# Genetic Rescue of Fragile X Syndrome Links FMRP Deficiency to Codon Optimality-Dependent RNA Destabilization

Huan Shu[1], Elisa Donnard[2], Botao Liu[1], and Joel D. Richter[1]

[1]Program in Molecular Medicine

[2]Bioinformatics and Integrative Biology

University of Massachusetts Medical School

Worcester, MA 01605

## Abstract

Fragile X syndrome (FXS) is caused by inactivation of *FMR1* gene and loss of its encoded product the RNA binding protein FMRP, which generally represses translation of its target transcripts in the brain. In mouse models of FXS (i.e., *Fmr1* knockout animals; *Fmr1* KO), deletion of *Cpeb1,* which encodes a translational activator, mitigates nearly all pathophysiologies associated with the disorder. Here we reveal unexpected wide-spread dys-regulation of RNA abundance in *Fmr1* KO brain cortex and its rescue to normal levels in *Fmr1/Cpeb1* double KO mice. Alteration and restoration of RNA levels are the dominant molecular events that drive the observed dys-regulation and rescue of translation as measured by whole transcriptome ribosome occupany in the brain. The RNAs down-regulated and rescued in these animal models are highly enriched for FMRP binding targets and have an optimal codon bias that would predict their stability in wild type and possible instability in FMRP knock-out brain. Indeed, whole transcriptome analysis of RNA metabolic rates demonstrates a codon optimality-dependent elevation of RNA destruction in FMRP knock-out cortical neurons. This elevated RNA destruction leads to a massive reshuffling of the identities of stabilizing versus destabilizing codons in neurons upon loss of FMRP. Our results show a widespread RNA instability in FXS, which results from the uncoupling of codon optimality, ribosome occupancy, and RNA degradation mechanisms. Re-establishment of the linkage among these events is likely required by the genetic rescue of the disorder.

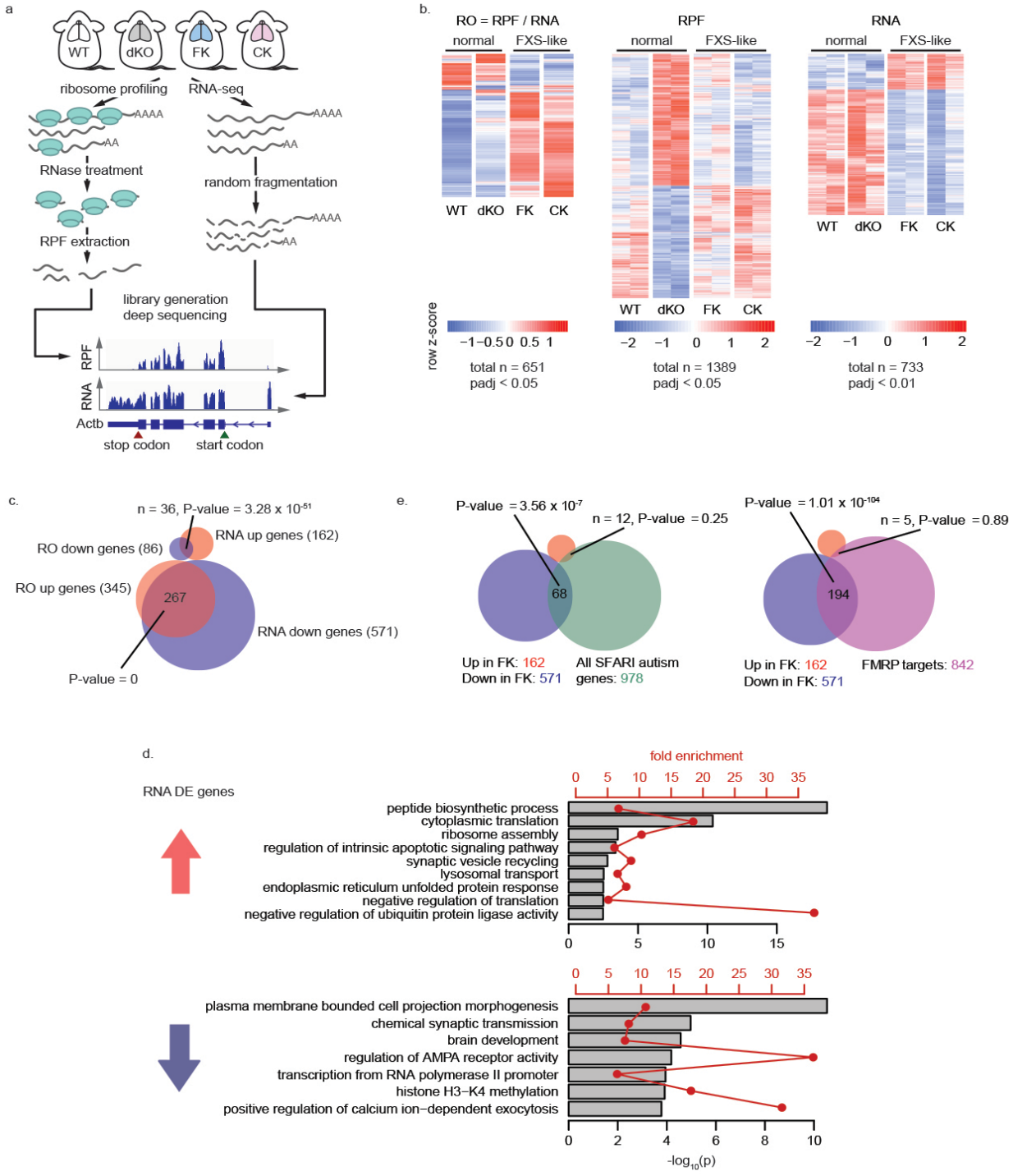

## Introduction

FXS is the most common form of inherited intellectual disability that is caused by a single gene mutation[1]. In addition to mild to severe intellectual disability, individuals with FXS often have increased susceptibility to seizures, autism-like behaviors, developmental delays, among other symptoms[1]. FXS is caused by the expansion of a CGG trinucleotide repeat in the 5'UTR of

*FMR1*, which results in transcriptional silencing of the gene and subsequent loss of its protein product, the fragile X mental retardation protein (FMRP)[2]. In the absence of FMRP, protein synthesis in the hippocampus (where most activities of FMRP have been studied) is elevated by ~20%[3,4], leading to the general hypothesis that this protein represses translation, possibly in dendritic spines as well as other regions of neurons. Stimulation of post-synaptic metabotropic glutamate receptors (mGluRs) results in a form of synaptic plasticity known as long-term depression (mGluR-LTD), which while normally protein synthesis-dependent, becomes protein-synthesis independent in the absence of FMRP. This causes aberrant synaptic plasticity (i.e., exaggerated LTD) and impaired learning and memory[5]. In mouse brain, FMRP binds mostly to coding regions of ~850 to 1000 mRNAs[6,7], and co-sediments with polyribosomes[6]. Because the ribosomes associated with many of these mRNAs are resistant to release by puromycin treatment and because these ribosomes translocate at faster rates in *Fmr1* KO brain compared to WT[6,8], FMRP is thought to repress translation by impedeing ribosme transit[6].

Most genetic or pharmacologic rescue paradigms of FXS in mouse models display re-establishment of disrupted translational homeostasis[3,4,8–10]. We have previously shown that depletion of CPEB1, which co-localizes and co-immunoprecipitates with FMRP and activates translation in response to synaptic stimulation[11], mitigates nearly all pathophysiologies associated with FXS in *Fmr1/Cpeb1* double knockout (dKO) mice, including the exaggerated mGluR-LTD and elevated protein synthesis[8]. However, the identities of the mRNAs whose translation is disrupted in the absence of FMRP but is restored in the dKO is unknown. Here, we have used ribosome profiling and RNA-seq to investigate the mRNAs whose expression is dys-regulated in the *Fmr1* KO cortex and rescued in the dKO animal. To our surprise, the apparent dys-regulation and rescue of translational activity (i.e., ribosome occupancy) in our FMRP- and CPEB1-depletion “disease” and rescue paradigm is largely driven by the dys-regulation and rescue at the RNA stability level. The RNAs that are up-regulated in *Fmr1* KO cortex are

64 enriched for those that encode ribosomal components and translational factors, and may
65 partially explain the excessive protein production in the *Fmr1* KO brain. Strikingly, the down-
66 regulated mRNAs in the *Fmr1* KO, which are enriched for those that encode factors involved in
67 neuronal and synaptic functions, are highly enriched for FMRP binding targets[6] and have a



strong bias for optimal codons (i.e., codons that are favored over other synonymous codons in highly expressed mRNAs; such RNAs tend to be stable[12–15]), suggesting that their levels are controlled by a post-transcriptional mechanism. These observations imply that in the cortex, FMRP directly or indirectly regulates RNA stability. Indeed, RNA metabolic profiling by 5-ethynyl uridine incorporation and whole transcriptome sequencing revealed wide-spread dysregulation in RNA degradation rates in *Fmr1*-KO cortical neurons while synthesis and processing rates remained substantially unchanged. We identified ~700 mRNAs that degrade significantly faster in *Fmr1* KO cortex compared to WT; those that favor optimal codons were particularly affected. This wide-spread codon-dependent dys-regulation in RNA degradation involves a massive reshuffle of the identities of stabilizing vs destabilizing codons, which is unlinked from codon bias. These results indicate that a primary consequence of FMRP depletion from the brain transcriptome is dys-regulated mRNA stability by uncoupling codon bias from the RNA destruction machinery. This uncoupling may be a general mechanism that underlies the FXS, and restoring the RNA stability landscape could be a key to ameliorating the disorder as implied by the restored RNA levels in the dKO brain.

**Figure 1: RNA and not ribosome footprint levels is the dominant molecular signature in Fragile X and in a CPEB1-deficient rescue paradigm. a,** Illustration of the experimental pipeline of ribosome profiling and RNA-seq for WT, *Fmr1*-deficient (FK), *CPEB1*-deficient (CK), *Fmr1/CPEB1* deficient (dKO) mouse brain cortices. RPF: ribosome protected fragments. **b,** Heatmaps showing genes identified having differential ribosome occupancies (RO; left) and RPF (middle) between any two genotypes of the four genotypes noted above, and genes expressing differential steady state RNA levels between the normal and FXS-like groups (right). Heatmap heights are proportional to the numbers of differential genes identified in each comparison. Red and blue shades show high or low z-scores calculated for each gene (row) across all samples. For RPF and RNA heatmaps both replicates are plotted separately for each genotype, and for RO a statistical summary of the two replicates were calculated using Xtail[17] and plotted. **c,** Venn diagram showing the overlap of genes up or down regulated at the ribosome occupancy (RO) level in FK compared to WT and at the RNA level in FXS-like group compared to the normal group. Numbers of genes in each group and in each overlap as well as p-values of enrichment (hypergeometric test, upper tail) are indicated. **d,** Representative Gene Ontology (GO) terms enriched for genes upregulated (upper) or down regulated (lower) at the RNA level in the FXS-like group. Grey bars and red point-and-lines show the $-\log_{10}$(P value) and fold enrichment of each of these GO terms, respectively. See **Tables S1 and S2** for full lists of enriched GO terms. **e,** Venn diagrams showing the overlap between the DE genes at the RNA level with all SFARI autism risk genes[20] (upper) and FMRP binding targets[6] (lower). Numbers of genes in each group and in each overlap as well as p-values of enrichment (hypergeometric test, upper tail) are indicated.

