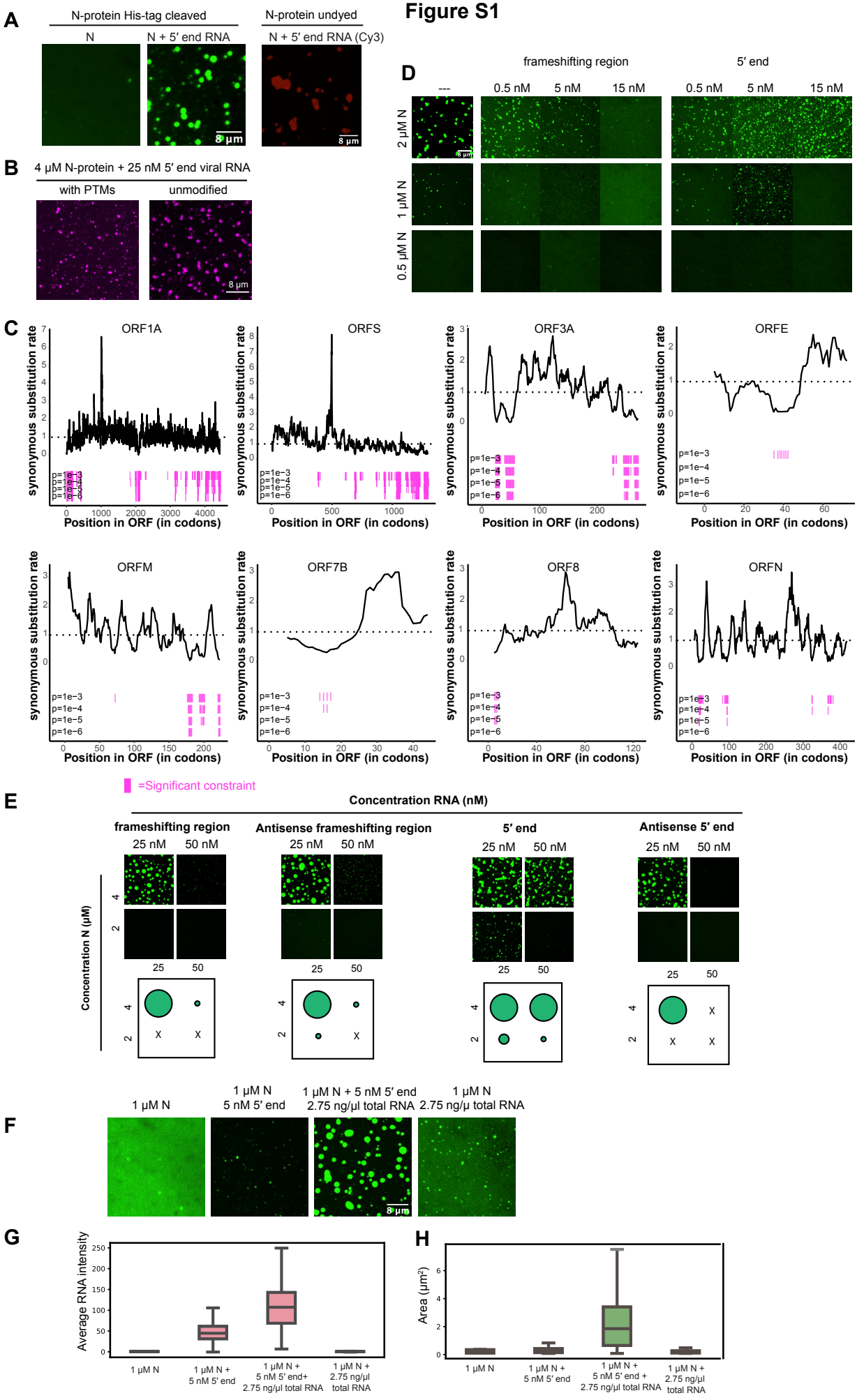
Molecular Cell, Volume *80*

Supplemental Information

Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid

Christiane Iserman, Christine A. Roden, Mark A. Boerneke, Rachel S.G. Sealfon, Grace A. McLaughlin, Irwin Jungreis, Ethan J. Fritch, Yixuan J. Hou, Joanne Ekena, Chase A. Weidmann, Chandra L. Theesfeld, Manolis Kellis, Olga G. Troyanskaya, Ralph S. Baric, Timothy P. Sheahan, Kevin M. Weeks, and Amy S. Gladfelter

