Supplementary information file

The Structure of the Influenza A Virus Genome

Bernadeta Dadonaite[1]*, Brad Gilbertson[2]*, Michael L. Knight[1], Sanja Trifkovic[2], Steven Rockman[2,3], Alain Laederach[4], Lorena E. Brown[2], Ervin Fodor[1], David L. V. Bauer[1]

[1]Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

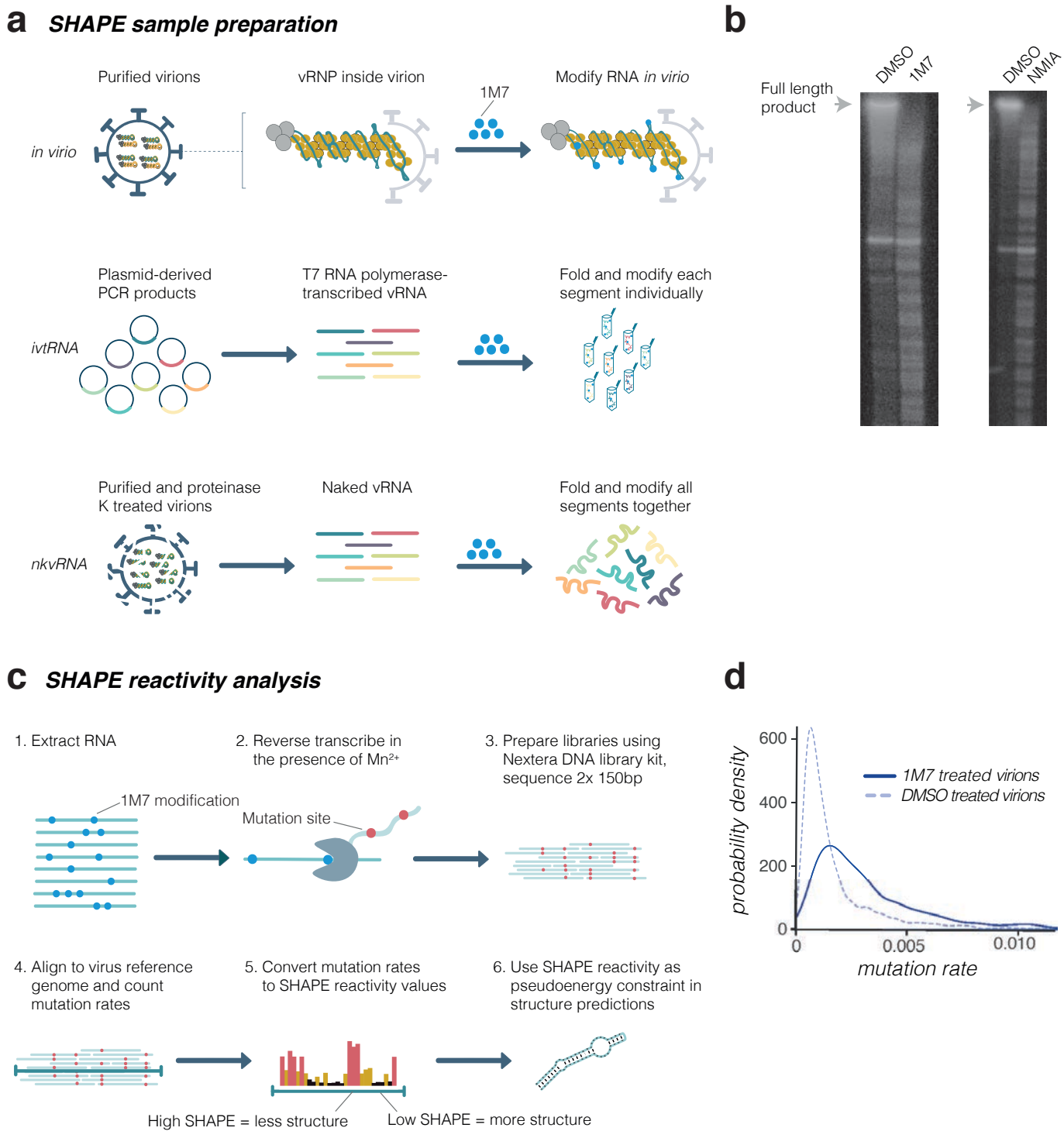
[2]Department of Microbiology and Immunology, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia

[3]Seqirus Ltd, 63 Poplar Road, Parkville, Victoria 3052, Australia

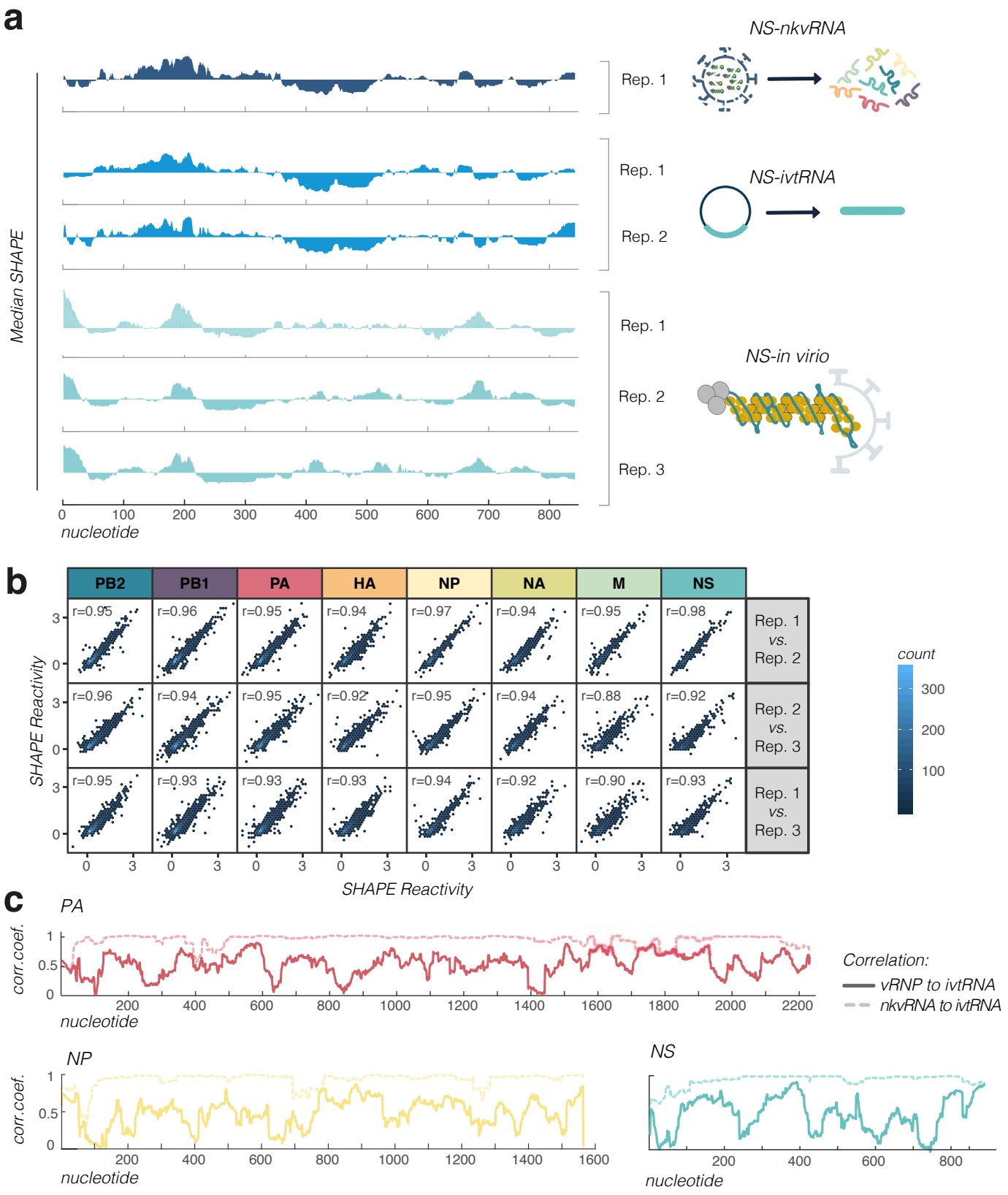
[4]Department of Biology, University of North Carolina, Chapel Hill, North Carolina, USA

