Partial loss of CFIm25 causes aberrant alternative polyadenylation and learning deficits

Callison E. Alcott[1,2,3]; Hari Krishna Yalamanchili[2,3]; Ping Ji[4]; Meike E. van der Heijden[2,6]; Alexander B. Saltzman[7]; Mei Leng[7]; Bhoomi Bhatt[7]; Shuang Hao[2,8]; Qi Wang[2,8]; Afaf Saliba[2,4]; Jianrong Tang[2,8]; Anna Malovannaya[7,9,10,11]; Eric J. Wagner[5]; Zhandong Liu[2,8,12]; Huda Y. Zoghbi[1,2,4,13,14,15]*

[1]Program in Developmental Biology, Baylor College of Medicine, Houston, USA
[2]Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, USA
[3]Medical Scientist Training Program, Baylor College of Medicine, Houston, USA
[4]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA
[5]Department of Biochemistry & Molecular Biology, University of Texas Medical Branch, Galveston, USA
[6]Department of Neuroscience, Baylor College of Medicine, Houston, USA
[7]Verna and Marrs McLean Department of Biochemistry and Molecular Biology Baylor College of Medicine, Houston, USA
[8]Section of Neurology, Department of Pediatrics, Baylor College of Medicine, Houston, USA
[9]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, USA
[10]Mass Spectrometry Proteomics Core, Baylor College of Medicine, Houston, USA
[11]Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, USA
[12]Graduate Program in Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, USA
[13]Department of Neuroscience, Baylor College of Medicine, Houston, USA
[14]Department of Pediatrics, Baylor College of Medicine, Houston, USA
[15]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, USA
* Correspondence should be addressed to: hzoghbi@bcm.edu

ABSTRACT

We previously showed that *NUDT21*-spanning copy-number variations (CNVs) are associated with intellectual disability (Gennarino et al., 2015). However, the patients' CNVs also included other genes. To determine if reduced *NUDT21* function alone can cause disease, we generated *Nudt21*$^{+/-}$ mice to mimic the human state of decreased expression. We found that although these mice have 50% reduced *Nudt21* mRNA, they only have 30% less of its cognate protein, CFIm25. Despite this partial protein-level compensation, the *Nudt21*$^{+/-}$ mice have learning deficits and cortical hyperexcitability. Further, to determine the molecular mechanism driving neural dysfunction, we partially inhibited *NUDT21* in human stem cell-derived neurons to reduce CFIm25 by 30%. This reduction in CFIm25 was sufficient to induce misregulated alternative polyadenylation (APA) and protein levels in hundreds of genes, dozens of which cause intellectual disability when mutated. Altogether, these results indicate that disruption of *NUDT21*-regulated APA events in the brain can cause intellectual disability.

INTRODUCTION

The brain is acutely sensitive to the dose of numerous proteins, such that even small perturbations in their levels can cause neurological disease. Proteins that affect the expression of other genes, such as transcriptional or translational regulators, are particularly critical (Rubeis et al., 2014; Vissers et al., 2015; Yin and Schaaf, 2017). Canonical examples of such broad regulators are the RNA-binding protein FMR1, which underlies fragile X syndrome, and the chromatin modulator MeCP2, whose loss or gain respectively causes Rett syndrome or *MECP2*-duplication syndrome (Amir et al., 1999; Pieretti et al., 1991; Verkerk et al., 1991).

Alternative polyadenylation (APA) is an important mechanism of transcript and protein-level regulation. Indeed, nearly 70% of human genes have multiple polyadenylation [poly(A)] sites, usually within the 3′ untranslated region (UTR) of the mRNA, allowing for transcript isoforms with different 3′ UTR lengths and content (Derti et al., 2012; Hoque et al., 2013). The general implication is that longer 3′ UTRs contain additional regulatory motifs, such as miRNA and RNA-binding protein binding sites. This allows for differential regulation of those gene products, typically through reduced mRNA stability, but also through other mRNA metabolism mechanisms, such as translation efficiency and localization [reviewed in (Tian and Manley, 2016)]. Alternative poly(A) site usage has been observed in multiple biological contexts, both in different disease states and in normal physiology and development (Hoque et al., 2013; Masamha and Wagner, 2017). In general, as cells differentiate, pathway-specific mRNAs increasingly use distal polyadenylation sites, while cells that are more proliferative use more proximal poly(A) sites (Hoque et al., 2013; Ji et al., 2009; Sandberg et al., 2008). In the brain, average 3' UTR length increases throughout neurogenesis, and mature, post-mitotic neurons have the longest average 3' UTR length of any cell type (Guvenek and Tian, 2018; Miura et al., 2013; Zhang et al., 2005).

Numerous genes regulate alternative polyadenylation, but *NUDT21* is among the most consequential (Gruber et al., 2012; Masamha et al., 2014; Tian and Manley, 2016). *NUDT21* encodes CFIm25, a component of the mammalian cleavage factor I (CFIm) complex (Kim et al., 2010; Rüegsegger et al., 1996; Yang et al., 2011). CFIm25 binds UGUA sequences in pre-mRNA and the CFIm complex helps recruit the enzymes required for cleavage and polyadenylation (Brown and Gilmartin, 2003; Rüegsegger et al., 1998; Yang et al., 2011, 2010; Zhu et al., 2018). The UGUA binding sites are often enriched at the distal polyadenylation sites

of *NUDT21*-regulated RNAs, so CFIm25 typically promotes the synthesis of longer mRNA isoforms (Zhu et al., 2018). Thus, when *NUDT21* expression is reduced, proximal cleavage sites are more frequently used. Indeed, CFIm25 downregulation in multiple human and mouse cell lines causes 3′ UTR shortening in hundreds of genes, and a consequent increase in protein levels of a subset of those genes (Brumbaugh et al., 2018; Gennarino et al., 2015; Gruber et al., 2012; Kubo et al., 2006; Li et al., 2015; Martin et al., 2012; Masamha et al., 2014). Notably, *MECP2* is among the most affected genes in these cell-line studies, and slight perturbations in MeCP2 levels cause neurological disease (Chao and Zoghbi, 2012). Moreover, *NUDT21* is a highly constrained gene. In the Genome Aggregation Database (gnomAD) of ~140,000 putatively healthy individuals, 125 missense and 13 loss of function variants would be expected in *NUDT21* if loss of function were not pathogenic, but instead there are only 15 missense and zero loss-of-function variants, suggesting that *NUDT21* loss of function is incompatible with health (Lek et al., 2016). Given this evidence, we hypothesized that *NUDT21* variants can cause neurological disease through APA misregulation of *MECP2* and other dose-sensitive genes in neurons.

Combining results from our previous work with data from the Decipher database, we have identified nine individuals with *NUDT21*-spanning duplications that have intellectual disability, and two patients with deletions that have both intellectual disability and seizures (Firth et al., 2009; Gennarino et al., 2015). However, the duplication patients also have three other genes common to their copy-number variations (CNVs) and the deletion patients have nearly 20 common genes (Gennarino et al., 2015). CNVs of these other genes could be causing their symptoms. Therefore, it is important to determine if *NUDT21* loss of function alone is sufficient to cause disease. Identifying the disease-causing genes within CNVs facilitates more accurate

diagnosis and prognosis, and allows for targeted therapy development. To that end, we generated *Nudt21*$^{+/-}$ mice to model the reduced CFIm25 expression observed in humans and assessed for phenotypes similar to the symptoms seen in the deletion patients. We found that *Nudt21* heterozygosity causes a 50% loss of wild-type *Nudt21* mRNA as expected, but only a 30% reduction of its cognate protein, CFIm25. Consistent with what we observed in the deletion patients, we found that *Nudt21* loss of function is sufficient to cause learning deficits in mice using a variety of behavioral assays. Further, to see how *NUDT21* loss might lead to disease, we analyzed *NUDT21*-depleted human neurons for aberrant alternative polyadenylation with 3' end sequencing and for protein level aberrations with quantitative mass spectrometry. These approaches showed that a 30% reduction of CFIm25 in human neurons induces widespread abnormal alternative polyadenylation and protein levels, including for dozens of dose-sensitive, disease-associated genes. Altogether, these results provide important in vivo and human-specific evidence that reduced *NUDT21* expression can lead to intellectual disability.

RESULTS

Nudt21* heterozygotes have 50% less *Nudt21* mRNA in their brain, but only 30% reduced CFIm25 protein.** *Nudt21* expression has not been well explored within organisms, particularly in the brain. Thus, we first sought to confirm that mice express *Nudt21* in the brain, and found that we could readily detect CFIm25 in neuronal nuclei in regions important for learning and memory—the cortex and hippocampus (Figure 1A***). We then generated *Nudt21* knockout mice by excising exons two and three, which removes the RNA-binding domain of CFIm25 and induces a frame shift mutation leading to nonsense-mediated decay of the transcript (***Figure 1B***) (Yang et al., 2011). We crossed the *Nudt21*$^{+/-}$ mice with each other and observed that of the 32

pups born, there were no homozygous null offspring, indicating that *Nudt21* is essential for life (***Figure 1C***). Unexpectedly, while the heterozygous mice have the anticipated 50% reduction in wildtype *Nudt21* mRNA in whole-brain extracts, there was only ~30% reduction in CFIm25 protein (***Figures 1D & E***). These results demonstrate the successful knockout of one *Nudt21* copy and reveal a partial post-transcriptional compensation of CFIm25 protein levels. Finally, we found that the *Nudt21*$^{+/-}$ mice had a normal life span, but consistently weighed ~10% less than their wild-type littermates, indicating that *Nudt21* heterozygosity might cause some abnormalities (***Figure 1—figure supplement 1***).

Partial loss of CFIm25 causes learning deficits. To determine if *Nudt21* heterozygosity affects cognitive function, we compared *Nud21*$^{+/-}$ mice and their wild-type littermates in several neurobehavioral assays starting at 30 weeks of age. We focused on learning and memory assays because intellectual disability is the most pronounced and consistent symptom seen in patients with *NUDT21*-spanning CNVs (Gennarino et al., 2015). We started with the conditioned fear test, which assesses the mice's ability to learn to associate an aversive event with a sensory context. In this test, we initially train the mice by exposing them to two tone-shock pairings in a novel chamber. Healthy mice will freeze as a behavioral expression of fear when they hear the tone for the second time. The following day, we return the mice to the same chamber, where healthy mice who learned to associate the chamber with the shock will freeze in recognition of the chamber. Lastly, we place the mice into a new chamber and replay the tone, where the healthy mice will remember that the tone is associated with the shock and freeze in fear. We found that *Nud21*$^{+/-}$ mice show abnormal behavior throughout the assay. Their initial fear learning was attenuated: when exposed to the reinforcement sound cue played again two minutes

after the initial tone-shock pairing during training, they froze on average ~50% less often than their wild-type littermates (***Figure 2Ai***). Furthermore, the following day, they froze on average ~30% less in both tests, indicating that *Nud21*$^{+/-}$ mice have reduced memory of the chamber or the sound cue compared to their wild-type littermates (***Figure 2Aii & iii***).

It is important to ensure that potential hyperactivity in the mutant mice does not confound the conditioned fear test results. If the mice are hyperactive, they may freeze less in the conditioned fear assay even if they remember the chamber or the cue. Therefore, we tested our mice in the open field assay. We place the mice in an open chamber and record the distance they travel in 30 minutes. We also record how much time they spend in the exposed center part of the chamber compared to the more hidden periphery. In this assay, hyperactive mice cover a greater distance, and anxious mice spend more time on the perimeter of the chamber (Crawley, 1985). We found that the *Nud21*$^{+/-}$ mice were not hyperactive, thus validating the conditioned fear results by showing they are not confounded by hyperactivity and indicating that partial loss of *Nudt21* function does indeed cause cognitive defects (***Figure 2Bi***). Intriguingly, the *Nud21*$^{+/-}$ mice spent ~15% more time in the exposed, center part of the open field, which suggests they have reduced anxiety and shows that they have neuropsychiatric features beyond learning and memory deficits (***Figure 2Bii***).

