bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

DDX3 depletion selectively represses translation of structured mRNAs

Lorenzo Calviello[1*], Srivats Venkataramanan[1*], Karol J. Rogowski[2], Emanuel Wyler[2], Malvika Tejura[1], Bao Thai[1], Jacek Krol[3], Witold Filipowicz[4], Markus Landthaler[2,5†], and Stephen N. Floor[1,6†]

[1]Department of Cell and Tissue Biology, University of California, San Francisco, San Francisco, California, USA, 94143
[2]Berlin Institute for Medical Systems Biology, Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, 13125, Berlin, Germany
[3]Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland
[4]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland
[5]IRI Life Sciences, Institute für Biologie, Humboldt Universität zu Berlin, Philippstraße 13, 10115, Berlin, Germany
[6]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California, USA, 94143
[*]Equal contributions
[†]Correspondence: stephen.floor@ucsf.edu, markus.landthaler@mdc-berlin.de

Abstract

DDX3 is an RNA chaperone of the DEAD-box family that regulates translation. Its yeast ortholog Ded1 controls the translation of nearly all mRNAs, whereas DDX3 is thought to regulate only a subset of mRNAs. However, the set of mRNAs that are regulated by DDX3 are unknown, along with the relationship between DDX3 binding and activity. Here, we use ribosome profiling, RNA-seq, and PAR-CLIP to define the set of mRNAs that are regulated by DDX3 in human cells. We find that while DDX3 binds most expressed mRNAs, depletion of DDX3 affects the translation level of only a small subset of the transcriptome. We further find that DDX3 binds a site on helix 16 of the human ribosome, placing it immediately adjacent to the mRNA entry channel and translation factor eIF4B. Translation changes caused by depleting DDX3 levels or through chemical inhibition mimicking a dominant negative allele are different. Taken together, our data defines the subset of the transcriptome that is responsive to DDX3 inhibition, with relevance for basic biology and disease states where DDX3 expression is altered.

Keywords: translational control, DEAD-box proteins, RNA, PAR-CLIP, ribosome profiling, post-transcriptional control

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Translation initiation is affected by mRNA structure. The DEAD-box RNA chaperone DDX3 and its yeast ortholog Ded1p facilitate translation initiation on mRNAs with RNA structures in their 5′ untranslated regions (UTRs) (Guenther et al. 2018; Lai et al. 2008; Oh et al. 2016; Soto-Rifo et al. 2012), a function that is essential in all eukaryotes (Sharma and Jankowsky 2014). Dysfunction in DDX3 is linked to numerous diseases and cancers, including medulloblastoma (Epling et al. 2015; Floor et al. 2016b; Jones et al. 2012; Kool et al. 2014; Oh et al. 2016; Pugh et al. 2012; Robinson et al. 2012; Valentin-Vega et al. 2016), many other cancer types (Sharma and Jankowsky 2014), and *de novo* developmental delay (Deciphering Developmental Disorders 2015; Lennox et al. 2018; Snijders Blok et al. 2015; Wang et al. 2018). Previous work studied how translation is altered by DDX3 variants found in medulloblastoma (Oh et al. 2016; Valentin-Vega et al. 2016), which are exclusively missense variants that preferentially target conserved residues. In contrast, hematological cancers like natural killer/T-cell lymphoma (Dufva et al. 2018; Jiang et al. 2015) and others (Wang et al. 2011; Schmitz et al. 2012) also have frequent variants in *DDX3X*, but they are mostly truncating or frameshift variants resulting in decreased expression. Changes in gene expression occurring as a result of decreased DDX3 levels remain incompletely understood.

Inactivation of Ded1 in yeast leads to polysome collapse and global downregulation of translation (Chuang et al. 1997; de la Cruz et al. 1997). More recent work showed that Ded1 is required for translation of most transcripts in yeast using genome-wide approaches (Guenther et al. 2018; Sen et al. 2015). In contrast, DDX3 depletion seems to only affect translation of a subset of expressed transcripts (Ku et al. 2018; Lai et al. 2008, 2010; Lee et al. 2008; Soto-Rifo et al. 2012). Despite the importance of DDX3 to normal function and its alteration in diverse disease states, the set of genes that depend on DDX3 for translation is not clearly defined. Moreover, it has been challenging to relate DDX3 binding to functional effects on bound mRNAs, and it was unclear if DDX3 is functioning outside of translation initiation given that binding was detected in coding sequences and 3′ UTRs (Oh et al. 2016; Valentin-Vega et al. 2016).

Here, we depleted DDX3 protein levels and measured alterations to translation and RNA abundance using ribosome profiling and RNA-seq. We also characterized DDX3 binding by PAR-CLIP, exploiting the presence of T>C mutations as a diagnostic hallmark of protein-RNA interactions. We observed robust interactions between DDX3 and transcript 5′ UTRs, as well as a specific and conserved site on the 40S ribosomal subunit. We found that transcripts with structured 5′ UTRs are preferentially affected by DDX3. We used an in vitro reporter system to conclude that decreases in ribosome occupancy upon DDX3 depletion are driven by 5′ UTRs. Taken together, our results support a model for DDX3 function where interactions with the small ribosomal subunit facilitate translation on messages with structured 5′ UTRs, which, when inactivated, pathologically deregulates protein synthesis.

Methods

NGS data pre-processing

Ribo-seq .fastq files were stripped of the adapter sequences using cutadapt. UMI sequences were removed and reads were collapsed to .fasta format. Reads shorter than 20 nt were removed. Reads were first aligned against rRNA (accession U13369.1), and to a collection of snoRNAs, tRNAs and miRNA (retrieved using the UCSC table browser) using bowtie2 in the 'local' alignment mode.

Read Alignment

Reads were mapped to the hg38 version of the genome (without scaffolds) using STAR 2.6.0a (Dobin et al. 2013) supplied with the GENCODE 25 .gtf file. The following parameter were used:

--alignEndsType EndToEnd (Ribo-seq and PAR-CLIP)
--outFilterMismatchNmax 2 --outFilterMismatchNoverLmax 0.04 (Ribo-seq)
--outFilterMismatchNmax 3 (PAR-CLIP and RNA-seq)
--outFilterMultimapNmax 20 --outSAMmultNmax 1 --outMultimapperOrder Random (All)

De-novo splice junction discovery was disabled for all datasets.

PAR-CLIP peak calling

Peak calling for PAR-CLIP reads was performed with PARalyzer v1.5 (Corcoran et al. 2011) in the "EXTEND_BY_READ" mode using the following parameters:
BANDWIDTH=3
CONVERSION=T>C
MINIMUM_READ_COUNT_PER_GROUP=5
MINIMUM_READ_COUNT_PER_CLUSTER=5
MINIMUM_READ_COUNT_FOR_KDE=5
MINIMUM_CLUSTER_SIZE=8
MINIMUM_CONVERSION_LOCATIONS_FOR_CLUSTER=1
MINIMUM_CONVERSION_COUNT_FOR_CLUSTER=1
MINIMUM_READ_COUNT_FOR_CLUSTER_INCLUSION=5
MINIMUM_READ_LENGTH=13
MAXIMUM_NUMBER_OF_NON_CONVERSION_MISMATCHES=0
Peaks with more than 10 reads were retained for subsequent analysis.

Additionally, wavClusteR (Comoglio et al. 2015) was used to call PAR-CLIP peaks using the "mrn" method and automatic thresholding calculation.

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Differential expression analysis

Count matrices for Ribo-seq and RNA-seq were built using reads mapping uniquely to CDS regions of protein-coding genes, using the Bioconductor packages GenomicFeatures, GenomicFiles and GenomicAlignments. DESeq2 (Love et al. 2014) was used to calculated log2FoldChange values between siRNA-treated cells and controls; only genes with BaseMean >20 for both RNA and Ribo-seq were retained. Differentially expressed genes ("RNA_up" and "RNA_down") were defined using the RNA differential expression, with an adjusted p-value cutoff of .01. Riborex (Li et al. 2017) was used to quantify translation regulation, using the DESeq2 engine and an adjusted p-value cutoff of .05 to define differentially translated genes ("TE_up" and "TE_down"). The "RNA_up" class was further divided in 2 groups: a mixture of 2 gaussians was fit to a vector comprising the difference between the log2FC of RNA and Ribo, using the mixtools package; a small difference is indicative of concordant RNA and Ribo change, while a positive value indicates a change in RNA larger than the change in the Ribo-seq (thus resembling the "TE_down" set). The means for the two gaussian distributions were set at the values of 0 and 0.4; genes belonging to the second distribution (according to the posterior probability of the fit) were removed. GO enrichment analysis was performed with the topGO package, using the fisher test with default parameters. Gene categories with a number of annotated genes between 3 and 3000 were retained.

Meta-transcript profiles

PARalyzer peaks (and peaks from the POSTAR2 repository (Zhu et al.)) were mapped on transcript coordinates using one coding transcript per gene: such transcript was chosen to have the longest 5′ UTR and the most common annotated start codon for that gene. Transcript positions were converted into bins using 33 bins for each 5′ UTR, 100 bins for each CDS and 67 bins for each 3′ UTR. Peak scores (ModeScore for PARalyzer peaks) were normalized for each transcript (to sum up to 1), and values were summed for each bin to build aggregate profiles. When plotting profiles for different RBPs. the aggregate profiles were further normalized to a sum of 1. To build the average meta-transcript profile in Figure 5 and S4, Conversion specificity values were averaged per transcript bin and transcript class.

Additional transcript features analysis

To compare read mapping locations within transcripts, a window of 25nt around the start codon was subtracted from annotated 5′ UTRs and CDS. 5′ UTRs and CDS regions were merged for each gene using the "reduce" command from the GenomicRanges package. Counts on 5′ UTR and CDS were first averaged between replicates. The ratio 5′ UTR to CDS of these counts were calculated for each gene, in the siRNA and controls condition. The log2 of the ratio siDDX3/control for those values represents the skew of counts towards 5′ UTR in the siDDX3 condition:

$$5'UTR\ ribosome\ skew = log_2\left[\left(\frac{RP^{5'UTR}}{RP^{CDS}}\right)_{siDDX3} \Big/ \left(\frac{RP^{5'UTR}}{RP^{CDS}}\right)_{control}\right]$$

RNA *in silico* folding was performed on 5′ UTRs sequences using RNAlfold (Lorenz et al. 2016) with default parameters. Average ΔG values per nucleotide were calculated averaging the ΔG values of each structure overlapping that nucleotide. Single-nucleotide values were averaged into bins using the same approach used to create meta-transcript profiles above.
%GC content and T>C transition specificity (defined as *ConversionEventCount* / (*ConversionEventCount*+*NonConversionEventCount*)) for each PAR-CLIP peak were derived using the clusters.csv output file from PARalyzer. Peak scores for the wavClusteR peaks (using the *RelLogOdds* column) and crosslinking efficiency were extracted from the wavClusteR output.
Gviz was used to plot tracks for RNA-seq, Ribo-seq and PAR-CLIP over different genomic regions. Intron size was set to a maximum of 150 nt.

Source code for all analysis programs can be found here: https://github.com/lcalviell/DDX3_manuscript_analysis.

Ribosome profiling

Flp-In T-REx HEK293 cells transfected with control siPool and with DDX3X targeting siPools (siTools Biotech) were washed with PBS containing 100 µg/ml cycloheximide, flash frozen on liquid nitrogen, lysed in lysis buffer (20 mM TRIS-HCl pH 7.4, 150mM NaCl, 5 mM MgCl2 1 % (v/v) Triton X-100, 25 U/ml TurboDNase (Ambion), harvested, centrifuged at 20,000 g for 4 min at 4ºC and supernatants were stored at -80ºC. Thawed lysates were treated with RNase I (Ambion) at 2.5 U/µl for 45 min at room temperature with slow agitation. Further RNase activity was stopped by addition of SUPERase:In (Ambion). Next Illustra MicroSpin Columns S-400 HR (GE Healthcare) were used to enrich for ribosome complexes. RNA was extracted from column flow throughs with TRIzol (Ambion) reagent. Precipitated nucleic acids were further purified and concentrated with Zymo-Spin IIC column (Zymo Research). Obtained RNA was depleted of rRNAs with Ribo-Zero Gold Kit (Human/Mouse/Rat) kit (Illumina), separated in 17% Urea gel and stained with SYBR Gold (Invitrogen). Gel slices containing nucleic acids 27 to 30 nucleotides long were excised and incubated in a thermomixer with 0.3 M NaCl at 4ºC over night with constant agitation to elute RNA. After precipitation nucleic acids were treated with T4 polynucleotide kinase (Thermo Scientific). Purified RNA was ligated to 3′ and 5′ adapters, reverse transcribed and PCR amplified. The amplified cDNA was sequenced on a HiSeq2000 (Illumina).

