Cell Reports, Volume *31*

Supplemental Information

Mapping of Influenza Virus RNA-RNA Interactions Reveals a Flexible Network

Valerie Le Sage, Jack P. Kanarek, Dan J. Snyder, Vaughn S. Cooper, Seema S. Lakdawala, and Nara Lee

Figure S1

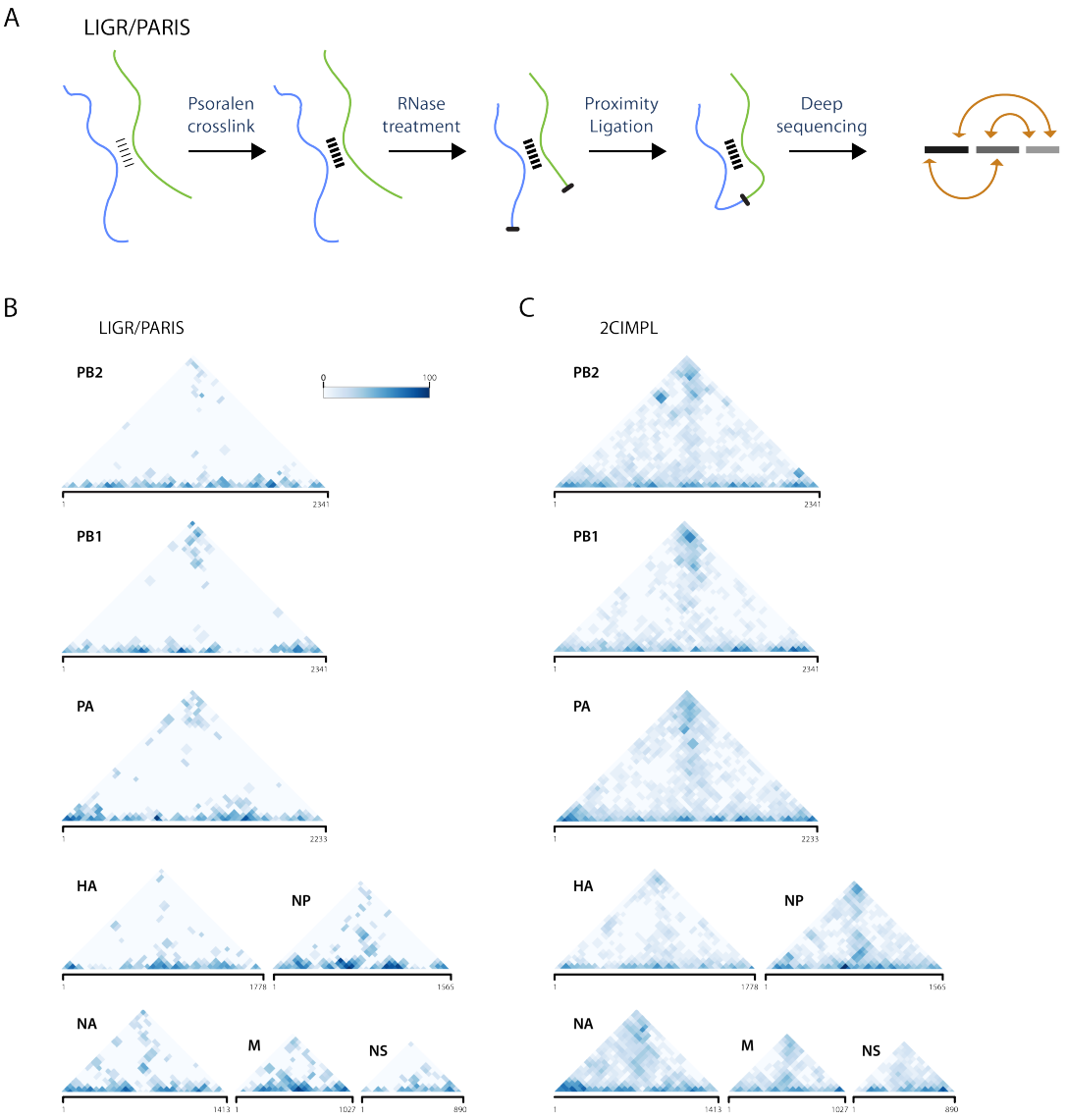
A
LIGR/PARIS
Psoralen crosslink
RNase treatment
Proximity Ligation
Deep sequencing
B
LIGR/PARIS
C
2CIMPL
PB2
PB1
PA
HA
NP
NA
M
NS
D
PB2
PA
M
NS
RNase-seq (Formaldehyde)
RNase-seq (UV 254 nm)
100
0

