Published in final edited form as:
Nature. 2018 August ; 560(7719): 441–446. doi:10.1038/s41586-018-0423-5.

Autism-like phenotype and risk gene-RNA deadenylation by CPEB4 mis-splicing

Alberto Parras[1,2], **Héctor Anta**[3,4], **María Santos-Galindo**[1,2], **Vivek Swarup**[5], **Ainara Elorza**[1,2], **José L. Nieto-Gonzalez**[2,6], **Sara Picó**[1,2], **Ivó H. Hernández**[1,2,7], **Juan I. Díaz-Hernández**[1,2], **Eulàlia Belloc**[4], **Annie Rodolosse**[4], **Neelroop N. Parikshak**[5], **Olga Peñagarikano**[5,8,9], **Rafael Fernández-Chacón**[2,6], **Manuel Irimia**[10,11], **Pilar Navarro**[3], **Daniel H. Geschwind**[5], **Raúl Méndez**[4,12,*], and **José J. Lucas**[1,2,*]

[1]Center for Molecular Biology “Severo Ochoa” (CBMSO) CSIC/UAM, Madrid 28049, Spain

[2]Networking Research Center on Neurodegenerative Diseases (CIBERNED). Instituto de Salud Carlos III, Madrid, Spain

[3]Cancer Research Program, Hospital del Mar Medical Research Institute (IMIM), Barcelona 08003, Spain

[4]Institute for Research in Biomedicine (IRB), Barcelona Institute of Science and Technology, 08028 Barcelona, Spain

[5]Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, UCLA, Los Angeles, CA 90095, USA

[6]Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/ Universidad de Sevilla and Departamento de Fisiología Médica y Biofísica, 41013 Sevilla, Spain

[7]Facultad de Ciencias, Departamento de Biología (Unidad Docente Fisiología Animal), Universidad Autónoma de Madrid, Madrid 28049, Spain

[8]Department of Pharmacology, School of Medicine, University of the Basque Country (UPV/EHU), 48940 Leioa, Spain

[9]Centro de Investigación Biomédica en Red en Salud Mental (CIBERSAM), Madrid, Spain

[*]Corresponding authors: José J. Lucas, Center for Molecular Biology “Severo Ochoa” (CBMSO), C/ Nicolás Cabrera, 1. Campus UAM de Cantoblanco, 28049 Madrid. Spain, Tel. +34 91 196 4552, jjlucas@cbm.csic.es; Raúl Méndez, ICREA and Institute for Research in Biomedicine (IRB), C/ Baldiri Reixac 10. 08028 Barcelona, Spain, Tel. +34 93 403 1900, raul.mendez@irbbarcelona.org.

Author Contributions. A.P. contributed to study design and was involved in all assays and data collection. H.A. and E.B. contributed to RIP and PolyU chromatography experiments and analyzed data from doxycycline treated mice. M.S.-G., S.P. and I.H.H. performed Western-Blot and qRT-PCR analysis. J.L.N.-G. performed electrophysiological recordings. A.R. optimized microarray processing. V.S., A.E., J.I.D.-H., N.N.P. and M.I. performed bioinformatics analyses. R.F.-C., P.N., O.P., M.I., R.M. and D.H.G. made intellectual contributions to experimental design and discussion. D.H.G., R.M. and J.J.L. directed the study and designed experiments. J.J.L. wrote the paper with input from all authors.

Competing interests. The authors declare no competing interests.

Data availability. The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Europe PMC Funders Author Manuscripts

[10]Centre for Genomic Regulation (CRG), Barcelona Institute for Science and Technology, 08003 Barcelona, Spain

[11]Universitat Pompeu Fabra, 08003, Barcelona, Spain

[12]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

Abstract

Common genetic contributions to autism spectrum disorder (ASD) reside in risk-gene variants that individually have minimal effect-sizes. Since neurodevelopment-perturbing environmental factors also underlie idiopathic-ASD, it is crucial to identify altered regulators able to orchestrate multiple ASD-risk genes during neurodevelopment. Cytoplasmic polyadenylation element binding proteins 1-4 (CPEB1-4) regulate translation of specific mRNAs by modulating their poly(A)-tail and participate in embryonic development and synaptic plasticity. Here we find that CPEB4 binds transcripts of most high-confidence ASD genes. Idiopathic-ASD brains show CPEB4 transcript isoform imbalance due to decreased inclusion of a neuronal-specific microexon together with a new molecular signature of global poly(A)-tail shortening that remarkably impacts high-confidence ASD-risk genes with concomitant reduction of their protein levels. Equivalent CPEB4 transcript isoform imbalance in mice mimics the mRNA-polyadenylation and protein level changes of ASD genes and induces ASD-like neuroanatomical, electrophysiological and behavioral phenotypes. Altogether, these data unravel CPEB4 as a novel regulator of ASD-risk genes.

Autism spectrum disorder (ASD) is highly heritable1. However, despite the importance of genetic determinants in ASD causality, neurodevelopment-perturbing environmental factors also contribute2–4. A minority of ASD cases correspond to syndromic forms caused by highly penetrant single-gene mutations or chromosomal abnormalities, often characterized by additional phenotypes, such as intellectual disability, epilepsy, craniofacial dysmorphology and others5. In contrast, the majority of cases correspond to idiopathic ASD for which the genetic causality resides in polygenic risk involving small effect-size variants in hundreds of genes5–8. A major question in understanding ASD is therefore to identify whether altered regulators in the brains of idiopathic ASD individuals could orchestrate pathogenic changes in numerous ASD-risk genes during neurodevelopment.

Cytoplasmic polyadenylation element binding proteins 1-4 (CPEB1-4) are RNA-binding proteins that repress or activate translation of mRNAs with CPE sequences in their 3’ untranslated regions (UTRs) by inducing cytoplasmic-shortening or -elongation of their poly(A)-tails9. CPEBs were discovered as regulators of certain mRNAs in response to embryonic environmental clues, such as hormones9,10; later, they were shown to be involved in learning and memory by modulating synaptic plasticity9,11,12. As FMR1/CPEB1 double-knockout (KO) rescues the fragile X-like phenotype of FMR1-KO mice13, it has been suggested that manipulating CPEB1 might have therapeutic value for this monogenic X-linked intellectual disability syndrome, in which up to 50% of cases also show autistic features. However, a role of CPEBs in the etiology of a broader range of neurodevelopmental disorders—including non-syndromic ASD—has not been studied.

ASD-risk gene mRNAs bear CPEs and bind CPEB4

To investigate the CPEB-bound brain transcriptome in a disease context, we performed CPEB1 and CPEB4 RNA-immunoprecipitation (RIP) with striatum (St) RNA from wild-type (WT) mice and from a Huntington's disease (HD) mouse model in which altered CPEB1- and CPEB4-levels correlate with transcriptomic poly(A)-tail length changes (Extended Data Fig. 1a-c). Regardless of genotype, 7.9% of transcripts were bound only by CPEB4, 5.8% only by CPEB1 and 7.0% by both (Fig. 1a, Supplementary Table 1a). Enrichment of CPE sequences on the 3'UTR of RIP-detected transcripts supported specificity of this binding (Extended Data Fig. 1b). When comparing CPEB-specific targets and HD-associated polyadenylation changes, we found that CPEB4-specific mRNAs were enriched within deadenylated transcripts (Extended Data Fig. 1c). Interestingly, the largest fold-change (FC) in this category corresponded to *Auts2*, a gene linked to ASD14, and several high confidence ASD genes (*Dyrk1a*, *Cul3*, and *Ptchd1*; categories 1-2 in SFARI database) were among the forty CPEB4-targets with most prominent poly(A)-shortening (FC ≤ –3.0) (Extended Data Fig. 1d, Supplementary Table 1b). The remarkable enrichment of CPEB4-specific deadenylated mRNAs for SFARI genes (Extended Data Fig. 1e), led us to hypothesize a role of CPEB4 in the expression of ASD-risk genes.

We then analysed the incidence of CPEB4 binders in a compiled set of ASD genetic risk candidates from the SFARI database (Supplementary Table 2). CPEB4 binders were enriched in ASD-risk genes and also in a smaller curated list of ASD-only genes, which cause ASD but not intellectual disability (Extended Data Fig. 1f; a weaker enrichment was also observed for CPEB1 target mRNAs). CPEB4 binders were also overrepresented within several functional co-expression modules that represent shared pathology in the ASD brain as identified in previous microarray15 and RNA-seq16 studies (Extended Data Fig. 1g).

Next, we found an increase of canonical CPEs in the 3′ UTRs of mRNAs of ASD genes in the SFARI highest confidence categories, in the 39 genes harbouring rare *de novo* protein disrupting mutations identified in two whole-exome sequencing studies for simplex ASD17,18 and in the equivalent 61 genes from a recent study19 ("ASD39" and "Takata 2018" lists; Supplementary Table 2) when compared with multiple control gene sets: brain-, synaptic- and neuronal-enriched transcriptomes (Fig. 1b). This enrichment remains after all and brain-, neuronal- and synaptic-enriched genes were stratified with respect to the ASD genes for 5'UTR, CDS or 3'UTR length, gene size, or ratio of neuron vs. glia expression (Extended Data Fig. 1h and Supplementary Table 5). We then confirmed that most high-confidence ASD genes corresponded to CPEB4 targets (Fig. 1c). Altogether, these data demonstrate that mRNAs of the majority of high-confidence ASD-risk genes contain CPEs and are bound by CPEB4.

CPEB4 alteration in idiopathic ASD brains

To assess whether CPEBs are altered in the brains of idiopathic ASD individuals, we analysed transcript levels in RNA-seq data16 from post-mortem cortex (Cx) (43 idiopathic ASD, 63 neurotypical control). We found no changes for *CPEB1* and *CPEB2,* a slight decrease for *CPEB3* and a slight increase for *CPEB4* (Extended Data Fig. 2a and Fig. 2a).

At the protein level, only CPEB4 was significantly altered in idiopathic ASD brains but, strikingly, it was decreased despite increased transcript levels (Fig. 2b and Extended Data Fig. 2b-c). This suggests that CPEB4 is heavily post-transcriptionally regulated, fitting the autoamplification loop20,21 via CPEB4 binding to its own transcript (Supplementary Table 1a), in line with the multiple CPE sequences in its 3' UTR21. The decreased levels of CPEB4 in ASD occur particularly in young individuals (≤35 years-old), where protein levels were higher compared to older individuals in control (CTRL) samples (Fig. 2b).

Splicing alterations15,22, particularly of microexons23, have been reported in ASD. We thus looked for potential splicing alterations in mRNAs of the different CPEBs in our published16 rMATS-analysis of cortical RNA-seq data. Only CPEB4 showed significant splicing alterations in ASD samples, involving different combinations of two consecutive alternatively spliced exons (Extended Data Fig. 2d): the 51-nt exon 3 and the 24-nt neuronal-specific microexon (exon 4) (Fig. 2c). To specifically investigate changes in the level of inclusion of these exons, we re-analyzed the RNA-seq data using vast-tools24. This revealed significantly less inclusion of exon 4 in ASD brains (ΔPSI=-7.6, Extended data Fig. 2e) which, parallel to the decreased CPEB4 protein levels, essentially occurs in individuals under 35 years old (ΔPSI=-8.8) (Fig. 2d). Interestingly, this microexon encodes the 8–amino acid B region25, which adds potential motifs for posttranslational modifications, such as phosphorylation by AKT, S6K, PKA or PKC (NETPHOS 2.0)25 (Fig. 2c). In contrast, a tendency to more exon 3 inclusion was found in individuals under 35 (Fig. 2d).

Isoform-specific exon-junction RNA-seq reads were used to estimate the relative abundance of the four possible isoforms resulting from alternative splicing of exons 3 and 4 (Full length-CPEB4, CPEB4Δ3, CPEB4Δ4 and CPEB4Δ3Δ4). This revealed a significant increase of the CPEB4Δ4 transcript in ASD (Extended Data Fig. 2f). RT-PCR with primers that simultaneously amplify the four isoforms confirmed the increase in Δ4 transcripts (CPEB4Δ4 and CPEB4Δ3Δ4) to the expense of Ex4+ isoforms (FL-CPEB4 and CPEB4Δ3) in young ASD individuals (Fig. 2e-f), a pattern that was further validated using conventional and digital-droplet (Fig. 2g-h and Extended Data Fig. 2g-i) absolute qRT-PCR. Overall, these data demonstrate an increased Δ4/Ex4+ transcript ratio of CPEB4 in ASD, concomitantly with slightly increased total RNA and decreased total protein levels.

Deadenylation of ASD gene mRNAs in ASD brains

We then explored potential genome-wide poly(A)-tail length changes in mRNA from post-mortem prefrontal Cx tissue of young controls and idiopathic ASD cases (Extended Data Fig. 3a). In ASD samples, 10.2% and 9.1% of transcripts showed poly(A)-tail lengthening and shortening, respectively (Fig. 3a "Total" column and Supplementary Table 3). Remarkably, transcripts deadenylated in ASD brains were significantly enriched for CPEB4 binders (Extended Data Fig. 3b). Gene ontology analysis detected "oxytocin signalling pathway" –implicated in social behaviour and proposed to be therapeutically relevant to ASD26– as the most significantly enriched term (among deadenyated CPEB4-binding transcripts, Extended Data Fig. 3c). Interestingly, the SFARI ASD-risk genes globally showed a deadenylation signature (Fig. 3a and Extended Data Fig. 3d-e). Strikingly, the poly(A)-tail shortening of ASD genes was progressively exacerbated with the increased

confidence in causality as defined by SFARI curated gene categories (Fig. 3a). This was not a by-product of enrichment of brain, neuronal or synaptic specific transcripts, or biased 5'UTR, CDS and 3'UTR lengths, gene sizes and ratio of neuronal vs. glial expression (Extended Data Fig. 3f-g and Supplementary Table 5). As expected, CPEB4 binders were overrepresented among deadenylated SFARI cat. 1–3 genes (Fig. 3b), thus suggesting a role of CPEB4 in the observed deadenylation of ASD genes.

We next analysed the protein levels of CPEB4-bound ASD genes in Cx tissue of young idiopathic ASD cases (Fig. 3c). The four studied SFARI cat. 1-2 genes (*PTEN*, *DYRK1A*, *FOXP1* and *WAC*) with significant poly(A) shortening (*PTEN* transcript deadenylation also validated by Hire-PAT, Extended Data Fig. 3h) showed significantly decreased protein levels in ASD brains despite unchanged transcript levels (Fig. 3d). A similar pattern was observed for tested SFARI cat. 3 genes showing a trend for deadenylation (*AUTS2, RBFOX1* and *ZBTB20*). Furthermore, *PCDH9* an SFARI cat. 4 gene displaying one of the most prominent and significant shortenings in poly(A)-tail length, also showed decreased protein levels despite increased transcript levels (Fig. 3c-d). Unaltered protein levels of neuronal- and glial-specific genes such as CALB1, D2R, SNAP25, TUBB3 and IBA1, whose poly(A) tails were not changed in ASD (Extended Data Fig. 3i), rule out an underlying non-specific decrease in protein translation in ASD. Together, these results are consistent with Δ4 CPEB4 isoforms favouring deadenylation (and concomitant decreased protein levels) of target transcripts, including CPEB4 itself, and multiple ASD-risk genes.

Altogether, the presented human data indicate that ASD-risk genes are enriched in CPE-containing and CPEB4-binder transcripts. Moreover, idiopathic ASD brains show (i) CPEB4 mis-splicing (exon 4 skipping) and reduced CPEB4 protein levels, (ii) a new molecular signature of mRNA deadenylation that remarkably impacts high-confidence ASD-risk genes and (iii) concomitant decreased protein levels of multiple CPEB4-target and deadenylated ASD-risk gene transcripts.

ASD-like poly(A) changes in TgCPEB4Δ4 mice

To determine if changes in CPEB4 splicing and/or protein levels can cause the observed changes in polyadenylation and translation of ASD-risk gene mRNAs, we used different mouse models to mimic the CPEB4 changes observed in idiopathic ASD brains. First, we used two models to emulate the decreased CPEB4 protein levels: a heterozygous CPEB4 $KO^{GT}/+$ model27 showing partial reduction of CPEB4 protein with unaltered isoform ratios (Extended Data Fig. 4a) and a homozygous CPEB4 KO model20 showing full suppression of CPEB4 protein (Extended Data Fig. 4b). Both models showed similar changes in global transcript polyadenylation, but in the opposite direction to what was observed in ASD cases (Extended Data Fig. 4c-f and Supplementary Table 4). CPEB4-deficient mice showed prominent and significant poly(A) lengthening in ASD-risk genes (Extended Data Fig. 4c-h). Therefore, we concluded that the decreased polyadenylation of ASD-risk mRNAs in ASD brains was unlikely a consequence of their reduced CPEB4 levels—on the contrary, decreased CPEB4 by itself had the opposite effect.

To explore the effects of the increase in Δ4 *CPEB4* transcript isoforms, we generated mice with conditional neuronal-specific overexpression of *CPEB4Δ4* transcript (TgCPEB4Δ4 mice, Fig. 4a), which did not display perinatal lethality–based on live births–and were indistinguishable from their WT and single-transgenic littermates from birth to weaning (Extended Data Fig. 5a). However, at weaning (3 weeks), up to 40% of TgCPEB4Δ4 mice begun to develop cranial dysmorphology suggestive of hydrocephalus. TgCPEB4Δ4 mice with cranial dysmorphology died prematurely, with a mortality peak at 7 weeks of age (Extended Data Fig. 5b). The TgCPEB4Δ4 mice with normal cranial morphology did not present any obvious abnormality nor die prematurely, but they were significantly smaller than their control littermates starting at 3 weeks of age (Extended Data Fig. 5c). In the remaining study, analysis was restricted to TgCPEB4Δ4 mice with normal cranial morphology. Transgene expression in TgCPEB4Δ4 mice takes place in neurons of forebrain structures, such as Cx and St (Extended Data Fig. 5d-e). Importantly, overexpression of CPEB4Δ4 transcript resulted in a Δ4/Ex4+ transcript ratio (Fig. 4b-c) similar to that observed in brains of idiopathic ASD individuals (Fig. 2h). Total CPEB4 protein levels in TgCPEB4Δ4 mice moderately increased in young adults (1.5 months), did not increase at 12 months and a tendency to decrease was observed in 2-year-old mice, despite increased transcript levels at all ages (Extended Data Fig. 5f-g). Likewise, as in human ASD brains, the other CPEBs were essentially unaltered in TgCPEB4Δ4 mice (Extended Data Fig. 5f-g). Strikingly, global poly(A)-tail length changes in TgCPEB4Δ4 mice significantly overlapped with those observed in idiopathic ASD individuals (Fig. 4d, Supplementary Table 4 and Extended Data Fig. 6a) and replicated the predominant deadenylation of ASD-risk genes, a pattern correlated with increasing ASD-risk gene confidence (Fig. 4e). The latter result was robust to different stratification analyses (Extended Data Fig. 6b-c and Supplementary Table 5). Thus, creating a CPEB4 transcript isoform imbalance in mice that mimics the increase in Δ4 isoforms observed in ASD individuals was sufficient to induce the ASD-associated poly(A) signature.