PAR-CLIP Experiments

Flp-In T-REx HEK293 cells expressing FLAG/HA-tagged DDX3X were labelled with 100 µM 4SU for 16h. After labelling PAR-CLIP was performed generally as described (Hafner et al. 2010). Briefly, cells were UV-crosslinked and stored. Obtained cell pellets were lysed in 3 times the cell pellet

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

volume of NP-40 lysis buffer (50 mM HEPES-KOH at pH 7.4, 150 mM KCl, 2mM EDTA, 1mM NaF, 0.5% (v/v) NP-40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche)), incubated 10 min on ice and centrifuged at 13,000 rpm for 15 min at 4ºC. Supernatants were filtered through 5 µm syringe filter. Next, lysates were treated with RNase T1 (Fermentas) at final concentration of 1U/µl for 10 min at room temperature. Immunoprecipitation of the DDX3X/RNA complexes was performed with Flag magnetic beads (Sigma) and was followed by second RNase T1 (Fermentas) digestion at final concentration of 10U/µl for 10 min at room temperature. After IP beads were incubated with calf intestinal phosphatase (NEB) and RNA fragments were radioactively end labelled using T4 polynucleotide kinase (Fermentas). Crosslinked protein–RNA complexes were resolved on a 4–12% NuPAGE gel (Invitrogen) and transferred to a nitrocellulose membrane (Whatman). Protein–RNA complex migrating at an expected molecular weight was excised. Next RNA was isolated from the membrane by proteinase K (Roche) treatment and phenol–chloroform extraction. Purified RNA was ligated to 3′ and 5′ adapter, reverse transcribed and PCR amplified. The amplified cDNA was sequenced on a HiSeq2000 (Illumina).

In vitro transcription, capping, and 2′-O methylation of reporter RNAs

Annotated 5′ UTRs for selected transcripts were cloned upstream of Renilla Luciferase (RLuc) under the control of a T7 promoter, with 60 adenosine nucleotides downstream of the stop codon to mimic polyadenylation. Untranslated regions were cloned using synthetic DNA (Integrated DNA Technologies) or by isolation using 5′ RACE (RLM-RACE kit, Invitrogen). Template was PCR amplified using Phusion polymerase from the plasmids using the following primers, and gel purified, as described (Floor and Doudna 2016).

pA60 txn rev: TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT CTG CAG
pA60 txn fwd: CGG CCA GTG AAT TCG AGC TCT AAT ACG ACT CAC TAT AGG

100 uL in vitro transcription reactions were set up at room temperature with 1-5 micrograms of purified template, 7.5mM ACGU ribonucleotides, 30mM Tris-Cl pH 8.1, 125mM $MgCl_2$, 0.01% Triton X-100, 2mM spermidine, 110mM DTT, T7 polymerase and 0.2 U/uL units of Superase-In RNase inhibitor (Thermo-Fisher Scientific). Transcription reactions were incubated in a PCR block at 37 degrees C for 1 hour. 1 uL of 1 mg/mL pyrophosphatase (Roche) was added to each reaction, and the reactions were subsequently incubated in a PCR block at 37 degrees C for 3 hours. 1 unit of RQ1 RNase-free DNase (Promega) was added to each reaction followed by further incubation for 30 minutes. RNA was precipitated by the addition of 200 uL 0.3M NaOAc pH 5.3, 15 ug GlycoBlue co-precipitant (Thermo-Fisher Scientific) and 750 uL 100% EtOH. Precipitated RNA was further purified over the RNA Clean & Concentrator-25 columns (Zymo Research). Glyoxal gel was run to assess the integrity of the RNA before subsequent capping and 2′ O-methylation.

20 ug of total RNA was used in a 40 uL capping reaction with 0.5mM GTP, 0.2mM s-adenosylmethionine (SAM), 20 units of Vaccinia capping enzyme (New England Biolabs), 100 units of 2′-O-Me-transferase (New England Biolabs) and 25 units RNasin Plus RNase inhibitor (Promega). The reactions were incubated at 37 degrees C for 1 hour, followed by purification over the RNA Clean & Concentrator-25 columns (Zymo Research) and elution in DEPC H_2O. Glyoxal gel was run to assess the integrity of the RNA before proceeding to in vitro translation reactions.

Transfection of siRNA for in vitro translation

HEK293T cells in 150mM plates were transfected with 20 uL of siRNA (against DDX3 or a non-targeting control) using Lipofectamine 2000 (Thermo Fisher Scientific), following manufacturer's instructions. Cells were harvested for preparation of cellular extracts after 48 hours.

Generation of DDX3 mutant cell lines

DDX3 WT and S228C mutant constructs were synthesized and cloned downstream of a CMV promoter (Twist Biosciences). 40 ug of plasmids were transfected into HEK293T cells using Lipofectamine 2000 (Thermo Fisher Scientific), following manufacturer's instructions. Cells were harvested for preparation of cellular extracts after 48 hours.

Preparation of cellular extracts for in vitro translation

Three to five 150mm plates of HEK293T cells were trypsinized and pelleted at 1000g, 4 degrees C. One cell-pellet volume of lysis buffer (10mM HEPES, pH 7.5, 10mM KOAc, 0.5mM $MgOAc_2$, 5mM DTT, and 1 tablet miniComplete EDTA free protease inhibitor (Roche) per 10 mL) was added to the cell pellet and was incubated on ice for 45 minutes. The pellet was homogenized by trituration through a 26G needle attached to a 1 mL syringe 13-15 times. Efficiency of disruption was checked by trypan blue staining (>95% disruption target). The lysate was cleared by centrifugation at 14000g for 1 minute at 4 degrees C, 2-5 ul was reserved for western blot analysis, and the remainder was aliquoted and flash frozen in liquid nitrogen.

Antibodies

Primary antibodies used in this study include anti-DDX3 (Bethyl A300-474A; Figure 1), rabbit polyclonal anti-DDX3 (custom made by Genemed Synthesis using peptide ENALGLDQQFAGLDLNSSDNQS; Figure 6; Lee et al. 2008), anti-actin HRP (Santa Cruz Biotechnology, sc-47778), anti-FLAG HRP (Sigma, A8592).

In vitro translation

5 uL in vitro translation reactions were set up with 2.5 uL of lysate and 20 ng total RNA (0.84mM ATP, 0.21mM GTP, 21mM Creatine Phosphate, 0.009units/mL Creatine phosphokinase, 10mM HEPES pH 7.5, 2mM DTT, 2mM MgOAc, 100mM KOAc, 0.008mM amino acids, 0.25mM spermidine, 5 units RNasin Plus RNase inhibitor (Promega) as described (Lee et al. 2015). Reaction tubes were incubated at

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

30 degrees C for 45 minutes, and expression of the reporter was measured using the Renilla Luciferase Assay System (Promega) on a GloMax Explorer plate reader (Promega). Drug treatments were performed by incubating the reaction with the indicated compound for 15 minutes at 30 degrees C prior to the addition of RNA.

Results

We performed ribosome profiling to determine the set of transcripts that are regulated by DDX3. *DDX3X* is an essential gene (Chen et al. 2016; Lee et al. 2008), so we transiently knocked down its expression using siRNA and collected ribosome protected footprints (Figure 1A,B). Measuring changes in both RNA abundance and ribosome occupancy enabled us to distinguish between different modes of DDX3-mediated regulation. High correlations between replicate experiments indicate reproducibility of the results (Figure S1A). Many genes with increased RNA expression also showed mild reduction in their translation level, small enough to not meet our threshold criteria. We used a Gaussian mixture model to deconvolve these groups (Methods; Figure S1B). Unlike its yeast ortholog Ded1p, we find that depletion of DDX3 only affects ribosome occupancy of two percent of messages (Figure 1C). Almost 90% of ribosome occupancy changes upon DDX3 depletion are decreases, consistent with a role of DDX3 in promoting translation of select mRNAs (Figure 1C,D). We note that transcripts whose RNA level changes after DDX3 depletion tend to have high basal expression level and small fold changes (Figure S1C). Interestingly, histone mRNAs modestly increase in translation upon DDX3 depletion (Figure 1D). Surveying the list of genes affected by DDX3 depletion revealed canonical genes that are subject to translational control, such as the oncogene *ODC1* (Figure 1E, Table S1). Diverse biological pathways are affected by DDX3 depletion, revealing the enrichments of histone in translationally upregulated mRNAs ("TE_up"), and genes related to neuronal branching in the translationally downregulated set ("TE_down") (Figure S1C).

DDX3 is thought to regulate translation through transcript 5′ UTRs, and we found genes that are regulated by their 5′ UTRs such as *ODC1* in the translationally downregulated set (Figure 1E; Auvinen et al. 1992; Steeg et al. 1991). We therefore measured the enrichment of ribosomes in transcript 5′ UTRs, under the hypothesis that depletion of DDX3 might lead to defective scanning and ribosome accumulation (Guenther et al. 2018), or selective ribosome depletion on coding sequences. Indeed, we find more ribosomes in transcript 5′ UTRs relative to coding sequences upon DDX3 depletion (Figure 2A), especially in mRNAs that show translational downregulation. An example gene with increased 5′ UTR ribosome occupancy is shown in Figures 2B and S2A (*HMBS)*, which is regulated by upstream ORF (uORF) translation. We considered features within the 5′ UTR that could potentially contribute to a transcript's sensitivity to reduced DDX3 levels, and observed that the length of 5′ UTRs that are sensitive or insensitive to DDX3 depletion are similar (Figure 2C). However, DDX3-dependent 5′ UTRs have higher GC content (Figure 2C), consistent with the relationship between 5′ UTR structure and ribosome scanning (Kozak 1986). Other regions of these DDX3-dependent mRNAs also show a high-GC signature (Figure S2B). We examined predicted RNA secondary structure in 5′ UTRs and found that mRNAs that decrease in translation upon DDX3 depletion have regions with lower predicted RNA folding free energy than transcripts that are not dependent on DDX3 (Figure 2D). The position of a structure in RNA is thought to influence the requirement for DDX3 (Soto-Rifo et al. 2012), so we examined predicted structure across 5′ UTRs. Interestingly, DDX3-sensitive transcripts have higher predicted structure through the entire 5′ UTR and not just at the 5′ end (Figure S2C).

To better define the set of DDX3-sensitive transcripts, we measured DDX3 binding sites with high specificity using PAR-CLIP. Previous work using a complementary method (iCLIP) to measure DDX3 binding sites identified 5′ UTR and ribosomal RNA binding. Curiously, even though DDX3 is thought to regulate translation initiation, binding was also identified in coding sequences and 3′ UTRs (Oh et al. 2016; Valentin-Vega et al. 2016). Here, we leveraged the additional specificity afforded by T>C transitions induced by protein adducts on crosslinked uridine residues in PAR-CLIP to reexamine these findings. We treated cells with 4-thiouridine (4sU) and used UV-A crosslinking at 365 nanometers to crosslink RNA and protein complexes (Hafner et al. 2010).

High-throughput sequencing of RNA fragments crosslinked to DDX3 identified a binding site for DDX3 on the 18S ribosomal RNA (Figure 3; Eliseev et al., 2018) with high replicability (Figure S3A). Interestingly, while there were many rRNA fragments sequenced following PAR-CLIP (Figure S3A), there were only two sites with high-confidence T>C transitions (Figure 3B). The major interaction site maps to helix 16 of the 18S rRNA, similar to where Ded1p crosslinks to rRNA in yeast (Guenther et al. 2018). Helix 16 is on the small ribosomal subunit facing incoming mRNA, which might provide DDX3 access to resolve mRNA secondary structures to facilitate inspection by the scanning 43S complex (Figure 3C). Interestingly, the crosslink site on h16 is just opposite the eIF4B binding site, another factor crucial in ribosome recruitment and scanning (Sen et al. 2016; Walker et al. 2013). This is consistent with observations that eIF4B and Ded1 cooperate in translation initiation on mRNAs (Sen et al. 2016), and consistent with a previously reported interaction between DDX3 and eIF3 (Lee et al. 2008).

