RESEARCH COMMUNICATION

Identification and characterization of conserved noncoding *cis*-regulatory elements that impact *Mecp2* expression and neurological functions

Yingyao Shao,[1,2,3] Sameer S. Bajikar,[1,3] Harini P. Tirumala,[1,3] Manuel Cantu Gutierrez,[2,4,5] Joshua D. Wythe,[2,4,5] and Huda Y. Zoghbi[1,2,3,6,7]

[1]Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Baylor College of Medicine, Houston, Texas 77030, USA; [2]Program in Developmental Biology, Baylor College of Medicine, Houston, Texas 77030, USA; [3]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA; [4]Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas 77030, USA; [5]Cardiovascular Research Institute, Baylor College of Medicine, Houston, Texas 77030, USA; [6]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA; [7]Department of Pediatrics, Baylor College of Medicine, Houston, Texas 77030, USA

While changes in MeCP2 dosage cause Rett syndrome (RTT) and *MECP2* duplication syndrome (MDS), its transcriptional regulation is poorly understood. Here, we identified six putative noncoding regulatory elements of *Mecp2*, two of which are conserved in humans. Upon deletion in mice and human iPSC-derived neurons, these elements altered RNA and protein levels in opposite directions and resulted in a subset of RTT- and MDS-like behavioral deficits in mice. Our discovery provides insight into transcriptional regulation of *Mecp2/MECP2* and highlights genomic sites that could serve as diagnostic and therapeutic targets in RTT or MDS.

Supplemental material is available for this article.

Received October 4, 2020; revised version accepted February 24, 2021.

Coding variants in hundreds of genes are known to alter protein levels and lead to intellectual disability (ID) (Schanze et al. 2018). However, only a handful of disease-causing mutations in noncoding *cis*-regulatory elements (CREs) (Soldner et al. 2016; Oz-Levi et al. 2019) have been identified, and our understanding of how these contribute to ID is limited. *MECP2* is an exemplar ID-causing, dosage-sensitive gene, with neurological dysfunction arising from both decreased (RTT) (Amir et al. 1999) and increased (MDS) (Van Esch et al. 2005) levels of MeCP2. This dosage sensitivity underscores the concept that precise control of MeCP2 levels is important for normal brain function (Chao and Zoghbi 2012; Sztainberg et al. 2015). Regulation of *MECP2* occurs at the transcriptional, post-transcriptional (Gennarino et al. 2015; Rodrigues et al. 2016), and post-translational levels (Lombardi et al. 2017; Yagasaki et al. 2018), with the latter two being the most well studied. Currently, little is known about the transcriptional regulation of *MECP2* beyond its core promoter region (Liu and Francke 2006; Nagarajan et al. 2006; Swanberg et al. 2009).

Results and Discussion

To identify potential CREs of *Mecp2* in the brain, we profiled open chromatin in the developing and adult mouse brain using the assay for transposase-accessible chromatin with deep sequencing (ATAC-seq) (Buenrostro et al. 2013). We restricted our search to accessible regulatory elements within the 100-kb genomic region of *Mecp2*, flanked upstream by the *Opn1mw* gene and downstream by the *Irak1* gene. This genomic region, containing only *Mecp2*, faithfully drives expression of human *MECP2* in a transgenic mouse line, rescuing all neurological defects in *Mecp2*-null mice (Collins et al. 2004). We identified six putative regulatory elements (Fig. 1A): one at the promoter (Peak-4), three in intron 2 that increase in accessibility during development (Peak-1, Peak-2, and Peak-3), one upstream of *Mecp2* (Peak-5) that decreases in accessibility in the adult brain compared with the postnatal day 6 brain, and one peak upstream of *Mecp2* (Peak-6) that is consistently accessible throughout development. Importantly, these same peaks are present in previously published ATAC-seq data sets from multiple neuronal cell types in the adult mouse brain (Supplemental Fig. S1; Mo et al. 2015), confirming our findings.