Similar to the observations in human ASD brains, shortened poly(A) tail-length of ASD-risk gene transcripts correlated with reduced protein levels in Cx and St of TgCPEB4Δ4 mice (Fig. 4f and Extended Data Fig. 6d). *Zbtb20, Tnrc6b*, Chd2, *Foxp1, Wac, Auts2* and *Gpc6* are among the deadenylated SFARI cat.1-4 genes (*Auts2* deadenylation also validated by Hire-PAT, Extended Data Fig. 6e) whose transcripts are bound by CPEB4, and they all showed significant decreases of their protein levels without decreased transcript levels (Fig. 4f-g and Supplementary Table 4). It is worth noting that RBFOX1 protein–one of the few splicing factors that enable microexon processing in neurons28 and that is known to regulate alternative splicing of *CPEB4*9– is also decreased in TgCPEB4Δ4 mice (Fig. 4h) and that *Cpeb4* itself shows poly(A) tail shortening in TgCPEB4Δ4 mice (Supplementary Table 4) which may explain why protein levels do not match the increased transcript levels. Unaltered protein levels of non-deadenylated neuronal- and glial-specific genes rules out a non-specific decrease in protein translation efficiency in TgCPEB4Δ4 mice (Extended Data Fig. 6f).

We also generated TgCPEB4Δ4 mice in a CPEB4-KO heterozygous background (Extended Data Fig. 7a). Notably, CPEB4 transcript isoform imbalance persists in TgCPEB4Δ4:CPEB4-KO^{GT}/+ mice without increased CPEB4 protein levels (Extended Data Fig. 7b-c). ASD genes that were diminished in the TgCPEB4Δ4 mice showed similar

decreases in protein levels in the TgCPEB4Δ4:CPEB4-KO^{GT}/+ mice, while control neuronal and glial genes were unaltered (Extended Data Fig. 7d-e). These data strongly suggest that the observed effects are due to the transcript isoform imbalance rather than to increased CPEB4 protein levels. Consistent with this, we did not observe decreased protein levels of ASD genes in CPEB4-KO^{GT}/+ mice -which only have decreased CPEB4 protein (Extended Data Fig. 7f).

ASD-related phenotypes of TgCPEB4Δ4 mice

We then tested whether TgCPEB4Δ4 mice showed potential ASD-related anatomical, electrophysiological and/or behavioural abnormalities. We observed reduced brain weight for TgCPEB4Δ4 mice (Extended Data Fig. 8a) as well as for TgCPEB4Δ4:CPEB4-KO^{GT}/+ mice, but not for CPEB4-KO^{GT}/+ mice, with volume reductions in Cx, St and hippocampus together with lateral ventricle enlargement (Fig. 5a) and a two-fold increase in caspase-3 positive cells, without decreased neuronal density (Extended Data Fig. 8b-c). Dendritic spine dysgenesis is frequent in ASD and mouse models30. We found a 9.2% decrease in total spine density in TgCPEB4Δ4 mice (Fig. 5b) but not in CPEB4 KO^{GT}/+ mice (Extended Fig. 8d). Whole-cell recordings of miniature excitatory postsynaptic currents (mEPSCs) in layer V pyramidal neurons of somatosensory Cx revealed no differences in the mean amplitude but a 32% reduction in the mean frequency in TgCPEB4Δ4 (Fig. 5c), resembling neurexin-dysfunction mice31. No such electrophysiological alterations were observed in CPEB4-KO^{GT}/+ mice (Extended Data Fig. 8e). Together with the deficit in spine density, these findings are compatible with a presynaptic shortfall in neurotransmitter release and/or a reduction in numbers of excitatory synapses.

We then assayed behaviour. In the open field (OF) test, TgCPEB4Δ4 mice showed stereotypical running at the cage periphery (Fig. 5d), which was not due to anxiety-related behaviour since, in the elevated plus-maze test, TgCPEB4Δ4 mice actually showed reduced levels of anxiety (Fig. 5e). In the ultrasonic vocalization (UsV) test, TgCPEB4Δ4 mice emitted significantly less UsVs (Fig. 5f). In the social approach (SA) test, TgCPEB4Δ4 mice did not show a preference to interact with the cage containing a mouse over the empty one, indicating a dysfunction in sociability (Fig. 5g). TgCPEB4Δ4:CPEB4-KO^{GT}/+ mice, but not CPEB4-KO^{GT}/+ mice, mimicked these ASD-like behaviours (Extended Data Fig. 8f-h), demonstrating that they are due to the altered transcript isoform ratio and not to altered total CPEB4 protein levels. Interestingly, TgCPEB4Δ4 mice treated with doxycycline to express the transgene only during embryonic and early postnatal life (ON/OFF-TgCPEB4Δ4 mice) displayed the stereotypic running in OF and the deficit in SA, while TgCPEB4Δ4 mice treated to express the transgene only after the age of 3 weeks (OFF/ON-TgCPEB4Δ4 mice) did not (Extended Data Fig. 9), indicating that these ASD-like phenotypes in TgCPEB4Δ4 mice originate during development. Together, the stereotypic running, the UsV-communication deficit, and the diminished social interaction indicate that TgCPEB4Δ4 mice display components of autistic-like behaviour that have been seen in multiple mouse models harbouring single ASD-risk mutations.

Hundreds of minimally penetrant mutations are implicated in ASD6 of which only a fraction will coincide in a given individual. Accordingly, environment-triggered misexpression of

multiple ASD-related genes may contribute4, and CPEB4 is well positioned to act as one such neurodevelopment perturbation-driven regulator. It is conceivable that development-modifying factors that have been proposed for non-genetic ASD causality1,32 could modulate CPEB4 developmental functions through inducing its mis-splicing. In fact, prenatal cytomegalovirus infection–which has been related to ASD33– has been shown to remodel RNA splicing and polyadenylation in a CPEB1-dependent manner and to increase CPEB4 expression34.

The new molecular signature of transcript-deadenylation in idiopathic ASD brains, combined with proteomics data and gene modules dysregulated in ASD brain, may unravel additional risk genes or pathophysiological pathways. Finally, preclinical studies have mainly relied on monogenic syndromic-ASD mouse models. Since CPEB4 mis-splicing orchestrates expression of a plethora of ASD-risk genes, preclinical testing in TgCPEB4Δ4 mice might be relevant for a wide range of ASDs.

Methods

Human brain tissue samples

Brain specimens used in immunoblot, RNA sequencing, absolute qRT-PCR, digital-droplet PCR and poly(U) chromatography in this study from frontal and temporal cortex of Autism Spectrum Disorder (ASD) patients and controls (CTRL) were provided by University of Maryland Brain and Tissue Bank, NIH NeuroBioBank (NBB) (Baltimore,MD) and the Autism Tissue Program (ATP) brain bank at The Harvard Brain and Tissue Bank (Belmont, MA). Written informed consent for brain removal after death for diagnostic and research purposes was obtained from brain donors and/or next of kin. Brain sample and donor metadata is available in Supplementary Table 6.

Animals

As HD mouse model, we used R6/1 mice transgenic for the human exon-1-Htt gene35 because our unpublished results show that these mice have altered levels of CPEBs (specifically CPEB1 and CPEB4) which correlate with changes in poly(A)-tail length of numerous mRNAs. R6/1 mice were in B6CBAF1/J background. Heterozygous CPEB4KO^{GT}/+ mice27 harbor a gene trap between exons 1 and 2 which prevents formation of the full length CPEB4 protein while allowing expression of the N-terminal low complexity domain, LCD)36, thus resulting in results in partial reduction of CPEB4 protein. CPEB4 KO mice20 harbor homozygous deletion of constitutive exon 2 resulting in a premature stop codon and full suppression of CPEB4 protein. CamkII-tTA (tTA)37, CPEB4KO^{GT}/+, and CPEB4 KO mice were in C57BL/6J background. Conditional mice expressing human CPEB4 lacking exon 4 (TgCPEB4Δ4) were generated (for details, see “Generation of TgCPEB4Δ4 mice” below) for this study and used in C57BL/6J background. All mice were housed in CBMSO animal facility. Mice were housed four per cage with food and water available ad libitum and maintained in a temperature-controlled environment on a 12/12 h light-dark cycle with light onset at 08:00. Animal housing and maintenance protocols followed the local authority guidelines. Animal experiments were performed under protocols (P15/P16/P18/P22) approved by the Centro de Biología Molecular Severo Ochoa

Institutional Animal Care and Utilization Committee (Comité de Ética de Experimentación Animal del CBM, CEEA-CBM), and Comunidad de Madrid PROEX 293/15.

RNA immunoprecipitation (RIP)

Four male WT and R6/1 mice (HD mice) were sacrificed by cervical dislocation at the age of 7-8 month-old. Total striatum (St) was quickly dissected on an ice-cold plate and cut into pieces. The pool of the four WT or R6/1 striata was washed twice with phosphate buffer solution (PBS), crosslinked with 0.5% formaldehyde in PBS for 7 min at room temperature and treated with glycine 1M for 5 min. After two washes with cold PBS, pool samples were homogenized in lysis buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% deoxicolate, 0.1% SDS, 1 mM EDTA, protease inhibitor (Complete, Roche, 11697498001), RNAse inhibitor (Ribolock, Life Technologies, EO0381)) and centrifuged at 13000g for 10 min at 4ºC. Dynabeads protein A (Life Technologies, 10001D) were washed twice with PBS and incubated with anti-CPEB4 antibody (10 µg, Abcam, ab83009), anti-CPEB1 antibody (5 µg, Proteintech, 13274-1-AP) or rabbit IgG (5 µg, Sigma, I5006) for 2h. Next, dynabeads were washed once with PBS and twice with triethanolamine 0.2 M pH 8.2, incubated with dimethyl pimelimidate 20 mM for 30 min, treated with Tris 50mM pH 8.0 and washed twice with lysis buffer. Lysates were precleared with unconjugated dynabeads for 20 min at 4ºC in a wheel, an aliquot was stored at -80ºC (“Input”) and the rest of the extract was immunoprecipitated with the antibody-conjugated dynabeads overnight at 4ºC in a wheel. Immunoprecipitates were washed six times in cold lysis buffer.

For protein extraction, samples were incubated with Laemmli buffer (10% SDS, 0.325 M Tris HCl pH7.5, Glycerol 25%) for 20 min at 60ºC. Dynabeads were removed with the help of a magnet and samples were boiled after adding DTT 0.1 M and bromophenol blue 0.1%.

For RNA extraction, immunoprecipitates were resuspended in 100 µl of proteinase K buffer (200 mM Tris pH 7.5, 100 mM NaCl, 10 mM EDTA, 1% SDS) containing 0.7 µg/µl of proteinase K (Roche, 03115852001) and incubated 1h at 42ºC and 1h at 65ºC. RNA was extracted using the TRIzol reagent (Invitrogen, 15596018). Briefly, samples were resuspended in 300 µl of TRIzol. Then, 20 µg of glycogen was added. Samples were vortexed, incubated for 5 min at room temperature and centrifuged at 14000g for 15 min at 4ºC. The aqueous phase was recovered, mixed with 1 volume of isopropanol, incubated for 5 min at room temperature and precipitated at 14000g 30 min at 4ºC. The pellet was washed with 300 µl of ice-cold 75% ethanol and centrifuged at 14000g 10 min at 4ºC. The pellets were resuspended in 100 µl of nuclease-free water. To ensure the complete removal of the phenol, the RNA was precipitated again. To this aim, 10 µl of sodium acetate 3 M pH 5.6 and 250 µl of 75% ethanol were added. Samples were vortexed and stored at -20ºC for 1h. Then, they were precipitated at 14000g for 30 min at 4ºC. The pellet was washed with 75% ethanol and centrifuged again at 14000g for 10 min at 4ºC. Pellets were air-dried for 5 min and resuspended in nuclease-free water.

RNA isolation, Whole Transcriptome Amplification (WTA) and microarrays processing

Input and immunoprecipitated RNA were purified using Agencourt RNAClean XP bead suspension (Beckman Coulter, A66514). Library preparation and amplification were

performed following the distributor's (Sigma-Aldrich) recommendations for WTA2 from purified immunoprecipitated or diluted input RNA. SYBR Green (Sigma-Aldrich, 163795-75-3) was added to the amplification reaction, which was performed in a CFX Real-time instrument (Bio-Rad) to monitor amplification yield. When the SYBR Green signal reached a plateau after 27 cycles, the reaction was stopped. Amplified cDNA was purified and quantified on a Nanodrop ND-1000 spectrophotometer (Thermo-Fischer). 8 μg of cDNA were subsequently fragmented by DNAseI and biotinylated by terminal transferase obtained from GeneChip Mapping 250K Nsp Assay Kit (Affymetrix, 900753). After hybridization for 16h at 45ºC, washing and staining was performed in the Affymetrix GeneAtlas Fluidics Station. The arrays were scanned in the GeneAtlas Imaging Station. All processing was performed according to manufacturer's recommendations. CEL files were generated from DAT files using Affymetrix Command Console software. To generate the log2 expression estimates, overall array intensity was normalized between arrays and the probe intensity of all probes in a probe set summarized to a single value using RMA (Robust Multichip Average) algorithm38.

In order to compare samples from different conditions, fold changes were computed after MA mean and variance normalization using a Generalized Additive Model (GAM). An empirical Bayes partial density model was then used to compute the posterior probability of differential expression. Differentially expressed genes were defined as those with a maximum False Discovery Rate (FDR) of 5% and a log2 fold change threshold of 1.75. We calculated the Input versus inmunoprecipitated fold change (FC) for CPEB1 and CPEB4 binders considering a transcript positive when at least one probe showed a FC above 1.75 in WT or HD mice.

Enrichment and co-expression network analysis of CPEB4 and CPEB1 binders

Enrichment analysis studies use one-sided Fisher's exact test to evaluate whether a gene set, in this case CPEB4 or CPEB1 binders determined by RIP, is enriched over background, providing a *P*-value and enrichment value. We used curated ASD candidate gene list from Simons Foundation Autism Research Initiative (SFARI) AutDB database, referred as ASD SFARI list and a more restrictive, smaller ASD only gene list, where genes linked to intellectual disability were removed. The gene set (CPEB4 and CPEB1 binders) was also used to study enrichment in functional co-expression modules that represent shared pathology in ASD brain. These gene modules derived from previous unbiased weighted gene co-expression network analysis (WGCNA) obtained by gene array15 and by RNA-seq16 studies from ASD postmortem samples.

Analysis of canonical and functional CPE sequences

3'UTR (untranslated region) sequences from selected gene sets were extracted from Ensembl (http://www.ensembl.org/)39 and incidence of canonical and functional CPE (cytoplasmic polyadenylation element) sequences was detected using the algorithm described in40 (http://genome.crg.es/CPE/). The list of brain genes was obtained from the human protein atlas (http://proteinatlas.org/humanproteome/brain), neuronal, astrocytic and oligodendrocyte-enriched genes from41, synaptic-enriched genes from42 and the ASD gene lists was obtained in July 2017 from SFARI database (categories 1 to 4, https://

gene.sfari.org/autdb/GS_Home.do). ASD39 gene list consists of the 39 genes harboring rare de novo protein disrupting mutations identified in the two most largest whole exome sequencing studies in simplex ASD17,43. Takata gene list consists of the 61 genes enriched for damaging *de novo* mutatios in ASD identificated by Takata et al19. All lists of ASD causing genes are shown in Supplementary Table 2.

Simulations

To compare feature enrichment in our subset of high confidence AS genes (SFARI cat. 1-2), we first selected several control gene groups: total genome, brain-enriched (from the human protein atlas), neuronal-enriched41 and synaptic-enriched42 and removed those genes previously linked to ASD (i.e. any SFARI category). Then, for each simulation and control group, each ASD gene was matched randomly with a gene from the control group based on its 5' UTR (+/- 75 nt), 3' UTR (+/- 150 nt) or CDS length (+/- 200 nt), genomic size (+/-2,000 bp) or ratio of Neuronal vs. Glial expression (+/- 0.1). For the latter, RNA-seq data for isolated populations of neurons, astrocytes, microglia, new oligodendrocytes, and oligodendrocyte precursors was obtained from44 and expression values calculated using vast-tools24. An average value for all glial cell types was calculated for each gene and the ratio between the expression in neurons vs. the glial (NvsG) average used for stratification. Next, the percentage of CPEs, CPEB4 binders and genes with poly(A)-tail shortened and lengthened in ASD human, TgCPEB4Δ4, CPEB4 KO^{GT}/+ and CPEB4 KO mice were calculated in each control and stratified test subset. This process was repeated 10,000 times and *P*-values were calculated as the number of times the stratified control show the same or higher percentage (or lower, if testing for depletion) than the test set, divided by 10,000. All results are shown in Supplementary Table 5.

Quantification of CPEBs gene expression in human postmortem ASD cases

CPEBs expression levels in idiopathic ASD (n=43 samples from 26 individuals) and CTRL (n=63 samples from 33 individuals) postmortem prefrontal (Ba9) and temporal (Ba41-42-22) cortex samples were evaluated from RNA-seq data16. Briefly, the paired-end raw reads were mapped to the human reference genome assembly GRCh37.73 using Tophat245, and the counts were quantified using HTSeq46. Gene length, G+C content and library size were normalized (referred to as "Normalized FPKM") using the cqn package in R47. Linear mixed effects (LME) were used, modeling to account for effects from biological covariates (condition, age, sex, brain region), technical variables related to sample processing (RIN, brain bank, sequencing batch), technical variables related to sequencing quality metrics and individual ID was set as a random effect accounting for the fact that multiple samples came from the same individual.

Quantification of CPEBs transcript splicing and differential splicing analysis

We used n= 81 CTRL and n = 82 ASD cortical prefrontal and temporal samples from n= 47 ASD and n = 44 CTRL individuals16. We computed percent spliced in (PSI) values using:

- Vast-tools (https://github.com/vastgroup/vast-tools). This software consists of multiple utilities to align and process raw RNA-seq reads to derive PSIs for all types of alternative Splicing24.

- Multivariate Analysis of Transcript Splicing (MATS, v3.08), which utilizes TopHat248 aligned reads and a custom splice-junction library. In order to account for the effects of covariates, we utilized PSI values in the linear mixed effects model described below for differential splicing analysis:

```
lme(PSI ~ diagnosis + age + sex + brain_region + sequencing.batch +
brain.bank.batch + RIN + seqSV1 + seqSV2, rand = ~1|individualID)
```

Where two sequencing surrogate variables (seqSV1 and seqSV2) were used as covariates.