Figure S1. Comparison between 2CIMPL and LIGR/PARIS. Related to Figure 1. (**A**) Schematic outline for LIGR/PARIS. Unlike 2CIMPL, LIGR/PARIS is performed with naked RNA, thus not taking into account spatial constraints imposed by *in vivo* RNA-protein interactions. Given that many of the enzymatic steps are identical between LIGR/PARIS and 2CIMPL, it is noteworthy that enhanced yield in RNA hybrids was observed with the latter when applied to influenza virions. (**B+C**) Upper triangle plots of all eight WSN segments showing intrasegmental RNA-RNA interactions identified by LIGR/PARIS (B) or 2CIMPL (C). Compared to LIGR/PARIS, the 2CIMPL approach retrieves an increased number of the expected end-to-end intrasegmental RNA-RNA interactions (top of the triangle). While we cannot rule out that some of the interactions stem from defective interfering (DI) particles, the panhandle structure is retrieved for all segments and DI particles are observed mainly for the PB2 and PA segments. For our LIGR/PARIS approach, we followed the experimental outline of PARIS, which includes a 2-D gel enrichment step not included in LIGR, until deep sequencing library preparation and used the user-friendly computational analysis pipeline of the LIGR protocol for data analysis. (**D**) Viral RNA is poorly digested by RNase A after formaldehyde crosslinking. Related to Figure 1. RNase-seq of WSN virions after formaldehyde (upper track) or UV light-mediated crosslinking (lower track). Virions were harvested after crosslinking with 0.2-1% formaldehyde for 10 min at RT or UV light irradiation as in HITS-CLIP. Virions were lysed in PXL buffer and lysate was treated with RNase A. Viral RNA was isolated and converted into an Illumina-compatible deep sequencing library. Four representative segments (PB2, PA, M, NS) are shown.

Figure S2

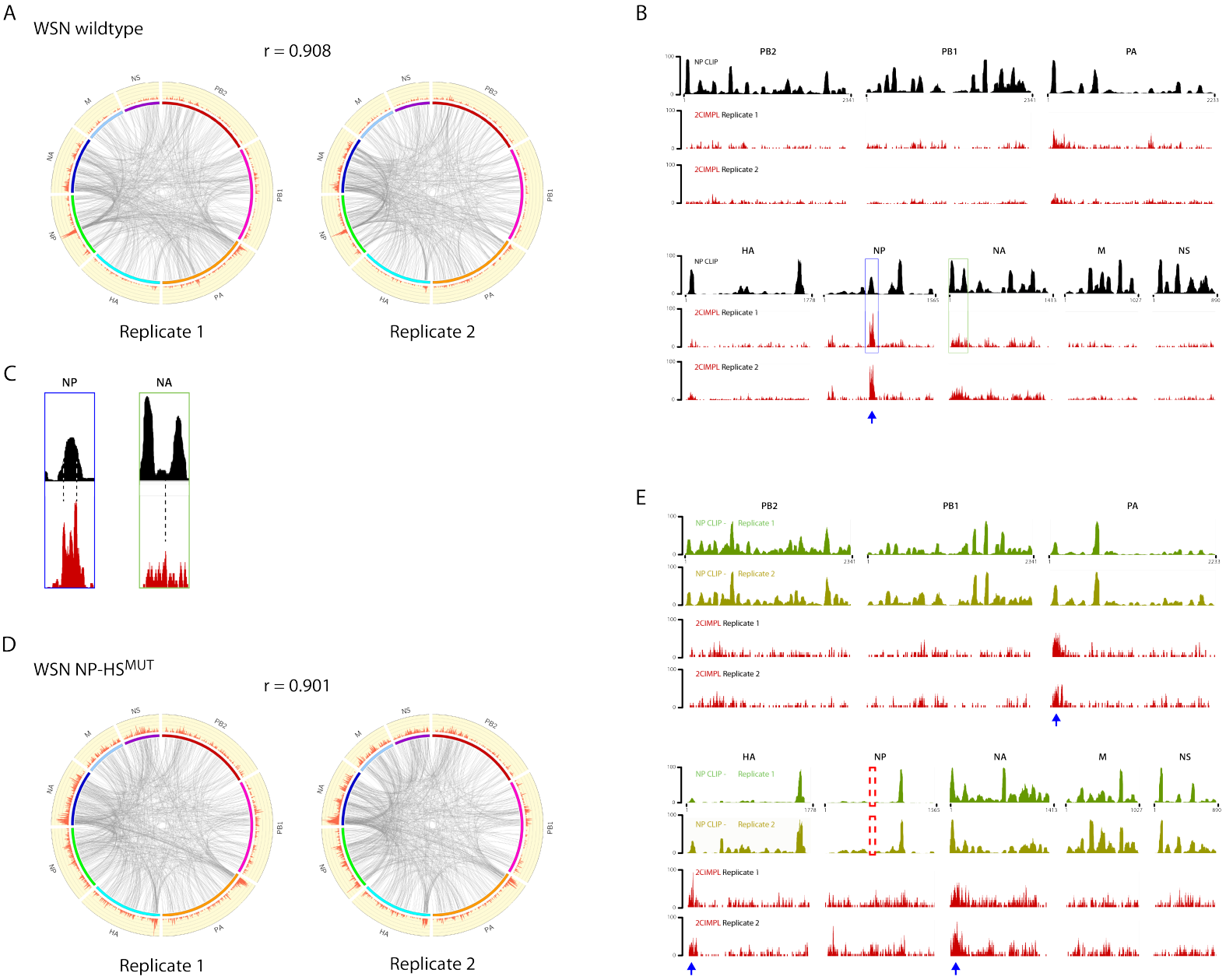
A
WSN wildtype
r = 0.908
Replicate 1
Replicate 2
B
NP CLIP
2CIMPL Replicate 1
2CIMPL Replicate 2
PB2
PB1
PA
HA
NP
NA
M
NS
C
NP
NA
D
WSN NP-HS^MUT
r = 0.901
Replicate 1
Replicate 2
E
NP CLIP - Replicate 1
NP CLIP - Replicate 2