To test whether these putative CREs regulate *Mecp2* expression in the mouse brain, we deleted the genomic regions corresponding to each individual peak using CRISPR-Cas9 genome editing (Supplemental Table S1), generating five unique mouse lines. We excluded Peak-4 from our analysis because it corresponds to the *Mecp2* promoter, whose deletion would likely ablate *Mecp2* expression. We measured *Mecp2* mRNA expression in the brains of our knockout (Peak KO) lines and found that *Mecp2* expression was decreased by ~20%–30% in Peak-$2^{KO/y}$, Peak-$3^{KO/y}$, and Peak-$5^{KO/y}$ mice, while it was increased by 50% in Peak-$6^{KO/y}$ mice and unchanged in Peak-$1^{KO/y}$ mice (Fig. 1B). Measurement of MeCP2 protein levels by Western blot showed 30% reduction ($P < 0.01$) in Peak-$2^{KO/y}$ mice and 70% increase ($P < 0.0001$) in Peak-$6^{KO/y}$ mice, while protein level changes in Peak-$1^{KO/y}$, Peak-$3^{KO/y}$, and Peak-$5^{KO/y}$ mice were not significant (Fig. 1C). Interestingly, Peak-2 and Peak-6 are the only two elements with strong sequence conservation at the nucleotide level between mice and humans, suggesting that

[*Keywords*: MeCP2; *cis*-regulatory elements; noncoding; neurological disorders]
Corresponding author: hzoghbi@bcm.edu
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.345397.120.

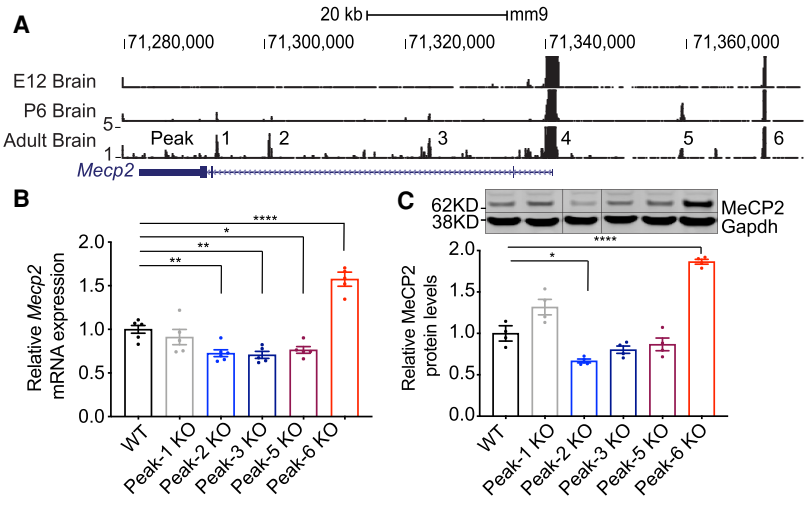


Figure 1. Identification of putative *cis*-regulatory elements of *Mecp2*. (*A*) ATAC-seq reads mapped to the *Mus musculus* genome (mm9) are shown from nuclei isolated from embryonic, postnatal, and adult mouse brains (specifically E12.5, P6, and 8 wk of age). Exons are denoted by solid blue rectangles, while introns are between exons. The 3′ UTR is represented by a thinner but longer blue block relative to the coding exons. Note that locus is transcribed in the opposite direction relative to the image shown (i.e., the promoter is at the *right*, and the 3′ UTR is at the *left*). Peaks 1, 2, 3, 5, and 6 were pursued for functional analysis of *cis*-regulatory elements, whereas Peak-4 (which spanned the proximal promoter) was not studied further. (*B*) RT-qPCR shows *Mecp2* mRNA expression in different knockout lines ($n = 5$). (*C*) Western blot analysis of mouse brain tissue shows MeCP2 protein expression levels in each different knockout line. Representative Western blot image of MeCP2 protein levels is shown with Gapdh as a loading control, and solid vertical lines indicate areas of the image that have been spliced to remove unneeded lanes ($n = 4$). All data were analyzed by one-way ANOVA followed by Dunnett post hoc test. Data are presented as mean ± SEM. (*) $P < 0.05$, (**) $P < 0.01$, (****) $P < 0.0001$.

evolutionary pressure maintained these two gene regulatory elements (Supplemental Fig. S2).

