A ubiquitous GC content signature underlies multimodal mRNA regulation by DDX3X

Ziad Jowhar[1,9,10,#], Albert Xu[1,9,10,#], Srivats Venkataramanan[1], Francesco Dossena[8], Mariah L Hoye[2], Debra L Silver[2,3,4,5,6] Stephen N Floor[1,7*] and Lorenzo Calviello[8*]

[1]Department of Cell and Tissue Biology, UCSF, San Francisco, United States

[2]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, United States

[3]Department of Cell Biology, Duke University Medical Center, Durham, United States

[4]Duke Regeneration Center, Duke University Medical Center, Durham, United States

[5]Department of Neurobiology, Duke University Medical Center, Durham, United States

[6]Duke Institute for Brain Sciences, Duke University Medical Center, Durham, United States

[7]Helen Diller Family Comprehensive Cancer Center, San Francisco, United States

[8]Human Technopole, Milan, Italy

[9]Medical Scientist Training Program, University of California, San Francisco, San Francisco, CA 94158, USA

[10]Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA

[#]These Authors contributed equally to this work

[*]Correspondence: Lorenzo.Calviello@fht.org, Stephen.Floor@ucsf.edu

Abstract

The road from transcription to protein synthesis is paved with many obstacles, allowing for several modes of post-transcriptional regulation of gene expression. A fundamental player in mRNA biology is DDX3X, an RNA binding protein that canonically regulates mRNA translation. By monitoring dynamics of mRNA abundance and translation following DDX3X depletion, we observe stabilization of translationally suppressed mRNAs. We use interpretable statistical learning models to uncover GC content in the coding sequence as the major feature underlying RNA stabilization. This result corroborates GC content-related mRNA regulation detectable in other studies, including hundreds of ENCODE datasets and recent work focusing on mRNA dynamics in the cell cycle. We provide further evidence for mRNA stabilization by detailed analysis of RNA-seq profiles in hundreds of samples, including a *Ddx3x* conditional knockout mouse model exhibiting cell cycle and neurogenesis defects. Our study identifies a ubiquitous feature underlying mRNA regulation and highlights the importance of quantifying multiple steps of the gene expression cascade, where RNA abundance and protein production are often uncoupled.

39 ## Introduction

40 The cytoplasmic fate of RNA molecules is impacted their subcellular localization, RNA binding
41 partners, and engagement with the ribosomal machinery. These aspects are strongly
42 interconnected[1], which poses a great challenge, as it increases the number of variables and
43 experimental approaches needed to answer many questions in mRNA biology. To this end,
44 many protocols couple biochemical isolation, or metabolic labeling, of RNA with high
45 throughput sequencing technologies, thus providing a snapshot of the transcriptome at
46 specific stages of the mRNA life cycle, with high throughput and sensitivity. For example, high-
47 throughput sequencing protocols, when coupled to ribosome isolation, such as in Ribo-seq[2],
48 metabolic labeling strategies in SLAM-seq[3], immunoprecipitation of RNA binding proteins
49 (RBP) as in CLIP-seq[4] and many others, have shed light on many regulatory mechanisms
50 pertaining to different aspects of post-transcriptional gene regulation.

51 DDX3X is a multifunctional RNA helicase that is highly expressed in many tissues and able to
52 unwind structured RNA to influence cytoplasmic post-transcriptional gene regulation[5].
53 Together with its ability to bind initiating ribosomes, DDX3X has been often described as a
54 translation regulator, specifically promoting translation of RNA with structured 5'UTRs[6,7].
55 However, as mentioned above, cytoplasmic processes like translation or mRNA decay are
56 intertwined, and connection between the two processes encompass different molecular
57 mechanisms, such as mRNA surveillance mechanisms like nonsense-mediated decay (NMD)[8],
58 ribosome-collision dependent mRNA cleavage[9], and others. In order to understand when and
59 how such processes are coupled, it is important to study the dynamics of such mechanisms.
60 For instance, it has been proposed that miRNA can first trigger translation suppression and
61 then mRNA deadenylation and decapping leading to RNA degradation[10].

62 Mutations in *DDX3X* are associated with a variety of human diseases including cancers and
63 developmental delay[11]. Variant types are disease selective in *DDX3X*, with cancers ranging
64 from primarily loss-of-function alleles in NK-TCL and other blood cancers to nearly exclusively
65 missense variants in medulloblastoma[11]. In *DDX3X* syndrome, missense variants are
66 phenotypically more severe than loss-of-function. Previously, we used functional genomics
67 approaches to identify mechanistic differences between depletion of DDX3X and expression
68 of missense variants[7]. We found that *DDX3X* missense variants predominantly affect
69 ribosome occupancy while DDX3X depletion impacts both ribosome occupancy and RNA
70 levels. However, it is unclear whether the changes in RNA levels constituted a cellular
71 response to translation suppression, often described as "buffering"[12].

72 mRNA regulation has been linked to neurogenesis during development, where multiple RNA
73 binding factors, including DDX3X, ensure correct protein synthesis as cells transition between
74 different fates and states[13]. To that end, it is important to think about the dynamics of gene
75 expression, as complex dynamics of cell proliferation and differentiation ensure correct
76 developmental patterning.

77 In order to access such complex interplays of a multitude of factors which shape gene
78 expression, large-scale consortia have provided a great resource for investigations into gene
79 regulation. While historically devoted to promoting investigation into transcriptional
80 regulation, recent efforts started to provide precious information into post-transcriptional
81 mechanisms, with hundreds of RBPs profiled in terms of both binding and function, by means
82 of CLIP-seq, and knockdown followed by RNA-seq[14]. As in biology many molecular processes
83 are interconnected, large-scale datasets and data amenable to re-analysis are at the very
84 heart of many research efforts[15].

85 Here, we identify how inactivation of DDX3X evolves over time to lead to acute and long-term
86 changes to post-transcriptional gene regulation. We here employ different analytical
87 approaches applied to newly generated experimental data and many previously published
88 studies related to mRNA regulation, to show that GC content is associated with mRNA stability
89 changes following DDX3X depletion. Our analyses indicate that this effect is widespread and
90 is associated with cell cycle changes in mRNA regulation, including RNA stability. This further
91 reinforces roles for DDX3X in RNA stability in addition to translation. Together, our work
92 represents a significant advancement in the understanding of a fundamental regulator, which
93 sits at the very heart of the gene expression cascade.

94

95 # Results

96 ## Time-resolved gene expression regulation by DDX3X.

97 To characterize the dynamics of DDX3X-dependent changes in the gene expression cascade,
98 we employed a previously validated auxin-degron system to efficiently deplete DDX3X protein
99 in the HCT116 colorectal cancer cell line[16], where we found near-complete rescue of gene
100 expression changes by DDX3X expression, thus being able to use this tool to monitor DDX3X-
101 dependent changes. We profiled RNA levels and translation using RNA-seq and Ribo-seq along
102 a time-course of DDX3X depletion, at 4, 8, 16, 24 and 48 hours after auxin or DMSO control
103 treatment. (Figure 1A). Efficiency of DDX3X depletion, together with quality control and
104 general statistics of the generated libraries, can be found in Supplementary Figure 1 and
105 Supplementary Table 1. As expected, the number of differentially expressed genes increased
106 along the time-course, with most changes supporting the role of DDX3X as a positive regulator

107 of translation (Figure 1B). Changes in translation were negatively correlated with changes in
108 mRNA levels, which together contributed to many changes in Translation Efficiency (TE),
109 calculated using Ribo-seq changes given RNA-seq changes (Methods). At a closer look, we
110 observed how “TE_down” mRNAs undergo translation suppression in the early time point
111 after DDX3X depletion, with their mRNA levels increasing in the later time points (Figure 1C).
112 The opposite behavior is observed for “TE_up” mRNAs, exhibiting higher ribosome occupancy
113 first, and lower mRNA levels later. Such behavior was more evident when showing time-point
114 specific changes and binning mRNAs in a 2D grid on the Ribo-seq/RNA-seq coordinate plane
115 (Figure 1D, Methods), which highlighted a common regulatory mode, with early translation
116 regulation followed by changes in mRNA levels.

117 This analysis shows the time-resolved dynamics of mRNA regulation by DDX3X, with hundreds
118 of mRNAs changing in their steady-state levels albeit showing the opposite directionality in
119 translation rates.