*These authors contributed equally to this work

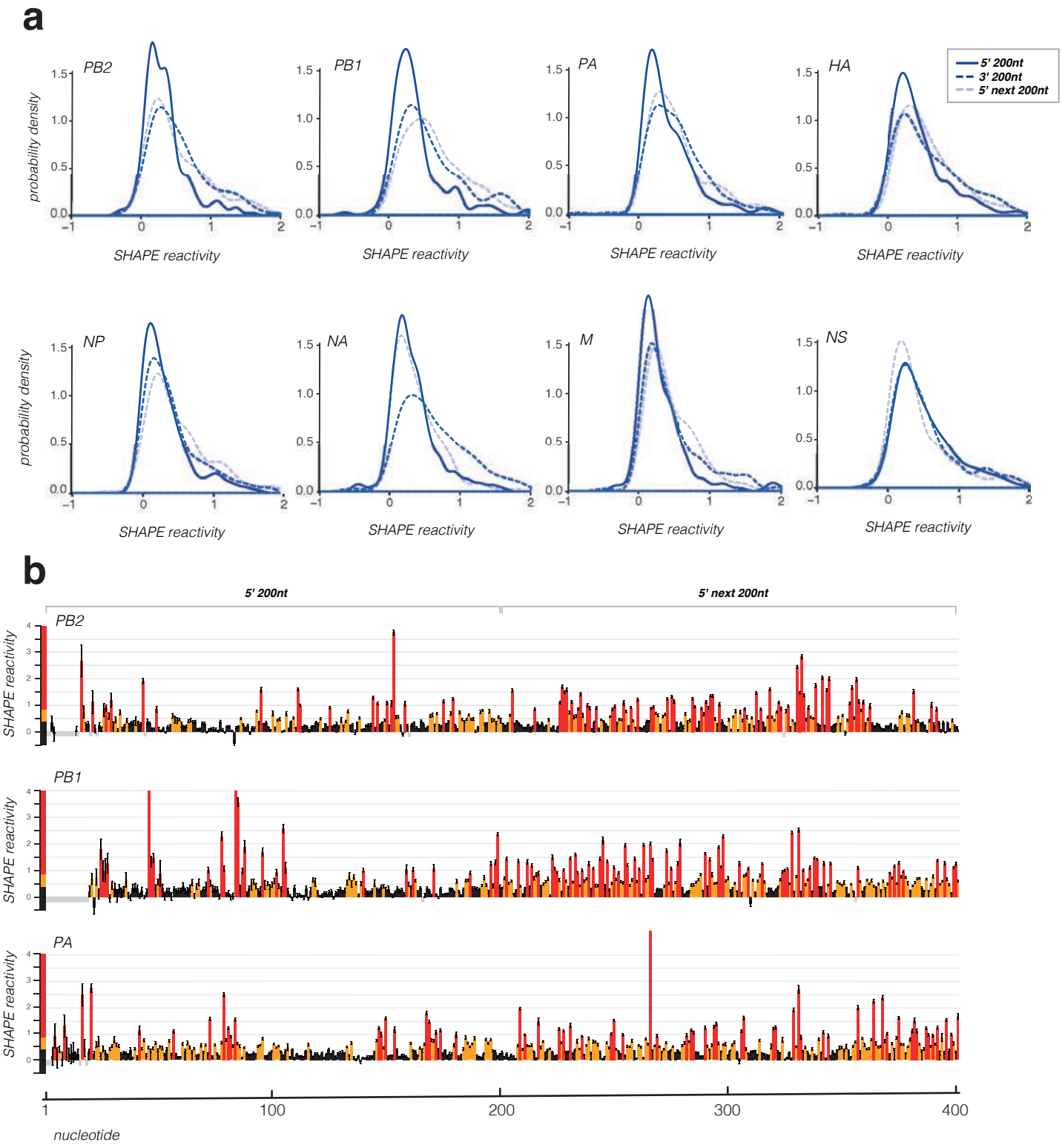
Contents



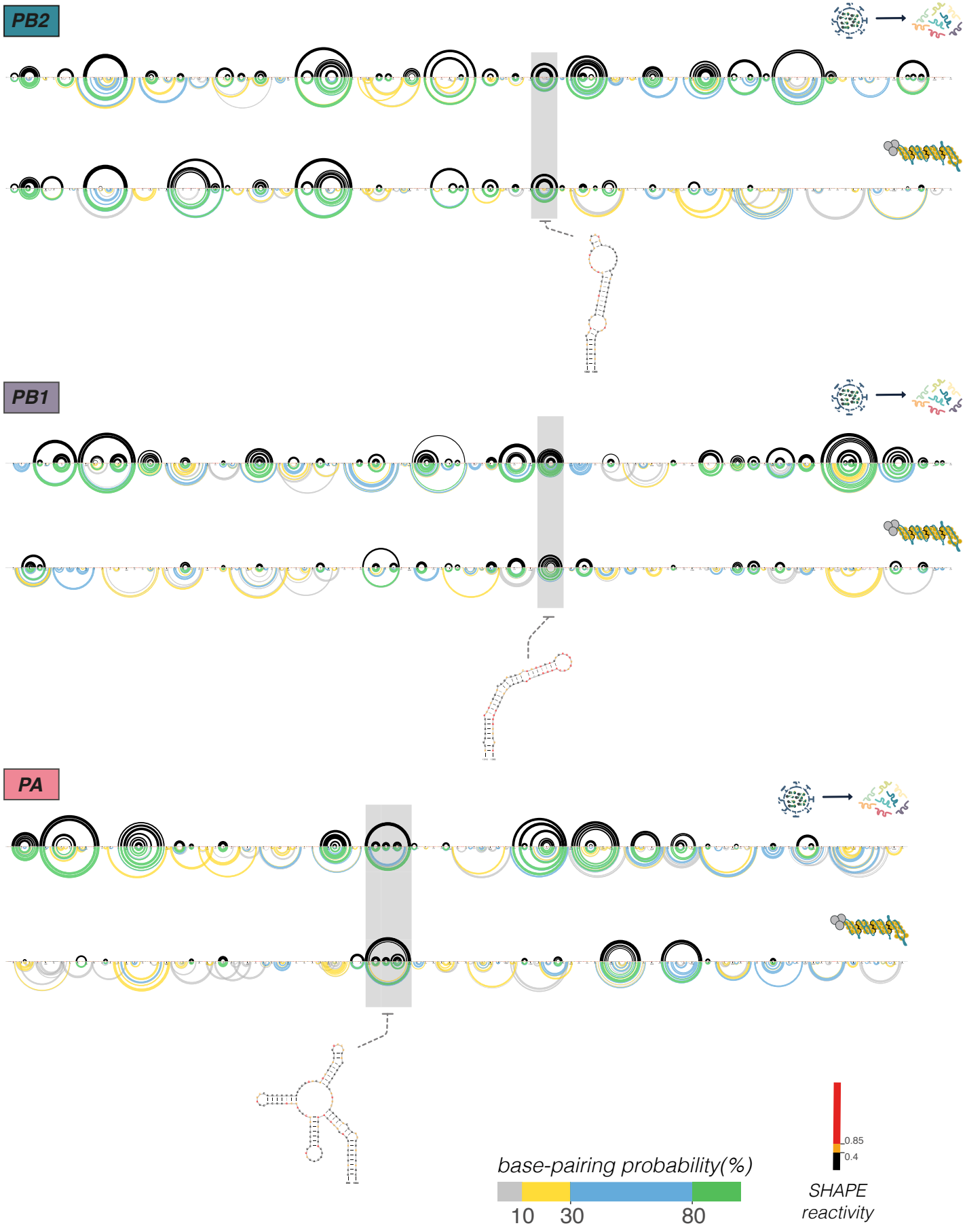
Supplementary Figure 1 | SHAPE-MaP analysis of the IAV genome. a, Schematic showing the vRNA samples used for SHAPE-MaP analysis. **b,** Reverse transcription primer extension reaction using vRNA that was extracted from SHAPE reagent (1M7 or NMIA) or DMSO treated viral samples and a ^{32}P-labelled primer targeting NA segment vRNA. Bands indicate the stalling of reverse transcriptase at the sites of SHAPE modifications. Image shown is representative of 3 (1M7) and 5 (NMIA) independent experiments. **c,** Schematic of SHAPE-MaP library preparation and sequencing data analysis. **d,** Mutation rates in DMSO versus 1M7 reagent treated samples. 1M7, 1-methyl-7-nitroisatoic anhydride; NMIA, N-methylisatoic anhydride; DMSO, dimethyl sulfoxide.



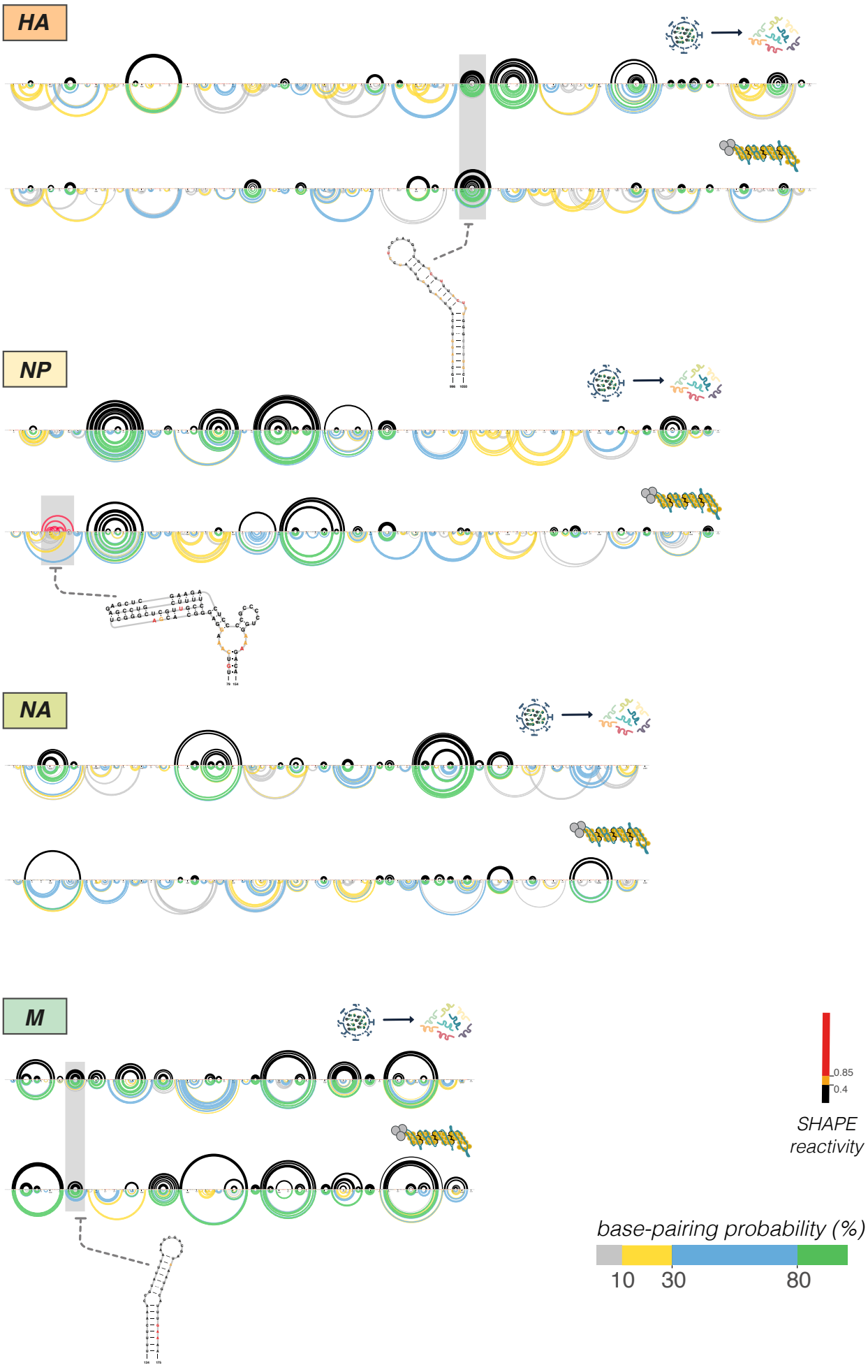
Supplementary Figure 2 | SHAPE reactivity variation between replicates and SHAPE reactivity correlation between different vRNA samples. a, Comparison between the *in virio* and *ex virio* (nkvRNA and ivtRNA) median SHAPE reactivities of NS segment vRNA across independent replicates. Medians were calculated over 50 nucleotide windows and plotted relative to the global median. Regions below zero tend to be more structured, while regions above zero indicate more flexible regions of RNA. **b,** Correlation of single-nucleotide SHAPE reactivities between each *in virio* replicate (rows) presented as a 2D histogram (scale at right) for each segment (columns) with Pearson's r indicated on each plot. **c,** Sliding window correlation of SHAPE reactivities between *ex virio* vRNA samples indicates very high correlation between different *ex virio* conditions across the different segments. Sliding window correlation between *in virio* and *ex virio* samples indicates regions of high SHAPE reactivity correlation, where the secondary RNA structures are likely maintained. Pearson correlation was calculated over 50 nucleotide windows. ivtRNA, *in vitro* transcribed RNA; nkvRNA, naked viral RNA; Rep., replicate.