# Results

## RNA dys-regulation and recovery correlate with Fragile X Syndrome and genetic rescue in mouse models

To identify mRNAs that are translationally dys-regulated in the *Fmr1*-KO mouse cortex and that are rescued in the *Fmr1-Cpeb1* dKO cortex[8], we performed ribosome profiling[16] and RNA-seq in this brain tissue from WT, *Fmr1* and *Cpeb1* single KO as well as dKO animals (**Fig 1a**). Ribosome occupancy (translational efficiency), defined as ribosome protected fragments (RPFs) normalized to mRNA levels, is a measure of translational activity[16] and in this sense serves as a proxy for protein synthesis. Accumulating evidence suggests that one mechanism whereby FMRP inhibits translation is by stalling ribosome transit[6,8] and indeed there is a moderately (10-15%) higher rate of protein synthesis in FMRP-deficient brain[3,4,9]. Using Xtail[17], an algorithm that tests for differential ribosome occupancies (DROs) between samples, we identified 651 genes with DROs among the four genotypes (FDR < 0.05; **Fig 1b, left**). Consistent with FMRP acting as a translation repressor, 345 out of 431 genes (80%) with DRO between *Fmr1* KO (FK) and WT were up-regulated. Importantly, 425 of these DROs were rescued in the dKO cortex. Unexpectedly, more than 50% of genes with DRO in *Cpeb1* KO (CK) (204 out of 359) also had DROs in FK, and were changed in the same direction (i.e., up or down). These molecular data are consistent with previous observations such as dendritic spine number and morphology, which are similarly aberrant in the two single KOs but rescued to normal in dKO animals. In this same vein, mGluR-LTD is enhanced in both of the single KOs but restored to WT in levels in dKO animals[8]. Because of the molecular similarities between WT and dKO, and between FK and CK, we henceforth refer to these two groups as “normal” and “FXS-like.”

To determine the underlying cause of DRO among the genotypes, we analyzed our RPF and RNA-seq data separately. Surprisingly, most RPFs were indistinguishable among FK, CK and

WT. Only 23 and 21 RPFs were significantly different (padj < 0.05) between FK and WT and between CK and WT. Conversely, the dKO was the most different from WT with 410 and 333 RPFs that were significantly higher or lower (**Fig 1b, middle**).

In contrast, the RNA-seq heatmap displayed a reverse mirror image of the DRO heatmap (**Fig S1a, Fig 1b right).** Compared to WT, the expression of 50 genes was dys-regulated in FK (padj < 0.05; 10 up-regulated, 40 down-regulated), 145 in CK (padj < 0.05; 13 up-regulated, 132 down-regulated), but only 2 in dKO (padj < 0.05; *Cpeb1* and *Fmr1*) (**Fig S1a**). The differentially expressed (DE) genes in FK and CK were largely identical. Among the 10 and 13 genes up-regulated in FK and CK, 7 overlap ($p = 8.72 \times 10^{-25}$, hypergeometric test, upper tail); among the 40 and 132 genes down-regulated, 35 overlap ($p = 3.91 \times 10^{-72}$, hypergeometric test, upper tail). Because the transcriptome profiles in FK and CK are so similar as are the WT and dKO (**Fig S1a**), we performed an unsupervised hierarchical clustering to test for sample to sample similarities (**Fig S1b).** FK and CK formed one cluster while WT and dKO formed another, validating the “FXS-like” vs “normal” grouping at the RNA level.

Having validated the grouping, we tested for DE genes in the FXS-like group (FK and CK) relative to the normal group (WT and dKO). The DE genes identified between the groups are changed the same way (i.e., up or down) in the single KOs and are rescued in the dKO to WT levels. We identified 733 genes dys-regulated in the FXS-like group (padj < 0.01), 162 (22.1%) up-regulated and 571 (77.9%) down-regulated (**Fig 1b right)**. Strikingly, over 77% of the genes with up-regulated ROs in FK vs WT (267 out of 345) were significantly reduced at the RNA level in the FXS-like group (p-value = 0, hypergeometric test, upper tail). Similarly, 42% of the genes with down-regulated ROs in FK (36 out of 86) were significantly increased at the RNA level (p-value = $3.28 \times 10^{-51}$, hypergeometric test, upper tail) (**Fig 1c**). We conclude that the observed dys-regulation and rescue ostensibly occurring at the translational level is largely driven by dys-regulation and rescue at the RNA level.

Gene Ontology (GO) analysis shows that many up-regulated RNAs have protein synthetic functions including ribosome biogenesis, translation, and protein folding, while the down-regulated RNAs have cell projection, synaptic transmission, as well as transcription and chromatin functions (**Fig 1d; Tables S1, S2).** Several important points come from this analysis. First, the down-regulation of many mRNAs may be "buffered" or compensated at the translational level by the up-regulation of other mRNAs that promote protein synthesis. Hence this could explain the net increase in protein output in FXS brain[3,4,9]. Indeed, we do observe increased ribosome occupancy of these RNAs (**Fig 1c**). Buffering in FMRP-deficient cells has been observed previously[18]. Second, we find that FMRP regulates the levels of mRNAs that encode chromatin modifying factors, which is reminiscent to other observations showing that FMRP controls the synthesis of epigenetic regulators in young neurons, albeit at the translational level[19]. Third, the brain and neuron components enriched in the GO terms of the down-regulated genes reflect the neural dysfunction that occurs in FXS. Indeed, the down-regulated genes are also significantly enriched for autism genes as compiled by SFARI[20] (**Fig 1e left**; $p = 3.56 \times 10^{-7}$, hypergeometric test, upper tail).

We examined whether FMRP might have a direct effect on the steady state levels of the brain transcriptome. Significantly, 199 genes out of 733 DE mRNAs are bound (i.e., by CLIP, UV Crosslink and Immunoprecipitation) by FMRP[6], and 194 of these were down-regulated, which is 34% of all the down-regulated genes in the FXS-like group ($p = 1.01 \times 10^{-104}$, hypergeometric test, upper tail; **Fig 1e right**). This result indicates that loss of FMRP may have a direct impact on the levels of a subset of the transcriptome important for the proper brain functions.

We found that the genes down-regulated in this study, as well as FMRP target mRNAs[6], were also reduced in other studies that examined various FMRP-deficient cell and tissue types from mouse to human (**Fig S1c**). Given that this dys-regulation is widespread in other FXS paradigms and that the RNA rescue parallels phenotypic rescue[8], it is axiomatic that

157 investigating RNA dys-regulation is fundamental to understanding and perhaps mitigating the
158 disorder.

159 **Down-regulated mRNAs have a strong bias for optimal codons**

160 Because of its strong cytoplasmic localization[21] and likely direct control of steady state levels of
161 its binding target, we surmised that FMRP would regulate mRNA stability[22]. How FMRP could
162 stabilize target RNAs is suggested by its role in stalling ribosomes during translation
163 elongation[6,8,23] (**Fig 2a**). In yeast, Dhh1p (DDX6) destabilizes mRNAs with low codon optimality
164 by sensing their slow ribosome decoding rate[24]. Codon optimality, a measure of the balance
165 between the demand and supply of charged tRNAs[25], is a major determinant of mRNA stability
166 from yeast to vertebrates[12–15,24,26]. Generally, mRNAs with more optimal codons (presumably
167 with faster decoding rates) are more stable than mRNAs using less optimal codons, connecting
168 translation regulation to mRNA stability. Consequently, we calculated the codon Adaptation
169 Index (cAI)[27] from our WT mouse cortex transcriptome data, which describes the codon usage
170 bias among synonymous codons for the highly expressed genes. We then derived the
171 geometric mean of the cAI of each codon in each gene, which is referred to as the gene cAI
172 score (see Materials and Methods). We considered the codon cAI score as a proxy of codon
173 optimality and the gene cAI score as a predictor of mRNA stability in WT mouse cortex; high
174 gene cAI scores predict stable mRNAs. In WT cortex, the transcripts have gene cAI score
175 ranging from 0.62(*Gm14431*) to 0.95 (*Rpl41*) (0.77 ± 0.04, mean ± S.D.).

176 Surprisingly, the RNA down-regulated genes (**Fig 2b**, blue) were significantly more optimal than
177 the overall transcriptome ($p = 1.51 \times 10^{-92}$, Wilcoxon test, two tail), and the RNA up-regulated
178 genes (red, **Fig 2b**) were significantly less optimal ($p = 5.80 \times 10^{-17}$, Wilcoxon test, two tail).
179 Indeed, the down-regulated genes were among the most optimal (gene cAI of 0.8 ± 0.03, mean
180 ± S.D.) in the transcriptome, while the up-regulated genes among the least optimal (cAI 0.74 ±
181 0.04, mean ± S.D.) (**Fig S2a,** left). These cAI scores of the DE genes are not mere reflections of

their transcript levels (**Fig S2a,** right); both gene groups were highly expressed. These values predict that the mRNAs down-regulated in FK cortex would be among the most stable in the WT cortex, while the up-regulated mRNAs the least stable.