To determine if the learning deficits we detected in the *Nud21*$^{+/-}$ mice extended to other learning paradigms, we tested spatial learning using the Morris water maze. In this assay, the mice are trained over four days to locate a hidden platform in a pool of water using visual cues provided on the perimeter of the room. We initially reveal the platform to the mice to confirm visual acuity. Then the mice are individually placed at different locations around the perimeter of the pool eight times per day and timed for how long it takes to locate the platform. Mice with

166 spatial learning deficits require more time to find the platform throughout training. For an
167 additional measurement, the platform is removed after the training and the mice are placed in a
168 novel location on the perimeter of the pool. We then record how much time, out of a minute,
169 they spend searching for the platform in the quadrant of the pool where it had been. Mice with
170 spatial learning and memory deficits spend relatively less time in the platform quadrant,
171 indicating that they do not remember its location as well. We found that the *Nudt21*$^{+/-}$ mice
172 indeed have spatial learning deficits. On average, they required more time to locate the platform
173 during their training trials, despite swimming slightly faster and covering a greater distance
174 (***Figure 2Ci & ii***). Moreover, when the platform was removed, the *Nudt21*$^{+/-}$ mice spent ~30%
175 less time in the correct quadrant (***Figure 2Ciii & iv***). Thus, the conditioned fear and Morris water
176 maze results show that *Nudt21*$^{+/-}$ mice have learning and memory deficits in multiple domains.
177
178 ***Nudt21*$^{+/-}$ mice have increased cerebral spike activity**. In addition to intellectual disability, the
179 patients with *NUDT21*-spanning deletions had seizures. In general, though, mice are much less
180 sensitive to seizure-causing mutations than humans, but seizure susceptibility can be assessed
181 using electroencephalography (EEG) (Amendola et al., 2014; Jiang et al., 1998; Kriscenski-Perry
182 et al., 2002; Miura et al., 2002). Despite a lack of detectable seizures in *Nudt21*$^{+/-}$ mice, we found
183 significantly more EEG spikes in their frontal cortex relative to their wild-type littermates. This
184 result indicates that *Nudt21* haploinsufficiency is alone sufficient to cause cortical
185 hyperexcitability, and suggests that *NUDT21* loss of function might increase seizure risk (***Figure***
186 ***3***).
187

188 ***NUDT21* depletion induces aberrant alternative polyadenylation and altered protein levels.**

189 The observations that individuals with *NUDT21*-spanning CNVs have intellectual disability (ID)

190 and that *Nudt21*$^{+/-}$ mice have learning deficits provides strong evidence that *NUDT21* loss of

191 function causes disease, but do not reveal the molecular pathology. Therefore, to understand how

192 partial loss of *NUDT21* function might sicken neurons and cause ID, we used shRNA to inhibit

193 *NUDT21* in human embryonic stem cell (ESC)-derived glutamatergic neurons and assessed them

194 for proteome and alternative polyadenylation (APA) dysregulation. To assess the proteome, we

195 used mass spectrometry, and to directly measure APA events, we used poly(A)-ClickSeq (PAC-

196 seq), an mRNA 3′-end sequencing method that allows for more accurate identification of

197 polyadenylation sites than standard RNA sequencing (Elrod et al., 2019; Routh et al., 2017). The

198 genotypes cluster by principle component in both assays, showing that both approaches are

199 reproducible (***Figure 4—figure supplement 1A & B***).

200 Like in the whole-brain extracts from *Nudt21*$^{+/-}$ mice, the human neurons had a 30%

201 reduction of CFIm25 protein, despite a greater reduction of *NUDT21* mRNA (***Figure 4A & B***).

202 This difference shows that there is also homeostatic stabilization of CFIm25 in human neurons.

203 Intriguingly, in addition to reduced *NUDT21* mRNA levels as expected, the PAC-seq data also

204 showed that *NUDT21* itself undergoes aberrant alternative polyadenylation following its

205 depletion (***Figure 4Ci***). These shortened *NUDT21* mRNAs have no stop codon, and are likely to

206 undergo non-stop decay, an mRNA decay mechanism that degrades mRNAs lacking an in-frame

207 stop codon (Frischmeyer et al., 2002; van Hoof et al., 2002).

208 As in the *NUDT21* CNV patient-derived lymphoblasts, *NUDT21* loss of function causes

209 3′ UTR shortening of *MECP2* in human neurons, as well as other genes strongly regulated by

210 *NUDT21* in cancer-cell-line assays, such as *VMA21*, *LAMC1*, and *PAK1* (***Figure 4Cii & iii***)

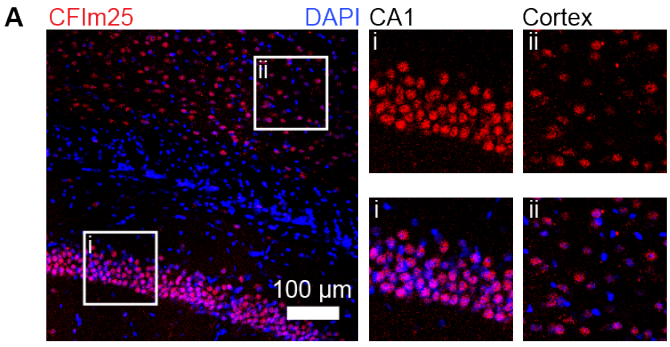
(Gennarino et al., 2015; Masamha et al., 2014). Also, similar to other cell-line assays, *NUDT21* loss of function causes widespread dysregulation of alternative polyadenylation, predominantly resulting in mRNA shortening (***Figure 4Ci-v & D***) (Brumbaugh et al., 2018; Gruber et al., 2012; Masamha et al., 2014). Of the misregulated genes, 230 have a gnomAD probability of loss-of-function intolerance over 0.9, 43 cause intellectual disability when mutated, and of those, 18 are autosomal or X-linked dominant (***Supplemental data 1***) (Lek et al., 2016; OMIM, 2019; Vissers et al., 2015). In addition to *MECP2*, misregulation of some of these genes is likely contributing to pathogenesis.

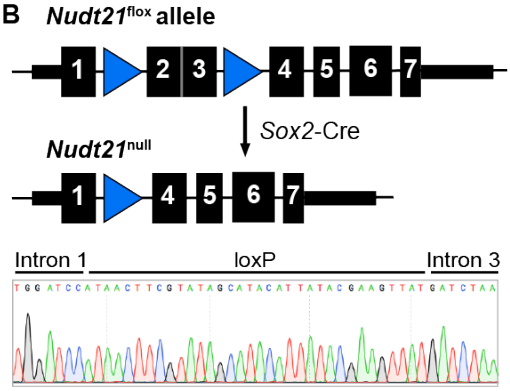
Notwithstanding disrupted RNA localization, the most apparent consequence of misregulated APA is altered protein levels (Tian and Manley, 2016). As expected, for many of the genes with significantly misregulated APA, we also see concordant changes in protein levels (***Figure 4E***). Most commonly, *NUDT21* loss results in shorter mRNAs and increased protein, such as with MeCP2 and VMA21, which respectively had protein-level increases of 50% and almost 300% (***Figure 4Cii-iii & E***). However, there are exceptions that further clarify the relationship between *NUDT21* loss and protein levels. Occasionally, the shortened mRNA will lose its stop codon and so likely undergo non-stop decay, resulting in reduced protein, such as with *ZCCHC6* (***Figure 4Civ & E***). Conversely, as with *KIF9*, *NUDT21* reduction occasionally induces synthesis of longer mRNAs, and this could paradoxically increase protein levels if the longer mRNAs gain a stop codon and cease to be degraded by non-stop decay (***Figure 4Cv & E***). Intriguingly, the degradation of KIF9 and others through non-stop decay in the control neurons shows an example of non-stop decay being used for normal gene regulation, rather than just as a quality control mechanism. We observed that ~15% of the genes with altered APA have a change in the percentage of mRNA isoforms undergoing non-stop decay. However, the majority

234 follow the trends seen in cancer cell lines: *NUDT21* loss of function in neurons causes
235 widespread 3′ UTR shortening and increased protein levels (***Figure 4D & E***).
236
237 **DISCUSSION**
238 Here we show that a partial loss of *Nudt21* function causes learning deficits and cortical
239 hyperexcitability in mice. Further, we show that partial loss of *NUDT21* function broadly
240 disrupts gene expression in human neurons via widespread misregulated alternative
241 polyadenylation and protein levels. Alongside our previous discovery that patients with
242 *NUDT21*-spanning deletions have intellectual disability and gnomAD data that *NUDT21* is a
243 highly constrained gene, our results provide strong evidence that partial loss of *NUDT21*
244 function causes intellectual disability (Gennarino et al., 2015; Lek et al., 2016).
245 Because a relatively small reduction in CFIm25 protein was sufficient to cause deficits,
246 these data suggest that individuals with missense variants in *NUDT21* that affect its function may
247 also have intellectual disability. Moreover, duplications of *NUDT21*, which we previously
248 showed are associated with ID, should lead to comparable dysregulation of alternative
249 polyadenylation and protein levels, including reduced MeCP2, and thus cause disease
250 (Gennarino et al., 2015).
251 Beyond providing insight into *NUDT21*-associated disease, these data provide useful
252 perspectives on the broader field of pediatric neurodevelopmental disease research. They
253 illustrate the importance of protein-level homeostasis. A mere 30% reduction in CFIm25 protein
254 was sufficient to cause learning deficits in mice, and mice are neurologically less sensitive to
255 genetic insult than humans (Tan and Zoghbi, 2018). Further, many genes associated with
256 neurodevelopmental diseases encode proteins that regulate or affect transcription (Rubeis et al.,

2014; Vissers et al., 2015; Yin and Schaaf, 2017). *NUDT21* provides another layer to this family of genes insofar as it broadly affects alternative polyadenylation and protein levels. As a group, these genes demonstrate how a partial loss or gain of function can result in large effects, and further show that neurons are particularly sensitive to protein-level disequilibrium.

Lastly, while the observed 50% increase of MeCP2 protein likely contributes to the symptoms seen in the *NUDT21* CNV patients, the transcriptional dysregulation in neurons following *NUDT21* loss is so widespread that there are probably other CFIm25 targets that mediate the patients' condition. Most likely, patient symptoms result from the additive effects of numerous misregulated genes, including the dozens that are ID associated. This then points to normalization of CFIm25 levels as the most viable molecular therapeutic strategy for treating this disease.