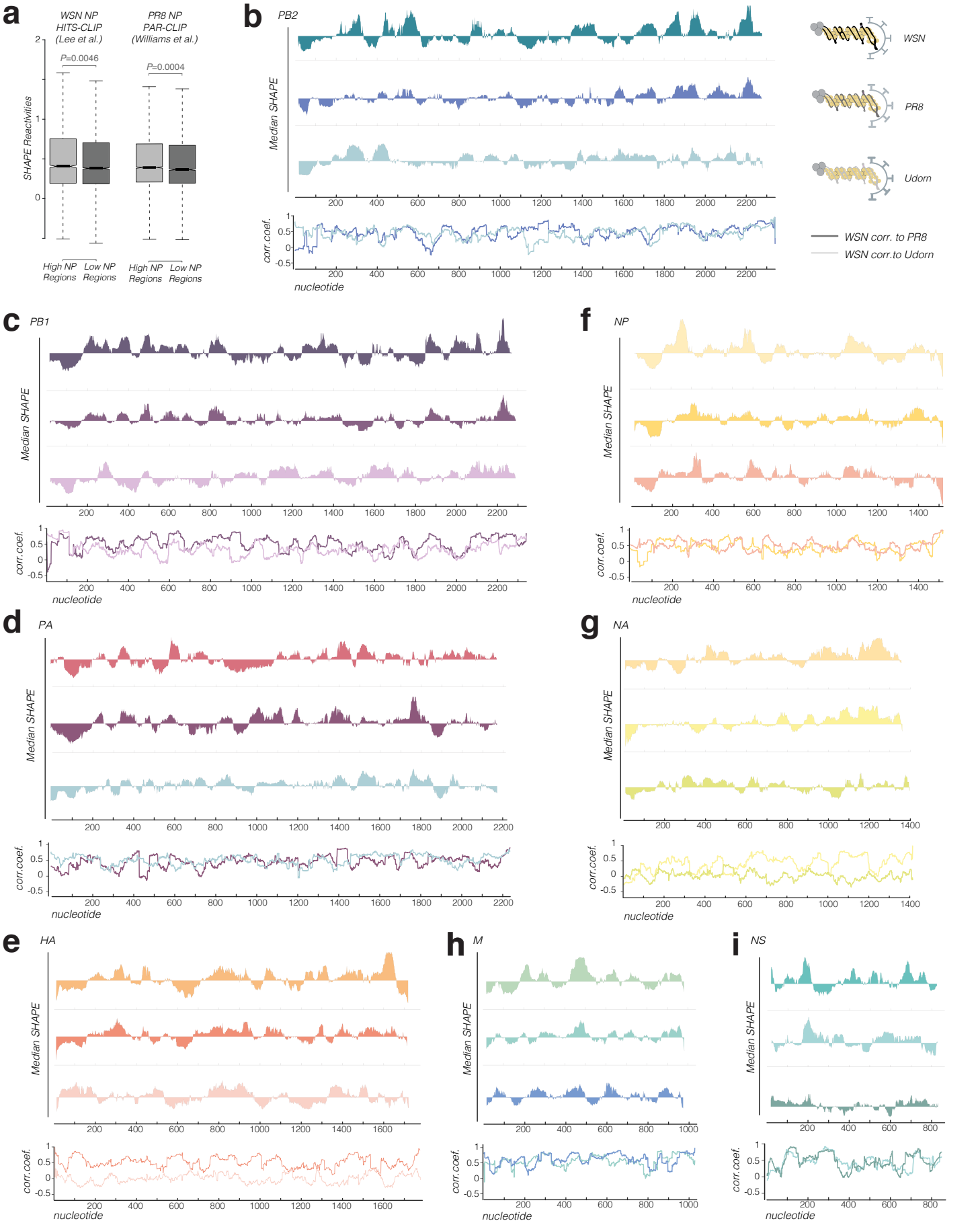
Supplementary Figure 3 | Analysis of the SHAPE reactivities in the WSN genome. a, Probability distribution of the SHAPE reactivities in different regions of individual viral segments. 5' 200nt, the first 200 nucleotides of the vRNA; 3' 200nt, the last 200 nucleotides of the vRNA; 5' next 200nt, the 200 nucleotides following the first 200 nucleotides of the vRNA. **b,** Individual SHAPE reactivities of the first 400 nucleotides of the WSN polymerase segments. High (red) SHAPE reactivities indicate flexible regions of RNA, low SHAPE reactivities (black) indicate constrained regions of RNA, intermediate reactivities are coloured yellow and grey regions indicate the nucleotides for which the reactivities are not available due to low sequencing coverage.