We grouped the detectable transcriptome into 10 equal sized bins of increasing gene cAI score and examined whether there was a global correlation with the change of the mRNAs in FK relative to WT for each bin (**Fig 2c,** left). We observed a global depletion of high relative to low cAI score RNAs upon loss of FMRP (**Fig 2c,** left). Globally, the log2FC ($log_2$ Fold Change) of RNA in the FXS-like group relative to the normal group has a strong negatively correlation with gene cAI scores (**Fig S2b**, left; Pearson's correlation coefficient = -0.34). Other transcript features that are often associated with RNA stability regulation[28], including coding sequencing (CDS) length, 5' and 3' UTR length, also correlated with mRNA level changes albeit not as strongly, except for coding sequence GC content (**Fig S2b**), which is a known confounding factor with codon optimality[29]. Given that FMRP target mRNAs strongly overlap with the down-regulated genes (**Fig 1d**), they are, not surprisingly, reduced in all cAI bins. However, the FMRP targets with

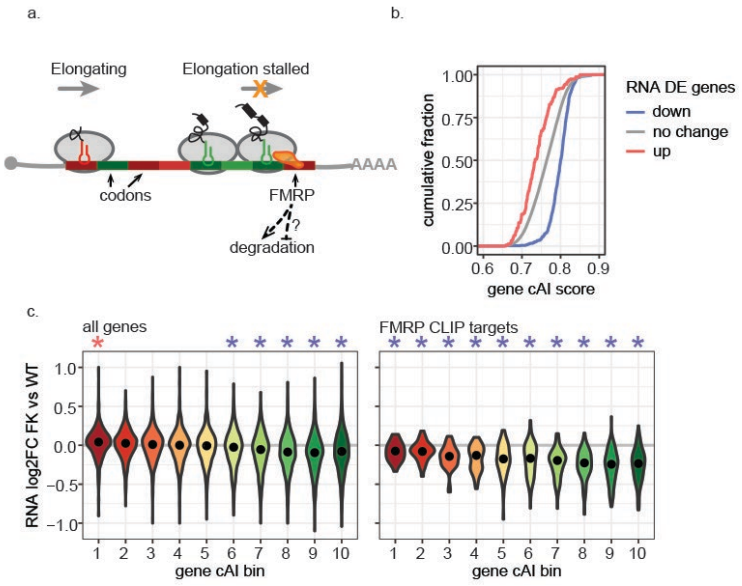



**Figure 2: RNA depletion upon loss of FMRP is a function of codon optimality. a,** Illustration of the links between FMRP, ribosomes, mRNA and codons. **b,** ECDF (empirical cumulative distribution function) plot of gene cAI scores for DE genes at the RNA level in the FXS-like group. **c,** All detectable protein coding genes were grouped into 10 equal bins based on their gene cAI scores. Bin 1 contains genes with gene cAI scores of the lowest quantile and bin 10 contains genes of the highest quantile. Violin plots show the RNA log2FoldChange (log2FC) in FK relative to WT of each bin for all expressed genes (left) and for only the FMRP CLIP targets[6] (right). The point in each violin denotes the median of the bin. Star indicates the median of the bin being greater (red) or less (blue) than 0 with a p-value < 0.01 (Wilcoxon test, one tail).

207 higher cAI scores are even more reduced than those with lower cAI scores (**Fig 2c**, right; **Fig**
208 **S2c**). These results show that loss of FMRP in mouse cortex leads to depletion of its target
209 mRNAs as well as mRNAs showing higher codon optimality. These results also predict a global
210 trend of destabilization of FMRP target mRNAs as well as stable mRNAs in FK brain cortex.

211 **RNA metabolism profiling reveals major disruption in RNA stability in FMRP-deficient**
212 **neurons**

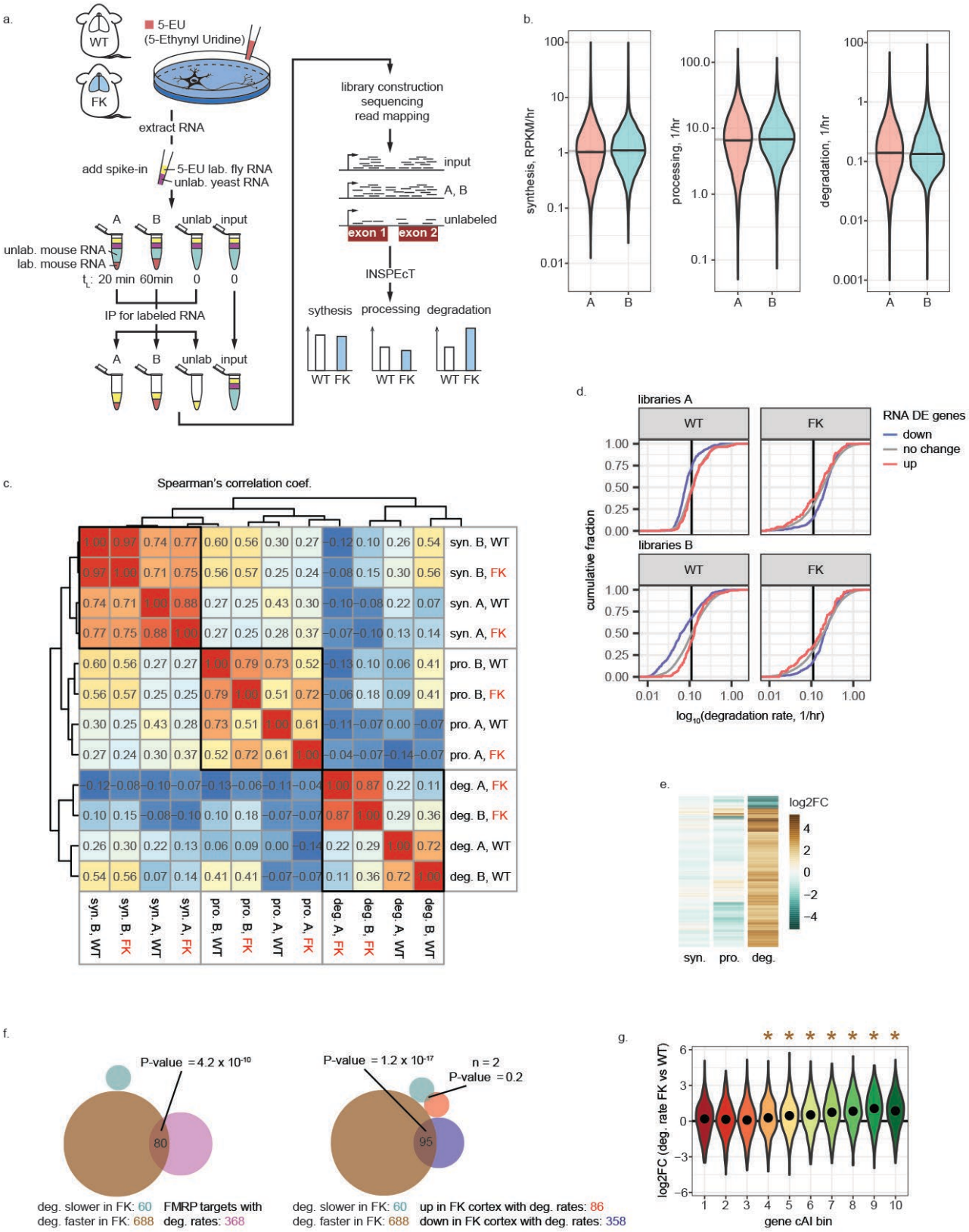
213 To determine whether loss of FMRP destabilizes mRNA, we incubated WT and FK mouse
214 cortical neurons (14 DIV) with 5-ethynyl uridine (5EU) for 0 (i.e., unlabeled control, or "unlab"),
215 20 (library A), or 60 min (library B), after which the RNA was "clicked" to biotin and purified by
216 streptavidin chromatography. The RNA was mixed with 5EU-labeled fly RNA and unlabeled
217 yeast RNA as a control and sequenced together with total unenriched RNAs as input samples
218 (**Fig 3a**). The spike-in RNAs for the libraries were used as quality control measures, showing
219 that the WT and FK libraries were of equal quality (**Fig S3a-c**). After filtering (**Fig S3d**), we
220 calculated RNA metabolism rates (synthesis, processing and degradation rates) by comparing
221 nascent and mature RNA concentrations in the 5EU-labeled and input total RNA libraries using
222 INSPEcT[30,31]. We obtained metabolism rate information for 8590 RNAs, which include 412
223 FMRP target mRNAs. The rates follow lognormal distributions with medians of 1.12 and 1.04
224 RPKM/hr, 6.84 and 6.60 $hr^{-1}$, and 0.13 and 0.14 $hr^{-1}$ for synthesis, processing, and degradation
225 for libraries A and B, respectively (**Fig 3b**), demonstrating the reproducibility of the assay.

226 We calculated Spearman's correlation coefficients for all three metabolism rates per genotypes
227 for both libraries (**Fig 3c**). For synthesis, processing, and degradation rates, we observed
228 decreasing correlation coefficients between WT and FK. For synthesis rates, WT and FK cluster
229 together for the same labeling parameter (library A or B), indicating that there is little genotype
230 difference. For libraries A and B, the correlation coefficients were 0.97 and 0.88 between WT
231 and FK, again demonstrating that the synthesis rates between the 2 genotypes are similar. For

processing rates, the two genotypes were also similar despite slightly lower Spearman's correlation coefficients between WT and FK (0.79 and 0.61 for libraries A and B). Strikingly, the correlation coefficients of degradation rates between WT and FK were substantially lower (0.22 and 0.36 for libraries A and B), indicating that there is a major difference in degradation between genotypes. The Spearman's correlation coefficients between libraries A and B for each genotype (0.87 and 0.72, respectively) indicates high reproducibility. Therefore, the degradation rates for the four libraries are separated by genotype (**Fig 3c**), demonstrating that RNA stability in FK neurons is disrupted.

We determined whether the RNA DE genes identified in the FXS-like group (**Fig 1b, right**) have altered RNA degradation rates. As predicted by gene cAI scores in WT cortical tissue (**Fig 2b**) and cultured cortical neurons (**Fig S3e**), the down-regulated RNAs (blue) are among those with the lowest degradation rates, i.e., the most stable (**Fig 3d, left**). On the contrary, in FK neurons, the down-regulated RNAs degrade significantly faster than the transcriptome in general (**Fig 3d,** right; $p = 0.00029$ and 0.026 for libraries A and B respectively; Wilcoxon test, one tail). The up-regulated RNAs (red) do not show a significant change in degradation rate (**Fig 3d**). Therefore, many mRNAs with optimal codons that are stable in WT cortical neurons become unstable in FK neurons.