Figure S2. 2CIMPL approach generates highly reproducible RNA-RNA interaction networks. Related to Figure 2 and 3. (**A**) Circos plots showing all intersegmental interactions of two independent biological replicates, which are highly reproducible with a Pearson correlation coefficient of 0.908. The high degree of reproducibility is reflected in the overall similar pattern in the histograms on the outer circle showing the frequency of junctions at a given site. (**B**) The histograms from part A are shown overlaid with the NP CLIP profile of the WSN wildtype strain (data previously published (Lee et al., 2017)). The blue box indicates the RNA-RNA interaction hotspot; the green box indicates the region in the NA segment, which interacts frequently with the hotspot region in NP. (**C**) The RNA hybrid junctions of the hotspot are adjacent to NP peaks. Zoomed-in view of blue and green boxes as shown in B. (**D**) Circos plots showing 2CIMPL results of two independent biological replicates for the WSN [NP-HS^{MUT}] strain. All intersegmental interactions are indicated. The Pearson correlation coefficient of 0.901 indicates a high degree of reproducibility. (**E**) The histograms of RNA hybrid junctions as in D are shown with the NP CLIP profiles of two independent biological replicates of the WSN [NP-HS^{MUT}] strain. The red box marks the location of the original hotspot region in the wildtype NP segment. New hotspots are indicated by arrows.