RNA extraction and cDNA synthesis

Total tissue RNA was extracted from prefrontal cortex - BA8/9 of CTRL (n = 15) and idiopathic ASD patients (n = 16) and striatum, cortex or forebrain from Control, CPEB4 KO^{GT}/+, TgCPEB4Δ4, CPEB4 KO/+ and TgCPEB4Δ4:CPEB4 KO^{GT}/+, mice using the Maxwell® 16 LEV simplyRNA Tissue Kit (Promega, AS1280). Quantification and quality of RNA was done on a Nanodrop ND-1000 spectrophotometer and Nanodrop 1000 v.3.7.1 (Thermo Scientific). Retrotranscription (RT) reactions were performed using the iScript cDNA Synthesis kit (Bio-Rad, PN170-8891) following manufacturer´s instructions. Briefly, 1000 ng of total RNA from each samples were combined with 10 μl of master mix (includes all necessary reagents among which a mixture of random primers and oligo- dT for priming). The reaction volume was completed up to 40 μl with DNAse/RNAse free distilled water (Gibco, PN 10977). Thermal conditions consisted of the following steps: 5 min at 25ºC; 20 min at 46ºC and 1 min at 95ºC.

CPEB4 PCR flanking primers

Specific primers were designed in CPEB4 exon 2 (Forward, 5'-ggacgtttgacatgcactcac-3') and exon 5 (Reverse, 5'-gaggttgatccccacggc-3') and we verify that amplified the four CPEB4 splicing isoforms (Full-Length, Δ4, Δ3 and Δ3Δ4) in human and mouse brain cDNA. PCR amplification protocol used: 10 min 94ºC + 33 cycles (30s at 94ºC + 30s at 58ºC + 2 min at 72ºC) + 10 min at 72ºC. PCR products according with four CPEB4 isoforms were resolved on 2.2% agarose/gelgreen (Biotium, 41004) gels run at 125V for 1.5h. Images were scanned with densitometer (Bio-Rad, GS-900) and quantificated with Image Lab 5.2 (Bio-Rad). Finally, the percentage of each CPEB4 isoform was calculated.

Digital-Droplet PCR

mRNAs of each CPEB4 splicing isoforms were measured by digital-droplet PCR (ddPCR) in a BioRad QX200™ Digital Droplet™ PCR system (Bio-Rad, 1864100). All PCR reactions were made in three replicates and assayed in 96 well plates. The PCR reaction volume was 22 μL using the QX200™ ddPCR™ Evagreen Supermix (Bio-Rad, 1864034). Each reaction included 11 μL of ddPCR™ Evagreen Supermix, forward and reverse primers at 0.9 μM each and 4 ng of template cDNA. The PCR reaction mixture was loaded into an eightwell DG8™ Cartridge for QX200™ (Bio-Rad, 1864008) and droplets with Oil for EvaGreen (Bio-Rad, 1864005) were formed with the Bio-Rad QX200™ Droplet Generator (Bio-Rad, 1864101), following the manufacturer's instructions. During emulsion, the QX200 droplet generator partitions the samples into 20,000 nanolitre-sized droplets. The

droplets were then transferred to a 96-well plate and sealed with a Bio-Rad PX1™ PCR Plate Sealer (Bio-Rad, 1814000). Optimal ddPCR annealing temperatures for the CPEB4 isoforms assays were determined by incorporating a temperature gradient from 55.6ºC to 66.6ºC. FL-CPEB4 and CPEB4Δ3 were amplified using the following cycling conditions: 95ºC for 5 min, 40x (95ºC for 30 s and an annealing–extension step at 60ºC for 1 min) and 90ºC for 5 min. CPEB4Δ4 and CPEB4Δ3Δ4 were amplified using: 95ºC for 5 min, 40x (95ºC for 30 s + 62ºC for 1 min) and 90ºC for 5 min. Finally, droplets were read on the QX200™ Droplet Reader (Bio-Rad, 1864003) and data were analyzed with Quantasoft Version 1.6.6.0320 (Bio-Rad). The primers used to amplify each CPEB4 isoform are detailed in Supplementary Table 7.

Real-time quantitative reverse transcriptase-PCR

Quantification was performed by real-time PCR using a CFX 384 Real Time System C1000 Thermal Cycler (Bio-Rad) in combination with SsoFast Eva Green (Bio-Rad, CN 172-5204) and 0.25 μM of primer pair was used. Data were analyzed by GenEx 5.3.7 software (Multid AnaLyses AB). The mRNA levels were normalized first relative to total RNA and then relative to the 18S ribosome subunit, β-*ACTIN*, *GAPDH* and β-*TUBULIN* gene expression in each sample. Absolute quantitative PCR was performed to determine the percentage of each CPEB4 splicing isoform in both human and mouse species using specific primers (Supplementary Table 7). For every primer couple, specificity was tested, PCR assay conditions were adjusted to obtain a single amplicon analyzed by both melting curve analysis and agarose gel electrophoresis. Amplicons of each CPEB4 isoform were serially diluted to generate a calibration curve. A duplication of this curve was made to give robustness. Next, total tissue RNA was extracted, and quantitative real-time RT-PCR was performed. Finally, the percentage of each CPEB4 isoform with respect to total CPEB4 copies was calculated.

Western blot

Samples from human brain were stored at -80ºC and were ground with a mortar in a frozen environment with liquid nitrogen to prevent thawing of the samples, resulting in tissue powder. For mouse, brains were quickly dissected on an ice-cold plate and the different structures stored at -80ºC. Human and mouse extracts were prepared by homogenizing the brain areas in ice-cold extraction buffer (20 mM HEPES pH 7.4, 100 mM NaCl, 20 mM NaF, 1% Triton X-100, 1 mM sodium orthovanadate, 1 μM okadaic acid, 5 mM sodium pyrophosphate, 30 mM β-glycerophosphate, 5 mM EDTA, protease inhibitors (Complete, Roche, Cat. No 11697498001)). Homogenates were centrifuged at 15000g for 15 min at 4ºC. The resulting supernatant was collected, and protein content determined by Quick Start Bradford kit assay (Bio-Rad, 500-0203). Between 10 and 20 μg of total protein were electrophoresed on 10% SDS-polyacrylamide gel, transferred to a nitrocellulose blotting membrane (Amersham Protran 0.45 μm, GE Healthcare Life Sciences, 10600002) and blocked in TBS-T (150 mM NaCl, 20 mM Tris–HCl, pH 7.5, 0.1% Tween 20) supplemented with 5% non-fat dry milk. Membranes were incubated overnight at 4ºC with the primary antibody in TBS-T supplemented with 5% non-fat dry milk, washed with TBS-T and next incubated with secondary HRP-conjugated anti-mouse IgG (1:2000, DAKO, P0447), anti-rabbit IgG (1:2000, DAKO, P0448) or anti-rat IgG-Fc fragment (1:5000, Bethyl,

A110-136P) and developed using the ECL detection kit (PerkinElmer, NEL105001EA). Images were scanned with densitometer (Bio-Rad, GS-900) and quantificated with Image Lab 5.2 (Bio-Rad).

Antibodies—Rabbit CPEB1 (1:350, Santacruz, sc-33193); rabbit CPEB2 (1:1000, Abcam, ab51069); rabbit CPEB3 (1:1000, Abcam, ab10883); rabbit CPEB4 (1:1000, Abcam, ab83009); rabbit PTEN (1:1000, Cell Signaling, 9559S); mouse DYRK1A (1:1000, Abnova, H00001859-M01); rabbit FOXP1 (1:2000 for mouse and 1:500 for human samples, Abcam, ab16645); rabbit WAC (1:500, Merk Millipore, ABE471); rabbit AUTS2 (1:750, Sigma, HPA000390); mouse RBFOX1 (1:2000 for mouse and 1:1000 for human samples, Merk Millipore, MABE985), rabbit PCDH9 (1:500, Abcam, ab171166); rabbit ZBTB20 (1:300, SantaCruz, sc-99728); mouse CALB1 (1:1000, Sigma, C9848); rabbit D2R (1:800, Calbiochem, 324396), rabbit SNAP25 (1:2500, abcam, ab5666), mouse TUBB3 (1:2500, Novus, NB120-11314), rabbit IBA1 (1:1000, Wako, 019-19741), rabbit TNRC6B (1:500, Merk Millipore, AB9913); rat CHD2 (1:750, Merk Millipore, MABE873); rabbit GPC6 (1:1000, Abcam, ab136295); mouse β-ACTIN (1:25000, Sigma, A2228).

Poly(U) chromatography

Human samples: brain specimens from prefrontal cortex - BA8/9 of ASD patients (n = 6) and CTRL (n = 5) males with age between 5-23 years old. To verify RNA integrity and poly(A)-tail quality, we performed Hire-PAT of a typical normalizer gene (*ACTB*) and found any alterations in CTRL or ASD brains.

Mouse samples: WT, CPEB4 KO^{GT}/+ and CPEB4 KO (n=2) and Control vs. TgCPEB4Δ4 mice (n=3) were sacrificed by cervical dislocation at the age of 6 weeks. The cortex and striatum together were quickly dissected on an ice-cold plate.

Human and mouse samples were homogenized and total RNA was extracted using the Maxwell® 16 LEV simplyRNA Tissue Kit (Promega, AS1280), and stored at -80ºC until use. The poly(A) RNA fraction was purified by poly(U) chromatography49. Poly(U)-agarose (Sigma, p8563) was suspended in swelling buffer (0.05 M Tris-HCl, pH 7.5, 1 M NaCl) 35 ml/g, incubated overnight at room temperature and loaded into the chromatography column. An aliquot of total RNA was stored at -80ºC (“Input”) and the rest was incubated with sample buffer (0.01 M Tris-HCl, pH 7.5, 1 mM EDTA, 1% SDS) for 5 min at 65ºC and chilled on ice. Binding buffer was added (0.05 M Tris-HCl, pH 7.5, 0.7 M NaCl, 10 mM EDTA, 25% [v/v] formamide) and then the sample was loaded into the poly(U)-agarose chromatography column (Mobitec, M1002s) and incubated for 30 min at room temperature (25ºC) with agitation. Next, the column containing the sample was washed three times at 25ºC and six times at 55ºC with washing buffer (0.05 M Tris-HCl, pH 7.5, 0.1 M NaCl, 10 mM EDTA, 25% [v/v] formamide). The 55ºC washes were collected and stored at -80ºC (“Short poly(A)-tail fraction”). The remaining poly(A) RNA (“Long poly(A)-tail fraction”) was eluted with elution buffer (0.05 M HEPES, pH 7, 10 mM EDTA, 90% [v/v] formamide) at 55ºC and stored at -80ºC. The RNA of the two poly(A) fractions was precipitated by adding 1 volume of isopropanol, 1/10th volumes of sodium acetate 3 M pH 5.2 and 20 μg of glycogen (Sigma, G1767). The samples were incubated at -20ºC for 20

min and centrifuged 15 min at 14000g at 4ºC. The supernatant was removed and the pellet was washed with 750 μL of ethanol and centrifuged at 14000g and 4ºC for 5 min. The supernatant was removed and the pellet was air-dried for 5 min. The RNAs were resuspended in 300 μL of nuclease-free water and then 300 μL of acid Phenol:Chloroform (5:1) were added to them. Samples were vortexed and centrifuged for 10 min at 14000g and 4ºC. The aqueous phase was recovered, mixed with 1 volume of chloroform, vortexed and centrifuged again. The aqueous phase was recovered and precipitated again using the isopropanol precipitation. When setting up the method, we perform digestion of the non-poly(A) mRNA regions followed by end-labelling of the poly(A) for each eluted fraction and Urea-PAGE to confirm the average length in each fragment. We also compared poly(A)-tail by HIRE-PAT assay of control genes in Input, Washed and Eluted fractions to verify that it worked properly.

Human PrimeView and GeneAtlas MG-430 PM microarrays analysis

cDNA library preparation and amplification were performed according to the manufacturer's instructions (Sigma-Aldrich) for the WTA2 kit from 25 ng starting material. The cDNA was amplified for 17 cycles and purified using PureLink Quick PCR Purification Kit (Invitrogen, K310001). Quantification of amplified cDNA was done on a Nanodrop ND-1000 spectrophotometer (Thermo-Fisher Scientific, Waltham, MA, USA). 8.5 ug of the cDNA from each sample were fragmented and labeling with GeneChip Mapping 250K Nsp assay kit (Affymetrix, 900753) following the instructions of manufacturer.

Human: samples ready to hybridize were denatured at 99ºC for 2 min prior to incubation into the GeneChip Human PrimeView arrays (Affymetrix, 901838). Hybridization was performed for 16h at 45ºC / 60 rpm in the GeneChip Hybridization Oven 645 (Affymetrix, 00-0331). Washing and stain steps after hybridization were performed in the GeneChip Fluidics Station 450 (Affymetrix, 00-0079), following the specific script for PrimeView arrays. Finally, the arrays were scanned with GeneChip Scanner GCS3000 (Affymetrix) using default parameters, and the generation of CEL files for bioinformatics analysis was done with Command Console software (Affymetrix).

Mouse: hybridization was performed using the GeneAtlas Hyb, Wash and Stain Kit for 3' IVT arrays. Samples ready to hybridize were denatured at 96ºC for 10 min prior to incubation into Mouse MG-430 PM Array Strip (Affymetrix, 901570), the hybridization was performed for 16 h at 45ºC in the GeneAtlas Hybridization Oven (Affymetrix, 00-0331). Washing and stain steps after hybridization were performed in the GeneAtlas Fluidics Station (Affymetrix, 00-0079), following the specific script for Mouse MG-430 PM Arrays. Finally, the arrays were scanned with GeneAtlas Scanner (Affymetrix) using default parameters, and the generation of CEL files for bioinformatics analysis was done with GeneAtlas software (Affymetrix).

Processing of microarray samples was performed using R50 and Bioconductor51. Raw CEL files were normalized using RMA background correction and summarization52. Standard quality controls were performed in order to identify abnormal samples53 regarding: a) spatial artifacts in the hybridization process (scan images and pseudo-images from probe level models); b) intensity dependences of differences between chips (MvA plots); c) RNA

quality (RNA digest plot); and d) global intensity levels (boxplot of perfect match log-intensity distributions before and after normalization and RLE plots). Probeset annotation was performed using the information available in Affymetrix web page (https://www.affymetrix.com/analysis/index.affx) using version na35.

Expression values were adjusted for technical biases as described in54 using a linear model and implemented with the R package "limma"55. For each biological replicate the log2 fold change was computed between "WASH" and "ELUTED" samples and used to find significant differences between WT vs. CPEB4 KO^{GT}/+ and CPEB4 KO mice (n = 2), control vs. TgCPEB4Δ4 mice (n = 3) and human CTRL (n = 5) vs ASD patients (n = 6). Differential expression was performed using a linear model with fluidics and amplification batch as covariates. *P*-values were adjusted with the Benjamini and Hochberg correction. We considered one transcript is shortened when *P*-value < 0.05 and FC is negative and lengthened when *P*-value < 0.05 and FC is positive, in at least one probe. If the same transcript showed opposite results for different probes, it was considered as not changed.

Differential gene expression array—Individual probeset expression values for each selected gene were annotated with Annmap webservice (http://annmap.cruk.manchester.ac.uk), using Homo sapiens v84 Primeview Human Gene Array and *Mus musculus* v84 Mouse Genome 430A 2.0 databases. Those probesets annotated as "reliable" were preferentially selected for analysis. For those genes lacking a reliable probeset, the whole group of probesets was taken for comparisons. Graph bars were plotted using the mean of RMA normalized expression values from the Primeview human gene array in case of human samples and Mouse genome 430 PM array in case of mice. The expression values were calculated using R50 and BioConductor51 packages.

High-Resolution poly(A) tail (HIRE-PAT) assay

USB® Poly(A) Tail-Length Assay Kit (Affymetrix, 76455) based on HIRE-PAT method, was used. Frontal cortex - BA8/9 of ASD patients (n = 3) and CTRL (n = 3) and total striatal RNA from 1.5 month-old control and TgCPEB4Δ4 mice (n = 3) was extracted using the Maxwell® 16 LEV simplyRNA Tissue Kit (Promega, AS1280) and stored at -80ºC until use. G/I tailing (1 μg of total RNA) and reverse transcription were performed according to the manufacturer's instructions. Poly(A) size was determined by subtracting the PCR amplicon size obtained with the Universal primer and forward specific primers. To verify that the measured poly(A) tail corresponds to specific gene three different forward specific primers were tested (Supplementary Table 7). PCR products were resolved on 2.5% agarose/gelgreen (Biotium, 41004) gels run at 120V for 1.5h.

Gene Ontology analysis

CPEB4 binders with poly(A)-tail changes in prefrontal cortex of ASD individuals were analyzed with DAVID Bioinformatics Resources 6.7, KEGG pathway annotation56.

Human-mouse altered poly(A) tail length geneset comparison

In order to compare transcripts with altered poly(A) tail in ASD patients and CPEB4 modified mice, we first converted the gene set from mouse into their human orthologous

(18649 total orthologous genes) using Ensembl Genes 85 *Mus musculus* GRCm38p4 Biomart (http://www.ensembl.org/biomart/)39. Then we calculated the statistical significance of the overlap between genes with poly(A) changes in human ASD and CPEB4-modified mice by hypergeometric distribution test. We considered overlapping when the representation factor is > 1 and *P*-value < 0.05, and dissimilar when the representation factor is < 1 and *P*-value < 0.05.

Generation of TgCPEB4Δ4 mice

Human CPEB4 cDNA lacking exon 4 (CPEB4Δ4) was cloned into a plasmid containing a bidirectional TetO sequence to also express LacZ reporter with a nuclear localization signal (pBI-G, Clontech, 631004). The construct was microinjected into single-cell CBAxC57BL/6 embryos and resulting TetO βGAL/CPEB4Δ4 founder mice were backcrossed with WT C57BL/6J mice (TetO βGAL/CPEB4Δ4). TetO βGAL/CPEB4Δ4 mice were crossed with CamkII-tTA (tTA) mice37 to obtain the conditional double transgenic mice with forebrain neuron expression of CPEB4Δ4 (TgCPEB4Δ4 mice). There are different CamKII-tTA transgenic mouse lines37 and, for this study, we chose one with expression starting at late embryonic age57. Upon observation of premature death of the subset of TgCPEB4Δ4 mice showing cranial dysmorphology, these were systematically culled when found.

To generate TgCPEB4Δ4 mice with transgene expression starting after weaning (OFF/ON-TgCPEB4Δ4 mice) pregnant females were isolated and kept on doxycycline (Sigma, D9891, 0.5g/L) until weaning of the litter. After weaning, the progeny was switched to plain water to allow expression of the transgene.

To generate TgCPEB4Δ4 mice that expressed the transgene only during embryonic and early postnatal development (ON/OFF-TgCPEB4Δ4 mice) pregnant females were isolated and kept on plain water until birth of the litter, when water was replaced by doxycycline solution (Sigma, D9891, 2g/L) so doxycycline intake starts in the pups through the milk. The progeny is kept on doxycycline throughout the rest of the experiment.