To perform gene level comparisons of RNA metabolism rates, we normalized the values between libraries A and B for WT and FK (**Fig S3f**). At an adjusted p-value cut-off of 0.01, we identified no RNA with different synthesis or processing rate, but 748 RNAs with different degradation rates, of which 688 (92%) degraded faster in FK compared to WT (**Fig 3d, Fig S3g**). Significantly, the RNAs that degraded faster in FK were highly enriched for FMRP targets and down-regulated RNAs in the cortex (**Fig 3f**), showing that faster degradation is mostly responsible for the reduced RNA levels in FK brain. We also determined the influence of codon optimality on global mRNA degradation upon the loss of FMRP. Consistent with the high gene




257

**Figure 3: Codon-dependent mRNA destabilization in FK neurons explains the steady state RNA depletion. a,** Illustration of the experimental pipeline of RNA metabolism profiling for WT and FK neurons. Unlab./lab.: unlabeled/labeled. $t_L$: time labeled. **b,** Violin plots for synthesis, processing and degradation rates estimated for libraries A and B. Each violin contains data from both WT and FK. Black horizontal lines denote median of each violin. Grey horizontal lines denote the median of both violins. **c,** Heatmap of Spearman's correlation coefficients between synthesis (syn.), processing (pro.), and degradation (deg.) rates estimated from RNA-seq libraries generated from WT and FK neurons labeled for 20 (A) or 60 (B) minutes. Dendrogram shows the unsupervised hierarchical clustering using their Spearman's correlation coefficients. **d,** ECDF plots of degradation rates for RNAs up- (red) or down- (blue) regulated in FXS-like group in WT or FK neurons estimated in libraries A and B. Black vertical lines denote the median degradation rates for all genes with an estimated degradation rate in WT for libraries A and B respectively. **e,** Heatmap of log2FC of synthesis (syn.), processing (pro.) and degradation (deg.) rates for genes with significant changes in any of these measures. **f,** Venn diagrams between genes with faster (brown) or slower (green) degradation rates and FMRP CLIP targets (left, pink) and genes with RNAs up- (red) or down- (blue) regulated in the FXS-like group in brain cortices (right). Numbers of genes in each group and each overlap, as well as the p-value of enrichment of each overlap, are indicated. **g,** Violin plots of log2FC of degradation rates in FK vs WT neurons for genes in each gene cAI score bins as described previously (**Fig 2c**). Brown star indicate the median of the bin greater than 0 with a p-value < 0.01 (Wilcoxon test, one tail). No bin had median less than 0.

cAI score-dependent reduction of steady state mRNA in the cortex as we observed previously (**Fig 2c, left**), there was indeed a preferential destabilization (higher degradation rate) of genes with high cAI scores (**Fig 3g, Fig S3h**).

## Disruption in RNA stability leads to massive reshuffling of stabilizing vs destabilizing codon identities in FMRP-deficient neurons

The codon-stability coefficient (CSC), which describes the link between mRNA stability and codon occurrence, has been calculated for each codon from yeast to human [12–15] . We determined whether this relationship is maintained in FK neurons by first calculating CSC in WT neurons. Here CSC values ranged from < -0.2 to > 0.2, which is comparable to previously reported CSC values for human cell lines and mouse embryonic stem cells (**Fig 4a**). This is unlike what has been described in fly where the neuronal CSC is attenuated relative to somatic cells[13]. Of the 60 non-start or -stop codons, 29 had CSCs greater than 0 (stabilizing codons) and 31 less than 0 (destabilizing codons) (**Fig 4b**, upper). Strikingly, 17 codons that are stabilizing in WT are destabilizing in FK neurons, and 21 codons changed the opposite way (**Fig 4b,** lower). Because optimal codons are generally frequently used in highly expression genes and are associated with positive CSCs (i.e., are stabilizing codons) [12–15], this reshuffling of the

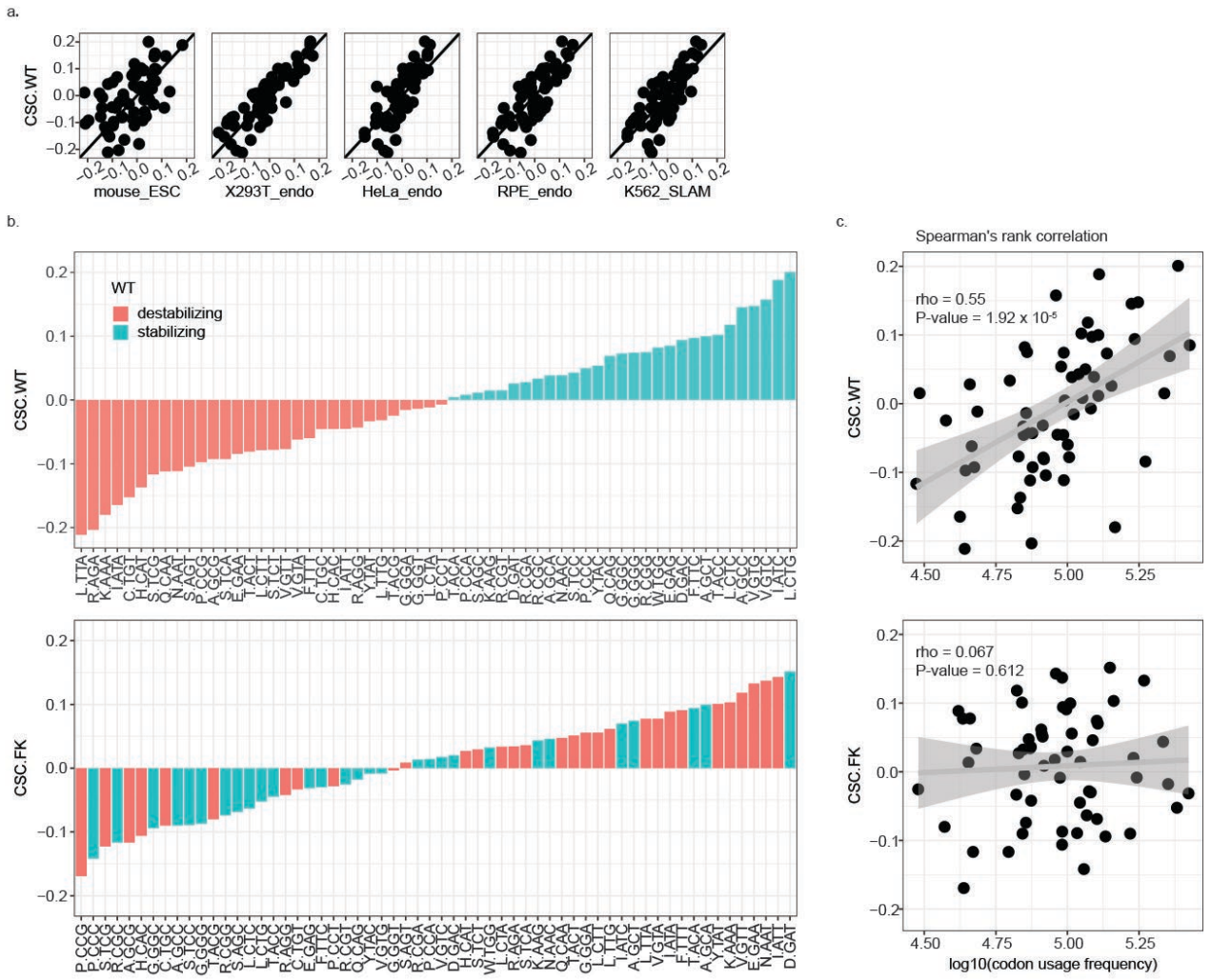



**Figure 4: Loss of FMRP leads to uncoupling of the link between codon usage bias and the identities of stabilizing and destabilizing codons. a,** Scatter plots comparing CSC (codon stability coefficient) of all non-start or -stop codons in WT neurons in this study (y axis) and that in several mouse and human cell lines shown by Wu et al[15]. CSCs in mouse ESC were calculated by Wu et al. based on data published by Herzog et al[62]. CSCs in human cells lines (293T, HeLa, and RPE) measured by either blocking transcription ("endo") or SLAM-seq ("SLAM") were generated by Wu et al. **b,** Bar graphs for Codon-Stability Coefficients (CSC) for each codon as arranged from minimum to maximum in WT (upper) and FK (lower) neurons. The color of each bar indicates the codon as stabilizing (CSC > 0, green) or destabilizing (CSC < 0, orange) in WT neurons. The amino acid for each codon is indicated. **c,** Scatter plots and linear regressions of the CSCs as a function of $\log_{10}$(codon usage frequency) of the top 10% of expressed genes in WT (upper) and FK (lower) neurons. Spearman's rank correlation coefficients and p-values of the correlations are indicated.

identities of stabilizing vs destabilizing codons in FMRP-deficient neurons could reflect changes in codon usage bias in highly expressed genes accordingly, i.e., the correlation between CSCs and codon usage bias is maintained, or alternatively, this reshuffling could result from a uncoupling of the link between codons' stabilizing or destabilizing properties and their usage bias. To examine which of the possibilities is the case, we tested Spearman's rank correlation

between CSCs and codon usage frequencies in the top 10% expressed mRNAs in WT and FK neurons. As expected, in WT neurons there is a positive correlation (r = 0.55, p-value = 1.92 x $10^{-5}$). However this correlation is largely lost in FK neurons, i.e., the correlation is almost random (r = 0.067, p-value = 0.612) (**Fig 4c**). The codon usage frequency in the top expressed genes in WT and FK neurons was largely unchanged (data not shown). These results show that the link between codons' stabilizing vs destabilizing properties and their usage bias is uncoupled in FMRP-deficient neurons.

## Discussion

Although it is widely assumed that FXS is caused by excessive protein synthesis[32], our study shows this postulate is over-simplistic. We find that steady state RNA levels are globally disrupted in the disorder, and that genetic rescue by *Cpeb1* deletion, and possibly in other rescue paradigms as well, mitigates this molecular dys-regulation. The loss of FMRP results in enhanced instability not only of its direct target substrates, but also of other mRNAs with an optimal codon bias transcriptome-wide. Our data show that RNA stability conferred by optimal codons requires trans-acting factors such as FMRP. This requirement leads to the massive reshuffling of the identities of stabilizing versus destabilizing codons in FMRP-deficient neurons.

FMRP could regulate codon dependent mRNA stability either directly or indirectly. Because FMRP seems to target transcripts with a bias for optimal codons (**Fig S2c**), and FMRP CLIP targets are generally reduced in FMRP deficient cortex (**Fig 2c**, right), FMRP could be recruited to optimal codons and its binding directly stabilize the target transcripts. Alternatively, FMRP could regulate mRNA stability indirectly through translation regulation. Translation and RNA decay are closely linked; aberrant translation activity could lead to accelerated mRNA decay[33]. Indeed, the genes with down-regulated transcript levels in FK cortex had generally up-regulated ribosome occupancy (**Fig 1c**), a measure for translation activity. In particular, repressing

translation elongation by applying translation elongation inhibitors such as cyclohexamide[34,35] and sordarin[35], by mutating the gene encoding eIF5A[36,37], or by simulating histidine starvation by treating cells with 3-amino-1,2,4-triazole[35], has been shown to stabilize mRNAs. This is reminiscent of the model where FMRP stalls translation elongation. Loss of FMRP could lead to derepression of translation elongation of its target transcripts and with it enhanced mRNA decay. However, these scenarios cannot explain the dys-regulated RNA degradation rates for mRNAs that are not FMRP CLIP targets. One possibility is that FMRP binds far more mRNAs than can be covalently crosslinked by UV irradiation. The FMRP CLIP experiments in the mouse cortex[6] did not use a nucleoside analog such as 4-thio uridine to enhance UV crosslinking[38], nor did they use formaldehyde[39], which does not rely on short-range proximity of FMRP to RNA to detect an association. Thus, the FMRP CLIP RNAs may be an underestimate of the number of transcripts bound by this protein.