Given that the magnitude of change in MeCP2 levels in both the Peak-2$^{KO/y}$ and Peak-6$^{KO/y}$ mice is less than that of the *Mecp2*$^{flox/y}$ allele (Samaco et al. 2008) and *MECP2* duplication mouse models (Collins et al. 2004), respectively, we wanted to identify and characterize any neurological dysfunction that may arise from more subtle changes in MeCP2 levels. Therefore, we performed a battery of behavioral tests on Peak-2$^{KO/y}$ and Peak-6$^{KO/y}$ male mice, after confirming peak knockout did not result in any gross histological abnormalities as assessed by Cresyl violet staining (Supplemental Fig. S3).

At 10 wk, Peak-2$^{KO/y}$ mice, which have ~30% reduction in *Mecp2* expression, were hyperactive compared with their wild-type littermates (Fig. 2A). At 24 wk, these mice had anxiety-like phenotypes (Fig. 2B) and social deficits (Fig. 2C). At 40 wk, these mice showed social dominance deficits (Fig. 2D). These behavioral deficits are reminiscent of those observed in *Mecp2*$^{flox/y}$ mice and some RTT (Moretti et al. 2005) and autism mouse models (Spencer et al. 2005; Kazdoba et al. 2016). Unlike *Mecp2*$^{flox/y}$ mice that express 50% of the normal MeCP2 level (Samaco et al. 2008), we did not observe any sensorimotor gating deficits, motor abnormalities, or learning and memory defects in the Peak-2$^{KO/y}$ mice (Supplemental Fig. S4A–F). Last, we measured the expression of several genes that are dysregulated in *Mecp2*-null mice (Boxer et al. 2020). Strikingly, expression of these genes in the cortex of Peak-2$^{KO/y}$ mice were altered in the same direction as reported in *Mecp2*-null animals (Supplemental Fig. S4G,H; Chahrour et al. 2008). These data suggest that a mild reduction in MeCP2 levels mimics a subset of behavioral and molecular changes observed in *Mecp2*-null mice.

In contrast, Peak-6$^{KO/y}$ mice, which display an ~70% increase in MeCP2, show hypoactivity (Fig. 3A) and anxiety-like phenotypes (Fig. 3B,C) at 10 wk. At 24 wk, these mice showed hippocampal-dependent contextual learning deficits in the contextual fear assay (Fig. 3D). These phenotypes resemble the behavioral profile of *MECP2-Tg1* mice (which have a 100% increase of MeCP2 protein). However, unlike the *MECP2-Tg1* mice, Peak-6$^{KO/y}$ mice showed no deficits in sensorimotor gating, motor function, or social behavior (Supplemental Fig. S5A–E). When we measured the expression of several genes known to be dysregulated in the *MECP2-Tg1* mouse model (Chahrour et al. 2008; Samaco et al. 2012), we found these MeCP2 targets were altered in the same direction as reported in *MECP2-Tg1* mice in the cortex of Peak-6$^{KO/y}$ mice (Supplemental Fig. 5F,G). These data suggest that a moderate increase in MeCP2 protein level recapitulates a subset of MDS-like behavioral and molecular phenotypes.

We next subjected the genomic sequences of these peaks to transcription factor motif