Immunohistochemistry

Mice were euthanized with CO_2. Brains were immediately removed and dissected on an ice-cold plate and left hemispheres, processed for histology, were placed in 4% paraformaldehyde in Sorensen's phosphate buffer overnight and then immersed in 30% sucrose in PBS for 72h. Once cryoprotected, the samples were included in optimum cutting temperature (OCT) compound (Tissue-Tek, Sakura Finetek Europe, 4583), frozen and stored at -80ºC until use. 30 μm sagittal sections were cut on a cryostat (Thermo Scientific), collected and stored free floating in glycol containing buffer (30% glycerol, 30% ethylene glycol in 0.02 M phosphate buffer) at -20ºC. Before staining, sections were washed with PBS to eliminate the cryoprotective buffer and immersed in 0.3% H_2O_2 in PBS for 30 min to quench endogenous peroxidase activity. Sections were immersed for 1h in blocking solution (PBS containing 0.5% Fetal Bovine Serum, 0.3% Triton X-100 and 1% BSA) and incubated overnight at 4ºC with the corresponding primary antibody diluted in blocking solution. After washing, brain sections were incubated first with biotinylated anti-rabbit or anti-mouse secondary antibody and then with avidin-biotin complex using the Elite Vectastain kit

(Vector Laboratories, PK-6101-2). Chromogen reactions were performed with diaminobenzidine (SIGMAFASTTM DAB, Sigma, D4293) for 10 min. Sections were mounted on glass slides and coverslipped with Mowiol (Calbiochem, 475904). Images were captured using an Olympus BX41 microscope with an Olympus camera DP-70 and Olympus cellSens Entry v.1.7 (Olympus Denmark A/S)

Antibodies—Rabbit CPEB4 (1:1000, Aviva, ARP41024_P050); rabbit β-GAL (1:2000, Invitrogen, A-11132); rabbit cleaved CASP3, Asp175 (1:60, Cell Signaling, 9661)

Golgi spine analysis

Control (n = 5) vs. TgCPEB4Δ4 mice (n = 4) three month-old mice, and WT (n = 3) vs. CPEB4 KO^{GT}/+ (n = 4) fifteen month-old were completely anesthetized with an intraperitoneal pentobarbital injection (60 mg/kg Dolethal®, Vetoquinol). The whole brain was extracted and immersed in Golgi-Cox staining solution (FD Rapid GolgiStainTM kit, FD Neurotechnologies, cat. PK401). 150 μm sagittal sections were obtained in a Leica VT1200S vibratome and mounted on gelatin-coated slides. Golgi staining was performed as manufacturer's instructions. Afterwards, all sections were counterstained with toluidine blue pH 4.0 (1 g/l Toluidine Blue (Sigma, 198161), 0.8 M glacial acetic acid) and coverslipped with DePeX (Amsbio, 18243.02). Pyramidal neurons from layer II/III of the cortex were identified by their distance from pia mater and their distinct morphologies. Secondary, tertiary and quaternary dendrites of these neurons were selected for analysis. Z-stacks of the entire apical dendritic tree of Golgi stained pyramidal neurons (up to 80 μm total on Z-axis, optical section thickness = 0.5 μm) were taken at 40x magnification with 2x optical zoom on a vertical Zeiss Axio Imager.Z1 M and analyzed by Laser Scanning Microscope LSM 510 v. 4.2 SP1 (Carl Zeiss). Spine density, length and classification were performed according to58, unbiased blinded to genotype.

Brain weight and volumetric analysis

1.5 month-old mice were completely anesthetized with an intraperitoneal pentobarbital injection (60 mg/kg Dolethal®, Vetoquinol). The whole brain was extracted and weighted in a precision scale (Metter Toledo, AB265-S). Left hemispheres were fixed in 4% paraformaldehyde, immersed in 30% sucrose, included in OCT compound (Tissue-Tek, Sakura Finetek Europe, 4583), frozen and stored at -80ºC. Sagittal sections (30 μm thick) were cut on a cryostat and every sixth section was counterstained with toluidine blue pH 4.0 (1g/l Toluidine Blue (Sigma, 198161), 0.8 M glacial acetic acid). Digital images were captured at a 2.5x magnification (Canon EOS 450D digital camera) and the hippocampal, striatal and motor and somatosensory cortical area from 20-22 sections for each animal was calculated by means of the ImageJ software59. Considering a separation of 180 μm between each section, total structure volume in each mouse was calculated.

Electrophysiology

For preparation of acute brain slices, we based on the N-methyl-D-glucamine (NMDG) protective recovery method according to60–62. Briefly, 5-6 weeks- old control vs. TgCPEB4Δ4 (n = 5) mice and WT vs. CPEB4 KO^{GT}/+ (n = 5) mice of both sexes were anesthetized with 2% tribromoethanol (0.15 ml/10 mg) and rapidly decapitated. The brains

were dissected out and transferred to NMDG ice-cold artificial cerebrospinal fluid (ACSF) composed of (in mM): 93 NMDG, 2.5 KCl, 1.2 NaH_2PO_4, 30 $NaHCO_3$, 25 D-glucose, 20 HEPES, 5 Na-ascorbate, 2 thiourea, 3 Na-pyruvate, 10 $MgSO_4$, and 0.5 $CaCl_2$. The pH of the solution was titrated to pH 7.3-7.4 with concentrated HCl (osmolality 310-315 $mOsmol \cdot kg^{-1}$) and bubbled with carbogen (5% CO_2 - 95% O_2). 350 μm coronal slices were cut on a Vibratome VT1200S (Leica) and transferred for initial recovery to NMDG ACSF at 33 ± 1ºC. Finally, slices were placed in a holding chamber at room temperature with normal ACSF composed of (in mM): 126 NaCl, 2.5 KCl, 2 $CaCl_2$, 2 $MgCl_2$, 1.25 NaH_2PO_4, 26 $NaHCO_3$, and 10 D-glucose (osmolality 305–315 $mOsmol \cdot kg^{-1}$), pH 7.4, when bubbled with carbogen (5% CO_2–95% O_2).

Recordings—For whole-cell patch-clamp recordings, slices were transferred into a recording chamber that was perfused with 33 ± 1ºC bubbled ACSF at 2–3 ml/min. Pyramidal neurons of the somatosensory cortex were visualized by a Nikon Eclipse FN1 microscope, a 40x water immersion objective (Nikon), and a USB 2.0 monochrome camera (DMK 31BU03.H, TheImagingSource). Whole-cell recordings were performed using a double patch clamp EPC10 plus amplifier (HEKA). Under voltage-clamp conditions, the patch-pipettes for excitatory postsynaptic currents recording (EPSCs) contained (mM): 120 K-gluconate, 10 KCl, 10 phosphocreatine disodium salt, 2 MgATP, 0.3 NaGTP, 0.1 EGTA, 10 HEPES, pH 7.2 adjusted with KOH, osmolality 280-290 $mOsmol \cdot kg^{-1}$. Recording of miniature EPSCs (mEPSCs) were done in the presence of tetrodotoxin (1 μM) and picrotoxin (50 μM) to block sodium channels and $GABA_A$ receptors, respectively. Cells were held in voltage-clamp mode at a holding potential (Vhold) of -70 mV, while resistance was compensated by 70% (lag 10 μs). Recordings were discontinued if series resistances increased by > 50% or exceeded 15 MΩ.

Currents were low-pass filtered at 3 kHz, digitized at 20 kHz, and acquired using PatchMaster software (HEKA). All miniature postsynaptic currents were analyzed with the program Stimfit63. Recordings were first digitally filtered at 1 kHz. For each cell, all events were inspected to avoid false-positive events, and then an average of all events detected was made.

Cleaved caspase-3 quantification

1.5 month-old control and TgCPEB4Δ4 mice (n = 6) were analyzed. The total number of immunopositive cells with apoptotic shape was quantified in the cortex of three sections per animal using an Olympus BX41 microscope with an Olympus camera DP-70 (Olympus Denmark A/S). Means of the three sections were calculated.

Stereology

Sagittal sections (30 μm thick) counterstained with toluidine blue pH 4.0 (1g/l Toluidine Blue (Sigma, 198161), 0.8 M glacial acetic acid) from the volumetric analysis were used. Sections containing striatum were selected and the 10 most central sections were analyzed. One randomly selected 60 μm x 60 μm optical dissector at 60x magnification with an Olympus BX41 microscope with an Olympus camera DP-70 (Olympus Denmark A/S) was analyzed in each section. Total neuronal cell number per dissector was assessed by a

researcher blind to genotype. Striatal neuronal cell density was calculated and compared for control (n = 19) and TgCPEB4Δ4 mice (n = 5).

Behavioral testing

Open Field—Locomotor activity was measured in 5 week-old mice in clear Plexiglas® boxes measuring 27.5 cm x 27.5 cm, outfitted with photo-beam detectors for monitoring horizontal and vertical activity. Activity levels were recorded with a MED Associates' Activity Monitor (MED Associates, St. Albans, VT) and were analyzed with the MED Associates' Activity Monitor Data Analysis v.5.93.773 software. Mice were placed in the center of the open-field apparatus and left to move freely. Data were individually recorded for each animal during 15 min. Distance walked in the periphery (3.5 cm from the edges) and in the center of the box was measured.

Ultrasonic Vocalization (UsV)—Numbers of UsVs were measured at the age of 3, 6, 9 and 12 postnatal days in mice. Dam was removed from a temperature-controlled home cage where the pups remained. Then, pups were removed individually and placed in a plate equipped to record UsV for 5 min (Avisoft Recorder). To avoid potential confounding effects due to temperature, the room was maintained at 21ºC and body temperature was measured with an axillary probe after the 5 min test. UsV was analyzed with Avisoft SASLab Pro v.5.2.09 software.

Social Approach—Social interaction was examined in 5 week-old mice. The first day (training), mice were allowed to explore an empty Plexiglas® box measuring 45 cm x 45 cm during 10 min. The next day (test), mice were placed in the same box containing two wire cages placed in opposite corners, one empty and the other with an unknown (gender paired) mouse on it. Mice were recorded during 10 min and the time expended interacting with each cage was measured.

Elevated plus maze (EPM)—Anxiety-like behavior was examined in 5 week-old mice. Animals were tested in a 5 min single trial EPM in which the mouse was allowed to move freely along the apparatus under a constant intense white light. Animal movement was recorded and the total time spent standing or walking on the open and closed arms was measured. The criterion was the head, forelimbs and hindlimbs being placed on open or closed arms. Maze consists of four arms (two open without walls and two enclosed by 15 cm high walls) 26 cm long and 5 cm wide, and it is elevated 40 cm off the floor.

Data analysis

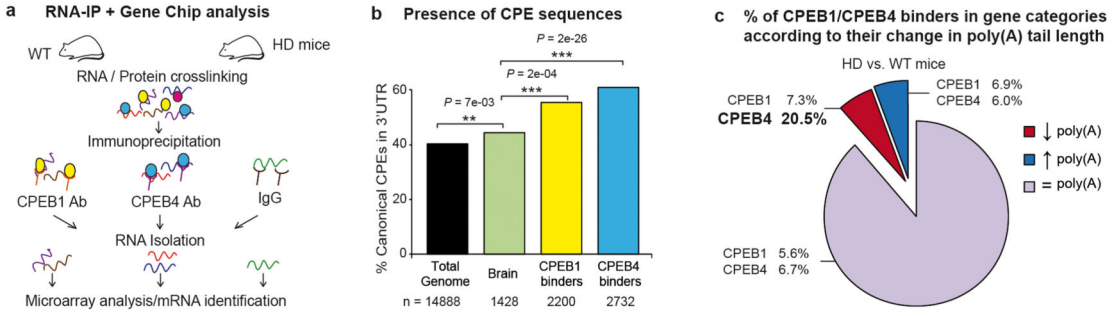
Statistical analysis was performed with SPSS 21.0 (SPSS® Statistic IBM®). Data are represented as Mean ± s.e.m (Standard Error of the Mean) with 95% confidence interval. In box plots, box segments show median, 25th and 75th percentiles, whiskers above and below show the locations of the minimum and maximum. Higher or lower points (outliers) are plotted individually or not plotted. The normality of the data was analyzed by Shapiro-Wilk test ($n < 50$) or o Kolmogorov-Smirnov ($n > 50$). Homogeneity of variance was analyzed by Levente test. For comparison of two independent groups two-tail unpaired t-Student's test (data with normal distribution), Mann-Whitney-Wilcoxon or Kolmogorov-Smirnov tests

(with non-normal distribution) was performed. To compare dependent measurements, we used a paired t-test (normal distribution) or Wilcoxon signed-rank tests (non-normal). For multiple comparisons, data with a normal distribution were analyzed by one way-ANOVA test followed by a Tukey's or a Games-Howell's post-hoc test. Statistical significance of non-parametric data for multiple comparisons was determined by Kruskal-Wallis One-way ANOVA test. Enrichment tests were carried out by using one-sided Fisher's exact test. A critical value for significance of $P < 0.05$ was used throughout the study.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

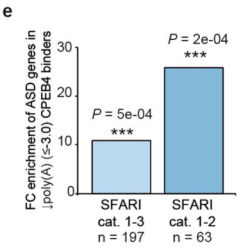
Extended Data

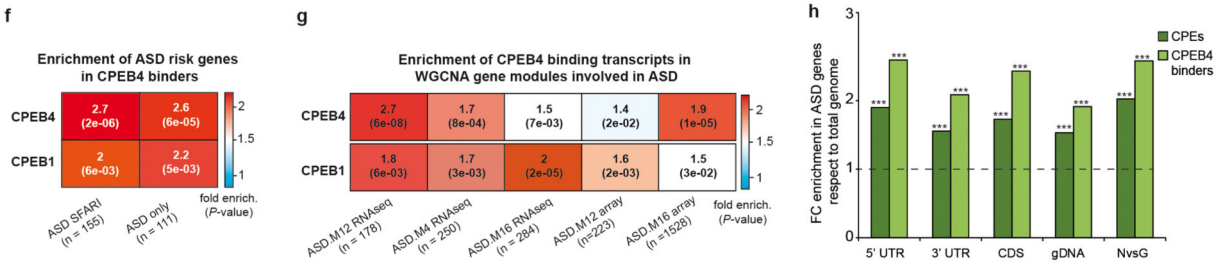


d CPEB4 binders with shortened poly(A)-tail length (FC ≤ -3.0)

Nº	FC	Symbol	Gene Name	CPEB1-binding
1	-5.4	Auts2	Autism susceptibility candidate 2	Y
2	-5.3	Akap10	A kinase (PRKA) anchor protein 10	Y
3	-4.9	Rock1	Rho-associated coiled-coil containing protein kinase 1	Y
4	-4.5	Ktn1	Kinectin 1	Y
5	-3.9	Slk	STE20-like kinase	Y
6	-3.9	Kpnb1	Karyopherin (importin) beta 1	Y
7	-3.9	Ssr3	Signal sequence receptor, gamma	Y
8	-3.8	Rp2h	Retinitis pigmentosa 2 homolog (human)	N
9	-3.8	Bach2	BTB and CNC homology 2	Y
10	-3.7	Cpsf6	Cleavage and polyadenylation specific factor 6	N
11	-3.7	Pan3	PAN3 polyA specific ribonuclease subunit homolog	N
12	-3.6	Tmx3	Thioredoxin-related transmembrane protein 3	N
13	-3.6	Ube2d3	Ubiquitin-conjugating enzyme E2D 3	N
14	-3.6	Epha3	Eph receptor A3	N
15	-3.6	Zfx	Zinc finger protein X-linked	Y
16	-3.5	Crbn	Cereblon	Y
17	-3.4	Rnd3	Rho family GTPase 3	N
18	-3.4	Dyrk1a	Dual-specificity tyrosine-phosphorylation regulated kinase 1a	N
19	-3.4	Hey2	Hairy/enhancer-of-split related with YRPW motif 2	N
20	-3.4	Mynn	Myoneurin	N
21	-3.3	Cers6	Ceramide synthase 6	Y
22	-3.2	Klhl28	Kelch-like 28	N
23	-3.2	Plagl1	Pleiomorphic adenoma gene-like 1	Y
24	-3.2	Dtwd2	DTW domain containing 2	N
25	-3.2	9630041G16Rik	RIKEN cDNA 9630041G16 gene	N
26	-3.2	Arl6ip6	ADP-ribosylation factor-like 6 interacting protein 6	N
27	-3.2	Ccdc88a	Coiled coil domain containing 88A	N
28	-3.2	Sp3	Trans-acting transcription factor 3	N
29	-3.2	Ctso	Cathepsin O	Y
30	-3.2	Cul3	Cullin 3	Y
31	-3.2	2810474O19Rik	RIKEN cDNA 2810474O19 gene	Y
32	-3.1	Tmco5	Transmembrane and coiled-coil domains 5	N
33	-3.1	Camk4	Calcium/calmodulin-dependent protein kinase IV	Y
34	-3.1	Zhx1	Zinc fingers and homeoboxes 1	Y
35	-3.1	Cep290	Centrosomal protein 290	N
36	-3.0	Eif4a2	Eukaryotic translation initiation factor 4A2	Y
37	-3.0	Wnt5a	Wingless-related MMTV integration site 5A	Y
38	-3.0	Metap2	Methionine aminopeptidase 2	N
39	-3.0	Ptchd1	Patched domain containing 1	N
40	-3.0	Birc6	Baculoviral IAP repeat-containing 6	Y

ASD causing genes





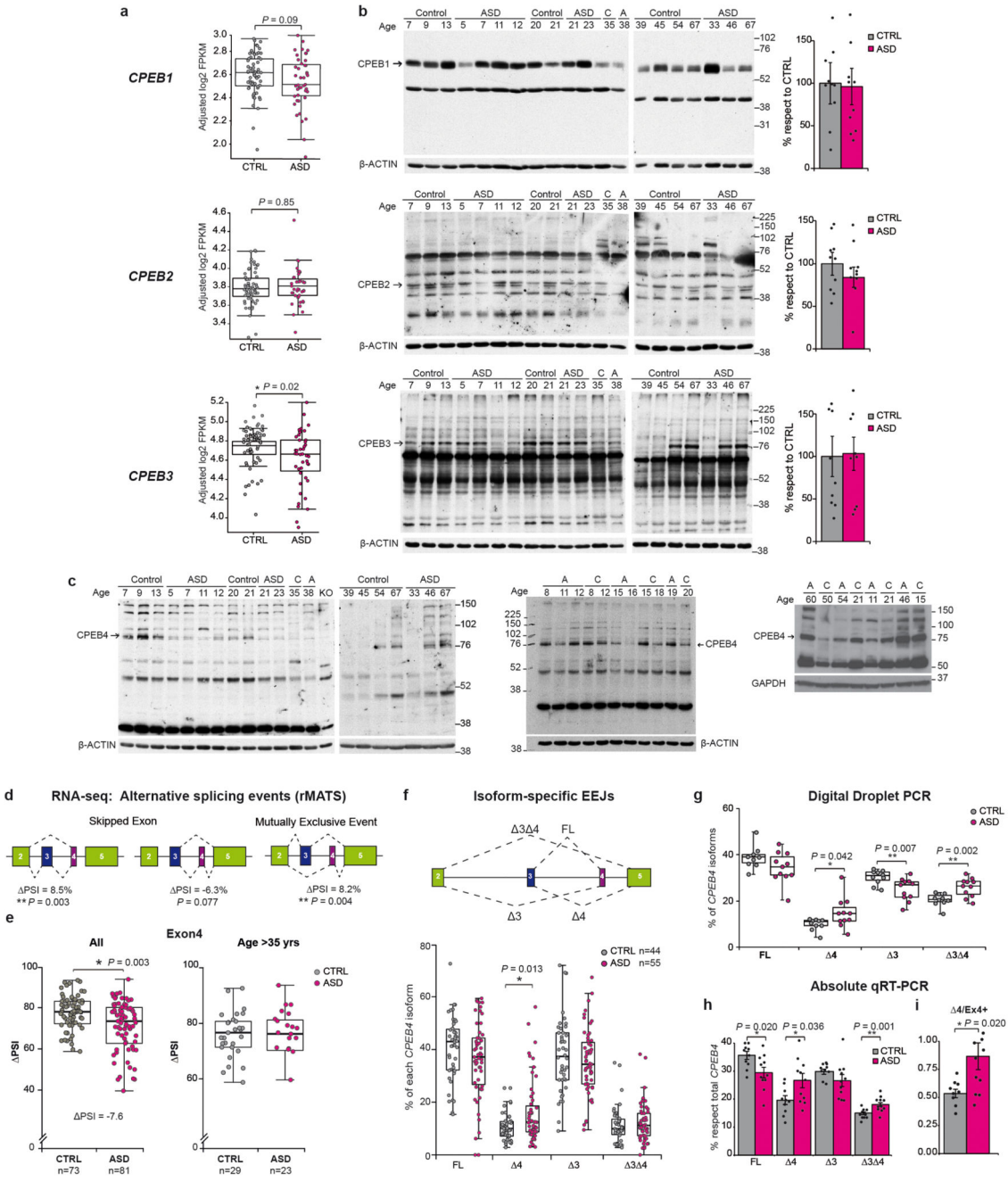
Extended Data Fig. 1. Enrichment in ASD-risk genes among CPEB1-4 binding transcripts whose poly(A)-tail is shortened in an HD mouse model with altered CPEBs.

