Genes with 5′ terminal oligopyrimidine tracts preferentially escape global suppression of translation by the SARS-CoV-2 NSP1 protein

Shilpa Rao, Ian Hoskins*, P. Daniela Garcia*, Tori Tonn, Hakan Ozadam, Elif Sarinay Cenik, Can Cenik

[1] Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712, USA

[*] Equal contribution

Key words: SARS-CoV-2, NSP1, MeTAFlow, translation, ribosome profiling, RNA-Seq, 5′ TOP, Ribo-Seq, gene expression

Abstract

Viruses rely on the host translation machinery to synthesize their own proteins. Consequently, they have evolved varied mechanisms to co-opt host translation for their survival. SARS-CoV-2 relies on a non-structural protein, NSP1, for shutting down host translation. Despite this, it is currently unknown how viral proteins and host factors critical for viral replication can escape a global shutdown of host translation. Here, using a novel FACS-based assay called MeTAFlow, we report a dose-dependent reduction in both nascent protein synthesis and mRNA abundance in cells expressing NSP1. We perform RNA-Seq and matched ribosome profiling experiments to identify gene-specific changes both at the mRNA expression and translation level. We discover a functionally-coherent subset of human genes preferentially translated in the context of NSP1 expression. These genes include the translation machinery components, RNA binding proteins, and others important for viral pathogenicity. Importantly, we also uncover potential mechanisms of preferential translation through the presence of shared sites for specific RNA binding proteins and a remarkable enrichment for 5′ terminal oligo-pyrimidine tracts. Collectively, the present study suggests fine tuning of host gene expression and translation by NSP1 despite its global repressive effect on host protein synthesis.

Introduction

Translation of viral mRNAs is a key step in the life cycle of all viruses, including SARS-CoV-2, a member of the Coronaviridae family with a positive-sense RNA genome. Viruses rely on the host translation machinery and have evolved mechanisms to divert it to translate their own mRNAs (Stern-Ginossar et al., 2019). SARS-CoV-2 encodes a protein, NSP1, which is thought to achieve this function by inhibiting translation of host genes. NSP1 is a non-structural protein of 180 amino acids which is formed by the proteolytic cleavage of a precursor polypeptide (Narayanan et al. 2015; Thoms et al. 2020). SARS-CoV-2 NSP1 shares 84% amino acid identity with that of SARS-CoV NSP1, suggesting its functional conservation (Thoms et al. 2020).

SARS-CoV NSP1 inhibits translation by binding to the 40S ribosomal subunit, thereby reducing translation initiation (Kamitani et al., 2006, 2009). Similarly, structural analysis of SARS-CoV-2 NSP1 revealed its ability to dock into the mRNA binding tunnel of the 40S ribosomal subunit to exclude mRNA binding (Thoms et al. 2020; Schubert et al. 2020, Lapointe et al., 2020). Additionally, NSP1 stably associates with intermediate states of the translation initiation complex and also with distinct sets of translationally inactive 80S ribosomes (Thoms et al. 2020). However, little is currently known about how this inhibition shapes the host gene expression profile. One critical question raised by these studies is whether translation of all host genes is impacted to a similar extent upon NSP1 expression or if certain host genes, perhaps those important for viral replication, can preferentially escape this repression.

Proteomic analysis of SARS-CoV-2 infected cells reported modest changes in global translation activity. Yet, functionally-related proteins in several pathways involved in translation, splicing, proteostasis, nucleic acid metabolism, and carbon metabolism were differentially impacted. In addition, translation inhibition with cycloheximide treatment was proposed as a possible anti-viral treatment, suggesting that the translation machinery could be a potential drug target (Bojkova et al., 2020). Given the critical role of NSP1 in modulating the translation machinery, a comprehensive analysis of host gene expression and translation is necessary to evaluate the role of NSP1 in viral pathogenesis.

In addition to its role in translation inhibition, SARS-CoV NSP1 has been shown to mediate host endonucleolytic mRNA cleavage while simultaneously protecting its own mRNAs via a 5′ untranslated leader sequence (Huang et al., 2011; Nakagawa et al., 2018; Tanaka et al., 2012). *In vitro* translation experiments with SARS-CoV NSP1 mapped the cleavage sites within the first 30-nt of many capped mRNA templates, suggesting that the cleavage occurs immediately after recruitment of the NSP1-bound 40S ribosomal subunit to the transcript (Huang et al., 2011).

Moreover, mutations in the N-terminal region of SARS-CoV NSP1 that abolish mRNA cleavage do not impact its translation inhibition function, thereby ruling out the possible dependence of translation regulatory function on its mRNA cleavage activity (Lokugamage et al., 2012). Importantly, the extent to which SARS-CoV NSP1-catalyzed mRNA cleavage activity shapes host mRNA expression, and whether a similar mechanism applies to SARS-CoV-2 NSP1, remains unknown.

In this study, we introduce a novel method called MeTAFlow to analyze the translation of cells in response to ectopic expression of SARS-CoV-2 NSP1. We demonstrate that NSP1 globally reduces nascent polypeptide synthesis and total mRNA levels in an expression dependent manner. To identify whether all genes are affected similarly in response to the global suppression of protein synthesis, we carried out matched RNA-Seq and ribosome profiling experiments. Surprisingly, functionally related genes—including components of the translation machinery, those involved in viral replication, the host immune response, protein folding chaperones, nucleocytoplasmic transport, and mitochondrial components preferentially escape NSP1-dependent translation inhibition. Further, we uncover a potential role of Poly(rC)-binding proteins in translational regulation upon NSP1 expression. Most importantly, the highly translated genes overwhelmingly have 5′ terminal oligopyrimidine (TOP) motifs, suggesting a mechanism for their selective translational response.

Together, our results show that NSP1 globally decreases translation in accordance with its expression, but specific host genes avoid this suppression through shared regulatory features. These results reveal how the SARS-CoV-2 may hijack the host translation machinery to suppress global translation while promoting viral translation and escape from host immune responses.

Results

SARS-CoV-2 NSP1 reduces host protein synthesis and mRNA content in an expression dependent manner

To analyze whether expression of SARS-CoV-2 NSP1 affects translation and mRNA abundance in host cells, we developed a FACS-based assay called MeTAFLow (Measurement of Translation Activity by Flow cytometry) **(Figure 1A).** This method uses flow cytometry to measure single cell translation as the ratio of nascent polypeptide abundance to poly-adenylated mRNA content. To quantify nascent polypeptide synthesis, we leveraged an analogue of puromycin, O-Propargyl Puromycin (OPP), that incorporates into growing polypeptide chains, releases the polypeptide from the translation machinery, and terminates translation (Liu et al., 2012). In fixed cells, the nascent polypeptides with incorporated OPP are fluorescently-labeled via a copper-catalyzed

azide-alkyne cycloaddition (CuAAC) reaction that labels the alkyne modification of OPP with an azide-conjugated fluorescent probe. To measure mRNA abundance, we designed a molecular beacon (MB) targeting the poly(A) tails of mRNAs. Molecular beacons are hairpin-shaped oligos with a fluorophore and quencher at close proximity (Tyagi and Kramer, 1996). Poly(A)-bound MBs consequentially result in fluorescence, which is proportional to the target mRNA concentration. While OPP labeling and MBs have each been used extensively, our goal was to develop an approach that could combine these two modalities in order to report translation activity from individual cells.

To benchmark MeTAFlow, we carried out several control experiments using both *in vitro* and cellular tests. First, to determine the sensitivity of our MB, an *in vitro* assay was performed by incubating the MB with varying concentrations of a synthetic oligo, N10A40 (10 nucleotides of randomized bases followed by 40 adenines). With increasing N10A40, we observed a linear increase in fluorescence representative of the typical range of mRNA concentration of a human cell (~200 nM) **(Supplementary Figure 1A)**. Further, to estimate the specificity of the MB in the cellular context, fixed cells were incubated with a MB of a random sequence (non-targeting) or an MB targeting the poly(A) tails of mRNAs (targeting). A high signal to noise ratio (~50X) was observed, which is indicative of the specificity of the designed MBs for cellular mRNAs **(Supplementary Figure 1B)**. Additionally, MeTAFlow analysis of different cell lines (HEK293T, K562, Calu-3, and CaCo-2) showed a linear relationship between the observed nascent protein levels and the total mRNA abundance (as estimated by the OPP signal and MB signal, respectively) **(Figure 1 and Supplementary Figure 2)**, further suggesting the effectiveness of the assay in estimating the translation activity of cells of different origin.

To investigate the global effect of SARS-CoV-2 NSP1 expression on host translation and mRNA abundance, NSP1 and NSP2 were independently expressed in HEK293T (human embryonic kidney cell line) cells for 24 hours followed by MeTAFlow analysis to assess their effect on translation activity. NSP2, a SARS-CoV-2 nonstructural protein with an unknown role, was used as a transfection control alongside untransfected HEK293T cells. MeTAFlow analysis of NSP1-expressing cells revealed reduced polypeptide synthesis and total mRNA levels as compared to the untransfected control cells **(Figure 1B,D).** NSP2-expressing cells were similar in both nascent polypeptide and mRNA abundance to the untransfected control cells, and therefore was used as a control in all our further analyses **(Supplementary Figure 3)**. Further, the polypeptide and mRNA reductions seen in NSP1-expressing cells could be directly correlated with the abundance of NSP1 in the cells as measured by StrepTactin-based detection of the Strep-tag fused to NSP1 **(Figure 1E)**. Specifically, cells that expressed higher amounts of NSP1

experienced the most significant reductions in polypeptide synthesis **(Figure 1F)** and also had lower total mRNA levels as estimated by the MB signal **(Figure 1G)**. However, no such effect was observed for NSP2-expressing cells. Our MeTAFlow assay therefore suggests a global downshift in the translation and total mRNA abundance of cells expressing NSP1. Given the viral reliance on host factors for their own replication, it is unclear as to how the cells continue to support viral replication despite global reductions in translation.

Ribosome profiling with matched RNA sequencing from cells expressing NSP1 reveals unaltered distribution of ribosomes across transcripts

Considering previous structural studies and our results with ectopically-expressed NSP1, we expected a depletion of translationally active ribosomes in NSP1-expressing cells. While mRNAs within the cell may be equally suppressed in their translation activity in the presence of NSP1, a reduced pool of active ribosomes can potentially have gene-specific impacts (Liu et al., 2017; Mills and Green, 2017; Raveh et al., 2016). To further analyze the effect of NSP1 on translation at both global and gene-specific levels, we carried out matched ribosome profiling and RNA sequencing (RNA-Seq) analysis using external spike-in controls (Munro et al., 2014).

We generated three replicates each of ribosomal profiling and RNA-Seq analysis from HEK293T cells expressing NSP1, NSP2, or no exogenous protein (untransfected). We avoided rRNA depletion to limit confounding variability that might bias the library composition. First, we assessed the technical quality of our RNA-Seq experiments by comparing replicate-to-replicate similarity and clustering of samples using transcript-level quantifications **(Supplementary Figure 4,5)**. High replicate similarity and sample clustering by treatment group suggests that technical variability was lower than biological variability between conditions.

Similarly, quality metrics for ribosome profiling revealed the expected tight length distribution **(Supplementary Figure 6A)**, robust three nucleotide periodicity, and characteristic footprint enrichment around the start and stop codons **(Supplementary Figure 6B,C)**. The vast majority of ribosome profiling reads originated from the coding regions of annotated transcripts **(Supplementary Figure 7)**. Replicate experiments for each condition were highly similar when comparing the number of coding region mapping reads for each transcript (Spearman correlation rho >0.99; **Supplementary Figure 8**). Replicates of each condition cluster together suggesting that biological variability was higher among conditions than between replicates **(Supplementary Figure 9)**. Taken together, these analyses suggest highly reproducible measurements of gene-specific ribosome occupancy and RNA expression.

While most previous studies implicated NSP1 in translation inhibition, structural analyses also uncovered several 80S structures with NSP1, raising the possibility that NSP1 may have additional roles on modulating other stages of translation. To understand whether NSP1 leads to major changes in translation at the elongation phase, we first analyzed the distribution of ribosome footprints along mature mRNAs at each nucleotide position; an analysis that was aided by incorporating 12-nucleotide UMI (unique molecular identifier) sequences at the 5' ends of the ribosome footprints. The incorporation of UMIs in sequencing libraries enables unambiguous identification of reads arising from unique molecules as opposed to those generated by PCR duplication.

Additionally, when investigating ribosome distributions, we restricted our analyses to 1493 transcripts with the highest density of ribosome footprints per nucleotide (see Methods for details; **Supplementary Figure 10, Supplementary Figure 11 and Supplementary Table 2**). Given that nucleotide-resolution ribosome footprint analyses are sufficiently powered for a subset of highly covered transcripts, we calculated the Pearson correlation of the nucleotide coverage for each transcript in all pairwise combinations of experiments and compared the distribution between replicates of the same condition or across conditions **(Figure 2A,B, Supplementary Figure 12A,B, Supplementary Figure 13)**. We observed a median correlation coefficient of 0.60 for replicates of NSP1, which compares favorably to several previously published studies (Diament and Tuller, 2016). Also, we found that the median correlation between one of the NSP1 replicates and all NSP2 experiments was 0.63 **(Figure 2B)**. These data suggest that the overall distribution of ribosomes across the most abundant transcripts between NSP1 replicates is as similar as the distribution between NSP1 and control (NSP2) cells.

Given the overall similarity of ribosome distribution between NSP1- and NSP2-expressing cells, we next wondered whether there were any specific transcript locations with evidence for differential ribosome speed or pausing. Ribosome pausing is an important factor in protein homeostasis and is known to regulate protein folding (Buskirk and Green, 2017; Collart and Weiss, 2020). We observed that the vast majority of the 6339 candidate pause sites (see Methods) overlap between untransfected controls and NSP1- or NSP2-expressing cells (**Supplementary Figure 14A**). Furthermore, very few candidate pause sites with high read counts were uniquely identified in any particular condition (**Supplementary Figure 14B**) suggesting that there is not a major rearrangement of ribosome pausing events across transcripts in cells expressing NSP1 as compared to controls.

One common candidate pausing event was observed in the *OAZ1* gene, a gene demonstrated to be regulated by ribosomal frameshifting (Ivanov et al., 2018). Similar to many

viruses, coronaviruses including SARS-CoV rely on ribosome frameshifting to synthesize viral proteins (Atkins et al., 2016; Dinman, 2012; Irigoyen et al., 2016; Plant and Dinman, 2008; Su et al., 2005). Hence, we wondered whether there was any evidence of frameshift in NSP1-expressing cells. We detected a prominent peak of longer footprints at codons Ser-68 and Asp-69 which are immediately upstream of the known frameshift site **(Supplementary Figure 15A,C)**. However, we did not detect any frameshift after the pause site (**Supplementary Figure 15B**) suggesting that NSP1 expression did not lead to frameshifting.

We then focused on identifying pause sites that might quantitatively differ between conditions. We found only a small subset of modest quantitative pausing changes (55 in total), including COX7A2L **(Figure 2C, Supplementary Figure 12C, Supplementary Tables 4 & 5)**, a mitochondrial enzyme that is regulated by mTOR through a 5' TOP element. Taken together, our results suggest that NSP1 does not alter the broad distribution of ribosomes on transcripts in agreement with its proposed mechanism of action at the translation initiation step.

Ribosome profiling from NSP1-expressing cells reveals differentially preferential translation of transcripts involved in protein synthesis and folding

Having established that the ribosome occupancy distribution at nucleotide resolution remains unaltered in NSP1-expressing cells, we next sought to identify any gene-specific changes both at the RNA and translation level using our RNA-Seq and ribosome profiling data. Differential RNA expression analyses with ERCC spike-ins revealed that our study is well powered to detect 2-fold changes in RNA expression (Point Estimate of Limit of Detection of Ratios; 130 average read count; **Supplementary Figure 16**; see Methods). We identified 810 transcripts with differential RNA expression (5% False Discovery Rate; **Supplementary Table 6**). Most changes had a relatively small magnitude; only 100 genes exhibited a minimum absolute fold change greater than two (**Supplementary Table 6)**. Among the genes with increased RNA expression, there was a significant enrichment for those associated with mRNA processing/splicing and histone methyltransferase activity (**Figure 3A**; **Supplementary Table 7)**. Interestingly, a previous study exploring proteomic changes upon SARS-COV-2 infection also identified splicing components to have increased protein abundance (Bojkova et al., 2020). Our RNA-Seq data also identified many ribosomal protein genes to have lower expression compared to the control. This is in accordance with another transcriptomics study with NSP1 that reports similar observations (Yuan et al., 2020). Taken together, our results reveal that NSP1 expression remodels RNA expression and leads to modest gene-specific RNA expression changes.

We then took advantage of ERCC spike-in mix controls to test whether NSP1 expression leads specifically to a depletion of the mRNA fraction. We calculated the jackknife estimate (Quenouille, 1949; Efron and Stein, 1981) of the ratio of the reads for each spike-in RNA while considering the replicate-to-replicate variability in these ratios (see Methods for details). If the input mRNA fraction were similar, we would expect these ratios to be highly similar across the pairwise comparisons. This analysis revealed that NSP1-expressing cells had lower mRNA content compared to untransfected controls but not compared to cells expressing NSP2 (**Supplementary Figure 16**).

Next, to identify gene-specific changes at the translational level, we determined which transcripts displayed significant changes in ribosome occupancy while controlling for RNA abundance (Methods). This metric is typically referred to as translation efficiency (TE) and we adopt this nomenclature for the rest of the study. We identified 177 transcripts with differential TE when comparing NSP1-expressing cells to those expressing the control viral protein (NSP2) (**Figure 3B-C; Supplementary Table 8**). Interestingly, 166 of these 177 transcripts had higher TE in NSP1-expressing cells (referred to as the “high-TE” set). It is essential to consider the compositional nature of sequencing studies, hence these results should be interpreted in relative terms and do not indicate absolute changes (Quinn et al., 2018). In other words, transcripts with relatively high translation efficiency in NSP1-expressing cells could have lower absolute rate of protein synthesis even though they have higher translational output compared to other genes.

Remarkably, the set of 166 transcripts with high-TE in NSP1-expressing cells were >690-fold enriched for components of the ribosome (p-value $1.34 \times 10_{-55}$ and $3.54 \times 10_{-85}$ for small and large subunits, respectively; **Supplementary Table 9**). In fact, 29 ribosomal small subunit and 43 ribosomal large subunit proteins were among the high-TE set. Furthermore, four members of eukaryotic elongation factor complex 1 (EEF1A1, EEF1A2, EEF1B2, EEF1D), eukaryotic elongation factor 2 (EEF2), 8 translation initiation factors (EIF2A, EIF3E, EIF3F, EIF3G, EIF3H, EIF3L and EIF4B) and methionyl-tRNA synthetase (MARS) had increased relative translation efficiency in cells expressing NSP1. In total, 86 of the 166 genes with high translation efficiency were components of the host translation machinery (**Figure 3B-C**).

Interestingly, we also identified *NOB1*, *NHP2* and *BOP1* among the high-TE genes **(Figure 2)**. These genes are involved in rRNA processing (Henras et al., 2015; Sloan et al., 2017), suggesting that proteins involved in ribosome biogenesis, in addition to structural constituents of ribosomes, escape NSP1-mediated translational repression. In addition, three chaperones—*BAG6*, *HSPA8*, and *HSP90AB1*—were among high-TE genes. These chaperones may play critical roles as cells producing viral protein may require sustained chaperone activity to ensure

proper folding (Aviner and Frydman, 2020; Binici and Koch, 2014; Wan et al., 2020). Taken together, NSP1 expression in human cells is associated with a remarkably coherent translational program that differentially sustains the availability of host translation machinery and protein folding capacity.

While the translation components predominate the list of high-TE genes, we identified other genes with potential roles in the viral life cycle. These included translation regulatory RNA binding proteins, such as cold shock domain containing E1 (CSDE1) (Unr), Caprin1, poly(rC) binding protein 2 (PCBP2), Staufen-1 (STAU1), FXR1 and DexH-box helicase 30 (DHX30) (Antonicka and Shoubridge, 2015; Brocard et al., 2020; Gaete-Argel et al., 2019; Guo et al., 2020). Genes involved in the maintenance of iron homeostasis, including both the ferritin heavy (FTH1) and light (FTL) chains, were observed to have high-TE. Additionally, components of the nucleocytoplasmic transport such as IPO5, IPO7 and THOC5 showed high-TE. IPO5 and IPO7 are importins involved in the import of some ribosomal proteins and other non-ribosomal substrates (Fassati et al., 2003; Hutchinson et al., 2011). Another interesting group of genes in the high-TE set are mitochondrial components, including genes of the mitochondrial solute carrier family (SLC): SLC25A3, SLC25A5, SLC25A6 and SLC25A15 (Jakel, 1998). Mitochondrial SLC genes are involved in the exchange of solutes such as ADP/ATP or phosphate from the cytoplasm and through the inner mitochondrial membrane (Nury et al., 2006). Furthermore, we also identified components of the translocase of the outer mitochondrial membrane (TOMM), TOMM22 and TOMM40, among the high-TE set. These two proteins are essential for mitochondrial import of precursor proteins (Kravic et al., 2018). Most of the above differentially expressed genes with high-TE are implicated in other viral infections (Chen et al., 2020; Dixit et al., 2016; Fassati et al 2003; Labaronne et al 2017; Lenarcic et al., 2013; Liu et al., 2010; Sun et al., 2018; Ye et al., 2010). However, characterization of their precise functions in response to SARS-CoV-2 infection remains to be elucidated.

Overall, only 11 genes displayed lower translation efficiency in NSP1 expressing cells than in the control. These included *ATF4*, which is a key regulator of the cellular adaptive stress response downstream of eIF2α phosphorylation (Wek et al., 2006). Genome wide knockout studies of SARS-CoV-2 infected cells suggests GCN2 to be an antiviral protein (Wei et al., 2020) whose activation causes eIF2α phosphorylation and increased translation of *ATF4*. Given this link, we speculate a GCN2 inactivation in response to viral infection. This may be a strategy for the virus to evade PERK-ATF4-CHOP driven apoptosis (Ikebe et al., 2020), or alternatively, avoid eIF2α phosphorylation through an independent stress response pathway (Fraser et al., 2016). However, we note that ATF4 may also be proviral (Caselli et al., 2012; Lee et al., 2018; Qian et

al., 2012). Other low-TE genes include *FAXDC2*, *STYX*, *SHOC2* which are involved in the Ras/Raf/Mitogen-activated protein kinase/ERK Kinase (MEK)/Extracellular-signal-Regulated Kinase (ERK) cascade (Jin et al., 2016; Reiterer et al., 2013; Rodriguez-Viciana et al., 2006). Importantly, it is commonly involved in viral infection, including that of coronaviruses (Kumar et al., 2018).

Shorter CDS and 3' UTR lengths and less structured 5' UTRs may confer higher translation efficiency upon NSP1 expression

We hypothesized that transcripts with higher translation efficiency under NSP1 expression could share common sequence features that facilitate their selective translation. We first compared features for high-TE genes, low-TE genes, and control genes with no evidence of differential translation efficiency (non-DE genes). We compared the CDS and UTR lengths, GC content, and predicted secondary structure at the 5' end of the transcript and surrounding the translation initiation site (TIS) (see Methods, **Supplementary Table 10**). Information content at the TIS revealed consensus Kozak sequences for both the high-TE and non-DE genes.

We found that the CDS and 3' UTR lengths were significantly shorter among the high-TE versus non-DE genes **(Figure 4A)**, with a mean length difference of 574 nt and 633 nt for the CDS and 3' UTR, respectively **(**Dunn's adjusted p-values 2.29×10^{-15} and 1.11×10^{-14}). The GC content of the 5' UTR was also slightly lower for the high-TE genes compared to the non-DE genes (**Figure 4B**, 62.7% for high-TE and 66.2% for non-DE, p-value=2.04×10^{-5}). Because secondary structure in the 5' UTR or near the TIS is known to regulate translation initiation, we compared the predicted minimum free energy of folding between the high-TE, low-TE, and non-DE genes for a 60 nt window at the start of the transcript or spanning the TIS **(Supplementary Figure 17)**. We found that the high-TE genes were less structured than non-DE genes at the 5' terminus (Kruskal-Wallis p-value=0.008; Dunn's Z=2.93; adjusted p-value=0.010) and at the TIS (Kruskal-Wallis p-value=0.001; Dunn's Z=2.81; adjusted p-value=0.015).