Figure S1

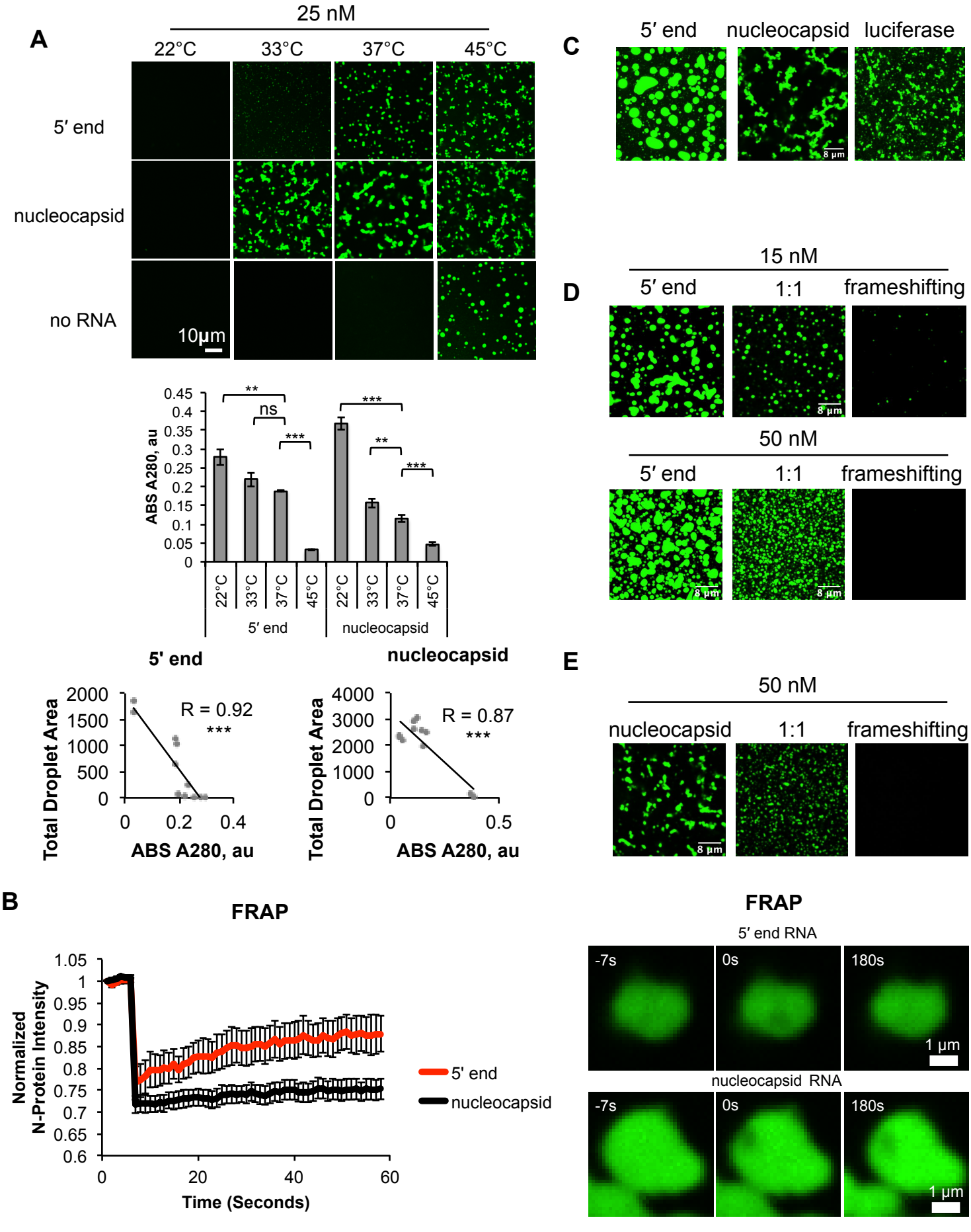

A
N-protein His-tag cleaved
N
N + 5′ end RNA
N-protein undyed
N + 5′ end RNA (Cy3)
8 µm
B
4 µM N-protein + 25 nM 5′ end viral RNA
with PTMs
unmodified
8 µm
D
frameshifting region
5′ end

0.5 nM
5 nM
15 nM
2 µM N
1 µM N
0.5 µM N
C
ORF1A
ORFS
ORF3A
ORFE
ORFM
ORF7B
ORF8
ORFN
synonymous substitution rate
Position in ORF (in codons)
p=1e−3
p=1e−4
p=1e−5
p=1e−6
=Significant constraint
E
Concentration RNA (nM)
frameshifting region
Antisense frameshifting region
5′ end
Antisense 5′ end
25 nM
50 nM
Concentration N (µM)
F
1 µM N
1 µM N
5 nM 5′ end
1 µM N + 5 nM 5′ end
2.75 ng/µl total RNA
1 µM N
2.75 ng/µ total RNA
8 µm
G
Average RNA intensity
1 µM N
1 µM N +
5 nM 5′ end
1 µM N +
5 nM 5′ end+
2.75 ng/µl total RNA
1 µM N +
2.75 ng/µl
total RNA
H
Area (µm²)
1 µM N
1 µM N +
5 nM 5′ end
1 µM N +
5 nM 5′ end +
2.75 ng/µl total RNA
1 µM N +
2.75 ng/µl
total RNA

Supplemental Figure 1: **N-protein LLPS is concentration and RNA species dependent. Related to Figure 1.**