a, Experimental design of RIP from WT and HD mice (with altered CPEB1 and CPEB4, see methods). **b,** Percentage of CPE sequences in the 3' UTR of total genome, brain genes and CPEB1 and CPEB4 binders from RIP experiment. **c,** Percentage of CPEB1- or CPEB4-only binders with shortened (red), lengthened (blue) or unaltered (purple) poly(A) tail. **d,** Symbol and gene names of CPEB4 binders in WT St with the most shortened poly(A)-tail (FC ≤

Europe PMC Funders Author Manuscripts

-3.0) in HD mice. The last column indicates whether they are also CPEB1 binders (Y, yes; N, no). High-confidence ASD-risk genes (SFARI cat. 1–3) are highlighted in pink. **e,** Fold-change enrichment of high-confidence ASD genes (SFARI cat. 1–3 and cat. 1–2) in CPEB4 binders whose poly(A) tail is shortened in HD mice (FC ≤ -3.0). Heatmaps of CPEB4/CPEB1 binders **f,** in SFARI ASD genes or removing intellectual disability genes (ASD only) and **g,** in weighted gene co-expression network analysis (WGCNA) modules involved in ASD. **h,** Fold change enrichment of percentage of CPE sequences and CPEB4 binders of ASD genes (SFARI cat. 1-2, n = 63) vs. total genome stratified by 5'UTR, 3'UTR, CDS, gDNA length and ratio neuronal/glial expression. **b, e-g,** One-sided Fisher's exact test. **c,** Pearson's chi-squared test. **h,** Statistical details in simulations in method section. ** $P < 0.01$, *** $P < 0.001$.

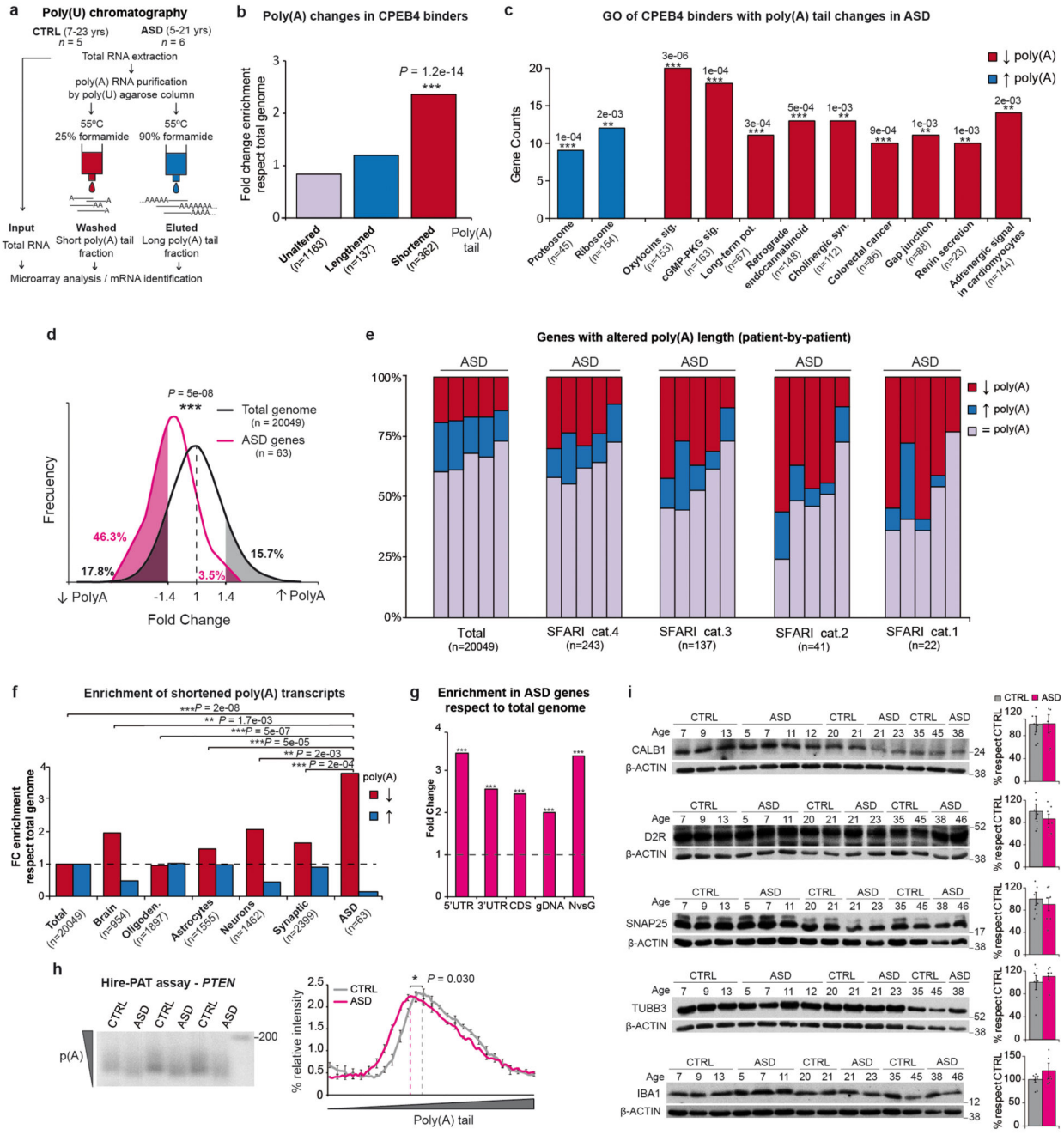
Europe PMC Funders Author Manuscripts



Extended Data Fig. 2. mRNA and protein levels of CPEBs in Cx of idiopathic ASD individuals and features of CPEB4 mis-splicing.

a, *CPEB1-3* mRNA expression levels according to RNA-seq data (n = 63 for control, n = 43 for ASD). **b,** CPEB1-3 protein levels (n = 10) and **c,** CPEB4 protein levels (n = 20 for control and n = 19 for ASD). **d,** Diagram representing the alternative splicing events of CPEB4 by rMATS. Percent Spliced in (PSI) is shown under each event (n = 81 for CTRL and n = 82 ASD cortical prefrontal and temporal samples). **e,** *CPEB4* exon 4 inclusion level (PSI) in all (left panel) and over 35-year-old (right panel) individuals, and **f,** percentage of

each *CPEB4* splicing isoform by vast-tools analysis of isoform-specific EEJs (exon-exon junctions). Percentage of each *CPEB4* splicing isoform by **g,** Digital Droplet PCR and **h,** Absolute qRT-PCR, i, Δ4/Ex4+ *CPEB4* isoform ratio in Cx of idiopathic ASD cases (n = 11) and CTRL (n = 10) under 35-year-old. For gel source data, see Supplementary Figure 1. **a, d-g, i,** Two-sided Mann -Whitney-Wilcoxon test. **b, h,** Two-sided unpaired t-test. Box plots show median, 25th, 75th percentiles. Data are mean ± s.e.m. 95% CIs. **P* < 0.05, ***P* < 0.01.

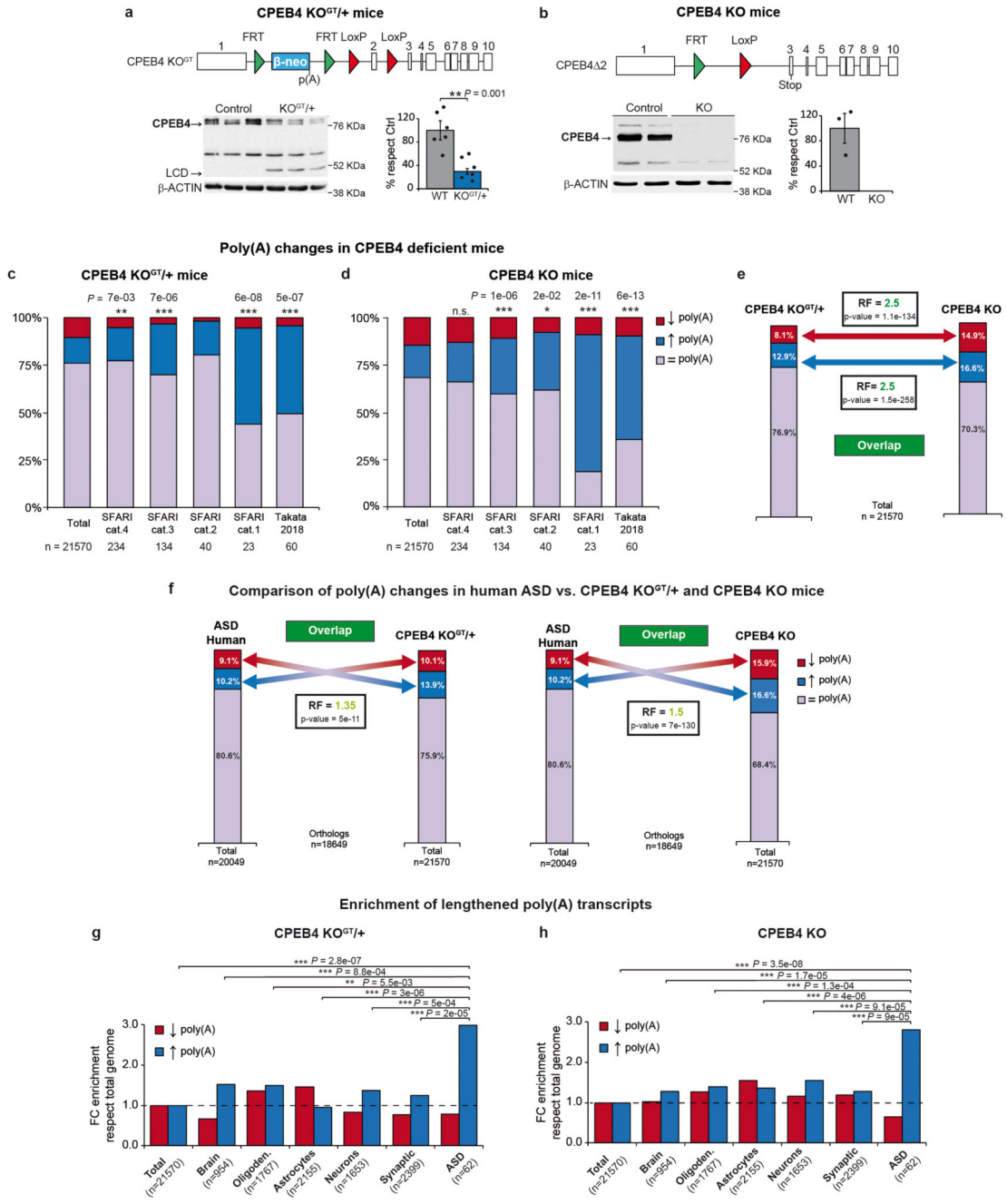


Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Extended Data Fig. 3. Supplemental data of global poly(A)-alteration and protein levels in idiopathic ASD brains.

a, Experimental design. **b,** Poly(A) changes of CPEB4 binders. **c,** Gene counts histogram from Gene Ontology (GO) analysis (KEGG pathways) of genes with poly(A) tail changes. **d,** Frequency distribution of fold changes of poly(A) alteration of total genes (in black) and ASD genes (SFARI cat. 1-2, in pink). **e,** Percentage of genes with shortened (red), lengthened (blue) or unaltered (purple) poly(A)-tail length in the whole transcriptome and ASD genes (SFARI cat. 4 to cat. 1) patient-by-patient. **f,** Fold change enrichment of brain, oligodendrocytic, astrocytic, neuronal, synaptic and ASD specific genes (SFARI cat.1-2) with shortened poly(A)-tail respect to total genome. **g,** Fold change enrichment of ASD (SFARI cat. 1-2) genes shortened in ASD human vs. total genome stratified by 5'UTR, 3'UTR, CDS, gDNA length and ratio neuronal/glial expression. **h,** Hire-PAT assay of *PTEN* poly(A)-tail in CTRL- and ASD cases (n = 3). **i,** Protein levels of neuronal and astrocytic specific genes in Cx of idiopathic ASD cases and CTRL (n = 7). For gel source data, see Supplementary Figure 1. **b, f,** One-sided Fisher's exact test, **c,** FDR Benjamini-Hochberg. **d,** Two-sided Mann-Whitney-Wilcoxon test. **f,** *P*-values of genes with shortened poly(A) in each group respect to ASD genes. **g,** Statistical details in simulations in method section. **h, i,** Two-sided unpaired t-test. Data are mean ± s.e.m. 95% CIs. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

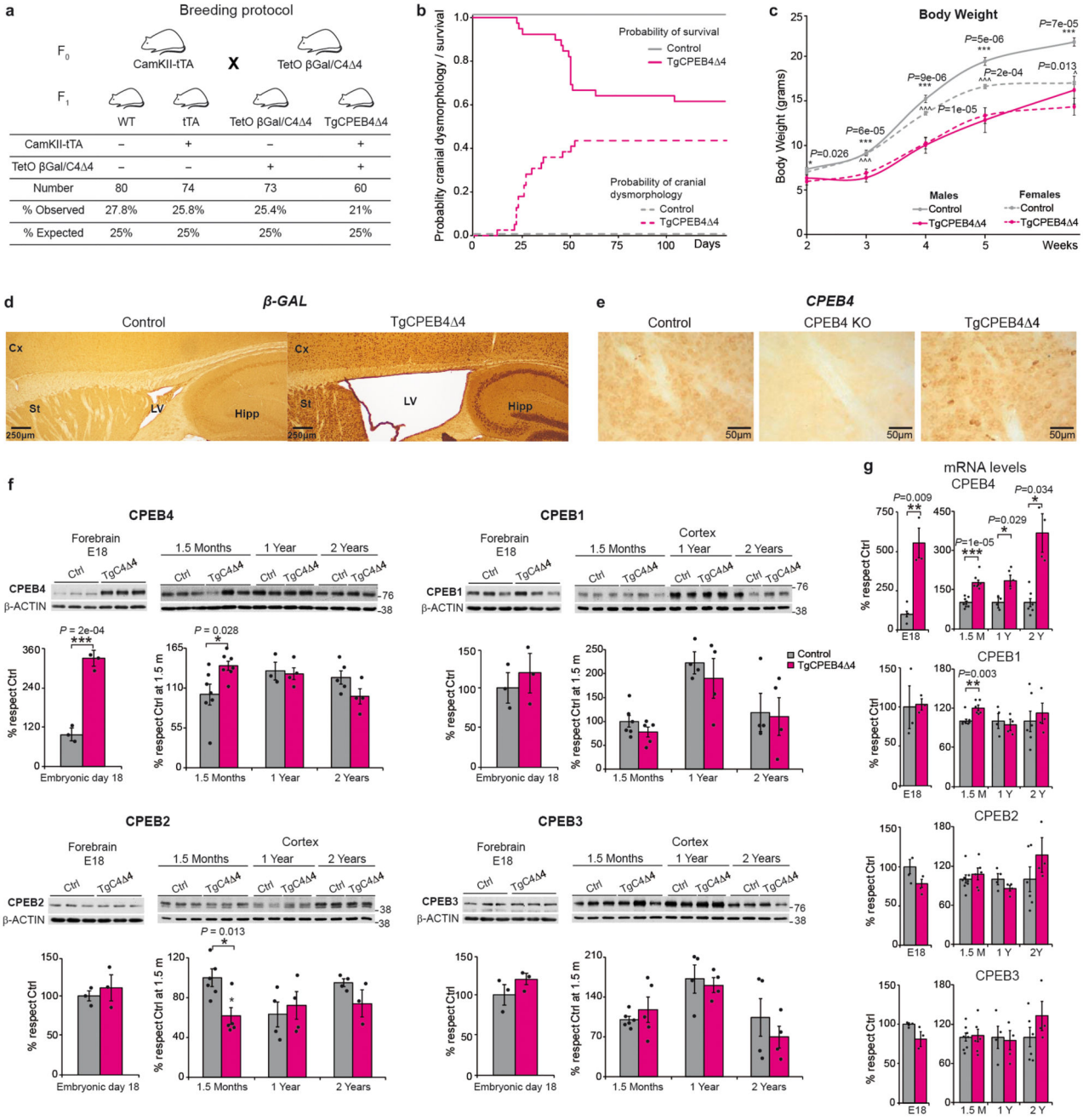


Extended Data Fig. 4. Poly(A) changes in CPEB4-deficient mice.