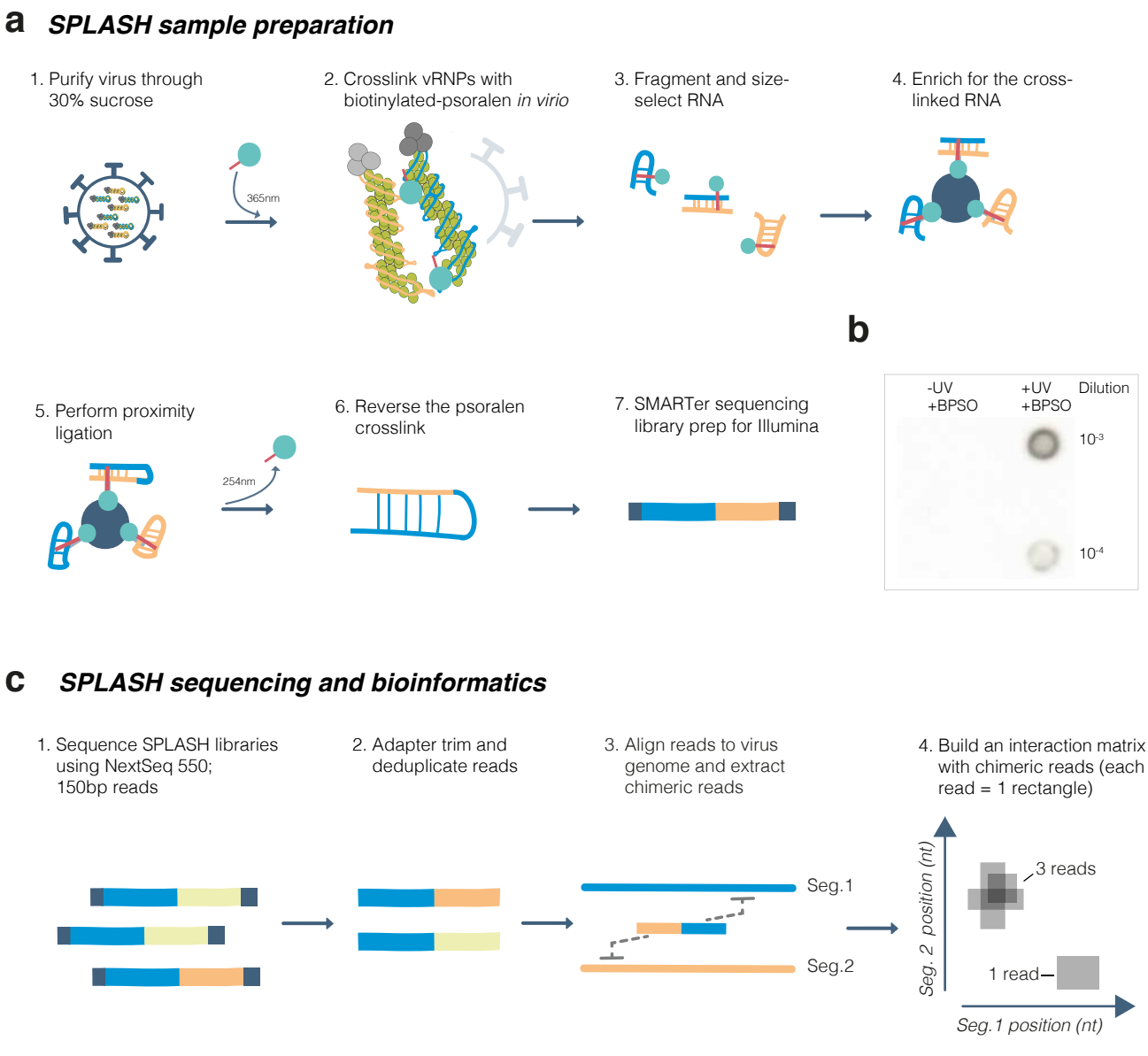
Supplementary Figure 4 | SHAPE-informed secondary RNA structure of the IAV polymerase segments. Black arcs indicate the maximum expected accuracy RNA structures; only the arcs associated with greater than 80% base-pairing probabilities are shown. Coloured arcs show base-pairing probabilities. Secondary structure examples of highlighted regions are shown.



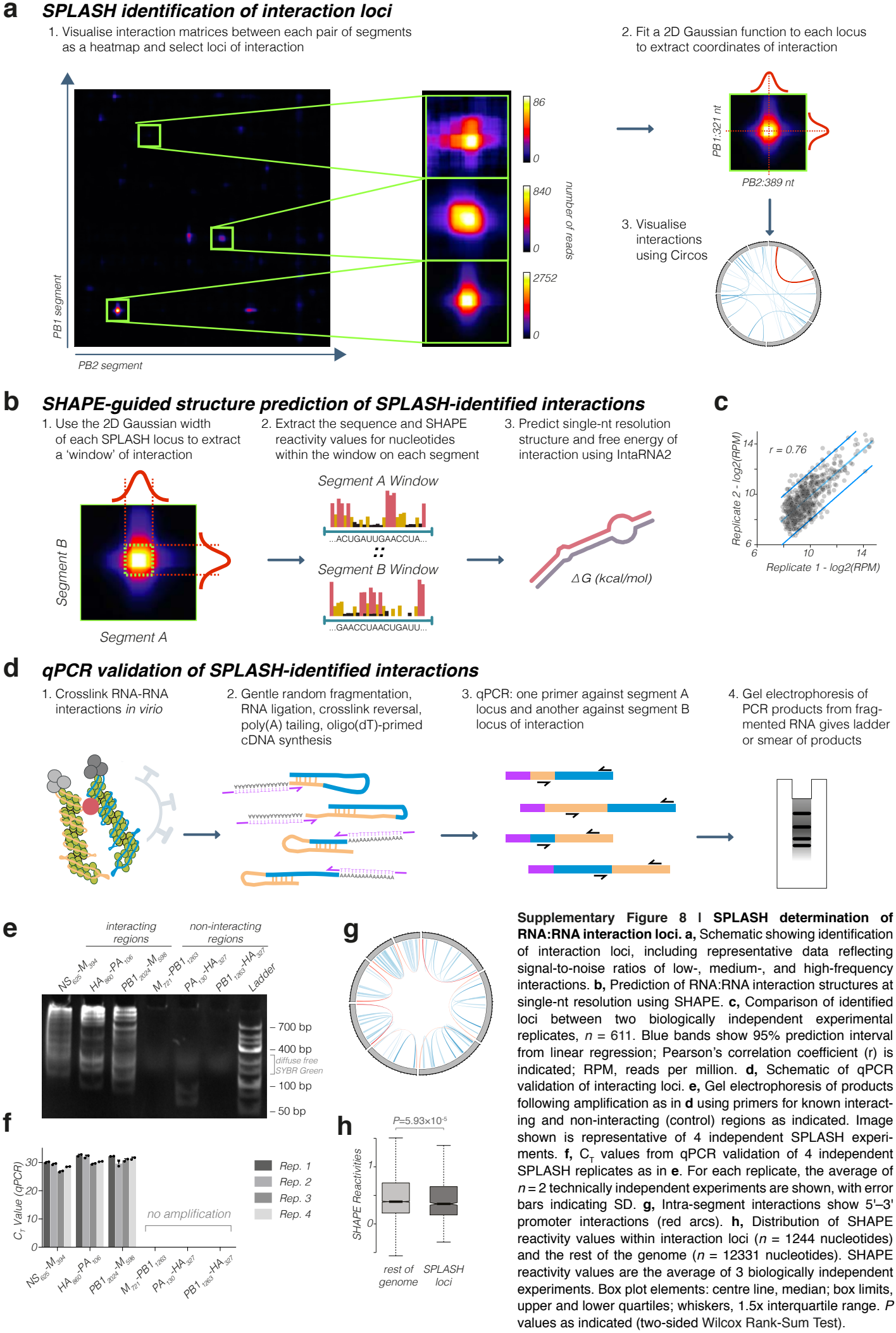
Supplementary Figure 5 | SHAPE-informed secondary RNA structure of the IAV HA, NP, NA and M segments. Black arcs indicate the maximum expected accuracy RNA structures; only the arcs associated with greater than 80% base-pairing probabilities are shown. Coloured arcs show base-pairing probabilities. Secondary structure examples of highlighted regions are shown. The previously-predicted hairpin in M[22] and pseudoknot in NP[23] are highlighted.