A) Left-most panels, 5′ end RNA drives LLPS of HIS-tag cleaved N-protein (green). Right panel, 5′ end (red) drives LLPS of undyed protein**. B)** Unmodified N-protein and N-protein containing post-translational modifications (PTMs) phase separate with viral RNA (Cy3, magenta). **C)** Multiple regions across SARS-CoV-2 genome (x axis) show elevated synonymous constraint; y axis less than 1. Magenta bars indicate regions with significant synonymous constraints. Row 1 $p<0.0001$, Row 2 $p< 0.00001$, Row 3 $p<0.00001$, Row 4 $p< 0.000001$. **D)** Related to **Figure 1G**. Frameshifting-region RNA promotes dissolution of N-protein condensates (green) whereas 5′ end RNA drives LLPS relative to N-protein alone (left-most panels). **E)** GC content does not dictate LLPS with N-protein. Frameshifting region, antisense frameshifting region, and antisense 5′ end RNAs undergo LLPS only at a narrow concentration range. Sense 5′ end RNA drives LLPS at multiple protein and RNA concentrations. **F)** Non-viral lung total cell RNA enhances N-protein (green) LLPS with 5′ end RNA. **G)** Quantification of average droplet RNA intensity from **(F)**. **H)** Quantification of droplet area from **(F)**. Scale bar, 8 μm unless otherwise noted.

Figure S2

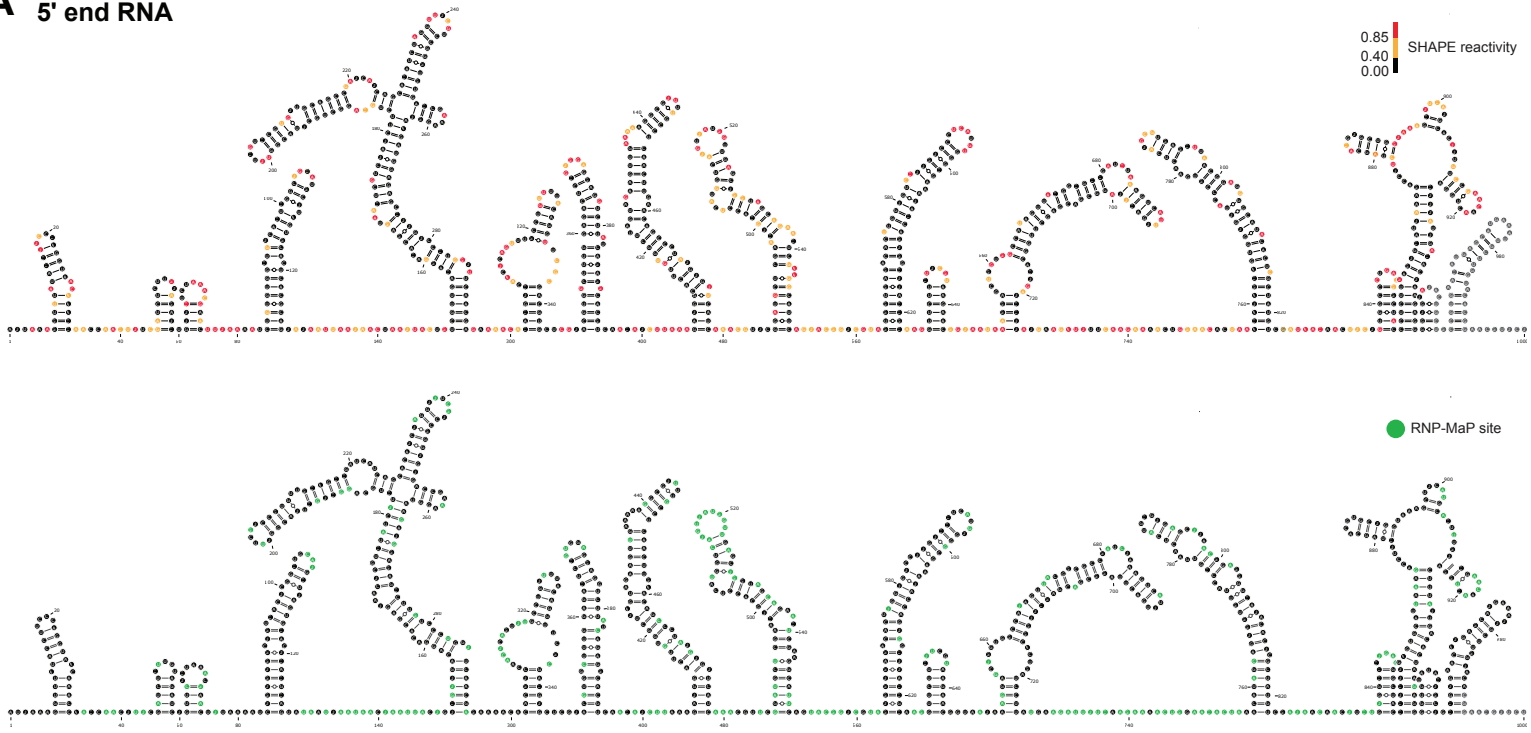

A
25 nM
22°C
33°C
37°C
45°C
5′ end
nucleocapsid
no RNA
10μm
ABS A280, au
ns
5′ end
nucleocapsid
5' end
nucleocapsid
Total Droplet Area
R = 0.92
R = 0.87
ABS A280, au
B
FRAP
Normalized N-Protein Intensity
Time (Seconds)
5' end
nucleocapsid
C
5′ end
nucleocapsid
luciferase
8 μm
D
15 nM
5′ end
1:1
frameshifting
50 nM
E
50 nM
nucleocapsid
1:1
frameshifting
FRAP
5′ end RNA
-7s
0s
180s
1 μm
nucleocapsid RNA

Supplemental Figure 2: N-protein LLPS is temperature dependent with RNA sequence encoded material properties. Related to Figure 2.