120

121

122

123

124

125

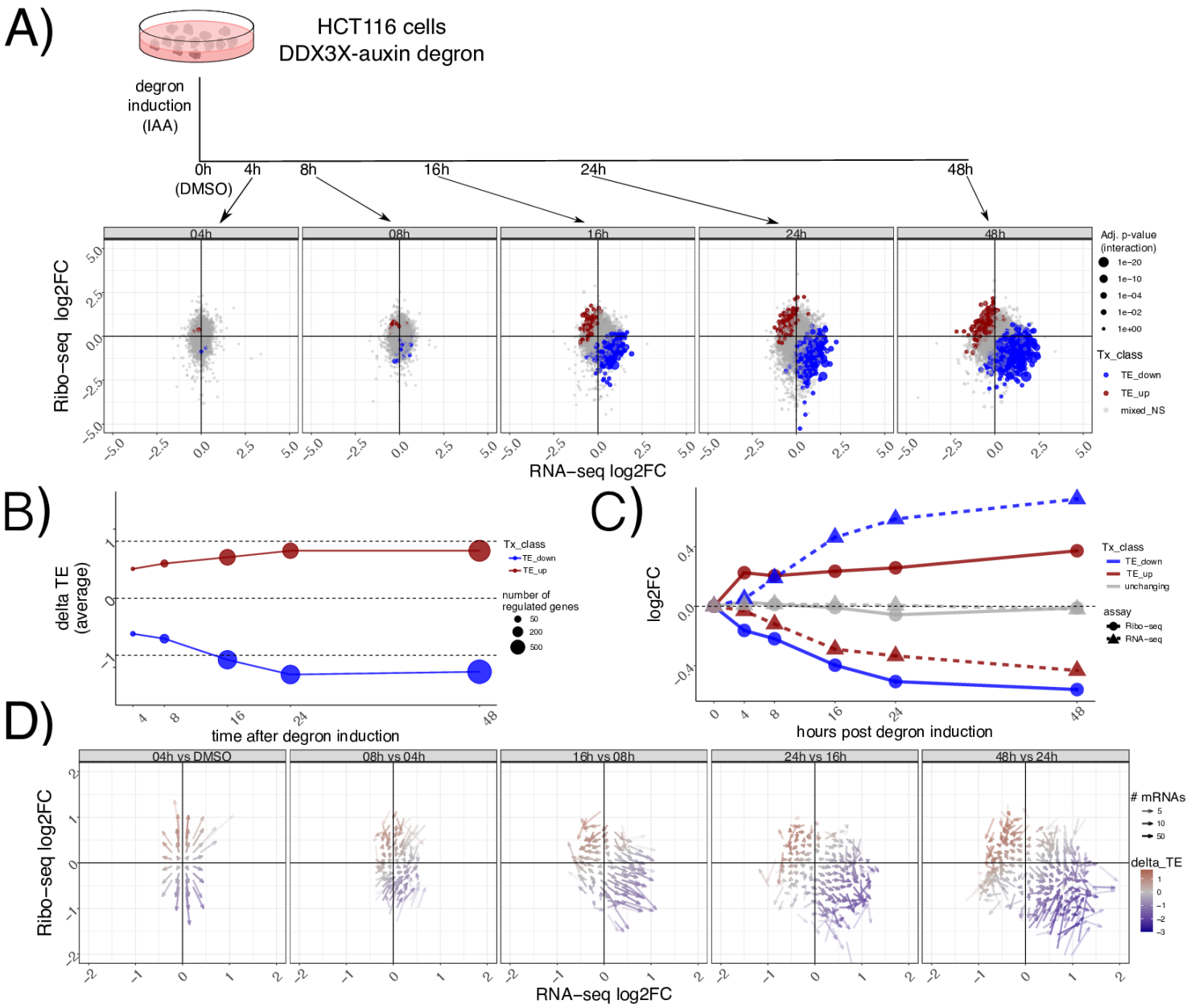


Figure 1: Dynamics of mRNA regulation by DDX3X

In A) a description of the experimental design. Below Ribo-seq and RNA-seq fold changes at different time points. Different regulated classes are shown in different colors. The size of the dots indicates the adjusted P-values for differential translation efficiency test (Methods). TE: translation efficiency, NS: not significant. In B) average delta TE values (differences in TE values) for each class along the time course. The size of the dots indicates the number of significantly changing mRNAs. C) progression along the time course for mRNA regulated 48h post degron induction. RNA-seq and Ribo-seq fold changes are shown at each time point. D) Differences in Ribo-seq or RNA-seq fold changes between each time point and the previous one, shown as a vector plot. Magnitude of changes shown as a color gradient, while transparency of the vectors indicates the number of mRNAs in each coordinate bin (Methods).

127 ## Translation suppression by DDX3X is coupled with mRNA stabilization.

128 Changes to transcript levels can result from changes in transcription rates or post-
129 transcriptional regulation. To identify the relative contribution of different processes to RNA
130 levels, we used our time-course dataset to calculate changes in transcription, processing and
131 stability using INSPEcT[17]. INSPEcT uses the proportion of intronic versus exonic reads to
132 identify nascent vs. mature transcripts, and uses a system of ordinary differential equations
133 (ODEs) to infer rates of RNA synthesis, processing and decay. Compared to non-regulated
134 mRNAs, regulated mRNAs showed modest changes in transcription rates, suggesting
135 transcription changes are not the major contributor to RNA level changes following DDX3
136 depletion, In contrast, we found more pronounced changes in mRNA stability as evidenced
137 by TE down transcripts (Figure 2A). As our initial RNA-seq protocol was not designed to
138 capture pre-mRNA molecules, we validated our estimated mRNA stability changes by
139 employing the 4sU metabolic labeling SLAM-seq protocol[3] in our degron system after 8 hours
140 of DDX3X depletion, in a way to detect changes in mRNA stability at early time points. Briefly,
141 cells were incubated with 4sU to comprehensively label transcribed RNAs, and their
142 abundance was followed after 8h of DDX3X degron activation, using DMSO as control. 4sU
143 treatment induces T>C conversions in the sequenced cDNA molecules, which can be used to
144 monitor mRNA stability changes after a uridine chase, as shown in Figure 2B. As expected, we
145 observed a drastic drop in T>C harboring reads after the chase, which reflects mRNA decay
146 rates (Supplementary Figure 2). As shown in Figure 2B, after a labeling time of 24 hours, the

147 percentage of reads harboring T>C
148 mutations was different for the regulated
149 categories (Methods) after only 8 hours of
150 degron induction, confirming the
151 stabilization of translationally suppressed
152 mRNAs upon DDX3X depletion. While the
153 modest depth and resolution of our SLAM-
154 seq dataset (Supplementary Figure 2)
155 couldn't allow for more detailed insights
156 on mRNA changes, it represented an
157 important validation of mRNA stability
158 regulation by DDX3X. In addition, we
159 profiled RNA abundance via qPCR
160 combining our DDX3X degron system with
161 ActD treatment, to measure RNA stability
162 changes. We selected few target genes:
163 JUND was identified in our data as a
164 stabilized RNA, while EIF2A was identified
165 to be degraded. RACK1, LGALS1, and
166 PFN1 were used as controls to normalize
167 with via RT-PCR with Taq-man probes.

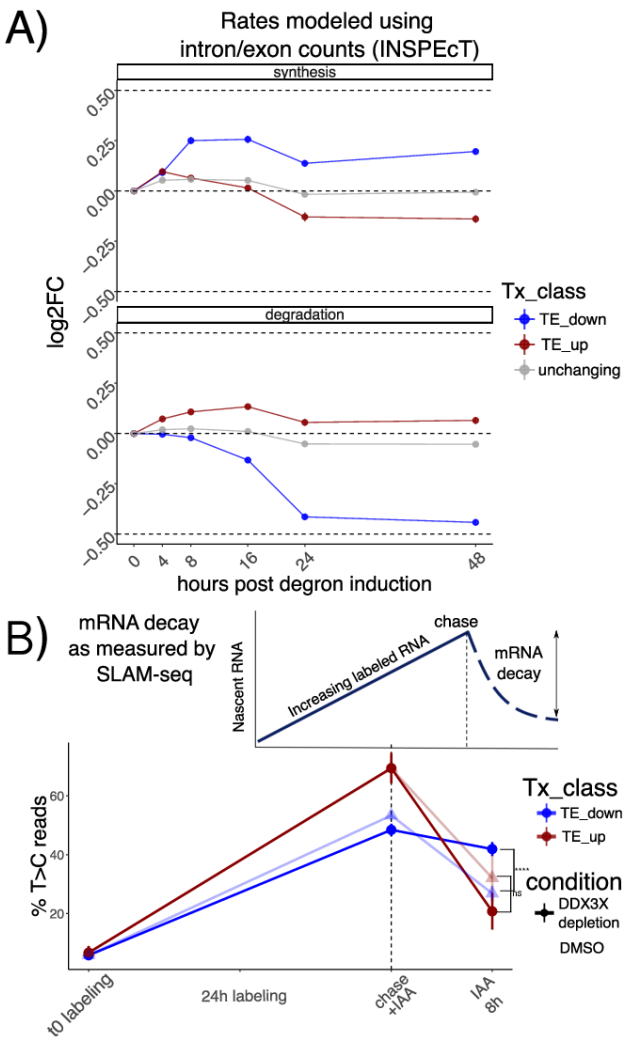


Figure 2: Stabilization of untranslated mRNAs

A) Synthesis and decay as inferred by INSPEcT: different regulated classes in different colors along the time course. Log2FC of estimated rates with respect to control are shown on the y axis. B) Schematic of a SLAM-seq experiment (above). Real data shown at the bottom: percentage of T>C-containing reads on the y axis after labeling and chase. DDX3X degron (using DMSO as a control) was triggered together with the chase reaction to monitor differences in decay rates upon DDX3X depletion. Significance values from a one-sided Wilcoxon test.

168 JUND RNA was stabilized after 24 hours with knock down of DDX3 and Actinomycin D (ACTd)
169 treatment (Supplementary Figure 3A); EIF2A RNA was more degraded after 24 hours with
170 knock down of DDX3 and ACTD (Supplementary Figure 3B). These results show an overall

171 good agreement between the qPCR and the sequencing-based assays, despite the difficulty
172 arising from choosing control genes and the modest fold changes observed in the sequencing
173 data.

174 By profiling ribosome occupancy, steady state transcript levels, and mRNA decay, this analysis
175 shows that DDX3X depletion triggers multiple modes of post-transcriptional regulation,
176 involving translation suppression and a subsequent wave of mRNA stabilization.

177

178

179

GC-rich coding sequences underlie mRNA regulation by DDX3X.