Figure S3

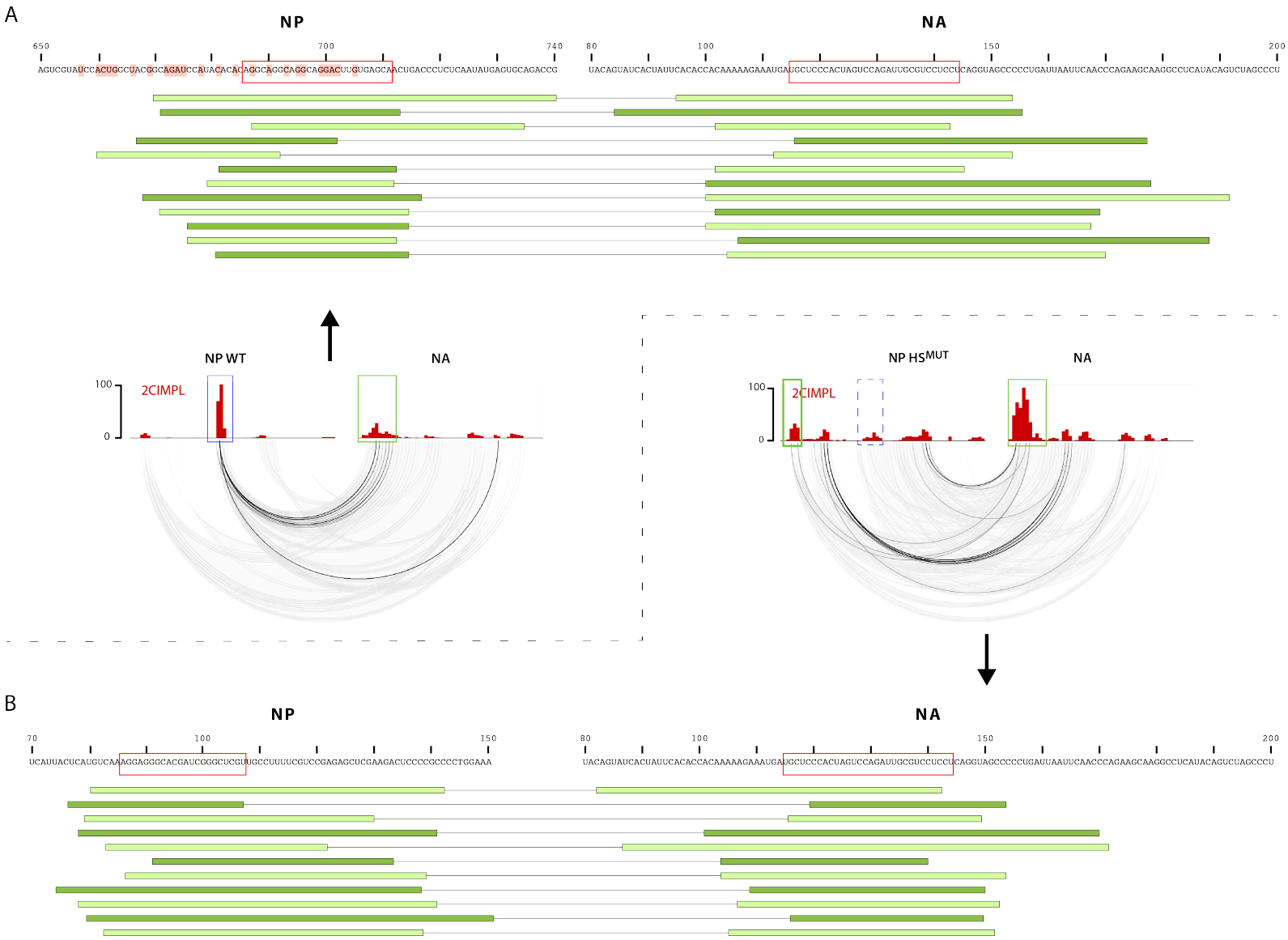
A
NP
NA
650
700
740
80
100
150
200
NP WT
NA
2CIMPL
100
0
NP HS^MUT
NA
2CIMPL
100
0
B
NP
NA
70
100
150
80
100
150
200

Figure S3. Identification of base pairs coordinating NP and NA interactions. Related to Figure 2 and 3. (**A**) The two regions of a single sequencing read that map to both NP and NA segments are shown. The nucleotides mapping to the RNA hybrids were used to predict the precise segment regions involved in the interaction. The nucleotides mutated in the WSN [HS^{MUT}] strain are highlighted. Nucleotides in red boxes are shown in the predicted RNA duplex in **Figure 2D**. Green boxes underneath the NP and NA segments correspond to the mapped regions of a single sequence read of an RNA hybrid. (**B**) Junctions between the NP segment of the WSN [NP-HS^{MUT}] strain and the 5′ NA segment. Nucleotides in red boxes are shown in the predicted RNA duplex in **Figure 3F**.