Another possibility could be that FMRP regulates the codon-dependent stability of the transcriptome via its interaction with other protein binding partners. In yeast, nonoptimal codons induce ribosome pileup, which is recognized by Dhh1p, an RNA helicase that leads to mRNA destruction[24]. However, we have no evidence for ribosome pileup in mouse brain cortex. Moreover, in FMRP-deficient brain, the destabilized RNAs have increased ribosome occupancies (**Fig 1c**), not decreased. Consequently, a cause-and-effect relationship among ribosome occupancy, codon optimality, and RNA destruction as illustrated in yeast may not precisely apply to the mammalian brain. However, it is curious to note that FMRP interacts with the mammalian ortholog of Dhh1p, DDX6[40,41], and that DDX6/Dhh1p CLIPs predominantly to mRNA coding regions and 3'UTRs[42,43]. It is worth noting the strong correlation between CDS GC content and RNA changes in the FXS-like group (**Fig S2b**), as well as the high GC content of the most destabilizing codons in FK neurons (**Fig 4b**, lower). Several factors have been found to regulate mRNA stability depending on GC content associated with codon optimality, including

328 DDX6[44] and ILF2[45] in human cells. It is tempting to speculate that in FMRP-deficient brain,
329 DDX6 might mediate the destabilization of the down-regulated RNAs.

330 Lastly, we cannot exclude the possibility that loss of FMRP could impact the availability of
331 charged tRNAs of certain anticodons. If such were the case, the balance between supply
332 (charged tRNA) and demand (codon usage) would be lost, leading to the dys-regulation of
333 translation elongation and mRNA decay, and therefore the uncoupling of the link between
334 stabilizing/destabilizing codon and codon usage bias.

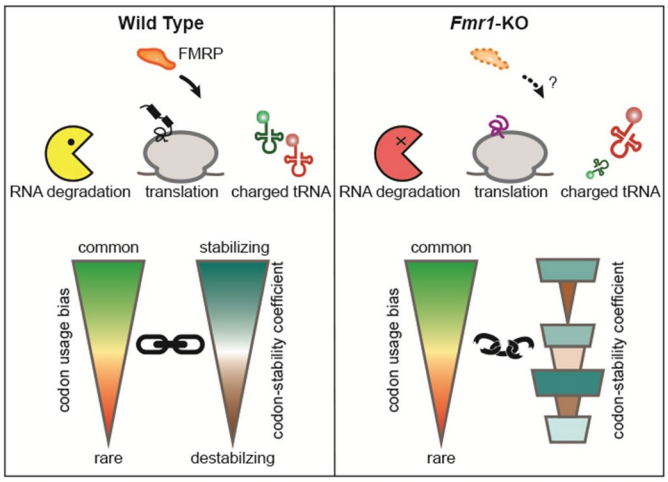



**Figure 5: Model of FMRP as the link between codon usage bias and the identities of stabilizing vs destabilizing codons in neurons.** In WT neurons (left), FMRP maintains the link between codon usage bias in the transcriptome and the rank of stabilizing and destabilizing codons. In *Fmr1*-KO neurons (right), loss of FMRP may lead to the dysregulation of one or more posttranscriptional regulatory mechanisms such as RNA degradation, translation, or availability of charged tRNAs, which in turn leads to the uncoupling of codon usage bias and codon-stability coefficients.

335 Our study establishes FMRP as a link between stabilizing/destabilizing codons and codon
336 usage bias in the neuronal transcriptome (**Fig 5**). Given the broad similarity between the WT
337 and dKO transcriptional profiles, we speculate that genetic rescue by CPEB1 ablation likely
338 causes a realignment of CSC to resemble that of WT. This realignment could be a key to
339 ameliorating the Fragile X disorder.

# Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

## Animals

WT, FK, CK and dKO mice were as used previously[8]. Specifically, FK (JAX stock# 004624) and its WT controls (JAX stock# 004828) were purchased from the Jackson Lab. CK were created in-lab[46]. Mice were bred as previously described[8]. All mice were maintained in a temperature- (25°C), humidity- (50–60%) and light-controlled (12 hr light-dark cycle) and pathogen-free environment. Animal protocols were approved for use by the University of Massachusetts Medical School Institutional Animal Care and Use Committee (IACUC).

## Ribosome profiling and RNA-seq in cortex

Two mice per genotype were used for ribosome profiling and RNA-seq. The brain was rapidly removed from P28–P35 mice, rinsed in ice-cold dissection buffer (1× HBSS + 10 mM HEPES-KOH), rapidly dissected in dissection buffer ice-liquid mixture to collect cerebral cortex as described previously[47]. Both cortex hemispheres were homogenized in 900µl of homogenization buffer[47] (10 mm HEPES-KOH, pH 7.4, 150 mm KCl, 5 mm MgCl2, 0.5 mm DTT, 100 µg/ml cycloheximide and 2 µg/ml harringtonine), containing protease and phosphatase inhibitors (cOmplete, EDTA-free Protease Inhibitor Cocktail and PhosSTOP from Roche/Sigma, cat. no. 11836170001 and 4906837001), in a pre-chilled 2-mL glass Dounce homogenizer, 20 strokes loose, 20 strokes tight, and centrifuged at low speed (2000 rcf 10 min 4°C) to pellet insoluble material[8]. Five hundred microliters of the resulted ~700µl supernatant (cytoplasmic lysate) were used for ribosome profiling, the rest for RNA-seq. For ribosome profiling, the lysate was digested by 60 U RNase T1 (Thermo Scientific, cat. no. EN0541) and 100ng RNase A (Ambion,

cat. no. AM2270) per A260 unit[48] for 30 min at 25°C with gentle mixing. Digestion was stopped by adding 30 µl SUPERase·In (Invitrogen, cat. no. AM2694). Digested lysate was separated by sedimentation through sucrose gradients. Monosome fractions were identified, pooled, and extracted with TRIzol LS (Invitrogen, cat. no. 10296028).

For RNA-seq, cytoplasmic RNA was extracted from the lysate using TRIzol LS. Ten micrograms of RNA were depleted of rRNA using Ribo-Zero Gold rRNA Removal Kit, Human/Mouse/Rat (illumina, discontinued), and fragmented by incubating with PNK buffer (NEB, cat. no. M0201S) for 10 min at 94°C. Fragmented RNA as separated on 15% urea-polyacrylamide gel, and 50-60nt fraction was collected.

Ribosome profiling and RNA-seq libraries were prepared following published protocols[49] and sequenced with Illumina NextSeq.

**Spike-in RNA for RNA metabolism profiling**

*D. melanogaster* (fly) Schneider 2 (S2) cells were grown in 12 ml Schneider's insect medium (Sigma-Aldrich, cat. no. S0146) containing 10% (v/v) of Fetal Bovine Serum (FBS, Sigma-Aldrich, cat. no. F2442) at 28°C until confluent. Cells were incubated with 200 µM 5-EU for 24 hr, and were washed, pelleted and snap frozen in liquid nitrogen. RNA was extracted using TRIzol.

*S. cerevisiae* (yeast) cells were grown in 10 ml YEP medium containing 3% glucose at 30°C until $OD_{600\ nm}$ reaches 0.5. Cells were then pelleted and RNA was extracted using hot acidic phenol[50].

**RNA metabolism profiling with cortical neuron cultures**

Cortical cell suspension were obtained by dissociating cerebral cortices from E18 embryos using the Papain Dissociation System (Worthington, cat. no. LK003150). One million live cells

were plated in 5 ml complete Neurobasal culture medium (Neurobasal™ Medium (Gibco, cat. no. 21103049), 1x B-27 supplement (Gibco, cat. no. 17504044), 1x Antibiotic-Antimycotic (Gibco, cat. no. 15240096), 1x GlutaMAX (Gibco, cat. no. 35050061)) per 60 mm poly-L-lysine treated cell culture dish. Neurons were fed by half-replacing the complete Neurobasal culture medium twice per week. DIV14 neurons were incubated with 200 μM 5-EU (Click-iT™ Nascent RNA Capture Kit, Invitrogen, cat. no. C10365) for 0 (input and "unlab"), 20 ("A"), and 60 ("B") min. Neurons were then washed with ice cold 1x PBS buffer, and RNA was extracted using TRIzol (Invitrogen, cat. no. 15596018). Five-EU labeled RNA was enriched and RNA-seq library was prepared by adapting the Coller lab (Case Western Reserve University) protocol (personal communication). Specifically, mouse neuron RNA was spiked-in with 10% (w/w) 5-EU labeled fly RNA and 10% (w/w) yeast RNA. The mixed RNA was depleted of rRNA using the Ribo-Zero Gold rRNA Removal Kit (Human/Mouse/Rat) and fragmented using NEBNext® Magnesium RNA Fragmentation Module (NEB, cat. no. E6150S) for 5 min. RNA samples for libraries unlab, A and B were biotinylated by Click-iT chemistry and pulled-down using the Click-iT™ Nascent RNA Capture Kit. The pulled-down samples, together with the input sample, were subjected to library construction the same as for ribosome profiling libraries[49]. Here, for the pulled-down samples, all reactions were performed directly on beads until after the reverse transcription step. Sequencing library fraction with insert size 50-200nt was collected and sequenced with Illumina NextSeq. For each WT and FK, two independent batches of neurons were prepared, and each batch resulted in one of each input, unlab, A and B libraries (**Tab S3**).

**Differential translation and RNA expression analysis**

Brain cortex ribosome profiling and RNA-seq reads were processed as previously described[11], which includes the following steps: 1) reads were separated based on sample barcode sequences; 2) known 3' adapter sequences and low quality bases were removed with Cutadapt[51] using parameters -O 2 -q 15 -a

412 "TGGAATTCTCGGGTGCCAAGGAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG"; 3) reads mapped to rRNA and
413 tRNA genes were removed using Bowtie2[52] with parameter -N 1; 3) remaining reads were
414 mapped to the mm10 genome using TopHat2[53]; and 4) PCR duplicates were removed based on
415 Unique Molecular Identifier sequences.

416 Uniquely mapped reads were then used as input to RSEM[54] for quantification of gene
417 expression and either mapped to RefSeq (v69) mouse coding sequences (RPF) or to whole-
418 transcriptome (RNA-seq). Genes were filtered to have a minimum of 10 TPM (transcripts per
419 million) in at least one sample. We used log transformed TPM expression values to correct for
420 batch effects using ComBat[55] (v3.18.0). Corrected values were transformed back to read counts
421 using the expected size of each transcript informed by RSEM. Batch-corrected counts were
422 used to identify differentially translated/expressed genes with DESeq2[56] (RPF and RNA) or
423 Xtail[10] (ribosome occupancy, RO).