With hundreds of mRNAs post-transcriptionally regulated after DDX3X depletion, we aimed to identify specific features belonging to up- or downregulated targets. We therefore built regression models to quantitatively predict levels of TE changes (Methods, Supplementary Table 2). We used different biophysical properties of genes and mRNAs, (e.g. length and GC content) and several gene and transcript features (e.g. introns, 3'UTR, etc.., Methods) as

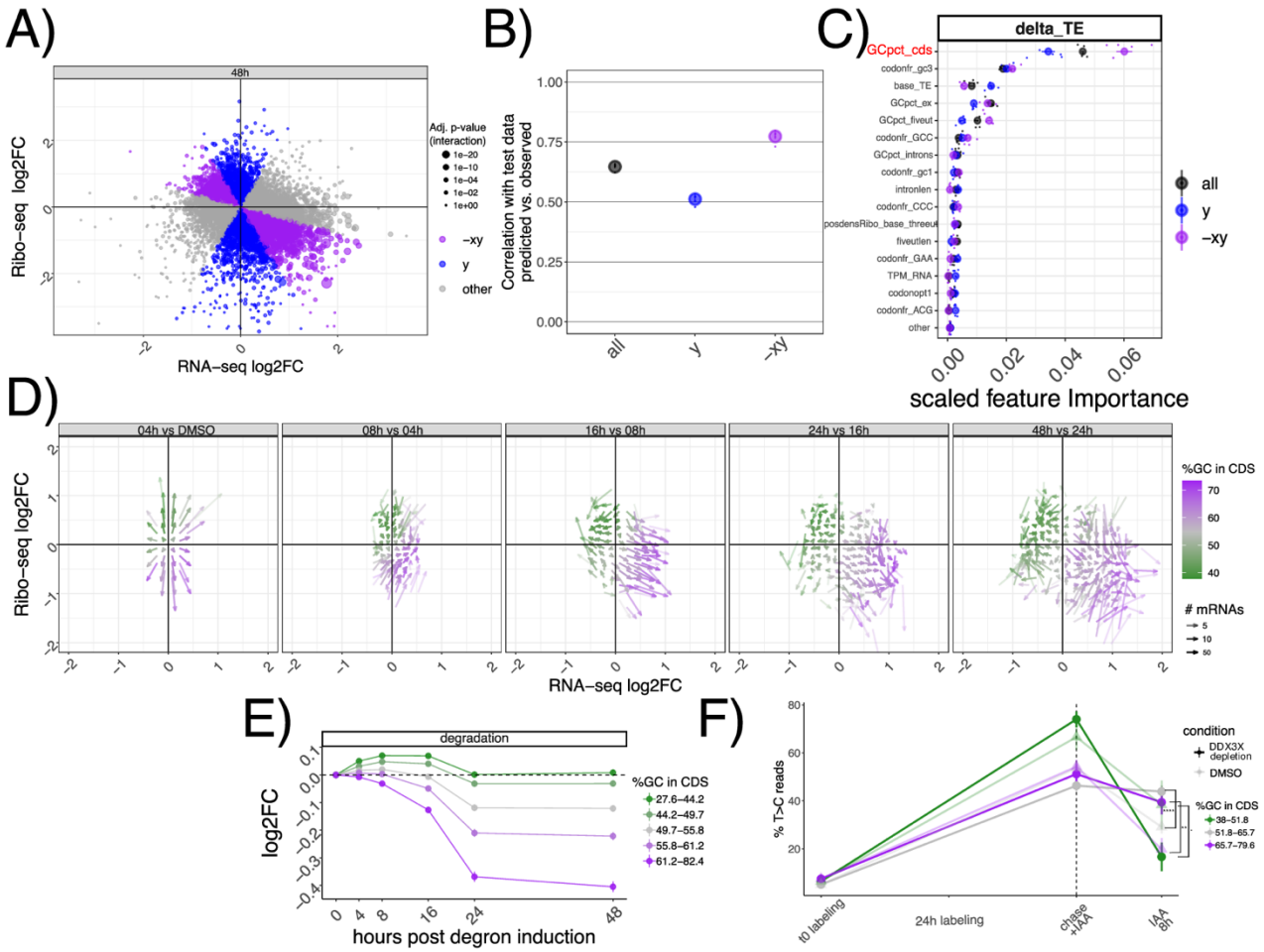


Figure 3: GC content in the coding sequence predicts regulation by DDX3X

A) Classification of different mRNAs according to their change in mRNA levels or translation. In B) model performance (correlation between predicted vs. real values) on unseen test data of the random forest regression model for transcript classes as defined in A). C) Predictive power of most informative features, with their importance values (Methods) plotted on the x axis. Feature pertaining to GC content in different section of transcripts (GCpct*), baseline translation levels (base_TE), codon frequencies (codonfr*), positional read density (posdens*), and length features (intronlen) are displayed. D) Vector plot as in Figure 1D, highlighting GCcds values. Partition of inferred degradation rate (E) or SLAM-seq profiles (F,) for mRNAs partitioned by GCcds values. Significance values for SLAM-seq from a one-sided Wilcoxon test.

186 features for a Random Forest regression model. Given the extensive literature on codon-
187 mediated mRNA stability regulation, we added codon frequencies and previously validated
188 codon optimality calculations[18]. Also, we added measured GC-content at 1st, 2nd or 3rd codon
189 position, as it was recently shown to potentially play a role in mRNA stability regulation[19,20].
190 In addition, to pinpoint features predictive of mRNA stability changes rather than translation
191 changes exclusively, we divided transcripts according to their position in the Ribo-seq/RNA-
192 seq coordinate system, to capture mRNAs where changes between assays agreed or not
193 (Figure 3A, Methods). Interestingly, the categories differed in their DDX3X binding pattern
194 (Supplementary Figure 4): re-analysis of our previously published PAR-CLIP data showed how
195 stabilized targes (x,-xy groups) have a lower T>C conversion signal in their 5'UTRs, and a higher
196 signal in CDS peaks, with the opposite being true for true translation targets (y group). This
197 analysis suggests that stabilized mRNAs might be regulated differently than "canonical"
198 translationally suppressed targets.

199 As shown in Figure 3B, the Random Forest model predicted TE changes with high precision,
200 especially in cases where mRNA stability and translation were anti-correlated (-xy group). In
201 addition, this model calculated the predictive power of each input feature (Figure 3C,
202 Methods), which highlighted GC content in the coding sequence (which we will refer to as
203 GCcds) as the most important feature. Feature selection is a very important method to select
204 predictive features, especially when facing high levels of multicollinearity (Supplementary
205 Figure 5). To validate the results from the Random Forest regression, we used Lasso
206 regression (Methods), another well-known method for feature selection. Results from the
207 Lasso regression were similar, and also identified GC content in the coding sequence as the
208 most relevant feature in predicting TE changes (Supplementary Figure 6). GC content in the

209 CDS remained the top predictor when using additional features, such as GC content in
210 different sections of the CDS, or amino acid frequencies (Supplementary Figure 7).

211 In the light of these results, we tested whether GCcds was associated with the DDX3X-
212 dependent transcriptome dynamics reported above. As shown in Figure 3D, mRNAs
213 partitioned on the Ribo-seq/RNA-seq coordinate system based on their GCcds value.
214 Moreover, stability values from both INSPEcT and SLAM-seq partitioned according to GCcds
215 values (Figure 3E-F). A similar, albeit weaker, separation was observed for predicted
216 transcription and processing rates (Supplementary Figure 8).

217 By using multiple analytical approaches, we here show how GCcds, not just GC content in
218 general, or in other sections of the transcriptome, is a predominant feature of stabilized, yet
219 untranslated, mRNAs following DDX3X depletion.

220

221 ## GC content in the coding sequence is a ubiquitous signal in mRNA regulation.

222 Given the extensive connections between different aspects of mRNA regulation by thousands
223 of regulators, we tested the breadth of the influence of features such as GCcds in other
224 studies of RNA regulators. We re-analyzed >2000 RNA-seq samples (Methods) from the
225 recent ENCODE RBPome[14] study encompassing >200 RBP knockdowns, and performed
226 differential analysis followed by predictive modeling using the same methods and features as

227 described in the previous section, this time aiming at predicting changes in mRNA levels

228 (Figure 4A).