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In addition to ribosomal RNA binding, we found that DDX3 binds primarily to coding transcripts (Figure 4A). We find high confidence peaks for DDX3 binding in 67 percent of expressed mRNAs (Figure 4B, S4A, Table S2), considerably more than are translationally regulated (Figure 1C). To identify where DDX3 binds mRNAs, we averaged reads across all expressed transcripts in a metagene analysis. We found that DDX3 primarily binds to transcript 5′ UTRs, with a small number of reads mapping in the coding sequence and 3′ UTR (Figure 4C). We used available CLIP data to compare the binding pattern of DDX3 to other known mRNA binders (Zhu et al. 2019). We selected three RNA-binding proteins to compare to: eIF3b is a member of the multi-subunit initiation factor eIF3 (Aitken et al. 2016), FMR1 interacts with elongating ribosomes (Chen et al. 2014), and MOV10 is involved in 3′ UTR-mediated mRNA decay (Gregersen et al. 2014; Sievers et al. 2012). The binding pattern of DDX3 most closely resembles the initiation factor eIF3b (Figure 4C). However, we detected some DDX3 binding within coding sequences and even 3′ UTRs. We used the frequency of T>C transitions at each site as a measure of the specificity of protein-RNA interaction (Figure S4B; Mukherjee et al. 2018). Interestingly, high specificity crosslinks with frequent T>C transitions reside most often in 5′ UTRs, especially near the start codon (Figure S4C). We also find DDX3 binding to transcript 5′ UTRs whose translation decreases upon DDX3 depletion, such as *PRKRA* and *ODC1* (Figure 4D and S4D). Confirmation of this binding pattern comes from an independent assay of protein-RNA interaction, as measured by eCLIP (Figure S4E; Van Nostrand et al. 2016). Taken together, we conclude that DDX3 primarily functions as a translation initiation regulator by binding to transcript 5′ UTRs.

Since we found that DDX3 binds to transcript 5′ UTRs of messages that decrease in translation upon DDX3 depletion such as *PRKRA* (Figure 4D), we sought to determine if there was a relationship between DDX3 binding and translation changes. DDX3 binds a large majority of cellular transcripts (Figure S4A) but only regulates a small fraction of them (Figure 1), suggesting there are differences between the transcripts that are translationally controlled by DDX3 versus those that are not. We hypothesized that there is a relationship between the structure of a transcript's 5′ UTR and its dependence on DDX3, as has been observed in the yeast ortholog Ded1p (Guenther et al. 2018; Gupta et al. 2018) and for individual mRNAs (Ku et al. 2018; Lai et al. 2008, 2010; Soto-Rifo et al. 2012). We find that there is increased T>C conversion specificity in the 5′ UTR of transcripts whose translation decreases upon DDX3 depletion (Figure 5A). An example of coordinate binding and regulation by DDX3 is the *ODC1* gene (Figure S4D). We analyzed the average GC content in PAR-CLIP peaks across mRNAs stratified by the nature of translation changes upon DDX3 depletion, and find that the set of translationally downregulated genes have elevated GC content in peaks in their 5′ UTRs (Figure 5B). Despite a clear reported preference of DDX3 binding to structured RNAs, we reasoned that such a sequence enrichment might also result from inherent biases of PAR-CLIP in determining binding specificity. We observed a consistent, albeit weak, correlation between peak scores and %GC content for published PAR-CLIP peaks (Figure S5A) for 50 RNA-binding proteins (Zhu et al. 2018), confirming the presence of possible biases inherent to the experimental protocol and/or analysis (Friedersdorf and Keene 2014). However, a similar binding pattern, enriched in 5′ UTRs of downregulated transcripts, was confirmed using a different PAR-CLIP peak calling strategy (Figure S5B,C). Taken together, we conclude that while T>C conversion specificity is potentially confounded by GC content, the set of mRNAs that DDX3 regulates nevertheless have increased binding relative to all expressed transcripts.

Our data suggest that DDX3 directly affects translation of a subset of mRNAs. However, ribosome profiling measures ribosome occupancy and does not distinguish between elongating or stalled ribosomes. Given that DDX3 binds primarily to 5′ UTRs, and that DDX3 binding is enriched in transcripts it regulates (Figure 5, S5), we reasoned the ribosome occupancy changes observed here are likely due to altered translation initiation at the main start codon. To test this, we cloned DDX3-sensitive 5′ UTRs from this work and previous work (Oh et al. 2016) upstream of a Renilla luciferase reporter, which is not sensitive to DDX3 depletion (Figure 6A, Table S3). As the annotated *DVL2* 5′ UTRs are nested, we cloned the prevalent isoforms in HEK 293T cells using 5′ RACE. We then made translation extracts from HEK 293T cells transfected with a nontargeting control siRNA or a DDX3 siRNA (Figure 6B). Next, we in vitro transcribed, capped, and 2′-O methylated the reporter RNAs and performed in vitro translation in wild-type or DDX3-depleted lysate. We find that 5′ UTRs from DDX3-sensitive mRNAs (Figure 1D, 4D) also confer DDX3 dependence to the luciferase reporter (Figure 6C; *PRKRA*, *DVL2*); translation of the Rluc control, containing a generic 5′ UTR sequence, was not affected. Therefore, based on these reporter experiments and DDX3 binding pattern, we interpret ribosome occupancy changes upon DDX3 depletion as a result of mis-regulated translation initiation dynamics.

However, translation of *RPLP1* was unchanged upon DDX3 depletion (Figure 6C). *RPLP1* was identified as an mRNA with uORF occupancy changes upon mutant DDX3 expression in prior work (Oh et al. 2016). This suggests there may be different effects on translation between depletion of DDX3 and expression of a dominant negative mutant. Interestingly, patients carrying inactivating point mutations in DDX3 display more

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

severe clinical symptoms than patients with truncating mutations (Lennox et al. 2018). Point mutations in DDX3 associated with medulloblastoma are dominant negative and act by preventing enzyme closure of DDX3 towards the high-RNA-affinity ATP-bound state (Epling et al. 2015; Floor et al. 2016b; Jones et al. 2012; Kool et al. 2014; Pugh et al. 2012; Robinson et al. 2012). We therefore tested chemical inhibition of DDX3, which functionally mimics a dominant negative mutation by blocking ATP binding. We were unable to inhibit translation in a DDX3-dependent manner using RK-33 (Bol et al. 2015), and instead found that it acted as a general translation inhibitor (Figure S6). We therefore used a recently established chemical genetic approach for inhibiting DEAD-box proteins by introducing a cysteine near the active site (Barkovich et al. 2018; Floor et al. 2016a). We made translation extracts with wild-type DDX3 or DDX3 S228C, which is sensitive to the electrophilic inhibitor AMP-acrylate (Figure 6D,E). Interestingly, we find that chemical inhibition of DDX3 decreases translation of all tested reporters but not the generic 5′ UTR sequence in the Rluc control (Figure 6E). Taken together, we find that DDX3 sensitivity for translation is preserved in translation extracts and that depletion of DDX3 appears to have different outcomes on translation than inhibition or dominant negative variants.

Discussion

DDX3X is an essential human gene that is altered in numerous disease states. Here, we use a set of transcriptomics approaches and biochemistry to show that DDX3 regulates a subset of the human transcriptome, likely through resolving RNA structures. We further show that inactivating point mutations in DDX3 yield different outcomes than depletion. Reporter experiments that show these 5′ UTRs are sufficient to confer DDX3 sensitivity onto unrelated coding sequences. We conclude that DDX3 affects translation initiation through interacting with transcript 5′ UTRs. Our data firmly establish that the major role of DDX3 is in translation initiation, and reveal translation differences between mutated and haploinsufficient DDX3 expression.

We identified binding between DDX3 and helix 16 (h16) on the human 40S ribosome. This is similar to binding sites identified previously (Guenther et al. 2018; Oh et al. 2016; Valentin-Vega et al. 2016), but strengthened here using T>C transitions defined by PAR-CLIP. Interestingly, histone mRNAs increased translation upon DDX3 knockdown (Figure 1D), and their translation is dependent on mRNA binding to the rRNA h16 helix (Martin et al. 2016). Our data suggests the possibility that there is competition for h16 between DDX3 and histone mRNAs, and their translation increases upon DDX3 knockdown due to increased accessibility to h16. However, histone mRNAs contain highly repetitive sequences and lack a poly-A tail, perhaps requiring a more tailored approach to precisely investigate the mechanisms of their regulation. The set of mRNAs that require h16 for their translation will be an interesting direction to pursue in the future.

Despite primarily affecting translation, some genes exhibited mild changes in steady-state RNA abundance upon DDX3 depletion. RNA-level changes could be mediated by indirect effects of a DDX3-dependent translation target or reflect additional mechanisms by which DDX3 regulates gene expression. For instance, the up-regulation of mRNAs encoding secretory proteins (*CHGA*, *CHGB*, *SCG3*) or of highly abundant RNAs (Figure 1D, S1C) might hint to possible roles of DDX3 in ER-mediated RNA processes or as a more general regulator of cytoplasmic RNP complexes. For instance, DDX3 is an important factor in stress-granule complexes (Shih et al. 2012). Alternatively, RNA-level changes could reflect differences in co-translational RNA decay pathways. Potential interactions between DDX3 regulation of both ribosome occupancy and RNA levels will be further explored in future studies.

DDX3 is an abundant protein, with approximately 1.4 million copies per HeLa cell (Kulak et al. 2014), or about half the abundance of ribosomes (Duncan and Hershey 1983). We have interpreted data in this work by hypothesizing that DDX3 is functioning *in cis* by binding to the 40S ribosome and facilitating translation initiation on the associated mRNA. It is also possible that DDX3, alone or in combination with other abundant DEAD-box proteins like eIF4A, functions *in trans* by activating an mRNA prior to 43S complex loading and then binds to the 43S complex. Future work defining the binding site of DDX3 on the ribosome could enable separation of *cis* and *trans* functions to test these two models, although we note that the functional consequences of DDX3 depletion we have observed here are independent of its functioning in *cis* or *trans.*

We find 249 mRNAs change in translation level upon DDX3 depletion following joint replicate analysis (Figure 1C), although DDX3 binds to over 9,000 mRNAs (Figure 4B, S4A). Therefore, not all mRNAs bound by DDX3 are strongly affected upon DDX3 depletion. This is markedly different than the observed requirement for Ded1 in translation of most yeast mRNAs (Guenther et al. 2018; Sen et al. 2015; Chuang et al. 1997; de la Cruz et al. 1997). mRNAs that require DDX3 for translation have a molecular signature involving increased 5′ UTR secondary structure (Figure 5). Genes that produce mRNAs that depend on DDX3 for translation include potent regulators of cellular state, such as disheveled 1 and 2 (*DVL1* and *DVL2*), many transcription factors (e.g. *E2F1, E2F4, TCF3, HES6,* and *ELK1*), and transforming growth factor beta 1 (*TGFB1;* Table S1). However, the set of regulated mRNAs will change between cell types and states depending on the set of expressed transcripts that contain this molecular signature. Thus, it becomes

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

challenging to establish directness for phenotypic changes that result from DDX3 knockdown, since DDX3 could be directly mediating the phenotype or indirectly mediating the phenotype through translational deregulation of a phenotypic effector.

DDX3 is altered in numerous human diseases, including cancers and developmental disorders (Sharma and Jankowsky 2014). Some diseases are characterized by missense variants (Jones et al. 2012; Kool et al. 2014; Pugh et al. 2012; Robinson et al. 2012), while others involve predominantly nonsense or frameshift variants (Dufva et al. 2018; Jiang et al. 2015), and still others present with a mixture of variant types (Deciphering Developmental Disorders 2015; Lennox et al. 2018; Wang et al. 2018). Our work here suggests that haploinsufficient variants in DDX3 that deplete protein levels may result in different translation changes than missense variants that yield dominant negative proteins. We attempted to directly compare translational changes upon DDX3 depletion identified in this work with previous expression of mutant DDX3 but stopped due to confounding variability in biological sample and library preparation and sequencing protocols. An intriguing direction for the future is to determine in an isogenic background how different mutation types in DDX3 affect gene expression, the underlying molecular mechanisms, and potential therapeutic interventions.

Data access

Sequencing data can be retrieved using GEO accession number GSE125114. Source code for all analysis programs can be found here: https://github.com/lcalviell/DDX3_manuscript_analysis.

Acknowledgments

We thank members of the Floor lab for feedback on the manuscript and Megan Moore and Kevan Shokat for advice and the kind gift of AMP-acrylate. Computation was supported by the UCSF Wynton computing infrastructure. This work was supported by the UCSF Program for Breakthrough Biomedical Research, funded in part by the Sandler Foundation (to SNF), the California Tobacco-Related Disease Research Grants Program 27KT-0003 (to SNF), and the National Institutes of Health DP2GM132932 (to SNF), and DFG Priority Project SPP1935 (to KR and ML).

Author contributions

Conceptualization, L.C., E.W., W.F., M.L., and S.N.F.; Investigation, L.C., K.R., S.V., and E.W.; Writing – Original Draft, S.N.F.; Writing – Review & Editing, S.N.F., S.V., L.C., E.W., and M.L.; Resources, B.T., M.T., J.K., and W.F.; Funding Acquisition, M.L., and S.N.F.; Supervision, W.F., M.L., and S.N.F.