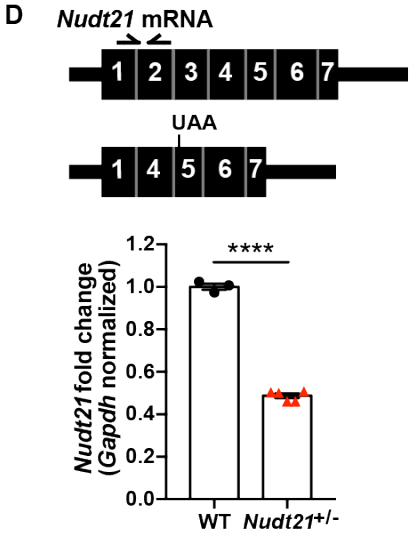


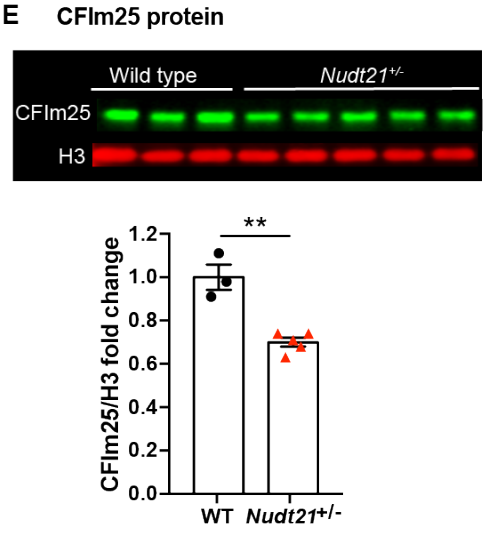
C

$Nudt21^{+/-}$ ♂ X $Nudt21^{+/-}$ ♀

P0 pups	Observed	Expected
$Nudt21^{+/+}$	12	8
$Nudt21^{+/-}$	20	16
$Nudt21^{-/-}$	0	8

df=2, $\chi^2(0.05)(2)=11$, P=0.004





270

Figure 1. *Nudt21* heterozygotes have 50% less *Nudt21* mRNA in their brains, but only 30% reduced CFIm25 protein. (A) Immunofluorescence showing CFIm25 expression in the (i) mouse hippocampus and (ii) cortex. **(B)** Schematic of the floxed and recombined *Nudt21* alleles (top), and sequencing showing successful recombination (bottom). **(C)** Observed and expected offspring counts born with each possible genotype from $Nudt21^{+/-}$ mating pairs. No homozygous *Nudt21* null offspring were born, when eight would be expected if loss of *Nudt21* did not affect survival: P=0.004, analyzed by two-tailed, chi-square test. **(D)** Schematic of wild-type and recombined *Nudt21* mRNA (top). Arrows indicate RT-qPCR primer binding sites, and UAA shows site of induced premature stop codon after recombination. RT-qPCR analysis shows expected 50% reduction of whole-brain, *Gapdh*-normalized, wild-type *Nudt21* mRNA in five-week-old mice with one wild-type *Nudt21* allele and one recombined, null allele (bottom): P<0.0001, n=3-5/genotype. **(E)** Western blot image comparing five-week-old $Nudt21^{+/-}$ mice CFIm25 protein levels with their WT littermates (top). Western blot analysis showing ~30% reduction of H3-normalized CFIm25 protein levels in $Nudt21^{+/-}$ mice: P=0.0012, n=3-5/genotype. We confirmed that CFIm25 does not regulate H3. For all charts, error bars indicate SEM. All data analyzed by unpaired, two-tailed t-test unless otherwise stated. **P<0.01; ****P<0.0001. Weights of the heterozygous animals are shown in ***Figure 1—figure supplement 1***.

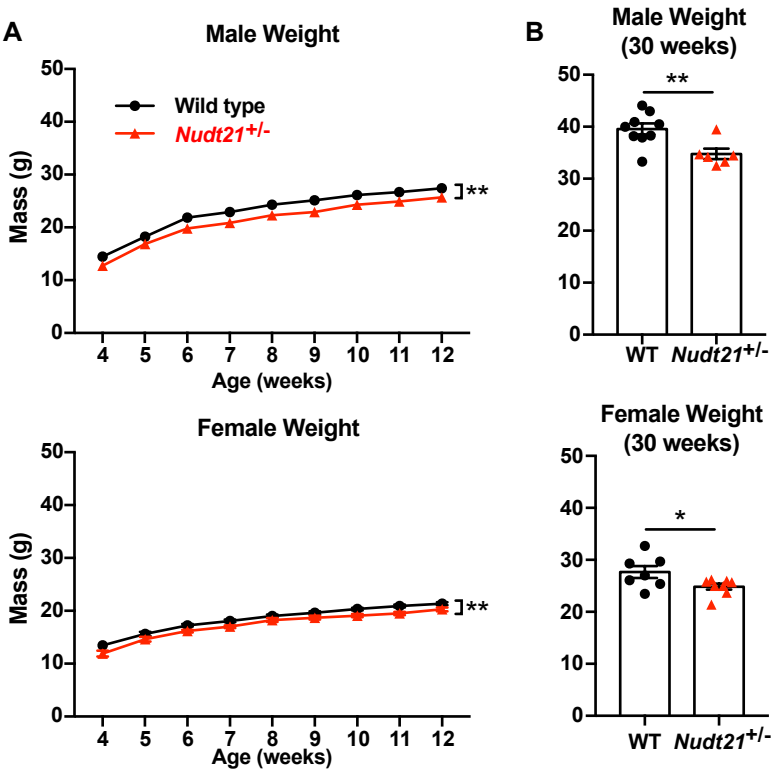
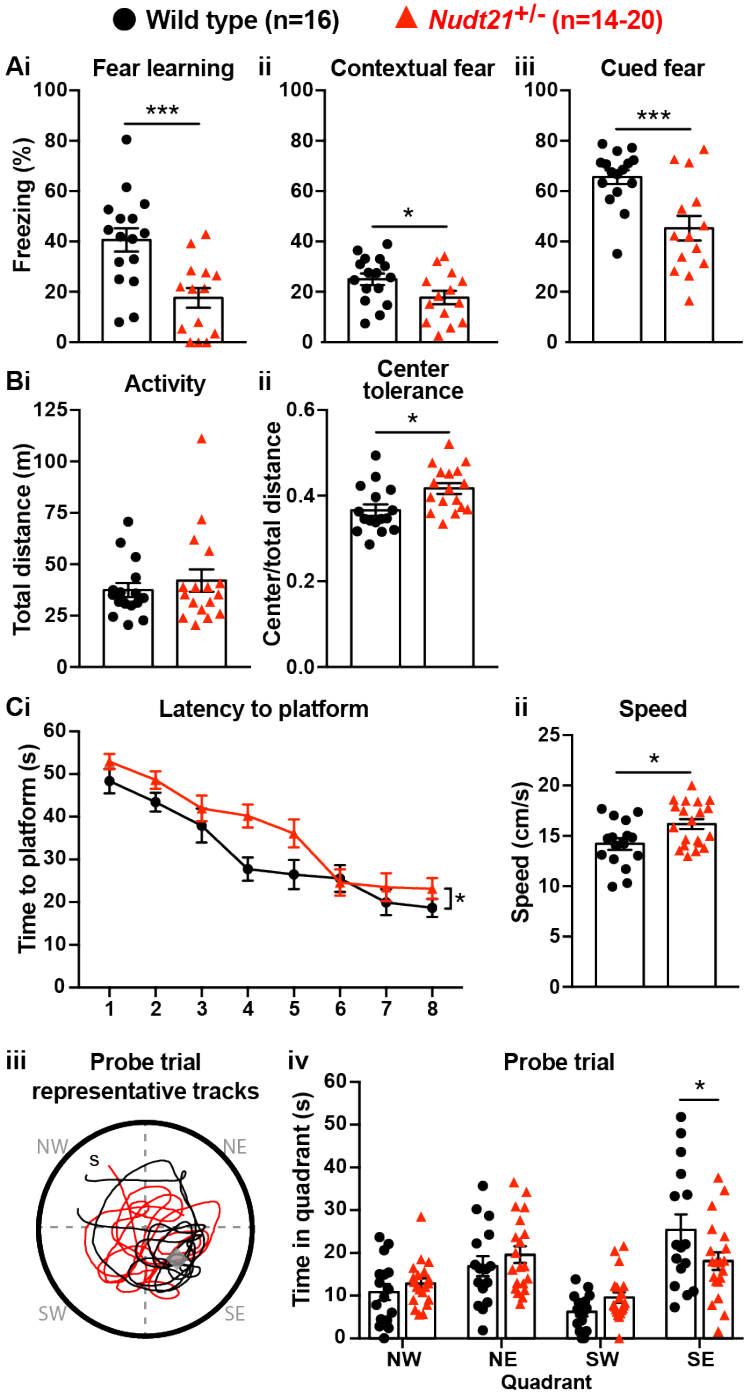


Figure 1—figure supplement 1. *Nudt21*$^{+/-}$ mice weigh less. (A) Male (top) and female (bottom) *Nudt21*$^{+/-}$ mice weigh less than their wild-type littermates from weening to twelve weeks: P=0.002 (males) and P=0.004 (females), analyzed by two-way repeated measures ANOVA (genotype*week). n=15-26/genotype. **(B)** Reduced weight persists up to at least 30 weeks in male (top) and female (bottom) *Nudt21*$^{+/-}$ mice: P=0.008 (males) and P=0.04 (females), analyzed by unpaired, two-tailed t-test. n=6-9/genotype. Error bars indicate SEM. *P<0.05; **P<0.01.

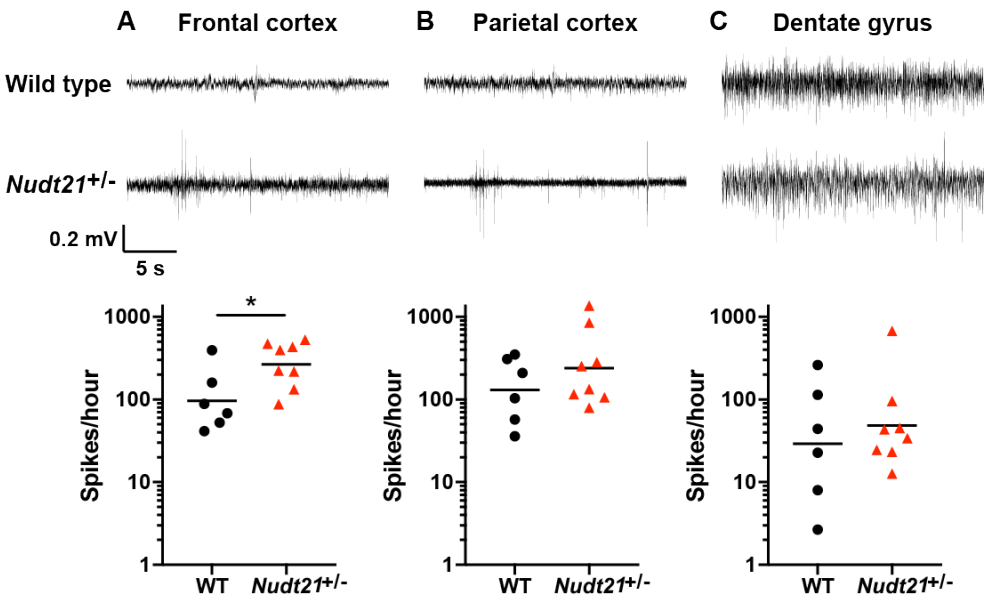


300

301

Figure 2. Partial loss of *Nudt21*-encoded CFIm25 protein causes learning deficits. (A) *Nudt21*$^{+/-}$ mice have conditioned fear learning deficits: fear learning, P=0.0009; contextual fear, P=0.045; cued fear, P=0.0009. **(B)** The open field assay shows that *Nudt21*$^{+/-}$ mice are no more active **(i)** but spend relatively more time in the center of the open field **(ii)**, indicating reduced anxiety: P=0.01. **(C)** *Nudt21*$^{+/-}$ mice have spatial learning deficits in the Morris water maze. **(i)** They take longer to find the hidden platform during the training blocks (P=0.026, two-way, repeated measures ANOVA), **(ii)** despite swimming faster and farther (P=0.013). **(iii & iv)** When the hidden platform is removed in the probe trial, they spend less time in the quadrant that previously had the platform (P=0.039, Sidak's multiple comparisons test). For all assays, mice were between 30-40 weeks of age. Error bars indicate SEM. Representative tracks are from the animal with the median result in each genotype. All data were analyzed by unpaired, two-tailed t-test unless otherwise stated. *P<0.05; ***P<0.001.