424 **GO analysis**

425 GO enrichment analysis was performed using Cytoscape with the ClueGo[57] plug-in (v2.3.3),
426 with genes that are expressed in the mouse cortex as the reference gene set. Specifically,
427 biological function GO terms of levels 6-13 were tested for enrichment at adjusted p-value <
428 0.05 (Right-sided hypergeometric test). Enriched GO terms that are similar were then fused to a
429 group based on their Kappa score which quantifies percentage of common genes between
430 terms. The leading group terms, which are the terms with highest significance in each group,
431 are presented in **Fig 2c**. All enriched terms are in **Tab. S1** and **S2**.

432 **Codon adaptation index (cAI)**

433 The codon adaptation index was calculated for a given sample as described by Sharp & Li (ref
434 1987). Briefly, for each sample, a set of the top 10% expressed genes was defined using batch-
435 corrected TPM; the relative synonymous codon usage was then calculated, dividing the

436 observed frequency of each codon by the frequency expected assuming all synonymous
437 codons for a given amino acid are used equally; the codon adaptation index (cAI) is then
438 calculated by comparing the frequency of each codon to the frequency of the most abundant (or
439 optimal) codon for a given amino acid. All codes used to perform this analysis are available on
440 GitHub (https://github.com/elisadonnard/CodonOPT).

441 **RNA metabolism profiling analysis**

442 Reads generated from the RNA metabolism profiling libraries were processed as for cortical
443 RNA-seq libraries described above, except that a mouse-fly-yeast merged genome (mm10 +
444 dm6 + sacCer3) was used as the reference genome for reads mapping by hisat2. The mapping
445 statistics here were used for quality control and filtering purposes (**Fig S3a-d**). Uniquely mapped
446 reads that are depleted of rRNA, tRNA sequences and PCR duplicates were again mapped to
447 mm10 genome with hisat2. Intron and exon read quantification, and RNA metabolism rates
448 (synthesis, processing and degradation) estimation was performed using INSPEcT[30] (v1.10.0),
449 with the degDuringPulse parameter set to TRUE. One set of libraries, which was of low
450 complexity (**Tab S3**), was still used to confirm the global shift of degradation rates in FK
451 neurons. This reproducible global shift allowed us to normalize WT and FK libraries separately
452 for our gene level analysis (**Fig S3g**). Specifically, raw RNA metabolism rates estimated by
453 INSPEcT were normalized between libraries A and B for WT and FK neurons separately using
454 the limma package[58] with the "cyclicloess" method. After normalization, genes with different
455 metabolism rates were tested using the limma package.

456 **Codon-stability coefficient (CSC) analysis**

457 CSCs were calculated as previously described[8,9,12,14]. Specifically, a Pearson's correlation
458 coefficient was calculated for each of the 60 non-start and -stop codons between the
459 frequencies of this codon in all the genes that use this codon, and the stability of these genes.

460 The stability of a gene (y), which is the inverse of its degradation rate (x), is expressed as
461 follows:

$$y = -(\log_2 x_A + \log_2 x_B)/2$$

462 Here $x_A$ and $x_B$ are the normalized degradation rates from library A and B respectively. The
463 highest expressed isoform of each gene was used to calculate the usage frequencies of each
464 codon.

465 **Quantification and statistical analysis**

466 Blinding or randomization was not used in any of the experiments. The number of independent
467 biological replicates used for an experiment is indicated in the respective figure legends or main
468 text. The statistical tests and P values used for the interpretation of data are mentioned in the
469 figure legends or main text.

470 **Code availability**

471 All codes used to perform cAI analysis are available on GitHub
472 (https://github.com/elisadonnard/CodonOPT). Other customized R scripts for data analysis are
473 available from the corresponding authors upon request.

474 **Data availability**

475 The data supporting the findings of this study have been deposited in the Gene Expression
476 Omnibus (GEO) repository with the accession code GSE0000000. All other data are available
477 from the corresponding authors upon reasonable request.

478 # References

479 1. Santoro, M. R., Bray, S. M. & Warren, S. T. Molecular Mechanisms of Fragile X Syndrome: A
480 Twenty-Year Perspective. *Annu. Rev. Pathol. Mech. Dis.* **7**, 219–245 (2012).

2. Pieretti, M. *et al.* Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66**, 817–822 (1991).

3. Dölen, G. *et al.* Correction of fragile X syndrome in mice. *Neuron* **56**, 955–962 (2007).

4. Bhattacharya, A. *et al.* Genetic Removal of p70 S6 Kinase 1 Corrects Molecular, Synaptic, and Behavioral Phenotypes in Fragile X Syndrome Mice. *Neuron* **76**, 325–337 (2012).

5. Bear, M. F., Huber, K. M. & Warren, S. T. The mGluR theory of fragile X mental retardation. *Trends Neurosci.* **27**, 370–377 (2004).

6. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).

7. Maurin, T. *et al.* HITS-CLIP in various brain areas reveals new targets and new modalities of RNA binding by fragile X mental retardation protein. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky267

8. Udagawa, T. *et al.* Genetic and acute CPEB1 depletion ameliorate fragile X pathophysiology. *Nat. Med.* **19**, 1473–1477 (2013).

9. Gross, C. *et al.* Increased Expression of the PI3K Enhancer PIKE Mediates Deficits in Synaptic Plasticity and Behavior in Fragile X Syndrome. *Cell Rep.* **11**, 727–736 (2015).

10. Richter, J. D., Bassell, G. J. & Klann, E. Dysregulation and restoration of translational homeostasis in fragile X syndrome. *Nat. Rev. Neurosci.* **16**, 595–605 (2015).

11. Udagawa, T. *et al.* Bidirectional control of mRNA translation and synaptic plasticity by the cytoplasmic polyadenylation complex. *Mol. Cell* **47**, 253–266 (2012).

12. Bazzini, A. A. *et al.* Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *EMBO J.* **35**, 2087–2103 (2016).

13. Burow, D. A. *et al.* Attenuated Codon Optimality Contributes to Neural-Specific mRNA Decay in Drosophila. *Cell Rep.* **24**, 1704–1712 (2018).

14. Presnyak, V. *et al.* Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* **160**, 1111–1124 (2015).

15. Wu, Q. *et al.* Translation affects mRNA stability in a codon-dependent manner in human cells. *eLife* **8**, e45396 (2019).

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

17. Xiao, Z., Zou, Q., Liu, Y. & Yang, X. Genome-wide assessment of differential translations with ribosome profiling data. *Nat. Commun.* **7**, 11194 (2016).

18. Liu, B. *et al.* Regulatory discrimination of mRNAs by FMRP controls mouse adult neural stem cell differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E11397–E11405 (2018).

19. Korb, E. *et al.* Excess Translation of Epigenetic Regulators Contributes to Fragile X Syndrome and Is Alleviated by Brd4 Inhibition. *Cell* (2017). doi:10.1016/j.cell.2017.07.033

20. SFARI Gene. *SFARI* (2017). Available at: https://www.sfari.org/resource/sfari-gene/. (Accessed: 15th November 2017)

21. Feng, Y. *et al.* Fragile X Mental Retardation Protein: Nucleocytoplasmic Shuttling and Association with Somatodendritic Ribosomes. *J. Neurosci.* **17**, 1539–1547 (1997).

22. Zalfa, F. *et al.* A new function for the fragile X mental retardation protein in regulation of PSD-95 mRNA stability. *Nat. Neurosci.* **10**, 578–587 (2007).

524 23. Chen, E., Sharma, M. R., Shi, X., Agrawal, R. K. & Joseph, S. Fragile X Mental Retardation
525 Protein Regulates Translation by Binding Directly to the Ribosome. *Mol. Cell* **54**, 407–417
526 (2014).
527 24. Radhakrishnan, A. *et al.* The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation
528 by Monitoring Codon Optimality. *Cell* **0**, (2016).
529 25. Pechmann, S. & Frydman, J. Evolutionary conservation of codon optimality reveals hidden
530 signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* **20**, 237–243 (2013).
531 26. Mishima, Y. & Tomari, Y. Codon Usage and 3' UTR Length Determine Maternal mRNA
532 Stability in Zebrafish. *Mol. Cell* **61**, 874–885 (2016).
533 27. Sharp, P. M. & Li, W. H. The codon Adaptation Index--a measure of directional synonymous
534 codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987).
535 28. Neymotin, B., Ettorre, V. & Gresham, D. Multiple Transcript Properties Related to
536 Translation Affect mRNA Degradation Rates in Saccharomyces cerevisiae. *G3*
537 *GenesGenomesGenetics* **6**, 3475–3483 (2016).
538 29. Bornelöv, S., Selmi, T., Flad, S., Dietmann, S. & Frye, M. Codon usage optimization in
539 pluripotent embryonic stem cells. *Genome Biol.* **20**, 119 (2019).
540 30. de Pretis, S. *et al.* INSPEcT: a computational tool to infer mRNA synthesis, processing and
541 degradation dynamics from RNA- and 4sU-seq time course experiments. *Bioinformatics* **31**,
542 2829–2835 (2015).
543 31. Mukherjee, N. *et al.* Integrative classification of human coding and noncoding genes
544 through RNA metabolism profiles. *Nat. Struct. Mol. Biol.* **24**, 86–96 (2017).

32. Kelleher, R. J. & Bear, M. F. The Autistic Neuron: Troubled Translation? *Cell* **135**, 401–406 (2008).

33. Roy, B. & Jacobson, A. The intimate relationships of mRNA decay and translation. *Trends Genet.* **29**, 691–699 (2013).

34. Beelman, C. A. & Parker, R. Differential effects of translational inhibition in cis and in trans on the decay of the unstable yeast MFA2 mRNA. *J. Biol. Chem.* **269**, 9687–9692 (1994).

35. Chan, L. Y., Mugler, C. F., Heinrich, S., Vallotton, P. & Weis, K. Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *eLife* **7**, e32536 (2018).

36. Saini, P., Eyler, D. E., Green, R. & Dever, T. E. Hypusine-containing protein eIF5A promotes translation elongation. *Nature* **459**, 118–121 (2009).

37. Zuk, D. & Jacobson, A. A single amino acid substitution in yeast eIF-5A results in mRNA stabilization. *EMBO J.* **17**, 2914–2925 (1998).

38. Hafner, M. *et al.* Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).

39. Singh, G., Ricci, E. P. & Moore, M. J. RIPiT-Seq: A high-throughput approach for footprinting RNA:protein complexes. *Methods* **65**, 320–332 (2014).

