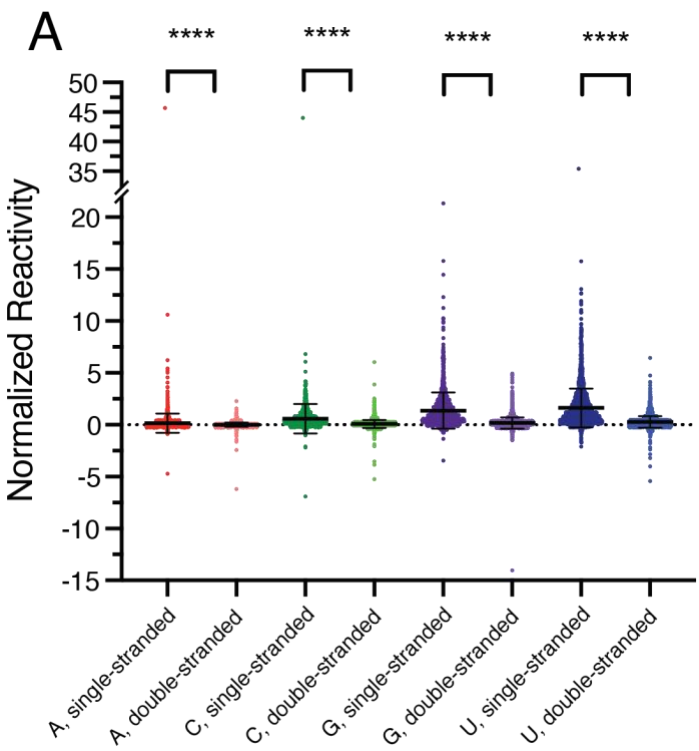
Nicholas C. Huston, Han Wan, Madison S. Strine, Rafael de Cesaris Araujo Tavares, Craig B. Wilen, and Anna Marie Pyle



Supplementary Figure 1. Analysis of correlation of normalized SHAPE reactivity reveals good agreement between biological replicates across Orf1ab, but not the subgenomic RNA region. Related to Figure 1. **A), B)** Correlation plot of normalized SHAPE reactivities from two biological replicates determined for the Orf1ab region or subgenomic RNA region, respectively. Lines represent linear regressions fit to the data. Pearson's correlation for each dataset is shown.



Supplementary Figure 2. Ensemble analysis of the region containing the SARS-CoV-2 PRF confirms that the canonical three-stem pseudoknot structure represents a minority conformation. Related to Figure 3. **A)** Base-pair probabilities calculated with the partition function implemented in the SuperFold pipeline for the region of the SARS-CoV-2 genome containing the PRF. Arcs corresponding to individual base-pairs are colored by base-pairing probability (green = >80%; blue = 80% > base-pair probability > 30%; yellow = 30% > base-pair probability > 10%; grey = 10% > base-pair probability > 3%). A black arc indicates the PRF pseudoknot. Shannon Entropies for individual nucleotides, represented as a heat map, are shown underneath the corresponding nucleotides in the arc plots. **B), C), D)** Structural clusters that comprise the conformational ensemble representing the SARS-CoV-2 PRF. Relative abundances of each cluster are shown as percentages. AS = Attenuator Stem; HSS = Heptanucleotide Slippery Sequence; SL1 = Stem Loop 1; SL2 = Stem Loop 2; SL3 = Stem Loop 3; Red lines indicate pseudoknot interaction.



Supplementary Figure 3. Reactivities separated by nucleotide identity and binned by stranded-ness reveals strong agreement between experimentally determined reactivities and the resulting structure prediction for each nucleotide. Related to Figure 4**. A)** Normalized SHAPE reactivity determined by ShapeMapper separated by nucleotide identity and binned by stranded-ness as determined in our consensus structure model. ****p<0.0001 by equal variance unpaired student t test.

Table S1. Well-determined region in Orf1ab region, related to **Figure 4**.