Given that GC content is entwined with secondary structure, we repeated these comparisons while controlling for differences in feature covariates using nonparametric matching (Ho et al., 2007) (**Supplementary Table 11,** Methods). With covariate matching, we found that the CDS and 3' UTR lengths were still negatively associated with the high-TE as compared to the matched, non-DE genes while controlling for one another (Wilcoxon p-values 2.14×10^{-6} and 9.29×10^{-9}, respectively). High-TE genes had a mean CDS length of 877 nt compared to 1312 nt for matched non-DE genes, and exhibited >3-fold shorter 3' UTRs (mean length 368 nt and 1034

nt). Therefore, a shorter CDS and 3' UTR were specifically associated with higher translation efficiency upon NSP1 compared to NSP2 expression.

In contrast, when matched on the length and GC content of the rest of the transcript (CDS, 3' UTR), 5' UTR GC content was no longer significantly different between gene sets. Similarly, with matching on length and nucleotide content, the 5' UTR terminus secondary structure also became nonsignificant. Nonetheless, after matching, the secondary structure at the TIS was lower for high-TE genes compared to non-DE genes (mean of -18.5 kcal/mol for non-DE genes and mean of -16.4 kcal/mol for high-TE genes; Wilcoxon p-value 0.015). Altogether, although several sequence features showed significant differences between high-TE and non-DE genes, when matching, we found that only the length of the CDS and 3' UTR and secondary structure near the TIS were associated with high-TE. We note these strong differences in length may simply be an inherent characteristic of ribosomal genes (Lima et al. 2017).

Preferentially translated genes have more RNA binding protein sites in the 3' UTR and may be regulated by PCBPs

In an effort to identify possible *trans*-regulatory factors influencing differential translation efficiency, we next characterized RNA binding protein sites in the high-TE and matched non-DE gene subsets. We reasoned that any enriched RBP sites in the high-TE transcripts may be important for their regulation and may suggest potential RBP regulators of SARS-CoV-2 translation. We selected six control gene sets with matched transcript lengths and GC content compared to the high-TE geneset. The 5' UTR and CDS of high-TE genes harbor fewer RBP sites, as defined in oRNAment database (Benoit Bouvrette et al., 2020), compared to control non-DE genes. Conversely, in the 3' UTR, high-TE genes exhibit enrichment for most RBPs **(Figure 4C, Supplementary Table 12)**. To identify candidate RBPs that may regulate translation concomitant with NSP1 expression, we determined which RBPs showed greater than or less than 2-fold sites in high-TE genes compared to matched non-DE gene sets. Many RBPs known to influence mRNA processing, export, splicing, and post-transcriptional regulation were identified in each transcript region **(Figure 4D, Supplementary Table 13-14)**.

Among RBPs with increased CDS sites in high-TE genes compared to controls were A1CF, a complementary factor of the APOBEC1 RNA editing complex (Blanc et al., 2019); RBM45, a RBP with roles in the DNA damage response and B19 Parvovirus splicing (Wang et al., 2020); and NUPL2, a nuclear pore protein involved in mRNA export (Kendirgi et al., 2005) and HIV-1 docking to the nuclear membrane (Le Rouzic et al., 2002). Apart from A1CF, a number of other RBP sites that were differential between genesets (TARDBP, FXR1, ILF2) suggested a

possible link to ADAR RNA editing (Quinones-Valdez et al., 2019). RNA editing in SARS-CoV-2 was recently suggested by Nanopore sequencing (Kim et al., 2020) and from mutation frequency data in natural isolates (Di Giorgio et al., 2020).

To explore the possibility that APOBEC1 C>U editing or ADAR1 A>I editing may regulate translation of host genes, we enumerated substitutions in the RNA-Seq data for the high-TE and non-DE coding and UTR regions. Comparing frequencies of relevant substitutions, we found no evidence of NSP1/2-induced RNA editing as a mechanism for high translation efficiency of host genes **(Supplementary Figure 18)**. This was despite slight frequency differences between the high-TE and non-DE genes, which are likely explained by variable sequence contexts of the loci.

Interestingly, in the 5' UTR, only a single RBP, PCBP1, had a higher number of binding sites among high-TE genes than non-DE genes. PCBP1 binds poly-cytosine motifs and has an established role in picornavirus replication and viral IRES-dependent translation initiation (Li et al., 2013, 2019; Luo et al., 2014). PCBPs also lead to ribosomal frameshifting in porcine reproductive and respiratory syndrome virus (Napthine et al., 2016). Further, this RBP is known to control mitochondrial antiviral signaling (Zhou et al., 2012), T-cell differentiation (Ansa-Addo et al., 2020), and tumorigenesis (Shi et al., 2018), highlighting its important role in both innate and adaptive immunity.

To further investigate the potential role of candidate RBPs in the regulation of host or viral translation efficiency, we selected RBPs with differential sites (PCBP1, FUBP3, TARDBP) and those RBPs which themselves had high translation efficiency (PCBP2, FXR1) for which there was enhanced crosslinking immunoprecipitation (eCLIP) data (Van Nostrand et al., 2020). Consistent with greater PCBP1 sites in high-TE genes, we confirmed a greater number of PCBP1 and PCBP2 eCLIP peaks in the high-TE exonic regions compared to controls **(Figure 4E, Supplementary Figure 19)**.

FXR1, an RBP with high-TE itself, also showed increased eCLIP peaks in the high-TE gene set as compared to controls. However, the variance in FXR1 eCLIP peaks across non-DE genesets was high as the fourth non-DE set harbored a gene (*UBR4*) that contained 127 peaks for FXR1. Nonetheless, the median FXR1 peaks across non-DE genesets was lower than the high-TE genes. Fewer peaks for FUBP3 was consistent with fewer predicted sites from the oRNAment database. However, TDP-43 (TARDBP) showed fewer predicted sites in the coding sequence of the high-TE genes but no difference in the number of eCLIP peaks. This indicates that while computational enumeration of RBP sites may help identify regulatory RBPs, it can fail to predict actual binding *in vivo*. Altogether, these results suggest the presence of increased RNA

binding protein sites in the 3' UTR and a potential role of PCBPs and FXR1 in positive regulation of high-TE genes.

NSP1 expression leads to selectively higher translation efficiency through the presence of 5' terminal oligopyrimidine (TOP) tracts

Having characterized the set of transcripts that had relatively increased translation efficiency upon NSP1 expression, we next wondered how NSP1 can mechanistically lead to gene-specific effects when structural analyses indicate a potential non-specific inhibition of translation initiation. While depletion of actively translating ribosomes can have gene-specific effects on translation due to intrinsic differences in translation dynamics across cells (Mills and Green, 2017), NSP1 may also modulate translation of specific transcripts via cooperative interactions with host factors.

To differentiate between these alternative modes of action, we sought to identify any potential mechanism of co-regulation at the translation level. The high-TE gene set contained many ribosomal proteins, which are known to be translationally regulated by 5' TOP motifs, defined by an oligopyrimidine tract of 7-14 residues at the 5' end of the transcript (Philippe et al., 2020). Indeed, when we calculated the pyrimidine content of the 5' UTR of high-TE transcripts, we found an increase compared to that of matched non-DE genes (**Fig 5A**; Wilcoxon p-value 4.35×10^{-4}; 7% increase in mean content).

Next, we investigated whether other 5' TOP mRNAs are in our list of high-TE genes. In addition to ribosomal proteins, the established 5' TOP mRNAs are VIM, TPT1, HNRNPA1, NAP1L1, PABPC1, EIF4B, EIF3E, EIF3F, EIF3H, EEF2, EEF1D, EEF1A1, EEF1B2, EIF3A, and FAU (Philippe et al., 2020) Dataset S1). Each of these genes were among high-TE genes with the exception of EIF3A and FAU. Both EIF3A and FAU, however, had high translation efficiency upon NSP1 expression but did not meet our statistical significance threshold (TE logFC 0.27 and 0.3, p-value 0.003 and 0.004, respectively).

In addition to the annotated 5' TOP mRNAs in the literature, hundreds of other transcripts may behave similarly given their 5' terminal sequence properties. Recent work leveraged transcription start site annotations to derive a 'TOPscore' and identified an expanded set of 5' TOP mRNAs (Philippe et al., 2020) that is predictive of regulation by mTORC1 and LARP1, an established feature of 5' TOP RNAs (Patursky-Polischuk et al., 2014; Tcherkezian et al., 2014). Remarkably, in this extended set of 25 additional mRNAs, a further 8 were in our list of high-TE genes (*CCNG1*, *EIF2A*, *EIF3L*, *IPO5*, *IPO7*, *NACA*, *OLA1*, *UBA52*). Furthemore, transcripts with high TOP scores were dramatically enriched among genes with high-TE **(Fig 5B-C)**.

Next, we aimed to identify whether high-TE genes behave as would be expected from the enrichment of 5' TOP sequences. Specifically, 5' TOP mRNAs are known to be translationally regulated by the mTOR pathway upon mitogen or anoxic stimulation, and also respond to nutrient availability (Meyuhas and Kahan, 2015). Hence, we determined how the high-TE genes change their translation initiation rates in response to amino acid starvation. We analyzed ribosome profiling data generated in the presence of lactimidomycin and puromycin in HEK293 cells (Gao et al., 2015). While lactimidomycin inhibits translation initiation, puromycin leads to polypeptide release from elongating ribosomes. Consequently, this particular antibiotic combination enriches for ribosome signatures on initiation sites, providing a quantitative readout of translation initiation (Gao et al., 2015).

We observed that under normal growing conditions, our set of high-TE genes had higher translation initiation efficiencies (**Fig 5D**; Wilcoxon rank sum test p-value = 0.006). As expected from the enrichment of 5' TOP mRNAs among the high-TE genes, we observed a significant suppression of translation initiation in response to starvation as measured by profiling in the presence of lactimidomycin and puromycin (**Fig 5D;** Wilcoxon rank sum test p-value $=1.3 \times 10^{-11}$). Taken together, these results reveal that the high-TE genes are largely composed of 5' TOP RNAs and show consistent physiological responses to nutrient deprivation.

Discussion

SARS-CoV-2 has emerged as the causative agent of a pandemic that has affected millions of people. A global effort is focused on understanding the viral infection mechanisms in hopes of stopping its spread. One of these mechanisms is the diversion of the host translational machinery that promotes viral replication. SARS-CoV-2 NSP1 is a viral protein that has been implicated in shutting down host translation by directly interacting with the mRNA entry tunnel of ribosomes (Lapointe et al., 2020; Thoms et al. 2020). However, a nonselective translation suppression of all host genes might be detrimental for the virus, which invariably relies on host factors for its lifecycle. In addition to inhibiting translation, NSP1 may alter the host transcriptome by mRNA degradation, as previously proposed for its counterpart in SARS-CoV (Huang et al., 2011; Yuan et al., 2020). Here, we attempted to characterize the changes upon ectopic expression of NSP1 on host mRNA expression and translation using matched RNA-Seq and ribosome profiling experiments.

To establish a model for this purpose, we first expressed NSP1 in HEK293T cells and simultaneously measured nascent polypeptide synthesis and total polyA mRNA abundance at single cell resolution using a novel FACS-based assay (MeTAFlow). We interestingly observed

an NSP1 expression level dependent modulation of both host translation and total mRNA abundance. One caveat with the MeTAFlow method is the potential difference in sensitivity between the modalities used for detection of polypeptide synthesis and mRNA abundance namely fluorescence signal from labelled OPP and the molecular beacon, respectively. For instance, subtle changes in cellular mRNA levels might robustly increase nascent protein production, making detection of newly synthesized proteins more sensitive than changes in the cellular mRNA levels. Further, the sensitivity of the molecular beacon to detect total mRNA can be affected by differences in accessibility or length of the mRNA poly (A) tail.

One caveat of using HEK293T cells as a model is that they are not the primary cell type infected by SARS-CoV-2. However, HEK293T cells are permissive to SARS-CoV-2 infection (Harcourt et al., 2020). Moreover, proteins identified in SARS-CoV-2 interactome studies using HEK293T cells were shown to have their highest expression in the lung tissue compared to others, indicating the relevance of this model for such studies (Gordon et al., 2020).

MeTAFlow in HEK293T cells revealed global changes in translation and mRNA abundance but did not give insight into gene-specific responses. To illuminate any potential gene-specific changes, we further analyze NSP1 effects on the translation and transcriptome of the host cells by ribosome profiling and RNA sequencing studies. A recent ribosome profiling study of SARS-CoV-2 infected Vero and Calu-3 cells revealed the high resolution map of the viral coding regions. However, lack of baseline characterization of uninfected host cells limited the ability to determine host translation response which we address in this study (Finkel et al., 2020).

By coupling ribosome profiling with RNA sequencing, we first show that the distribution of ribosomes remains broadly unchanged along transcripts upon NSP1 expression. This result suggests that elongation dynamics are relatively unaffected supporting the proposed role of NSP1 on translation initiation. Despite reduction in global mRNA and nascent protein abundance, gene specific analysis indicated relative changes of many genes in both directions. A critical aspect of any sequencing-based experiment is the compositional nature of the resulting data (Quinn et al., 2018). In other words, sequencing experiments report on only the relative expressions of the biomolecules analyzed. Therefore, a given transcript with relative increase in response to NSP1 expression could still have lower absolute abundance compared to control conditions. All the conclusions in the current study should be interpreted in this manner.

Despite the global translational shutdown, we identified 166 genes that have relatively high translation efficiency. Strikingly, 86 out of these 166 genes were components of the ribosome and the translation machinery. A proteomics study using SARS-CoV-2 infected cells also suggested an increase in translation machinery and rRNA processing components among other

pathways identified to be differentially regulated in the host (Bojkova et al., 2020). In addition to this, our study also revealed several high-TE genes to be involved in rRNA processing, protein folding, nucleocytoplasmic and mitochondrial transport. Many previous studies implicated these components in other viral infections but of unknown significance with respect to SARS-CoV-2 infection (Aviner and Frydman, 2020; Bianco and Mohr, 2019; Labaronne et al., 2017; Liu et al., 2010). Our study, therefore, indicates a potential role of NSP1 in altering the host gene expression through its impact on translation.

A critical next question is how could specific genes be differentially translated during a global shutdown of initiation. There are two prevailing hypotheses for preferential translation. First, shared regulatory features can specifically enable these genes to escape the translational inhibition. Second, a reduced pool of active ribosomes can potentially have gene-specific impacts due to intrinsic differences in translation efficiency (Fig. 5A) (Liu et al., 2017; Mills and Green, 2017; Raveh et al., 2016).

To test these two hypotheses, we characterized common *cis*-regulatory features of the high-TE genes. We observed that high-TE transcripts harbor shorter 3′ UTRs which is also reported to be correlated with reduced poly(A) tail length (Legnini et al., 2019), and in turn lead to higher translation efficiency (Lima et al., 2017). In addition, we determined a number of RNA binding proteins with differential sites in the high-TE genes. Of the RBPs analyzed, poly(rC) binding proteins (PCBPs) were particularly notable as the number of their binding sites, and the RBPs themselves, were enriched in the high-TE transcripts. PCBP2 is involved in cellular mRNA stability and translation regulation (Makeyev and Liebhaber 2002). Genome-wide CRISPR knockout studies using Vero cells infected with SARS-CoV-2 suggests an antiviral role of *PCBP2*, as knockout sensitises the host cells (Wei et al., 2020). However, PCBPs may also negatively regulate innate immune responses. Sumoylation of PCBP2 during viral invasion causes its nuclear export and degradation of MAVS (a mitochondrial antiviral signalling protein) (Xia et al. 2015). PCBP1 also negatively regulates MAVS (Zhou et al., 2012). Most importantly, PCBPs are known to directly regulate viral translation. Both PCBPs contribute to Flavivirus infection (Li et al., 2013, 2019; Luo et al., 2014) and bind to the 5′ UTRs of the poliovirus and EV71 (picornaviruses) to stimulate their translation and replication, respectively (Gamarnik and Andino, 2000; Li et al., 2013, 2019; Luo et al., 2014). Curiously, we also found the SARS-CoV-2 5′ untranslated leader sequence contains a consensus PCBP2 motif near the 5’ terminus at nucleotides 15-21 (NC_045512.2). This raises the possibility that PCBPs may positively regulate translation of all the SARS-CoV-2 ORFs containing this leader.

Our high-TE set contained many genes known to have intrinsically higher translation efficiency **(Figure 5D)** prompting us to explore the idea that a reduced pool of active ribosomes can potentially have gene-specific impacts due to intrinsic differences in translation efficiency. Strikingly, these same proteins are known to harbor 5′ TOP tracts. In this study, 85 out of 93 known TOP RNAs were identified among the high-TE set. Furthermore, the remaining genes were dramatically enriched for 5′ TOP-like sequences (as defined by (Philippe et al., 2020). La-related protein 1 (LARP1), a RNA binding protein, has emerged as a key regulator of 5′ TOP mRNA stability and 40S ribosomes are required for LARP-1 mediated stability of 5′ TOP mRNAs (Gentilella et al., 2017). Interestingly, our RNA-seq results also revealed that many 5′ TOP bearing ribosomal genes classified in the high-TE set had relatively lower mRNA abundance. In light of the above observations, we speculate that the limited availability of active ribosomes in response to NSP1 expression might be responsible for negatively influencing the stability of 5′ TOP mRNAs in the high-TE set as well as their increased relative translation. However, the precise mechanism by which translation initiation of 5′ TOP mRNAs is favoured under certain conditions is still unclear. 5′ TOP mRNAs along with inherent sequence features (shorter CDS and 3’ UTR lengths, RBP sites) may fine tune translation regulation of this set of genes. Taken together, our results provide support to both hypotheses regarding gene-specific translation changes upon global repression of translation initiation by NSP1 and provide a mechanistic explanation of the observed changes that may prove to be beneficial for viral life cycle.

Finally, we caution that expression of the SARS-CoV NSPs in isolation, with a single time point post-infection, and in a single cell line, may not recapitulate NSP1 translational control during natural infection. NSP1 is expressed early during infection (Ziebuhr, 2005) and studying a time course may portray the complete dynamic changes brought about by NSP1 in host cells. Also given its role in global translation repression the levels of NSP1 might possibly be self-regulating additionally allowing fine tuning of the host gene expression. We also propose that NSP1 expression during SARS-CoV-2 infection is regulated such that it remains lower in abundance than cellular ribosomal pools. This might create a limited pool of active 40S ribosomes which are preferentially used for translation of viral and host mRNA of intrinsically higher translation ability. Overall, our study indicates the possible role of SARS-CoV-2 NSP1 in tweaking the host translation machinery to its advantage.

Acknowledgements

This work was supported in part by the Texas Rising Star Award [to E.S.C] and National Institutes of Health [1R35GM138340-01 to E.S.C; CA204522 to C.C.] and Welch Foundation [F-2027-20200401 to C.C.]. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high performance computing and storage resources that have contributed to the research results reported within this paper. URL: http://www.tacc.utexas.edu.

Author Contributions

Shilpa Rao: Data Curation, Investigation, Methodology, Validation, Formal Analysis, Visualization, Writing- Original Draft, Writing- Review and Editing.
Ian Hoskins: Data Curation, Investigation, Validation, Formal Analysis, Software, Visualization, Writing- Original Draft, Writing- Review and Editing.
Daniela Garcia: Data Curation, Investigation, Validation, Methodology, Formal Analysis, Writing- Original Draft, Writing- Review and Editing.
Tori Tonn: Data Curation, Investigation, Methodology, Formal Analysis, Visualization, Writing- Original Draft, Writing- Review and Editing.
Hakan Özadam: Data Curation, Investigation, Formal Analysis, Software, Validation, Visualization, Writing- Original Draft, Writing- Review and Editing.
Elif Sarinay Cenik: Conceptualization, Resources, Project Administration, Writing- Review and Editing.
Can Cenik: Conceptualization, Data Curation, Formal Analysis, Software, Supervision, Funding Acquisition, Investigation, Methodology, Project Administration, Visualization, Writing- Original Draft, Writing- Review and Editing.

Declaration of interests

The authors declare no competing interests.

Figure Legends

Figure 1 - MeTAFlow assay of HEK293T cells expressing NSP1

(A) Schematic representation of MeTAFlow. Briefly, OPP molecules are incorporated into growing peptide chains then fluorescently labelled via CuAAC reaction. mRNA molecules are labelled

with fluorescent, poly(A)-targeting molecular beacons. Simultaneous measurement of nascent protein and mRNA abundance via their fluorescent signals in single cells is detected with flow cytometry. Translation efficiency is inferred as a ratio of nascent protein to mRNA amounts. **(B)** MeTAFlow analysis of untransfected (unt), **(C)** NSP2-transfected (NSP2), and **(D)** NSP1-transfected (NSP1) HEK293T cells. The populations in panels B-D are OPP-AF488+/MB-Cy5+ cells. Gates show a population of unchanged cells (P1), as established by the baseline untransfected cells, and a second population of cells (P2) with reduced nascent protein and mRNA abundance. The Pearson correlation coefficient (r) for the nascent protein levels to the total mRNA abundance is given for each condition. **(E)** Comparison of NSP1 expression levels in the NSP1-transfected cells' unchanged population (P1, gray) and population with reduced parameters (P2, purple) via NSP1 strep tag detection with Strep-Tactin DY549. **(F)** Nascent polypeptide levels as measured by OPP fluorescence for P1 (gray) and P2 (purple) of NSP1-expressing cells. **(G)** Measurements of total mRNA abundance by molecular beacon signal for P1 (gray) and P2 (purple) of NSP1-expressing cells. The populations in panels E-G are OPP-AF488+/MB-Cy5+ cells. Data shown are representative examples of experiments carried out in duplicates.

Figure 2 - Ribosome occupancy at nucleotide resolution

(A) Histogram of Pearson correlations between ribosome profiling replicates of NSP1-expressing cells. The distribution of the Pearson correlation coefficient was calculated for each pair of NSP1 replicates, for each gene. **(B)** Histogram of Pearson correlations between the first replicate of NSP1-expressing cells and all replicates of NSP2-expressing cells, calculated as in panel A. **(C)** Ribosome occupancy on *COX7A2L*. The mirrored graphs depict the number of ribosome footprints per million across the nucleotide positions of 950 to 1150 nt, in the second replicates of NSP1- and NSP2-expressing cells. A candidate, and quantitatively differential, pause site is highlighted at the nucleotide position 1060. The inset shows ribosome occupancy around the highlighted pause site. This pausing event has a fold change of 2.98 between the replicates of NSP1- and NSP2-expressing cells.

Figure 3 - Differential RNA expression and translation efficiencies upon NSP1 expression

(A) Volcano plot depicting RNA expression changes in HEK293T cells expressing NSP1 as compared to a control viral protein (NSP2). Representative genes belonging to functional groups of proteins with enriched differential RNA expression are highlighted. Non-significant genes are shown in gray. **(B)** Volcano plot depicting translation efficiency (TE) changes in HEK293T cells

expressing NSP1 as compared to a control viral protein (NSP2). Ribosome occupancy differences normalized by RNA expression changes (differential TE) were calculated from the same samples (see Methods). Highlighted genes belong to highly enriched groups of functionally-related genes. **(C)** Counts per million reads (cpm) were calculated for each of the highlighted genes using ribosome profiling and RNA-Seq experiments from NSP1- or NSP2-expressing cells. For each matched experiment pair, the ratio of cpm values were plotted for three replicates. Selected genes from the following categories are highlighted: ribosomal proteins, genes involved in rRNA processing, genes involved in nucleocytoplasmic transport (NCT), translation factors, RNA binding proteins (RBPs), and chaperones.