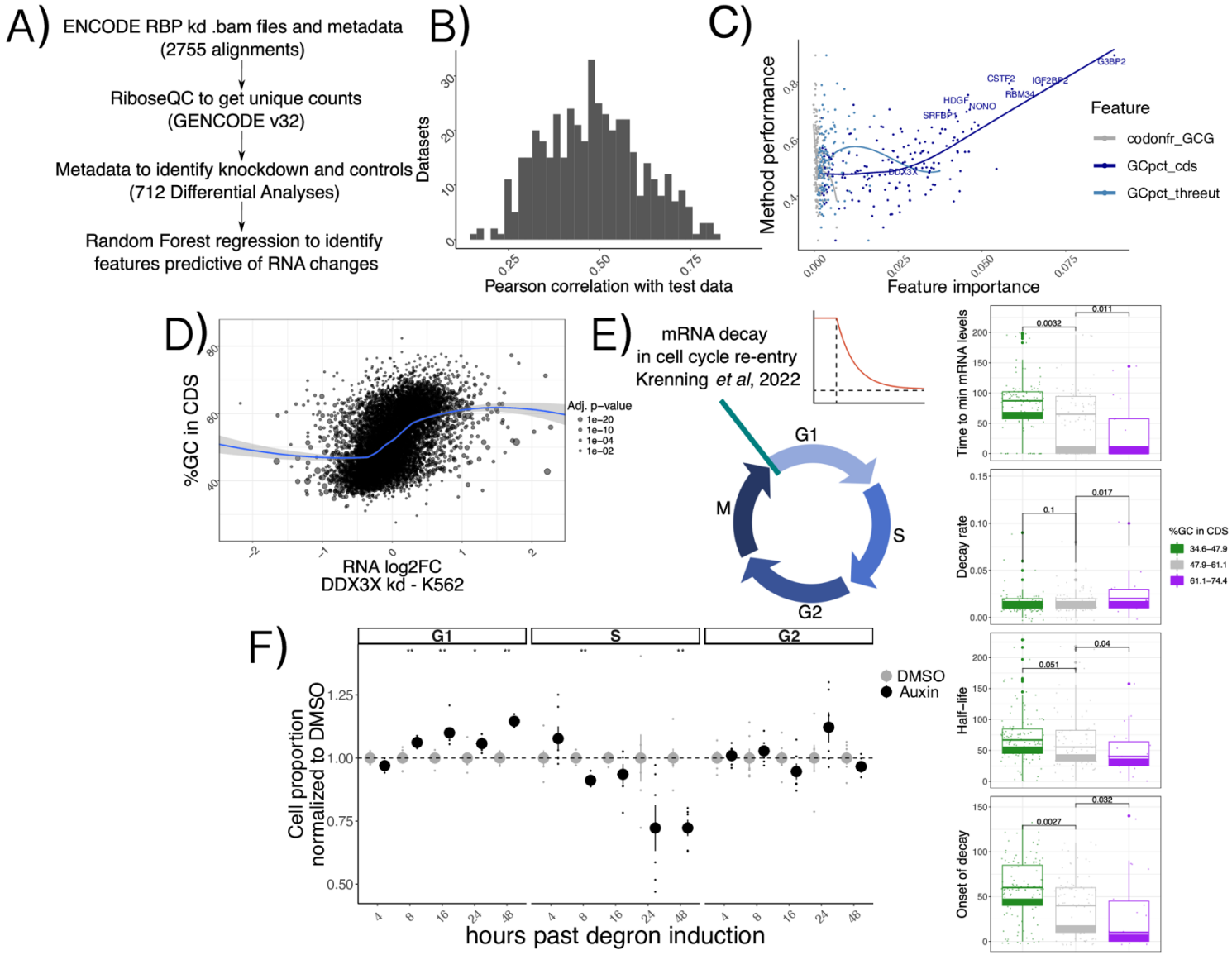


Figure 4: A ubiquitous feature in mRNA regulation

A) Schema describing the ENCODE analysis strategy. B) Histogram representing overall model performance across datasets. C) Model performance (spearman correlation between predicted and real values on unseen test data) on the y axis, with importance of 3 example features variables (indicating their predictive value) on the x axis. Top knockdown experiments, together with DDX3X, are show with labels. Data shown is from shRNA KD experiments in K562 cells. The linear relationship between GCcds importance and model performance indicates its relevance as the top predictor of RNA changes in dozens of datasets. D) mRNA level changes against GCcds values in a DDX3X knockdown experiment in the ENCODE dataset. E) Schematics of the cell cycle data used. Values for different kinetic parameters were partitioned according to GCcds values of their mRNAs and tested for significant differences. F) Normalized cell proportion (obtained by dividing cell percentages between Auxin treatment and DMSO) in different stages of the cell cycle along the degron time course. An increase in G1 and decrease in S phase can be observed at later time points. Significance values come from a Wilcoxon two-sided test (n=6 in each condition).

229 We first grouped datasets according to knockdown efficiency, which varied according to
230 knockdown method and cell line (Supplementary Figure 9, Methods). We selected the sample
231 with the highest knockdown efficiency for each RBP and called feature importance using our
232 analytical pipeline. Predictive power of our Random Forest regression strategy varied across
233 different datasets (Figure 4B). Once again, the strongest predictor of mRNA changes was
234 GCcds, whose predictive power dominated over other variables (Figure 4C, Supplementary
235 Figure 10). As expected, changes upon DDX3X knockdown in the ENCODE data also exhibited
236 a clear dependency over GCcds (Figure 4D), albeit to a lower degree compared to our degron
237 dataset, likely due to differences in DDX3X depletion strategies and, importantly, to our
238 translation profiling dataset, which allowed us to distinguish between specific classes (i.e.
239 "TE_down") of regulated mRNAs (Discussion).

240 Given the widespread relevance of GCcds as a predictor of post-transcriptionally regulated
241 targets, we reasoned that a major cellular process might mediate the observed mRNA
242 changes. We re-analyzed data from a recent study[21] focused on mRNA clearance during cell
243 cycle re-entry, where the authors used a FUCCI (fluorescent, ubiquitination-based cell-cycle
244 indicators) cell system coupling RNA labeling, scRNA-seq and single-molecule imaging
245 techniques to find extensive decay differences among different transcripts, potentially
246 related to poly-A tail mediated regulation. Despite a lower throughput when compared to
247 sequencing-based experiments, kinetic parameters estimated from their data (exemplified in
248 the decay curve in Figure 4E) showed significant differences when partitioned by GCcds values
249 (Figure 4E). mRNAs rich in GCcds showed lower half-life values, and fast decay kinetics at cell
250 cycle re-entry, with the opposite trend exhibited by mRNAs poor in GC content in their coding
251 sequence. Motivated by this finding, we decided to investigate differences in cell cycle
252 dynamics in our degron system, by using 5-ethynyl-2'-deoxyuridine (EdU) incorporation

253 followed by FACS analysis (Methods, Supplementary Figure 11). As shown in Figure 4F and
254 Supplementary Figure 12, DDX3X depletion resulted in cells staying more in G1 and less in S
255 phase when compared to controls, throughout the time course.

256 By re-analysis of thousands of RNA-seq samples, these results show the prevalence of GCcds
257 in post-transcriptional regulation and RBP functions, with a potential role for cell-cycle
258 dependent mRNA dynamics in shaping such a regulatory phenomenon.

259

260 A shift in 5'-3' RNA-coverage as a hallmark of mRNA stabilization.

261 In addition to gene-level aggregate measures of abundance, we investigated whether changes
262 in decay could be identified by taking advantage of the high resolution of RNA-seq
263 experiments across gene bodies, which has previously been employed to inform about mRNA
264 decay[19]. We leveraged our time-resolved degron dataset to investigate changes in 5'-3'
265 coverage, a known hallmark of RNA degradation often employed to verify overall integrity of
266 cellular mRNAs or to estimate transcript-level decay. We calculated 2 different metrics, using
267 the strategy illustrated in Figure 5.

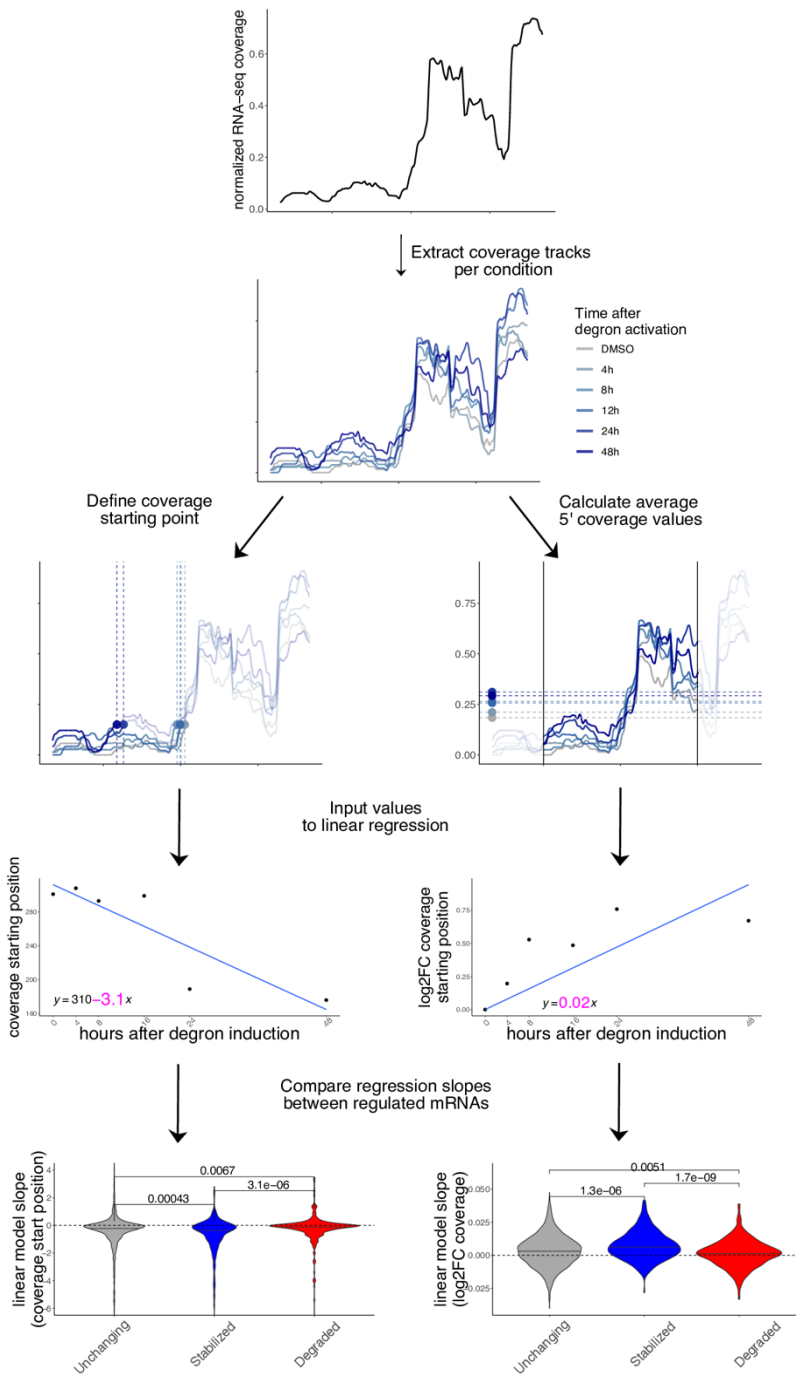


Figure 5: Coverage analysis of regulated mRNAs reveals changes in 5'-3' decay

Coverage analysis strategy in the degron dataset using a practical example (*CSRNP2* gene): coverage starting point is first identified using pooled data, then coverage tracks for each experiment are extracted. Coverage starting points (in transcript coordinates) and coverage values (log2FC to DMSO) are calculated for each time point and used as input to a linear model. The beta coefficient (shown in pink) for each model is then extracted for each mRNA and values are compared across different classes (stabilized vs unchanging vs degraded). More details are available in the Methods section. P-values from one-sided Wilcoxon test.

268 Initially, we pooled all samples to identify the major isoform for each gene (Methods), and

269 the first position at 15% of the maximum coverage. We then calculated such position for each
270 time point. Importantly, coverage values were normalized for each transcript, thus controlling
271 for expression level changes. Also, we did not observe similar changes at the 3' end of
272 transcripts (Supplementary Figure 13). We then used coverage starting points as input for
273 linear regression. The regression coefficient was extracted and compared across the top 250
274 stabilized, degraded, and control mRNAs, alongside 1500 control transcripts. As shown in
275 Figure 5, coverage values on stabilized mRNAs started as an earlier position in the transcripts,
276 with moderate albeit significant differences between categories, indicating a lower 5'-3'
277 decay along the DDX3X degron time course. The opposite trend was observed for degraded
278 transcripts. Similarly, we calculated average coverage values in a window of 300nt around the
279 coverage start and applied a similar strategy: 5' coverage values increased along the time
280 course, confirming the accumulation of translationally suppressed mRNA species otherwise
281 destined for degradation. Results were similar when using different cutoffs for the definition
282 of coverage starting point (Supplementary Figure 14).