40. Barbee, S. A. *et al.* Staufen- and FMRP-Containing Neuronal RNPs Are Structurally and Functionally Related to Somatic P Bodies. *Neuron* **52**, 997–1009 (2006).

41. Ayache, J. *et al.* P-body assembly requires DDX6 repression complexes rather than decay or Ataxin2/2L complexes. *Mol. Biol. Cell* **26**, 2579–2595 (2015).

566 42. Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat. Struct. Mol.*
567 *Biol.* **20**, 127–133 (2013).
568 43. Nostrand, E. L. V. *et al.* A Large-Scale Binding and Functional Map of Human RNA Binding
569 Proteins. *bioRxiv* 179648 (2018). doi:10.1101/179648
570 44. Courel, M. *et al.* GC content shapes mRNA decay and storage in human cells. *bioRxiv*
571 373498 (2019). doi:10.1101/373498
572 45. Hia, F. *et al.* Codon Bias Confers Stability to mRNAs via ILF2 in Humans. *bioRxiv* 585992
573 (2019). doi:10.1101/585992
574 46. Tay, J. & Richter, J. D. Germ Cell Differentiation and Synaptonemal Complex Formation Are
575 Disrupted in CPEB Knockout Mice. *Dev. Cell* **1**, 201–213 (2001).
576 47. Stefani, G., Fraser, C. E., Darnell, J. C. & Darnell, R. B. Fragile X Mental Retardation Protein Is
577 Associated with Translating Polyribosomes in Neuronal Cells. *J. Neurosci.* **24**, 7272–7276
578 (2004).
579 48. Cenik, C. *et al.* Integrative analysis of RNA, translation and protein levels reveals distinct
580 regulatory variation across humans. *Genome Res.* gr.193342.115 (2015).
581 doi:10.1101/gr.193342.115
582 49. Heyer, E. E., Ozadam, H., Ricci, E. P., Cenik, C. & Moore, M. J. An optimized kit-free method
583 for making strand-specific deep sequencing libraries from RNA fragments. *Nucleic Acids Res.*
584 **43**, e2–e2 (2015).
585 50. Collart, M. A. & Oliviero, S. Preparation of Yeast RNA. *Curr. Protoc. Mol. Biol.* **23**, 13.12.1-
586 13.12.5 (1993).

587 51. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
588 *EMBnet.journal* **17**, 10–12 (2011).
589 52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**,
590 357–359 (2012).
591 53. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions,
592 deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
593 54. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or
594 without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
595 55. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for
596 removing batch effects and other unwanted variation in high-throughput experiments.
597 *Bioinformatics* **28**, 882–883 (2012).
598 56. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
599 RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
600 57. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene
601 ontology and pathway annotation networks. *Bioinforma. Oxf. Engl.* **25**, 1091–1093 (2009).
602 58. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and
603 microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
604 59. Thomson, S. R. *et al.* Cell-Type-Specific Translation Profiling Reveals a Novel Strategy for
605 Treating Fragile X Syndrome. *Neuron* **95**, 550-563.e5 (2017).
606 60. Zhang, F. *et al.* Fragile X mental retardation protein modulates the stability of its m6A-
607 marked messenger RNA targets. *Hum. Mol. Genet.* **27**, 3936–3950 (2018).

608 61. Haenfler, J. M. *et al.* Targeted Reactivation of FMR1 Transcription in Fragile X Syndrome
609 Embryonic Stem Cells. *Front. Mol. Neurosci.* **11**, (2018).
610 62. Herzog, V. A. *et al.* Thiol-linked alkylation for the metabolic sequencing of RNA. *bioRxiv*
611 177642 (2017). doi:10.1101/177642

## 612 Acknowledgements

613 We thank Emmiliano Ricci for sharing his experience and protocol for ribosome profiling, Mariya
614 Ivshina for advice and help with the mouse breeding and Ruijia Wang for help with the
615 bioinformatics. We thank Lindsay Romo for sharing her protocol for 5-EU labeling of neuron
616 culture, and Jeff Coller (Case Western Reserve University) for Click-iT EU tagging – RNA-seq
617 protocol. We also thank Ariel Bazzini for his very helpful discussion and input. Nathan
618 Gioacchini and Yongjin Lee helped preparing yeast and fly spike-in RNA samples. This works
619 was supported by the National Institutes of Health (U54HD82013); the Simons Foundation, and
620 the Charles H. Hood Foundation (to JDR).

## 621 Author information

### 622 Contributions

623 H.S. and J.D.R. conceived the project and designed the experiments. H.S. performed most of
624 the experiments, B.L. generated RNA-seq libraries. H.S. and E.D. performed the bioinformatic
625 analysis with input from B.L. H.S. and J.D.R. wrote the manuscript with input from all authors.

### 626 Corresponding authors

627 Correspondence to Huan Shu or Joel Richter.

# Ethics declarations

## Competing interests

The authors declare no competing interests.

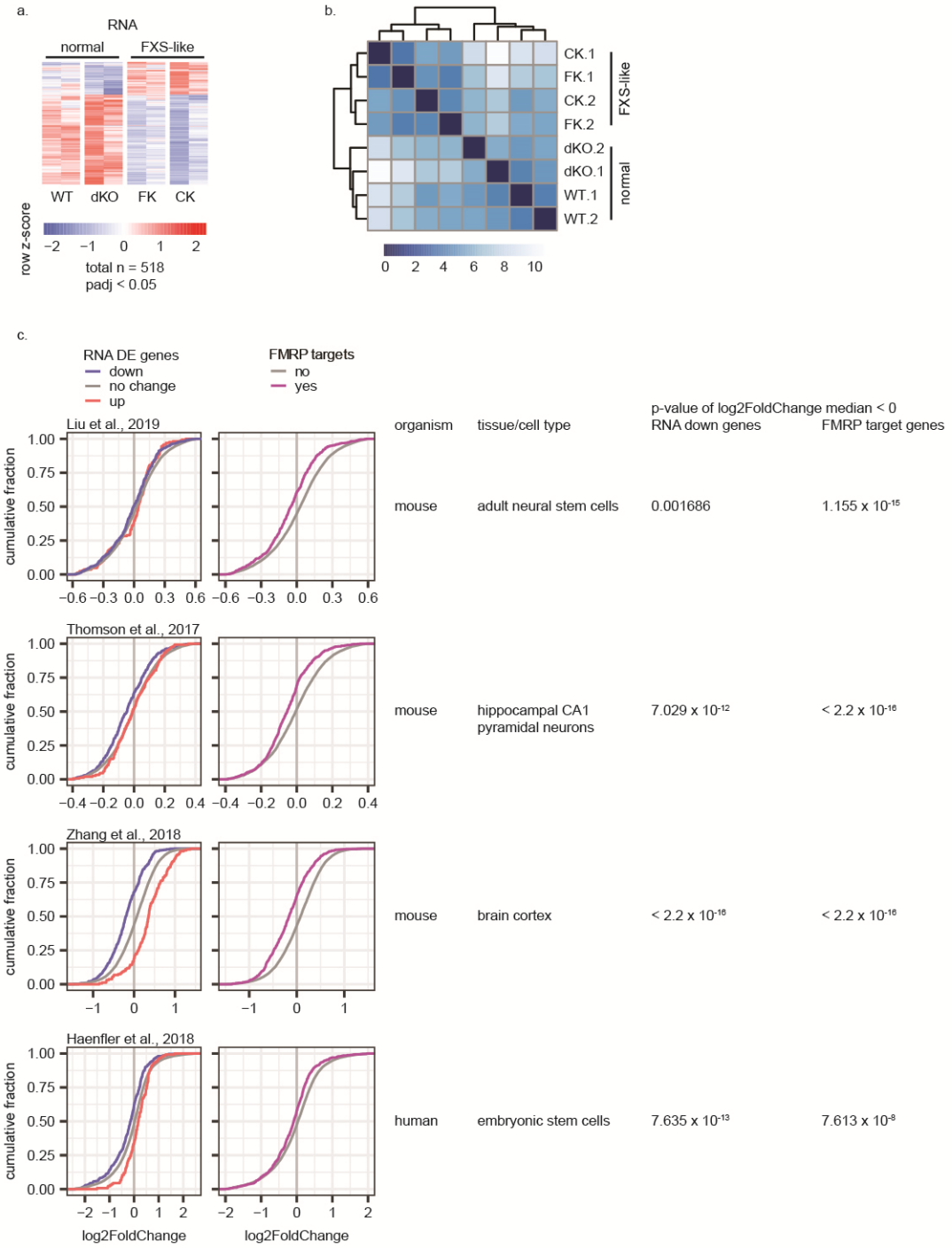



**Figure S1: RNA levels in Fragile X mouse and human models. a,** Heatmap showing DE genes at steady state RNA level between any two genotypes of the four genotypes (WT, *Fmr1*-deficient or FK, *CPEB1*-deficient or CK, *Fmr1/CPEB1* deficient or dKO). Red and blues shades show high or low z-scores for each gene (row) across all samples. Both replicates are plotted separately for each genotype.

**b,** Unsupervised hierarchical clustering of sample to sample distances measured by the Euclidean distance between each other using their top 1000 most variable genes. Darker to lighter shades of blue indicate closer to farther distance between samples. Dendrogram represents the clustering. **c,** ECDF (empirical cumulative distribution function) plots for log2FoldChange in published RNA-seq data sets of various FXS models[18,59–61] for genes up-(red) or down-(blue) regulated at the RNA level identified in this study (left), and for FMRP binding targets[6] (brown; right). The animal species and tissue/cell typed used in each of these studies is indicated. P-values were calculated for the log2FoldChange values of the downregulated genes identified in this study (blue) and of the FMRP targets (brown) to be smaller than 0 (Wilcoxon test, lower tail). For data from Thomson et al., 2017[59], genes were filtered for normalized counts between $10^{2.5}$ and $10^{4.25}$ as was done in the original publication. For data generated using human embryonic stem cells[61], only genes with unique mouse orthologs were considered.