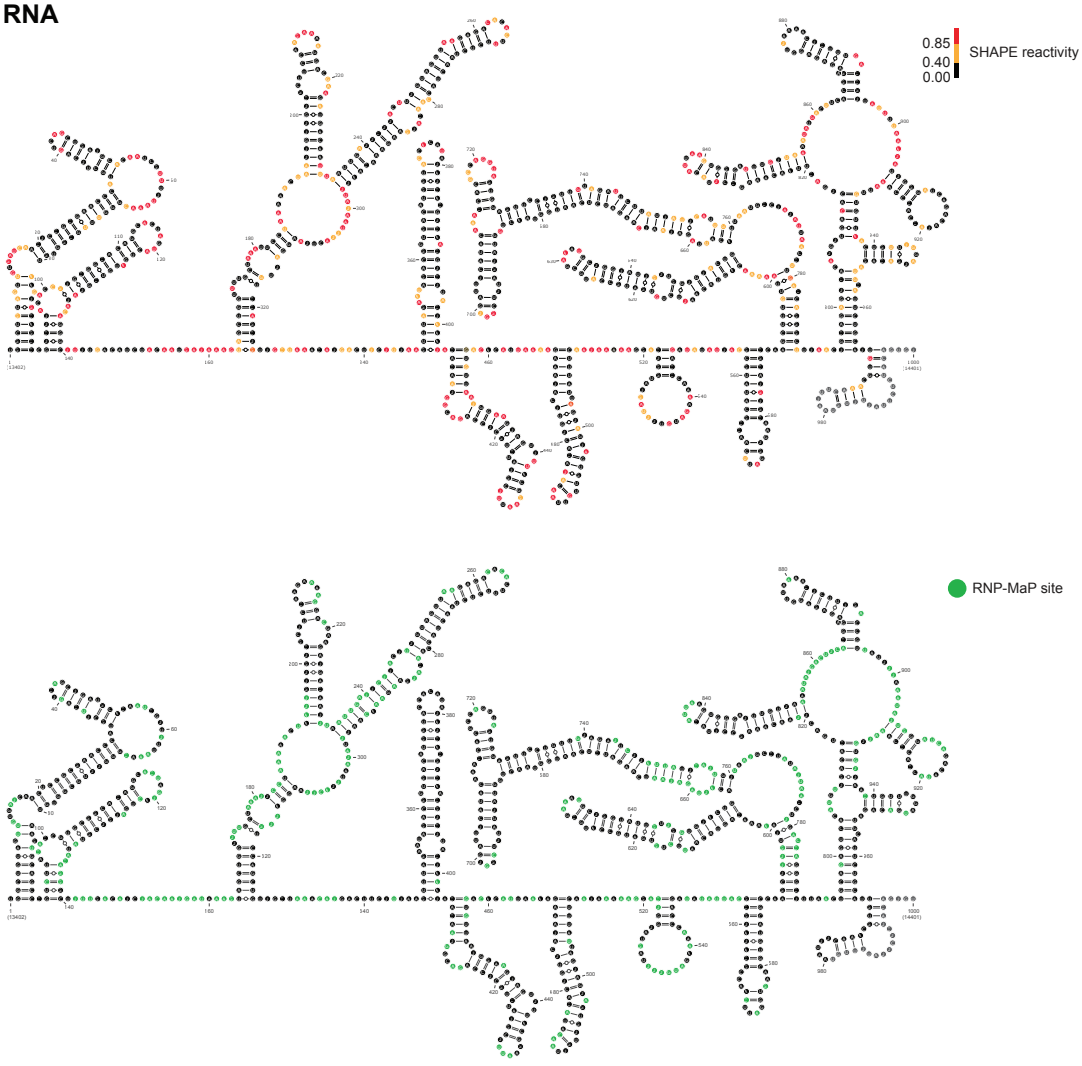
A) Addition of RNA to N-protein lowers temperature dependent LLPS independent of RNA sequence. Nucleocapsid RNA, and 5′ end RNA readily phase separate with purified N-protein (green) at 45, 37, and 33°C, but not at 22°C. N-protein without RNA only phase separates at 45°C. Addition of RNA (5′ end or nucleocapsid) and temperature alters the average A280 absorbance in the soluble phase. Bars indicate the average absorbance for 3 replicates with error bars indicating standard deviation. Average A280 absorbance is significantly ($p<0.001$ Pearson correlation) anti-correlated with surface area which is occupied by droplets. **B)** N-protein droplets made with 5′ end or nucleocapsid RNAs show different recovery of protein signal following photobleaching. Left panel: recovery curves for 8 (5′ end) or 9 (nucleocapsid) droplets. Error Bars represent standard error. Right panel: representative droplets for FRAP quantifications in the left panel. **C)** 5′ end /N-protein (green) droplets 5′ end RNA are with rounder than droplets with nucleocapsid RNA or luciferase RNA. **D)** 5′ end and frameshifting region RNA forms droplets of intermediate phenotypes when mixed together prior to N-protein addition (middle columns 1:1) compared to those of 5′ end (left) or frameshifting region RNA alone (right). **E)** Nucleocapsid RNA and N-protein form droplets of intermediate phenotypes (smaller size, less flocculated) when mixed with frameshifting region RNA. Scale bar, 8 μm unless otherwise noted.

