

Cell Systems, Volume 11

Supplemental Information

RNA-GPS Predicts SARS-CoV-2 RNA

Residency to Host Mitochondria and Nucleolus

Kevin E. Wu, Furqan M. Fazal, Kevin R. Parker, James Zou, and Howard Y. Chang

Supplementary Tables & Figures

Localization	APEX2 fusion complex
Nucleolus	GFP + APEX2 + NF- κ B-inducing kinase (NIK, three tandem nuclear targeting sequences)
Nucleus	V5 + APEX2 + nuclear localization sequence (NLS)
Nuclear lamina	V5 + APEX2 + N-terminus of prelamin-A/C (LMNA)
Nuclear pore	V5 + APEX2 + N-terminus of sentrin-specific protease 2 (SEN2)
Cytosol	FLAG + APEX2 protein + nuclear export signal (NES)
ER Membrane	Amino acids 1-27 of P450 oxidase 2C1 + APEX2 + V5
Mitochondrial matrix	Amino acids 1-24 of COX4 (component of mitochondrial electron transport chain) + V5 + APEX2
Outer mito membrane	FLAG + APEX2 + C-terminal amino acids 510-540 of mitochondrial antiviral signaling protein (MAVS)

Table S1 (related to Figure 1): APEX2 fusions used to measure localization of transcripts. APEX2 is responsible for labelling, while the protein (segments) it is fused to drive its localization. Additional information regarding the APEX-seq protocol and data can be found in the original APEX-seq manuscript (Fazal et al., 2019), particularly Figure S2. Transcripts picked up by APEX2 (both en route and upon arrival at each fusion's final destination) are used to train the RNA-GPS model.

Strain	Count	Proportion
Human coronavirus NL63	48	0.25
Human coronavirus 229E	22	0.12
Human coronavirus OC43	82	0.43
Human coronavirus HKU1	3	0.02
MERS coronavirus	20	0.10
SARS coronavirus (2003)	16	0.08
Total	191	1.00

Table S2 (related to STAR Methods): Viral strains comprising the human coronavirus baseline. The strains NL63, 229E, OC43, and HKU1 historically commonly infect humans worldwide, while the MERS and SARS coronavirus have been recently responsible for more severe outbreaks.

	ER membrane	Nuclear lamina	Mito matrix	Cytosol	Nucleolus	Nucleus	Nuclear pore	Outer mito membrane
ORF1ab	1.38E-05	1.00	4.13E-32	1.00	4.86E-24	1.00	1.00	1.00
S	3.54E-05	1.00	8.01E-41	1.00	1.04E-37	1.00	1.00	1.00
ORF3a	1.00	1.00	9.08E-61	1.00	5.89E-36	1.00	1.00	1.00
E	8.91E-01	2.86E-13	3.11E-65	1.00	1.38E-22	1.00	1.00	1.00
M	1.23E-03	2.14E-05	8.72E-53	1.00	1.73E-21	1.00	1.00	1.00
ORF6	1.00	7.60E-09	1.31E-47	1.00	3.47E-25	1.00	1.00	1.00
ORF7a	5.92E-02	1.00	1.01E-54	1.00	4.26E-25	1.00	1.00	1.00
ORF8	2.62E-01	7.38E-01	2.56E-61	1.00	7.50E-26	1.00	1.00	1.00
N	1.00	1.00	7.83E-64	1.00	1.15E-22	1.00	6.10E-01	1.00
ORF10	1.00	6.84E-03	5.04E-85	1.00	5.17E-05	1.00	7.26E-08	1.00
ORF7b	1.00	8.94E-01	6.77E-09	1.00	2.85E-05	1.00	1.00	1.00

Table S3 (related to Figure 1): Wilcoxon rank-sum test p-values comparing SARS-CoV-2 sgRNAs' residency predictions against those of human transcripts without significant measured localization. All p-values are Holm-adjusted. Values that exceed our significance cutoff of 0.05 are in bold. The SARS-CoV-2 sgRNA residency predictions towards the mitochondrial matrix and nucleolus both have consistently significant p-values, indicating that their predictions are significantly higher than that of unlocalized transcripts (for each respective compartment), suggesting significant predicted residency.