283 To test whether the suppression of 5'-3' decay of untranslated transcripts by DDX3X occurs *in*

284 *vivo*, we re-analyzed recent RNA-seq/Ribo-seq dataset in a conditional *Ddx3x* (cKO) mouse

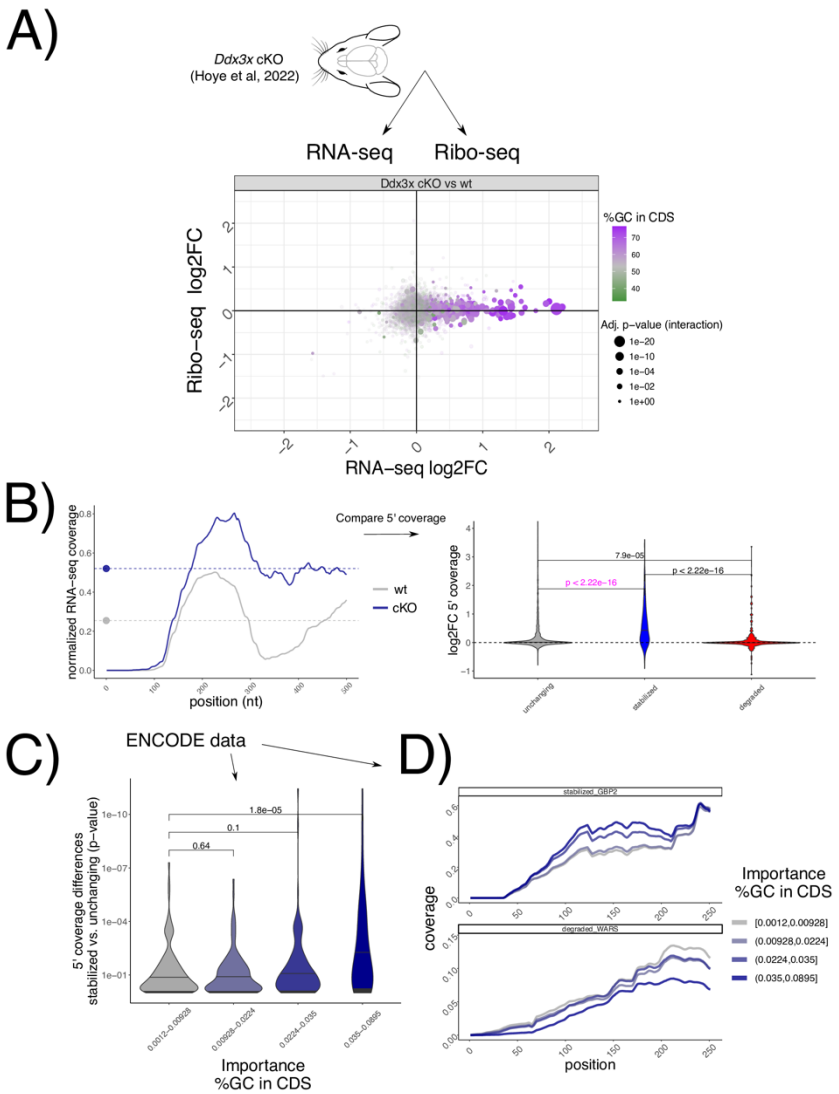


Figure 6. GCcds - mediated mRNA stabilization is detectable in vivo and across the ENCODE RBP database.

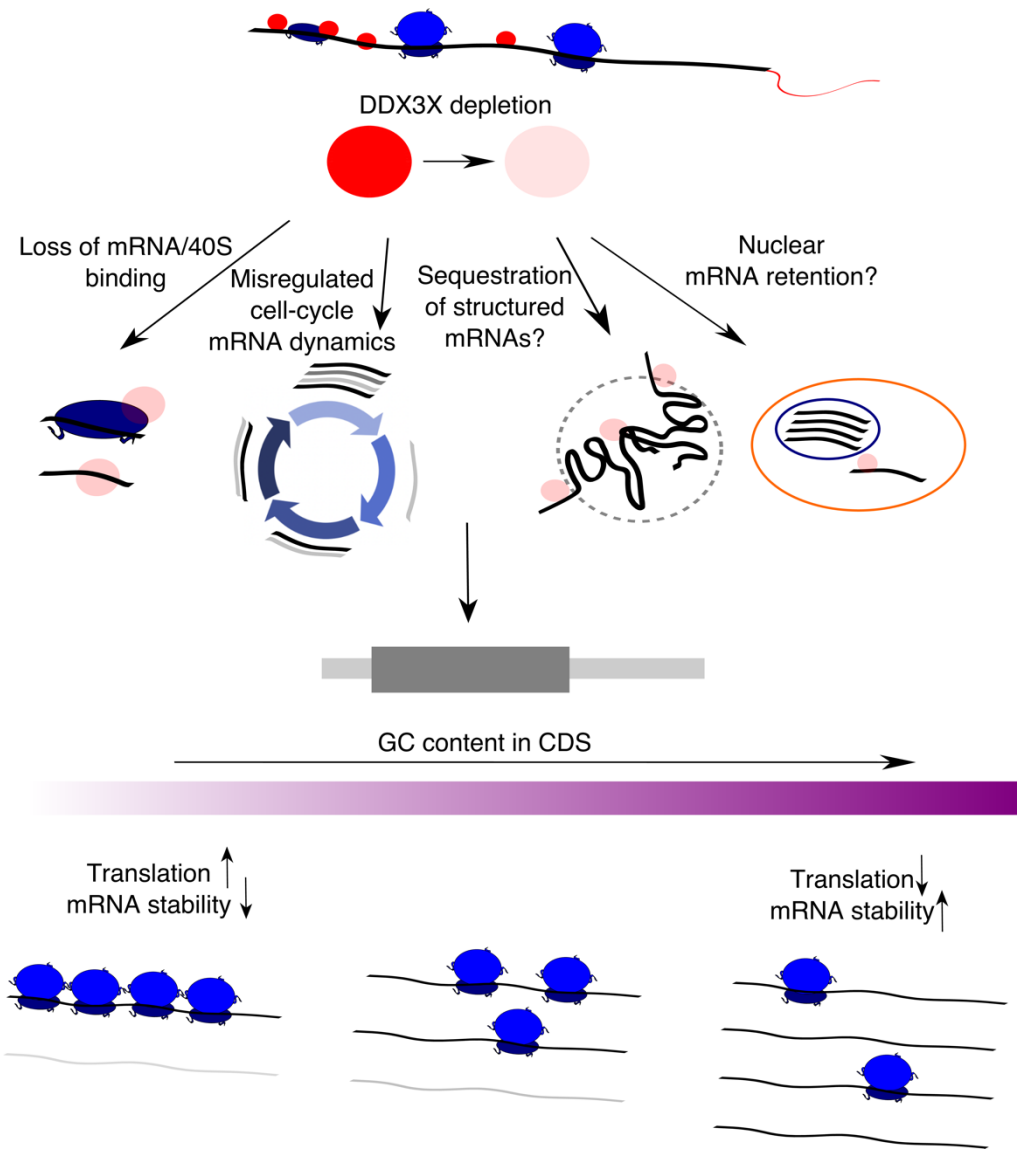
A) Changes in Ribo-seq and RNA-seq levels in a conditional Ddx3x mouse model, as in Figure 1A, showing GCcds values. B) Strategy for coverage analysis in the mouse Ddx3x cKO experiment, shown for the Ctxn1 gene. Differences in coverage values are extracted and compared across regulated mRNAs. In C) same strategy as in Figure 5A applied to each differential analysis followed by RBP knockdown in the ENCODE dataset. Differences in coverage values between stabilized and unchanging mRNAs (shown by p-values, as calculated as in panel B), in pink color) are plotted against GCcds importance (x axis). D) Example mean coverage on 2 mRNAs (1 stabilized and 1 degraded), partitioning RBP knockdown datasets by their GCcds importance. An increase in coverage can be observed for the stabilized mRNA, while the opposite trend is visible for a degraded transcript.

285 model[13](Figure 6), where cell cycle and neurogenesis defects are evident when *Ddx3x* is
286 depleted in neuronal progenitors. After applying our analytical pipeline, we observed that the
287 accumulation of untranslated transcripts is even more evident in this *in vivo* model, as is its
288 relationship with GCcds values (Figure 6A). Analogous to the strategy presented in Figure 5,
289 5' coverage values, as well as coverage starting points (Supplementary Figure 15), differed
290 significantly between wild type and *Ddx3x* cKO animals (Figure 6B) in regulated transcripts,
291 while no difference was detected at the 3'end (Supplementary Figure 16).

292

293 Leveraging again the power of hundreds of RNA-seq experiments, we examined 5' coverage
294 profiles in the ENCODE dataset, partitioning experiments by their dependency on GCcds
295 values. Differences between stabilized and control mRNAs are greater as the GCcds signature
296 is more predominant (Figure 6C). Aggregating different experiments according to their GCcds
297 dependency for example transcripts (Figure 6D) confirm this phenomenon, where both
298 coverage starting position and coverage values changed across different datasets, indicative
299 of mRNA decay regulation.

300 Taken together, we provide evidence for *in vivo* DDX3X-mediated stabilization of untranslated
301 transcripts, its dependence on GCcds values, and, supporting the different analyses reported
302 in this study (Figure 7) a high-resolution RNA-seq coverage analysis strategy to investigate
303 GCcds-related mRNA decay regulation, with support from hundreds of post-transcriptionally
304 perturbed transcriptomes.



305

306

Figure 7. A model for multimodal mRNA regulation by DDX3X

A schematic showing the effects of DDX3X depletion on GC-content related changes in translation and mRNA stability, highlighting potential molecular mechanisms underlying this phenomenon.

307

308

309 ## Discussion

310 The multifaceted role of DDX3X, described as involved in different molecular processes, often
311 hinders the ability to understand its functions, especially considering the interconnected
312 nature of molecular processes in the cell. Multiple mRNA features might underlie different
313 modes of regulation, as we previously showed and experimentally validated 5'UTR
314 dependencies underlying DDX3X translation regulation[7]. This outlines an unmet need for
315 studies linking multiple aspects of the gene expression cascade.