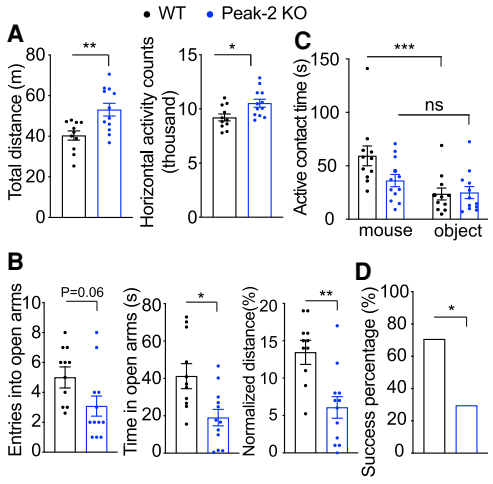


Figure 2. Peak-2$^{KO/y}$ mice show age-related behavioral deficits and abnormal RTT-like gene expression. (*A*) Peak-2$^{KO/y}$ mice traveled further and displayed more activity counts in the open field assay than their wild-type littermates in open field assay at 10 wk of age ($n = 12$). (*B*) Peak-2$^{KO/y}$ mice show anxiety-like phenotypes as measured by decreased entries, less time spent, and decreased distance in the open arms in the elevated plus maze at 24 wk of age ($n = 12$). (*C*) Peak-2$^{KO/y}$ mice show social deficits as measured by decreased time investigating a novel mouse in the three-chamber sociability assay at 24 wk of age ($n = 12$). (*D*) Peak-2$^{KO/y}$ mice show social dominance deficits as measured by decreased winning percentage in the tube test at 40 wk of age ($n = 12$). Data are presented as mean ± SEM. All data were analyzed by two-tailed *t*-test. (*) $P < 0.05$, (**) $P < 0.01$, (***) $P < 0.001$, (ns) not significant.

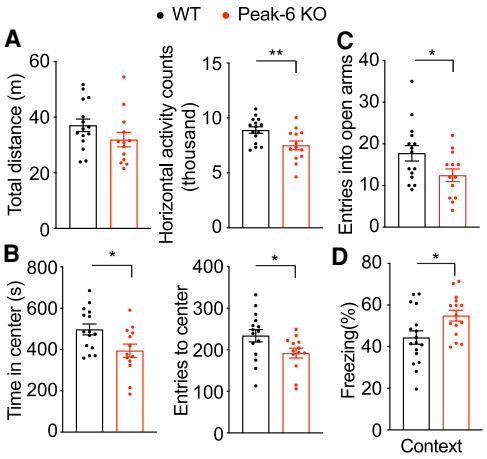


Figure 3. Peak-6$^{KO/y}$ mice show abnormal MDS-like gene expression and behavioral deficits. (*A*) Peak-6$^{KO/y}$ mice show hypoactivity, with fewer activity counts in open field assay at 9 wk of age (n = 13–15). (*B*) Peak-6$^{KO/y}$ mice show anxiety-like phenotypes as evidenced by decreased entries and time spent at the center in open field assay at 9 wk of age (n = 13–15). (*C*) Peak-6$^{KO/y}$ mice show anxiety-like phenotypes as measured by reduced entries to the open arms in the elevated plus maze at 10 wk of age (n = 13–15). (*D*) Peak-6$^{KO/y}$ mice show learning and memory deficits using a fear conditioning test at 36 wk of age (n = 13–15). Data are presented as mean ± SEM. All data were analyzed by two-tailed t-test. (*) $P < 0.05$, (**) $P < 0.01$.

analysis using HOMER to identify putative factors that bind within these peaks, and this analysis identified a potential CTCF binding site on Peak-6 (Supplemental Fig. S6A; Heinz et al. 2010). To validate this prediction, we conducted CTCF chromatin immunoprecipitation (ChIP) followed by qRT-PCR in the wild-type mouse frontal cortex and found CTCF enrichment on Peak-6 (Supplemental Fig S6B). Disruption of the structural protein CTCF can result in dysregulation of genes near its binding site (Dixon et al. 2012). We measured the expression of genes upstream of Peak-6 and downstream from *Mecp2* to see whether their expression is altered. We found the expression of *Irak1* (2 kb downstream) and *Bgn* (480 kb downstream) are increased but not *Zfp185*, which is 980 kb downstream from *Mecp2*. Similarly, the expression of *Emd* (140 kb upstream) and *Taz* (175 kb upstream) are elevated but not *Ikbkg* (310 kb upstream), indicating that disruption of this regulatory element can cause dysregulation of some neighboring genes (Supplemental Fig. S6C, D). However, the expression of these genes is not affected in Peak-2 KO mice (Supplemental Fig. S6E). Mouse genetic studies have demonstrated that increasing MeCP2 levels alone is sufficient to cause the neurological phenotypes seen in MDS (Ramocki et al. 2010). These data together with our discovery that the phenotypes seen in Peak-6$^{KO/y}$ mice are a subset of those seen upon doubling MeCP2 support the conclusion that the 70% increase in MeCP2 is the main driver of the disease phenotype in Peak-6$^{KO/y}$ mice.