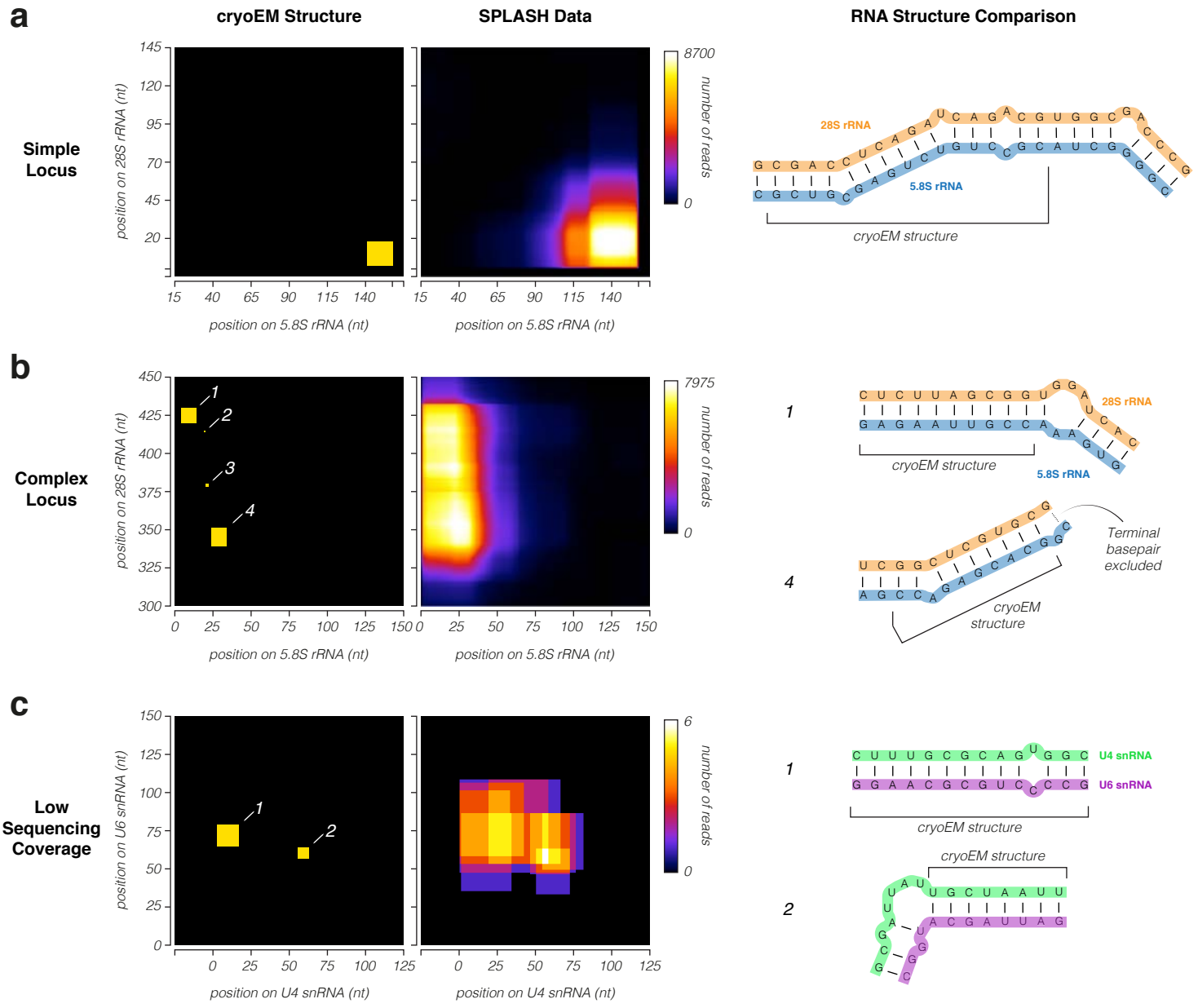
Supplementary Figure 6 | Comparison of SHAPE reactivities and NP occupancy in different IAV strains. a, Distribution of SHAPE reactivity values in high- and low-NP regions of WSN (as determined by Lee et al.[19], *n* = 3409 and 10166 nucleotides per respective region, data is an average of three biologically independent samples) and PR8 (as determined by Williams et al.[20], *n* = 6651 and 6931 nucleotides per respective region, data is from a single biological sample). Box plot elements: centre line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. *P* values were determined using two-sided Wilcox Rank-Sum Test. **b-i,** SHAPE reactivity plots for each segment and correlation between WSN, PR8, and Udorn strains, as indicated in the legend to the right of panel **b**. SHAPE reactivity medians were calculated over 50 nt windows and plotted relative to the global median. Pearson correlation was calculated over 50 nt windows. Data shown is the average of three biologically independent experiments (WSN) and of a single experiment (PR8 and Udorn); raw data is presented in **Supplementary Table 1**. WSN, A/WSN/1933 (H1N1); PR8, A/Puerto Rico/8/1934 (H1N1); Udorn, A/Udorn/307/72 (H3N2); nt, nucleotide; corr., correlation.