316 In addition to profiling RNA levels and translation, we further dissected dynamics of
317 cytoplasmic regulation by DDX3X, by employing a time course of efficient DDX3X depletion
318 (Figure 1A). Akin to previous studies observing translation suppression preceding mRNA
319 changes during miRNA-mediated regulation[10], we observed an accumulation of
320 translationally suppressed RNAs. This highlights the importance to profile not only mRNA
321 abundance but also translation levels, which, in absence of quantitative estimates of
322 regulated protein levels, can greatly help researchers understanding the functions of many
323 cryptic regulators often involved in multiple processes, like DDX3X and other RBPs[22]. Despite
324 relatively fast kinetics of DDX3X degradation from our degron system, more work needs to be
325 performed to pinpoint exactly what changes occur right after DDX3X depletion, and to more
326 precisely quantify the lag between translation suppression and mRNA stabilization.

327

328 By employing multiple techniques for feature selection, we identified a major feature
329 underlying mRNA regulation by DDX3X, as well as by many other post-transcriptional
330 regulators. An important area of investigation for the future is to employ more unbiased

331 approaches, akin to recent Natural Language Processing-inspired methods in transcription
332 regulation[23], in mRNA biology to accurately estimate the relevant features directly from data
333 rather than specified by potentially biased approaches. In our hands, the relevance of GCcds
334 is clearly picked up by both the Random Forest and the Lasso (Supplementary Figure 4).
335 Importantly, we included similar features, such as overall GC content[24], in UTRs, introns etc.,
336 alongside codon frequencies[20] and previously estimated values of codon optimality.

337 Our study suggests that data-driven approaches to functional transcriptomics are highly
338 needed, where data from multiple experiments are routinely re-analyzed to test hypotheses
339 and provide new insights into the complex world of mRNA biology. However, while profiling
340 translation allowed us to focus on specific mRNA classes and their features, no large-scale
341 translation profiling study exists yet, with few, precious small atlases recently appearing in
342 the literature[25]. The current ENCODE RBP series is of great value to many mRNA biology
343 researchers worldwide and it has been an invaluable resource for many recent studies[26,27],
344 yet an extension of these approaches which includes other aspects of post-transcriptional
345 regulation, such as translation and stability, is in great need.

346

347 In the original ENCODE RBP study[14], gene expression estimates were GC-corrected for each
348 sample, as GC content has been often reported as a confounder, especially when comparing
349 across sequencing technologies and labs. Given the presence of GC-related biases in
350 sequencing-based assays, we think that great caution must be taken when observing
351 expression changes driven by GC content features, especially when interpreted as direct
352 effects from single molecular factors. Our degron time course analysis, despite containing
353 dozens of features pertaining to GC content measures, detected GC content specifically in

354 coding sequence as a feature underlying regulation, and this region-specific effect is not
355 consistent with a general confounding role for GCcds. Moreover, our analysis focused on
356 differences upon a perturbation under a single sequencing platform and laboratory settings,
357 which are likely to have similar GC-related confounders, should there be any. Important
358 confirmation of the relevance of GCcds and its relationship to mRNA dynamics also came
359 from: employing SLAM-seq to estimate differences in stability (Figure 2), qPCR validations
360 (Supplementary Figure 3), re-analysis of *in vivo Ddx3x* cKO RNA-seq/Ribo-seq (Figure 6), re-
361 analysis of hundreds of RBP perturbations in human cell lines (Figure 4), and by analyzing
362 kinetics extracted by transcriptome dynamics in cell-cycle specific states (Figure 4).

363

364 Together with well-established differential analysis statistical methods, which allowed us to
365 robustly identify different classes of regulated mRNAs, we exploited the high resolution
366 offered by RNA-seq to analyze differences in 5'end coverage for thousands of individual
367 transcripts (Figure 5), as an additional metric reflecting active regulation of mRNA decay
368 mechanisms. We posit that popular analysis strategies for -omics techniques, despite their
369 popularity over more than a decade, often obscures information with regards to mRNA
370 processing and other molecular mechanisms, which can be uncovered by dedicated
371 computational methods. Importantly, such dynamics are invisible (or, worse, can significantly
372 distort quantification estimates) when performing gene-level analyses.

373

374 The mechanism, or mechanisms, by which GC content in coding regions shapes mRNA
375 dynamics is still to be determined. We speculate that complex RNA structures in the coding
376 sequence can form in the absence of active translation elongation, and such structure may

377 mediate degradation, helped by RNP complexes in the cytoplasm. However, recent literature
378 focused on the role of different codons in mediating such effect[18]. In our hands, codon-
379 mediated effects seem to be negligible when considering the overall GCcds values, but more
380 work needs to be done to identify cases where one or the other, or a mix of the two, can
381 mediate mRNA decay on different transcripts. The involvement of mRNA dynamics during the
382 cell cycle (Figure 4) suggests a model where, during cell cycle - dependent translation
383 suppression, mRNAs are able to fold structures in the coding sequence promoting decay, and,
384 when such processes are misregulated (e.g., by depleting multifunctional RNA helicases such
385 as DDX3X), this process is less efficient. The extent to which cell cycle changes might depend
386 on direct DDX3X binding and regulation remains to be elucidated. Further work needs to be
387 done to refine the exact function, together with the subcellular localization, of regulated
388 mRNAs. For instance, mRNA retention in the nucleus might be an additional
389 underappreciated mode of gene expression control[28], and is in line with our observation
390 about the untranslated status of regulated transcripts. However, we identified GC content in
391 the coding sequence as the hallmark feature for stabilized transcripts, a feature which is
392 defined by translation in the cytoplasm.

393 While RBP binding data remains an important starting point from which we can build testable
394 hypothesis, simple binding-to-function paradigms might also create bias when trying to
395 explain complex phenotypes arising from RBP misfunction. Moreover, we observed how
396 binding patterns might different between different regulated classes (Supplementary Figure
397 4). In our previous study we investigated the changes in translation and RNA abundance using
398 a DDX3X helicase mutant; one of the observations we made pertained to the lack of RNA
399 changes in our data, suggesting a potential function for the helicase activity in orchestrating
400 such changes.

401 Previous work implicated DDX3X in mediating cell cycle dynamics by a variety of
402 mechanisms[29], including a direct regulation of cyclin E1 translation[30], which however was not
403 among the most regulated mRNAs in our dataset (Supplementary Table 2). More work needs
404 to be done to accurately quantify mRNA dynamics and RBP functions in the cell cycle, where
405 translation regulation mechanisms[31,32] ensure controlled rates protein synthesis. The
406 connection between cell cycle, sequence content and mRNA regulation is reinforced by the
407 *in vivo* data, adding to the importance of studying post-transcriptional regulation along the
408 neurogenesis axis[33,34], where the equilibrium between proliferation, apoptosis and
409 differentiation[35] shapes the complexity of the developing brain.

410

411

412

413

414

415

416

417

418

419

420

421

422 # Methods

423

424 ## Ribo-seq and RNA seq experimental protocol

425

426 HCT116 cells with inducible degradation of DDX3X (as previously described[16]), were plated in
427 15cm plates at 20% confluency (~3.5x10 6 cells/plate). 48 hours post plating, when the cells
428 were at ~ 70% confluency, the media was changed and fresh media with 500 μM IAA (Indole-
429 3-acetic acid, the most common naturally occurring Auxin hormone) (Research Products
430 International, cat: I54000-5.0) or DMSO was added to cells. Cells were harvested at 0, 4, 8,
431 16, 24, and 48 hours post IAA addition. Cell number did not appreciably increase over the 48
432 hours of the experiment. To quantify DDX3X protein, we used an anti-DDX3X antibody
433 described in previous work[7] normalized to an anti-GAPDH antibody (Rockland
434 Immunochemicals, cat: 600-401-A33S).

435 Cells were treated with 100 μg/ml cycloheximide (CHX), washed with PBS containing 100
436 μg/ml CHX, and immediately spun down and flash frozen. Once all time-points were collected,
437 the cells were thawed and lysed in ice-cold lysis buffer (20 mM TRIS-HCl pH 7.4, 150mM NaCl,
438 5 mM MgCl2, 1mM DTT, 100 μg/ml CHX, 1 % (v/v) Triton X-100, 25 U/ml TurboDNase
439 (Ambion)). 240 μl lysate was treated with 6 μl RNase I (Ambion, 100 U/μl) for 45 minutes at
440 RT with gentle agitation and further digestion halted by addition of SUPERase:In (Ambion).
441 Illustra Microspin Columns S-400 HR (GE healthcare) were used to enrich for monosomes, and
442 RNA was extracted from the flow-through using Direct-zol kit (Zymo Research). Gel slices of
443 nucleic acids between 24-32 nts long were excised from a 15% urea-PAGE gel. Eluted RNA

444 was treated with T4 PNK and preadenylated linker was ligated to the 3' end using T4 RNA
445 Ligase 2 truncated KQ (NEB, M0373L).

446 Linker-ligated footprints were reverse transcribed using Superscript III (Invitrogen) and gel-
447 purified RT products circularized using CircLigase II (Lucigen, CL4115K). rRNA depletion was
448 performed using biotinylated oligos as described[36] and libraries constructed using a different
449 reverse indexing primer for each sample.

450 For the RNA-seq, RNA was extracted from 25 μl intact lysate (non-digested) using the Direct-
451 zol kit (Zymo Research) and stranded total RNA libraries were prepared using the TruSeq
452 Stranded Total RNA Human/Mouse/Rat kit (Illumina), following manufacturer's instructions.

453 Libraries were quantified and checked for quality using a Qubit fluorimeter and Bioanalyzer
454 (Agilent) and sequenced on a HiSeq 4000 sequencing system.