Given that these two CREs regulate *Mecp2* expression and their sequences are highly conserved in humans, we first mined ATAC-seq data generated from human prefrontal cortex to examine whether these peaks are present (Markenscoff-Papadimitriou et al. 2020). Our analyses revealed that both Peak-2 and Peak-6 are present, suggesting both peaks are conserved (Supplemental Fig. S7A). We next tested whether deletion of these conserved CREs also regulates *MECP2* expression in cultured human neurons. First, we deleted the two CREs in two male, human induced pluripotent stem cell (iPSC) lines using CRISPR/Cas9 (Supplemental Table S2). Next, we generated Vglut1-positive glutamatergic neurons (iNeurons) from these stem cells using directed differentiation via overexpression of *NEUROGENIN2* (*NGN2*) (Supplemental Fig. S8; Zhang et al. 2013). Consistent with our mouse models, *MECP2* mRNA expression is reduced in Peak-2$^{KO/y}$ iNeurons by ~50%, while it is increased in Peak-6$^{KO/y}$ iNeurons by ~30% (Fig. 4A). Protein levels of MeCP2 also showed a concomitant change (Fig. 4B,C; Supplemental Fig. S9). These data strongly suggest that Peak-2 and Peak-6 regulate *MECP2* expression. Furthermore, motif analysis also identified a potential CTCF binding site on the human sequence of Peak-6 (Supplemental Fig. S7B; Heinz et al. 2010), and we validated that CTCF binds within this region using ChIP-qPCR (Supplemental Fig. S7C).

To date, several mouse models with varying levels of MeCP2 have been well characterized; notably, *Mecp2*$^{flox/y}$ mice (50% MeCP2) and transgenic *MECP2-Tg1* (200% MeCP2) show progressive behavioral deficits (Collins et al. 2004; Samaco et al. 2008). Our two CRE KO mouse models (Peak-2 and Peak-6) perturb MeCP2 levels in similar directions (albeit milder) than the *Mecp2*$^{flox/y}$ mice and the *MECP2-Tg1* mice, respectively. Our in vivo studies show that even these subtle alterations are sufficient to produce some behavioral phenotypes observed in *Mecp2*$^{flox/y}$ and *MECP2-Tg1* mice (Fig. 5). Our CRE KO lines provide a unique allelic series demonstrating that disease severity directly corresponds with precise MeCP2 levels in the brain. We

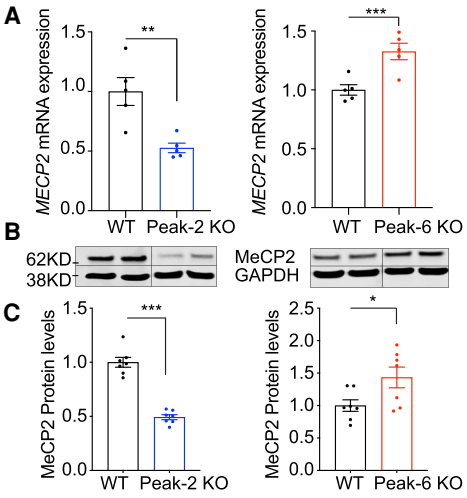


Figure 4. Deletion in conserved *cis*-regulatory elements of *MECP2* affect the mRNA and protein levels in human iPSC-derived neurons. (*A*) *MECP2* mRNA expression is reduced in Peak-2 KO and increased in Peak-6 KO iNeurons as measured by RT-qPCR (n = 5). (*B*) MeCP2 protein expression is reduced in Peak-2 KO and increased in Peak-6 KO iNeurons. Representative Western blot of MeCP2 levels in CRE deletion iNeurons with GAPDH displayed as loading control. Solid vertical line indicates a spliced region of the single gel image to remove unneeded lanes (n = 7). (*C*) Quantification of MeCP2 protein expression by Western blot normalized to GAPDH loading control (n = 7). All data were analyzed by two-tailed t-test. Data are presented as mean ± SEM. (*) $P < 0.05$, (**) $P < 0.01$, (***) $P < 0.001$.