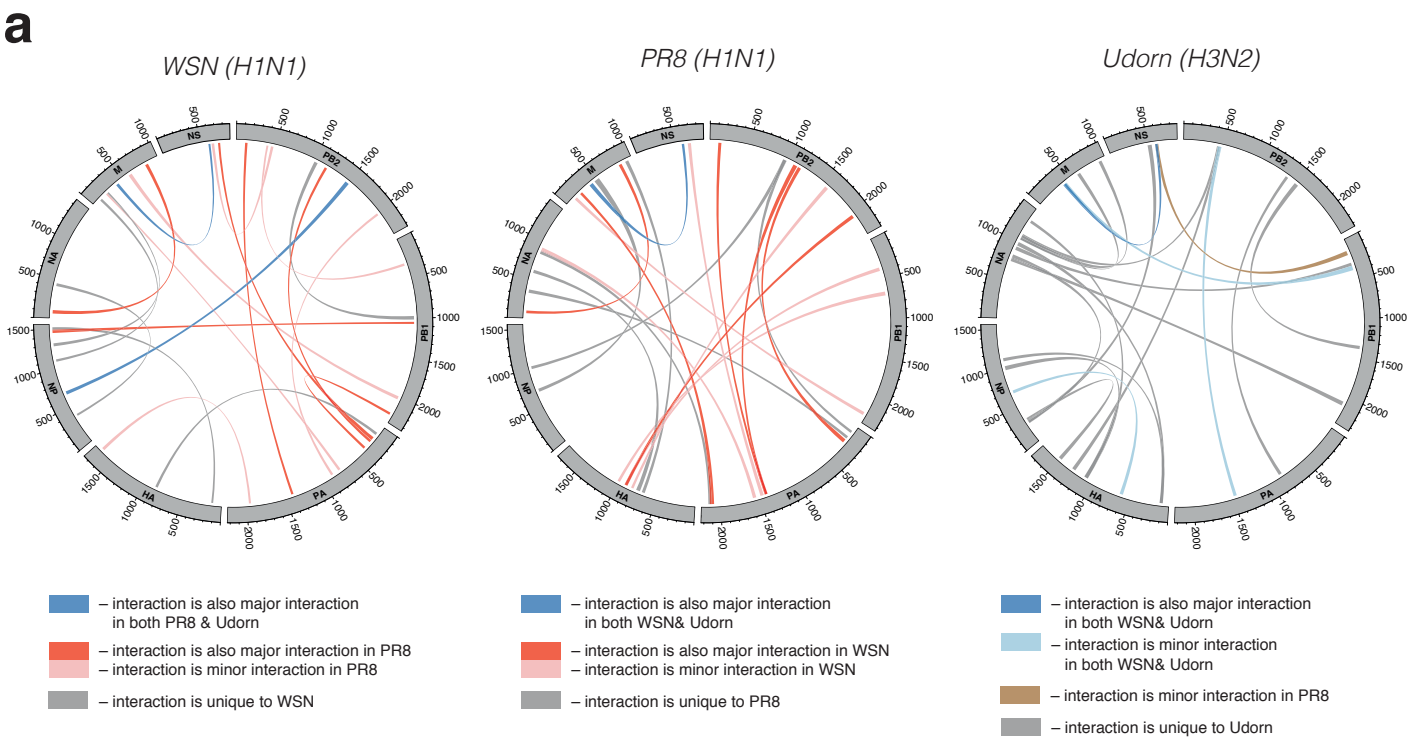
Supplementary Figure 7 | SPLASH samples and sequencing. a, Schematic showing SPLASH sample preparation. **b,** Anti-biotin dot blot analysis of RNA crosslinked with biotinylated psoralen and extracted from viral particles. Image shown is representative of 3 biologically independent experiments. **c,** Schematic showing the SPLASH sequencing method and bioinformatics analysis steps. **d,** Read overlaps between two experimental replicates. **e,** Chimeric reads aligned to the HA segment from two experimental replicates versus total RNA input coverage. Red trace indicates sample in which T4 RNA ligase 1 was omitted during proximity ligation (step 5 in **a**). BPSO, biotinylated psoralen; Rep., replicate.

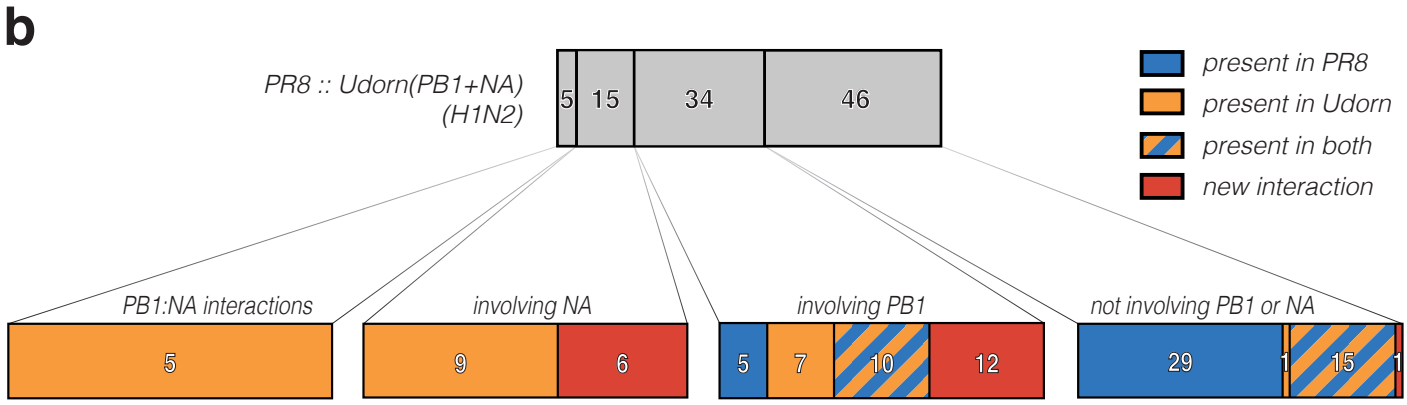


Supplementary Figure 8 | SPLASH determination of RNA:RNA interaction loci. a, Schematic showing identification of interaction loci, including representative data reflecting signal-to-noise ratios of low-, medium-, and high-frequency interactions. **b,** Prediction of RNA:RNA interaction structures at single-nt resolution using SHAPE. **c,** Comparison of identified loci between two biologically independent experimental replicates, n = 611. Blue bands show 95% prediction interval from linear regression; Pearson's correlation coefficient (r) is indicated; RPM, reads per million. **d,** Schematic of qPCR validation of interacting loci. **e,** Gel electrophoresis of products following amplification as in **d** using primers for known interacting and non-interacting (control) regions as indicated. Image shown is representative of 4 independent SPLASH experiments. **f,** C_T values from qPCR validation of 4 independent SPLASH replicates as in **e**. For each replicate, the average of $n = 2$ technically independent experiments are shown, with error bars indicating SD. **g,** Intra-segment interactions show 5'–3' promoter interactions (red arcs). **h,** Distribution of SHAPE reactivity values within interaction loci (n = 1244 nucleotides) and the rest of the genome (n = 12331 nucleotides). SHAPE reactivity values are the average of 3 biologically independent experiments. Box plot elements: centre line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. P values as indicated (two-sided Wilcox Rank-Sum Test).