455

456 Slam-seq experimental protocol

457

458 SLAM-seq was performed at 60-70% confluency for DDX3X-mAID tagged HCT116. Media was
459 changed and fresh media with 100µM 4-thiouridine (4sU) was added to cells and changed
460 every 3 hours for 24 hours. 8 hours prior to collection, growth medium was aspirated and
461 replaced. Uridine chase was performed where cells were washed twice with 1X PBS and
462 incubated with media containing 10 mM uridine and DMSO or 100µM IAA for 0 or 8 hours to
463 induce degradation of DDX3X. At respective time points, cells were harvested followed by
464 total RNA extraction using TRIzol (Ambion) following the manufacturer's instructions
465 (SLAMseq Kinetics Kit – Catabolic Kinetics Module, Lexogen). Total RNA was alkylated by

466 iodoacetamide for 15 min and RNA was purified by ethanol precipitation. 200ng alkylated
467 RNA were used as input for generating 3'-end mRNA sequencing libraries using a
468 commercially available kit (QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina,
469 Lexogen).

470

471 ## Ribo-seq data analysis

472

473 Reads were stripped of their adapter, collapsed, and UMI sequences were removed. Clean
474 reads were then mapped to rRNA, tRNA, snoRNA and miRNA sequences using bowtie2[37] using
475 sequences retrieved from UCSC browser and aligning reads were discarded. Remaining reads
476 were mapped to the genome and transcriptome using STAR[38] v2.7.9a supplied with the
477 GENCODE v32 GTF file. STAR parameters were: *--outFilterMismatchNmax 3 --*
478 *outFilterMultimapNmax 50 --chimScoreSeparation 10 --chimScoreMin 20 --chimSegmentMin*
479 *15 --outFilterIntronMotifs RemoveNoncanonicalUnannotated --alignSJoverhangMin 500 --*
480 *outSAMmultNmax 1 --outMultimapperOrder Random.*

481

482 ## SLAM-seq data analysis

483

484 Reads were mapped to the genome and transcriptome using same RNA-seq parameters,
485 except for *--outFilterMismatchNmax 10*. Reads containing T>C mutations were extracted
486 from the BAM file using *GenomicAlignments* and *GenomicFiles* Bioconductor[39] packages.

487

488 ## RNA-seq data analysis

489

490 Reads were mapped to the genome and transcriptome using STAR with same Ribo-seq
491 parameters. Synthesis, processing, and degradation rates were obtained using INSPEcT[17]
492 v1.17, using default settings. Genes significantly changing in their dynamics at a p-value cutoff
493 of .05 were used for subsequent analysis.

494

495 ## Differential analysis

496

497 Unique counts on different genomic regions were obtained using *RiboseQC*[40]. 5' end coverage
498 values were inspected using Bioconductor[39] packages such as *GenomicFeatures*[41] and
499 *rtracklayer*[42]. *DESeq2*[43] was used to obtain RNA-seq, Ribo-seq, and TE regulation, as described
500 previously[7]: changes in translation efficiency were calculated using DESeq2 by using assay
501 type (RNA-seq or Ribo-seq) as an additional covariate. Translationally regulated genes were
502 defined using an FDR cutoff of 0.05 from a likelihood ratio test, using a reduced model without
503 the assay type covariate, e.g. assuming no difference between RNA-seq and Ribo-seq counts.

504 A similar strategy was used to define significant changes in DDX3X-mediated stability from
505 SLAM-seq: count tables with T>C reads were built and analyzed using labeling (4sU/DMSO)
506 and degron status (8h. vs DMSO) as the two variables of interest; regulation in stability was
507 defined using a reduced model without the degron type covariate, e.g. assuming no
508 difference between DMSO and degron activation.

509 Translationally regulated genes (as defined by Ribo-seq/RNA-seq) and stability regulated
510 genes (as defined by SLAM-seq) were defined using a p-value cutoff of .05.

511 For Figure 1D and 3D, the coordinate system was divided into 70 bins on each axis. GCcds
512 values (for Figure 3D), or Ribo-seq and RNA-seq fold changes between each time point and
513 the previous one (for Figure 1D) were averaged across genes in the same bin. Only mRNAs
514 with significant changes in translation efficiency at 48h post degron induction were
515 considered.

516

517 ## Random Forest and Lasso regression

518

519 The Random Forest regression was run using the *randomForest*[44] package with default
520 parameters. Lasso regression was performed on scaled variables using the *glmnet*[45] package.
521 While the entire feature table is available in Supplementary Table 2, a short description of the
522 input features follows:

523 TPM values using RNA-seq (in log scale). Baseline TE levels, defined as ratio of Ribo to RNA
524 reads. Baseline RNA mature levels, defined as length-normalized ratio of RNA-seq reads in
525 introns versus exons. GC content, length (in log scale) and Ribo-seq/RNA-seq density in: 5'
526 UTRs, a window of 25nt around start and stop codons, CDS regions, non-coding internal
527 exons, introns, and 3' UTRs. Codon frequencies. Measures of gene-specific codon optimality,
528 previously calculated from a recent study[18]. GC-content at first, second, or third codon
529 position.

530

531 Feature importance (measured by mean decrease in accuracy for the random forest model
532 and by the lasso coefficients) and correlation between predicted and measured test data
533 were calculated on a 5-fold cross-validation scheme.

534

535

536 ## Analysis of cell cycle - dependent mRNA dynamics

537

538 Estimated mRNA decay kinetics at cell cycle re-entry were deposited as supplementary files
539 of the original study[21]. Genes were partitioned cutting their GCcds values into 3 groups given
540 the low number of quantified genes (total n=220).

541

542 ## Cell cycle staging

543

544 To measure DNA replication and cell cycle stage, EdU (5-ethynyl-2´-deoxyuridine) was added
545 to cells at 10nM for 1.5 hrs before harvesting. 1 confluent well of a 6-well plate of HCT116
546 cells were harvested and processed as per manufacturer's instructions for the Click-iT™ Plus
547 EdU Alexa Fluor™ 647 Flow Cytometry Assay Kit (Thermo Fisher cat: C10634). Per
548 manufacturer's instructions, FxCycle Violet DNA content stain (Thermo Fisher cat: F10347)
549 was added after the Click-iT reaction at 1:1,000 dilution before quantifying on a BD LSR Dual
550 Fortessa flow cytometer. Alexa Fluor 647 was measured in the 670-30 Red C-A Channel and

551 FxCycle Violet Stain was measured in the 450-50 Violet F-A Channel. Analysis was performed
552 using FACS DIVA and FlowJo V10 (FlowJo, LLC) software.

553

554 ## 5'end coverage analysis

555

556 Computation on single-nucleotide coverage values was performed using *rtracklayer*[42]. For
557 each differential analysis, we extracted the most 250 stabilized and the most 250 degraded
558 genes ranking P-values from RNA-seq differential analysis. 1500 control RNAs were randomly
559 sampled from non-regulated genes, using p-values >.2 and TPM values > 3. Coverage values
560 were 0-1 (min/max) normalized and the first position at value >.15 was identified as coverage
561 starting position. In addition, a general coverage starting point was selected by pooling all
562 samples, and a window of 250nt around such position was used to calculate average coverage
563 values around the coverage start. Log2 fold change with respect to the control condition were
564 then calculated.

565 For degron data, starting position and log2fc coverage values were extracted and used as
566 input for linear regression. For coverage values, intercept was omitted, as the first value was
567 0. Beta coefficients were then extracted and compared between stabilized, degraded, and
568 control mRNAs.

569 For mouse Ddx3x cKO and ENCODE data, differences between starting position (knockdown
570 vs wt) and log2FC (knockdown vs wt) in coverage values were used to compare stabilized,
571 degraded and control mRNAs, bypassing the regression step (2 values were calculated, as only
572 wt or knockdown conditions were present).

573 ## TaqMan RT-PCR

574 DDX3X-mAID tagged HCT116 cells were plated in 6 well plates at 30-40% confluency. 24 hours
575 post plating 500 µM IAA or DMSO was added to cells with or without 200nM Actinomycin D
576 (ActD) for respective conditions. Total RNA was extracted from cells at 60-70% confluency
577 using Direct-zol kit (Zymo Research) at 0 and 24 hours post-ActD and IAA or DMSO treatment.
578 TaqMan probes for JUND, EIF2A, RACK1, LGALS1, and PFN1 were predesigned and purchased
579 from ThermoFisher Scientific. Riboseq degraded (EIF2A) or stabilized genes (JUND) were
580 conjugated with FAM dye while control genes RACK1, LGALS1, and PFN1 were conjugated
581 with VIC dye. For the TaqMan real-time quantitative PCR amplification reactions, we
582 employed an Applied Biosystems QuantStudio 6 Real-Time PCR System instrument. Real-time
583 PCR was conducted using TaqMan Fast Virus 1-Step Master Mix from Applied Biosystems in
584 384-well plates, following the manufacturer's protocol. Each well contained either the genes
585 subject to riboseq degradation gene (EIF2A) or stabilization gene (JUND) along with control
586 genes (RACK1, LGALS1, or PFN1). All reactions were conducted in triplicate. Thermal cycling
587 conditions adhered to the manufacturer's recommended standard protocol. The
588 quantification of the target input amount was determined using the cycle threshold (CT)
589 value, which corresponds to the point at which the PCR amplification plot crosses the
590 threshold. Expression of ribose degraded and stabilized genes were normalized to each
591 control genes respectively.

592 Gene Species Chromosome Location Assay ID Dye

593 RACK1 HUMAN Chr.5: 181236928 - 181243906 on Build GRCh38 Hs00272002_m1 VIC-MGB

594 LGALS1 HUMAN Chr.22: 37675606 - 37679802 on Build GRCh38 Hs00355202_m1 VIC-MGB

595 PFN1 HUMAN Chr.17: 4945650 - 4949088 on Build GRCh38 Hs07291746_gH VIC-MGB

596 JUND HUMAN Chr.19: 18279694 - 18281656 on Build GRCh38 Hs04187679_s1 FAM-MGB

597 EIF2A HUMAN Chr.3: 150546678 - 150586016 on Build GRCh38 Hs00230684_m1 FAM-MGB

598 Details of TaqMan® real-time PCR assays obtained from ThermoFisher Scientific.

599

600 Data and code availability

601

602 Raw sequencing data for Ribo-seq, RNA-seq and SLAM-seq can be found under GEO accession
603 GSE218433, with token "ujmtquoulnirpgx". Encode accession numbers can be found in
604 Supplementary Table 3. *Ddx3x* knockout Ribo-seq and RNA-seq were analyzed from accession
605 number GSE203078, processed data can be found in Supplementary Table 4. Code to
606 reproduce all figures, together with processed data, can be found at
607 https://github.com/calviellolab/DDX3X_GC_paper .