Constructs design and CPEB4 protein levels of **a,** CPEB4 KO^{GT}/+ (n = 7), **b,** CPEB4 KO (n = 3). Low complexity domain (LCD) isoform. **c, d,** Percentage of transcripts with poly(A)-tail changes in c, CPEB4 KO^{GT}/+ **d,** CPEB4 KO Cx-St samples (n = 2), in whole transcriptome and in ASD gene-lists. **e,** Comparison of genes with poly(A) changes between CPEB4 KO^{GT}/+ and CPEB4 KO mice, representation factor (RF). **f,** Comparison of genes with poly(A) changes between ASD cases and CPEB4-deficient mice. **g, h,** Fold change enrichment of brain, oligodendrocytic, astrocytic, neuronal, synaptic and ASD specific genes

Europe PMC Funders Author Manuscripts

(SFARI cat. 1-2) with lengthened poly(A)-tail respect to total transcriptome in **g,** CPEB4 KO^{GT}/+ mice and **h,** CPEB4 KO mice. For gel source data, see Supplementary Figure 1. **a,** Two-sided unpaired t-test. **c, d,** One-sided Fisher's exact test, *P*-values of ASD transcripts with lengthened poly(A) vs. Total. **e, f,** Hypergeometric test. **g, h,** One-sided Fisher's exact test, *P*-values of genes with lengthened poly(A) in each group respect to ASD genes. Data are mean ± s.e.m. 95% CIs. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

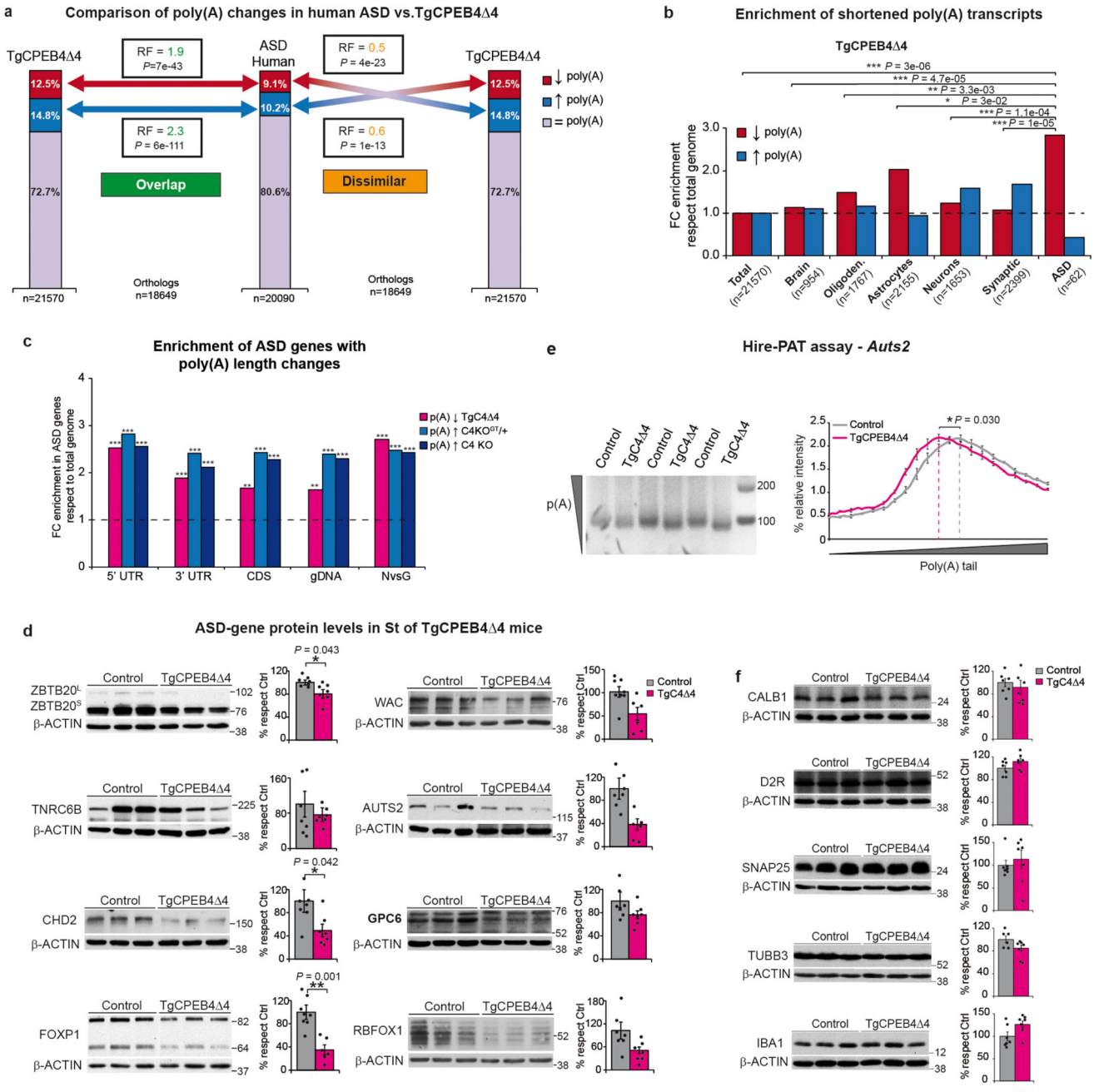


Extended Data Fig. 5. Supplemental characterization of TgCPEB4Δ4 mice.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

a, Breeding protocol to obtain TgCPEB4Δ4 mice. Number of mice and percentages of births observed and expected for the four experimental genotypes. **b,** Kaplan-Meier curve for cumulative survival (continuous line) and probability of developing cranial dysmorphology (dashed line), (n = 44 for control, n = 39 for TgCPEB4Δ4 mice). **c,** Evolution of mice body weight (grams). Males (continuous line), n = 25 controls, n = 9 TgCPEB4Δ4 mice. Females (dashed line), n = 26 control, n = 7 TgCPEB4Δ4 mice. **d,** β-GAL nuclear staining in forebrain neurons from 1.5-month-old controls (n = 6) and TgCPEB4Δ4 mice (n = 4). Cx, cortex; St, striatum; Hipp, hippocampus; LV, lateral ventricle. Scale bars represent 250 μm. **e,** St CPEB4 immunohistochemistry shows cytoplasm pattern in control (n = 6), no staining in CPEB4 KO (n = 2) and overexpressing neurons in TgCPEB4Δ4 mice (n = 4). Scale bars represent 50 μm. **f,** Protein and g, mRNA expression levels of CPEB1-4 in forebrain at embryonic day 18 (n = 3) and Cx at 1.5 months (n = 6), 1 year (n = 4) and 2 years (n = 5) of control and TgCPEB4Δ4 mice. For gel source data, see Supplementary Figure 1. **a,** Pearson's chi-squared test. **c, f, g,** Two-sided unpaired t-test. Data are mean ± s.e.m. 95% CIs. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

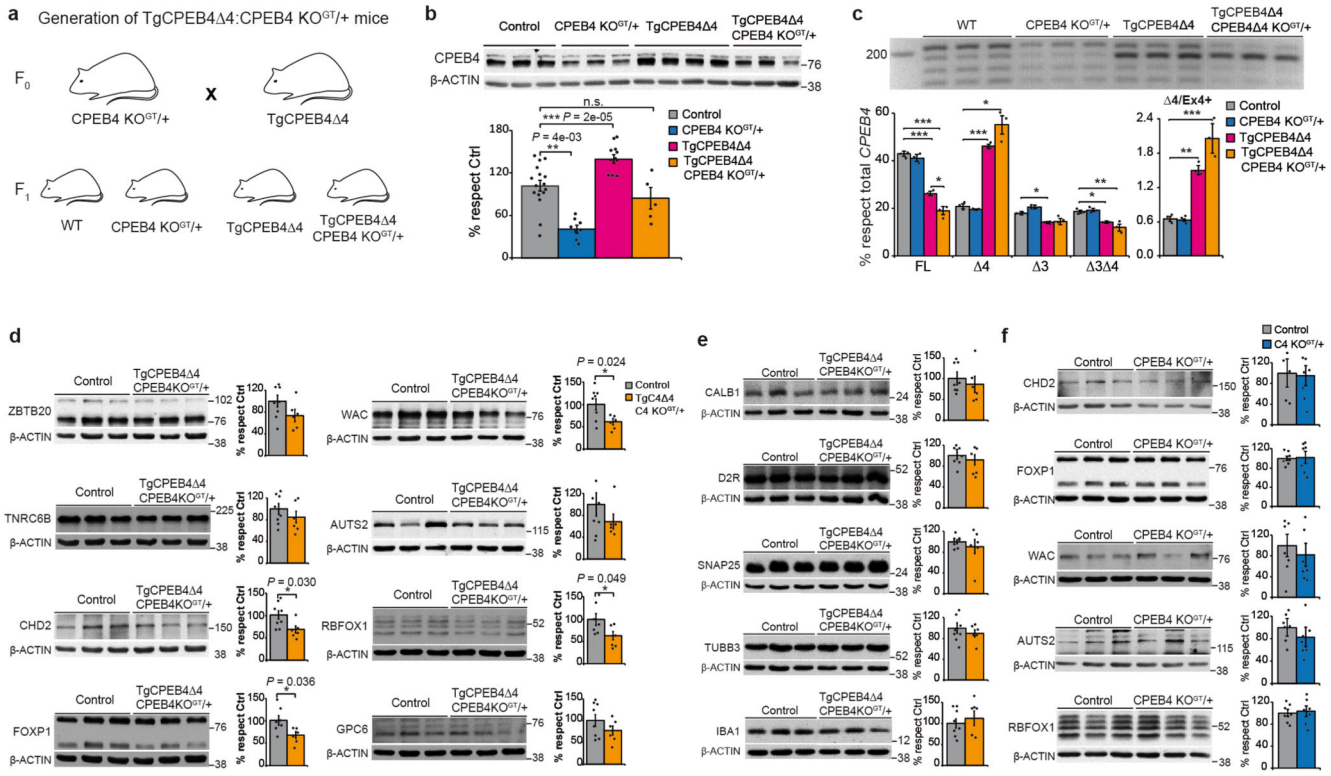


Extended Data Fig. 6. Supplemental data of global poly(A)-alteration and protein levels in TgCPEB4Δ4 mice.

a, Comparison of genes with poly(A) changes in the same or the opposite direction between human ASD cases and TgCPEB4Δ4 mice, representation factor (RF). **b,** Fold change enrichment of brain, oligodendrocytic, astrocytic, neuronal, synaptic and ASD specific (SFARI cat.1-2) genes with shortened poly(A)-tail respect to total genome in TgCPEB4Δ4 mice. **c,** Fold change enrichment of ASD (SFARI cat. 1-2, n = 62) genes shortened in TgCPEB4Δ4 mice and lengthened in CPEB4 KO^{GT}/+ and CPEB4 KO mice vs. total genome stratified by 5'UTR, 3'UTR, CDS, gDNA length and ratio neuronal/glial

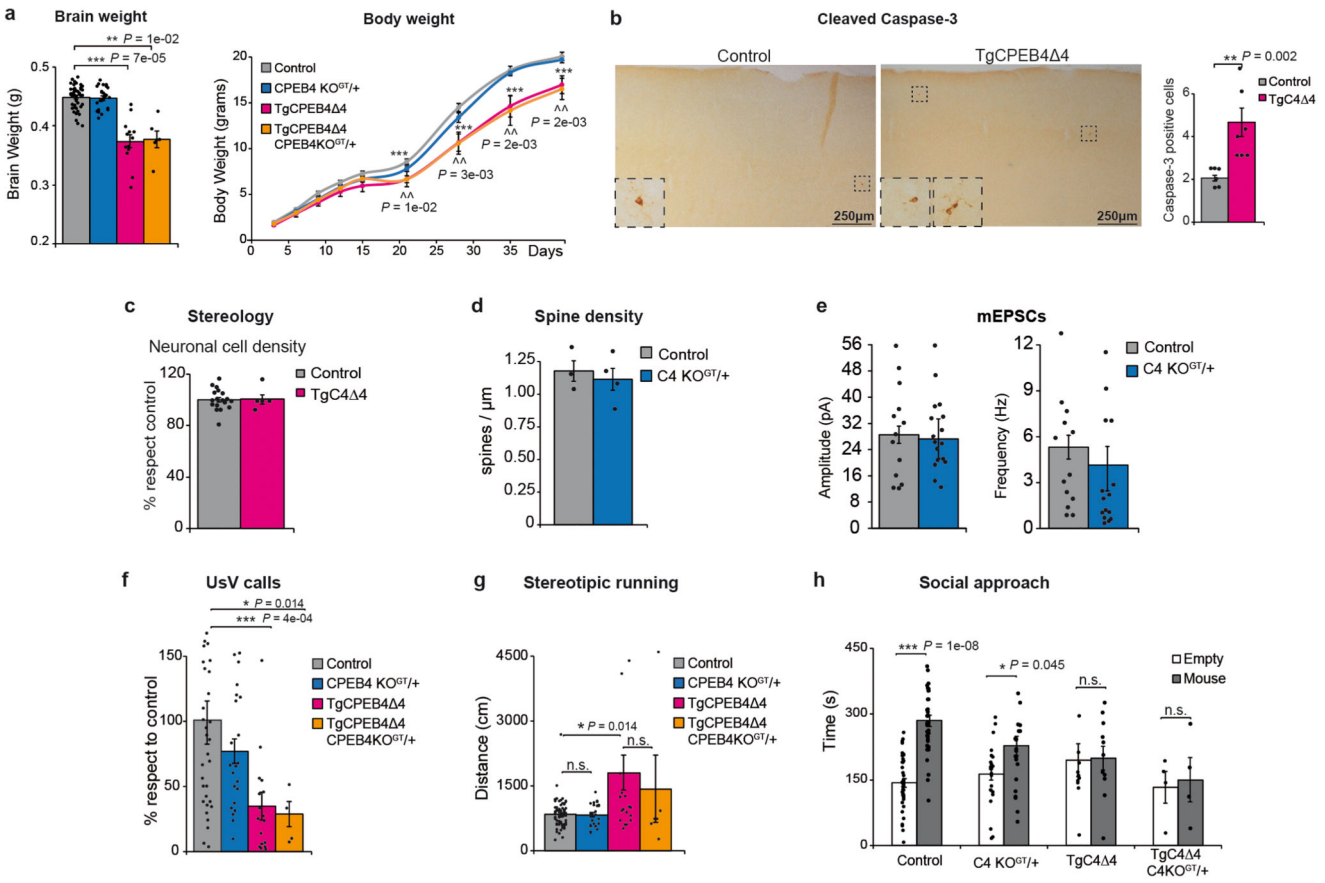
Europe PMC Funders Author Manuscripts

expression. **d,** Protein levels in St of 1.5-month-old control and TgCPEB4Δ4 mice (n = 7). **e,** Hire-PAT assay of *Auts2* poly(A)-tail in control and TgCPEB4Δ4 mice (n = 3). **f,** Protein levels of neuronal and astrocytic specific genes in Cx of control and TgCPEB4Δ4 mice (n = 7). For gel source data, see Supplementary Figure 1. **a,** Hypergeometric test. **b,** One-sided Fisher's exact test, *P*-values of genes with shortened poly(A) in each group respect to ASD genes. **c,** Statistical details in simulations in method section. **d-f,** Two-sided unpaired t-test. Data are mean ± s.e.m. 95% CIs. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.



Extended Data Fig. 7. TgCPEB4Δ4:CPEB4 KO^{GT}/+ mice but not CPEB4 KO^{GT}/+ mice show ASD gene protein changes.

a, Breeding protocol to obtain TgCPEB4Δ4:CPEB4 KO^{GT}/+ mice. **b,** CPEB4 protein levels in Cx of Control 1.5-month-old (n = 16), CPEB4 KO^{GT}/+ (n = 8) TgCPEB4Δ4 (n = 11) and TgCPEB4Δ4:CPEB4 KO^{GT}/+ mice (n = 5). **c,** Percentage of *CPEB4* splicing isoforms and Δ4/Ex4+ ratio in Cx of Control, CPEB4 KO^{GT}/+, TgCPEB4Δ4 and TgCPEB4Δ4:CPEB4 KO^{GT}/+ mice (n = 3) by PCR with primers annealing to exons 2 and 5. **d, f,** Protein levels of ASD genes in **d,** Control (n = 8) and TgCPEB4Δ4:CPEB4 KO^{GT}/+ mice (n = 6) and **f,** Control and CPEB4 KO^{GT}/+ mice (n = 7). **e,** Protein levels of neuronal and astrocytic specific genes in Cx of Control (n = 8) and TgCPEB4Δ4:CPEB4 KO^{GT}/+ mice (n = 6). For gel source data, see Supplementary Figure 1. **b,** One-way ANOVA followed by Games-Howell post hoc test. **c,** One-way ANOVA followed by Tukey's post hoc test. **d-f,** Two-sided unpaired t-test. Data are mean ± s.e.m. 95% CIs. **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

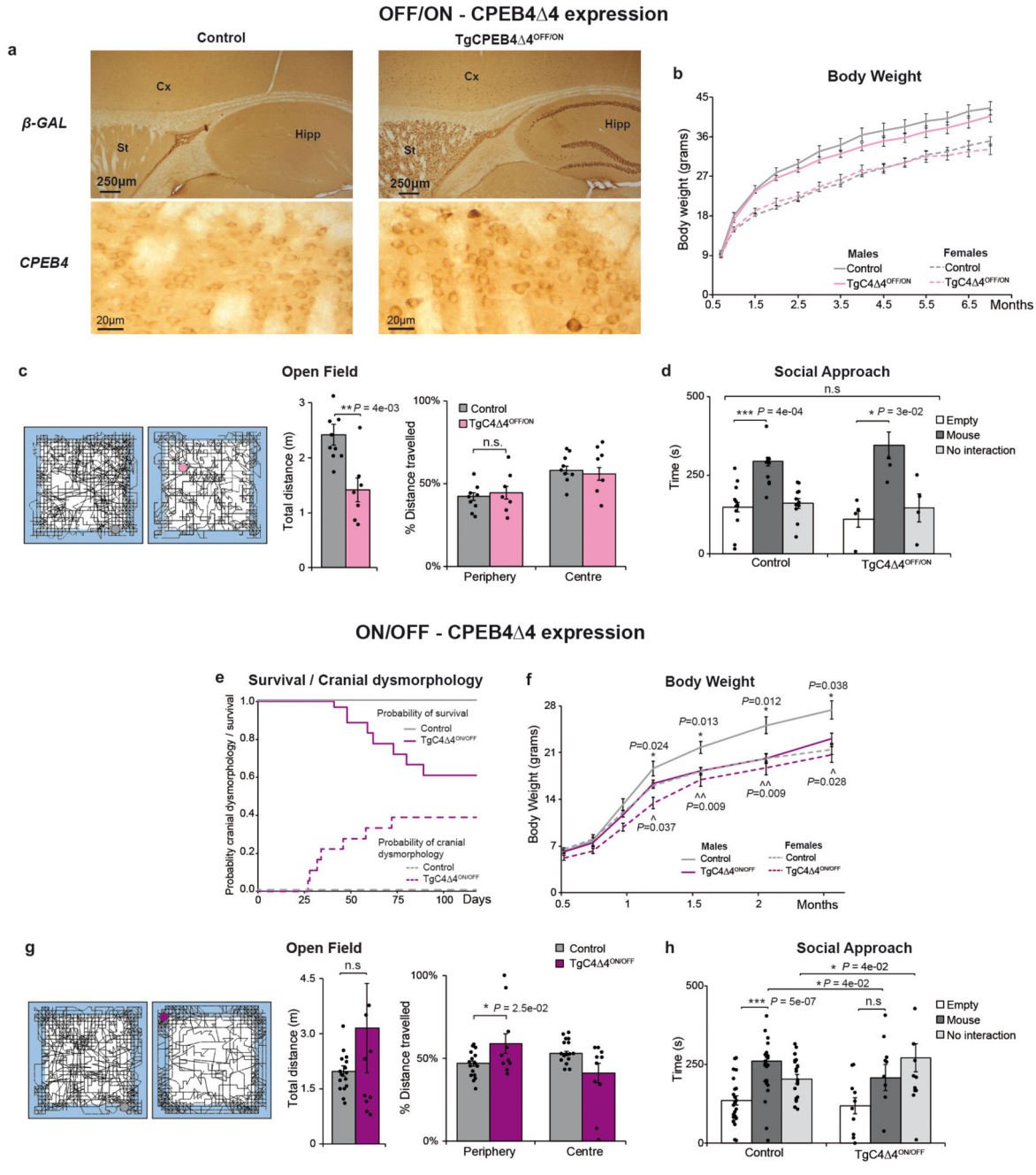


Extended Data Fig. 8. TgCPEB4Δ4:CPEB4 KO^{GT}/+ mice, but not CPEB4KO^{GT}/+ mice, show anatomical and behavioral alteration.