315



316

317

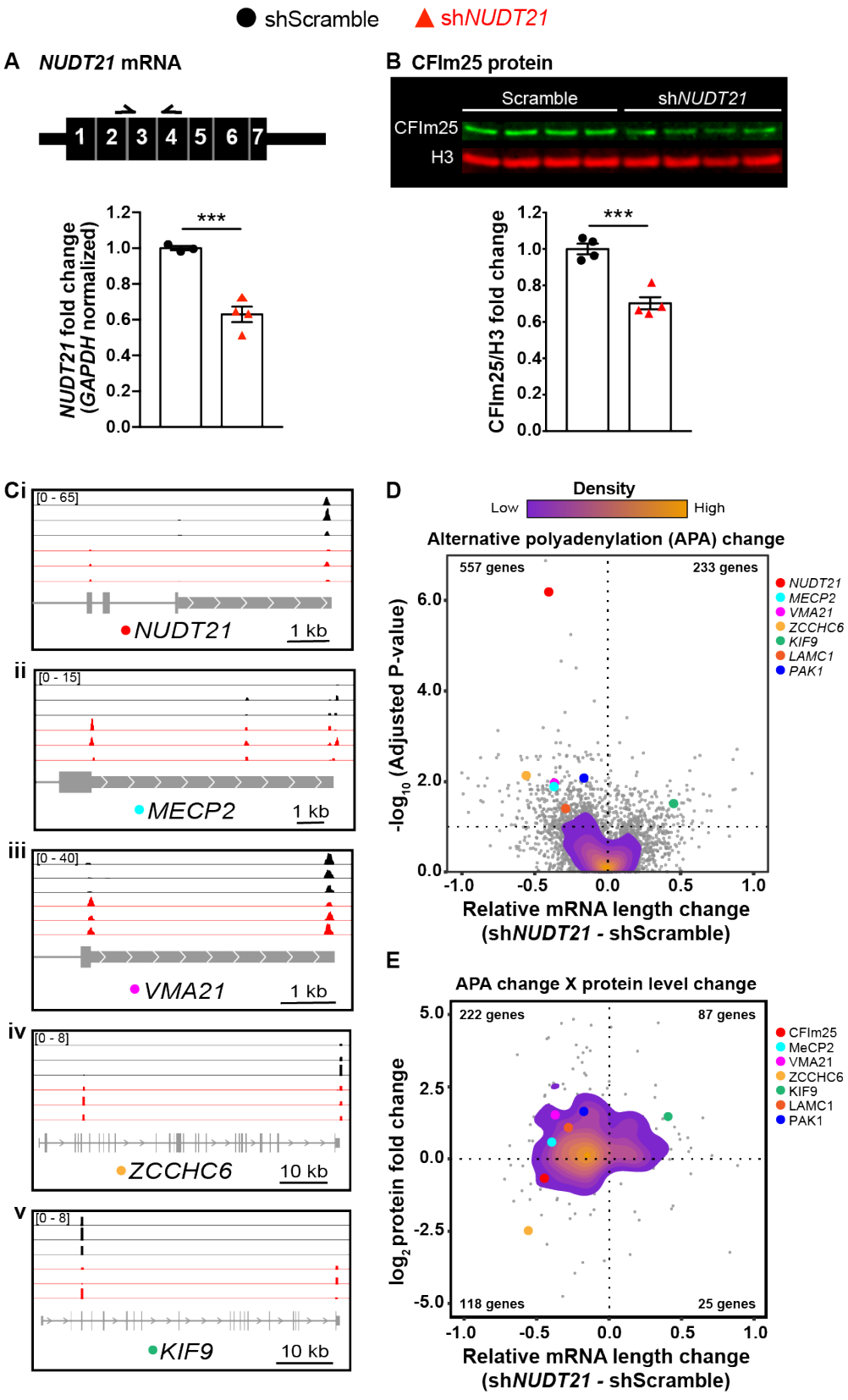
318 **Figure 3. *Nudt21*$^{+/-}$ mice have increased cerebral spike activity. (A)** 57-week-old *Nudt21*$^{+/-}$

319 mice have significantly increased spike activity in the frontal cortex by EEG (P=0.029), but not

320 in the parietal cortex (**B)** and dentate gyrus (**C**). Representative traces are on top and spike count

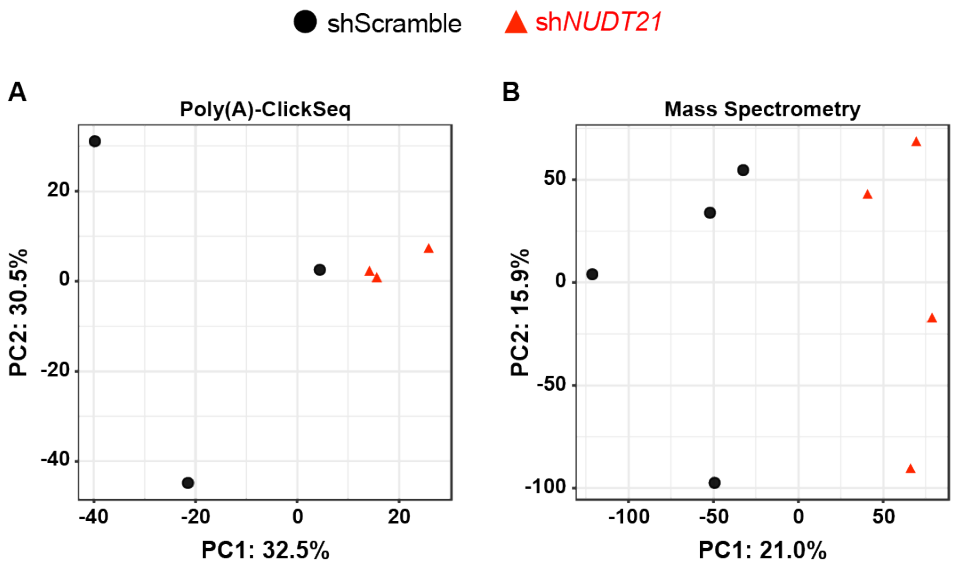
321 summaries below. All data analyzed by two-tailed Mann-Whitney test. Central tendency lines

322 show the geometric mean. *P<0.05.



323

Figure 4. *NUDT21* depletion induces aberrant alternative polyadenylation and altered protein levels. (A) Schematic of *NUDT21* mRNA with qPCR primers (top) and RT-qPCR quantification of *NUDT21* mRNA levels in human embryonic stem cell (ESC)-derived neurons infected with scrambled shRNA (shScramble) or shRNA targeting *NUDT21* (sh*NUDT21*). The neurons infected with sh*NUDT21* have a 40% reduction of *GAPDH*-normalized *NUDT21*: P=0.0009, n=3-4/treatment. **(B)** Western blot image and quantification showing a 30% reduction of H3-normalized CFIm25 protein in sh*NUDT21*-infected neurons: P=0.0005, n=4/treatment. We confirmed that CFIm25 does not regulate H3. **(C)** Poly(A)-ClickSeq (PAC-seq) tracks showing reduced *NUDT21* and altered alternative polyadenylation in example genes: n=3/treatment. Peaks are 3′ end sequencing reads. Multiple peaks in a gene indicate multiple mRNA isoforms with different cleavage and polyadenylation sites. Bracketed numbers show counts per million; kb stands for kilobases. **(D)** Volcano plot showing relative mRNA length change in sh*NUDT21*-infected neurons compared to shScramble-infected controls. The horizontal, dashed line shows $P_{adjusted}$=0.1, n=3/treatment. Principle component analysis shows sample clustering by treatment: (***Figure 4—figure supplement 1A***) **(E)** Mass spectrometry quantification of protein level fold change for genes with significantly altered APA ($P_{adjusted}$ <0.1). For bar charts, error bars indicate SEM and data analyzed by unpaired, two-tailed t-test. ***P<0.001. Principle component analysis shows sample clustering by treatment: (***Figure 4—figure supplement 1B***). Source files for the PAC-seq and mass spectrometry quantification data are available in ***Figure 4—source data 1***.



345 **Figure 4—figure supplement 1. Control and sh*NUDT21*-infected human neurons**
346 **transcriptomes and proteomes segregate by principle component analysis (PCA). (A)**
347 Control and sh*NUDT21*-infected human neurons segregate by principle component 1 (PC1) of
348 differentially expressed genes. Principle components (%): 32.5, 30.5, 16.9, 11.2, 8.8, 0. **(B)**
349 Control and sh*NUDT21*-infected human neurons segregate by PC1 of protein levels. Principle
350 components (%): 21, 15.9, 15.3, 13.6, 12.4, 11.1, 10.6, 0.

351

352 **Figure 4—source data 1. RNA length and protein level changes in genes with misregulated**
353 **APA following neuronal *NUDT21* inhibition.** Likelihood ratio (lr), degrees of freedom (df),
354 fold change (FC).

355

356 **Supplemental data 1. Intellectual disability associations of genes with misregulated APA**
357 **following neuronal *NUDT21* inhibition.** Probability of loss of function intolerance (pLI),
358 intellectual disability (ID), Online Mendelian Inheritance in Man (OMIM), autosomal recessive
359 (AR), autosomal dominant (AD), X-linked dominant (XLD), X-linked recessive (XLR)

METHODS

Generation of *Nudt21*$^{+/-}$ mice

The Baylor College of Medicine Institutional Animal Care and Use Committee approved all mouse care and manipulation (IACUC, protocol AN-1013). We generated C57BL/6J *Nudt21*$^{flox/flox}$ mice by inserting loxP sites flanking exons 2 and 3 of *Nudt21* by homologous recombination in embryonic stem cells (Ozgene) (Weng et al., 2019). We then crossed male *Nudt21*$^{flox/flox}$ mice with female C57BL/6J *Sox2*-Cre hemizygous mice to obtain *Nudt21*$^{+/-}$ mice. All oocytes from *Sox2*-Cre hemizygous females have Cre and can induce cre/lox recombination, even those haploid oocytes that do not have the Cre transgene (Hayashi et al., 2002). We confirmed successful cre/lox recombination by Sanger sequencing. To eliminate potential confounding from recombination mosaicism, we first crossed the *Nudt21*$^{+/-}$ F1 generation with wild-type C57BL/6J mice (Jackson Lab) before establishing mating pairs with F2s to generate experimental cohorts

Immunofluorescence

We dissected then hemisected brains from three seven-week-old *Nudt21*$^{flox/flox}$ C57BL/6J mice and drop-fixed their brains in 4% PFA overnight with gentle rocking at 4°C. After fixation, we cryoprotected the brains by gentle rocking in 30% sucrose solution in PBS at 4°C until the tissue sank. We then froze the brains in OCT, cryosected 40µm sections, and stored them at 4°C.

For immunolabeling, we first blocked free-floating sections in 5% normal goat serum, 0.5% Triton-X in PBS for one hour at room temperature. We then incubated the sections in CFIm25

primary antibody (1:50, Proteintech 10322-1-AP) in blocking buffer overnight at 4°C, followed by three washes and Alexa Fluor 488 (1:1000, Molecular Probes) secondary antibody incubation for two hours at room temperature. Lastly, we labeled the nuclei with DAPI (1:10,000) and mounted the sections with Vectashield. We used a Leica TCS SP5 confocal microscope for imaging and ImageJ Fiji for image processing. We confirmed the specificity of the CFIm25 antibody with si*NUDT21* knockdown in HeLa cells.

RT-qPCR

RNA extraction: We dissected and hemisected the brains of five-week-old *Nudt21*$^{+/-}$ mice and their wild-type littermates. We immediately flash froze them in liquid nitrogen and stored them temporarily at -80°C. We later homogenized the half brains in TRIzol Reagent (ThermoFisher Scientific) with the Polytron PT 10-35 GT (Kinematica). We then isolated the RNA by chloroform phase separation, precipitated it with 2-propanol, and washed it with 75% ethanol. We eluted the purified RNA in water.