608

609 Author contributions

610

611 Z.J. performed SLAM-seq and provided support for analysis of stabilized mRNAs. A.X.
612 performed cell cycle analysis along the degron time course. S.V. performed Ribo-seq and RNA-
613 seq. M.H. and D.L.S. provided crucial input with the *Ddx3x* mouse model dataset. S.N.F.
614 supervised the initial part of the project, secured funding, and provided significant input,
615 together with all authors, in writing the manuscript. L.C. conceived the project, performed all

616 computational analyses, prepared the figures, and wrote the manuscript, with support from
617 F.D. and all Authors.
618

619 Acknowledgements

620

621 L.C. wants to thank Piero Angela (1928-2022) for inspiring him, and countless other kids in
622 Italy, to pursue a career in academic science, in constant awe of the beautiful mechanisms
623 shaping the natural world. This work was supported by the National Institutes of Health
624 DP2GM132932 and R01NS120667 (to S.N.F.). S.N.F. is a Pew Scholar in the Biomedical
625 Sciences, supported by The Pew Charitable Trusts. ZJ was supported by the UCSF Moritz-
626 Heyman Discovery Fellow Program. AX was supported by NIH F30 Ruth L Kirschstein National
627 Research Service Award HD110250. ZJ and AX were supported by the UCSF Medical Scientist
628 Training Program (T32GM007618). Flow cytometry data was generated at the UCSF Parnassus
629 Flow CoLab (RRID:SCR_018206) and sequencing was performed at the UCSF CAT, supported
630 by UCSF PBBR, RRP IMIA, and NIH 1S10OD028511-01 grants. Mouse image in Figure 6A used
631 with permission from doi.org/10.5281/zenodo.3925903 . F.D. is a PhD student within the
632 European School of Molecular Medicine (SEMM).

633 Supplementary information

634

635

636 Supplementary Figures are provided separately.

637

638

639 **Supplementary Table 1**. Read mapping statistics for the Ribo-seq RNA-seq DDX3X time course

640 dataset.

641

642 **Supplementary Table 2**. Input to the Random Forest model for the DD3X3 time course

643 dataset.

644

645 **Supplementary Table 3**. Accession codes for the analyzed ENCODE datasets, with information

646 for each differential analysis. Multiple accession can be technical replicate of a biological

647 replicate.

648

649 **Supplementary Table 4**. Input to the Random Forest model for the cKO *Ddx3x* mouse dataset.

650

651

652

653

654

655

656 ## References

657

658 1. Shoemaker, C. J. & Green, R. Translation drives mRNA quality control. *Nat Struct Mol*
659 *Biol* **19**, 594–601 (2012).

660 2. Ingolia, N. T., Brar, G. a, Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome
661 profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-
662 protected mRNA fragments. *Nat Protoc* **7**, 1534–1550 (2012).

663 3. Herzog, V. A. *et al.* Thiol-linked alkylation of RNA to assess expression dynamics.
664 *Nature Methods 2017 14:12* **14**, 1198–1204 (2017).

665 4. Hafner, M. *et al.* CLIP and complementary methods. *Nature Reviews Methods Primers*
666 *2021 1:1* **1**, 1–23 (2021).

667 5. Sharma, D. & Jankowsky, E. The Ded1/DDX3 subfamily of DEAD-box RNA helicases.
668 *Critical Reviews in Biochemistry and Molecular Biology* Preprint at
669 https://doi.org/10.3109/10409238.2014.931339 (2014).

670 6. Oh, S. *et al.* Medulloblastoma-associated DDX3 variant selectively alters the
671 translational response to stress. *Oncotarget* (2016) doi:10.18632/oncotarget.8612.

672 7. Calviello, L. *et al.* DDX3 depletion represses translation of mRNAs with complex 5'
673 UTRs. *Nucleic Acids Res* **49**, 5336–5350 (2021).

674 8. Chang, Y.-F. F., Imam, J. S. & Wilkinson, M. F. The nonsense-mediated decay RNA
675 surveillance pathway. *Annu Rev Biochem* **76**, 51–74 (2007).

676 9. D'Orazio, K. N. *et al.* The endonuclease Cue2 cleaves mRNAs at stalled ribosomes
677 during No Go Decay. *Elife* **8**, (2019).

678 10. Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome profiling shows that miR-430
679 reduces translation before causing mRNA decay in zebrafish. *Science* **336**, 233–7
680 (2012).

681 11. Lennox, A. L. *et al.* Pathogenic DDX3X Mutations Impair RNA Metabolism and
682 Neurogenesis during Fetal Cortical Development. *Neuron* **106**, 404-420.e8 (2020).

683 12. Ingolia, N. T. Ribosome Footprint Profiling of Translation throughout the Genome. *Cell*
684 **165**, 22–33 (2016).

685 13. Hoye, M. L. *et al.* Aberrant cortical development is driven by impaired cell cycle and
686 translational control in a DDX3X syndrome model. *Elife* **11**, (2022).

687 14. Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-
688 binding proteins. *Nature 2020 583:7818* **583**, 711–719 (2020).

689 15. Hon, C. C. & Carninci, P. Expanded ENCODE delivers invaluable genomic encyclopedia.
690 *Nature 2021 583:7818* **583**, 685–686 (2020).

691 16. Venkataramanan, S., Gadek, M., Calviello, L., Wilkins, K. & Floor, S. N. DDX3X and
692 DDX3Y are redundant in protein synthesis. *RNA* **27**, rna.078926.121 (2021).

693 17. De Pretis, S. *et al.* INSPEcT: a computational tool to infer mRNA synthesis, processing
694 and degradation dynamics from RNA- and 4sU-seq time course experiments.
695 *Bioinformatics* **31**, 2829–2835 (2015).

696 18. Medina-Muñoz, S. G. *et al.* Crosstalk between codon optimality and cis-regulatory
697 elements dictates mRNA stability. *Genome Biol* **22**, 1–23 (2021).

698 19. Courel, M. *et al.* Gc content shapes mRNA storage and decay in human cells. *Elife* **8**,
699 (2019).

700 20. Hia, F. *et al.* Codon bias confers stability to human mRNAs. *EMBO Rep* **20**, e48220
701 (2019).

702 21. Krenning, L., Sonneveld, S. & Tanenbaum, M. Time-resolved single-cell sequencing
703 identifies multiple waves of mRNA decay during the mitosis-to-G1 phase transition.
704 *Elife* **11**, (2022).

705 22. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat*
706 *Rev Genet* **15**, 829–845 (2014).

707 23. Avsec, Ž. *et al.* Effective gene expression prediction from sequence by integrating
708 long-range interactions. *Nature Methods 2021 18:10* **18**, 1196–1203 (2021).

709 24. Thomas, A. *et al.* RBM33 directs the nuclear export of transcripts containing GC-rich
710 elements. *Genes Dev* **36**, 550–565 (2022).

711 25. Chothani, S. P. *et al.* A high-resolution map of human RNA translation. *Mol Cell* **82**,
712 2885-2899.e8 (2022).

713 26. Zhao, W. *et al.* POSTAR3: an updated platform for exploring post-transcriptional
714 regulation coordinated by RNA-binding proteins. *Nucleic Acids Res* **50**, D287–D294
715 (2022).

716 27. Van Nostrand, E. L. *et al.* Principles of RNA processing from analysis of enhanced CLIP
717 maps for 150 RNA binding proteins. *Genome Biol* **21**, 1–26 (2020).

718 28. Bahar Halpern, K. *et al.* Nuclear Retention of mRNA in Mammalian Tissues. *Cell Rep*
719 **13**, 2653–2662 (2015).

720 29. Kotov, A. A., Olenkina, O. M., Kibanov, M. V. & Olenina, L. V. RNA helicase Belle
721 (DDX3) is essential for male germline stem cell maintenance and division in
722 Drosophila. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1863**,
723 1093–1105 (2016).

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

726 31. Clemm von Hohenberg, K. *et al.* Cyclin B/CDK1 and Cyclin A/CDK2 phosphorylate
727 DENR to promote mitotic protein translation and faithful cell division. *Nature*
728 *Communications 2022 13:1* **13**, 1–14 (2022).

729 32. Tanenbaum, M. E., Stern-Ginossar, N., Weissman, J. S. & Vale, R. D. Regulation of
730 mRNA translation during mitosis. *Elife* **4**, (2015).

731 33. Hoye, M. L. & Silver, D. L. Decoding mixed messages in the developing cortex:
732 translational regulation of neural progenitor fate. *Curr Opin Neurobiol* **66**, 93–102
733 (2021).

734 34. Harnett, D. *et al.* A critical period of translational control during brain development at
735 codon resolution. *Nature Structural & Molecular Biology 2022 29:12* **29**, 1277–1290
736 (2022).

737 35. Pilaz, L. J. *et al.* Prolonged Mitosis of Neural Progenitors Alters Cell Fate in the
738 Developing Brain. *Neuron* **89**, 83–99 (2016).

739 36. Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The
740 ribosome profiling strategy for monitoring translation in vivo by deep sequencing of
741 ribosome-protected mRNA fragments. *Nat Protoc* **7**, 1534–50 (2012).

742 37. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat*
743 *Methods* (2012) doi:10.1038/nmeth.1923.

744 38. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
745 (2013).

746 39. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor.
747 *Nat Methods* **12**, 115–121 (2015).

748 40. Calviello, L., Sydow, D., Harnett, D. & Ohler, U. Ribo-seQC: comprehensive analysis of
749 cytoplasmic and organellar ribosome profiling data. doi:10.1101/601468.

750 41. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS*
751 *Comput Biol* **9**, e1003118 (2013).

752 42. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing
753 with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).

754 43. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
755 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

756 44. Wiener, A. L. and M. Classification and Regression by randomForest. *R News* **2**, 18–22
757 (2002).

45. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1 (2010).