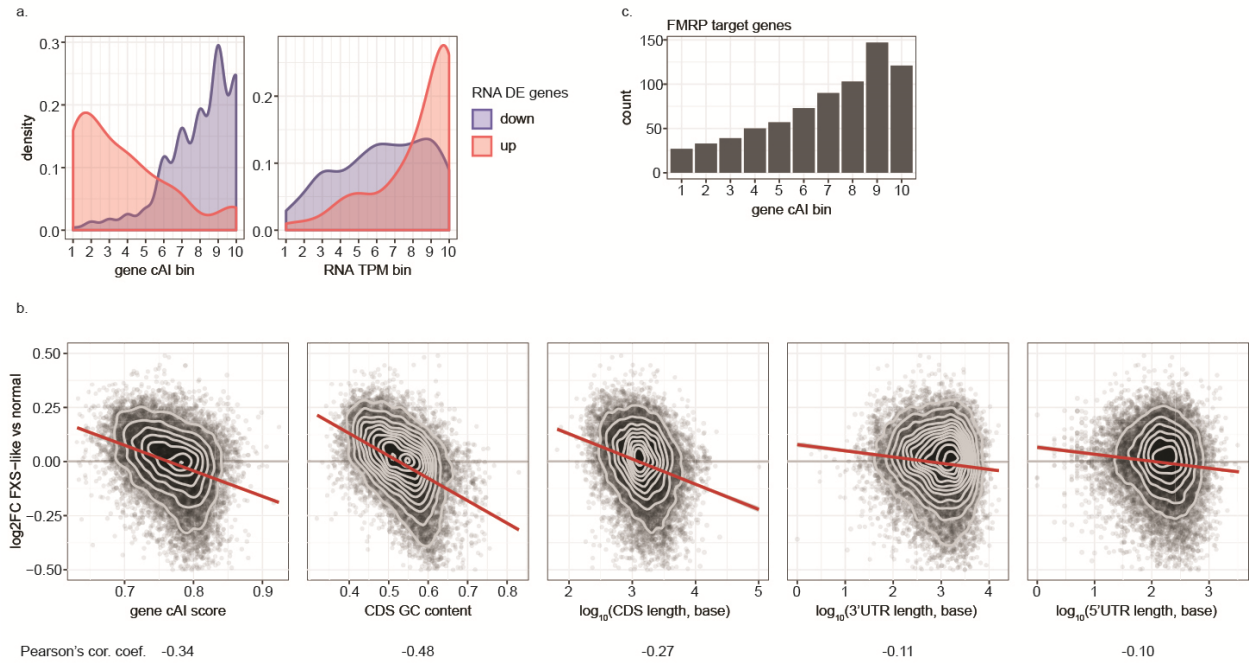



**Figure S2: RNA changes and codon optimality. a,** Density plots of the distribution of genes up- (red) or down- (blue) regulated at the RNA level in the FXS-like group over gene cAI score bins (left) and RNA transcript per million (TPM) bins (right). Gene bins were generated by dividing all detectable protein coding genes into 10 equal bins based on their gene cAI scores (left) or their TPM in WT brain (right). Bin 1 genes have gene cAI scores or TMPs of the lowest quantile and bin 10 genes of the highest quantile. **b,** Scatter and 2D density contour plots of RNA log2FC in FXS-like vs normal group as a function of gene cAI scores, CDS GC content, and log10 of CDS lengths, 3'UTR lengths, and 5'UTR lengths of all detectable genes. The highest expressed isoform of each gene was selected to calculate the gene cAI score, CDS GC content, and lengths of each feature. The red straight line shows the linear regression of the data points. Pearson's product−moment correlation coefficients are indicated. **c,** Bar graph of count of FMRP target genes in each gene cAI score bin.

Shu et al, Supplementary Figures

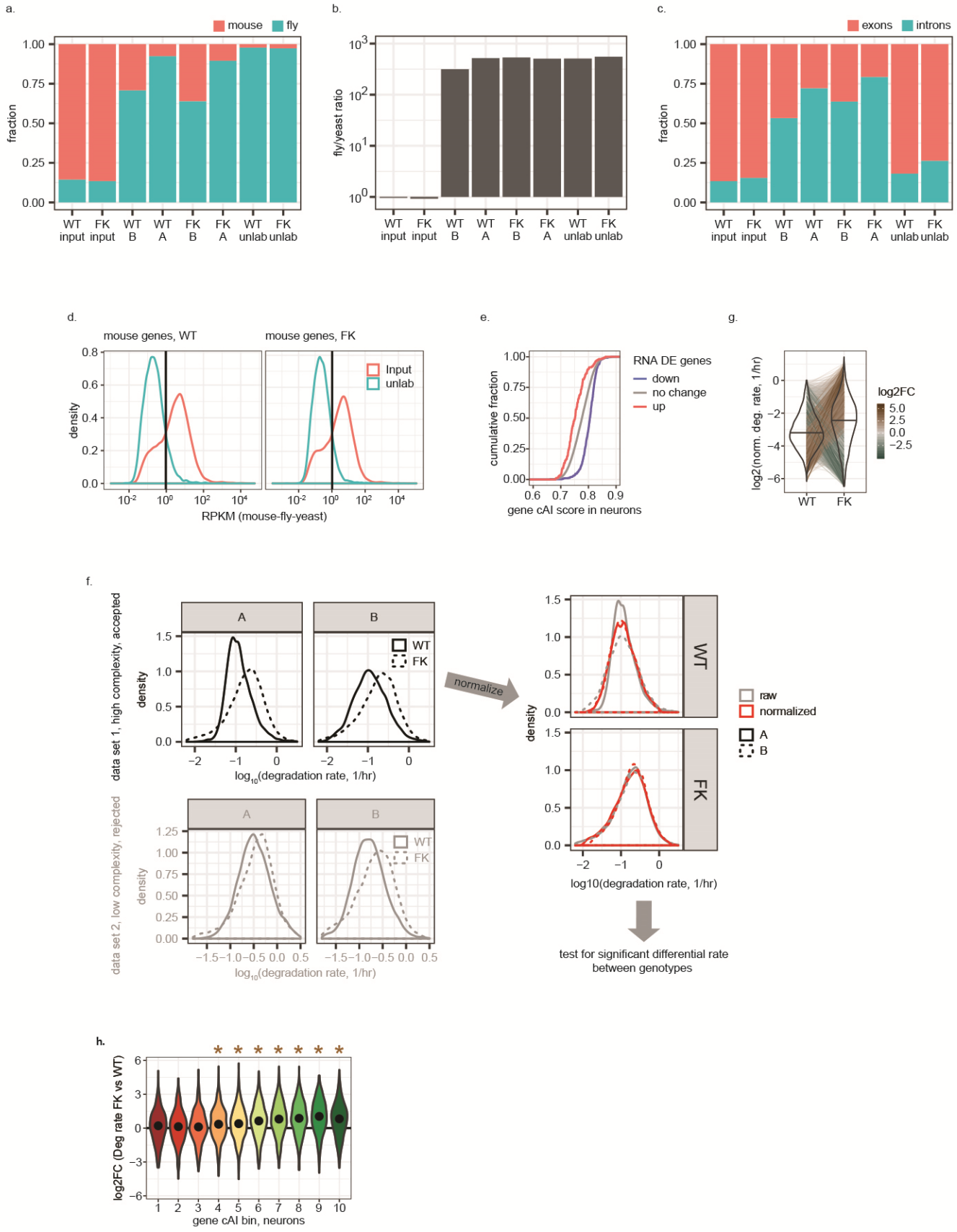


**Figure S3: RNA metabolic profiling in WT and FK neurons. a,** Bar graph of fractions of reads uniquely mapped to mouse (orange) or *Drosophila* (green) transcriptome in each library. As expected, the pulled-down libraries (A, B and unlab) were enriched in reads mapped to the fly transcriptome (5-EU labeled to saturation) over that of mouse (5-EU labeled only for a brief pulse). The input libraries were not subjected to pull-down and had more reads mapped to mouse compared to fly. Unlab libraries had the smallest ratio of reads that mapped to mouse, demonstrating minimum background to the pull-down process. Accordingly, libraries from mouse neurons that are labeled for a shorter time (20min, libraries A) had smaller ratios of reads mapped to mouse transcriptome than that labeled for longer (60min, libraries B). **b,** Bar graph of ratio of reads that uniquely mapped to *Drosophila* transcriptome *vs* that to yeast in each library. Ratios are scaled so that the mean of this ratio in WT input and in FK input libraries is 1. Similar to panel **a**, the high *Drosophila* to yeast ratio demonstrates specific pull-down to enrich for 5-EU labeled RNA. **c,** Bar graph of fractions of reads that uniquely mapped to exons (orange) and introns (green) among those uniquely mapped to the mouse transcriptome. As expected, input libraries are composed mostly of mature mRNAs and therefore had predominantly exon reads. Similarly, the exon/intron ratio for unlab libraries represents nonspecific signal that originates from the input RNA pool. Libraries from mouse neuron RNAs that are labeled for short (20min, A) or longer (60min, B) are mostly composed of nascent transcripts and therefore had more introns. Accordingly libraries labeled for a shorter time (A) had more introns than that labeled for longer (B). **d,** Density plots of RPKM (read per kb per million reads uniquely mapped to mouse-fly-yeast combined genome) of each mouse gene in input (orange) and unlab (green) libraries in WT (left) and FK (right) neurons. Filtering thresholds (black vertical lines) were identified for WT and FK at 0.95 and 1.05 RPKM, respectively. Genes were filtered for those that had RPKM higher than threshold in input (i.e., that are expressed) and lower than threshold in unlab libraries (i.e., that do not have high nonspecific pull-down background). Data of genes that survive filtering in both WT and FK libraries are analyzed by the INSPEcT program[30] to estimate RNA metabolism rates. **e**, ECDF plot of gene cAI scores calculated using WT neuron transcriptome (input) for DE genes at the RNA level in the FXS-like group, similar to **Fig 2b**. **f,** Pipeline

for normalization and statistical tests for genes with differential RNA metabolic rates, using the degradation rate as the example that is shown. We observed reproducible global faster degradation rates in FK than WT in both libraries A and B and in both data set 1 (high quality libraries, presented in this study) and data set 2 (independent data set, low complexity lacks statistical power and was rejected for gene-level analysis (gray shaded graphs) but was sufficient for confirming global-level shift) (left). To capture the global shift between genotypes while testing for genes with significantly different metabolism (i.e., synthesis, processing, and degradation) rates, we considered library A and B in data set 1 as pseudo-replicates and normalized them using the Limma package[58] for each genotype separately. With normalized RNA metabolic rates, genes with significantly different rates between genotypes were then called (right). **g,** Violin-and-line plot for the means of log2 of normalized degradation rates in libraries A and B for all genes with degradation rates inferred by INSPEcT[30] in WT and FK neurons. The black horizontal line in each violin denotes the median. Thin lines span WT and FK connect the values of the same genes in both genotypes. Brown-grey-green shades of the thin lines indicate the log2FC of the normalized degradation rates of each gene. **h,** Violin plots of log2FC of degradation rates in FK vs WT neurons for genes in each gene cAI score bins calculated using WT neuronal transcriptome. Brown star indicates the median of the bin greater than 0 with a p-value < 0.01 (Wilcoxon test, one tail). No bin had median less than 0.