We extracted RNA from shRNA-infected, human ESC-derived neurons in a 12-well plate: 4 wells non-silencing scrambled (RHS4348) and 8 wells sh*NUDT21* (V2LHS_253272 and V2LHS_197948) (Dharmacon). We lysed the neurons in the tissue-culture plate with TRIzol Reagent (ThermoFisher Scientific) and immediately transferred the lysate to microfuge tubes for trituration. We then isolated the RNA by chloroform phase separation, precipitation with 2-propanol, washing with 75% ethanol, and eluting in water as with the mice.

RT-qPCR: We synthesized first-strand cDNA with M-MLV reverse transcriptase (Life Technologies), and performed qPCR with PowerUp SYBR Green Master Mix (ThermoFisher Scientific) using the CFX96 Real-Time PCR Detection System (Bio-Rad). We designed exon-spanning primers using the UCSC genome browser mm10 and hg38 assemblies (http://genome.ucsc.edu/) (Kent et al., 2002), and Primer3 (http://bioinfo.ut.ee/primer3/) (Koressaar and Remm, 2007; Untergasser et al., 2012). We performed all RT-qPCR reactions in triplicate and determined relative cDNA levels by *NUDT21* threshold cycle (Ct) normalized to *GAPDH* Ct using the delta Ct method: relative expression (RQ) = 2^-(*NUDT21* average Ct - *GAPDH* average Ct). We present the data from human neurons infected with shRNA clone V2LHS_253272 because we used it for the PAC-seq experiment. We analyzed the data by two-tailed, unpaired t-test, and present it as mean ± SEM. *, **, ***, and **** denote P<0.05, P<0.01, P<001, P<0.0001.

Primers

Nudt21:

Forward: 5′- TACATCCAGCAGACCAAGCC - 3′

Reverse: 5′- AATCTGGCTGCAACAGAGCT - 3′

Gapdh:

Forward: 5′- GGCATTGCTCTCAATGACAA - 3′

Reverse: 5′- CCCTGTTGCTGTAGCCGTAT - 3′

NUDT21:

Forward: 5′- CTTCAAACTACCTGGTGGTG - 3′

Reverse: 5′- AAACTCCATCCTGACGACC - 3′

GAPDH:

Forward: 5′- CGACCACTTTGTCAAGCTCA - 3′

Reverse: 5′- TTACTCCTTGGAGGCCATGT - 3′

Western blot

Protein extraction: For the mice, we dissected and hemisected the brains of five-week-old *Nudt21*$^{+/-}$ mice and their wild-type littermates. We immediately flash froze them in liquid nitrogen and stored them temporarily at -80°C. We later homogenized the half brains with the Polytron PT 10-35 GT (Kinematica) in lysis buffer: 2% SDS in 100 mM Tris-HCl, pH 8.5, with protease and phosphatase inhibitors (ThermoFisher Scientific) and universal nuclease (Pierce). We incubated the lysates on ice for ten minutes, then rotated them for 20 minutes at room temperature. We spun the samples at top speed for 20 minutes to remove membrane, then quantified the protein levels with the Pierce Protein BCA Assay kit (ThermoFisher Scientific). For the ESC-derived neurons infected with shRNA targeting *NUDT21*, our extraction protocol was similar, except we lysed them directly in their 12-well plate and rocked them for 20 minutes at room temperature.

Western blot: We diluted the protein to 1 μg/μL in reducing buffer (LDS and sample reducing agent (ThermoFisher Scientific)) and ran 10 μg/sample. We imaged the membranes with the LI-COR Odyssey and analyzed the data with LI-COR Image Studio (LI-COR Biosciences),

451 comparing H3-normalized CFIm25 levels. We confirmed the specificity of the CFIm25 antibody
452 by sh*NUDT21* knockdown in HEK293T cells. We present the data from human neurons infected
453 with shRNA clone V2LHS_253272 because we used it for the PAC-seq experiment. We
454 analyzed the data by two-tailed, unpaired t-test, and present it as mean ± SEM. *, **, ***, and
455 **** denote P<0.05, P<0.01, P<001, P<0.0001.
456
457 *Antibodies:*
458 CFIm25: NUDT21 (2203C3): sc81109 (Santa Cruz)
459 H3: Histone H3 (D1H2) XP (Cell Signaling Technology)
460
461 **Mouse husbandry and handling**
462 The Baylor College of Medicine Institutional Animal Care and Use Committee approved all
463 mouse care and manipulation (IACUC, protocol AN-1013). We housed the mice in an
464 AAALAS-certified level three facility on a 14-hour light cycle. The mice had ad libitum access
465 to standard chow and water. We weighed a cohort of mice weekly from 4-12 weeks and a
466 different cohort at 30 weeks. We tested both male and females in all experiments and only
467 present the weight data separately because we never otherwise detected a difference between
468 them. We were blinded to the mice's genotype during all handling.
469
470 *Genotyping*
471 We determined the mice's genotypes by PCR amplification of tail lysates.
472
473 Primers:

474 *Nudt21*^{null}

475 Forward: 5′- ACAGATTAGCTGTTAGTACAGG - 3′

476 Reverse: 5′- GAAGAACCAGAGGAAACGTGAG - 3′

477

478 Wild-type *Nudt21*

479 Forward: 5′- AGGAGGCTGACATGGATTGTT - 3′

480 Reverse: 5′- TCTTCTCCTGGGTTAAGTTCCC - 3′

481

482 **Behavioral tests**

483 Because neurodevelopmental disease loss-of-function mouse models typically have more
484 pronounced phenotypes when they are older, we started our behavioral battery on 30-week-old
485 mice. We performed the assays on two cohorts. We tested the mice during their light cycle,
486 typically between 11 AM and 5 PM. Prior to each assay, we habituated the mice to the testing
487 facility for 30-60 minutes. The investigators were blind to the mice's genotypes during all
488 assays.

489

490 ***Conditioned fear***

491 We first habituate the mice for 30 minutes in an adjacent room on each day of the test. On day
492 one, we conditioned the mice by placing them in the habitest operant cage (Coulbourn) for a
493 training session. The training consists of two minutes habituation, then a 30 second 85 dB tone
494 followed by a foot shock of 1.0 mA for two seconds. After another two minutes, the 30 second
495 85 dB tone is played again (the training day cue). Throughout the experiment, except for the two
496 seconds during the foot shock, the FreezeFrame3 system (Coulbourn/Actimetrics) recorded the

497 mice's movement and freezing episodes. On day two, we performed the contextual and cued fear
498 assays. For contextual fear, we returned the mice to the test chambers precisely as we had done
499 during the training, and recorded their freezing in the chamber for five minutes. We waited two
500 hours before beginning the cued fear assay. We first changed the holding cages and test chamber
501 shape, color, texture, scent, and lighting to make the experience as unrecognizable as possible to
502 the mice. We then placed them in the modified chamber, and after three minutes played the 30
503 second 85 dB tone, then recorded their freezing for the following three minutes. We analyzed the
504 difference in freezing between the two groups after the second sound cue (training cue) on day
505 one, throughout the contextual fear test, and after the sound cue in the cued fear test, each by
506 unpaired, two-tailed t-test. We excluded data for all the mice from one contextual fear trial. We
507 suspect there was a technical error in the collection of those data: the FreezeFrame3 system
508 recorded them as freezing far more than the mice in any other trial (50-80%) and two of the three
509 were significant outliers in the Grubbs test.
510
511 ***Open field***
512 We lit the room to 200 lux and set the ambient white noise to 60 dB during habituation and
513 throughout the test. We placed each mouse in the open field, a 40 x 40 x 30 cm chamber
514 equipped with photobeams (Accuscan Instruments), and recorded their activity for 30 minutes.
515 We analyzed total distance and center tolerance (center distance/total distance) by unpaired, two-
516 tailed t-test.
517
518 ***Morris water maze***

519 Our Morris water maze experiment took pl ace in a 120 cm diameter pool of water. We hid a 10
520 cm X 10 cm platform 0.5-1 cm underwater in the Southeast quadrant. On each wall of the testing
521 room, we taped brightly colored shapes that the mice can use for orientation. Our experiment
522 spanned four days. Each day, we set the lighting to 60 lux and habituated the mice in the testing
523 room in their home cages for 30 minutes, then ten minutes in holding cages. Prior to the first
524 day's experiment, we introduced the mice to the invisible platform by placing them on it for ten
525 seconds. We next pulled the mice into the water and let them swim for ten seconds to ensure they
526 could swim, then placed them directly in front of the platform to confirm they could climb back
527 on to it. We tested the mice in two training blocks per day for the four days. Each training block
528 consisted of four trials. For each trial, we placed the mice in a different quadrant of the pool
529 (North, South, East, or West); the quadrant order was the same for every mouse in the trial, but
530 different for every trial. We removed the mice after they found the platform, or if they did not
531 find that platform, we guided them to the platform to rest on it for ten seconds before removing
532 them. We used an EthoVision XT automated video tracking system (Noldus Information
533 Technology) to track the mice's location, speed, and latency to find the platform. After the
534 second training block on the fourth day, we immediately performed the probe trial: we removed
535 the platform and placed the mice in a new location, the Northwest quadrant, and tracked them for
536 60 seconds. We analyzed their speed by unpaired, two-tailed t-test; their latency to find the
537 platform by two-way, repeated measures ANOVA (genotype*block); and their time in each
538 quadrant in the probe trial by two-way, repeated-measures ANOVA (genotype*quadrant).
539
540 ***Animal behavior statistical analysis***

From previous experience, we know a sample size of 14 animals is sufficient to detect meaningful phenotypic differences in a neurobehavioral battery (Chao et al., 2010; Lu et al., 2017; Samaco et al., 2012). We analyzed all the data with Prism 7 (Graphpad), following Graphpad's recommendations. We used unpaired, two-tailed t-tests for all simple comparisons, and two-way repeated measures ANOVAs for all two-factor comparisons. We present all data as mean ± SEM. *, **, ***, and **** denote P<0.05, P<0.01, P<0.001, P<0.0001.

Video electroencephalography (EEG) and spike counting

Surgery and data recordings: The Baylor College of Medicine Institutional Animal Care and Use Committee approved all research and animal care procedures. We tested eight *Nudt21*$^{+/-}$ mice and six wild-type littermate controls. Experimenters were blind to the mouse genotype. We secured 54-week-old mice on a stereotaxic frame (David Kopf) under 1-2% isoflurane anesthesia. Each mouse was prepared under aseptic condition for the following recordings: the cortical EEG recording electrodes of Channels 1 and 2 were made of Teflon-coated silver wires (bare diameter 127 µm, A-M systems) and implanted in the subdural space of the parietal cortex and frontal cortex, respectively, with reference at the midline over the cerebellum. The electrode of the third channel, made of Teflon-coated tungsten wire (bare diameter 50 µm, A-M systems) was stereotaxically aimed at the hippocampal dentate gyrus (1.9 mm posterior, 1.7 mm lateral, and 1.8 mm below the bregma) with reference in the ipsilateral corpus callosum (Paxinos and Franklin, 2001). In addition, Teflon-coated silver wires were used to record the electromyogram (EMG) in the neck muscles to monitor mouse activity. All of the electrode wires together with the attached miniature connector sockets were fixed on the skull by dental cement. After two weeks of post-surgical recovery, mice received three two-hour sessions of EEG/EMG recordings

564 over a week. Signals were amplified (100x) and filtered (bandpass, 0.1 Hz - 1 kHz) with the
565 1700 Differential AC Amplifier (A-M Systems), then digitized at two kHz and stored on disk for
566 off-line analysis (DigiData 1440A and pClamp10, Molecular Devices). The time-locked mouse
567 behavior was recorded by ANY-maze tracking system (Stoelting Co.).
568
569 *EEG data analysis*: Abnormal synchronous discharges were manually identified when the sharp
570 positive deflections exceeding twice the baseline and lasting 25-100 ms (Roberson et al., 2011).
571 We counted the number of abnormal spikes over the recording period using Clampfit 10 software
572 (Molecular Devices, LLC) and averaged the spike numbers across sessions for each animal.
573 Since the data follow a lognormal distribution, we statistically compared the genotypes with the
574 Mann-Whitney test using Prism 7 (Graphpad). The measure of central tendency is the geometric
575 mean and * indicates $P<0.05$.
576
577 **shRNA lentivirus production and titer assessment**
578 *Virus production*. We made several viruses that express shRNAs targeting *NUDT21*. We
579 transfected 45 ug DNA into 80%–90% confluent, low-passage HEK293T cells (ATCC CRL-
580 3216; RRID:CVCL_0063) in 150 mm dishes at a 4:3:1 ratio of pGIPz, psPAX2, pMD2.G with
581 TransIT-293 transfection reagent (Mirus, MIR 2706). The following day, we changed the media
582 to 10 mL. At 48 and 72 hours, we collected and pooled their media, then centrifuged at 4000 x g
583 for ten minutes and filtered the supernatant through a Poly-ethersulfone filter (VWR, 28145-505)
584 to remove cellular debris. We concentrated the virus 100-fold with Lenti-X concentrator
585 (Clontech, 631231) following the manufacturer's recommendations before aliquoting and
586 freezing at -80C.