Disclosure declaration

The authors declare no conflicts of interest.

References

Aitken CE, Beznosková P, Vlčkova V, Chiu W-L, Zhou F, Valášek LS, Hinnebusch AG, Lorsch JR. 2016. Eukaryotic translation initiation factor 3 plays distinct roles at the mRNA entry and exit channels of the ribosomal preinitiation complex. *eLife* **5**.

Auvinen M, Paasinen A, Andersson LC, Hölttä E. 1992. Ornithine decarboxylase activity is critical for cell transformation. *Nature* **360**: 355–358.

Barkovich KJ, Moore MK, Hu Q, Shokat KM. 2018. Chemical genetic inhibition of DEAD-box proteins using covalent complementarity. *Nucleic Acids Res* **46**: 8689–8699.

Bol GM, Vesuna F, Xie M, Zeng J, Aziz K, Gandhi N, Levine A, Irving A, Korz D, Tantravedi S, et al. 2015. Targeting DDX3 with a small molecule inhibitor for lung cancer therapy. *EMBO Mol Med* **7**: 648–69.

Chen CY, Chan CH, Chen CM, Tsai YS, Tsai TY, Wu Lee YH, You LR. 2016. Targeted inactivation of murine Ddx3x: essential roles of Ddx3x in placentation and embryogenesis. *Hum Mol Genet*.

Chen E, Sharma MR, Shi X, Agrawal RK, Joseph S. 2014. Fragile X Mental Retardation Protein Regulates Translation by Binding Directly to the Ribosome. *Mol Cell* **54**: 407–417.

Chuang RY, Weaver PL, Liu Z, Chang TH. 1997. Requirement of the DEAD-Box protein ded1p for messenger RNA translation. *Science* **275**: 1468–71.

Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U. 2011. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol* **12**: R79.

de la Cruz J, Iost I, Kressler D, Linder P. 1997. The p20 and Ded1 proteins have antagonistic roles in eIF4E-dependent translation in Saccharomyces cerevisiae. *Proc Natl Acad Sci U A* **94**: 5201–6.

Deciphering Developmental Disorders S. 2015. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**: 223–8.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Dufva O, Kankainen M, Kelkka T, Sekiguchi N, Awad SA, Eldfors S, Yadav B, Kuusanmäki H, Malani D, Andersson EI, et al. 2018. Aggressive natural killer-cell leukemia mutational landscape and drug profiling highlight JAK-STAT signaling as therapeutic target. *Nat Commun* **9**: 1567.

Duncan R, Hershey JW. 1983. Identification and quantitation of levels of protein synthesis initiation factors in crude HeLa

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cell lysates by two-dimensional polyacrylamide gel electrophoresis. *J Biol Chem* **258**: 7228–7235.

Eliseev B, Yeramala L, Leitner A, Karuppasamy M, Raimondeau E, Huard K, Alkalaeva E, Aebersold R, Schaffitzel C. 2018. Structure of a human cap-dependent 48S translation pre-initiation complex. *Nucleic Acids Res* **46**: 2678–2689.

Epling LB, Grace CR, Lowe BR, Partridge JF, Enemark EJ. 2015. Cancer-Associated Mutants of RNA Helicase DDX3X Are Defective in RNA-Stimulated ATP Hydrolysis. *J Mol Biol* **427**: 1779–1796.

Floor SN, Barkovich KJ, Condon KJ, Shokat KM, Doudna JA. 2016a. Analog sensitive chemical inhibition of the DEAD-box protein DDX3. *Protein Sci* **25**: 638–49.

Floor SN, Condon KJ, Sharma D, Jankowsky E, Doudna JA. 2016b. Autoinhibitory Interdomain Interactions and Subfamily-specific Extensions Redefine the Catalytic Core of the Human DEAD-box Protein DDX3. *J Biol Chem* **291**: 2412–21.

Floor SN, Doudna JA. 2016. Tunable protein synthesis by transcript isoforms in human cells. *Elife* **5**.

Friedersdorf MB, Keene JD. 2014. Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. *Genome Biol* **15**: R2.

Gregersen LH, Schueler M, Munschauer M, Mastrobuoni G, Chen W, Kempa S, Dieterich C, Landthaler M. 2014. MOV10 Is a 5′ to 3′ RNA Helicase Contributing to UPF1 mRNA Target Degradation by Translocation along 3′ UTRs. *Mol Cell* **54**: 573–585.

Guenther U-P, Weinberg DE, Zubradt MM, Tedeschi FA, Stawicki BN, Zagore LL, Brar GA, Licatalosi DD, Bartel DP, Weissman JS, et al. 2018. The helicase Ded1p controls use of near-cognate translation initiation codons in 5′ UTRs. *Nature* **559**: 130.

Gupta N, Lorsch JR, Hinnebusch AG. 2018. Yeast Ded1 promotes 48S translation pre-initiation complex assembly in an mRNA-specific and eIF4F-dependent manner eds. N. Sonenberg and J.L. Manley. *eLife* **7**: e38892.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp A-C, Munschauer M, et al. 2010. Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* **141**: 129–141.

Jiang L, Gu ZH, Yan ZX, Zhao X, Xie YY, Zhang ZG, Pan CM, Hu Y, Cai CP, Dong Y, et al. 2015. Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. *Nat Genet* **47**: 1061–6.

Jones DT, Jager N, Kool M, Zichner T, Hutter B, Sultan M, Cho YJ, Pugh TJ, Hovestadt V, Stutz AM, et al. 2012. Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**: 100–5.

Kool M, Jones DT, Jager N, Northcott PA, Pugh TJ, Hovestadt V, Piro RM, Esparza LA, Markant SL, Remke M, et al. 2014. Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. *Cancer Cell* **25**: 393–405.

Kozak M. 1986. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl Acad Sci* **83**: 2850–2854.

Ku Y-C, Lai M-H, Lo C-C, Cheng Y-C, Qiu J-T, Tarn W-Y, Lai M-C. 2018. DDX3 Participates in Translational Control of Inflammation Induced by Infections and Injuries. *Mol Cell Biol* MCB.00285-18.

Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. 2014. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **11**: 319–24.

Lai M-C, Chang W-C, Shieh S-Y, Tarn W-Y. 2010. DDX3 Regulates Cell Growth through Translational Control of Cyclin E1. *Mol Cell Biol* **30**: 5444–5453.

Lai M-C, Lee Y-HW, Tarn W-Y, Walter P. 2008. The DEAD-Box RNA Helicase DDX3 Associates with Export Messenger Ribonucleoproteins as well asTip-associated Protein and Participates in Translational Control. *Mol Biol Cell* **19**: 3847–3858.

Lee ASY, Kranzusch PJ, Cate JHD. 2015. eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* **522**: 111–114.

Lee CS, Dias AP, Jedrychowski M, Patel AH, Hsu JL, Reed R. 2008. Human DDX3 functions in translation and interacts with the translation initiation factor eIF3. *Nucleic Acids Res* **36**: 4708–18.

Lennox AL, Jiang R, Suit L, Fregeau B, Sheehan CJ, Aldinger KA, Moey C, Lobach I, Mirzaa G, Afenjar A, et al. 2018. Pathogenic DDX3X mutations impair RNA metabolism and neurogenesis during fetal cortical development. *bioRxiv*.

Li W, Wang W, Uren PJ, Penalva LOF, Smith AD. 2017. Riborex: fast and flexible identification of differential translation from Ribo-seq data. *Bioinforma Oxf Engl* **33**: 1735–1737.

Lorenz R, Hofacker IL, Stadler PF. 2016. RNA folding with hard and soft constraints. *Algorithms Mol Biol AMB* **11**: 8.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Martin F, Ménétret J-F, Simonetti A, Myasnikov AG, Vicens Q, Prongidi-Fix L, Natchiar SK, Klaholz BP, Eriani G. 2016. Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. *Nat Commun* **7**: 12622.

Mukherjee N, Wessels H-H, Lebedeva S, Sajek M, Ghanbari M, Garzia A, Munteanu A, Yusuf D, Farazi T, Hoell JI, et al. Deciphering human ribonucleoprotein regulatory networks. *Nucleic Acids Res*. https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gky1185/5230955 (Accessed January 16, 2019).

Oh S, Flynn RA, Floor SN, Purzner J, Martin L, Do BT, Schubert S, Vaka D, Morrissy S, Li Y, et al. 2016. Medulloblastoma-associated DDX3 variant selectively alters the translational response to stress. *Oncotarget* **7**: 28169–28182.

Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J, Carneiro MO, Carter SL, Cibulskis K, Erlich RL, et al. 2012. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* **488**: 106–10.

Robinson G, Parker M, Kranenburg TA, Lu C, Chen X, Ding L, Phoenix TN, Hedlund E, Wei L, Zhu X, et al. 2012. Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488**: 43–8.

Schmitz R, Young RM, Ceribelli M, Jhavar S, Xiao W, Zhang M, Wright G, Shaffer AL, Hodson DJ, Buras E, et al. 2012. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* **490**: 116–20.

Sen ND, Zhou F, Harris MS, Ingolia NT, Hinnebusch AG. 2016. eIF4B stimulates translation of long mRNAs with structured 5′ UTRs and low closed-loop potential but weak dependence on eIF4G. *Proc Natl Acad Sci* **113**: 10464–10472.

Sen ND, Zhou F, Ingolia NT, Hinnebusch AG. 2015. Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Res* **25**: 1196–1205.

Sharma D, Jankowsky E. 2014. The Ded1/DDX3 subfamily of DEAD-box RNA helicases. *Crit Rev Biochem Mol Biol* **49**: 343–360.

Shih J-W, Wang W-T, Tsai T-Y, Kuo C-Y, Li H-K, Wu Lee Y-H. 2012. Critical roles of RNA helicase DDX3 and its interactions with eIF4E/PABP1 in stress granule assembly and stress response. *Biochem J* **441**: 119–129.

Sievers C, Schlumpf T, Sawarkar R, Comoglio F, Paro R. 2012. Mixture models and wavelet transforms reveal high confidence RNA-protein interaction sites in MOV10 PAR-CLIP data. *Nucleic Acids Res* **40**: e160–e160.

Snijders Blok L, Madsen E, Juusola J, Gilissen C, Baralle D, Reijnders MR, Venselaar H, Helsmoortel C, Cho MT, Hoischen A, et al. 2015. Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. *Am J Hum Genet* **97**: 343–52.

Soto-Rifo R, Rubilar PS, Limousin T, Breyne S de, Décimo D, Ohlmann T. 2012. DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. *EMBO J* **31**: 3745–3756.

Steeg HV, Oostrom CTMV, Hodemaekers HM, Peters L, Thomas A a. M. 1991. The translation in vitro of rat ornithine decarboxylase mRNA is blocked by its 5′ untranslated region in a polyamine-independent way. *Biochem J* **274**: 521–526.

Valentin-Vega YA, Wang Y-D, Parker M, Patmore DM, Kanagaraj A, Moore J, Rusch M, Finkelstein D, Ellison DW, Gilbertson RJ, et al. 2016. Cancer-associated DDX3X mutations drive stress granule assembly and impair global translation. *Sci Rep* **6**: 25996.

Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K, et al. 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**: 508–514.

Walker SE, Zhou F, Mitchell SF, Larson VS, Valasek L, Hinnebusch AG, Lorsch JR. 2013. Yeast eIF4B binds to the head of the 40S ribosomal subunit and promotes mRNA recruitment through its N-terminal and internal repeat domains. *RNA* **19**: 191–207.

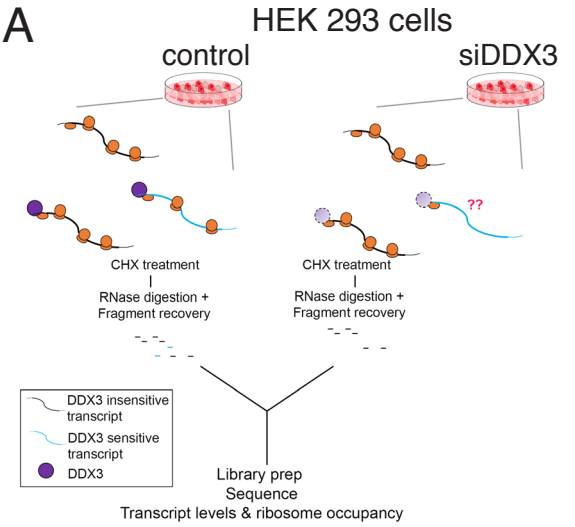
Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, et al. 2011. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med* **365**: 2497–506.