Figure S3

A 5' end RNA



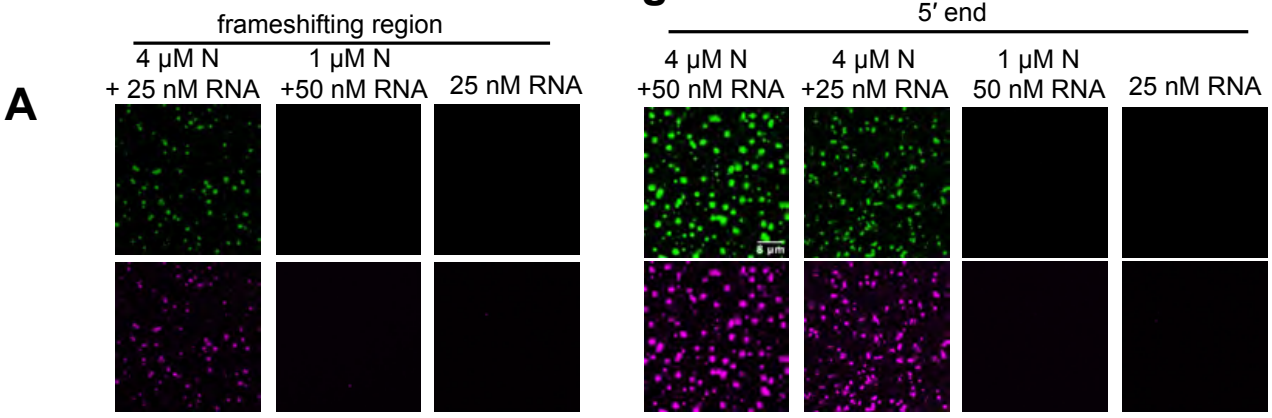
B frameshifting region RNA

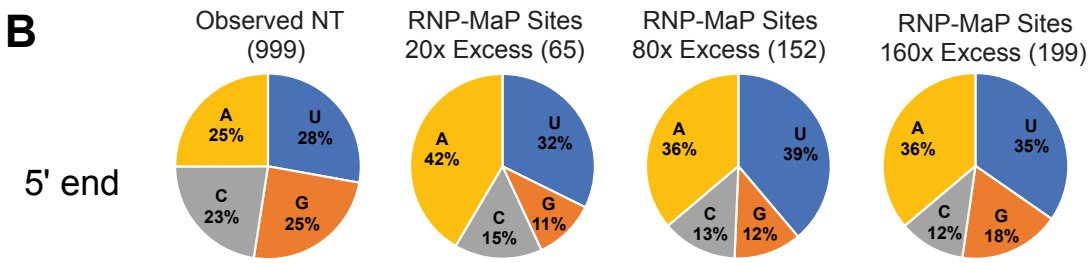


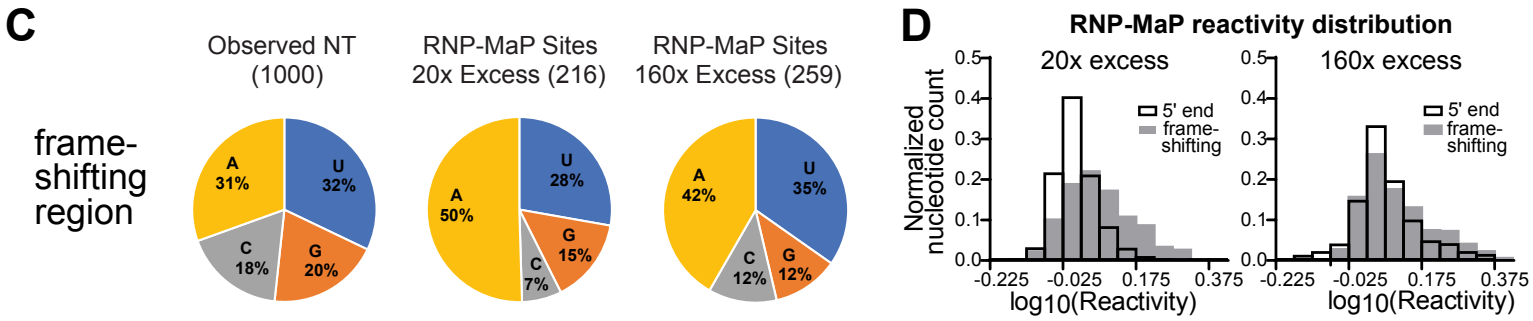
Supplemental Figure 3: High-resolution RNA structures with RNP-MaP reactivity. Related to Figure 3.