587

588 *Titer assessment*.

589 We measured the viral titer using Open Biosystems pGIPZ method (Thermo Fisher Scientific).
590 We plated 5 X 10^4 HEK293T cells in a 24-well plate. The following day, we made a serial
591 dilution of the virus in a 96-well plate, and when the HEK293T cells reached ~50% confluency,
592 we infected them with the diluted virus. We cultured the cells for two days, then counted the
593 tGFP colonies with Axiovert 25 microscope (Zeiss) microscope X-cite 120 lamp (ExFo) to
594 determine the viral titer. Our viruses had 10^9 transducing units/mL.

595

596 *Confirmation*

597 Compared to non-silencing scrambled control virus (RHS4348, Dharmacon), we confirmed
598 *NUDT21* knockdown efficacy in HEK293T cells (ATCC CRL-3216; RRID:CVCL_0063), and
599 selected the two most efficient shRNAs for our studies: V2LHS_197948 and V2LHS_253272
600 (Dharmacon).

601

602 **Human embryonic stem cell (hESC)-derived neuron culture**

603 We used WA09 (H9; RRID:CVCL_9773) female embryonic stem cells (ESCs) to generate
604 human neurons as previously described (Jiang et al., 2019). We differentiated neural progenitors
605 into human neurons over three weeks, changing the media every 3 d. Afterwards, we passaged
606 the neurons with trypsin. Three days after passaging, we infected the neurons with lentiviruses
607 containing pGIPZ shRNA clones at a multiplicity of infection of 10. We verified the tropism and
608 infectivity of the virus using the tGFP reporter signal. At day 3 after infection, we treated the
609 neurons with puromycin (0.75–1.25 g/ml) for 6 days to select for infected cells. We cultured the

610 cells for 60 days after infection, changing the media three times per week. We then aspirated all
611 the media and washed the cells with PBS before freezing them at -80C for later RNA and protein
612 extraction.
613
614 **Poly(A) click-seq**
615 *RNA extraction*
616 We extracted RNA from shRNA-infected, human ESC-derived neurons in a 12-well plate: 4
617 wells non-silencing scrambled (RHS4348) and 8 wells sh*NUDT21* (V2LHS_253272 and
618 V2LHS_197948) (Dharmacon). We know from past experience that we need at least three
619 samples for RNA- and PAC-seq analysis (Elrod et al., 2019; Maio et al., 2018; Routh et al.,
620 2017; Tan et al., 2016). We prepared four samples per genotype to allow for loss of one sample.
621 We lysed the neurons in the tissue-culture plate with TRIzol Reagent (ThermoFisher Scientific)
622 and immediately transferred the lysate to microfuge tubes for trituration. We then isolated the
623 RNA by chloroform phase separation, precipitation with 2-propanol, washing with 75% ethanol,
624 and eluting in water.
625
626 *Library preparation and sequencing*
627 We prepared sequencing libraries as previously described (Routh et al., 2017). We reverse
628 transcribed 1 ug of total RNA with the partial P7 adapter (Illumina_4N_21T) and dNTPs with
629 the addition of spiked-in azido-nucleotides (AzVTPs) at 5:1. We click-ligated the p5 adapter
630 (IDT) to the 5′ end of the cDNA with CuAAC. We then amplified the cDNA for 21 cycles with
631 Universal primer and 3′ indexing primer and purified it on a 2% agarose gel by extracting

amplicon from 200-300 base pairs. We pooled the libraries and sequenced single-end, 75 base-pair reads on a Nextseq 550 (Illumina).

P7 adapter (Illumina_4N_21T):

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNTTTTTTTTTTTTTTTTTTT)

P5 adapter (IDT):

5′HexynylNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGT CGCCGTATCATT

Universal primer:

AATGATACGGCGACCACCGAG

Example 3′ indexing primer:

CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGT

PAC-seq data analysis

We sequenced each library to a depth of ~20 million, 75 base pair (bp), single-end reads. For each sample, we obtained raw reads in a zipped fastq format. We used fastp for initial quality control (Chen et al., 2018). We filtered adapter contamination (AGATCGGAAGAGC) using the -a option. We trimmed the first 6 nucleotides and any PolyG tail nucleotides with the -f and -g options. And removed reads shorter than 40 nucleotides (nt) after trimming, using the -l option.

Read alignment

We downloaded raw FASTA sequences and annotations of the Human genome build GRCh38 from the UCSC table browser tool (Karolchik et al., 2004), and aligned trimmed reads to the reference genome with Bowtie 2 version 2.2.6 with parameters: -D 20 -R 3 -N 0 -L 20 -i S,1,0.50 (very-sensitive-local) (Langmead and Salzberg, 2012). We indexed the reference genome using bowtie2-build default settings, and saved sample-wise alignments as Sequence Alignment Map (SAM) files. We then used SAMtools V0.1.19 "*view"*, "*sort*" and "*index*" modules to convert SAM files to Binary Alignment Maps (BAM), coordinate sort, and index (Li et al., 2009).

Sample quality

We obtained gene level counts as the sum of total reads mapped to their respective genes. shRNA clone V2LHS_253272 gave more consistent *NUDT21* knockdown, so we used those samples in our analysis. We sequenced four samples from both the control and *NUDT21* knockdown conditions, but excluded one control sample with <10 million reads, and one knockdown sample with unusually low *NUDT21* levels (<2% of other knockdown samples). We then performed principle component analysis (PCA) on gene counts as previously described and confirmed that the treatment groups clustered (***Figure 4—figure supplement 1A***) (Yalamanchili et al., 2017).

Alternative poly-adenylation (APA) analysis

We computed strand-specific coverage of the features (mRNA cleavage sites) from BAM files using bedtools v2.25.0 (Quinlan and Hall, 2010) "*genomecov"* module. We pooled sample-wise feature files to get a comprehensive list of polyadenylation (p(A)) sites across all the samples.

We merged overlapping features and features within 15 nucleotides using the bedtools "*merge*" module. We mapped features to genes if they were within the gene body or within 16 kilobases (kb) downstream of the annotated transcription end site and do not overlap with any other gene, which is far enough to detect the longest known 3′UTRs (Miura et al., 2013; Quinlan and Hall, 2010). We then mapped these features to known, annotated human polyadenylation sites downloaded from PolyA_DB3 (Wang et al., 2017). Because the annotations were in the hg19 coordinate system, we converted them to the GRCh38 coordinate system with the liftOver program from the UCSC genome browser portal. With a read length of 75 nucleotides, we may not have reached the actual cleavage sites. Thus, each annotated polyadenylation site was considered a 70-nucleotide feature for mapping our features to annotated p(A) sites. Our features that did not overlap with at least 1 bp of any of the annotated p(A) sites were considered novel p(A) sites.

For calling novel p(A) sites, we used a sliding window approach to filter out any potential misprimed sites, i.e. potential cleavage sites where the poly-d(T) reverse transcription primer bound genomic adenines in the body of the mRNA rather than the poly(A) tail. We filtered out features with more than 12 genomic adenines in a 20 bp window 10 nt up- and 50 nt downstream of the potential cleavage site.

We quantified resultant novel and annotated p(A) sites across all samples using featureCounts (Liao et al., 2013). To minimize potential bias from under-covered p(A) sites, we retained only features with at least 5 reads in two samples and 1 read in the remaining sample in either of the conditions. We also filtered out p(A) sites accounting for less than 10% of the reads mapped to a

gene in both the control and *NUDT21* knockdown conditions. We considered genes with mean read counts <5 in either condition not expressed and so excluded them.

Because about half of genes that undergo APA have greater than two p(A) sites, we wanted to use a statistic that could account for changes in all of the p(A) sites (Derti et al., 2012). Thus, we identified changes in alternative polyadenylation site usage using the Dirichlet multinomial test in the DRIMSeq package, and computed adjusted p-values as previously described (Nowicka and Robinson, 2016; Zhao et al., 2013). We quantified the magnitude and direction of change in polyadenylation site usage as an mRNA lengthening score. We summed the weighted read counts per million (cpm) from each p(A) site in a gene to get that gene's mRNA length score. We assigned the weights in decreasing order from the most distal p(A) site as below:

$$W_i = 1 - \left(\left(\frac{1}{NS-1}\right)(i-1)\right)$$

where W_i is the weight of i^{th} distal polyadenylation site of a gene and NS is the of total number of polyadenylation sites mapped to it. Thus, in a gene with 3 p(A) sites, the cpm at the most promoter-distal site would all be counted (weight of 1) in the length score, the cpm at the intermediate site would be weighted by 0.5, and the cpm at the proximal site would not be counted (weight of 0). We then averaged the scores across samples and took the difference of treated to controls for the relative mRNA length change.

Disease association determination

For the probability of loss of function intolerance (pLI) of all the genes, we used gnomAD (Lek et al., 2016). In addition, we cross-referenced the misregulated genes with published genetic studies to see which were known to cause intellectual disability when mutated, then looked to the

724 Online Mendelian Inheritance in Man (OMIM) database to see if the pathological variants were
725 dominant or recessive (OMIM, 2019; Vissers et al., 2015).
726
727 **Mass spectrometry**
728 We processed, measured, and analyzed the sample as previously described (Saltzman et al.,
729 2018). We summarize the main steps below.
730
731 *Lysis and Digestion*
732 We pelleted and lysed the shRNA-infected human neurons with three freeze (LN2) and thaw
733 (42ºC) cycles in 50 µL of ammonium bicarbonate + 1mM $CaCl_2$. We then boiled the lysate at
734 95°C for two minutes with vortexing at 20 second intervals, and then centrifugation at 21,000 rcf.
735 We digested 50 µg of total protein with a 1:20 solution of 1 µg/µL trypsin:protein overnight at
736 37°C with shaking and then again with a 1:100 solution of 1 µg/µL trypsin:protein for 4 hours.
737 We extracted the peptides with 80% acetonitrile + 0.1% formic acid solution, followed by
738 centrifugation at 10,000 g, and vacuum drying.
739
740 *Off-Line Basic pH Reverse Phase Peptide Fractionation*
741 A "15F5R" protocol was used for off-line fractionation as described before(Saltzman et al., 2018).
742 We filled a 200 µl pipette tip with 6 mg of C18 matrix (Reprosil-Pur Basic C18, 3 µm, Dr.
743 Maisch GmbH) on top of a C18 disk plug (EmporeTM C18, 3M). We dissolved 50 µg of
744 vacuum-dried peptides with 150 µl of pH10 ABC buffer and loaded it onto the pre-equilibrated
745 C18 tip. We eluted bound peptides with fifteen 2%-step 2-30% gradient of ACN non-
746 contiguously combined into five pools, and then vacuum dried.