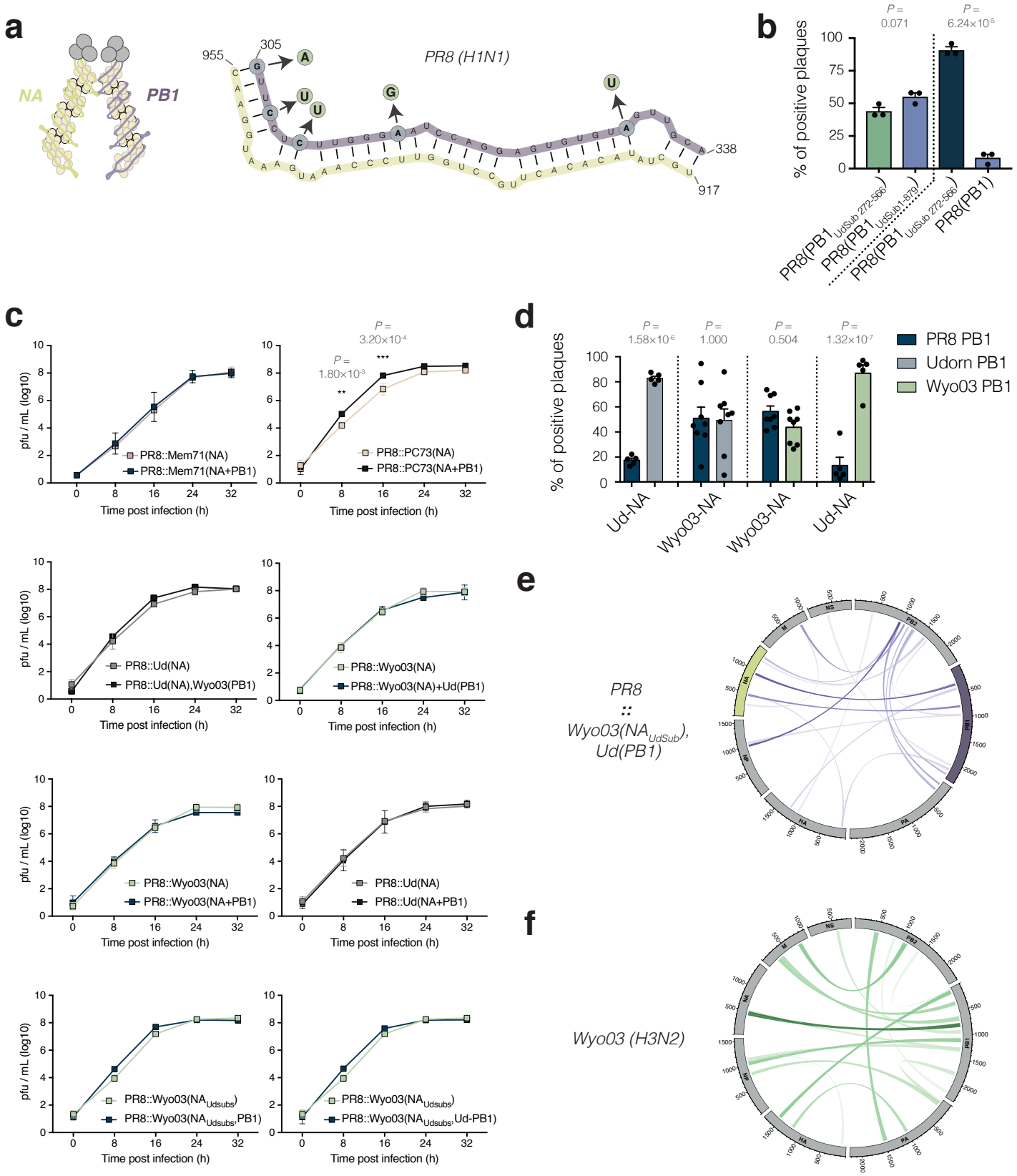


Supplementary Figure 9 | Benchmarking of SPLASH-based RNA:RNA interaction structure prediction against structured cellular RNAs. a-c, Inter-segment RNA structures determined from published cryoEM structures of the 80S ribosome[43] (PDB: 6EK0) and the U4/U6.U5 tri-snRNP spliceosomal complex[44] (EMDB: EMD-2966) are shown in the left column with base-paired regions indicated (yellow rectangles). SPLASH sequencing reads for the corresponding region are shown in the second column as a heatmap (scale as indicated), and the predicted RNA structures using the SPLASH data and the IntaRNA 2.0 algorithm (see Methods) are shown in the right column and compared to the cryoEM structures. In all cases, the SPLASH data is sufficient to correctly predict the core RNA structure (as indicated). **a-b,** Inter-segment interactions present in the 80S ribosome (PDB: 6EK0) comprise a set of five discrete interactions between the 5.8S and 28S rRNA segments, and form two loci: **a,** a simple locus between the end of the 5.8S rRNA and the beginning of the 28S rRNA, and **b,** a complex locus composed of four base-pairing stretches (numbered in panel) between the beginning of the 5.8S rRNA and the ~325-425nt region of the 28S rRNA. Two of these stretches ('2' and '3') are comprised of a 1-nt and 2-nt basepair, respectively, and are too short to be detected as discrete interactions by IntaRNA. **c,** Inter-segment interactions present in the U4/U6.U5 tri-snRNP spliceosomal complex (EMDB: EMD-2966) comprise a set of two discrete interactions between the U4 and U6 snRNAs. Low SPLASH sequencing coverage of the U4/U6 snRNA (9 reads total) does not prevent accurate prediction of the core RNA structures.





Supplementary Figure 10 | Common loci involved in inter-segment interactions in different IAV strains. a, The top 20 interactions in WSN, PR8, and Udorn are shown. Loci in common with other strains are highlighted. Note that some interactions are in common with another strain when considering the top 10%, but do not appear on the corresponding strain's plot since they are not in the top 20 interactions. A full table of interactions and comparisons is presented in **Supplementary Data Table 2**. **b,** Analysis of the origin of the top 100 interactions in the PR8::Udorn(PB1+NA) reassortant virus, classified according to which segments are involved in the interaction (see **Figure 3** for Circos plot).



Supplementary Figure 11 | Properties of viruses containing different sources of NA and PB1 segments. a, Structure prediction for the H3N2-origin PB1-NA interaction with circled nucleotides highlighting the bases on the PB1 segment that differ between Udorn (H3N2) and PR8 (H1N1) strains. **b,** Competitive reverse-engineering of influenza viruses with a PR8::Udorn(NA) background and chimeric PR8 PB1s with Udorn region as indicated, demonstrates that the region containing the interaction shown in panel **a** is responsible for PB1-NA co-segregation; *P* values as indicated, ANOVA with Tukey correction for multiple testing, $n = 3$ biologically independent experiments for each indicated competition; bar plot centre represents the mean, error bars indicate SEM. **c,** Replication kinetics of viruses on a PR8 background. Relative fitness was assessed for different virus pairs representing the potential progeny of the different competitive reverse-engineering experiments shown in Fig. 4; data shown is the mean of $n = 3$ biologically independent experiments, except for PR8::Ud(NA), where $n = 4$, error bars indicate SD. P values as indicated, two-way ANOVA with Sidak correction for multiple testing. **d,** Competitive reverse-engineering experiment showing Wyo03 NA segment, not its PB1 segment, is responsible for the failure of Wyo03 PB1 and NA segments to co-segregate; *P* values as indicated, ANOVA with Sidak correction for multiple testing, $n = 5$ (Ud-NA) and 8 (Wyo03-NA) biologically independent experiments for each indicated competition; bar plot centre represents the mean, error bars indicate SEM. **e,** Inter-segment RNA interactions in the PR8::Wyo03(NA_{UdSub}),Ud(PB1) virus. **f,** Inter-segment RNA interactions in the Wyo03 virus. The larger width of the arcs reflects the low sequencing coverage of this experiment. Mem71, A/Memphis/1/71; Wyo03, A/Wyoming/3/03; PC73, A/Port Chalmers/73; Udorn, A/Udorn/307/72; PR8, A/Puerto Rico/8/34.