SHAPE-directed secondary structure models for **(A)** the 5′ end RNA and **(B)** the frameshifting region RNA showing nucleotide identities and colored by SHAPE reactivity (upper panel, black are low, yellow are intermediate and red are high SHAPE reactivity) or RNP-MaP reactive nucleotides (lower panel, green marked nucleotides). The lower panels of **(A)** and **(B)** are also shown in **Figure 3A** and **B**.

Figure S4


A
frameshifting region
5′ end
4 μM N
+ 25 nM RNA
1 μM N
+50 nM RNA
25 nM RNA
4 μM N
+50 nM RNA
4 μM N
+25 nM RNA
1 μM N
50 nM RNA
25 nM RNA

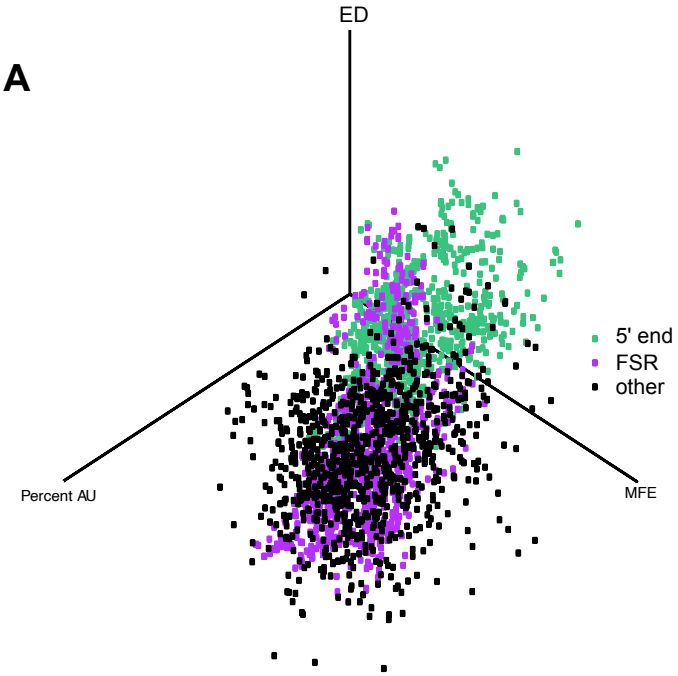

B
5' end
Observed NT
(999)
A 25%
U 28%
C 23%
G 25%
RNP-MaP Sites
20x Excess (65)
A 42%
U 32%
C 15%
G 11%
RNP-MaP Sites
80x Excess (152)
A 36%
U 39%
C 13%
G 12%
RNP-MaP Sites
160x Excess (199)
A 36%
U 35%
C 12%
G 18%

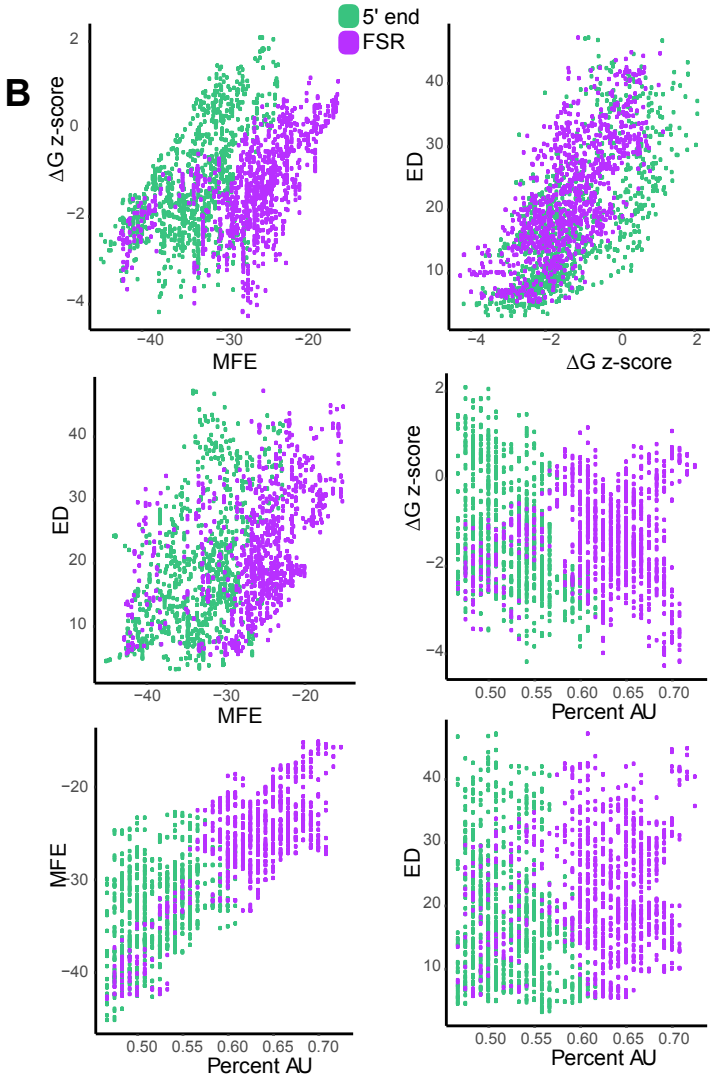

C
frame-
shifting
region
Observed NT
(1000)
A 31%
U 32%
C 18%
G 20%
RNP-MaP Sites
20x Excess (216)
A 50%
U 28%
C 7%
G 15%
RNP-MaP Sites
160x Excess (259)
A 42%
U 35%
C 12%
G 12%
D
RNP-MaP reactivity distribution
20x excess
160x excess
Normalized
nucleotide count
5' end
frame-
shifting
0.5
0.4
0.3
0.2
0.1
0.0
-0.225
-0.025
0.175
0.375
log10(Reactivity)

Supplemental Figure 4: RNP-MaP reveals viral RNA-specific and condition-dependent N-protein binding site characteristics. Related to Figure 3.