747

748 *Mass Spectrometry*

749 We analyzed fractionated peptides on an Orbitrap Fusion mass spectrometer coupled with the
750 Nanospray Flex ion sources and an UltiMate 3000 UPHLC (Thermo Fisher Scientific). For each
751 run, we loaded approximately one microgram of peptide onto a two cm 100 µm ID pre-column
752 and resolved it on a twelve cm 100 µm ID column, both packed with sub-twi µm C18 beads
753 (Reprosil-Pur Basic C18, Catalog #r119.b9.0003, Dr. Maisch GmbH). We maintained a constant
754 flow rate for 100-minute 2-28% B gradient elutions, where A is water and B is 90% acetonitrile,
755 both with 0.1% formic acid.

756

757 *Proteome Discoverer (Mascot-based) Search and Protein Inference/Quantification*

758 We used the Proteome Discoverer software suite (PD version 2.0.0.802; Thermo Fisher
759 Scientific) to search the raw files with the Mascot search engine (v2.5.1, Matrix Science),
760 validate peptides with Percolator (v2.05), and provide MS1 quantification through Area Detector
761 Module (Käll et al., 2007; Perkins et al., 1999). We matched MS1 precursors in a 350-10,000
762 mass range against the tryptic RefProtDB database digest (2015-06-10 download) with Mascot
763 permitting up to 2 missed cleavage sites (without cleavage before P), a precursor mass tolerance
764 of 20 ppm, and a fragment mass tolerance of 0.5 Da. We allowed the following dynamic
765 modifications: Acetyl (Protein N-term), Oxidation (M), Carbamidomethyl (C), DeStreak (C), and
766 Deamidated (NQ). For the Percolator module, we set the target strict and relaxed FDRs for PSMs
767 at 0.01 and 0.05 (1% and 5%), respectively. We used gpGrouper (v1.0.040) for gene product
768 inference and label-free iBAQ quantification with shared peptide distribution.

769

Data Normalization

Because *NUDT21* inhibition consistently increases protein levels across the proteome, normalizing to total protein levels in each sample would introduce bias. Instead, we normalized each sample to a group of ~20 consistently expressed genes that we know are not directly regulated by *NUDT21* from the PAC-seq data.

Analysis

We performed a PCA as previously described to confirm that the treatment groups separated (***Figure 4—figure supplement 1B***) (Yalamanchili et al., 2017). We then computed protein level changes as $\log_2$ fold change values and plotted them against their relative mRNA length change (***Figure 4E***).

Data availability

The PAC-seq data are available in the NCBI Gene Expression Omnibus, accession number GSE135384. The analysis code can be found at http://liuzlab.org/iNeuron_PACSeq.zip. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD014842 (Perez-Riverol et al., 2018).

ETHICS

For animal experimentation, we housed up to five mice per cage on a 12-hour light cycle in a level 3, AALAS-certified facility and provided water and standard rodent chow ad libitum. The

Institutional Animal Care and Use Committee for Baylor College of Medicine and Affiliates approved all procedures carried out in mice under protocol AN-1013.

ACKOWLEDGEMENTS

We would like to thank the Zoghbi and Wagner lab members for critical feedback, particularly Nathan Elrod for his suggestions on analyzing the PAC-seq data. We would also like to thank the cores that provided services for the project: the Microscopy, Neuroconnectivity, Animal Behavior, and Human Neuronal Differentiation cores from the Jan and Dan Duncan Neurological Research Institute and BCM Intellectual and Developmental Disabilities Research Center (NIH U54 HD083092 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development); the Next Generation Sequencing core at the University of Texas Medical Branch; and BCM Mass Spectrometry Core (supported by NIH P30 CA125123 and CPRIT RP170005). Lastly, we would like to thank our funders who made the work possible: the NIH National Institute of Neurological Disorders and Stroke—F30NS095449 (CEA), the Howard Hughes Medical Institute (HYZ), and the Intellectual and Developmental Disabilities Research Center—NIH U54 HD083092 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (HYZ). The content is solely the responsibility of the authors and does not represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Institutes of Health, or the Howard Hughes Medical Institute.

REFERENCEES

Amendola E, Zhan Y, Mattucci C, Castroflorio E, Calcagno E, Fuchs C, Lonetti G, Silingardi D, Vyssotski AL, Farley D, Ciani E, Pizzorusso T, Giustetto M, Gross CT. 2014. Mapping Pathological Phenotypes in a Mouse Model

of CDKL5 Disorder. *Plos One* **9**:e91613. doi:10.1371/journal.pone.0091613

Amir RE, den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics* **23**:185–188. doi:10.1038/13810

Brown KM, Gilmartin GM. 2003. A Mechanism for the Regulation of Pre-mRNA 3′ Processing by Human Cleavage Factor Im. *Mol Cell* **12**:1467–1476. doi:10.1016/s1097-2765(03)00453-2

Brumbaugh J, Stefano B, Wang X, Borkent M, Forouzmand E, Clowers KJ, Ji F, Schwarz BA, Kalocsay M, Elledge SJ, Chen Y, Sadreyev RI, Gygi SP, Hu G, Shi Y, Hochedlinger K. 2018. Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling. *Cell* **172**. doi:10.1016/j.cell.2017.11.023

Chao H-T, Chen H, Samaco RC, Xue M, Chahrour M, Yoo J, Neul JL, Gong S, Lu H-C, Heintz N, Ekker M, Rubenstein JL, Noebels JL, Rosenmund C, Zoghbi HY. 2010. Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468**:263–269. doi:10.1038/nature09582

Chao H-T, Zoghbi HY. 2012. MeCP2: only 100% will do. *Nature Neuroscience* **15**:176–177. doi:10.1038/nn.3027

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinform Oxf Engl* **34**:i884–i890. doi:10.1093/bioinformatics/bty560

Crawley JN. 1985. Exploratory behavior models of anxiety in mice. *Neurosci Biobehav Rev* **9**:37–44. doi:10.1016/0149-7634(85)90030-2

Derti A, Garrett-Engele P, MacIsaac KD, Stevens RC, ram S, Chen R, Rohl CA, Johnson JM, Babak T. 2012. A quantitative atlas of polyadenylation in five mammals. *Genome Res* **22**:1173–1183. doi:10.1101/gr.132563.111

Elrod NR, Jaworski EA, Ji P, Wagner EJ, Routh A. 2019. Development of Poly(A)-ClickSeq as a Tool Enabling Simultaneous Genome-wide Poly(A)-site identification and Differential Expression Analysis. *Methods*. doi:10.1016/j.ymeth.2019.01.002

Firth HV, Richards SM, Bevan PA, Clayton S, Corpas M, Rajan D, Vooren S, Moreau Y, Pettett RM, Carter NP. 2009. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *The American Journal of Human Genetics* **84**:524–533. doi:10.1016/j.ajhg.2009.03.010

Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, Dietz HC. 2002. An mRNA Surveillance Mechanism That Eliminates Transcripts Lacking Termination Codons. *Science* **295**:2258–2261.

doi:10.1126/science.1067338

Gennarino VA, Alcott CE, Chen C-AA, Chaudhury A, Gillentine MA, Rosenfeld JA, Parikh S, Wheless JW, Roeder ER, Horovitz DD, Roney EK, Smith JL, Cheung SW, Li W, Neilson JR, Schaaf CP, Zoghbi HY. 2015. NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. *eLife* **4**. doi:10.7554/eLife.10782

Gruber AR, Martin G, Keller W, Zavolan M. 2012. Cleavage factor Im is a key regulator of 3' UTR length. *RNA biology* **9**:1405–12. doi:10.4161/rna.22570

Guvenek A, Tian B. 2018. Analysis of alternative cleavage and polyadenylation in mature and differentiating neurons using RNA-seq data. *Quantitative Biology* 1–14. doi:10.1007/s40484-018-0148-3

Hayashi S, Lewis P, Pevny L, McMahon AP. 2002. Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. *Mech Develop* **119**:S97–S101. doi:10.1016/s0925-4773(03)00099-6

Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park J, Yehia G, Tian B. 2013. Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. *Nature Methods* **10**:133. doi:10.1038/nmeth.2288

Ji Z, Lee J, Pan Z, Jiang B, Tian B. 2009. Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proceedings of the National Academy of Sciences* **106**:7028–7033. doi:10.1073/pnas.0900028106

Jiang X, Zhang C, Chen J, Choi S, Zhou Y, Zhao M, Song X, Chen X, Maletić-Savatić M, Palzkill T, Moore D, Wang MC, Wang J. 2019. Quantitative Real-Time Imaging of Glutathione with Subcellular Resolution. *Antioxid Redox Sign* **30**:1900–1910. doi:10.1089/ars.2018.7605

Jiang Y, Armstrong D, Albrecht U, Atkins CM, Noebels JL, Eichele G, Sweatt JD, Beaudet AL. 1998. Mutation of the Angelman Ubiquitin Ligase in Mice Causes Increased Cytoplasmic p53 and Deficits of Contextual Learning and Long-Term Potentiation. *Neuron* **21**:799–811. doi:10.1016/s0896-6273(00)80596-6

Käll L, Canterbury JD, Weston J, Noble W, MacCoss MJ. 2007. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **4**:nmeth1113. doi:10.1038/nmeth1113

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent JW. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**:D493–D496. doi:10.1093/nar/gkh103

Kent JW, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, and Haussler D. 2002. The Human Genome Browser at UCSC. *Genome Research* **12**:996–1006. doi:10.1101/gr.229102

Kim S, Yamamoto J, Chen Y, Aida M, Wada T, Handa H, Yamaguchi Y. 2010. Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. *Genes Cells* **15**:1003–1013. doi:10.1111/j.1365-2443.2010.01436.x

Koressaar T, Remm M. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**:1289–1291. doi:10.1093/bioinformatics/btm091

Kriscenski-Perry E, Applegate CD, Serour A, Mhyre TR, Leonardo CC, Pearce DA. 2002. Altered Flurothyl Seizure Induction Latency, Phenotype, and Subsequent Mortality in a Mouse Model of Juvenile Neuronal Ceroid Lipofuscinosis/Batten Disease. *Epilepsia* **43**:1137–1140. doi:10.1046/j.1528-1157.2002.16002.x

Kubo T, Wada T, Yamaguchi Y, Shimizu A, Handa H. 2006. Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3′-UTRs. *Nucleic Acids Research* **34**:6264–6271. doi:10.1093/nar/gkl794

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357. doi:10.1038/nmeth.1923

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine A, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won H-H, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, harf J, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Consortium E. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**:285. doi:10.1038/nature19057

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. doi:10.1093/bioinformatics/btp352

Li W, You B, Hoque M, Zheng D, Luo W, Ji Z, Park J, Gunderson SI, Kalsotra A, Manley JL, Tian B. 2015. Systematic Profiling of Poly(A)+ Transcripts Modulated by Core 3' End Processing and Splicing Factors Reveals