a, Brain weight in 6-week-old control (n = 45), CPEB4 KO^{GT}/+ (n = 25), TgCPEB4Δ4 (n = 13) and TgCPEB4Δ4:CPEB4 KO^{GT}/+ (n = 6) mice and evolution of body weight of control (n = 74), CPEB4 KO^{GT}/+ (n = 27), TgCPEB4Δ4 (n = 18) and TgCPEB4Δ4:CPEB4 KO^{GT}/+ (n = 6) mice. **b,** Immunohistochemistry against anti-cleaved caspase-3 in Cx (n = 3 slices from six controls and six TgCPEB4Δ4 mice). Scale bars represent 250 µm. **c,** Striatal neuronal cell density in Control (n = 19) and TgCPEB4Δ4 mice (n = 5). **d,** Spine density (spines/µm) in cortical layers II/III of pyramidal neurons in CPEB4 KO^{GT}/+ mice (n = 5 cells from three controls, and n = 5 cells from four CPEB4 KO^{GT}/+ mice). **e,** Amplitude (pA) and frequency (Hz) of mEPSCs recorded from pyramidal neurons of the somatosensory Cx, in CPEB4 KO^{GT}/+ mice (n = 13 cells from five controls, and n = 17 cells from six CPEB4 KO^{GT}/+ mice). **f,** Ultrasonic calls of pups during 5 min after separation from their mothers as mean of data from postnatal days 6 and 12 in control (n = 36), CPEB4 KO^{GT}/+(n = 22), TgCPEB4Δ4 (n = 17) and TgCPEB4Δ4:CPEB4 KO^{GT}/+ (n = 4) pups. **g,** Stereotypical running represented as distance travelled (cm) in the periphery in the OF-test in control (n = 74), CPEB4 KO^{GT}/+ (n = 25), TgCPEB4Δ4 (n = 19) and TgCPEB4Δ4:CPEB4 KO^{GT}/+ (n = 6) mice. **h,** Time interacting with empty cage or an unfamiliar mouse during 10 min. Control (n = 40), CPEB4 KO^{GT}/+ (n = 24), TgCPEB4Δ4 (n = 11) and TgCPEB4Δ4:CPEB4 KO^{GT}/+ (n = 4) mice. **a,** One-way ANOVA followed by

Europe PMC Funders Author Manuscripts

Games-Howell post hoc test. **b,** Two-sided Mann-Whitney-Wilcoxon test. **c-e,** Two-sided unpaired t-test. **f, g,** Kruskal-Wallis one-way ANOVA test. **h,** Two-sided Wilcoxon signed-rank test. Data are mean ± s.e.m. 95% CIs. n.s non-significative, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.



Extended Data Fig. 9. Effect on ASD-like behaviors of doxycycline-mediated temporal regulation of transgene expression in TgCPEB4Δ4 mice.

a-d, TgCPEB4Δ4 mice with transgene expression starting at the age of 3 weeks (OFF/ON-TgCPEB4Δ4 mice) do not display ASD-like behavioral phenotypes. **a,** β-GAL nuclear

Europe PMC Funders Author Manuscripts

staining in forebrain neurons and CPEB4 immunohistochemistry in 3 month-old control and TgCPEB4Δ4 mice (n = 3). Cx, cortex; St, striatum; Hipp, hippocampus. **b,** Evolution of body weight (grams) of males (n = 29 controls, n = 11 OFF/ON-TgCPEB4Δ4) and females (n = 29 control, n = 10 OFF/ON-TgCPEB4Δ4). No premature death nor cranial dysmorphology was observed in OFF/ON-TgCPEB4Δ4 mice. **c,** Total distance travelled by control (n = 9) and OFF/ON-TgCPEB4Δ4 (n = 7) mice and percentage of their distance in the periphery and in the center in OF test. **d,** Time interacting with either an empty cage, an unfamiliar mouse or without any interaction during 10 min. Control (n = 12) and OFF/ON-TgCPEB4Δ4 mice (n = 5). **e-h,** Silencing transgene expression in TgCPEB4Δ4 mice which have expressed the transgene during embryonic development does not revert ASD-like behaviors (ON/OFF-TgCPEB4Δ4 mice). **e,** Kaplan-Meier curve for cumulative survival (solid line) and percentage of mice developing cranial dysmorphology (dashed line), n = 21 for controls, n = 16 for ON/OFF-TgCPEB4Δ4. **f,** Evolution of body weight (grams): males (n = 19 controls and n = 10 ON/OFF-TgCPEB4Δ4), females (n = 12 control and n = 6 ON/OFF-TgCPEB4Δ4). **g,** Total distance travelled by control (n = 16) and ON/OFF-TgCPEB4Δ4 (n = 10) mice and percentage of their distance in the periphery and in the center in OF test. **h,** Time interacting with either an empty cage, an unfamiliar mouse or without any interaction during 10 min. Control (n = 20) and ON/OFF-TgCPEB4Δ4 mice (n = 13). **b-d, f,** Two-sided unpaired t-test. **g-h,** Two-sided Mann-Whitney-Wilcoxon test. **h,** Two-sided Wilcoxon signed-rank test. Data are mean ± s.e.m. 95% CIs. n.s non-significative, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by grants: ISCIII-CiberNed-PI2013/09-&PI2015-2/06 (J.J.L., R.F.-C), -FEDER-PI14/00125&PI17/00199 (P.N.); MINECO-SAF2012-34177&SAF2015-65371-R (J.J.L.) -FEDER-BFU2014-54122-P (R.M.) -BFU2014-55076-P (M.I.) -BFU2016-76050-P (R.F.-C.), -SEV-2012-0208 to CRG by European Union FEDER (M.I.); NIMH 5R37 MH060233, 5R01 MH09714 and 5R01 MH100027 (D.H.G.); Junta de Andalucía-P12-CTS-2232&CTS-600 (R.F.-C.); Generalitat de Catalunya-2014/SGR/143 (P.N.); ERC-StG-LS2-637591 (M.I.); and by Fundación Botín-Banco Santander/Santander Universities Global Division, Fundación BBVA, and Fundación Ramón Areces; A.P. was recipient of a MICINN FPI-fellowship; N.N.P. (NRSA F30 MH099886, UCLA Medical Scientist Training Program) and V.S. (Larry Hillblom Postdoctoral Fellowship).

We thank the computing facilities of Extremadura Research Centre for Advanced Technologies (CETA-CIEMAT/Government of Spain), which is funded by ERDF. Tissue, biological specimens or data used in this research were obtained from the Autism BrainNet (formerly the Autism Tissue Program), which is sponsored by the Simons Foundation, and the University of Maryland Brain and Tissue Bank (a component of the NIH NeuroBioBank). We are grateful to the patients and families who participate in the tissue donation programs. We thank Dr. R. García-Escudero (CIEMAT, Madrid) for bioinformatics advice, Miriam Lucas for technical assistance, and the following core facilities: CBMSO-Genomics & Massive Sequencing, CBMSO-Animal Facility, CBMSO-Confocal Microscopy, CBMSO-CNB Mouse Transgenesis, CNB-Proteomics, IRB-Functional Genomic and IRB-Bioinformatics/Biostatistics.

References

1. Hallmayer J, et al. Genetic heritability and shared environmental factors among twin pairs with autism. Archives of general psychiatry. 2011; 68:1095–1102. [PubMed: 21727249]
2. Sandin S, et al. The familial risk of autism. Jama. 2014; 311:1770–1777. [PubMed: 24794370]

3. De Rubeis S, Buxbaum JD. Genetics and genomics of autism spectrum disorder: Embracing complexity. Human molecular genetics. 2015
4. Kim YS, Leventhal BL. Genetic epidemiology and insights into interactive genetic and environmental effects in autism spectrum disorders. Biological psychiatry. 2015; 77:66–74. [PubMed: 25483344]
5. Sztainberg Y, Zoghbi HY. Lessons learned from studying syndromic autism spectrum disorders. Nature neuroscience. 2016; 19:1408–1417. [PubMed: 27786181]
6. Gaugler T, et al. Most genetic risk for autism resides with common variation. Nature genetics. 2014; 46:881–885. [PubMed: 25038753]
7. Geschwind DH, State MW. Gene hunting in autism spectrum disorder: on the path to precision medicine. The Lancet. 2015; 14:1109–1120. [PubMed: 25891009]
8. Willsey AJ, State MW. Autism spectrum disorders: from genes to neurobiology. Current opinion in neurobiology. 2015; 30:92–99. [PubMed: 25464374]
9. Ivshina M, Lasko P, Richter JD. Cytoplasmic polyadenylation element binding proteins in development, health, and disease. Annual review of cell and developmental biology. 2014; 30:393–415.
10. Sarkissian M, Mendez R, Richter JD. Progesterone and insulin stimulation of CPEB-dependent polyadenylation is regulated by Aurora A and glycogen synthase kinase-3. Genes & development. 2004; 18:48–61. [PubMed: 14724178]
11. Si K, et al. A neuronal isoform of CPEB regulates local protein synthesis and stabilizes synapse-specific long-term facilitation in aplysia. Cell. 2003; 115:893–904. [PubMed: 14697206]
12. Fioriti L, et al. The Persistence of Hippocampal-Based Memory Requires Protein Synthesis Mediated by the Prion-like Protein CPEB3. Neuron. 2015; 86:1433–1448. [PubMed: 26074003]
13. Udagawa T, et al. Genetic and acute CPEB1 depletion ameliorate fragile X pathophysiology. Nature medicine. 2013; 19:1473–1477.
14. Sultana R, et al. Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. Genomics. 2002; 80:129–134. [PubMed: 12160723]
15. Voineagu I, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature. 2011; 474:380–384. [PubMed: 21614001]
16. Parikshak NN, et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature. 2016; 540:423–427. [PubMed: 27919067]
17. Iossifov I, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014; 515:216–221. [PubMed: 25363768]
18. De Rubeis S, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014; 515:209–215. [PubMed: 25363760]
19. Takata A, et al. Integrative Analyses of De Novo Mutations Provide Deeper Biological Insights into Autism Spectrum Disorder. Cell reports. 2018; 22:734–747. [PubMed: 29346770]
20. Calderone V, et al. Sequential Functions of CPEB1 and CPEB4 Regulate Pathologic Expression of Vascular Endothelial Growth Factor and Angiogenesis in Chronic Liver Disease. Gastroenterology. 2016; 150:982–997 e930. [PubMed: 26627607]
21. Igea A, Mendez R. Meiosis requires a translational positive loop where CPEB1 ensues its replacement by CPEB4. The EMBO journal. 2010; 29:2182–2193. [PubMed: 20531391]
22. Xiong HY, et al. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science (New York, N.Y. 2015; 347:1254806.
23. Irimia M, et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell. 2014; 159:1511–1523. [PubMed: 25525873]
24. Tapial J, et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. Genome research. 2017; 27:1759–1768. [PubMed: 28855263]
25. Theis M, Si K, Kandel ER. Two previously undescribed members of the mouse CPEB family of genes and their inducible expression in the principal cell layers of the hippocampus. Proceedings

of the National Academy of Sciences of the United States of America. 2003; 100:9602–9607. [PubMed: 12871996]

26. Yamasue H, Domes G. Oxytocin and Autism Spectrum Disorders. Curr Top Behav Neurosci. 2017
27. Hu W, Yuan B, Lodish HF. Cpeb4-mediated translational regulatory circuitry controls terminal erythroid differentiation. Developmental cell. 2014; 30:660–672. [PubMed: 25220394]
28. Li YI, Sanchez-Pulido L, Haerty W, Ponting CP. RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. Genome research. 2015; 25:1–13.
29. Pedrotti S, et al. The RNA-binding protein Rbfox1 regulates splicing required for skeletal muscle structure and function. Human molecular genetics. 2015; 24:2360–2374. [PubMed: 25575511]
30. Chen JA, Penagarikano O, Belgard TG, Swarup V, Geschwind DH. The emerging picture of autism spectrum disorder: genetics and pathology. Annual review of pathology. 2015; 10:111–144.
31. Rabaneda LG, Robles-Lanuza E, Nieto-Gonzalez JL, Scholl FG. Neurexin dysfunction in adult neurons results in autistic-like behavior in mice. Cell reports. 2014; 8:338–346. [PubMed: 25017069]
32. Kalkbrenner AE, Schmidt RJ, Penlesky AC. Environmental chemical exposures and autism spectrum disorders: a review of the epidemiological evidence. Current problems in pediatric and adolescent health care. 2014; 44:277–318. [PubMed: 25199954]
33. Maeyama K, et al. Congenital Cytomegalovirus Infection in Children with Autism Spectrum Disorder: Systematic Review and Meta-Analysis. J Autism Dev Disord. 2018; 48:1483–1491. [PubMed: 29185167]
34. Batra R, et al. RNA-binding protein CPEB1 remodels host and viral RNA landscapes. Nature structural & molecular biology. 2016
35. Mangiarini L, et al. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. Cell. 1996; 87:493–506. [PubMed: 8898202]
36. Shin J, Salameh JS, Richter JD. Impaired neurodevelopment by the low complexity domain of CPEB4 reveals a convergent pathway with neurodegeneration. Scientific reports. 2015; 6 29395.
37. Mayford M, et al. Control of memory formation through regulated expression of a CaMKII transgene. Science (New York, N.Y. 1996; 274:1678–1683.
38. Irizarry RA, et al. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003; 31:e15. [PubMed: 12582260]
39. Yates A, et al. Ensembl 2016. Nucleic Acids Res. 2016; 44:D710–716. [PubMed: 26687719]
40. Pique M, Lopez JM, Foissac S, Guigo R, Mendez R. A combinatorial code for CPE-mediated translational control. Cell. 2008; 132:434–448. [PubMed: 18267074]
41. Cahoy JD, et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J Neurosci. 2008; 28:264–278. [PubMed: 18171944]
42. Cajigas IJ, et al. The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. Neuron. 2012; 74:453–466. [PubMed: 22578497]
43. De Rubeis S, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014; 515:209–215. [PubMed: 25363760]
44. Zhang Y, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci. 2014; 34:11929–11947. [PubMed: 25186741]
45. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36. [PubMed: 23618408]
46. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31:166–169. [PubMed: 25260700]
47. Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics. 2012; 13:204–216. [PubMed: 22285995]
48. Shen S, et al. MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. Nucleic Acids Res. 2012; 40:e61. [PubMed: 22266656]
49. Belloc E, Mendez R. A deadenylation negative feedback mechanism governs meiotic metaphase arrest. Nature. 2008; 452:1017–1021. [PubMed: 18385675]

50. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; 2014.
51. Gentleman RC, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004; 5:R80. [PubMed: 15461798]
52. Irizarry RA, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4:249–264. [PubMed: 12925520]
53. Gentleman RC, Carey VJ, Huber W, Irizarry R, Dudoit S. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer-Verlag; New York: 2005.
54. Eklund AC, Szallasi Z. Correction of technical bias in clinical microarray data improves concordance with known biological information. Genome Biol. 2008; 9:R26. [PubMed: 18248669]
55. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47. [PubMed: 25605792]
56. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. [PubMed: 19131956]
57. Lucas JJ, et al. Decreased nuclear beta-catenin, tau hyperphosphorylation and neurodegeneration in GSK-3beta conditional transgenic mice. The EMBO journal. 2001; 20:27–39. [PubMed: 11226152]
58. Risher WC, Ustunkaya T, Singh Alvarado J, Eroglu C. Rapid Golgi analysis method for efficient and unbiased classification of dendritic spines. PLoS One. 2014; 9:e107591. [PubMed: 25208214]
59. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012; 9:671–675. [PubMed: 22930834]
60. Peca J, et al. Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. Nature. 2011; 472:437–442. [PubMed: 21423165]
61. Ting JT, Daigle TL, Chen Q, Feng G. Acute brain slice methods for adult and aging animals: application of targeted patch clamp analysis and optogenetics. Methods Mol Biol. 2014; 1183:221–242. [PubMed: 25023312]
62. Zhao S, et al. Cell type-specific channelrhodopsin-2 transgenic mice for optogenetic dissection of neural circuitry function. Nat Methods. 2011; 8:745–752. [PubMed: 21985008]
63. Guzman SJ, Schlogl A, Schmidt-Hieber C. Stimfit: quantifying electrophysiological data with Python. Front Neuroinform. 2014; 8:16. [PubMed: 24600389]

Europe PMC Funders Author Manuscripts

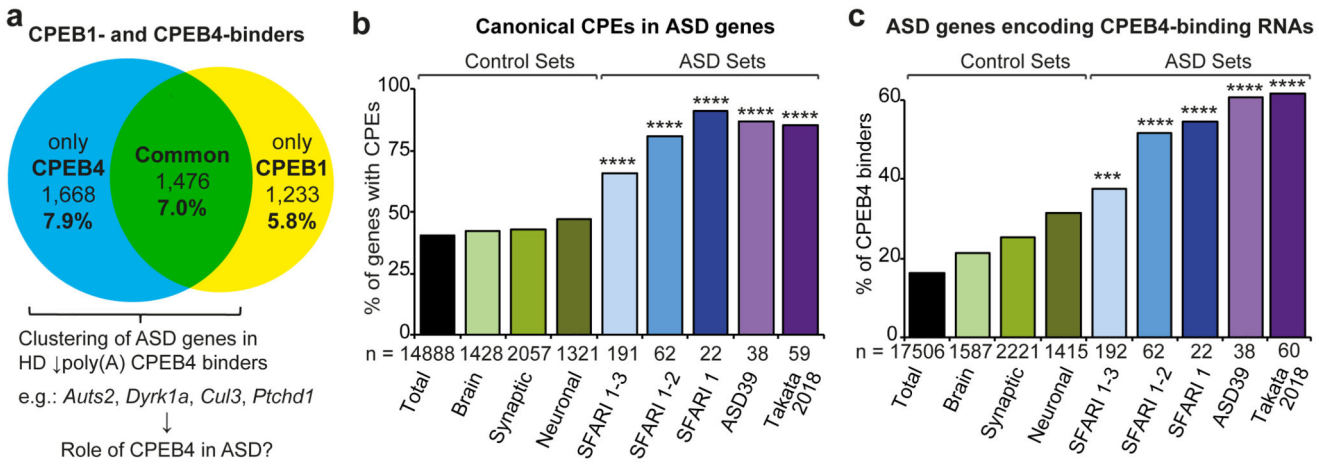


Fig. 1. ASD-risk gene mRNAs bear CPEs and bind CPEB4.

a, Percentage of CPEB1- and/or CPEB4-binder transcripts in mouse St. **b-c,** Percentage of transcripts, with **b,** canonical CPEs and **c,** bound by CPEB4 in control gene sets and high-confidence ASD-risk genes (SFARI cat. 1-3, ASD39 and Takata lists). **b-c,** One-sided Fisher's exact test, ***significant (P<0.05) respect total, brain and synaptic transcriptomes, ****significant (P<0.05) respect each control set.