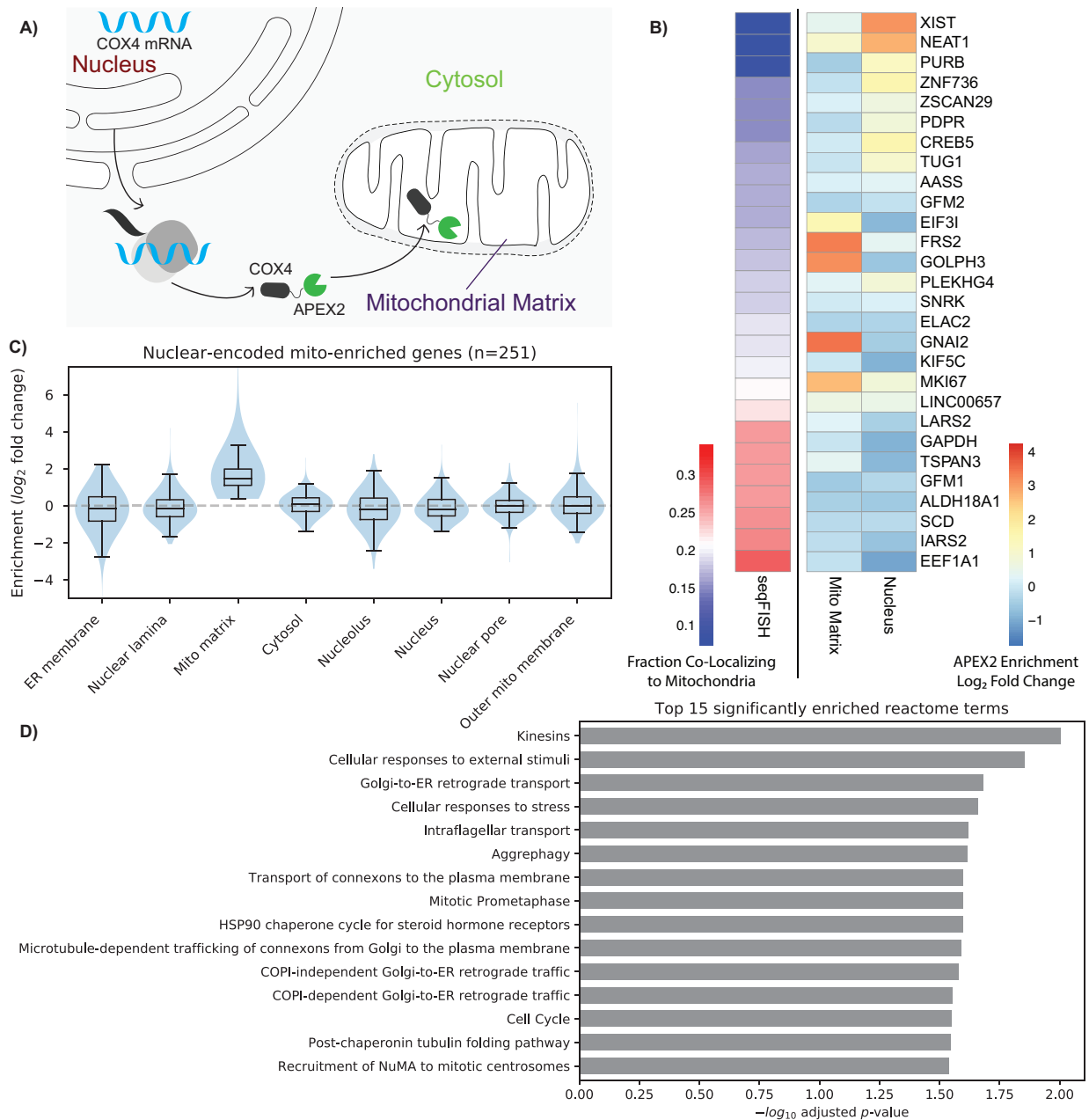


Figure S1 (related to Figure 2): Analysis of APEX-seq mitochondrial transcripts used to train RNA-GPS. (A) COX4 is a nuclear-encoded protein that localizes within the mitochondria (Richter-Dennerlein et al., 2016), and is used to localize APEX to the mitochondria as shown in this illustration. Many transcripts thus picked up by COX4 that nominally localize at the mitochondrial matrix are actually nuclear-encoded. We hypothesize that these are picked up as the APEX2-COX4 fusion is transported from cytosol to mitochondria (final arrow). (B) Sequential FISH data showing fraction of transcripts colocalizing at the mitochondria (using the mitochondrial-resident MTND5 RNA as a mitochondrial marker, as described in (Fazal et al., 2019)). Nuclear transcripts like XIST and NEAT1 do not show mitochondrial enrichment, while transcripts known to localize to the outer surface of the mitochondria like SCD and IARS2 are enriched, providing negative and positive controls, respectively. Within this range, “non-canonical” nuclear-