Figure 4 - Sequence features and RNA binding protein analysis

(A) Lengths of coding sequence (CDS) and 3′ untranslated region (3′ UTR) for high translation efficiency (high-TE), low translation efficiency (low-TE) and non-differentially expressed (non-DE) control genes. Outliers greater than (1.5 x interquartile range) were omitted for clarity. Asterisks indicate significant Dunn's post-hoc tests at a significance level of 0.01. **(B)** 5' untranslated region (5′ UTR) GC content of gene sets. Outlier removal, statistical tests, and significance level were applied as in panel A. **(C)** Distribution of the median log2 fold changes (FC) in RBP sites between high-TE and non-DE gene sets. Across the x-axis, Kernel density estimation of the median log2 FCs across six non-DE gene sets for each RBP site in the oRNAment database is plotted, colored by transcript region. **(D)** Differential RBP motif site analysis for high-TE and matched non-DE gene sets for oRNAment database RBPs. The log2 FC is between high-TE and non-DE genes. Each point corresponds to a unique comparison with a different non-DE control gene set. Blue lines indicate the median log2 FC. The heat scale indicates the mean log2 sites between the high-TE group and each control set. Numerous filters were applied (see Methods). **(E)** Enhanced crosslinking-immunoprecipitation (eCLIP) peaks for candidate RBPs in the high-TE versus non-DE gene sets. Heat scale shows the median of the log2 reproducible peaks found in exonic regions of each gene set in K562 and HepG2 cells. Reproducible peaks were those identified by an irreproducible discovery rate (IDR) (Van Nostrand et al., 2020).

Figure 5 - 5′ terminal oligopyrimidine tracts in the 5' UTRs of differential translation efficiency genes

(A) Pyrimidine content in the 5' UTR for high translation efficiency (high-TE), low translation efficiency (low-TE) and non-differentially expressed (non-DE) control genes. Outliers were removed for clarity. * demarcates comparisons with significant Dunn's post-hoc tests at a p-value

cutoff of 0.01. **(B)** HEK293 TOPscores for all genes in the high-TE set retrieved from Phillippe et al., 2020. Blue and pink dots represent previously annotated or unannotated 5'TOP RNAs, respectively. A kernel density estimator of the TOPscore distribution is plotted along with median (P_{50}) and 90th percentile (P_{90}) across all human genes. **(C)** Log2 fold-change of translation efficiency comparing NSP1- vs NSP2-expressing HEK293T cells as determined by matched ribosome profiling and RNA-Seq (see Methods). Each gene is colored according to its TOPscore (Philippe et al., 2020) **(D)** Estimated translation initiation rate (Gao et al., 2015) of high-TE genes (blue) and RNA expression matched controls (gray) derived from ribosome profiling data from HEK293 cells generated in the presence of lactimidomycin and puromycin. Open circles represent genes with outlier counts beyond the axis limit. The boxplot summarizes the distribution of read counts for each set. **(E)** The fold change of estimated translation initiation in HEK293 cells subjected to amino acid starvation prior to ribosome profiling in the presence of lactimidomycin and puromycin. The boxplot and open circles represent read count distribution and outliers, respectively, as in panel D.

Supplementary Figure Legends

Supplementary Figure 1 - Validation of sensitivity and specificity of the molecular beacon
(A) *In vitro* molecular beacon (MB) assay targeting the 10N40A synthetic oligo. The gray region indicates the estimated concentration of mRNA in a cell. **(B)** Signal to noise ratio of the MB signal using a random (gray) and specific MB (purple) with HEK293T cells.

Supplementary Figure 2 - MeTAFlow assay of different cell lines
(A) MeTAflow assay of K562 cell line, **(B)** Calu-3 cell line, and **(C)** CaCo-2 cell line after OPP (50 µM) treatment for 10 minutes (K562 cell line) or for 30 minutes (Calu-3 & CaCo-2 cell line).

Supplementary Figure 3 - Gating strategy for MeTAFlow assay
(A) FSC-A & SSC-A gating followed by doublet discrimination using the **(B)** SSC-H & SSC-W plots and **(C)** FSC-A & FSC-H. **(D)** The DY-549 signal in NSP1-, NSP2-, and N-transfected cells during the MeTAFlow assay. The DY-549 positive population in the NSP1 and NSP2 conditions were gated (dotted line) based on the negative control StrepTactin-DY549 background signal (NC, gray). **(E)** AF488 signal in untransfected control (unt), NSP1, and NSP2 conditions during the MeTAFlow assay. The OPP-/AF488+ population in NSP1, NSP2, and control conditions were gated (dotted line) based on the negative control AF488 background signal (NC, green). **(F)** The

Cy5 molecular beacon signal in control, NSP1-, and NSP2-transfected HEK293T cells during the MeTAFlow assay. Compensation was carried out using single stained HEK293T controls. OPP-treated cells, N-transfected cells, and cells incubated with MB were used as AF488, DY549, and Cy5 single stained controls, respectively.

Supplementary Figure 4 - RNA-Seq correlation plots

(A) Pairwise comparison of gene expression data generated with untransfected (unt), **(B)** NSP2-expressing, and **(C)** NSP1-expressing cells. The Spearman correlation coefficients (ρ) are provided and the density of transcripts is represented with the color scale. The axes correspond to the number of reads mapping to the CDS region of a gene for the replicates being compared.

Supplementary Figure 5 - Clustering of RNA-Seq correlations across replicates

Hierarchical clustering of the pairwise Spearman correlation values between RNA-Seq replicates 1, 2, and 3 of untransfected (unt), NSP2-transfected (NSP2), and NSP1-transfected conditions. RNA-Seq read counts mapping to the coding region of each analyzed transcript were used to calculate pairwise Spearman correlation coefficients.

Supplementary figure 6 - **Quality control metrics for ribosome profiling data**

(A) Read length distributions. Ribosome profiling read lengths for untransfected (unt), NSP2, and NSP1 conditions were quantified as fractions of the total reads. All experiments were performed in triplicate. Only reads which mapped to the CDS region of the genes were used. **(B)** Metagene plots of the translation start site across conditions. Position 0 denotes the start site and is flanked by a 50 nt region up and downstream. Mapped reads, with matching positions relative to the translation start site, were aggregated and normalized to counts per 1M reads (see methods). **(C)** Metagene plots of the translation stop site across conditions. These plots were generated as in panel B, but position 0 denotes the stop site. For panels B and C, the 5' end of the ribosome footprints were plotted, causing a shift of ~15 nucleotides to the left.

Supplementary figure 7 - Percentages of region counts as basic QC for ribosome profiling

Representation of the percentage of ribosome profiling reads that mapped to different transcript regions: 5' UTR, CDS, and 3' UTR. Reads with lengths 28-35 nt were plotted for three replicates of untransfected control (unt), NSP2, and NSP1 conditions.

Supplementary Figure 8 - Ribosome profiling correlation plots

(A) Pairwise comparison of ribosome occupancy data generated with untransfected (unt), **(B)** NSP2-expressing, and **(C)** NSP1-expressing cells. The Spearman correlation coefficients (ρ) are provided and the density of transcripts is represented with the color scale. The axes correspond to the number of ribosome footprints mapping to the CDS region of a gene for the replicates being compared.

Supplementary Figure 9 - **Clustering of ribosome profiling correlations across replicates**
Hierarchical clustering of the pairwise Spearman correlation values between ribosome profiling replicates 1, 2, and 3 of untransfected (unt), NSP2-transfected (NSP2), and NSP1-transfected conditions. Ribosome profiling read counts mapping to the coding region of each analyzed transcript were used to calculate pairwise Spearman correlation coefficient.

Supplementary Figure 10 - Sets of highly covered transcripts from ribosome profiling
Transcripts with CDS density 1 (total number of CDS mapping ribosome footprints divided by CDS length) in untransfected (unt), NSP2, and NSP1 conditions. For nucleotide coverage analyses, we used the union of all three sets.

Supplementary Figure 11 - Levels of highly covered transcripts

Average CDS coverage by number of transcripts. For each transcript, the total number of footprints on the CDS was by the length of CDS, giving us the average number of reads per nucleotide position, shown in the x-axis. The y-axis shows the number of genes exceeding the given average reads per nucleotide. In other words, the existence of a point (x, y) means that there are y genes whose average per nucleotide is x.

Supplementary Figure 12 - Ribosome occupancy at nucleotide resolution
(A) Histogram of the Pearson correlations for all pairwise replicate combinations of untransfected (unt) cells. The Pearson correlation coefficient was calculated for each pair of replicates, for each gene, and plotted against the frequency of occurrence. **(B)** Histogram of the Pearson correlations between all pairwise replicate combinations of NSP2-expressing cells. **(C)** Ribosome occupancy of the CDS region of COX7A2L. The mirrored graphs depict ribosome occupancy across the entire CDS region, generated from all replicates of NSP1- and NSP2-expressing cells, and the

number of ribosome footprints per million. A candidate, and quantitatively differential, pause site is highlighted at nucleotide position 1060. The inset shows ribosome occupancy around the highlighted pause site. This pausing event has a fold change of 2.9 between the replicates of NSP1- and NSP2-expressing cells.

Supplementary Figure 13 - Examples of CDS coverages of varying Pearson correlations

(A) Ribosome footprint coverage of RPL41, **(B)** SERP1, and **(C)** ANAPC11 at nucleotide resolution. The coverages above the x-axis come from NSP1 replicate 1 and the coverages below the x-axis come from NSP1 replicate 3. The x-axis shows the nucleotide position and the y-axis shows the number of ribosome footprints. For each gene, Pearson correlation coefficients, between replicate 1 and 3, were calculated using the reads mapping to the CDS region only. RPL41, SERP1, and ANAPC11 have Pearson correlations of 0.93, 0.60, and 0.43, respectively.

Supplementary Figure 14 - Pause sites

(A) Pause sites from the three different conditions. We modeled the distribution of the reads mapping to CDS using the negative binomial distribution and combined their p-values for each condition. Using a cutoff of 10-7on the combined p-values we obtained a list of candidate pause sites for each condition. We used the union of these 3 sets for further analysis. **(B)** Percentage of overlapping Pause Sites versus Pause Site Means. The y-axis shows the percentage of pause sites that exist both in NSP1 and NSP2 expressing replicates. The x-axis shows the average number of reads coming from the pause sites in log2 scale.

Supplementary Figure 15 - OAZ1 pause site analysis

(A) CDS Coverage of the OAZ1 gene. The x-axis shows the nucleotide position and the y-axis shows the number of ribosome footprints. The pause site, spanning the positions from 259 to 270 nt, is highlighted in red. Coverages from the first replicate of each condition are plotted individually. **(B)** Frame shift analysis. The coding sequence was partitioned into bins of 3 nucleotides (0 nt, 1 nt and 2 nt). For EEF1A1 and DHX15, the ribosome footprints were aggregated at each frame position, in the entire CDS. Then, the percentages of the frame positions were calculated. For OAZ1, ribosome footprints to the right of the pause site shown in A were aggregated. (**C)** Coverage per read length at the pause site. For each read length from 28 to 35 nt, the number of reads at the pause site were divided by the number of all reads having the same read length, resulting in the ratio shown on the y-axis.

Supplementary Figure 16 - ERCC spike-in RNA controls

(A) Jackknife estimate of the observed ratio for each spike-in control was calculated for RNA sequencing libraries prepared from NSP1-expressing (NSP1) HEK293T cells and untransfected (unt) controls in the presence of ERCC Spike-in Control Mix 1 or **(B)** NSP2-expressing HEK293T cells prepared in the presence of ERCC Spike-in Control Mix 2. An equal mRNA fraction between samples would lead the ratio estimates to be centered at 1. A higher ratio indicates lower mRNA content in cells expressing NSP1. Different spike-in controls are designed to be at four distinct ratios indicated by the dashed lines. **(C)** Differential expression p-value for each spike-in plotted against its average count. ERCCdashboard was used to estimate a limit of detection of ratios for each of the spike-in RNAs. Vertical lines indicate the threshold read count above which differential expression of the given magnitude can be confidently identified.

Supplementary Figure 17 - Sequence features of gene sets

(A) 5' and 3' untranslated region (5'UTR, 3'UTR) and coding sequence (CDS) lengths for high and low translation efficiency (high-TE, low-TE) and non-differentially expressed (non-DE) control genes. **(B)** GC content. **(C)** Predicted minimum free energy (MFE) in -kcal/mol of the first 60 nts of the 5' UTR for comparative genesets. **(D)** Predicted MFE for a window between -30 to +30 around the start codon. **(A-E)** Asterisks indicate significant Dunn's post-hoc tests at a significance level of 0.01. Outliers greater than (1.5 * interquartile range) were omitted for clarity.

Supplementary Figure 18 - Investigation of RNA editing in high-TE genes

Base substitution frequencies in RNA-Seq alignments. Substitutions were enumerated relative to the annotated transcript strands, requiring a base quality >= 35. The counts were normalized to the total mapped read count of the library. Data was then unit-normalized by dividing by the sum of all depth-normalized matches and substitutions. The log proportions are shown for A:G, A:T (ADAR editing) and C:T (APOBEC editing).

Supplementary Figure 19 - Raw data for eCLIP

Processed eCLIP data. Lines represent the median values which correspond to the heat fill in Figure 4E. Narrowpeak BED files from eCLIP experiments (Van Nostrand et al., 2020) for pertinent RNA binding proteins were intersected with exonic regions of the high-TE and six matched non-DE gene sets. Only BED features that were labeled with "IDR", indicating reproducible peaks, were used in analysis. Peaks were summed across the geneset and a pseudocount was added to generate each point.

Supplementary Files and Tables

Supplementary Table 1 - Basic Mapping Statistics

This table contains some critical mapping statistics such as the depth of the sequencing libraries, number of mapped reads and number of reads after quality filtering and deduplication.

Supplementary Table 2 - List of highly covered transcripts

This table contains the list of transcripts having CDS density 1. CDS density is defined as the total number of footprints mapping to CDS divided by the length of CDS.

Supplementary Table 3 - Potential differential pause sites

This table contains 402 potential differential pause sites. It was obtained by comparing the read counts at the pause sites and calling the differential pause sites using EdgeR with a p-value cut-off 0.5. Note that our finalized pause sites belong to a subset of this list provided in other tables. The coordinate position of the pause sites are appended to the name of the genes after an underscore ("_"). The positions are given with respect to the start sites. In other words, the first nucleotide of each start site has the coordinate 0.

Supplementary Table 4 - Differential Pause Sites with no CDS occupancy Influence

This table contains the 39 pausing events whose differential behavior doesn't come from the overall differential ribosome occupancy.

Supplementary Table 5 - Differential Pause Sites with low CDS occupancy Influence

This table contains the list of 16 pausing sites having differential ribosome occupancy. However, the absolute difference between the log fold changes of the pause site and the ribosome occupancy of the gene are above 0.8. Thus, the magnitude of the pausing differential is significantly higher than that of ribosome occupancy.

Supplementary Table 6 - Differential RNA expression between HEK293T cells expressing NSP1 compared to NSP2

This table contains the list of transcripts with significant expression changes between NSP1 and NSP2 expressing cells.

Supplementary Table 7 - List of enriched gene ontology terms for genes with differential RNA expression

This table contains the gene ontology analysis results for genes with differential expression. FuncAssociate (Berriz et al., 2009) was used to determine significantly enriched gene ontology (GO) terms. A permutation based multiple hypothesis testing correction was adopted to calculate adjusted p-values (Berriz et al., 2009). HGNC gene identifiers were used for all analyses.

Supplementary Table 8 - Differential translation efficiency between HEK293T cells expressing NSP1 compared to NSP2

This table contains the list of transcripts with significant translation efficiency changes between NSP1 and NSP2 expressing changes.

Supplementary Table 9 - List of enriched gene ontology terms for high TE genes

This table contains the gene ontology analysis results for genes with differential translation efficiency. FuncAssociate (Berriz et al., 2009) was used to determine significantly enriched gene ontology (GO) terms. A permutation based multiple hypothesis testing correction was adopted to calculate adjusted p-values (Berriz et al., 2009). HGNC gene identifiers were used for all analyses.

Supplementary Table 10 - Unmatched sequence feature statistical tests

This table contains the summary of Kruskal-Wallis and Dunn's post-hoc test results for comparative gene sets. Dunn's p-values were adjusted by the Benjamini-Hochberg (BH) procedure.

Supplementary Table 11 - Covariate-matched sequence feature statistical tests

This table contains the summary of significant matched comparison by t-test and Wilcoxon rank sum tests. Matched_on field indicates which covariates (feature - regions) were matched when selecting control non-DE genes for comparison to the high-TE genes.

Supplementary Table 12 - oRNAment RBP site counts- Filter C

This table contains the RBP site counts with a matrix similarity score >= 0.9 for each RBP and region. The annotation field represents either the high-TE genes or matched, mutually exclusive non-DE gene sets.

Supplementary Table 13 - oRNAment RBP site log2 FCs- Filter C

This table contains log2 fold-changes in RBP sites between the high-TE set and each non-DE set. The log2 mean sites between comparative groups are included. See Methods for Filter C details. Reference Supplementary Table 15.

Supplementary Table 14 - oRNAment RBP site log2 FCs- Filter D

This table contains log2 fold-changes in RBP sites between the high-TE set and each non-DE set. The log2 mean sites between comparative groups are included. See Methods for Filter D details.

Supplementary Table 15 - Excluded high-TE genes in RBP analysis

This table contains the high-TE genes that were dropped due to missing regions, non-real values, or discrepancies between the APPRIS annotations and the Bioconductor Ensembl db v79 release.

Supplementary File 1 - .ribo file

The ribo file contains ribosome profiling data, at nucleotide resolution, and RNA-Seq data. For details, see (Ozadam et al., 2020).

Materials and Methods

Plasmids and Cloning

pLVX-EF1alpha-SARS-CoV-2-nsp1-2XStrep-IRES-Puro plasmid (Addgene, 141367), pLVX-EF1alpha-SARS-CoV-2-nsp2-2XStrep-IRES-Puro plasmid (Addgene, 141368) and pLVX-EF1alpha-SARS-CoV-2-N-2XStrep-IRES-Puro (Addgene, 141391) were obtained from Addgene. The IRES-Puro elements in the above plasmids were replaced with IRES-TagBFP from TRE-KRAB-dCAS9-IRES-BFP (Addgene, 85449) to make them compatible with the MeTAFlow assay. Briefly, IRES-Tag BFP was PCR amplified (Fp 5'-TCCGCAGTTTGAAAAGTAAGCCCCTCTCCCTCCCCCCC-3' and Rp 5'-CAGAGGTTGATTGTTCCAGATTAATTAAGCTTGTGCCCCAGTTTGCTAGGG-3'). The pLVX-EF1alpha-SARS-CoV-2-nsp1/nsp2-2XStrep-IRES-Puro plasmids were digested with BamHI (NEB) and MluI (NEB) to remove the IRES-Puromycin fragment followed by gel purification. The purified vector was ligated with the IRES-TagBFP PCR amplified product using Gibson assembly cloning (NEB). For pLVX-EF1alpha-SARS-CoV-2-N-2XStrep-IRES-Puro plasmid, due to

presence of internal MluI site in N ORF, a three fragment Gibson cloning was performed. The plasmid was digested with MluI and two PCR amplified products corresponding to a fragment of N ORF and IRES-Tag BFP were PCR amplified using the following oligos. (Fragment 1 Fp 5' 5'CTCTTGAATAAACACATAGACGCGTATAAGACGTTTCC 3′ and Rp GGGAGAGGGGCGGGATCCTTACTTTTCAAAC 3′ and Fragment 2 5' AAGGATCCCGCCCCTCTCCCTCCCCCCC 3′ and Rp 5' CAGAGGTTGATTGTTCCAGATTAATTAAGCTTGTGCCCCAGTTTGCTAGGG3′ The sequences of the resulting plasmid clones were confirmed by Sanger sequencing.

Cell culture and Transfection

HEK293T cell line were obtained from ATCC and maintained in Dulbecco's modified Eagle's media (DMEM, GIBCO) supplemented with 10% Fetal Bovine Serum (FBS, GIBCO, Life Technologies) and 1% Penicillin and Streptomycin (GIBCO, Life Technologies) at 37°C in 5% CO_2 atmosphere. Cell lines were tested for mycoplasma contamination every six months and were consistently found negative.

For the MeTAFlow Assay, HEK293T cells were plated at a density of $3X10_5$ cells in a 6-well plate. The following day, 2.5 µg of pLVX-NSP1/NSP2/N plasmids were transfected using 3.5 µL of Lipofectamine 3000 (GIBCO, Life Technologies). The media was changed after 8 h followed by MeTAFlow assay 24 h post transfection.

MeTAFlow Assay and Flow Cytometry

24 h post-transfection, cells were treated with 50 µM O-Propargyl Puromycin (OPP, Click Chemistry Tools) for 10 min at 37 °C. Cells were washed with phosphate-buffered saline without calcium or magnesium (PBS, GIBCO, Life Technologies) to remove free OPP followed by centrifugation at 300 x g for 5 min at 4 °C. Chilled 70% ethanol was added drop by drop with intermittent vortexing followed by an overnight incubation at -20°C for fixation. Before FACS analysis, cells were again washed with PBS, followed by Click Chemistry reaction to label the OPP incorporated into the nascent polypeptide chains. Briefly ~2 million cells were incubated in Dulbecco's phosphate-buffered saline (DPBS) buffer containing 1 mM $CuSO_4$ (Sigma), 5 mM BTTA (Click Chemistry Tools) and 50 nM of Picolyl Azide AF488 (Click Chemistry Tools). After a 2 min incubation, 2.5 mM sodium ascorbate (Sigma) was added followed by addition of 1 µM molecular beacon (IDT). Sequence of the molecular beacon used are 5Cy5CTCGCTTTTTTTTTTTTTTTTTTTGCGAG/3IAbRQSp and that of the random control

MB is 5Cy5CTCGCCGACAAGCGCACCGATAGCGAG/3IAbRQSp (Wile et al., 2014). The cells were incubated at 37 °C for 1 h, washed with PBS followed by labelling of the Strep-tag using 0.5 µL of Strep-Tactin®XT, conjugated to DY-549 (IBA Lifesciences) in DBPS at 4 °C for 20 min for cells transfected with NSP1/NSP2 and Nucleocapsid (N) protein. Cells were then washed with PBS containing 3% Bovine Serum Albumin (BSA, Sigma). Following MeTAflow assay the cells were passed through a strainer cap to achieve a single cell suspension and immediately analyzed using BD FACS Aria Fusion SORP Cell Sorter. Compensation was performed using singly stained controls with OPP-AF488, MB-Cy5 and N protein for StrepTactin 549. N protein expressed cells were used for compensation due to high fluorescence intensity compared NSP1 and NSP2. Data was later analyzed by Flowjo 10.6.1. Briefly, singlets were gated using a SSC-H vs SSC-W plot followed by sequential gating using a FSC-A vs FSC-H plot. SInglets were gated for OPP-AF488 positive population (OPP Fl.) using background signals obtained from AF488 stained cells. Further, Strep-tag positive cells were obtained by gating using a background signal obtained from with Strep-Tactin®XT, conjugated to DY-549 stained untransfected cells.

In Vitro Molecular Beacon assay

The assay was performed by incubating the MB (0.5 µM) with varying concentrations of a synthetic oligo (IDT), N10A40 (10 nucleotides of randomized bases followed by 40 adenines) for 1 h at 37 °C. The fluorescence was measured using a Tecan M1000 Plate reader at 647 nm excitation wavelength.