Figure S4

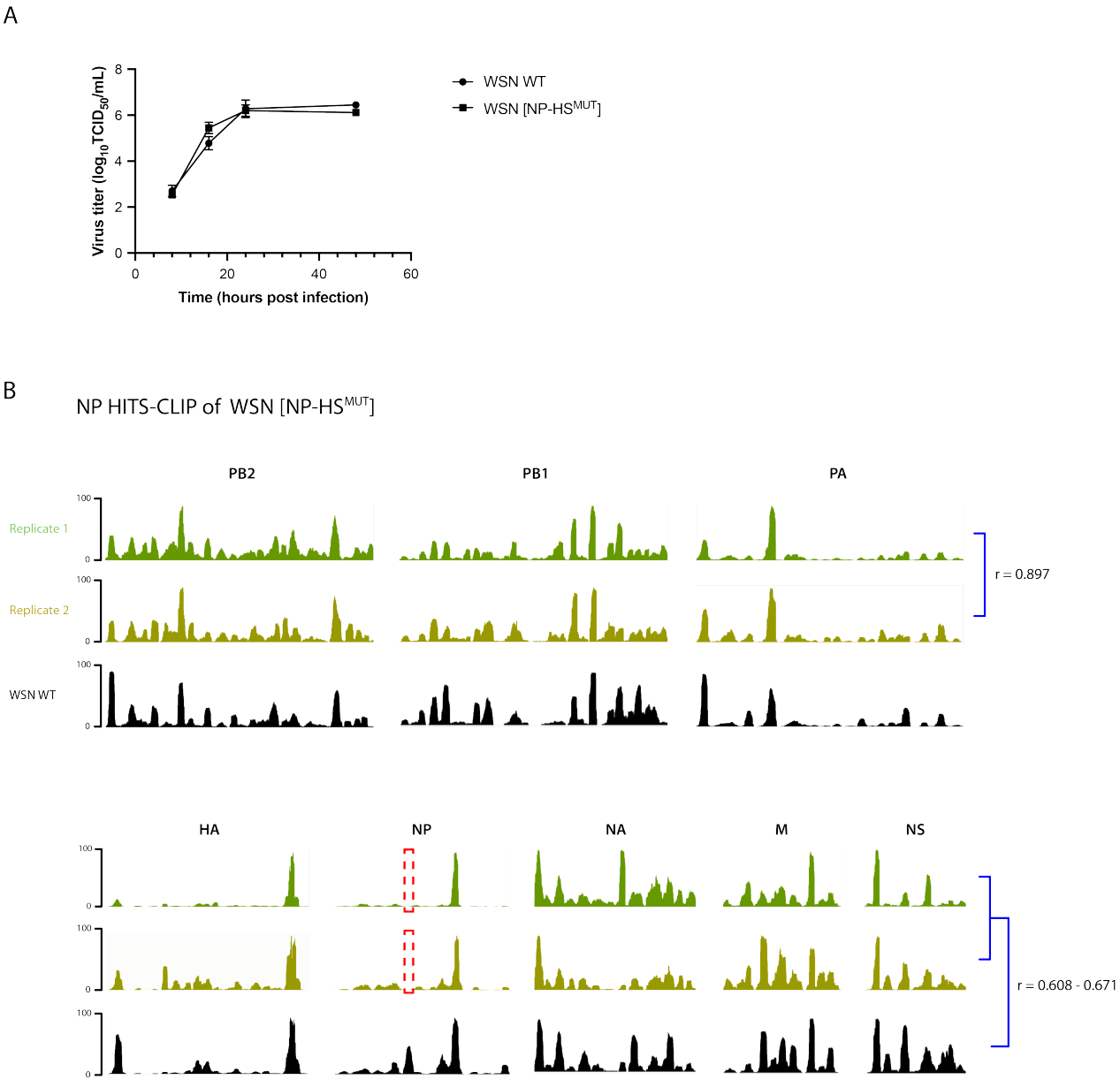

Figure S4. Genome-wide NP binding profile is altered in the WSN [NP-HS^{MUT}] strain. Related to Figure 3. (**A**) Replication kinetics of wildtype and WSN [HS^{MUT}] strains. MDCK cells were infected in triplicate at a MOI of 0.01. Supernatants were collected at the indicated time points and virus titers were determined using $TCID_{50}$ assays. (**B**) NP CLIP profiles of two independent biological replicates with a Pearson correlation coefficient of 0.897. These mutant NP profiles have a Pearson correlation coefficient of 0.608 and 0.671, respectively, with the wildtype strain.