Europe PMC Funders Author Manuscripts

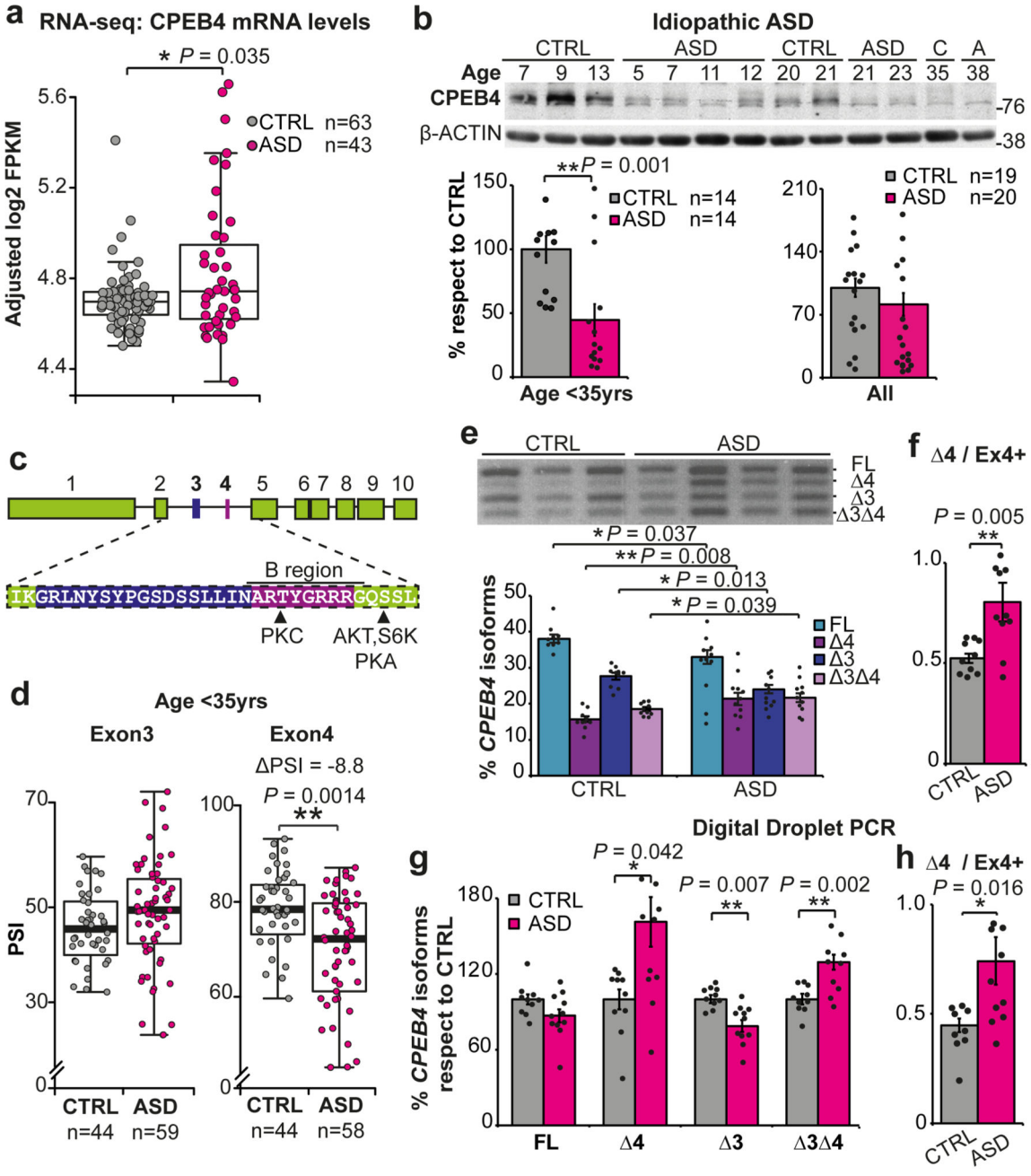


Fig. 2. CPEB4 alteration in idiopathic ASD brains.

CPEB4 **a,** mRNA **b,** protein levels in Cx. **c,** Alternatively spliced exons (3 and 4) of *CPEB4*, Δputative phosphorylation sites. **d,** Percent Spliced in (PSI). **e-h,** RT-PCR in CTRL (n=10) and idiopathic ASD cases (n=11) under 35-year-old, **e-f,** with external primers **e,** *CPEB4* isoforms percentage, **f,** exon4–excluding (Δ4)/exon4-including (Ex4+) isoform ratio. **g-h,** Digital-droplet PCR **g,** *CPEB4* isoform percentage normalized respect CTRL, **h,** Δ4/Ex4+ ratio. For gel source data, see Supplementary Figure 1. **a, d, f, h,** Two-sided Mann-Whitney-

Europe PMC Funders Author Manuscripts

Wilcoxon test. **b, e, g,** Two-sided unpaired t-test. Box plots show median, 25th, 75th percentiles. Data are mean ± s.e.m. 95% confidence intervals (CIs).

Europe PMC Funders Author Manuscripts

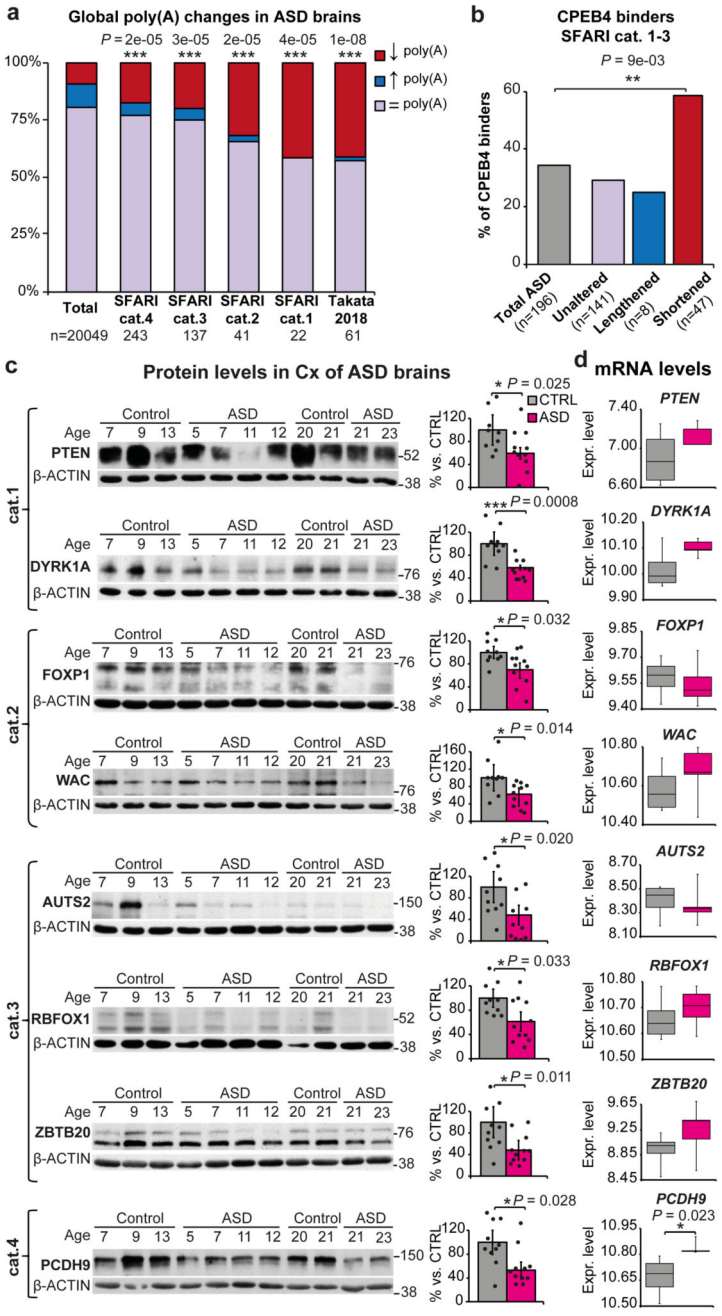


Fig. 3. ASD-risk gene mRNA deadenylation and decreased protein levels.

a, Poly(A)-tail length changes in Cx of ASD cases (n=6) vs. CTRL (n=5) in whole transcriptome and in ASD gene-lists. **b,** CPEB4 binders in ASD genes according to their poly(A)-tail change. **c,** Protein levels in Cx of idiopathic ASD (n=11) and CTRL (n=10) under 35-year-old. **d,** mRNA levels, ASD (n=6), CTRL (n=5). **a,** One-sided Fisher's exact test, *P*-values of ASD deadenylated transcripts vs. Total. **b,** One-sided Fisher's exact test. **c, d,** Two-sided unpaired t-test. Box plots show median, 25th, 75th percentiles. Data are mean ± s.e.m. 95% CIs.

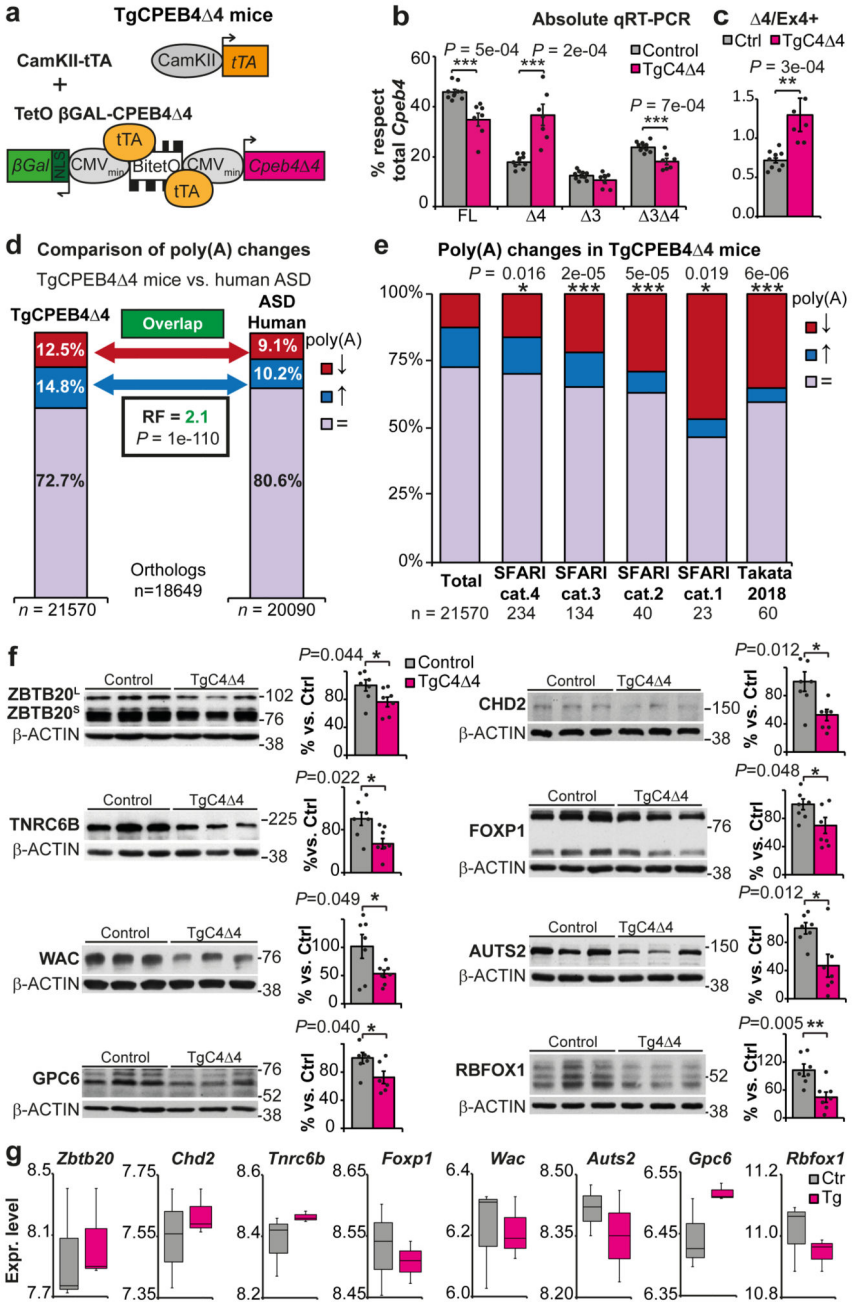


Fig. 4. ASD-like poly(A) changes in TgCPEB4Δ4 mice.

a, Transgenesis construct design. *CPEB4* **b,** splicing isoform percentage and **c,** Δ4/Ex4+ ratio in St of 1.5-month-old control (n=9), TgCPEB4Δ4 (n=7). **d,** Comparison of poly(A) changes in ASD cases vs. TgCPEB4Δ4 mice, representation factor (RF). **e,** Transcripts with poly(A)-tail changes in Cx-St of controls vs. TgCPEB4Δ4 (n=3) in whole transcriptome and ASD gene-lists. **f,** Protein levels in Cx of 1.5-month-old control and TgCPEB4Δ4 (n=7), **g,** mRNA levels (n=3). **b, f-g,** Two-sided unpaired t-test. **c, f,** Two-sided Mann-Whitney-Wilcoxon test. **d,** Hypergeometric test. **e,** One-sided Fisher's exact test, *P*-values of ASD

Europe PMC Funders Author Manuscripts

deadenylated transcripts vs. Total. Box plots show median, 25th, 75th percentiles. Data are mean ± s.e.m. 95% CIs.

Europe PMC Funders Author Manuscripts

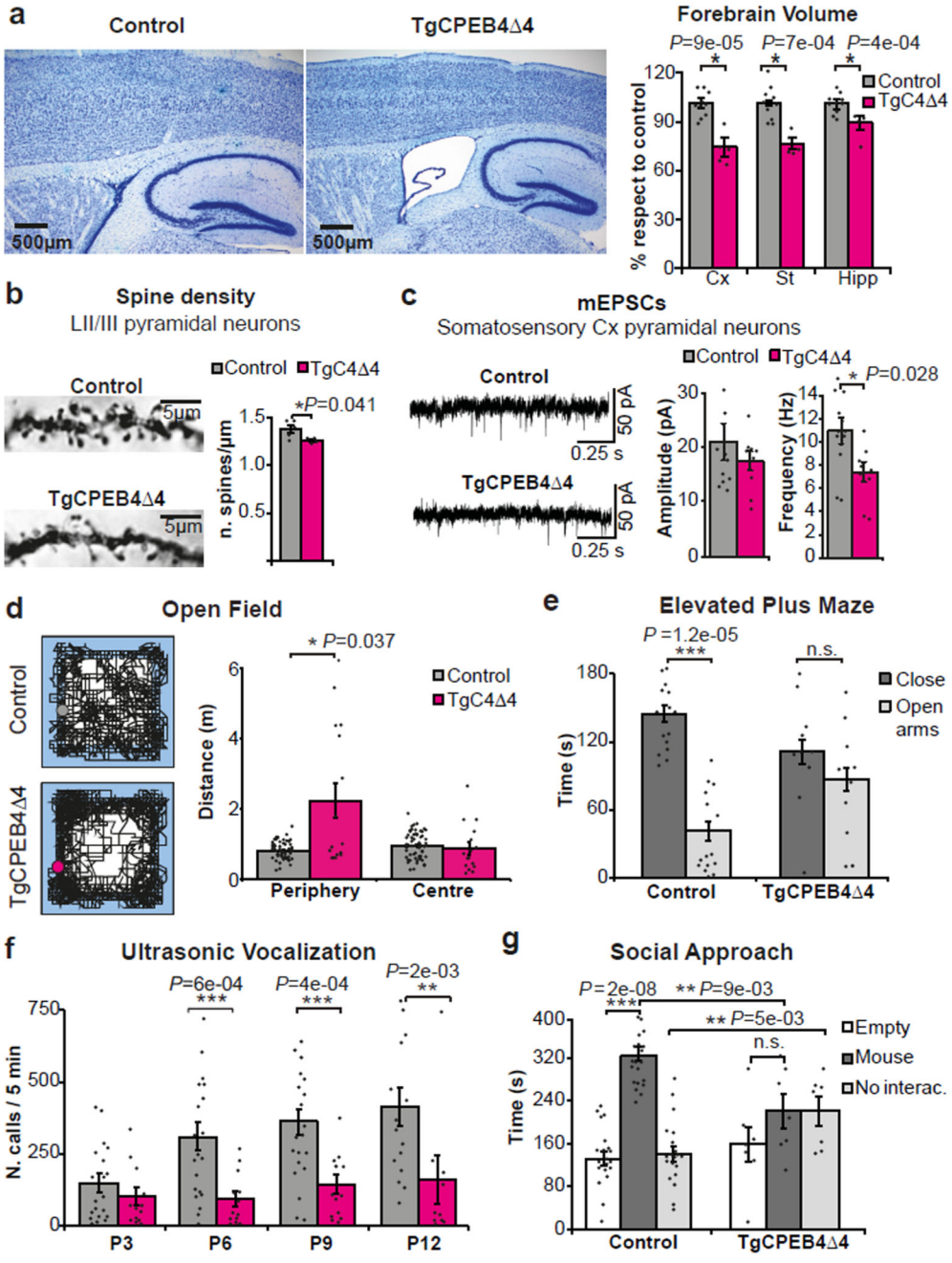


Fig. 5. ASD-like phenotypes in TgCPEB4Δ4 mice.

a, Forebrain Volume, control (n=10), TgCPEB4Δ4 (n=5). **b,** Spine density (n=14 cells from five controls, n=12 cells from four TgCPEB4Δ4). **c,** mEPSCs (n=11 cells from five controls, n=9 cells from five TgCPEB4Δ4). **d,** Distance travelled, control (n=60), TgCPEB4Δ4 (n=16). **e,** Time spent in closed/open arms, control (n=15),TgCPEB4Δ4 (n=10). **f,** Ultrasonic calls, control (n=20), TgCPEB4Δ4 (n=13). **g,** Time interacting with empty cage and unfamiliar mouse, control (n=20), TgCPEB4Δ4 (n=7). **a-c, f,** Two-sided unpaired t-test. **d, f-**

g, Two-sided Mann-Whitney-Wilcoxon test. **e,** Two-sided paired t-test. **g,** Two-sided Wilcoxon signed-rank test. Data are mean ± s.e.m. 95% CIs, n.s non-significative.

Europe PMC Funders Author Manuscripts