Ribosome Profiling

Three million HEK293T cells were plated in a 10 cm_2 flask followed by transfection to express SARS-CoV-2 NSP1 and NSP2 (see above). Untransfected and transfected cells (~8 million) were washed twice with 10 mL of ice-cold PBS. The plates were immediately placed on ice and 400 µL of lysis buffer (Tris-Cl, pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 100 µg/mL Cycloheximide, 1% Triton-X) was added to each plate, cells were scraped and transferred to 1.5 ml tubes. Cells were lysed on ice by pipetting up and down ~5 times every five minutes for a total of 10 min. All experiments were done in triplicate. The lysate was clarified by centrifugation at 160 x g for 10 min at 4°C. Ten percent of the clarified lysate by volume was separated, mixed with 700 µl QIAzol and stored at -20°C for RNA-Seq experiments (see below). The rest of the supernatant was immediately processed for ribosome profiling. Briefly, 7 µL of RNaseI (Invitrogen AM2249) was added to the clarified lysates and digestion was done for 1 h at 4 °C. The digestions were stopped with ribonucleoside vanadyl complex (NEB S1402S) at a final concentration of 20 mM. Digested

lysates were layered on a 1 M sucrose cushion (Tris, pH 7.4, NaCl 150 mM, $MgCl_2$ 5 mM, DTT 1 mM, 1 M sucrose) and the ribosomes were pelleted by centrifugation in a SW 41 Ti rotor (Beckman Coulter) at 38K rpm and 4°C for 2.5 h. RNA was extracted from the ribosome pellets with the addition of 700 µl QIAzol followed by chloroform and ethanol precipitation. RNA isolated from the pellets were size-selected by running 5 µg of each sample in a 15% polyacrylamide TBE-UREA gel. The 21-34 nt RNA fragments were excised and extracted by crushing the gel fragment in 400 µL of RNA extraction buffer (NaOAc [pH 5.5] 300 mM, EDTA 1 mM, SDS 0.25% v/v) followed by a 30 min incubation on dry ice and an overnight incubation at room temperature. The sample was passed through a Spin X filter (Corning 8160) and the flow through was ethanol precipitated in the presence of $MgCl_2$ 5 mM and 1 µL of Glycoblue (Invitrogen AM9516). The RNA pellet was resuspended in 10 µL of RNase-free water and immediately processed for library preparation.

Ribosome profiling library preparation

Ribosome profiling libraries were prepared with the D-Plex Small RNA-Seq Kit (Diagenode). This method incorporates a 3′ end dephosphorylation step, 5′ unique molecular identifiers (UMI) and a template switching reverse transcriptase that improves the quality of our libraries. Briefly, 25 ng of size-selected ribosome footprints were prepared following the manufacturer's instructions with some modifications. The cDNA was amplified for 9 cycles and the resulting DNA libraries were pooled in equimolar amounts (~425 nM each). The library pool was cleaned with the AMPure XP beads (Beckman Coulter A63880) and eluted with 30 µL of RNase-free water. To enrich for ~30 bp fragments in the libraries, 3 µg of the cleaned libraries were size-selected in a 3% agarose precast-gel (Sage Science) with the BluePippin system (Sage Science) using 171-203 nt range with tight settings. The resulting size-selected libraries were analyzed with the Agilent High Sensitivity DNA Kit (Agilent) and sequenced with the NovaSeq 6000 S1 SE 75 (Illumina).

RNA sequencing

Total RNA was extracted with QIAzol and ethanol precipitation from 10% of the lysate volume (see above). Sequencing libraries were generated using the SMARTer Stranded RNA-Seq Kit which uses a template switching reverse transcriptase (Takara Bio 634837). Briefly, 100 ng of total RNA were mixed with 1 µL of a 1:10 dilution of ERCC RNA Spike-In Mix controls (Thermo Fisher 4456740). ERCC mix 1 was added to HEK293T and HEK293T-NSP1 and mix 2 was added to HEK293T-NSP2 samples. RNA hydrolysis was done for 4 min and half of the cDNA was amplified for 10 cycles. Samples were sequenced with NovaSeq 6000 S1 PE 100 (Illumina).

Preprocessing and Quality Control of Sequencing Data

Ribosome profiling and RNA-Seq data were preprocessed using RiboFlow (Ozadam et al., 2020). Quantified data were stored in ribo files and analyzed using RiboR and RiboPy (Ozadam et al., 2020). Source code is available at https://github.com/ribosomeprofiling. A brief description of the specifics are provided here for convenience.

For ribosome profiling data, we extracted the first 12 nucleotides and discarded the following 4 nucleotides, of the form NGGG (nucleotides 13 to 16 from the 5' end of the read). Next, we trimmed the 3' adapter sequence AAAAAAAAAACAAAAAAAAAA. For RNA-Seq data, we trimmed 30 nucleotides from either end of the reads, yielding 40 nucleotides to be used in downstream processing. After trimming, reads aligning to rRNA and tRNA sequences were discarded. The remaining reads were mapped to principal isoforms obtained from the APPRIS database (Rodriguez et al., 2018) for the human transcriptome. Next, UMI-tools (Smith et al., 2017) was used to eliminate PCR duplicates from the ribosome profiling data. Two ribosome footprints mapping to the same position are deemed PCR duplicates if they have UMIs with Hamming Distance of at most one. We note that deduplication via UMI-tools is an experimental feature of RiboFlow as of this study and was not a feature of the stable release at the time of its publication. Finally, the alignment data, coming from ribosome profiling and RNA-Seq, were compiled into a single ribo file (Ozadam et al., 2020) for downstream analyses. Unless otherwise specified, ribosome footprints of length between 28 and 35 (both inclusive) nucleotides were used for all analyses. Additionally, to quantify aligned reads, we counted the nucleotide position on the 5′ end of each ribosome footprint. Basic mapping statistics are provided in Supplementary Table 1.

Differential Expression and Translation Efficiency Analysis

Read counts that map to coding regions were extracted from the 'ribo' files for all experiments. For RNA-Seq analyses, these counts were merged with a table of read counts for each of the ERCC spike-in RNAs. ERCC data analyses were done using the erccdashboard R package (Munro et al., 2014). The jackknife estimator of the ratios for each ERCC spike-in RNA was calculated by assuming arbitrary pairing of the libraries and previously described methods by Quenouille and Durbin (Durbin, 1959).

Ribosome occupancy and RNA-Seq data was jointly analyzed and normalized using TMM normalization (Robinson and Oshlack, 2010). Transcript specific dispersion estimates were calculated and differentially expressed genes (Supplementary Table 6) were identified using edgeR (Robinson et al., 2010). To identify genes with differential ribosome occupancy

(Supplementary Table 8) while controlling for RNA differences, we used a generalized linear model that treats RNA expression and ribosome occupancy as two experimental manipulations of the RNA pool of the cells analogous to previously described (Cenik et al., 2015). We used an adjusted p-value threshold of 0.05 to define significant differences. Gene set enrichment analyses for gene ontology terms was carried out using FuncAssociate (http://llama.mshri.on.ca/funcassociate/) with default settings (Berriz et al., 2009). R packages cowplot, pheatmap, EnhancedVolcano, ggpubr, ggplot2, and reshape2 were used for analyses and plotting (Wilke 2016; Kolde 2012; Blighe et al. 2019; Kassambara 2018; Wickham 2011; Wickham 2012).

Nucleotide Resolution Analyses of Ribosome Occupancy and Pause Site Detection

For all nucleotide resolution and pause site detection analyses, we used the ribosome footprints mapping to the CDS (28 to 35 nucleotides). We restricted our analyses to a subset of genes that have a minimum density of coding sequence ribosome footprint density (CDS density). We define CDS density as the ratio of total CDS mapping footprints to the CDS length. For each of the three experimental conditions, we selected genes having a CDS density 1 in all of its replicates. Then, the union of the three sets coming from the conditions NSP1, NSP2 and WT was analyzed further (1493 transcripts; **Supplementary Table 2**). Consequently, for any given analyzed transcript, there is at least one condition such that this transcript has a CDS density 1 in all replicates of this condition.

Per nucleotide coverage across the CDS was used to compute the Pearson correlation between all pairwise combinations of replicates of each condition. The distribution of correlation coefficients (**Supplementary Figure 12**) between replicates were generated to assess reproducibility. The medians of Pearson correlation coefficients were determined after combining the Pearson correlations coming from the 3 pairs of experiments being compared.

To identify candidate pause sites, we first fitted a separate Negative Binomial (NB) to each transcript using its distribution of ribosome footprint counts at each nucleotide. We assumed that a statistically outlier number of footprints at position would be a candidate ribosome pause site given the expected read count from that transcripts fitted NB distribution. More specifically, for each transcript, we first estimate parameters of the NB distribution. Let x be the number of footprints for a given nucleotide position in a given transcript. Then, the p-value, assigned to this position, is the probability of observing x or more footprints given this transcript's fitted NB distribution. We compiled a list of candidate pause sites by aggregating p-values across the 3

replicates of the given condition using Fisher's method. Nucleotide positions with a combined p-values less than 10-7 were considered as candidate pause sites yielding a list of 6339 elements.

Using EdgeR (Robinson et al., 2010), we determined candidate pause sites which are significantly different between NSP1 and NSP2. We collected ribosome occupancy counts from the candidate pause sites into a single table. After normalizing the counts using the TMM method, we extracted a list of potential differential pause sites using an adjusted p-value threshold of 5x10-2 yielding 402 elements (**Supplementary Table 3**). In this list, differential pausing events might be a consequence of transcript-wide differential ribosomal occupancy. Therefore, we determined differential CDS occupancy of ribosome footprints across transcripts using EdgeR with an adjusted p-value cutoff 5x10-2. In our list of potential differential pause sites, transcripts corresponding to 39 sites (**Supplementary Table 4**) were not in our differential CDS occupancy list. For remaining sites, we determined the absolute value of the difference between the log fold change of the candidate pause site and its corresponding transcript. We found 16 sites with absolute log fold change difference exceeding 0.8 (**Supplementary Table 5**). Combined together, we identified a total of 55 differential pausing events.

Translation initiation rate estimation

Ribosome profiling data with lactimidomycin and puromycin treatment (named QTI-Seq) was retrieved from (Gao et al., 2015). Among the 166 genes with relatively increased translation efficiency upon NSP1 expression, 144 were quantified in QTI-Seq experiments in HEK293 cells. There were 381 potential translation start sites in this set of genes. We restricted our analyses to the subset of annotated start sites. To generate a control set with a similar distribution of RNA expression, we used the MatchIt R package (Ho et al., 2007). Specifically, we used the nearest neighbor matching algorithm and used the RNA expression measurements obtained by RNA sequencing from the same cells that were used for QTI-Seq experiments.

Sequence feature analysis and statistical testing

Gene names from the high-TE, low-TE, and non-DE genesets were converted to Ensembl gene IDs with the Bioconductor Ensembl v79 release version 2.99.0 (Johannes Rainer, 2017). Then these IDs were used to extract sequences from the APPRIS principal isoforms used as the reference in ribosome profiling analysis. Due to discrepancies in APPRIS and Bioconductor gene IDs, 8 genes were dropped from the high-TE geneset in sequence feature analysis (**Supplementary Table 15**). Transcript region lengths were extracted from the APPRIS annotations via custom shell scripts. Region sequences were extracted using bedtools v2.29.1

(Quinlan and Hall, 2010) and nucleotide content was computed via the Bioconductor Biostrings package v2.54.0 (Pagès et al., 2017). Minimum free energy (MFE) was computed via the PyPi package seqfold (https://pypi.org/project/seqfold/), which uses a nearest-neighbor dynamic programming algorithm for the minimum energy structure (Zuker and Stiegler, 1981); (Turner and Mathews, 2010). MFE was computed for a 60 nt window at the start of the 5' UTR and for a 60 nt window spanning the start codon (between -30 and +30). Sequence feature differences were tested using the Kruskal-Wallis test and Dunn's post-hoc tests. p-values from Dunn's tests were adjusted with the Benjamini-Hochberg procedure and a significance level of 0.01 was for all comparisons. For matched comparisons, the MatchIt package (Ho et al., 2007) was used with the nearest neighbor method and the logit for computing distance. All analyses utilized data.table version 1.12.8 (Dowle et al., 2013).

Analysis of RNA-binding protein sites and eCLIP data

Five additional high-TE genes were discarded that had no 5' UTR or had non-real MFE values, resulting in a total of 153 high-TE genes for further analysis (**Supplementary Table 15).** Then six matched non-DE control genesets were generated from the set of genes with mRNA expression but no differential translation efficiency, with a sample size equivalent to the high-TE geneset. Non-DE genesets were matched on length and GC content for each region in the transcript by MatchIt. The number of RBP sites was computed for each geneset after filtering for a matrix similarity score >= 0.9 using records extracted from the full oRNAment database (Benoit Bouvrette et al., 2020) (Filter A). The log2 fold change in the number of RBP sites was computed between the high-TE genes and each non-DE group. Comparisons (high-TE and matched non-DE set) were dropped if no sites were found for a RBP in the high-TE or non-DE genesets (Filter B). Candidate differential RBP sites were defined as having at least one high-TE:non-DE comparison with a fold change of sites >= 2 and with sum of sites for high-TE and matched non-DE set >= 20 (Filter C). To generate Figure 4D, RBPs with ambiguous nomenclature, isoforms, or homologs with identical fold changes and RBPs that had at least one non-DE set with a FC in the opposite direction were manually filtered out (Filter D). All RBP data under Filters C and D are available in **Supplementary Tables 13-14**.

To analyze experimental eCLIP data, RBP narrowpeak BED files (Van Nostrand et al., 2020) were downloaded from ENCODE project website (https://www.encodeproject.org/). eCLIP peaks were intersected with exonic regions of the high-TE and non-DE control genesets using bedtools intersect with the -wb option (Quinlan and Hall, 2010). Only reproducible eCLIP peaks that passed the irreproducible discovery rate filter (Van Nostrand et al., 2020) were counted.

Substitution frequency analysis

All single nucleotide substitutions and matched bases with base quality >= 35 were enumerated in the processed RNA-seq alignments from NSP1, NSP2, and untreated conditions. Substitutions were reported relative to the annotated transcript strands and were normalized to the total number of reads in each library. Then relative log proportions for each substitution, including matches, was computed and compared for A>G, A>T, and C>T, the predominant substitutions expected from RNA editing. We note that while base quality information was considered, this method is likely to contain noise from polymerase errors during sequencing and/or library preparation.

Sequencing Data and Supplementary Tables and Files

Deep sequencing files of ribosome profiling and RNA-Seq experiments, together with the supplementary tables and files, will be available from GEO. As soon as GEO assigns an accession number, the supplementary data will be publicly available.

References

Ansa-Addo, E.A., Huang, H.-C., Riesenberg, B., and Iamsawat, S. (2020). RNA binding protein PCBP1 is an intracellular immune checkpoint for shaping T cell responses in cancer immunity. Science Advances *6*.

Antonicka, H., and Shoubridge, E.A. (2015). Mitochondrial RNA Granules Are Centers for Posttranscriptional RNA Processing and Ribosome Biogenesis. Cell Rep. *10*, 920–932.

Atkins, J.F., Loughran, G., Bhatt, P.R., Firth, A.E., and Baranov, P.V. (2016). Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. Nucleic Acids Res. *44*, 7007–7078.

Aviner, R., and Frydman, J. (2020). Proteostasis in Viral Infection: Unfolding the Complex Virus-Chaperone Interplay. Cold Spring Harb. Perspect. Biol. *12*.

Benoit Bouvrette, L.P., Bovaird, S., Blanchette, M., and Lécuyer, E. (2020). oRNAment: a database of putative RNA binding protein target sites in the transcriptomes of model species. Nucleic Acids Res. *48*, D166–D173.

Berriz, G.F., Beaver, J.E., Cenik, C., Tasan, M., and Roth, F.P. (2009). Next generation software for functional trend analysis. Bioinformatics *25*, 3043–3044.

Bianco, C., and Mohr, I. (2019). Ribosome biogenesis restricts innate immune responses to virus infection and DNA. Elife *8*.

Binici, J., and Koch, J. (2014). BAG-6, a jack of all trades in health and disease. Cell. Mol. Life Sci. *71*, 1829–1837.

Blanc, V., Xie, Y., Kennedy, S., Riordan, J.D., Rubin, D.C., Madison, B.B., Mills, J.C., Nadeau, J.H., and Davidson, N.O. (2019). Apobec1 complementation factor (A1CF) and RBM47 interact in tissue-specific regulation of C to U RNA editing in mouse intestine and liver. RNA *25*, 70–81.

Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., Ciesek, S., Cinatl, J., and Münch, C. (2020). Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. Nature *583*, 469–472.

Brocard, M., Iadevaia, V., Klein, P., Hall, B., Lewis, G., Lu, J., Burke, J., Willcocks, M.M., Parker, R., Goodfellow, I.G., et al. (2020). Norovirus infection results in eIF2α independent host translation shut-off and remodels the G3BP1 interactome evading stress granule formation. PLoS Pathog. *16*, e1008250.

Buskirk, A.R., and Green, R. (2017). Ribosome pausing, arrest and rescue in bacteria and eukaryotes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*.

Caselli, E., Benedetti, S., Grigolato, J., Caruso, A., and Di Luca, D. (2012). Activating transcription factor 4 (ATF4) is upregulated by human herpesvirus 8 infection, increases virus replication and promotes proangiogenic properties. Archives of Virology *157*, 63–74.

Cenik, C., Cenik, E.S., Byeon, G.W., Grubert, F., Candille, S.I., Spacek, D., Alsallakh, B., Tilgner, H., Araya, C.L., Tang, H., et al. (2015). Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. Genome Res. *25*, 1610–1621.

Chen, G., Ma, L.-C., Wang, S., Woltz, R.L., Grasso, E.M., Montelione, G.T., and Krug, R.M. (2020). A double-stranded RNA platform is required for the interaction between a host restriction factor and the NS1 protein of influenza A virus. Nucleic Acids Res. *48*, 304–315.

Collart, M.A., and Weiss, B. (2020). Ribosome pausing, a dangerous necessity for co-translational events. Nucleic Acids Res. *48*, 1043–1055.

Diament, A., and Tuller, T. (2016). Estimation of ribosome profiling performance and reproducibility at various levels of resolution. Biol. Direct *11*, 24.

Di Giorgio, S., Martignano, F., and Torcia, M.G. (2020). Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. Science Advances *6*.

Dinman, J.D. (2012). Mechanisms and implications of programmed translational frameshifting. Wiley Interdiscip. Rev. RNA *3*, 661–673.

Dixit, U., Pandey, A.K., Mishra, P., Sengupta, A., and Pandey, V.N. (2016). Staufen1 promotes HCV replication by inhibiting protein kinase R and transporting viral RNA to the site of translation and replication in the cells. Nucleic Acids Res. *44*, 5271–5287.

Dowle, M., Short, T., Lianoglou, S., Srinivasan, A., and Saporta, R. (2013). data. table: Extension of data. frame for fast indexing, fast ordered joins, fast assignment, fast grouping and list columns. R Package Version *1*.

Durbin, J. (1959). [No title]. Biometrika *46*, 477–480.

Efron, B., and Stein, C. (1981). The Jackknife Estimate of Variance. Ann. Stat. *9*, 586–596.

Fassati, A., Görlich, D., Harrison, I., Zaytseva, L., and Mingot, J.-M. (2003). Nuclear import of HIV-1 intracellular reverse transcription complexes is mediated by importin 7. EMBO J. *22*, 3675–3685.

Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Melamed, S., Weiss, S., Israely, T., et al. (2020). The coding capacity of SARS-CoV-2.

Fraser, J.E., Wang, C., Chan, K.W.K., Vasudevan, S.G., and Jans, D.A. (2016). Novel dengue virus inhibitor 4-HPR activates ATF4 independent of protein kinase R–like Endoplasmic Reticulum Kinase and elevates levels of eIF2α phosphorylation in virus infected cells. Antiviral Res. *130*, 1–6.

Gaete-Argel, A., Márquez, C.L., Barriga, G.P., Soto-Rifo, R., and Valiente-Echeverría, F. (2019). Strategies for Success. Viral Infections and Membraneless Organelles. Front. Cell. Infect. Microbiol. *9*, 336.

Gamarnik, A.V., and Andino, R. (2000). Interactions of Viral Protein 3CD and Poly(rC) Binding Protein with the 5′ Untranslated Region of the Poliovirus Genome. Journal of Virology *74*, 2219–2226.

Gao, X., Wan, J., Liu, B., Ma, M., Shen, B., and Qian, S.-B. (2015). Quantitative profiling of initiating ribosomes in vivo. Nat. Methods *12*, 147–153.

Gentilella, A., Morón-Duran, F.D., Fuentes, P., Zweig-Rocha, G., Riaño-Canalias, F., Pelletier, J., Ruiz, M., Turón, G., Castaño, J., Tauler, A., et al. (2017). Autogenous Control of 5′TOP mRNA Stability by 40S Ribosomes. Mol. Cell *67*, 55–70.e4.

Gismondi, A., Caldarola, S., Lisi, G., Juli, G., Chellini, L., Iadevaia, V., Proud, C.G., and Loreni, F. (2014). Ribosomal stress activates eEF2K-eEF2 pathway causing translation elongation inhibition and recruitment of terminal oligopyrimidine (TOP) mRNAs on polysomes. Nucleic Acids Res. *42*, 12668–12680.

Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature *583*, 459–468.

Guo, A.-X., Cui, J.-J., Wang, L.-Y., and Yin, J.-Y. (2020). The role of CSDE1 in translational reprogramming and human diseases. Cell Commun. Signal. *18*, 14.

Harcourt, J., Tamin, A., Lu, X., Kamili, S., Sakthivel, S.K., Murray, J., Queen, K., Tao, Y., Paden, C.R., Zhang, J., et al. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States. Emerg. Infect. Dis. *26*, 1266–1273.

Henras, A.K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A., and Gleizes, P.-E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdiscip. Rev. RNA *6*, 225–242.

Ho, D.E., Imai, K., King, G., and Stuart, E.A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Polit. Anal. *15*, 199–236.

Hopkins, K.C., Tartell, M.A., Herrmann, C., Hackett, B.A., Taschuk, F., Panda, D., Menghani, S.V., Sabin, L.R., and Cherry, S. (2015). Virus-induced translational arrest through 4EBP1/2-dependent decay of 5′-TOP mRNAs restricts viral infection. Proc. Natl. Acad. Sci. U. S. A. *112*, E2920–E2929.

Huang, C., Lokugamage, K.G., Rozovics, J.M., Narayanan, K., Semler, B.L., and Makino, S. (2011). SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. PLoS Pathog. *7*, e1002433.

Hutchinson, E.C., Orr, O.E., Man Liu, S., Engelhardt, O.G., and Fodor, E. (2011). Characterization of the interaction between the influenza A virus polymerase subunit PB1 and the host nuclear import factor Ran-binding protein 5. J. Gen. Virol. *92*, 1859–1869.

Ikebe, E., Matsuoka, S., Tezuka, K., Kuramitsu, M., Okuma, K., Nakashima, M., Kobayashi, S., Makiyama, J., Yamagishi, M., Oyadomari, S., et al. (2020). Activation of PERK-ATF4-CHOP pathway as a novel therapeutic approach for efficient elimination of HTLV-1--infected cells. Blood Advances *4*, 1845–1858.

Irigoyen, N., Firth, A.E., Jones, J.D., Chung, B.Y.-W., Siddell, S.G., and Brierley, I. (2016). High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling. PLoS Pathog. *12*, e1005473.

Ivanov, I.P., Shin, B.-S., Loughran, G., Tzani, I., Young-Baird, S.K., Cao, C., Atkins, J.F., and Dever, T.E. (2018). Polyamine Control of Translation Elongation Regulates Start Site Selection on Antizyme Inhibitor mRNA via Ribosome Queuing. Mol. Cell *70*, 254–264.e6.

Jakel, S. (1998). Importin beta , transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. The EMBO Journal *17*, 4491–4502.

Jin, Q., Ren, Y., Wang, M., Suraneni, P.K., Li, D., Crispino, J.D., Fan, J., and Huang, Z. (2016). Novel function of FAXDC2 in megakaryopoiesis. Blood Cancer J. *6*, e478.

Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., Kubo, H., and Makino, S. (2006). Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. Proc. Natl. Acad. Sci. U. S. A. *103*, 12885–12890.

Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K.G., and Makino, S. (2009). A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. Nat. Struct. Mol. Biol. *16*, 1134–1140.