A) Representative conditions for RNP-MaP experiments showing diffuse and condensed conditions. Frameshifting region and 5′ end RNAs (magenta) were mixed with N-protein (green) or droplet buffer and imaged prior to processing for RNP-MaP analysis. Scale bar, 8 μm. **B and C)** N-protein–RNA interactions as defined by RNP-MaP sites are enriched in A/U nucleotides (pie charts). The total number of nucleotides (NT) are indicated in parenthesis. **D)** Comparison of RNP-MaP reactivity distributions for 5′ end and frameshifting region reveal the frameshifting region RNA to be more broadly interacting with N-protein (larger distribution of higher reactivity nucleotides) at both 20X and 160X excess protein.

Figure S5





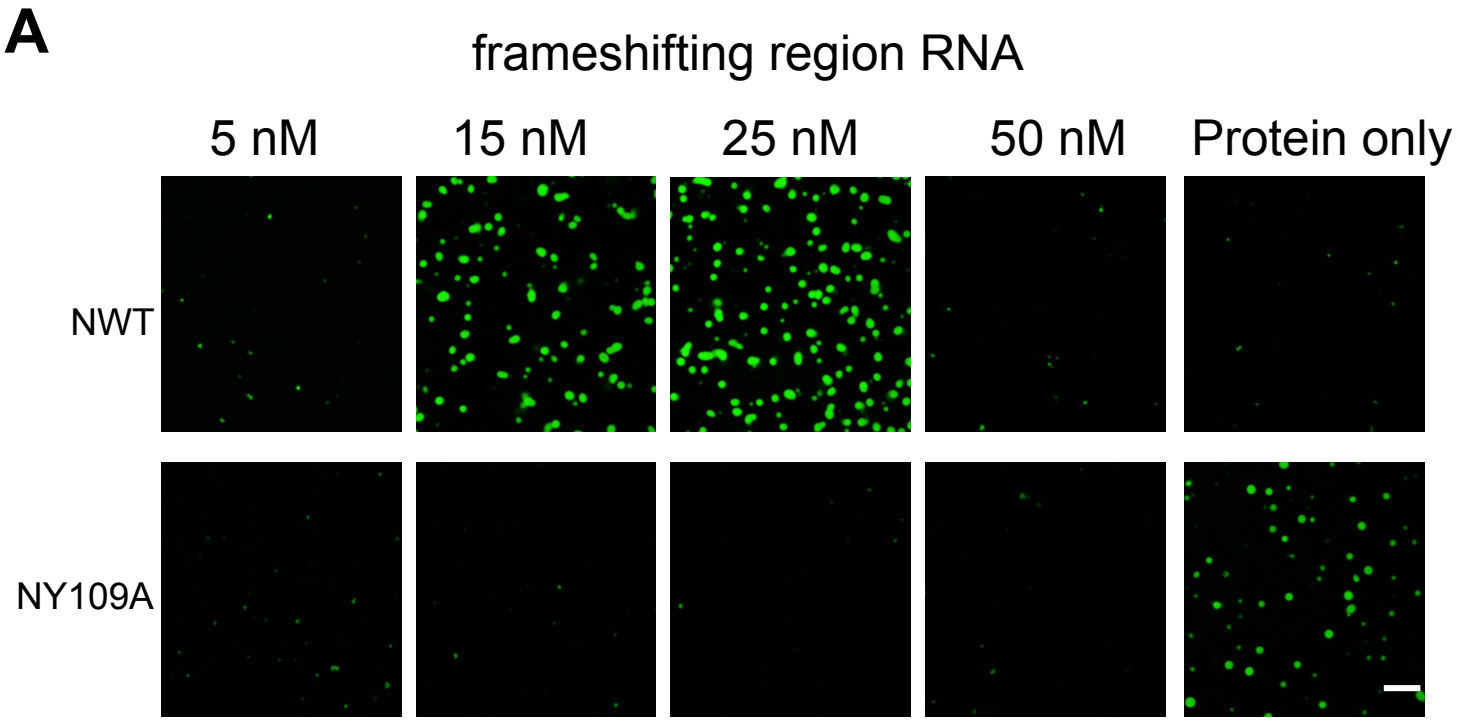
C

	AU content	MFE	ΔG z-s.	ED
5′ end mean	0.52	-33.19	-1.24	18.32
FSR mean	0.62	-26.07	-1.38	22.00
other mean	0.62	-25.86	-1.51	21.29
Wilcoxon pval (5′ end vs. FSR)	<2.2e-16	<2.2e-16	0.36	<2.2e-16
Wilcoxon pval (5′ end vs. other)	<2.2e-16	<2.2e-16	6.8e-05	<2.2e-16
Wilcoxon pval (FSR vs. other)	5.2e-07	0.48	0.01	0.003

Supplemental Figure 5: Comparison of SARS-CoV-2 genome to frameshifting region and 5′ end. Related to Figure 4.