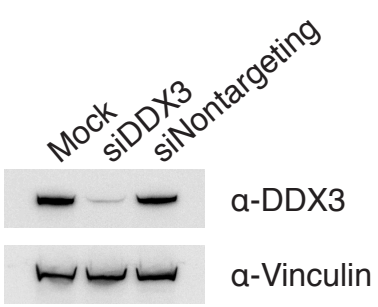
Wang X, Rosenfeld JA, Bacino CA, Scaglia F, Immken L, Harris JM, Hickey SE, Mosher TM, Slavotinek A, Zhang J, et al. 2018. Phenotypic expansion in DDX3X — a common cause of intellectual disability in females. *bioRxiv*.

Zhu Y, Xu G, Yang YT, Xu Z, Chen X, Shi B, Xie D, Lu ZJ, Wang P. POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res*. https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gky830/5098585 (Accessed December 7, 2018).

Zhu Y, Xu G, Yang YT, Xu Z, Chen X, Shi B, Xie D, Lu ZJ, Wang P. 2019. POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res* **47**: D203–D211.

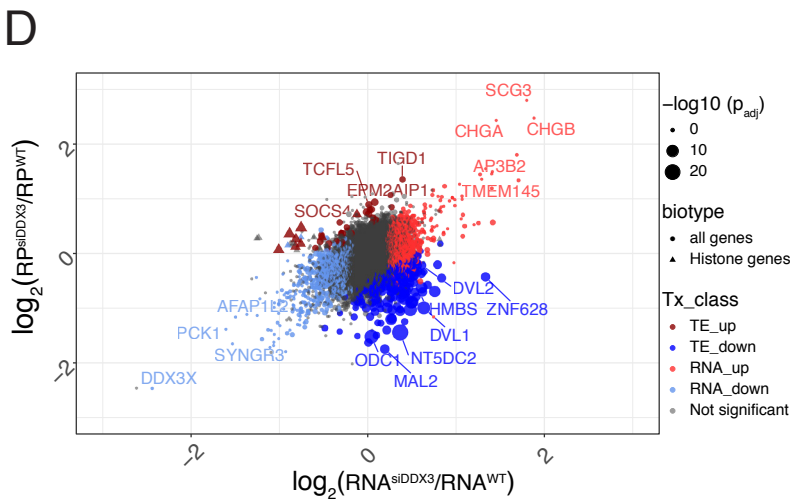
bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.





C

	Not_significant	RNA_down	RNA_up	TE_down	TE_up
number of genes	10365	792	913 (617)	249	34
% of total	83.91	6.41	7.39 (4.99)	2.02	0.28



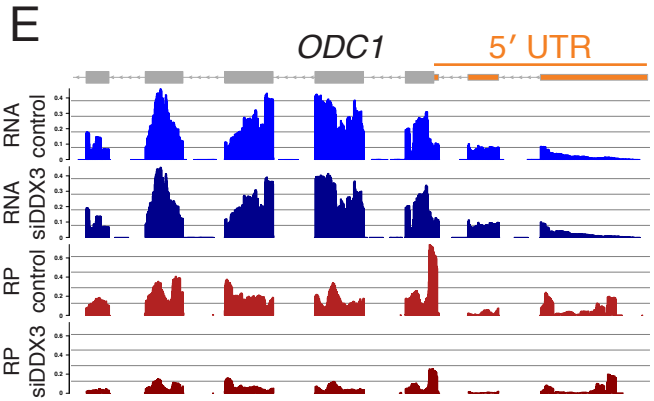


Figure 1: Translational changes upon DDX3 depletion. (A) A workflow of the experimental system. (B) siRNA knockdown efficiency of DDX3 analyzed by western blot. (C) Information on the number of genes that change translation upon DDX3 depletion. TE: translation efficiency. Parentheses: RNA_up genes after mixture model filtering (see Methods). (D) Differential expression analysis of RNA and ribosome profiling changes upon DDX3 depletion. RP: ribosome profiling. Point size indicates differential translation (as defined by Riborex). (E) Diminished translation of ODC1 mRNAs following DDX3 depletion (compare RP_siDDX3 to RP_ctrl).

Calviello, Venkataramanan *et al*. Figure 1

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

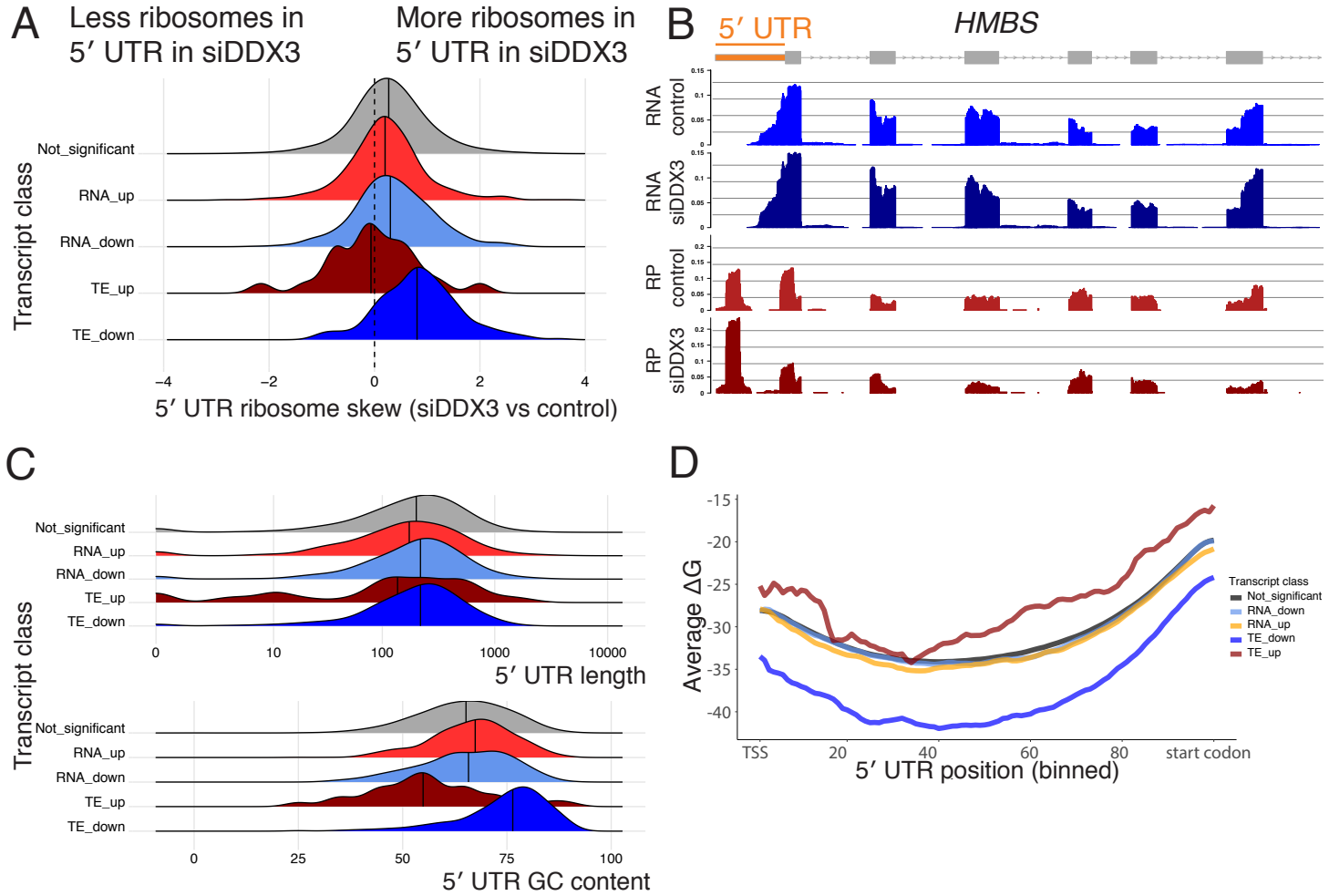


Figure 2: DDX3-sensitive transcripts have complex 5′ leaders. (A) The fold-change of the ratio in ribosome occupancy in the 5′ UTR versus the coding sequence upon DDX3 depletion (Methods). (B) An example gene (*HMBS*) with increases in 5′ UTR ribosomes versus coding sequence ribosomes. (C) DDX3-sensitive mRNAs (TE_down) have similar length 5′ UTRs but higher GC content. (D) DDX3-sensitive mRNAs (TE_down) have more predicted 5′ UTR structure than other mRNAs.

Calviello, Venkataramanan *et al*. Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

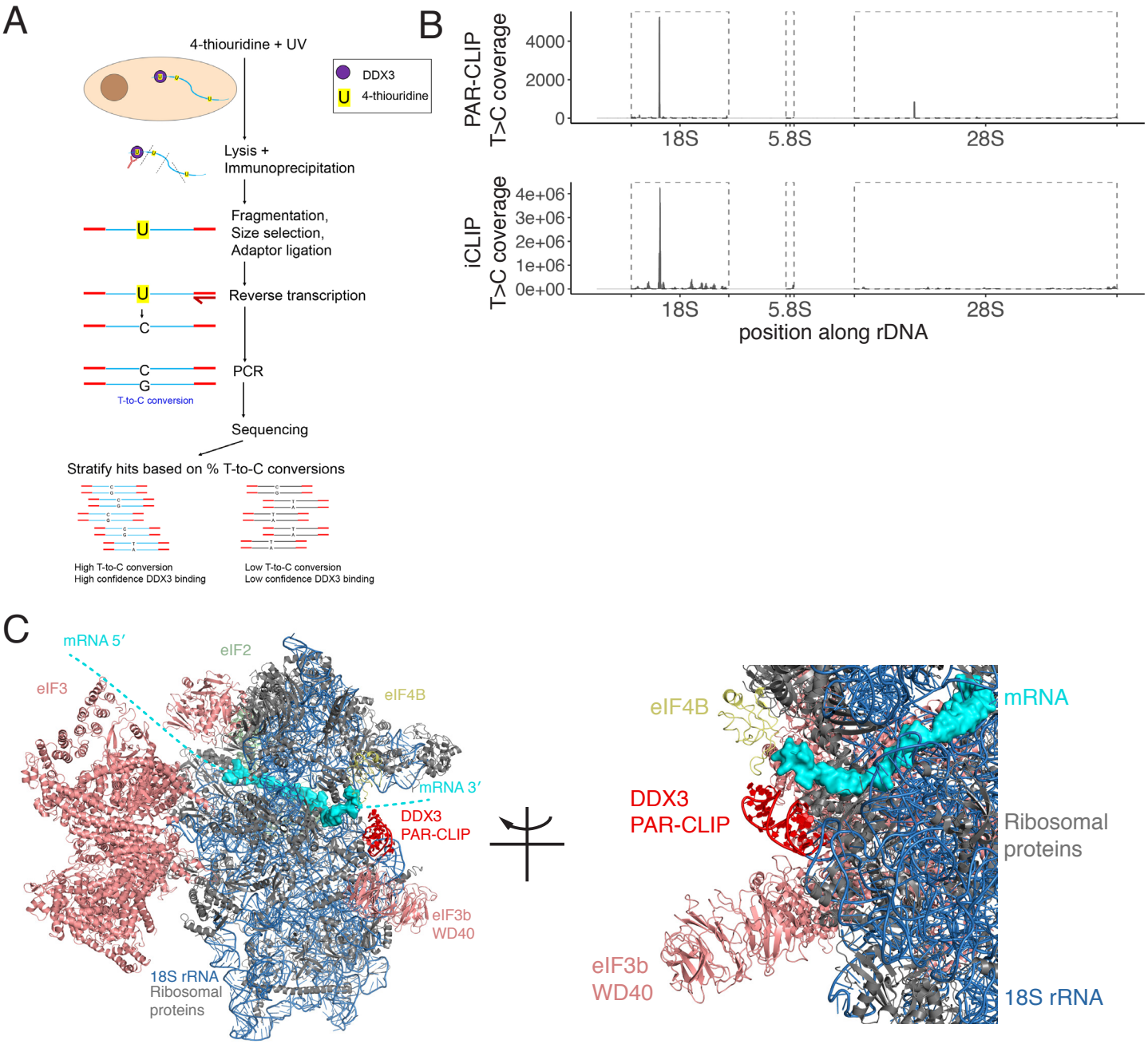
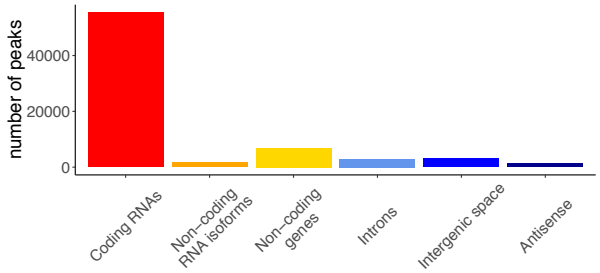


Figure 3: DDX3 binding sites on rRNA identified by PAR-CLIP. (A) a workflow of the PAR-CLIP experiment. (B) PAR-CLIP peaks on human rDNA (top) compared to iCLIP from previous work (bottom). (C) The PAR-CLIP peak on the 18S rRNA is visualized in red on the structure of the 48S ribosome (PDB 6FEC), with a zoom on the right. A portion of the mRNA was observed in the structure shown in cyane surface with additional mRNA modeled as a dashed line.