Kendirgi, F., Rexer, D.J., Alca zar-Roman, A.R., Onishko, H.M., and Wente, S.R. (2005). Interaction between the Shuttling mRNA Export Factor Gle1 and the Nucleoporin hCG1: A Conserved Mechanism in the Export of Hsp70 mRNA. Molecular Biology of the Cell *16*, 4303–4315.

Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J.W., Kim, V.N., and Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. Cell *181*, 914–921.e10.

Kravic, B., Harbauer, A.B., Romanello, V., Simeone, L., -Nora Vögtle, F., Kaiser, T., Straubinger, M., Huraskin, D., Böttcher, M., Cerqua, C., et al. (2018). In mammalian skeletal muscle, phosphorylation of TOMM22 by protein kinase CSNK2/CK2 controls mitophagy. Autophagy *14*, 311–335.

Kumar, R., Khandelwal, N., Thachamvally, R., Tripathi, B.N., Barua, S., Kashyap, S.K., Maherchandani, S., and Kumar, N. (2018). Role of MAPK/MNK1 signaling in virus replication. Virus Res. *253*, 48–61.

Labaronne, A., Milles, S., Donchet, A., Jensen, M.R., Blackledge, M., Bourhis, J.-M., Ruigrok, R.W.H., and Crépin, T. (2017). Structural analysis of the complex between influenza B nucleoprotein and human importin-α. Sci. Rep. *7*, 17164.

Lapointe, C.P., Grosely, R., Johnson, A.G., Wang, J., Fernández, I.S., and Puglisi, J.D. (2020). Dynamic competition between SARS-CoV-2 NSP1 and mRNA on the human ribosome inhibits translation initiation.

Lee, S.-D., Yu, K.-L., Park, S.-H., Jung, Y.-M., Kim, M.-J., and You, J.-C. (2018). Understanding of the functional role (s) of the Activating Transcription Factor 4 (ATF4) in HIV regulation and production. BMB Rep. *51*, 388.

Legnini, I., Alles, J., Karaiskos, N., Ayoub, S., and Rajewsky, N. (2019). FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. Nat. Methods *16*, 879–886.

Lenarcic, E.M., Landry, D.M., Greco, T.M., Cristea, I.M., and Thompson, S.R. (2013). Thiouracil cross-linking mass spectrometry: a cell-based method to identify host factors involved in viral amplification. J. Virol. *87*, 8697–8712.

Le Rouzic, E., Mousnier, A., Rustum, C., Stutz, F., Hallberg, E., Dargemont, C., and Benichou, S. (2002). Docking of HIV-1 Vpr to the nuclear envelope is mediated by the interaction with the nucleoporin hCG1. J. Biol. Chem. *277*, 45091–45098.

Li, D., Li, S., Sun, Y., Dong, H., Li, Y., Zhao, B., Guo, D., Weng, C., and Qiu, H.-J. (2013). Poly(C)-Binding Protein 1, a Novel Npro-Interacting Protein Involved in Classical Swine Fever Virus Growth. J. Virol. *87*, 2072–2080.

Li, D., Zhang, J., Yang, W., He, Y., Ru, Y., Fu, S., Li, L., Liu, X., and Zheng, H. (2019). Poly (rC) binding protein 2 interacts with VP0 and increases the replication of the foot-and-mouth disease virus. Cell Death Dis. *10*, 516.

Lima, S.A., Chipman, L.B., Nicholson, A.L., Chen, Y.-H., Yee, B.A., Yeo, G.W., Coller, J., and Pasquinelli, A.E. (2017). Short poly(A) tails are a conserved feature of highly expressed genes. Nat. Struct. Mol. Biol. *24*, 1057–1063.

Liu, J., Xu, Y., Stoleru, D., and Salic, A. (2012). Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin. Proc. Natl. Acad. Sci. U. S. A. *109*, 413–418.

Liu, T.-Y., Huang, H.H., Wheeler, D., Xu, Y., Wells, J.A., Song, Y.S., and Wiita, A.P. (2017). Time-Resolved Proteomics Extends Ribosome Profiling-Based Measurements of Protein Synthesis Dynamics. Cell Syst *4*, 636–644.e9.

Liu, X.-Y., Wei, B., Shi, H.-X., Shan, Y.-F., and Wang, C. (2010). Tom70 mediates activation of interferon regulatory factor 3 on mitochondria. Cell Res. *20*, 994–1011.

Lokugamage, K.G., Narayanan, K., Huang, C., and Makino, S. (2012). Severe acute respiratory syndrome coronavirus protein nsp1 is a novel eukaryotic translation inhibitor that represses multiple steps of translation initiation. J. Virol. *86*, 13598–13608.

Luo, Z., Dong, X., Li, Y., Zhang, Q., Kim, C., Song, Y., Kang, L., Liu, Y., Wu, K., and Wu, J. (2014). PolyC-binding protein 1 interacts with 5’-untranslated region of enterovirus 71 RNA in membrane-associated complex to facilitate viral replication. PLoS One *9*, e87491.

Meyuhas, O., and Kahan, T. (2015). The race to decipher the top secrets of TOP mRNAs. Biochim. Biophys. Acta *1849*, 801–811.

Mills, E.W., and Green, R. (2017). Ribosomopathies: There's strength in numbers. Science *358*, eaan2755.

Munro, S.A., Lund, S.P., Pine, P.S., Binder, H., Clevert, D.-A., Conesa, A., Dopazo, J., Fasold, M., Hochreiter, S., Hong, H., et al. (2014). Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures. Nat. Commun. *5*, 5125.

Nakagawa, K., Narayanan, K., Wada, M., Popov, V.L., Cajimat, M., Baric, R.S., and Makino, S. (2018). The Endonucleolytic RNA Cleavage Function of nsp1 of Middle East Respiratory Syndrome Coronavirus Promotes the Production of Infectious Virus Particles in Specific Human Cell Lines. J. Virol. *92*.

Napthine, S., Treffers, E.E., Bell, S., Goodfellow, I., Fang, Y., Firth, A.E., Snijder, E.J., and Brierley, I. (2016). A novel role for poly(C) binding proteins in programmed ribosomal frameshifting. Nucleic Acids Res. *44*, 5491–5503.

Narayanan, K., Ramirez, S.I., Lokugamage, K.G., and Makino, S. (2015). Coronavirus nonstructural protein 1: Common and distinct functions in the regulation of host and viral gene expression. Virus Res. *202*, 89–100.

Nury, H., Dahout-Gonzalez, C., Trezeguet, V., Lauquin, G.J.M., Brandolin, G., and Pebay-Peyroula, E. (2006). Relations between structure and function of the mitochondrial ADP/ATP carrier. Annu. Rev. Biochem. *75*, 713–741.

Ozadam, H., Geng, M., and Cenik, C. (2020). RiboFlow, RiboR and RiboPy: an ecosystem for analyzing ribosome profiling data at read length resolution. Bioinformatics *36*, 2929–2931.

Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2017). Biostrings: Efficient manipulation of biological strings. R Package Version *2*.

Patursky-Polischuk, I., Kasir, J., Miloslavski, R., Hayouka, Z., Hausner-Hanochi, M., Stolovich-Rain, M., Tsukerman, P., Biton, M., Mudhasani, R., Jones, S.N., et al. (2014). Reassessment of the role of TSC, mTORC1 and microRNAs in amino acids-meditated translational control of TOP mRNAs. PLoS One *9*, e109410.

Philippe, L., van den Elzen, A.M.G., Watson, M.J., and Thoreen, C.C. (2020). Global analysis of LARP1 translation targets reveals tunable and dynamic features of 5' TOP motifs. Proc. Natl. Acad. Sci. U. S. A. *117*, 5319–5328.

Plant, E.P., and Dinman, J.D. (2008). The role of programmed-1 ribosomal frameshifting in coronavirus propagation. Front. Biosci. *13*, 4873–4881.

Qian, Z., Xuan, B., Chapa, T.J., Gualberto, N., and Yu, D. (2012). Murine cytomegalovirus targets transcription factor ATF4 to exploit the unfolded-protein response. J. Virol. *86*, 6712–6723.

Quenouille, M.H. (1949). Problems in Plane Sampling. Ann. Math. Stat. *20*, 355–375.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Quinn, T.P., Erb, I., Richardson, M.F., and Crowley, T.M. (2018). Understanding sequencing data as compositions: an outlook and review. Bioinformatics *34*, 2870–2878.

Quinones-Valdez, G., Tran, S.S., Jun, H.-I., Bahn, J.H., Yang, E.-W., Zhan, L., Brümmer, A., Wei, X., Van Nostrand, E.L., Pratt, G.A., et al. (2019). Regulation of RNA editing by RNA-binding proteins in human cells. Commun Biol *2*, 19.

Raveh, A., Margaliot, M., Sontag, E.D., and Tuller, T. (2016). A model for competition for ribosomes in the cell. J. R. Soc. Interface *13*, 20151062.

Reiterer, V., Fey, D., Kolch, W., Kholodenko, B.N., and Farhan, H. (2013). Pseudophosphatase STYX modulates cell-fate decisions and cell migration by spatiotemporal regulation of ERK1/2. Proc. Natl. Acad. Sci. U. S. A. *110*, E2934–E2943.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rodriguez, J.M., Rodriguez-Rivas, J., Di Domenico, T., Vázquez, J., Valencia, A., and Tress, M.L. (2018). APPRIS 2017: principal isoforms for multiple gene sets. Nucleic Acids Res. *46*, D213–D217.

Rodriguez-Viciana, P., Oses-Prieto, J., Burlingame, A., Fried, M., and McCormick, F. (2006). A phosphatase holoenzyme comprised of Shoc2/Sur8 and the catalytic subunit of PP1 functions as an M-Ras effector to modulate Raf activity. Mol. Cell *22*, 217–230.

Schubert, K., Karousis, E.D., Jomaa, A. *et al.* (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat Struct Mol Biol*

Shi, H., Li, H., Yuan, R., Guan, W., Zhang, X., Zhang, S., Zhang, W., Tong, F., Li, L., Song, Z., et al. (2018). PCBP1 depletion promotes tumorigenesis through attenuation of p27Kip1 mRNA stability and translation. J. Exp. Clin. Cancer Res. *37*, 187.

Sloan, K.E., Warda, A.S., Sharma, S., Entian, K.-D., Lafontaine, D.L.J., and Bohnsack, M.T. (2017). Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. RNA Biol. *14*, 1138–1152.

Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. *27*, 491–499.

Stern-Ginossar, N., Thompson, S.R., Mathews, M.B., and Mohr, I. (2019). Translational Control in Virus-Infected Cells. Cold Spring Harb. Perspect. Biol. *11*.

Su, M.-C., Chang, C.-T., Chu, C.-H., Tsai, C.-H., and Chang, K.-Y. (2005). An atypical RNA pseudoknot stimulator and an upstream attenuation signal for -1 ribosomal frameshifting of SARS coronavirus. Nucleic Acids Res. *33*, 4265–4275.

Sun, Y., Bao, Q., Xuan, B., Xu, W., Pan, D., Li, Q., and Qian, Z. (2018). Human Cytomegalovirus Protein pUL38 Prevents Premature Cell Death by Binding to Ubiquitin-Specific Protease 24 and Regulating Iron Metabolism. J. Virol. *92*.

Tanaka, T., Kamitani, W., DeDiego, M.L., Enjuanes, L., and Matsuura, Y. (2012). Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mRNA. J. Virol. *86*, 11128–11137.

Tcherkezian, J., Cargnello, M., Romeo, Y., Huttlin, E.L., Lavoie, G., Gygi, S.P., and Roux, P.P. (2014). Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5’TOP mRNA translation. Genes Dev. *28*, 357–371.

Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., Kratzat, H., Hayn, M., Mackens-Kiani, T., Cheng, J., et al. (2020) Science Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. 369, 1249-1255

Turner, D.H., and Mathews, D.H. (2010). NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. Nucleic Acids Res. *38*, D280–D282.

Tyagi, S., and Kramer, F.R. (1996). Molecular beacons: probes that fluoresce upon hybridization. Nat. Biotechnol. *14*, 303–308.

Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.-Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins Nature *583*, 711–719.

Wan, Q., Song, D., Li, H., and He, M.-L. (2020). Stress proteins: the biological functions in virus infection, present and challenges for target-based antiviral drug development. Signal Transduct Target Ther *5*, 125.

Wang, J., Ganaie, S.S., Cheng, F., Xu, P., Ning, K., Wang, X., Kleiboeker, S., Cheng, S., and Qiu, J. (2020). RNA Binding Motif Protein RBM45 Regulates Expression of the 11-Kilodalton Protein of Parvovirus B19 through Binding to Novel Intron Splicing Enhancers. MBio *11*.

Wei, J., Alfajaro, M.M., Hanna, R.E., DeWeirdt, P.C., Strine, M.S., Lu-Culligan, W.J., Zhang, S.-M., Graziano, V.R., Schmitz, C.O., Chen, J.S., et al. (2020). Genome-wide CRISPR screen reveals host genes that regulate SARS-CoV-2 infection. bioRxiv.

Wek, R.C., Jiang, H.-Y., and Anthony, T.G. (2006). Coping with stress: eIF2 kinases and translational control. Biochem. Soc. Trans. *34*, 7–11.

Wile, B.M., Ban, K., Yoon, Y.-S., and Bao, G. (2014). Molecular beacon-enabled purification of living cells by targeting cell type-specific mRNAs. Nat. Protoc. *9*, 2411–2424.

Ye, P., Liu, S., Zhu, Y., Chen, G., and Gao, G. (2010). DEXH-Box protein DHX30 is required for optimal function of the zinc-finger antiviral protein. Protein Cell *1*, 956–964.

Yuan, S., Peng, L., Park, J.J., Hu, Y., Devarkar, S.C., and Dong, M.B. (2020). Nonstructural protein 1 of SARS-CoV-2 is a potent pathogenicity factor redirecting host protein synthesis machinery toward viral RNA. bioRxiv.

Zhou, X., You, F., Chen, H., and Jiang, Z. (2012). Poly(C)-binding protein 1 (PCBP1) mediates housekeeping degradation of mitochondrial antiviral signaling (MAVS). Cell Res. *22*, 717–727.

Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. *9*, 133–148.

Author Contributions

Shilpa Rao: Data Curation, Investigation, Methodology, Validation, Formal Analysis, Visualization, Writing- Original Draft, Writing- Review and Editing.
Ian Hoskins: Data Curation, Investigation, Validation, Formal Analysis, Software, Visualization, Writing- Original Draft, Writing- Review and Editing.
Daniela Garcia: Data Curation, Investigation, Validation, Methodology, Formal Analysis, Writing- Original Draft, Writing- Review and Editing.
Tori Tonn: Data Curation, Investigation, Methodology, Formal Analysis, Visualization, Writing- Original Draft, Writing- Review and Editing.
Hakan Özadam: Data Curation, Investigation, Formal Analysis, Software, Validation, Visualization, Writing- Original Draft, Writing- Review and Editing.
Elif Sarinay Cenik: Conceptualization, Resources, Project Administration, Writing- Review and Editing.
Can Cenik: Conceptualization, Data Curation, Formal Analysis, Software, Supervision, Funding Acquisition, Investigation, Methodology, Project Administration, Visualization, Writing- Original Draft, Writing- Review and Editing.

Figure 1

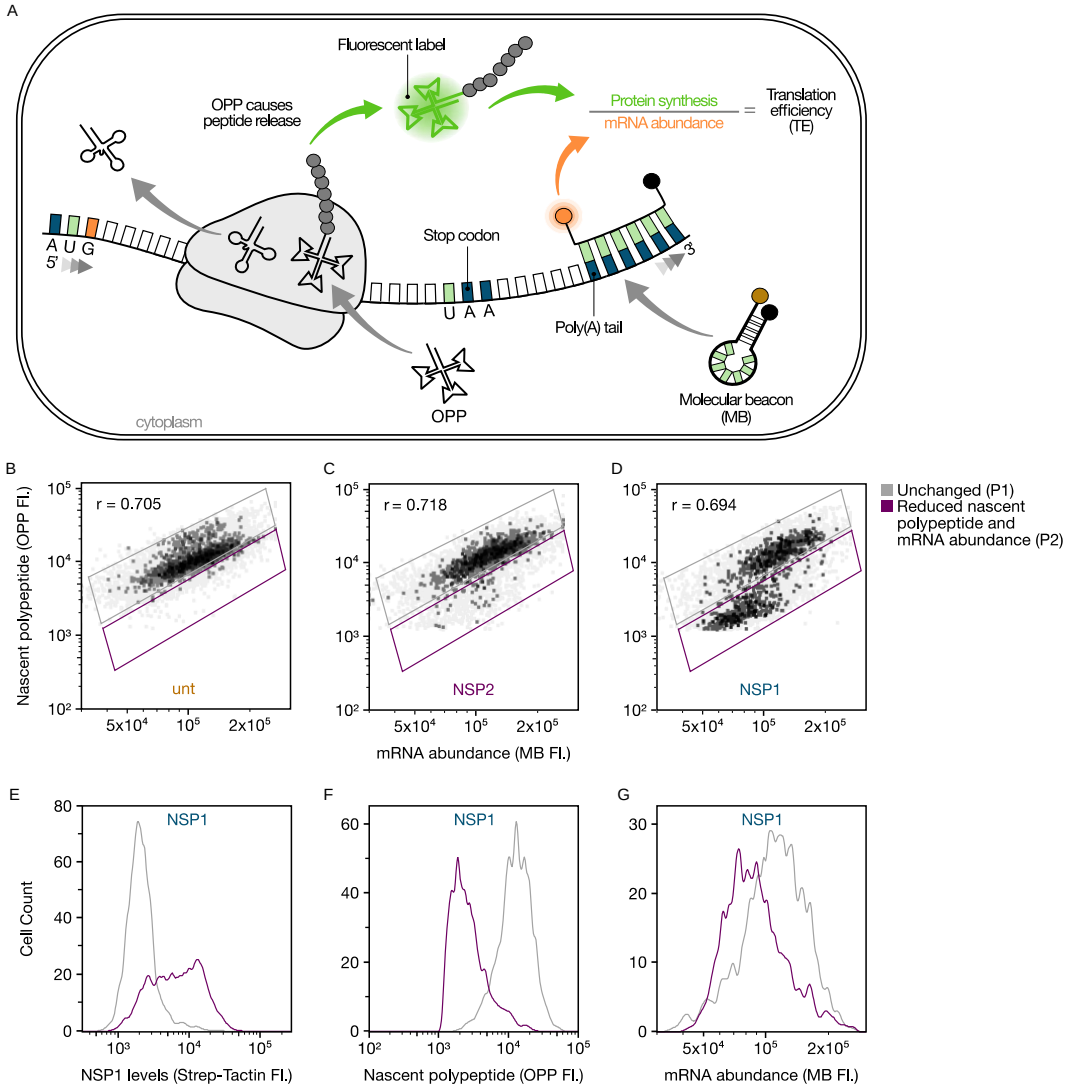

A
Fluorescent label
OPP causes
peptide release
Protein synthesis
mRNA abundance
=
Translation
efficiency
(TE)
A U G
5'
Stop codon
U A A
3'
Poly(A) tail
OPP
Molecular beacon
(MB)
cytoplasm
B
r = 0.705
unt
Nascent polypeptide (OPP Fl.)
C
r = 0.718
NSP2
D
r = 0.694
NSP1
mRNA abundance (MB Fl.)
Unchanged (P1)
Reduced nascent
polypeptide and
mRNA abundance (P2)
E
NSP1
Cell Count
NSP1 levels (Strep-Tactin Fl.)
F
NSP1
Nascent polypeptide (OPP Fl.)
G
NSP1
mRNA abundance (MB Fl.)

Figure 2

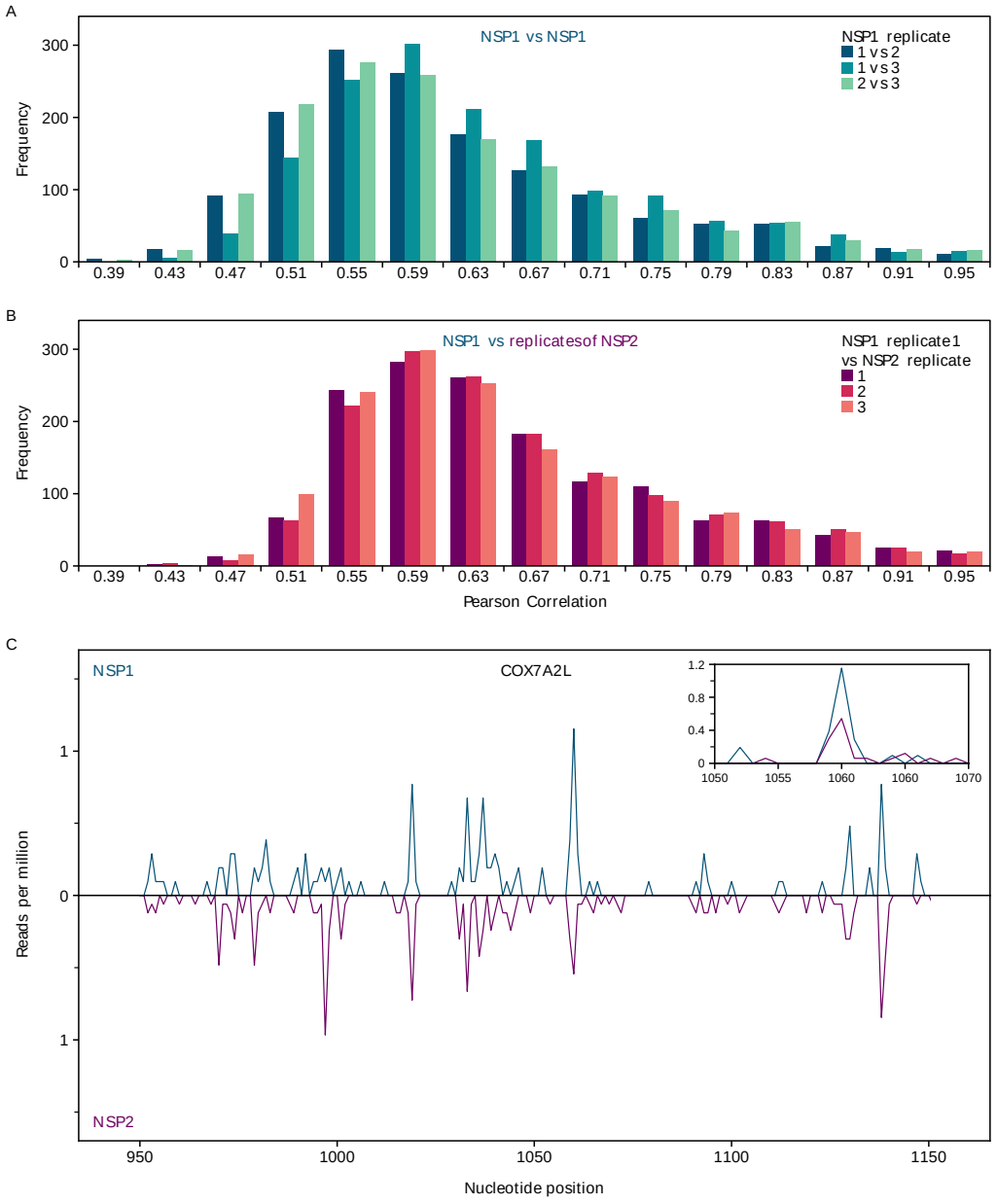

A
NSP1 vs NSP1
NSP1 replicate
1 v s 2
1 v s 3
2 v s 3
Frequency
0
100
200
300
0.39
0.43
0.47
0.51
0.55
0.59
0.63
0.67
0.71
0.75
0.79
0.83
0.87
0.91
0.95
B
NSP1 vs replicatesof NSP2
NSP1 replicate1
vs NSP2 replicate
1
2
3
Frequency
Pearson Correlation
C
NSP1
COX7A2L
NSP2
Reads per million
1
0
1
950
1000
1050
1100
1150
Nucleotide position
1.2
0.8
0.4
1050
1055
1060
1060
1070