Region	Window	Start	End	Size	Protein Domain	Median Shannon Entropy
1	1	1	622	622	5'UTR, Nsp1	7.69E-05
2	1	944	1026	83	Nsp2	1.75E-06
3	1	1478	1572	95	Nsp2	2.62E-02
4	1	1968	2188	221	Nsp2	5.07E-04
5	1	2682	2800	119	Nsp2, Nsp3	9.19E-03
6	1	3416	3597	182	Nsp3	5.24E-05
7	1	4169	4232	64	Nsp3	3.01E-06
8	1	4471	4713	243	Nsp3	4.82E-04
9	1	4791	5162	372	Nsp3	4.55E-04
10	1	5693	6013	321	Nsp3	2.07E-03
11	1	6116	6549	434	Nsp3	4.24E-04
12	1	6786	6887	102	Nsp3	4.88E-03
13	1	7025	7072	48	Nsp3	1.06E-03
14	2	7413	7518	106	Nsp3	1.55E-04
15	2	7717	8230	514	Nsp3	1.29E-03
16	2	8350	8512	163	Nsp3	4.90E-04
17	2	8702	8789	88	Nsp4	6.45E-04
18	2	9295	9398	104	Nsp4	2.76E-02
19	2	9612	9886	275	Nsp4	1.70E-02
20	2	10129	10312	184	Nsp5	3.00E-04
21	2	10630	10687	58	Nsp5	3.53E-04
22	2	10798	11039	242	Nsp5, Nsp6	7.67E-04
23	2	11221	11470	250	Nsp6	2.34E-03
24	2	11552	11908	357	Nsp6, Nsp7	2.09E-03
25	2	12230	12686	457	Nsp8, Nsp9	1.86E-03
26	2	12895	13030	136	Nsp9, Nsp10	1.55E-02
27	2	13594	13920	327	Nsp12	7.39E-04
28	2	13993	14230	238	Nsp12	1.19E-02
29	3	14444	14532	89	Nsp12	4.27E-04
30	3	14557	14641	85	Nsp12	1.05E-05
31	3	14973	15136	164	Nsp12	9.07E-04
32	3	15510	15608	99	Nsp12	1.59E-04
33	3	15767	16005	239	Nsp12	4.29E-04
34	3	16114	16260	147	Nsp12, 13	2.69E-03
35	3	17580	17677	98	Nsp13	1.64E-04
36	3	17854	17938	85	Nsp13	6.09E-04
37	3	19373	19550	178	Nsp14	1.52E-02
38	3	19665	19735	71	Nsp15	2.32E-05
39	3	20248	20408	161	Nsp15	7.72E-03
40	3	20668	20792	125	Nsp16	4.91E-06

Table S2. Gene-specific RT primers, related to STAR Methods.

Primer Name	**Sequence**
RT_SC2_Amplicon_1	TTAGTCAAATTCTCAGTGC
RT_SC2_Amplicon_2	TTTGTTGACTATCATCATC
RT_SC2_Amplicon_3	AAACATAAAATGTTTTACC
RT_SC2_Amplicon_4	AATTAGACATTAAAACACC
RT_SC2_Amplicon_5	TACCAACTGCACTAAAAAC
RT_SC2_Amplicon_6	TATCTAAAACGGCAATTCC
RT_SC2_Amplicon_7	AAGCAGTTTGTGTAGTACC
RT_SC2_Amplicon_8	TTAGTAAGTGCAGCTACTG
RT_SC2_Amplicon_9	TAACATTATCGCTACCAAC
RT_SC2_Amplicon_10	TAACTCTGGAAAAATCTGT
RT_SC2_Amplicon_11	AACCACCTAACTGACTATG
RT_SC2_Amplicon_12	TAATACCTATTGGCAAATC
RT_SC2_Amplicon_13	AATCATTTCATCTGTGAGC
RT_SC2_Amplicon_14	TAACATGTTCAACACCAGT
RT_SC2_Amplicon_15	ATGTTGAGTACATGACTGT
RT_SC2_Amplicon_16	TTTTTTTTTGTCATTCTCC

Table S3. Gene-specific PCR Primers, related to STAR Methods.