Calviello, Venkataramanan *et al*. Figure 3

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

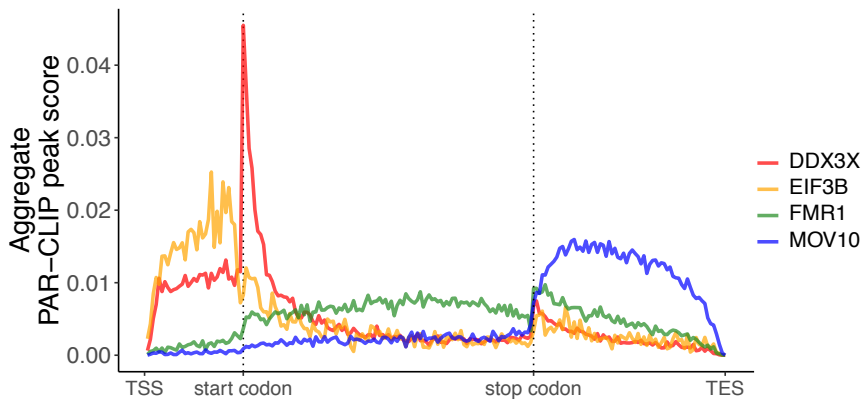
A



B

	all_genes	RNA_up	RNA_down	TE_up	TE_down
Detected genes	12054	617	792	34	249
DDX3–bound (% of Detected genes)	9443 (78.3%)	522 (84.6%)	519 (65.5%)	32 (94.1%)	219 (88%)
DDX3–bound, high T>C (% of DDX3–bound)	6827 (72.3%)	380 (72.8%)	375 (72.3%)	24 (75%)	186 (84.9%)

C



D

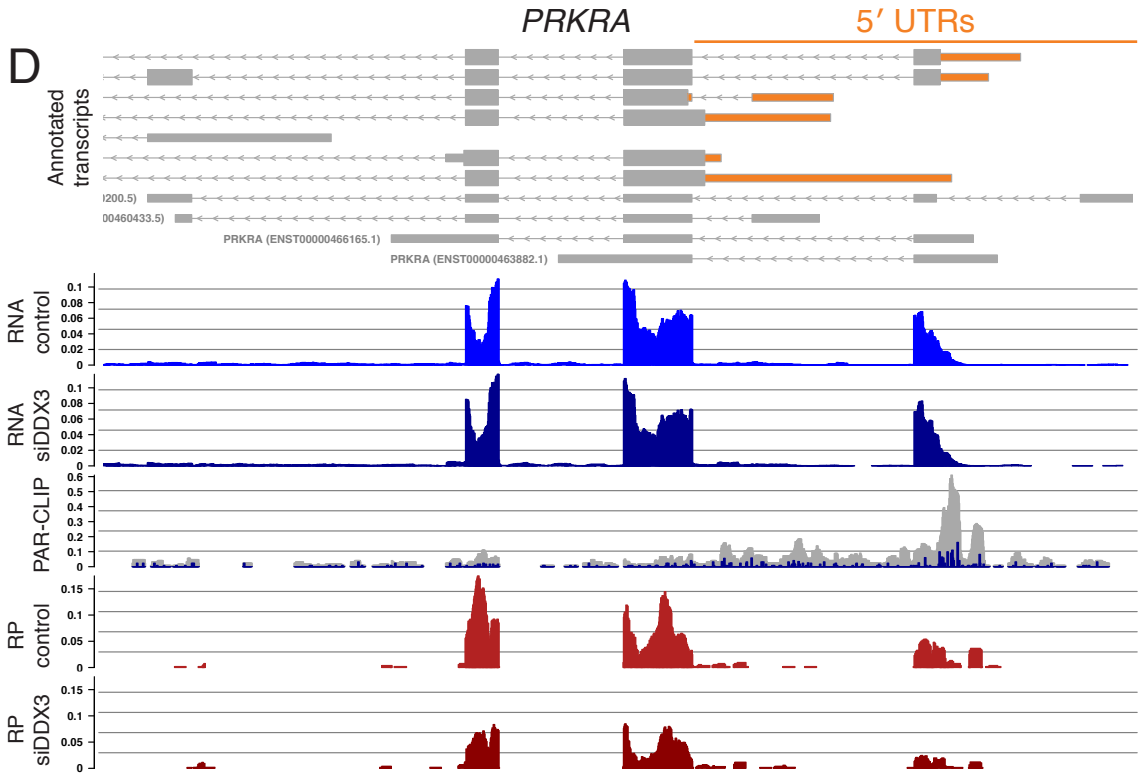


Figure 4: DDX3 binds to mRNA 5' leaders, similar to canonical initiation factors. (A) Most of the non-rRNA bound by DDX3 is coding RNA. (B) DDX3 binds to many coding mRNAs, many more than it translationally regulates. However, high confidence T>C transitions are moderately enriched in the genes that are sensitive to DDX3 knockdown. (C) A metagene plot of DDX3 and other PAR-CLIP data. eIF3B is a canonical initiation factor, FMR1 binds elongating ribosomes, and MOV10 is a 3′ UTR binding factor. (D) An example of a gene that is bound and translationally controlled by DDX3. Note there is another transcript of *PRKRA* with an extended 5′ UTR that spans the first intron above.

Calviello, Venkataramanan *et al*. Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

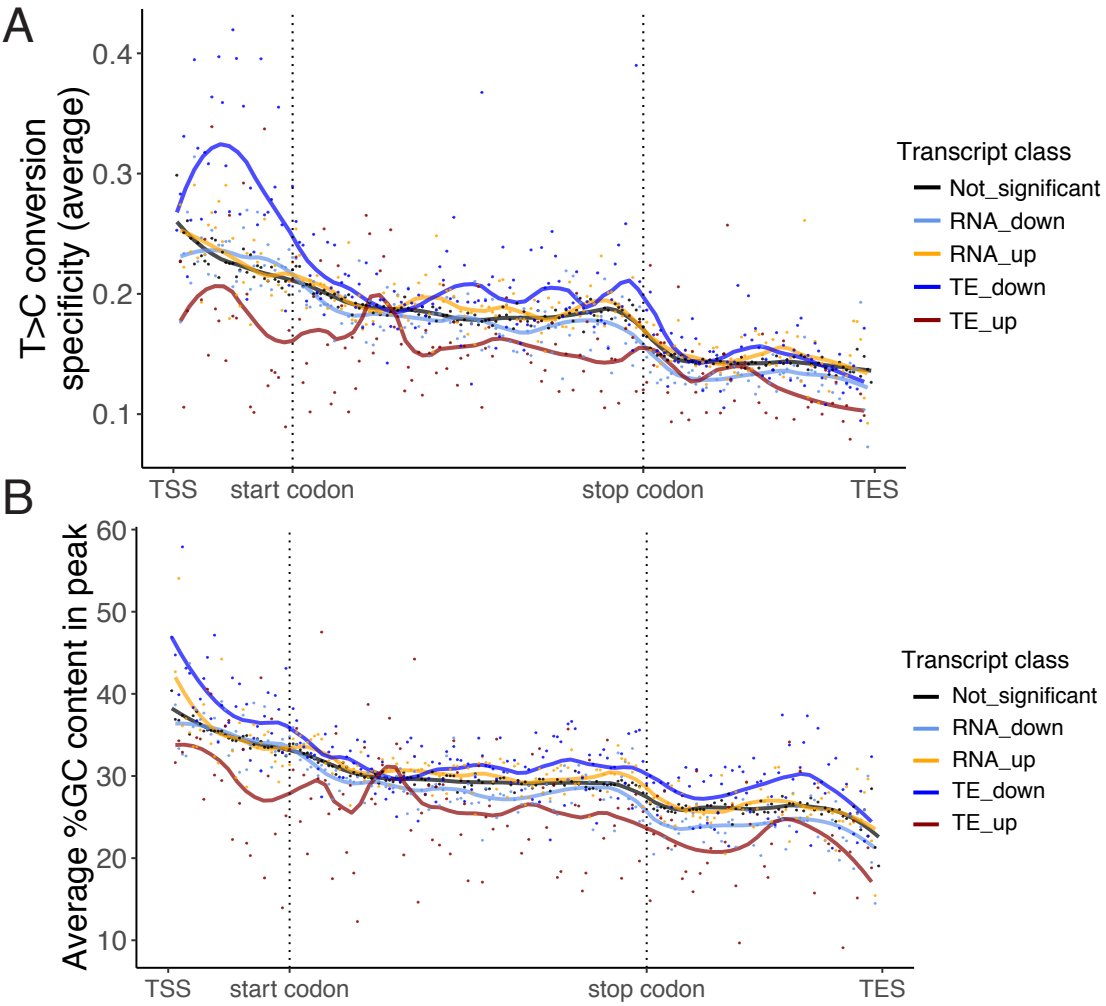
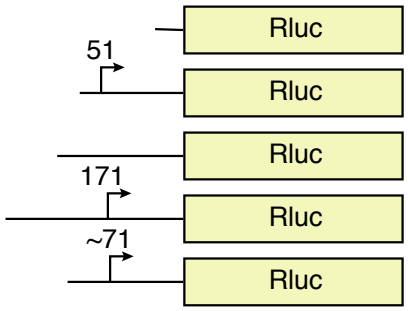


Figure 5: DDX3 binding regulates translation of GC-rich transcripts.
(A, B) Metagene tracks of T>C conversion specificity (A) and GC content within PAR-CLIP peaks (B) for different gene classes across transcript bins.

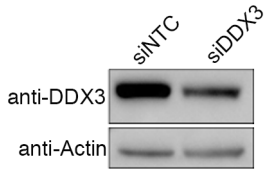
bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A

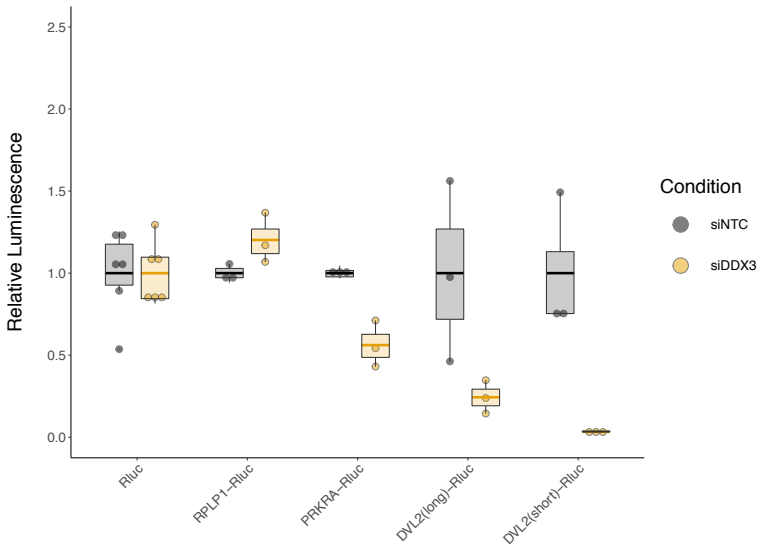
	GC content	length	75nt region ΔG min (kcal mol^{-1})	Upstream start sites (TISdb)
control	66.7	9	0	
RPLP1	66.1	165	-33.7	51
PRKRA	74.6	201	-40.1	
DVL2 (long)	71.6	285	-48.0	171
DVL2 (short)	73.0	185	-40.3	~71