Figure 3

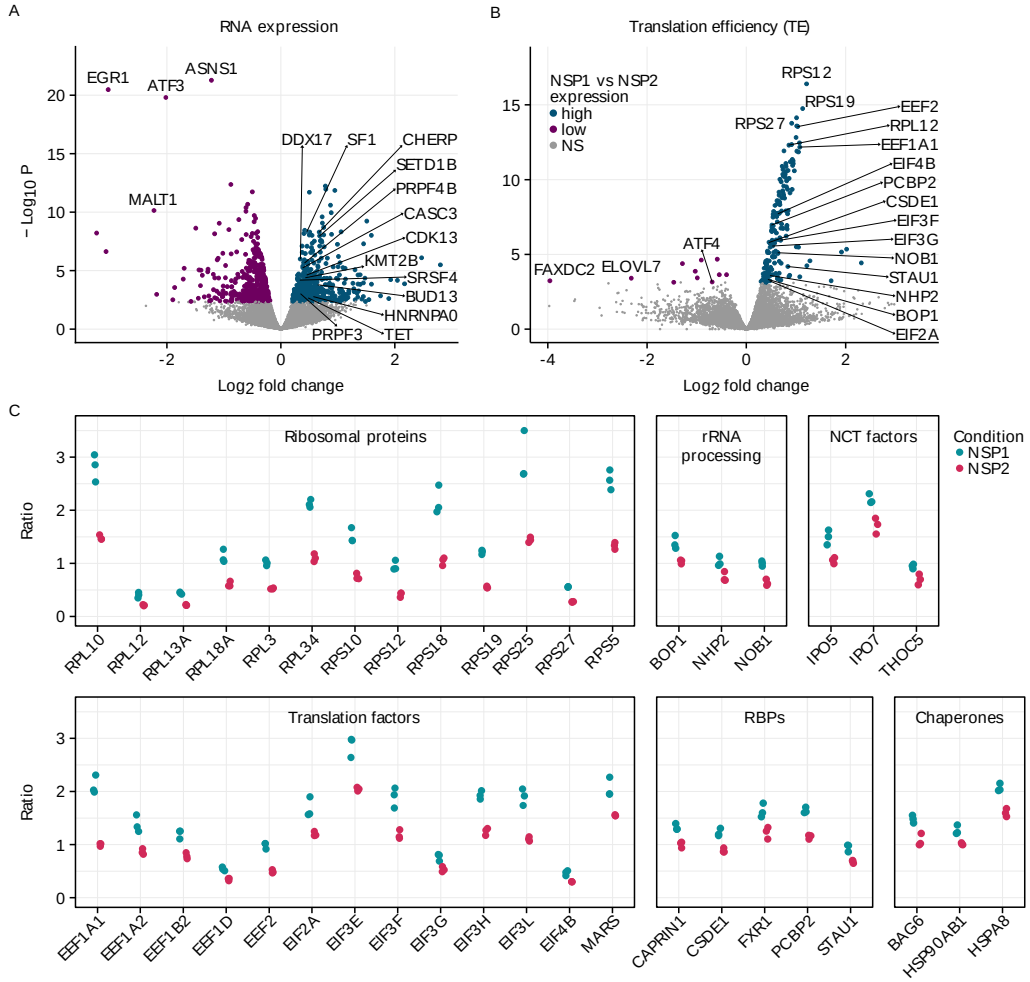

A
RNA expression
EGR1
ATF3
ASNS1
MALT1
DDX17
SF1
CHERP
SETD1B
PRPF4B
CASC3
CDK13
KMT2B
SRSF4
BUD13
HNRNPA0
PRPF3
TET
−Log10 P
Log2 fold change
B
Translation efficiency (TE)
NSP1 vs NSP2 expression
high
low
NS
RPS12
RPS19
RPS27
EEF2
RPL12
EEF1A1
EIF4B
PCBP2
CSDE1
EIF3F
EIF3G
NOB1
STAU1
NHP2
BOP1
EIF2A
ATF4
FAXDC2
ELOVL7
Log2 fold change
C
Ribosomal proteins
rRNA processing
NCT factors
Condition
NSP1
NSP2
Ratio
RPL10
RPL12
RPL13A
RPL18A
RPL3
RPL34
RPS10
RPS12
RPS18
RPS19
RPS25
RPS27
RPS5
BOP1
NHP2
NOB1
IPO5
IPO7
THOC5
Translation factors
RBPs
Chaperones
EEF1A1
EEF1A2
EEF1B2
EEF1D
EEF2
EIF2A
EIF3E
EIF3F
EIF3G
EIF3H
EIF3L
EIF4B
MARS
CAPRIN1
CSDE1
FXR1
PCBP2
STAU1
BAG6
HSP90AB1
HSPA8

Figure 4

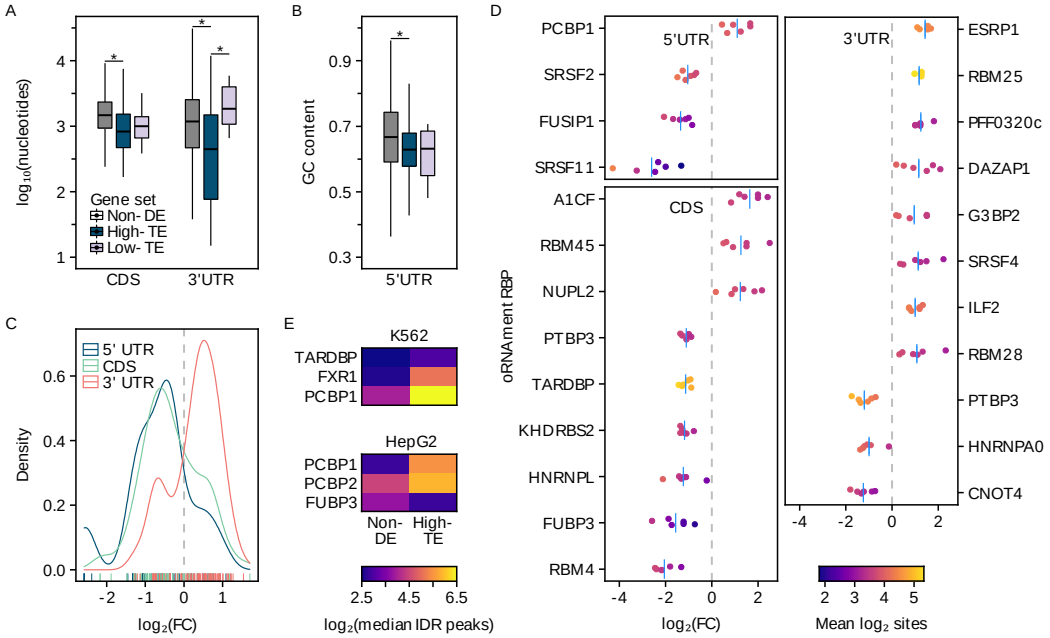
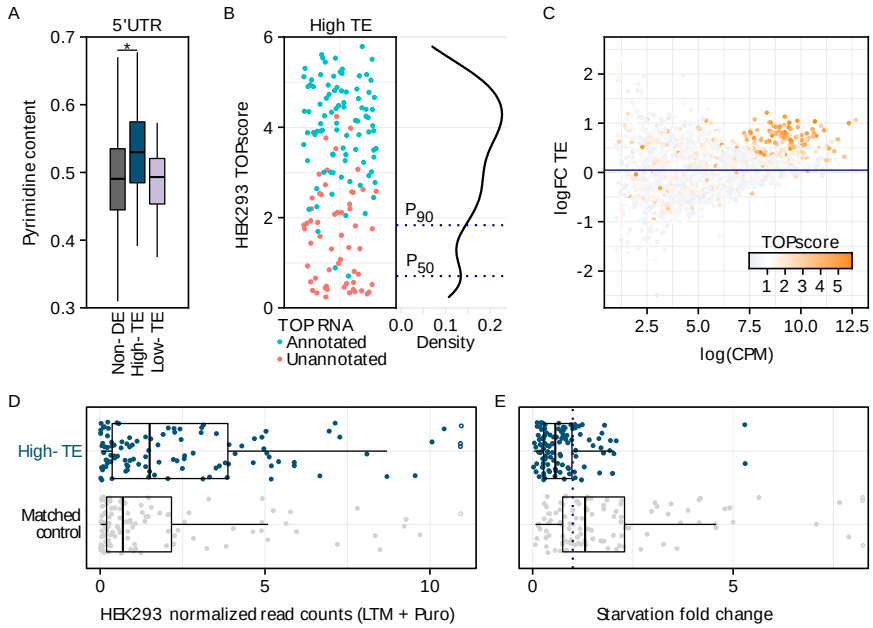
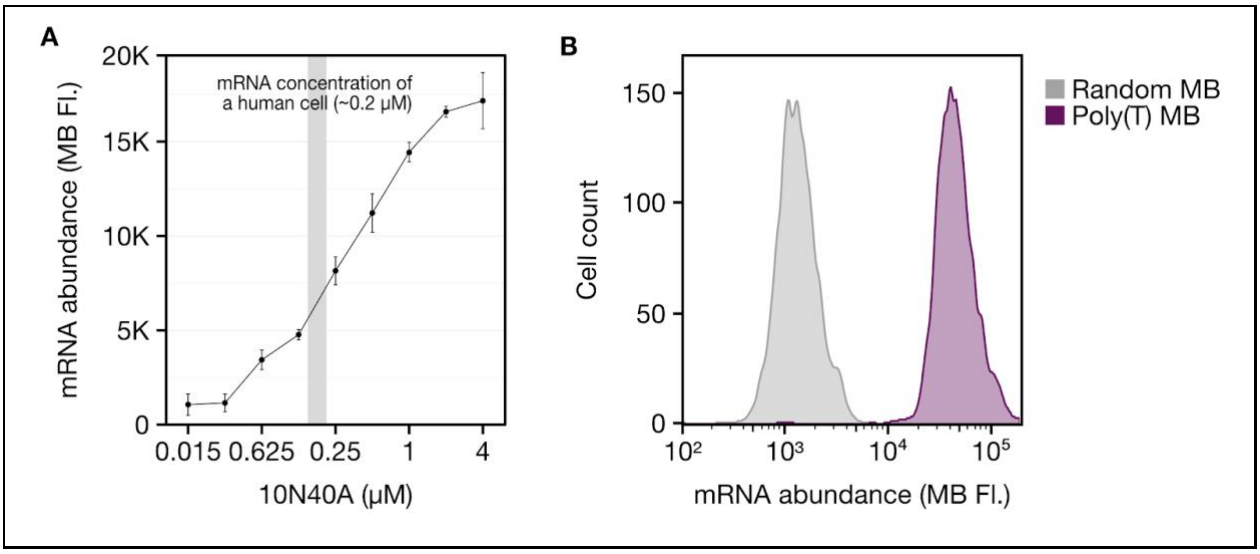

A
log10(nucleotides)
Gene set
Non- DE
High- TE
Low- TE
CDS
3'UTR
B
GC content
5'UTR
C
Density
5' UTR
CDS
3' UTR
log2(FC)
E
K562
TARDBP
FXR1
PCBP1
HepG2
PCBP1
PCBP2
FUBP3
Non- DE
High- TE
log2(median IDR peaks)
D
oRNAment RBP
5'UTR
CDS
3'UTR
PCBP1
SRSF2
FUSIP1
SRSF11
A1CF
RBM45
NUPL2
PTBP3
TARDBP
KHDRBS2
HNRNPL
FUBP3
RBM4
ESRP1
RBM25
PFF0320c
DAZAP1
G3BP2
SRSF4
ILF2
RBM28
PTBP3
HNRNPA0
CNOT4
log2(FC)
Mean log2 sites

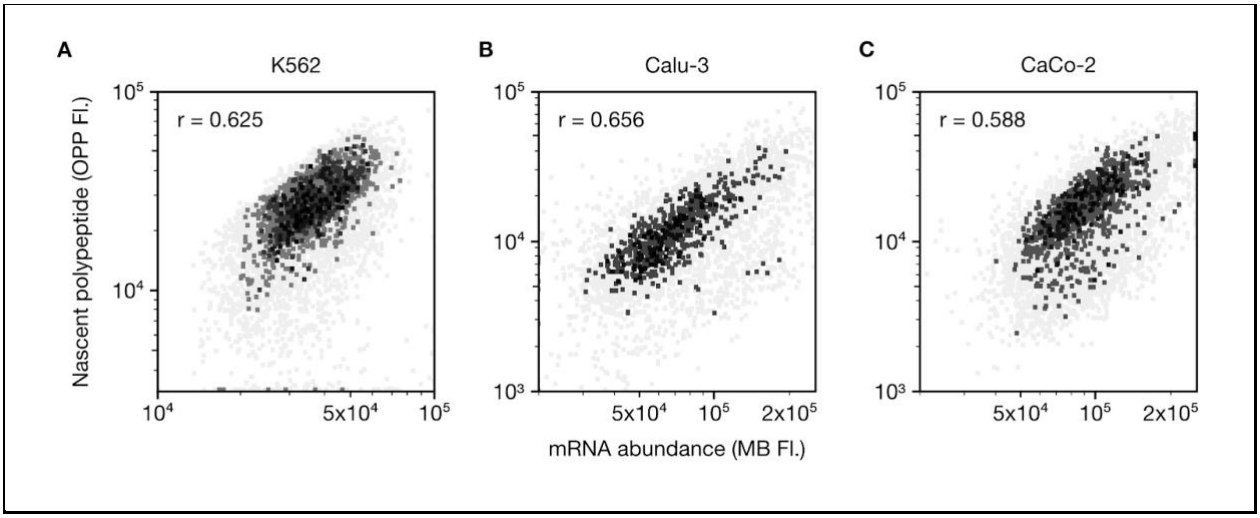
Figure 5


A
5'UTR
Pyrimidine content
0.7
0.6
0.5
0.4
0.3
*
Non- DE
High- TE
Low- TE
B
High TE
HEK293 TOPscore
6
4
2
0
TOP RNA
Annotated
Unannotated
0.0
0.1
0.2
Density
P90
P50
C
logFC TE
2
1
0
-1
-2
TOPscore
1 2 3 4 5
2.5
5.0
7.5
10.0
12.5
log(CPM)
D
High- TE
Matched control
0
5
10
HEK293 normalized read counts (LTM + Puro)
E
0
5
Starvation fold change