Figure S5

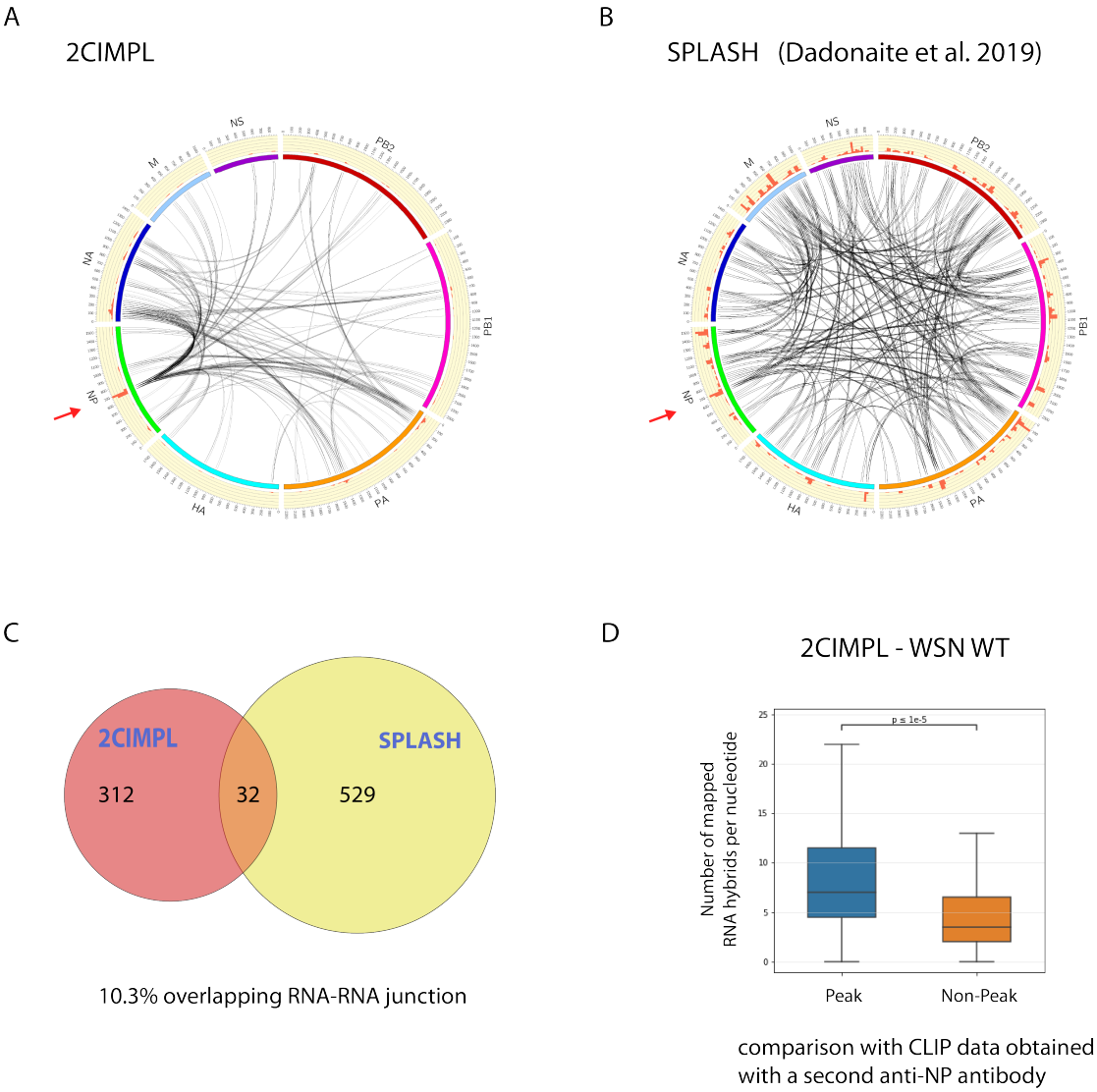

Figure S5. Comparison between 2CIMPL and the recently published SPLASH result. Related to Figure 5. (**A+B**) The SPLASH data (Dadonaite et al., 2019) was analyzed using the same analysis pipeline as the 2CIMPL data and displayed as a circos plot. The hotspot region in the NP segment is indicated by an arrow. While the SPLASH data does not exhibit a hotspot like the 2CIMPL data, the hotspot region is one of the major RNA-RNA interaction sites detected by SPLASH. (**C**) Overlap between 2CIMPL and SPLASH data. (**D**) Comparison of 2CIMPL data with a second NP HITS-CLIP data set. NP peak and non-peak regions as determined by a previously reported second NP HITS-CLIP data set (Lee et al., 2017) were overlaid with RNA duplex forming regions as in Figure 5. The values represent the absolute number of mapped RNA hybrids at each nucleotide, which were plotted based on their classification of overlapping with either a called NP peak or non-peak. Box plots were generated using *seaborn*, a Python library that builds off *matplotlib*. The bottom and top of the boxes represent the 25th and 75th percentiles; the bar in the middle represents the median value, and the whiskers represent 1.5x the interquartile range extended out from the 25th and 75th percentiles. P-values were determined using the Wilcoxon signed-rank test.