A) Relationship between number of possible structures (ensemble diversity (ED), z axis), Percent AU (x axis), and MFE (y axis) for 5′ end (green), frameshifting region (FSR, purple), and (black) other (remaining genome sequences which are not 5′ end or frameshifting region, black). Each dot is 120 nucleotides derived from a sliding window (sliding 1 nt for FSR and 5′ end RNAs and 30 nt for remaining genome sequences to ease visualization). Most of the genome resembles frameshift region sequences. **B)** Comparison of 5′ end (green) and frameshifting region (FSR, purple) RNA sequence features. Comparisons run clockwise from the upper left panel as follows; (1) ∆G z-score vs. MFE. (2) ED vs. ∆G z-score (3) ∆G z-score vs. Percent AU. (4) ED vs. Percent AU. (5) Percent AU vs. MFE. (6) ED vs. MFE. MFE is the minimum free energy, ∆G z-score is (MFE_{native}-MFE_{random}), ED is a quantification of the variety of different structures which are present in the ensemble, and AU is the percent of nucleotides which are A or U. MFE, ∆G z-score, and ED are derived from and are defined in (Andrews et al., 2020). **C)** Related to panels A and B: the mean values for 5′ end, frameshifting region, and other (remaining genome sequences which are not 5′ end or frameshifing region) sequences for AU content, MFE, ∆G z-score, and ED. Frameshifting region sequences are significantly different from 5′ end sequences and the remaining gRNA (other).

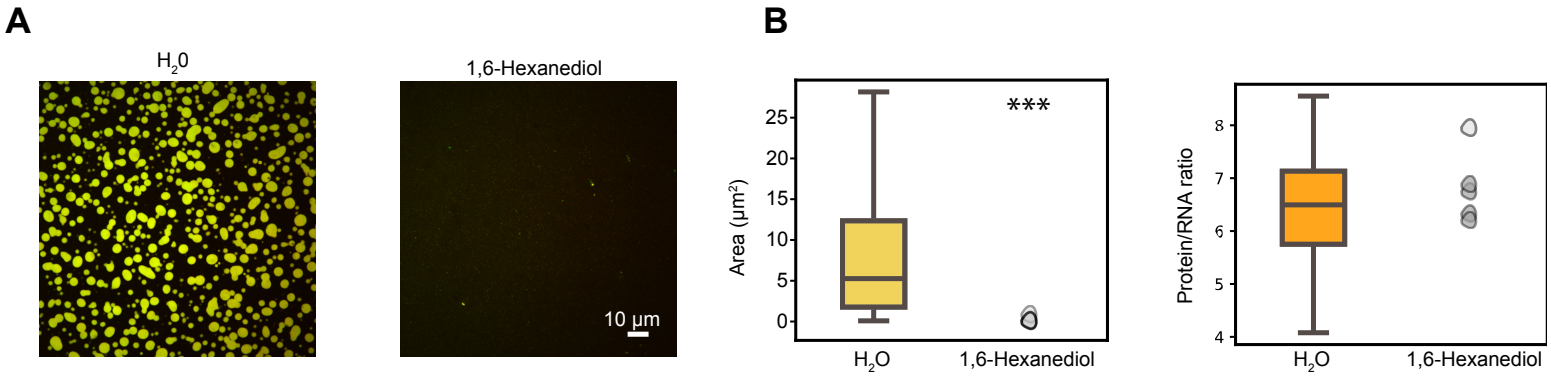
Figure S6


A
frameshifting region RNA
5 nM
15 nM
25 nM
50 nM
Protein only
NWT
NY109A

Supplemental Figure 6: N-protein Y109A droplets are dissolved by frameshifting region RNA. Related to Figure 5.

A) Comparison of wildtype N-protein and Y109A mutant protein droplets made with frameshifting region RNA. Y109A is solubilized by frameshifting region RNA relative to a water control. Images show protein (green) signal.

Figure S7


A
H20
1,6-Hexanediol
10 µm
B
***
Area (µm²)
25
20
15
10
5
0
H2O
1,6-Hexanediol
Protein/RNA ratio
8
7
6
5
4
H2O
1,6-Hexanediol

Supplemental Figure 7: 1,6-Hexanediol dissolves frameshifting region RNA/N-protein droplets, Related to Figure 6.

A) 9% 1,6-hexanediol prevents N-protein/frameshifting region RNA LLPS relative to water vehicle. Images show merge of protein (green) and RNA (red) signals. **B)** For 1,6-hexanediol, size but not protein:RNA ratio is reduced relative to vehicle. Left: quantification of condensate area depicted in **(A**). Right: quantification of protein to RNA ratio.