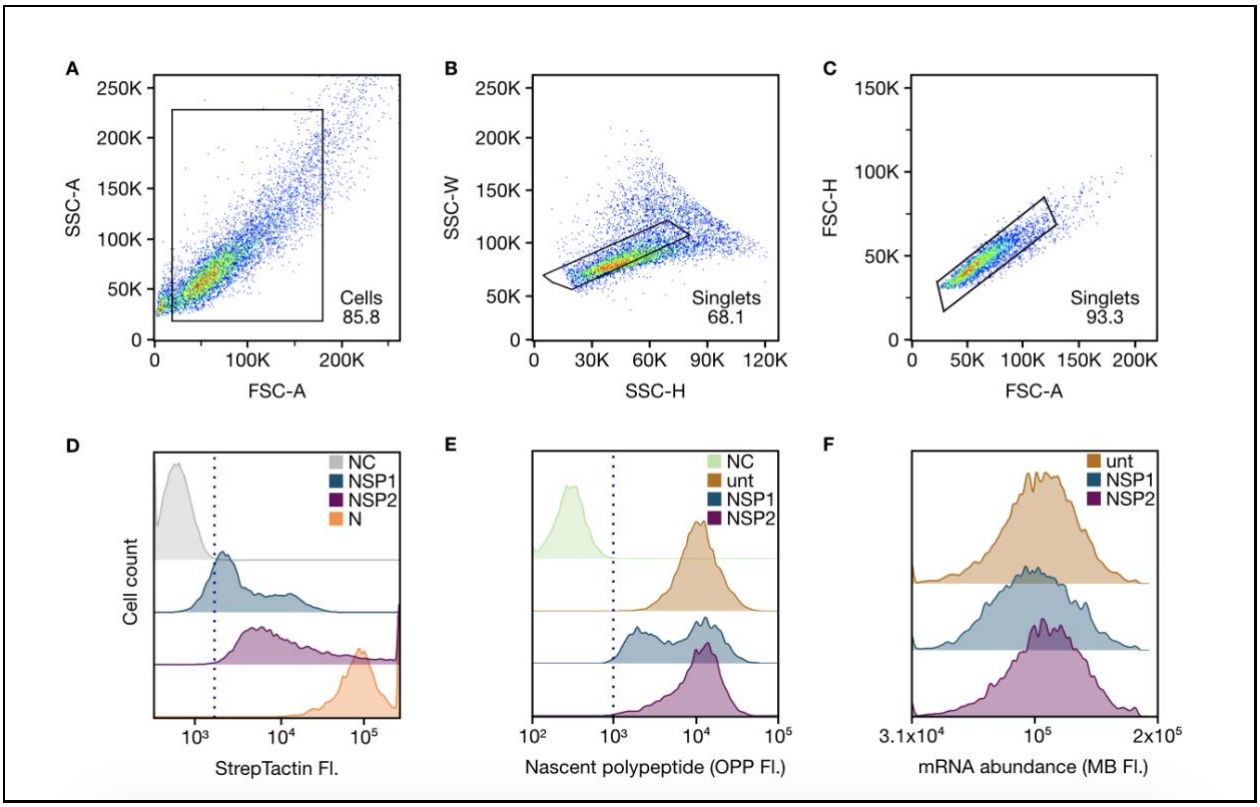
Supplementary Figure 1



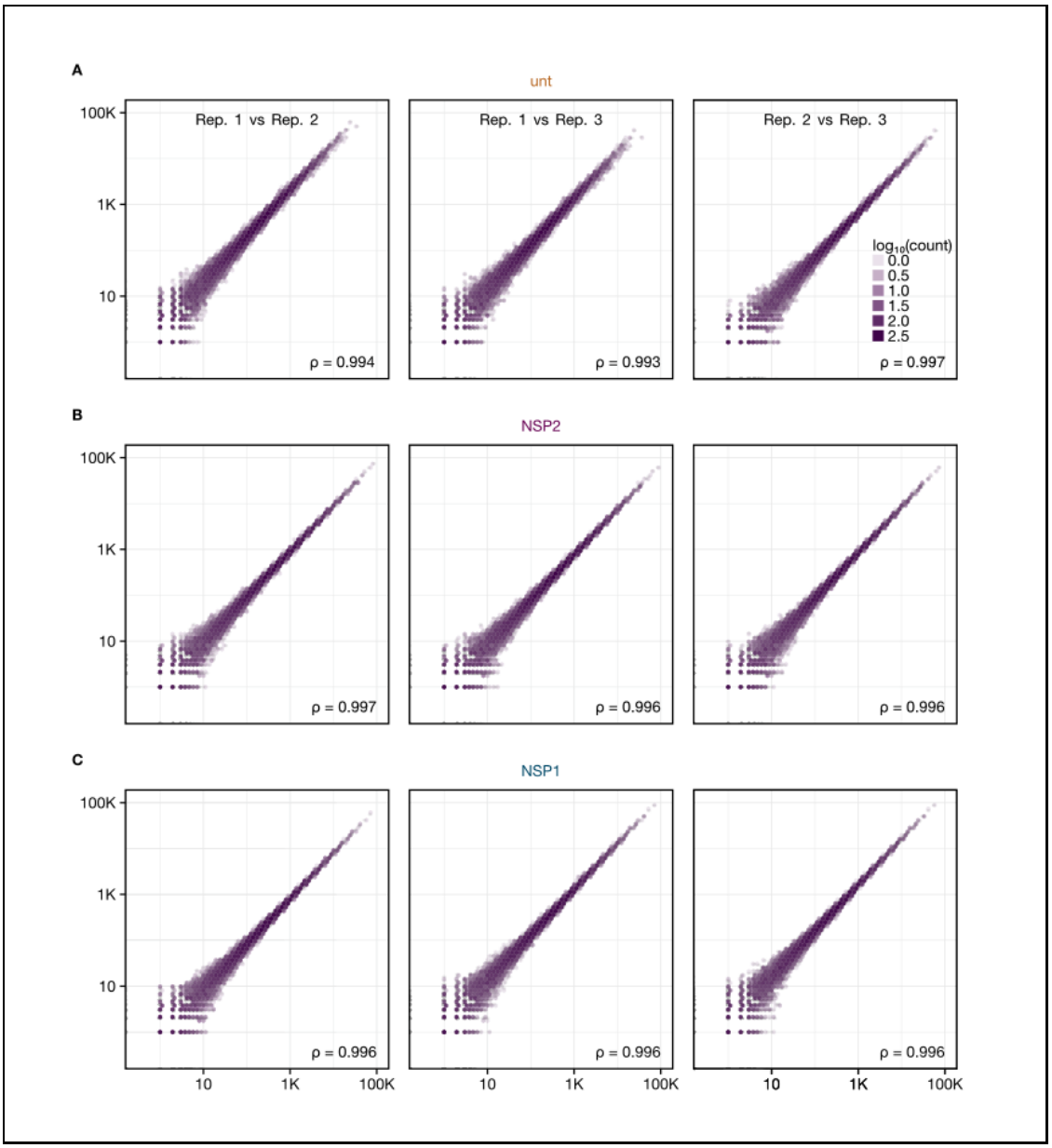
Supplementary Figure 2



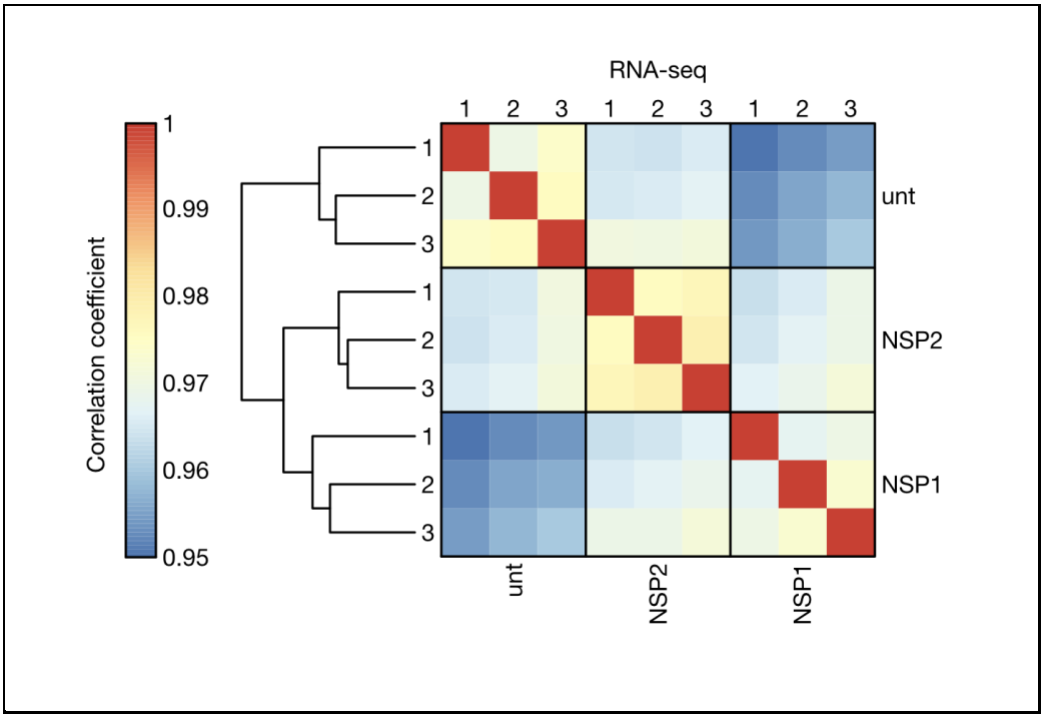
Supplementary Figure 3



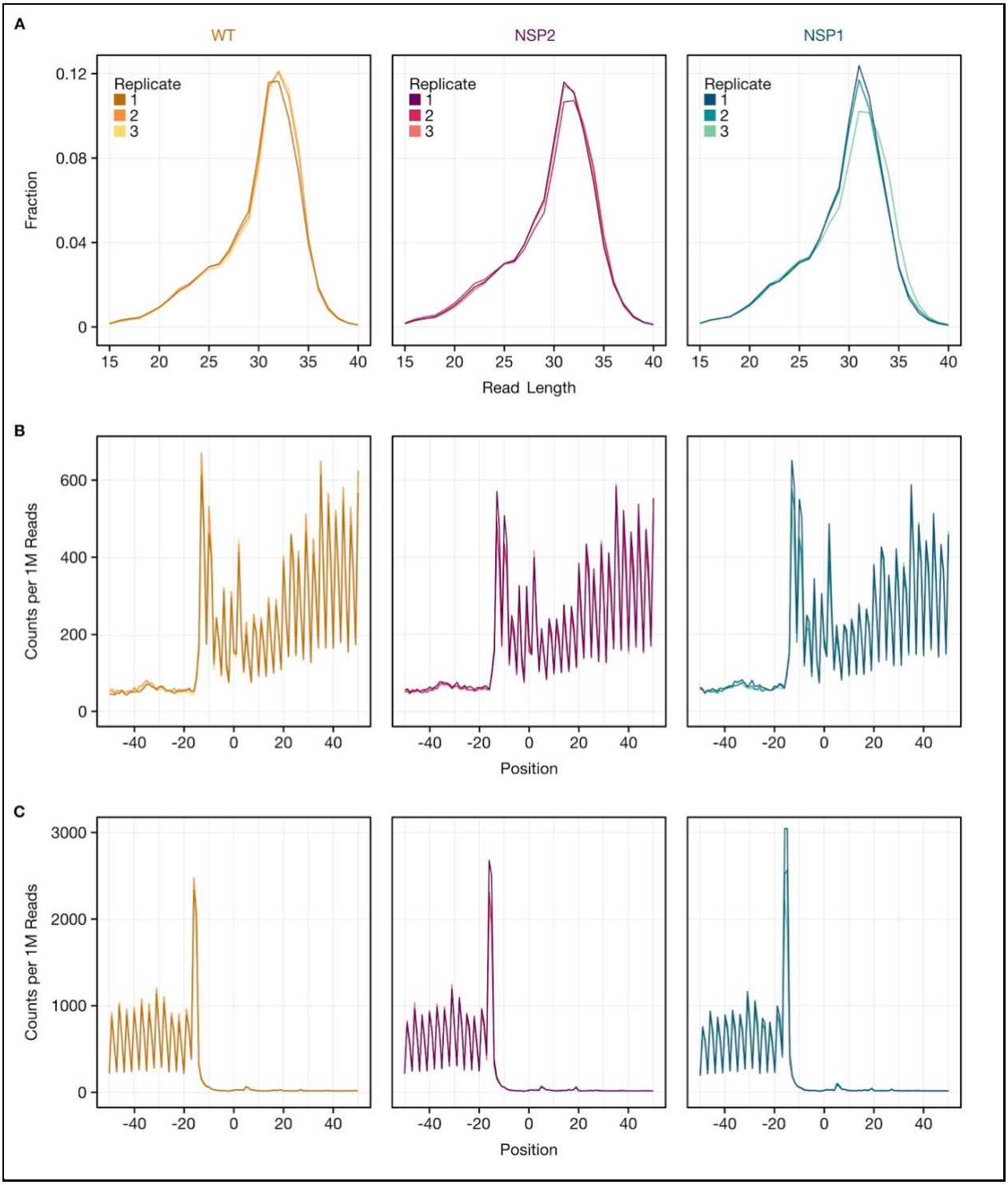
Supplementary Figure 4



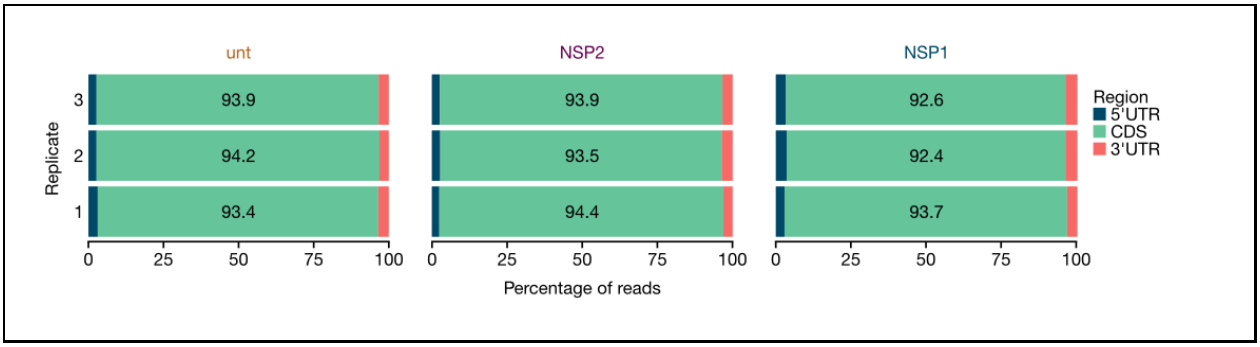
Supplementary Figure 5



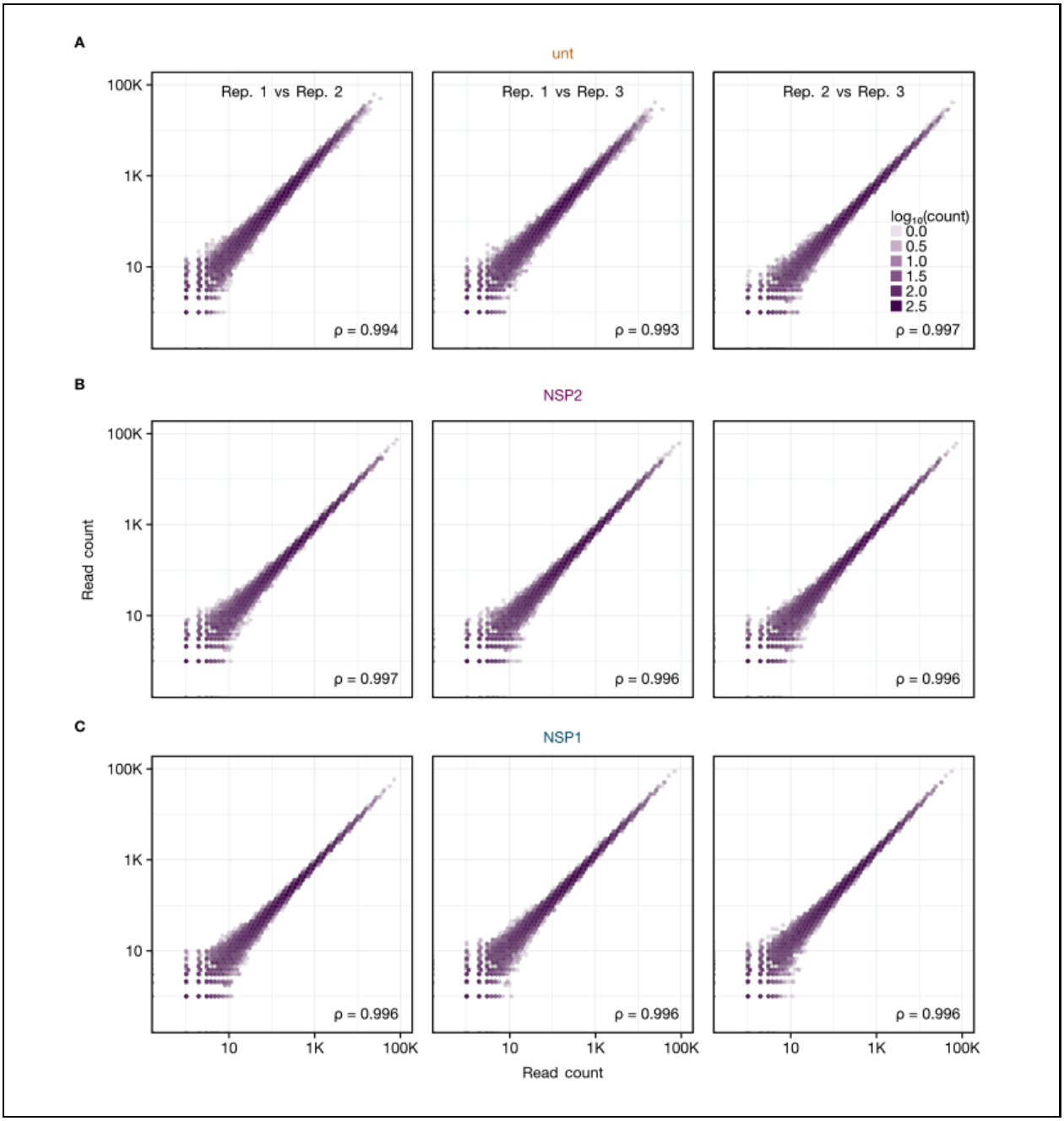
Supplementary Figure 6



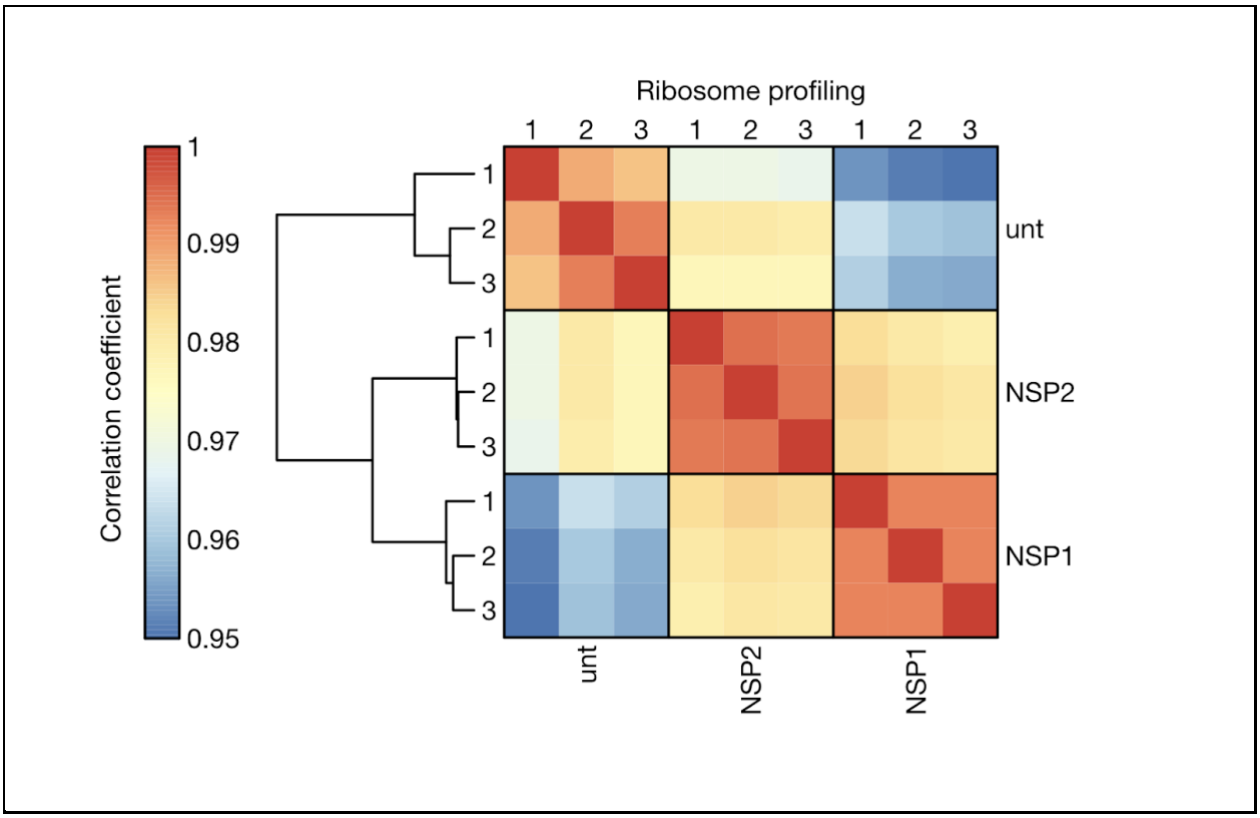
Supplementary Figure 7



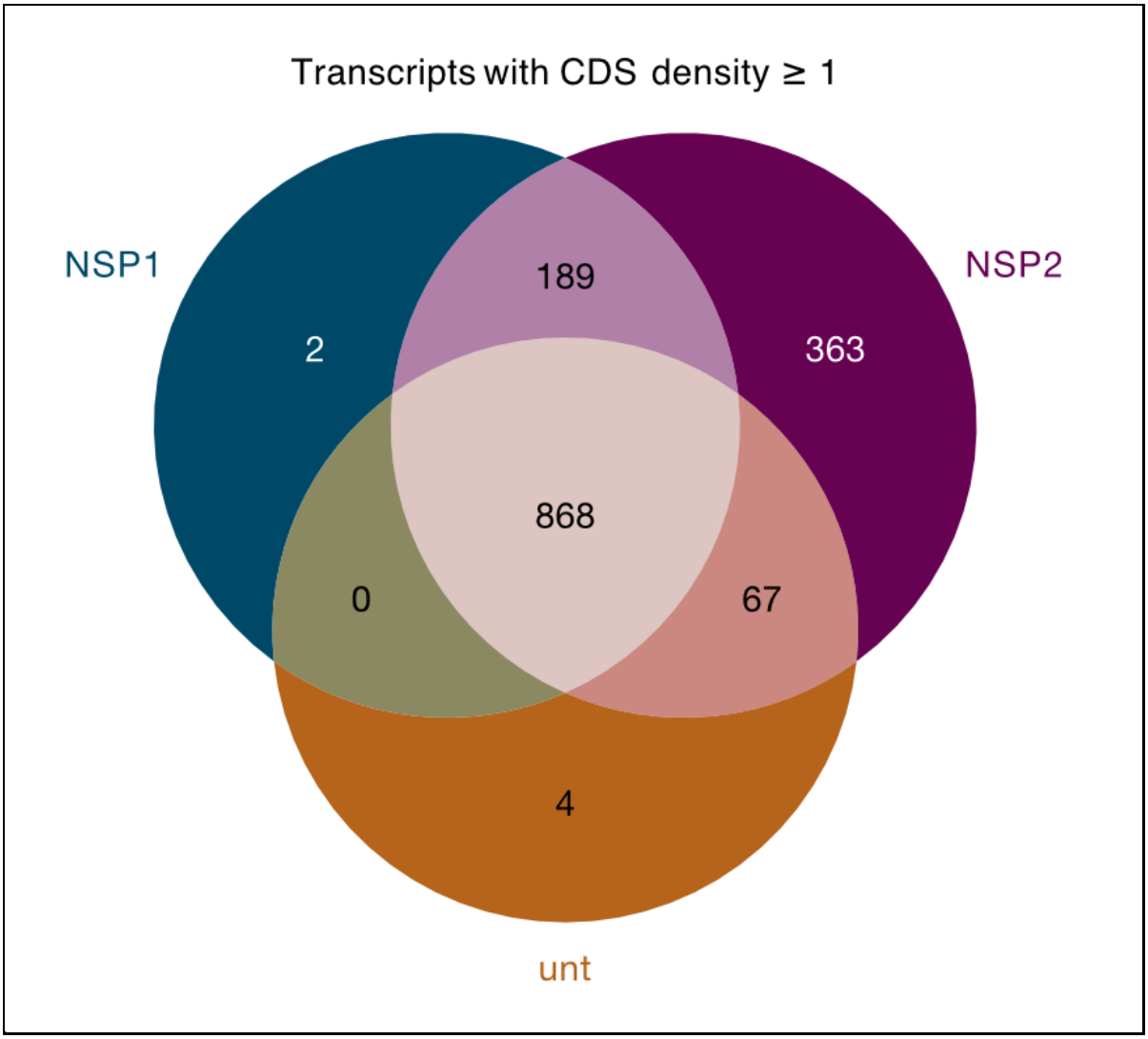
Supplementary Figure 8



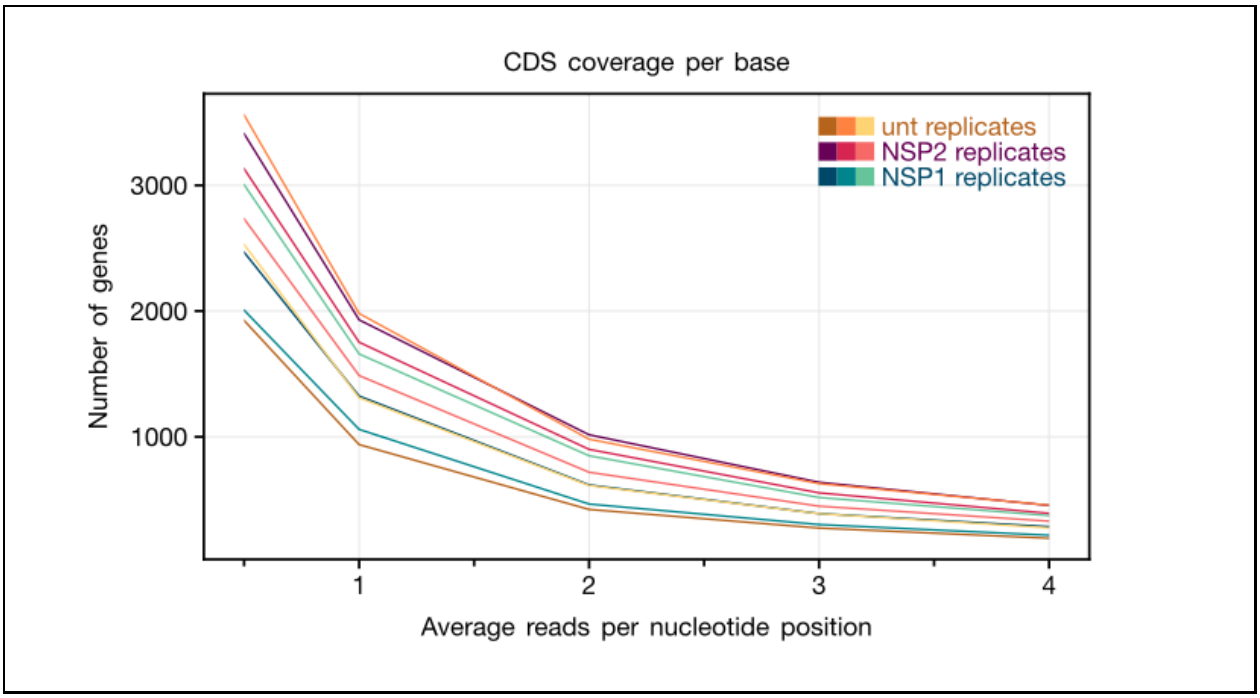
Supplementary Figure 9