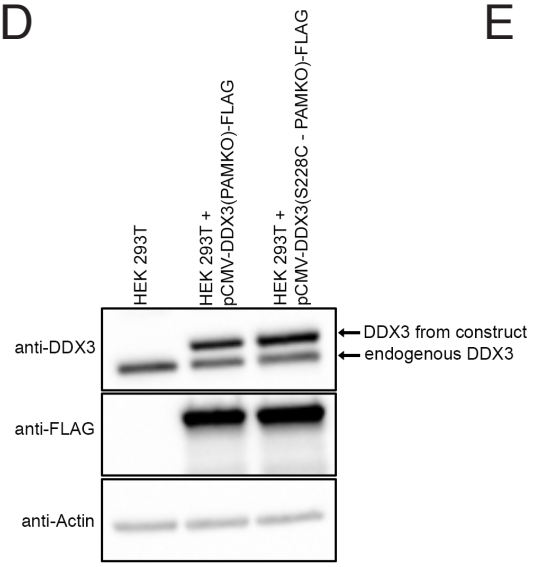
B



C



D



E

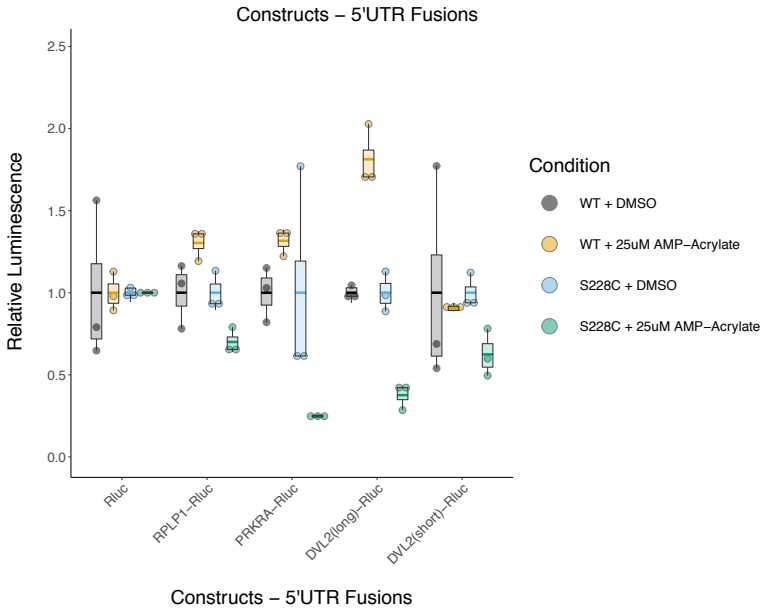
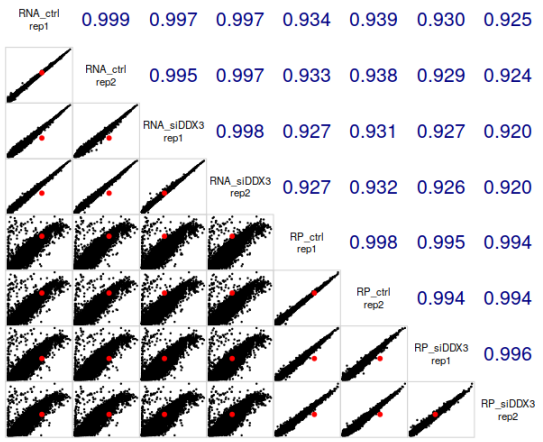


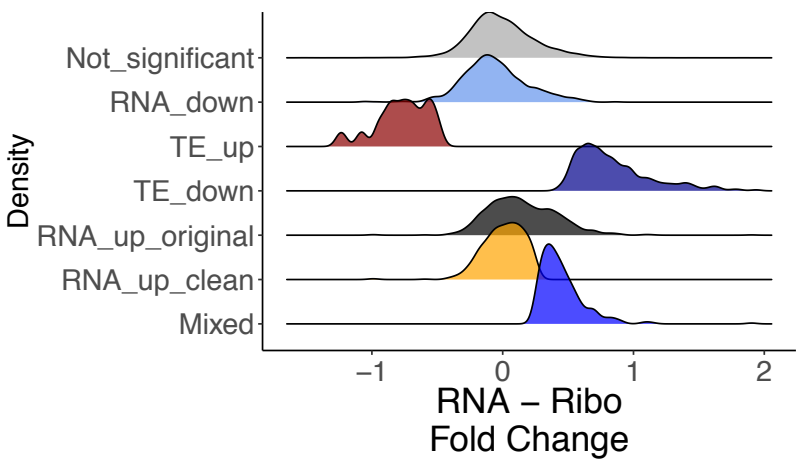
Figure 6: Chemical inhibition of DDX3 affects translation differently than knockdown.
(A) Diagrams of the reporter constructs used to test DDX3-dependent translation. (B) Knockdown of DDX3 using siRNA to make translation lysates. (C) Translation of in vitro transcribed reporter RNAs in siNTC (non targeting control) or siDDX3 HEK 293T lysates. (D) Expression of either WT DDX3 or DDX3 S228C, which is sensitive to electrophile inhibitors. (E) Addition of AMP-acrylate decreases translation of a different set of DDX3-sensitive mRNAs than knockdown.

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A



B



C

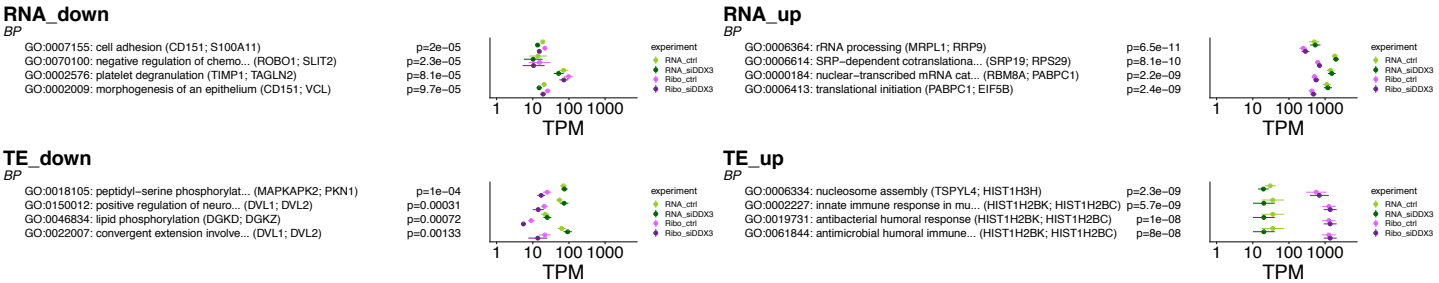


Figure S1: Related to Figure 1 ribosome profiling.
(A) Replicate correlations for RNAseq and ribosome profiling libraries. Red dots: the *DDX3X* gene. (B) Correction of RNA_up genes by mixture modeling (methods). (C) Gene ontology terms enriched in various gene sets.

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

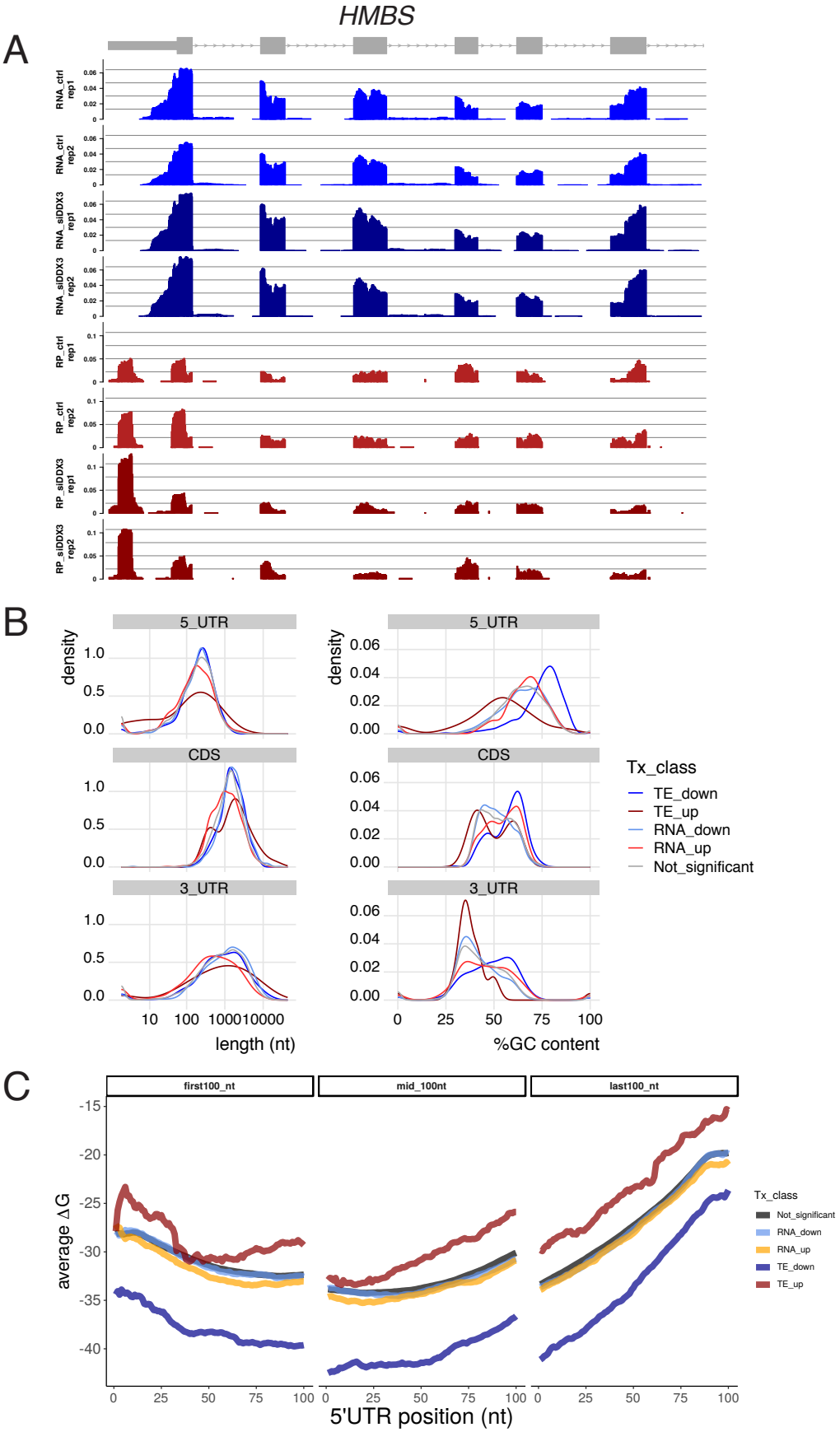


Figure S2: Related to Figure 2 ribosome profiling feature analysis

(A) RNA-seq and Ribo-seq tracks for individual replicates across the *HMBS* gene. (B) mRNA features associated with indicated gene sets. (C) Predicted RNA structure for the first, middle, and last 100 nucleotides of 5′ untranslated regions in indicated gene sets.

Calviello, Venkataramanan *et al*. Figure S2

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

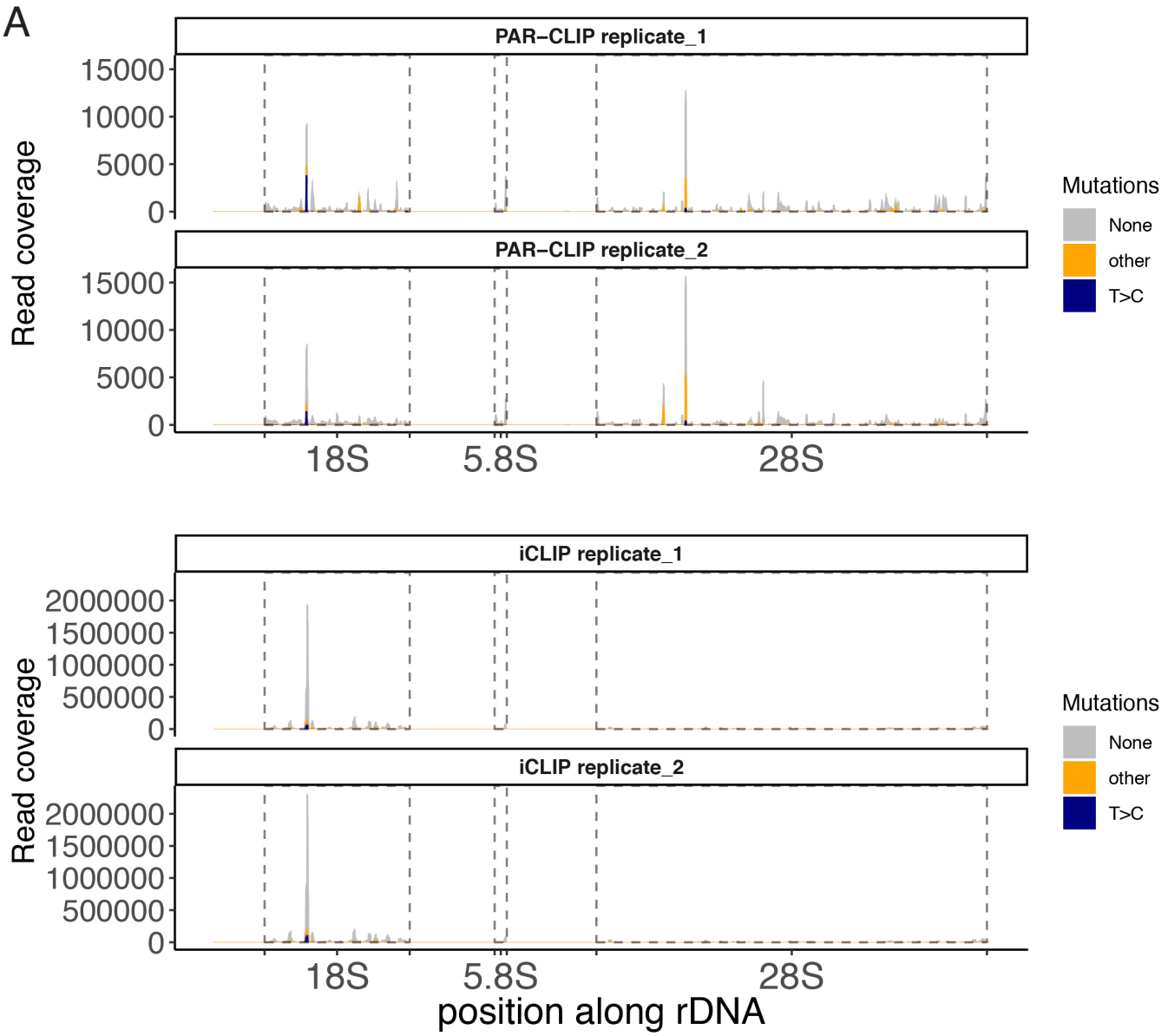


Figure S3: Related to PAR-CLIP.
(A) rDNA tracks for replicate experiments of PAR-CLIP (this work) and iCLIP from Oh et al 2016.