encoded transcripts like GOLPH3 show intermediate FISH values. This confirms their presence, which likely arises as these transcripts are labelled in the cytosol while COX4 makes its way to the mitochondria. (C) Shows a plot of APEX-seq log fold-change enrichment scores at each compartment for the 251 mitochondrial-enriched, nuclear-encoded “non-canonical” transcripts used to build RNA-GPS. We see that these transcripts have enrichment centered around 0 for all but the mitochondrial matrix, indicating that while these transcripts are nuclear-encoded, the APEX-seq labelling technology consistently and uniquely associates them with the mitochondrial matrix, and are thus not noise. These transcripts are also biologically meaningful, as shown by a reactome ontology analysis of the 100 most enriched (by p-value) nuclear-encoded mitochondrial matrix transcripts (D). There is a clear emphasis for cytoskeletal and intracellular transport terms (e.g. kinesins, post-chaperonin tubulin folding pathway, recruitment of NuMA to mitotic centrosomes; adjusted $p < 0.05$). This supports the interpretation that many of these non-canonical transcripts are picked up as the APEX-seq protein is itself trafficked to the mitochondria.

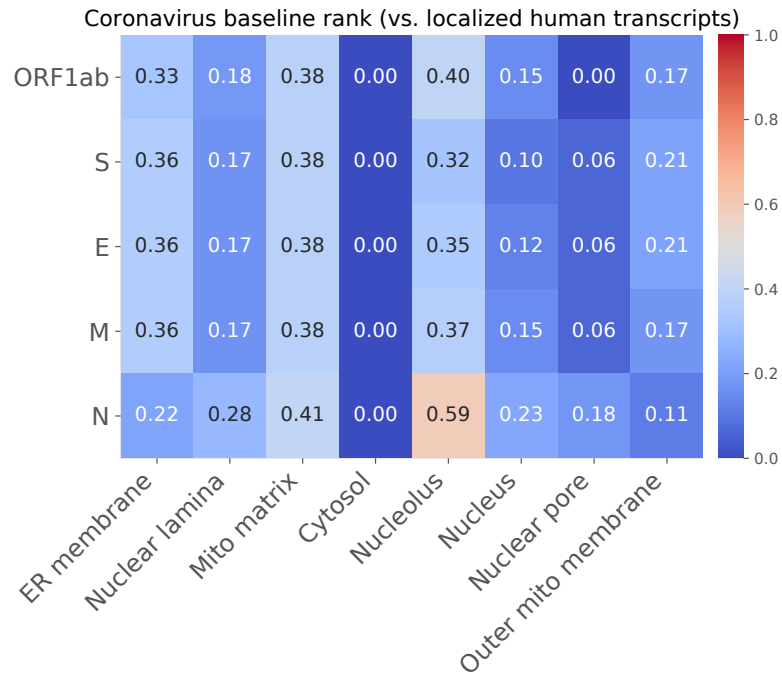


Figure S2 (related to Figure 1): Summary of residency patterns aggregated across all transcripts comprising the human coronavirus baseline. We see that coronaviruses in general primarily exhibit residency towards the nucleolus, mitochondrial matrix, and ER membrane – a pattern similar to that seen in SARS-CoV-2's sgRNAs (albeit less dramatic).