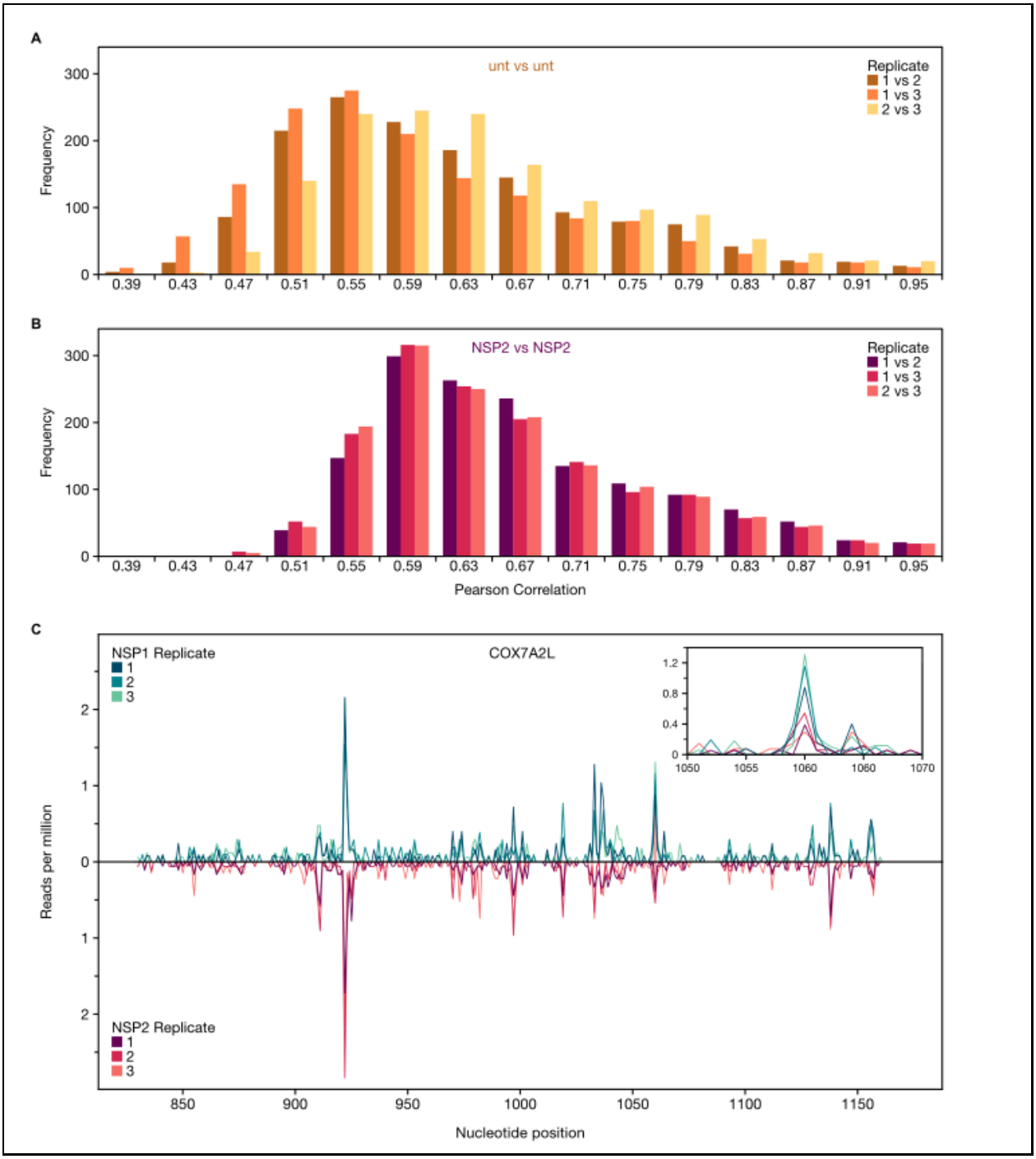
Supplementary Figure 10



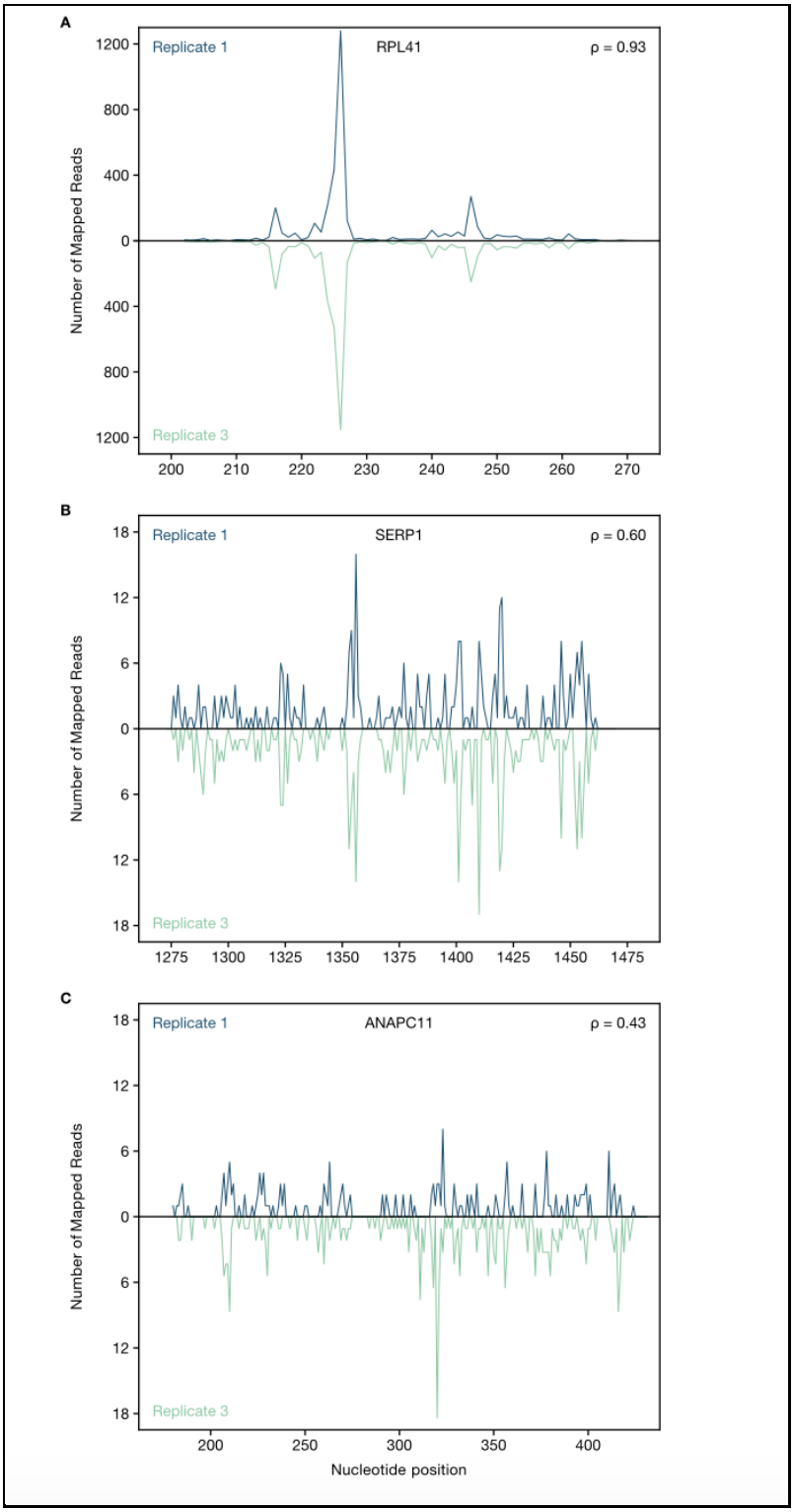
Supplementary Figure 11



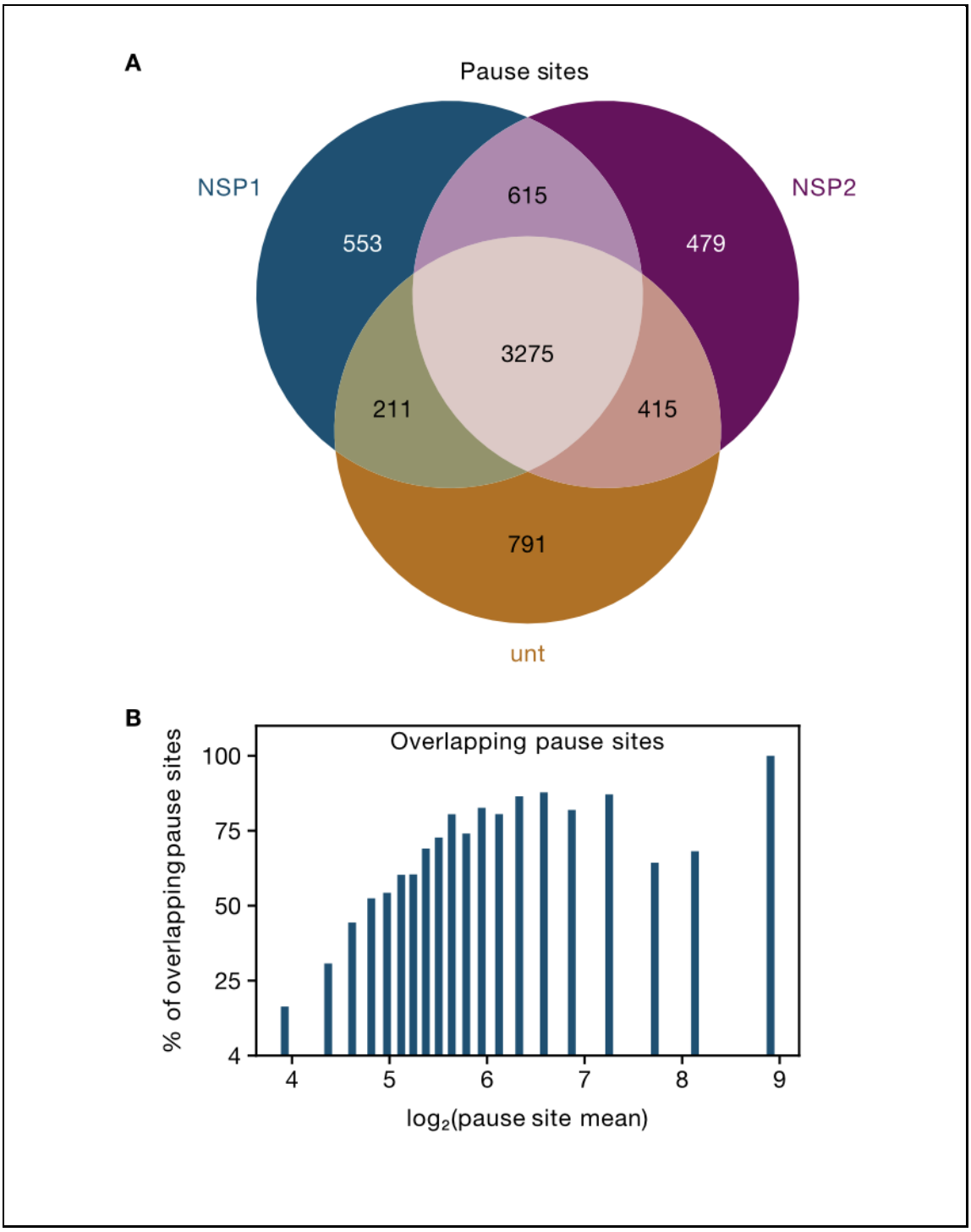
Supplementary Figure 12



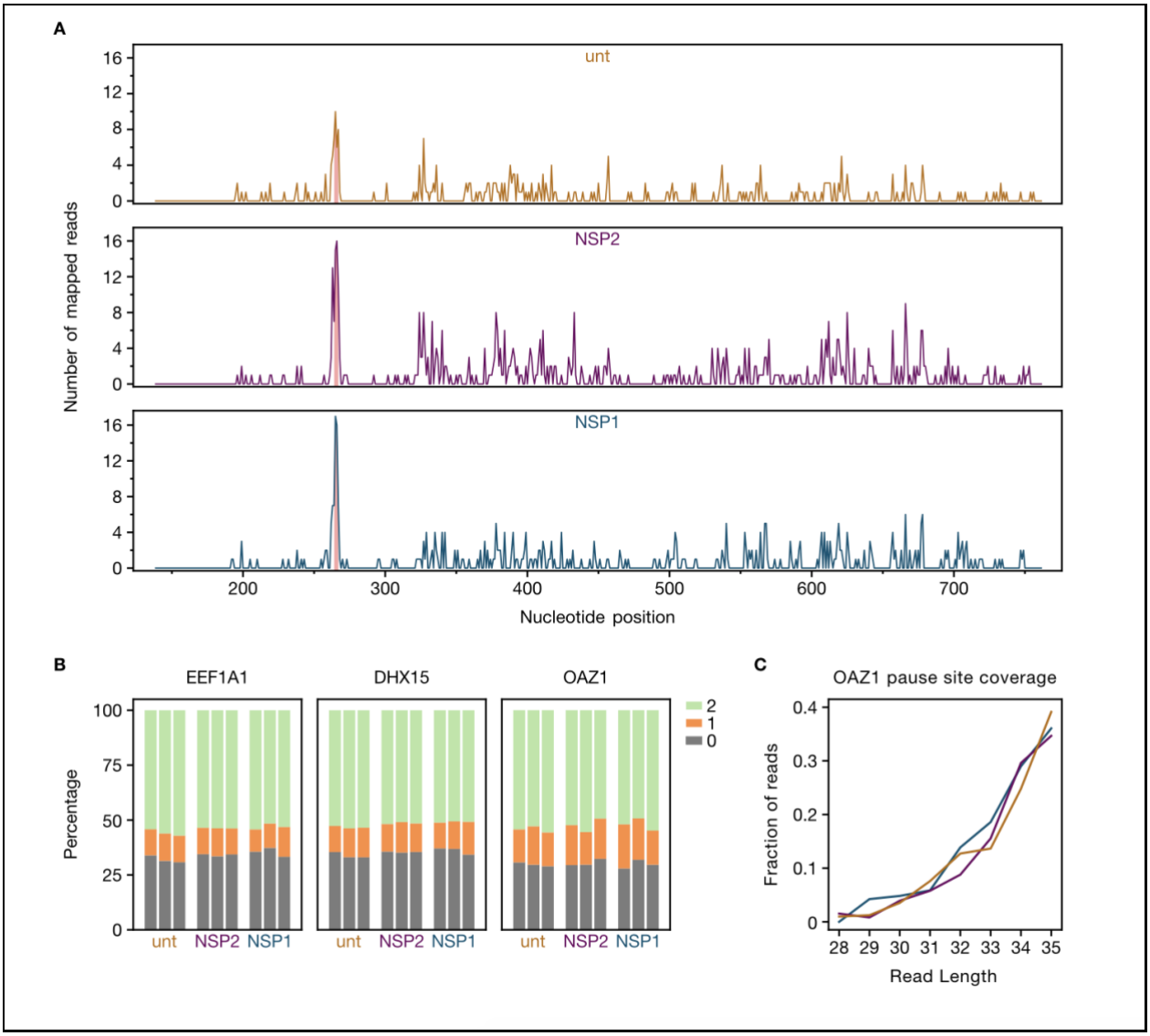
Supplementary Figure 13



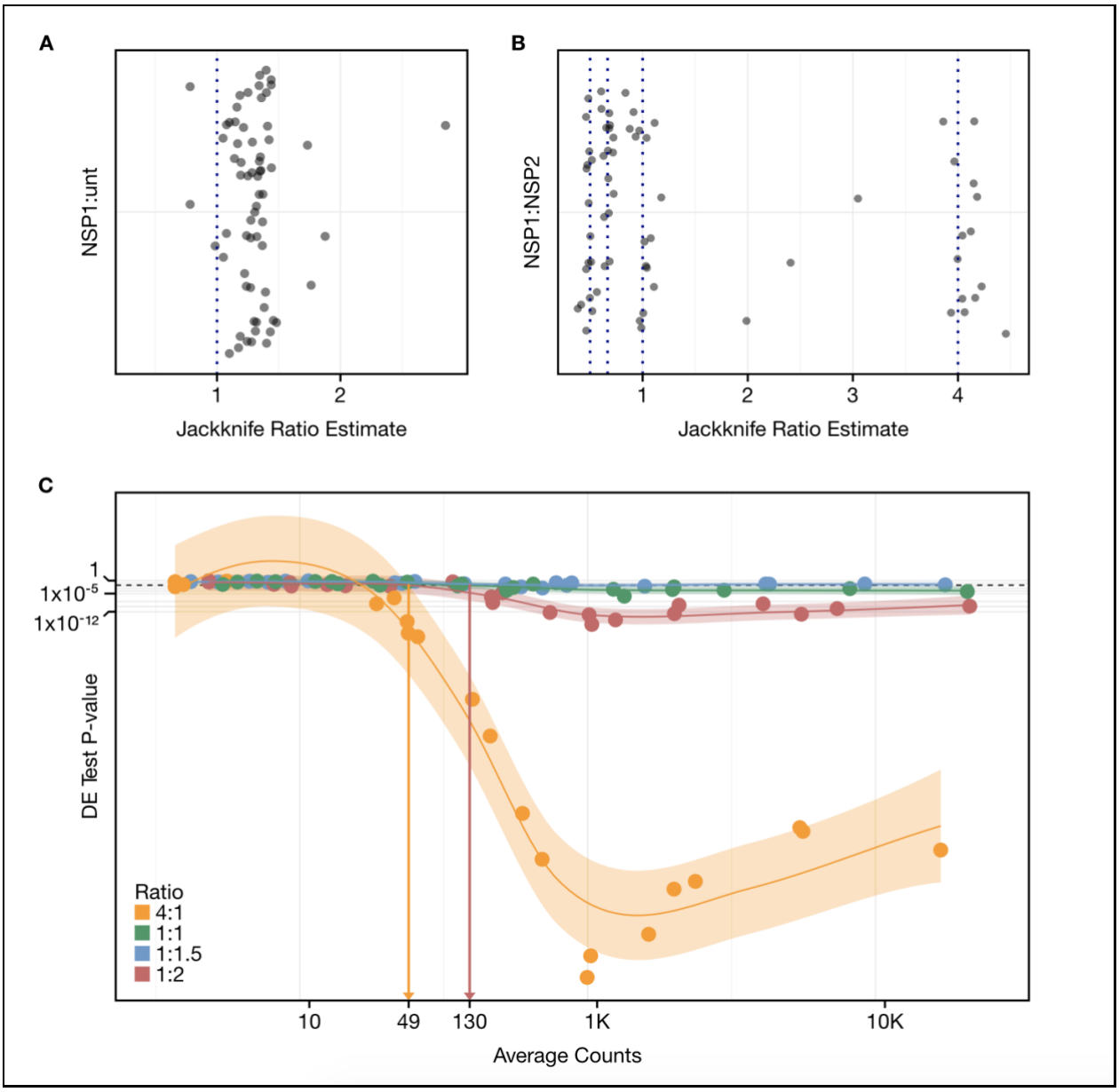
Supplementary Figure 14



Supplementary Figure 15

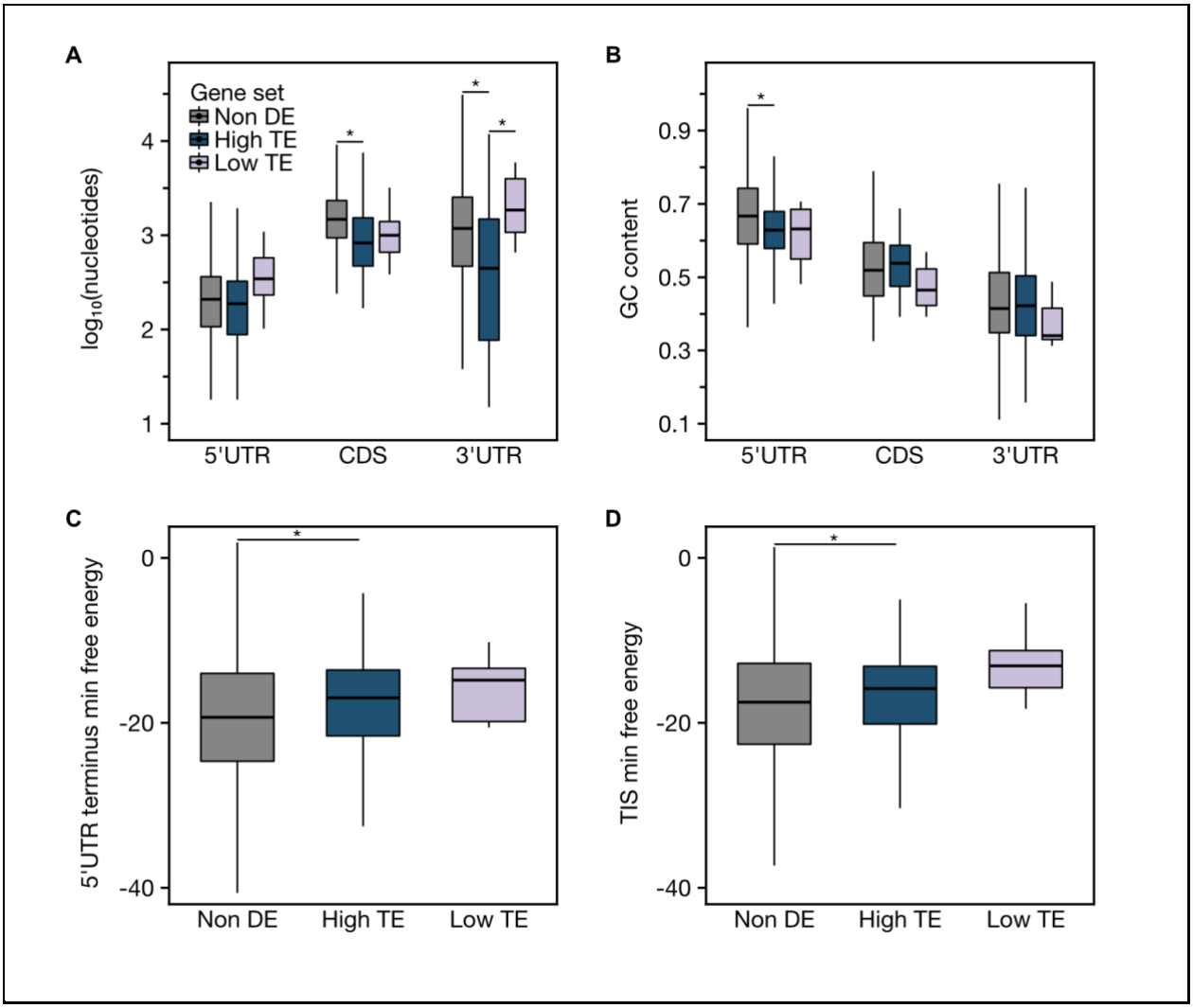


Supplementary Figure 16

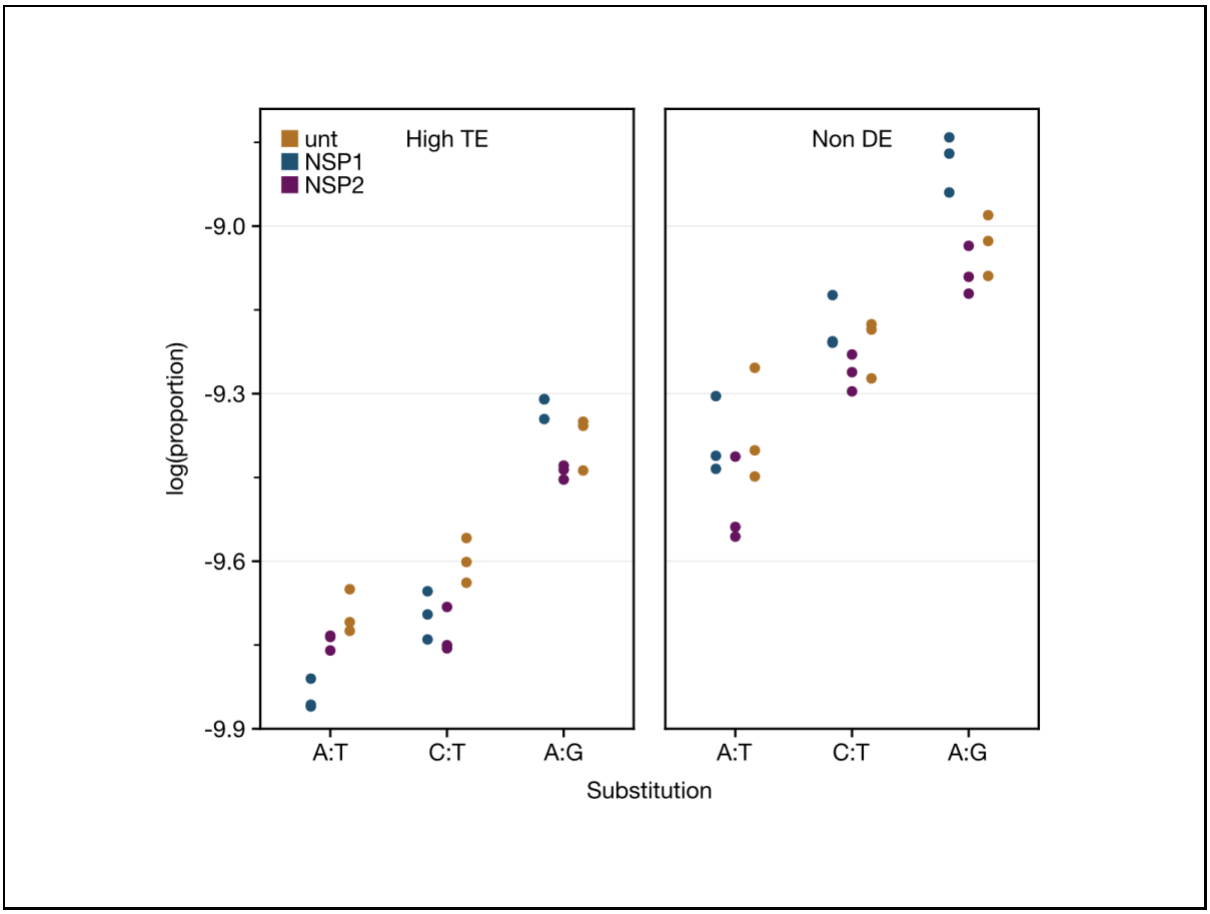
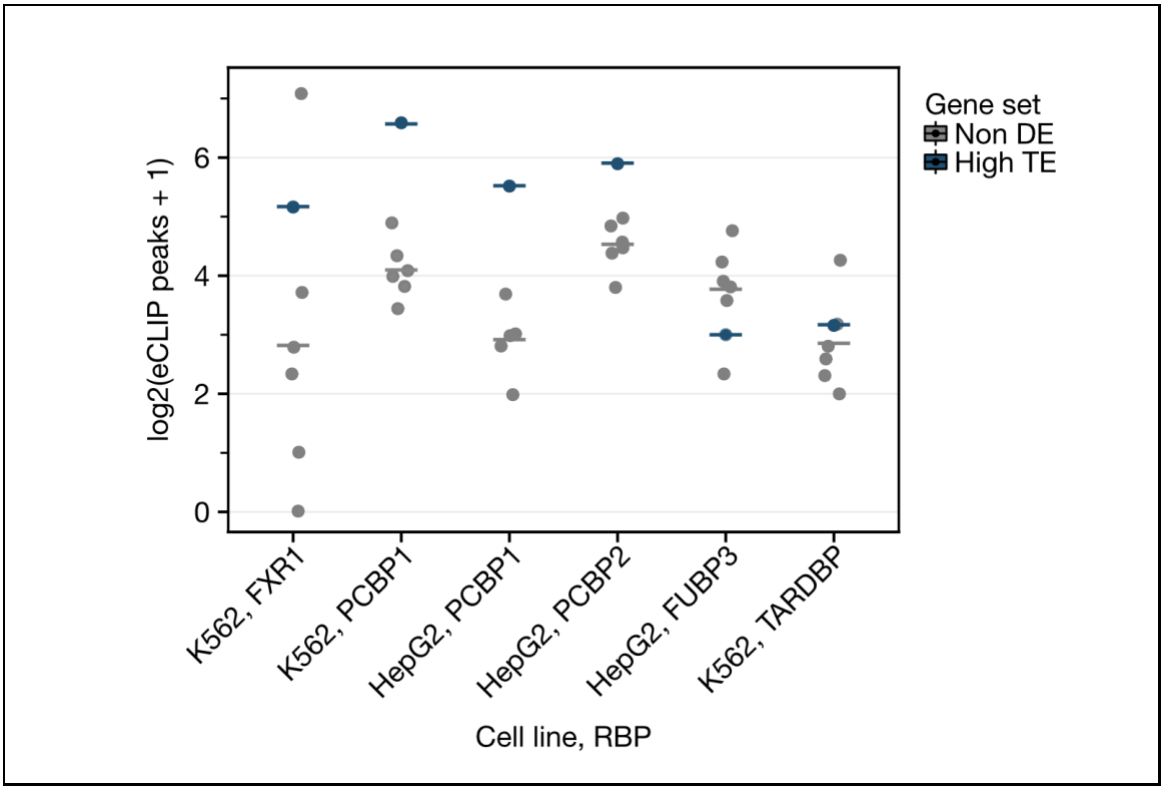


Supplementary Figure 17

.



Supplementary Figure 18



Supplementary Figure 19