927 Regulatory Rules of Alternative Cleavage and Polyadenylation. *PLOS Genetics* **11**:e1005166.
928 doi:10.1371/journal.pgen.1005166
929
930 Liao Y, Smyth GK, Shi W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote.
931 *Nucleic Acids Res* **41**:e108–e108. doi:10.1093/nar/gkt214
932
933 Lu H-CC, Tan Q, Rousseaux MW, Wang W, Kim J-YY, Richman R, Wan Y-WW, Yeh S-YY, Patel JM, Liu X, Lin
934 T, Lee Y, Fryer JD, Han J, Chahrour M, Finnell RH, Lei Y, Zurita-Jimenez ME, Ahimaz P, Anyane-Yeboa K,
935 Maldergem L, Lehalle D, Jean-Marcais N, Mosca-Boidron A-LL, Thevenon J, Cousin MA, Bro DE, Lanpher BC,
936 Klee EW, Alexander N, Bainbridge MN, Orr HT, Sillitoe RV, Ljungberg M, Liu Z, Schaaf CP, Zoghbi HY. 2017.
937 Disruption of the ATXN1-CIC complex causes a spectrum of neurobehavioral phenotypes in mice and humans.
938 *Nature genetics* **49**:527–536. doi:10.1038/ng.3808
939
940 Maio A, Yalamanchili H, Adamski CJ, Gennarino VA, Liu Z, Qin J, Jung SY, Richman R, Orr H, Zoghbi HY.
941 2018. RBM17 Interacts with U2SURP and CHERP to Regulate Expression and Splicing of RNA-Processing
942 Proteins. *Cell Reports* **25**:726-736.e7. doi:10.1016/j.celrep.2018.09.041
943
944 Martin G, Gruber AR, Keller W, Zavolan M. 2012. Genome-wide analysis of pre-mRNA 3' end processing reveals
945 a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell reports* **1**:753–63.
946 doi:10.1016/j.celrep.2012.05.003
947
948 Masamha CP, Wagner EJ. 2017. The contribution of alternative polyadenylation to the cancer phenotype.
949 *Carcinogenesis* **39**:2–10. doi:10.1093/carcin/bgx096
950
951 Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu A-B, Li W, Wagner EJ. 2014. CFIm25 links alternative
952 polyadenylation to glioblastoma tumour suppression. *Nature* **510**:412–416. doi:10.1038/nature13261
953
954 Miura K, Kishino T, Li E, Webber H, Dikkes P, Holmes GL, Wagstaff J. 2002. Neurobehavioral and
955 Electroencephalographic Abnormalities in Ube3aMaternal-Deficient Mice. *Neurobiol Dis* **9**:149–159.
956 doi:10.1006/nbdi.2001.0463
957
958 Miura P, Shenker S, Andreu-Agullo C, Westholm JO, Lai EC. 2013. Widespread and extensive lengthening of 3′
959 UTRs in the mammalian brain. *Biotechfor* **23**:812–825. doi:10.1101/gr.146886.112
960
961 Nowicka M, Robinson MD. 2016. DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in
962 genomics. *F1000research* **5**:1356. doi:10.12688/f1000research.8900.1
963

964 Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins
965 University (Baltimore, MD), 7/22/2019. World Wide Web URL: https://omim.org/
966
967 Paxinos G, Franklin K. 2001. The Mouse Brain in Stereotaxic Coordinates Academic Press San Diego. *CA, USA*.
968
969 Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G,
970 Eisenacher M, Pérez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yılmaz Ş, Tiwary S, Cox J, Audain E, Walzer M,
971 Jarnuczak AF, Ternent T, Brazma A, Vizcaíno J. 2018. The PRIDE database and related tools and resources in
972 2019: improving support for quantification data. *Nucleic Acids Res* **47**:gky1106-. doi:10.1093/nar/gky1106
973
974 Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. 1999. Probability-based protein identification by searching
975 sequence databases using mass spectrometry data. *Electrophoresis* **20**:3551–3567. doi:10.1002/(sici)1522-
976 2683(19991201)20:18<3551::aid-elps3551>3.0.co;2-2
977
978 Pieretti M, Zhang F, Fu Y-H, Warren ST, Oostra BA, Caskey CT, Nelson DL. 1991. Absence of expression of the
979 FMR-1 gene in fragile X syndrome. *Cell* **66**:817–822. doi:10.1016/0092-8674(91)90125-i
980
981 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*
982 **26**:841–842. doi:10.1093/bioinformatics/btq033
983
984 Roberson ED, Halabisky B, Yoo JW, Yao J, Chin J, Yan F, Wu T, Hamto P, Devidze N, Yu G-Q, Palop JJ, Noebels
985 JL, Mucke L. 2011. Amyloid-β/Fyn–Induced Synaptic, Network, and Cognitive Impairments Depend on Tau Levels
986 in Multiple Mouse Models of Alzheimer's Disease. *J Neurosci* **31**:700–711. doi:10.1523/jneurosci.4152-10.2011
987
988 Routh A, Ji P, Jaworski E, Xia Z, Li W, Wagner EJ. 2017. Poly(A)-ClickSeq: click-chemistry for next-generation
989 3′-end sequencing without RNA enrichment or fragmentation. *Nucleic acids research* **45**:e112.
990 doi:10.1093/nar/gkx286
991
992 Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek EA, Kou Y, Liu L, Fromer M, Walker S, Singh T,
993 Klei L, Kosmicki J, Fu S-C, Aleksic B, Biscaldi M, Bolton PF, Brownfeld JM, Cai J, Campbell NG, Carracedo A,
994 Chahrour MH, Chiocchetti AG, Coon H, Crawford EL, Crooks L, Curran SR, Dawson G, Duketis E, Fernandez BA,
995 Gallagher L, Geller E, Guter SJ, Hill SR, Ionita-Laza I, Gonzalez P, Kilpinen H, Klauck SM, Kolevzon A, Lee I, Lei
996 J, Lehtimäki T, Lin C-F, Ma'ayan A, Marshall CR, McInnes AL, Neale B, Owen MJ, Ozaki N, Parellada M, Parr
997 JR, Purcell S, Puura K, Rajagopalan D, Rehnström K, Reichenberg A, Sabo A, Sachse M, Sanders SJ, Schafer C,
998 Schulte-Rüther M, Skuse D, Stevens C, Szatmari P, Tammimies K, Valladares O, Voran A, Wang L-S, Weiss LA,
999 Willsey JA, Yu TW, Yuen RK, udy T, for Autism H, Consortium U, Consortium T, Cook EH, Freitag CM, Gill M,

1000 Hultman CM, Lehner T, Palotie A, Schellenberg GD, Sklar P, State MW, Sutcliffe JS, Walsh CA, Scherer SW,
1001 Zwick ME, Barrett JC, Cutler DJ, Roeder K, Devlin B, Daly MJ, Buxbaum JD. 2014. Synaptic, transcriptional and
1002 chromatin genes disrupted in autism. *Nature* **515**:209. doi:10.1038/nature13772
1003
1004 Rüegsegger U, Beyer K, Keller W. 1996. Purification and Characterization of Human Cleavage Factor I Involved in
1005 the 3′ End Processing of Messenger RNA Precursors. *Journal of Biological Chemistry* **271**:6107–6113.
1006 doi:10.1074/jbc.271.11.6107
1007
1008 Rüegsegger U, Blank D, Keller W. 1998. Human Pre-mRNA Cleavage Factor Im Is Related to Spliceosomal SR
1009 Proteins and Can Be Reconstituted In Vitro from Recombinant Subunits. *Molecular Cell* **1**:243–253.
1010 doi:10.1016/s1097-2765(00)80025-8
1011
1012 Saltzman AB, Leng M, Bhatt B, Singh P, Chan DW, Dobrolecki LE, Chandrasekaran H, Choi J, Jain A, Jung S,
1013 Lewis MT, Ellis MJ, Malovannaya A. 2018. gpGrouper: A Peptide Grouping Algorithm for Gene-Centric Inference
1014 and Quantitation of Bottom-Up Proteomics Data. *Molecular & Cellular Proteomics* **17**:mcp.TIR118.000850.
1015 doi:10.1074/mcp.tir118.000850
1016
1017 Samaco RC, McGraw CM, Ward CS, Sun Y, Neul JL, Zoghbi HY. 2012. Female Mecp2(+/-) mice display robust
1018 behavioral deficits on two different genetic backgrounds providing a framework for pre-clinical studies. *Human*
1019 *molecular genetics* **22**:96–109. doi:10.1093/hmg/dds406
1020
1021 Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008. Proliferating Cells Express mRNAs with Shortened
1022 39 Untranslated Regions and Fewer MicroRNA Target Sites. *Science* **320**:1643–1647. doi:10.1126/science.1155390
1023
1024 Tan Q, Yalamanchili H, Park J, Maio A, Lu H-C, Wan Y-W, White JJ, Bondar VV, Sayegh LS, Liu X, Gao Y,
1025 Sillitoe RV, Orr HT, Liu Z, Zoghbi HY. 2016. Extensive cryptic splicing upon loss of RBM17 and TDP43 in
1026 neurodegeneration models. *Hum Mol Genet* ddw337. doi:10.1093/hmg/ddw337
1027
1028 Tan Q, Zoghbi HY. 2018. Mouse models as a tool for discovering new neurological diseases. *Neurobiol Learn Mem*.
1029 doi:10.1016/j.nlm.2018.07.006
1030
1031 Tian B, Manley JL. 2016. Alternative polyadenylation of mRNA precursors. *Nature reviews Molecular cell biology*.
1032 doi:10.1038/nrm.2016.116
1033
1034 Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new
1035 capabilities and interfaces. *Nucleic Acids Res* **40**:e115–e115. doi:10.1093/nar/gks596
1036

van Hoof A, Frischmeyer PA, Dietz HC, Parker R. 2002. Exosome-Mediated Recognition and Degradation of mRNAs Lacking a Termination Codon. *Science* **295**:2262–2264. doi:10.1126/science.1067272

Verkerk A, Pieretti M, Sutcliffe JS, Fu Y-H, Kuhl D, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang F, Eussen BE, van Ommen G-JB, Blonden L, Riggins GJ, Chastain JL, Kunst CB, Galjaard H, Caskey TC, Nelson DL, Oostra BA, Warren ST. 1991. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**:905–914. doi:10.1016/0092-8674(91)90397-h

Vissers L, Gilissen C, Veltman JA. 2015. Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics* **17**:9–18. doi:10.1038/nrg3999

Wang R, Nambiar R, Zheng D, Tian B. 2017. PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res* gkx1000-. doi:10.1093/nar/gkx1000

Weng T, Ko J, Masamha CP, Xia Z, Xiang Y, Chen N, Molina JG, Collum S, Mertens TC, Luo F, Philip K, Davies J, Huang J, Wilson C, Thandavarayan RA, Bruckner BA, Jyothula S, Volcik KA, Li L, Han L, Li W, Assassi S, Karmouty-Quintana H, Wagner EJ, Blackburn MR. 2019. Cleavage factor 25 deregulation contributes to pulmonary fibrosis through alternative polyadenylation. *J Clin Invest*. doi:10.1172/jci122106

Yalamanchili H, Wan Y, Liu Z. 2017. Current Protocols in Bioinformatics. *Curr Protoc Bioinform* **59**:11.15.1-11.15.21. doi:10.1002/cpbi.33

Yang Q, Coseno M, Gilmartin GM, Doublié S. 2011. Crystal Structure of a Human Cleavage Factor CFIm25/CFIm68/RNA Complex Provides an Insight into Poly(A) Site Recognition and RNA Looping. *Structure* **19**:368–377. doi:10.1016/j.str.2010.12.021

Yang Q, Gilmartin GM, Doublié S. 2010. Structural basis of UGUA recognition by the Nudix protein CFIm25 and implications for a regulatory role in mRNA 3 processing. *Proceedings of the National Academy of Sciences* **107**:10062–10067. doi:10.1073/pnas.1000848107

Yin J, Schaaf CP. 2017. Autism genetics – an overview. *Prenatal Diagnosis* **37**:14–30. doi:10.1002/pd.4942

Zhang H, Lee J, Tian B. 2005. Biased alternative polyadenylation in human tissues. *Genome Biology* **6**:R100. doi:10.1186/gb-2005-6-12-r100

Zhao K, Lu Z, Park J, Zhou Q, Xing Y. 2013. GLiMMPS: robust statistical model for regulatory variation of

alternative splicing using RNA-seq data. *Genome Biology* **14**:R74. doi:10.1186/gb-2013-14-7-r74

Zhu Y, Wang X, Forouzmand E, Jeong J, Qiao F, Sowd GA, Engelman AN, Xie X, Hertel KJ, Shi Y. 2018. Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. *Molecular Cell* **69**. doi:10.1016/j.molcel.2017.11.031