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

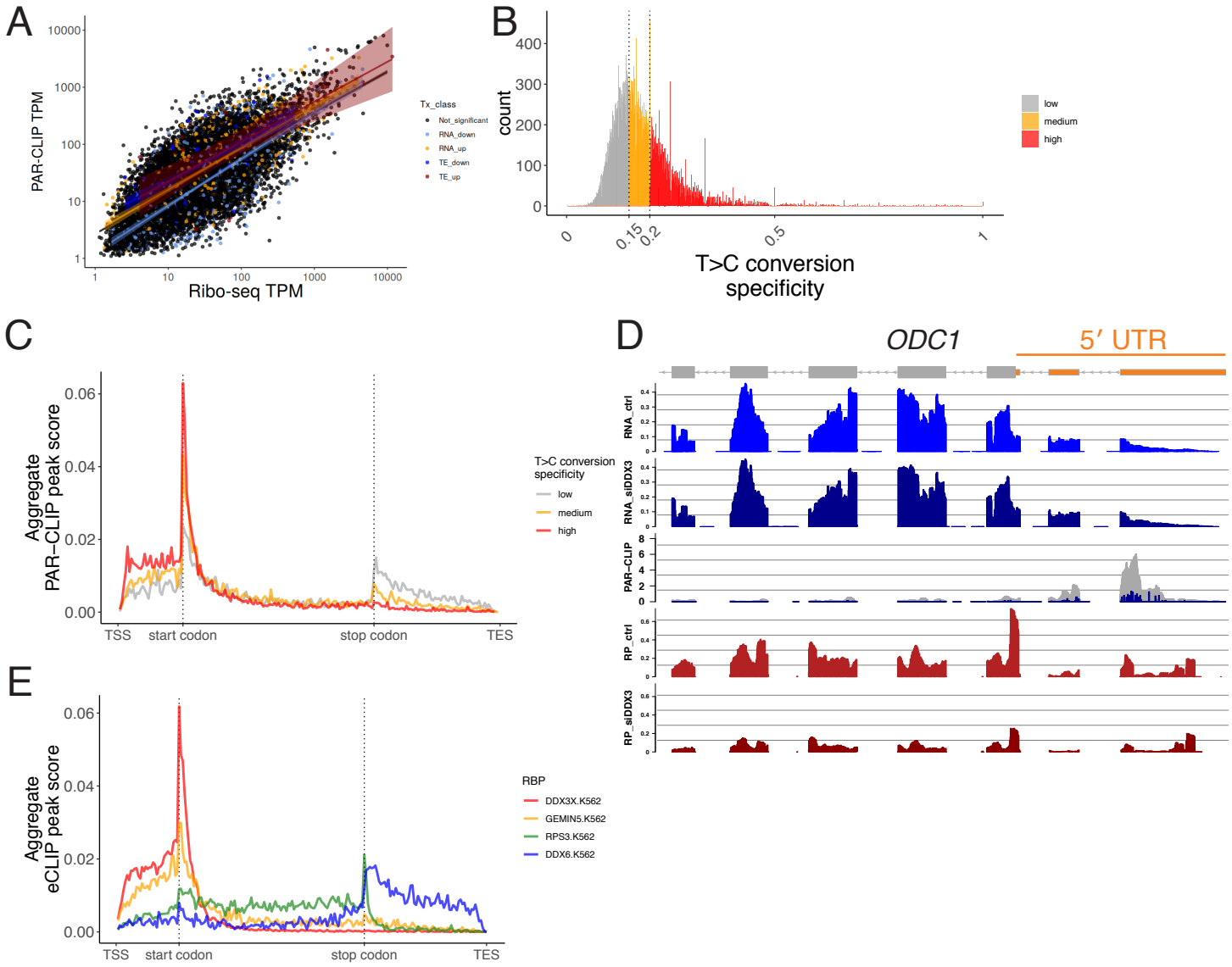


Figure S4: Related to Figure 4 PAR-CLIP - mRNA analysis
(A) Scatter plot of PAR-CLIP TPMs and Ribo-seq TPMs showing that DDX3 binding is proportional to ribosome occupancy for most genes. (B) A histogram of T>C conversion specificity for each PARalyzer PAR-CLIP peak. (C) Aggregate PAR-CLIP peak scores from PARalyzer stratified by T>C conversion specificity thirtiles across transcript regions. (D) Ribo-seq, RNA-seq and PAR-CLIP reads across the *ODC1* gene. (E) Plotting aggregate RBP binding data from eCLIP shows that the DDX3 mRNA binding pattern is similar in an unrelated cell type (K562) and using a complementary approach (eCLIP).

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

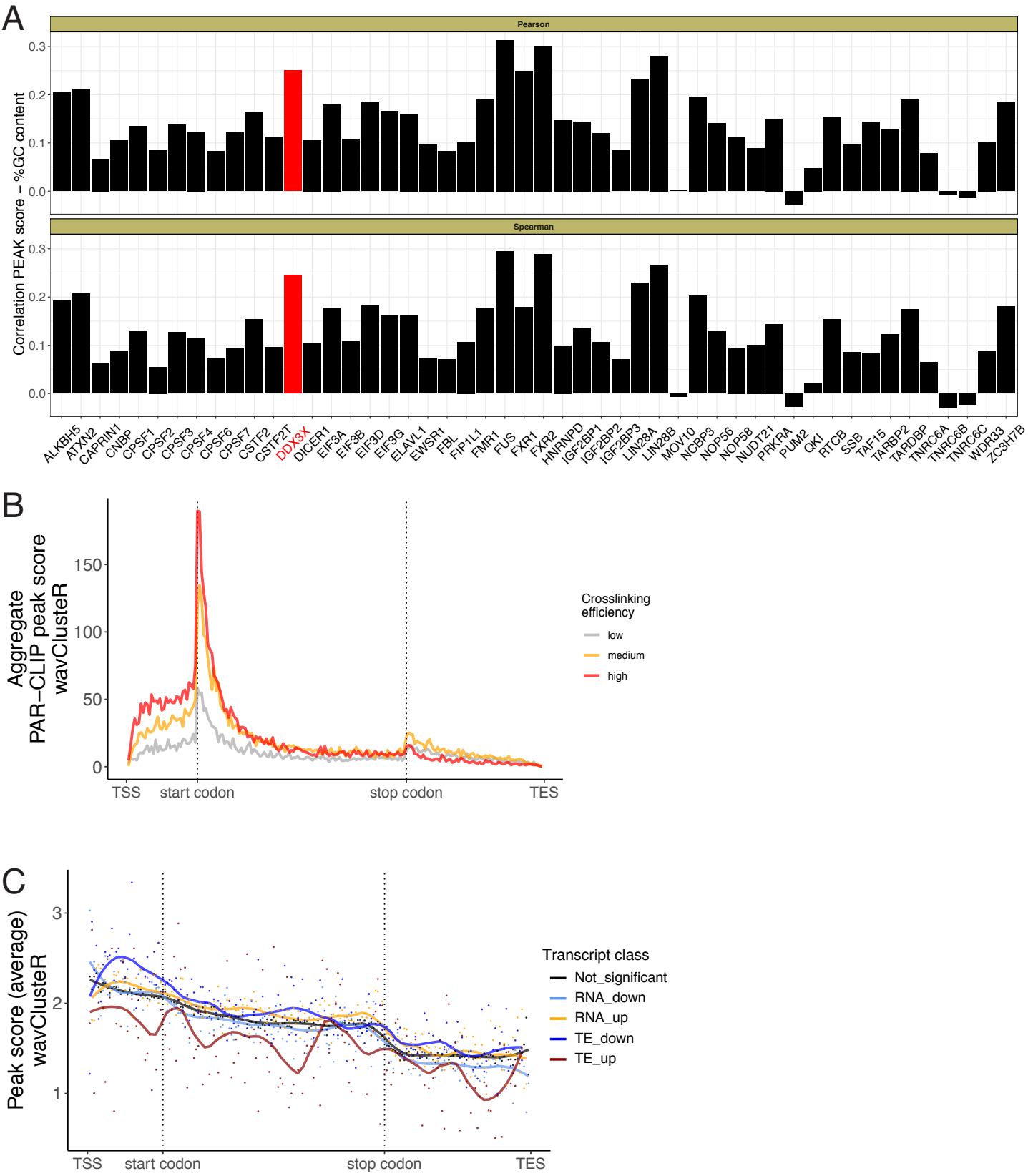


Figure S5: Related to Figure 5

(A) Pearson (top) and spearman (bottom) correlations between peak scores and GC content across 50 published PAR-CLIP datasets with scores from the POSTAR2 database. (B) Aggregate PAR-CLIP peak scores from wavClusteR stratified by crosslinking efficiency thirtiles across transcript regions. (C) Average peak scores from PAR-CLIP called by wavClusteR for indicated transcript groups across transcript regions.

Calviello, Venkataramanan *et al*. Figure S5

bioRxiv preprint doi: https://doi.org/10.1101/589218; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

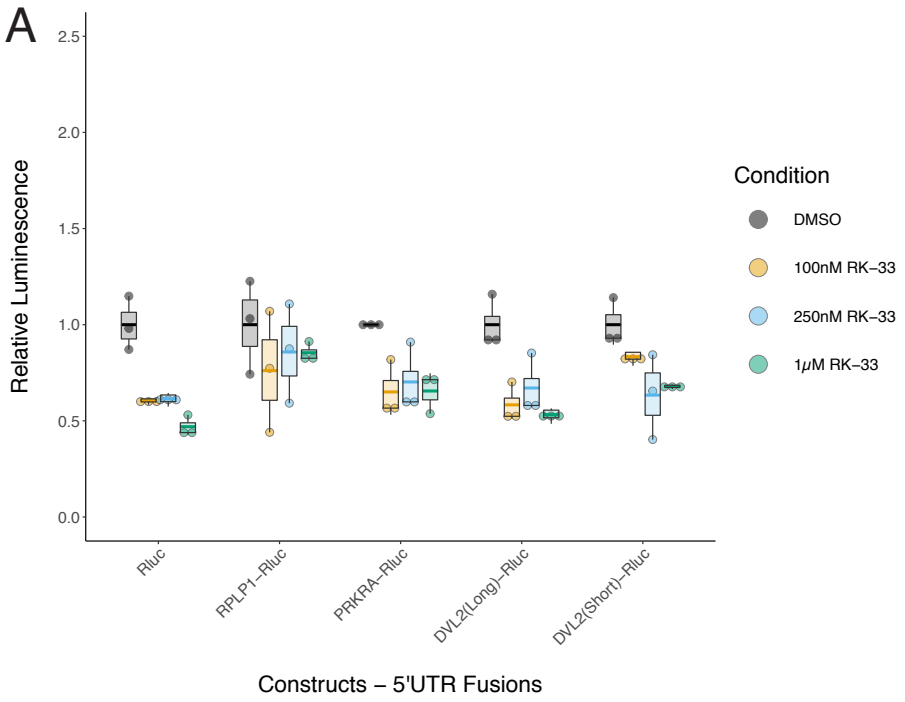


Figure S6: Related to Figure 6 in vitro translation

(A) In vitro translation with the small molecule RK-33 reveals DDX3-independent inhibition of mRNA translation (compare Rluc control to DDX3-specific 5' UTR reporters).