Primer Name	Sequence
F_PCR_SC2_Amplicon_1	ATTAAAGGTTTATACCTTCCCAG
F_PCR_SC2_Amplicon_2	CTCATGAAGTGTGATCATTGTGG
F_PCR_SC2_Amplicon_3	GATTACCAAGGTAAACCTTTGGA
F_PCR_SC2_Amplicon_4	TATGGACAACAGTTTGGTCCAAC
F_PCR_SC2_Amplicon_5	ATAAATATTATAATTTGGTTTTTACTATTA
F_PCR_SC2_Amplicon_6	AAGAGAAGTGGGTTTTGTCG
F_PCR_SC2_Amplicon_7	TGTGGCTATGAAGTACAATTATG
F_PCR_SC2_Amplicon_8	TGTAACAGCTTTAAGGGCCAATT
F_PCR_SC2_Amplicon_9	TAAGGAATTACTTGTGTATGCTG
F_PCR_SC2_Amplicon_10	TTATTGTAAATCACATAAACCAC
F_PCR_SC2_Amplicon_11	ACAGCTAGGTTTTTCTACAGGTG
F_PCR_SC2_Amplicon_12	TTAGAATTAGCTATGGATGAATT
F_PCR_SC2_Amplicon_13	TATATTCTAAGCACACGCCTATT
F_PCR_SC2_Amplicon_14	GTGATTGCCTTGGTGATATT
F_PCR_SC2_Amplicon_15	TCTGGAGTAAAAGACTGTGTTGT
F_PCR_SC2_Amplicon_16	GTCACGCCTAAACGAACATG
R_PCR_SC2_Amplicon_1	AGTCAAATTCTCAGTGCCACAA
R_PCR_SC2_Amplicon_2	TGTTGACTATCATCATCTAACCA
R_PCR_SC2_Amplicon_3	ACATAAAATGTTTTACCTTCATG
R_PCR_SC2_Amplicon_4	TTAGACATTAAAACACCTAAAGC
R_PCR_SC2_Amplicon_5	CCAACTGCACTAAAAACTCTAGG
R_PCR_SC2_Amplicon_6	CTAAAACGGCAATTCCAGTT
R_PCR_SC2_Amplicon_7	CAGTTTGTGTAGTACCGGCA
R_PCR_SC2_Amplicon_8	GTAAGTGCAGCTACTGAAAAGCA
R_PCR_SC2_Amplicon_9	CATTATCGCTACCAACACATGTA
R_PCR_SC2_Amplicon_10	CTCTGGAAAAATCTGTATTATTAGG
R_PCR_SC2_Amplicon_11	CACCTAACTGACTATGACTAAAA
R_PCR_SC2_Amplicon_12	TACCTATTGGCAAATCTACCAAT
R_PCR_SC2_Amplicon_13	CATTTCATCTGTGAGCAAAG
R_PCR_SC2_Amplicon_14	CATGTTCAACACCAGTGTCTGTA
R_PCR_SC2_Amplicon_15	TTGAGTACATGACTGTAAACTACAT
R_PCR_SC2_Amplicon_16	TTTTTTGTCATTCTCCTAAGAAG

Table S4. LNAs used in this study (LNA bases are indicated with a "+" on the left), related to STAR Methods.

Region	LNA	LNA content	%GC	RNA Tm
PRF, SL1	+A+C+GG+GC+TGC+ACT+TA+CA+C+C+G	57.9	63.2	90
PRF, SL2	+C+A+GTAC+TAG+TG+CC+TG+TGC+C+G+C	52.4	61.9	89
Region 15	+A+C+AAA+CCC+TTG+CCG+AG+CT+G+C+T	52.4	57.1	91
Region 15, Control	+G+T+TT+TCA+ACT+TTG+TTA+TAG+G+T+G	50.0	31.8	86
Region22	+G+T+CTA+ACA+ACA+TCA+AA+AG+G+T+G	52.4	38.1	86
Region 22, Control	+G+C+TAC+AG+TGG+CAA+GAG+AA+G+G+T	52.4	52.4	86
Scrambled LNA	+G+C+GGC+ACG+TTG+CG+AGT+A+C+T	52.6	63.2	N/A