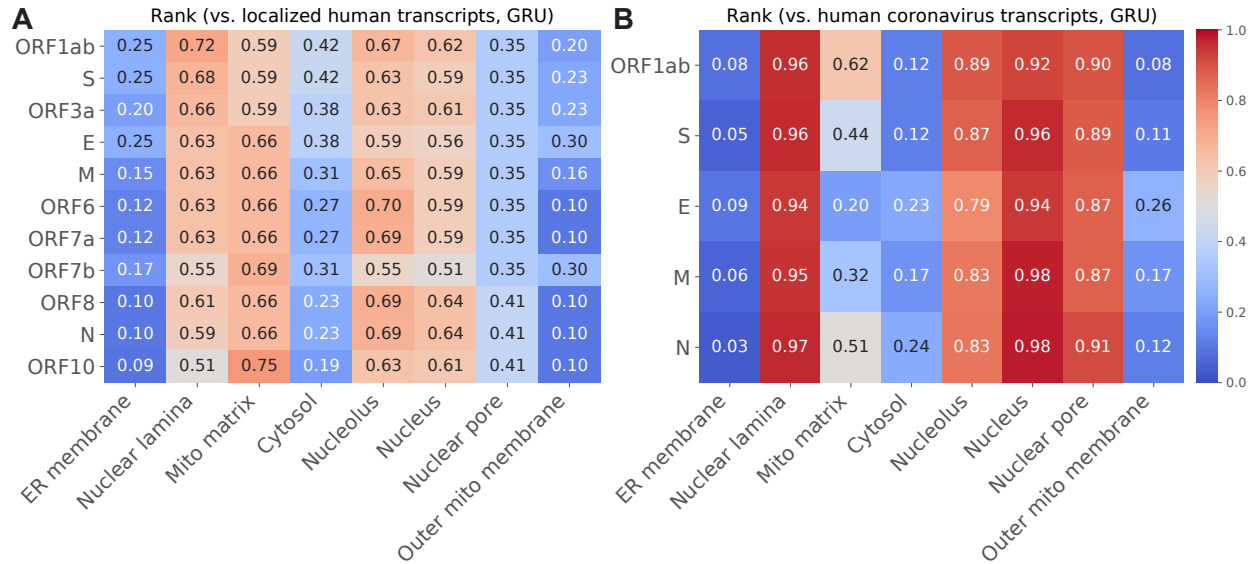


Figure S3 (related to Figure 1): Heatmaps of rank scores of SARS-CoV-2 residency predictions, relative to localized human transcripts (A) and other coronavirus genomes (B), according to a deep-learning recurrent model (GRU). This model takes a very different computational approach to predicting residency compared to RNA-GPS, and thus serves as an orthogonal computational support of results covered in our primary figures. (A) Recapitulates that mitochondrial matrix and nucleolus are among the two most prominent residency signals for SARS-CoV-2. (B) Recapitulates that compared to other coronaviruses, SARS-CoV-2 generally exhibits a stronger nuclear residency signal.

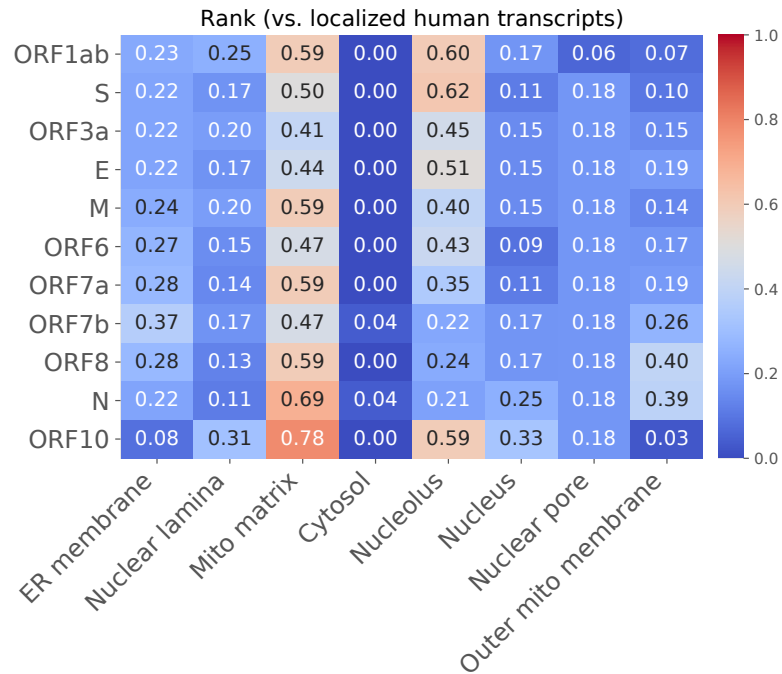


Figure S4 (related to Figure 1): Residency of negative strand sgRNA precursors. Figure 1C shows that the positive strand sgRNA transcripts tend to exhibit residency towards the mitochondrial matrix and nucleolus. Here, we look at the negative-strand precursors to those sgRNAs and observe that these transcripts share similar mitochondrial matrix and nucleolus residency patterns. This suggests another layer of conservation of this predicted residency signal.