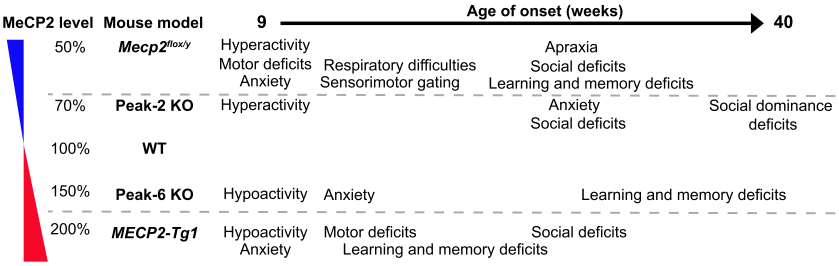


Figure 5. Summary of Peak KO phenotypes and comparison with MeCP2 50% loss (*Mecp2*$^{flox/y}$) or MDS mouse model (*MECP2-Tg1*).

demonstrate, in agreement with previous work, that normal MeCP2 dosage (100%) is required for normal brain function in mice, and either a 20%–30% decrease or a 50%–70% increase in MeCP2 may lead to some neuropsychiatric phenotypes (altered activity, anxiety, and social and learning deficits), while 50% decrease or 100% increase of MeCP2 leads to severe neurodevelopmental disorders like RTT or MDS, respectively.

MeCP2 expression has a unique spatial and temporal pattern, where it is the lowest in the liver and highest in the brain (Supplemental Fig. S10A) and increases significantly postnatally (Supplemental Fig. S10B). Accurate postnatal MeCP2 level is critical for normal brain function. Generating ATAC-seq profiles during developmental stages that coincide with the timing of MeCP2 increase identified regions with accessible chromatin structure at the *MECP2* gene locus. These accessible regions may contain specific chromatin remodelers or transcription factors that regulate *MECP2* expression. Further studies will be needed to identify the putative transcription factor or factors that bind to Peak-2 or Peak-6 to regulate *MECP2* expression.

Understanding the phenotypic range that results from varying MeCP2 levels is critical for a number of reasons: first, assessing the type and severity of clinical manifestations of *MECP2* dosage-related disorders; second, classifying and treating non-RTT patients with neuropsychiatric disorders due to mild *MECP2* mutations or noncoding mutations that affect MeCP2 levels (e.g., in CREs); and third, predicting the clinical benefits of therapeutic interventions based on the degree of modulation of MeCP2 levels. Studies on existing mouse models across a range of MeCP2 levels have shown that phenotypic severity and symptom onset vary in correlation with the deviation from normal (Chao and Zoghbi 2012). Importantly, our CRE KO models provide information about the previously unexplored intermediary range between 50% and 100% (Peak-2 KO) and between 150% and 200% (Peak-6 KO). Our study suggests that reducing the MeCP2 protein level from 200% in MDS patients by even a small degree to 150%–170% could improve many phenotypes, such as motor and sensorimotor deficits. Similarly, in patients with a <50% MeCP2 protein level, boosting MeCP2 level to just 70% could potentially improve some RTT phenotypes. This provides hope that even treatments that slightly correct MeCP2 protein level in MDS or RTT patients could achieve clinically relevant phenotypic improvement.

Noncoding regions account for >98% of the human genome, yet we have limited understanding of their contribution to disease. Here, we demonstrate that disruption of CREs alters MeCP2 levels in cultured human iPSC-derived neurons. Human mutations in these *cis*-regulatory regions could contribute to atypical RTT patients without mutations in *MECP2* coding regions or patients with ID, autism, or neuropsychiatric disorders. Many genes that cause neurodevelopmental disorders and autism (Satterstrom et al. 2020) are dosage sensitive (Han et al. 2013; Rocha et al. 2016; Rice and McLysaght 2017; Raveau et al. 2018; Schnabel et al. 2018). It is possible that unidentified mutations in regulatory elements of these genes may also affect their abundance and lead to disease phenotype. Our research highlights the importance of whole-genome sequencing (WGS) to achieve diagnoses in cases in which exome sequencing fails. Furthermore, the catalog of mutations by WGS on the X chromosome is underrepresented due to the reduced effective population size of sequenced X chromosomes, as males only carry one copy (Telenti et al. 2016). Our study also demonstrates that representation of the noncoding regions on the X-linked gene *MECP2* in WGS needs to be reexamined to detect any potential variants in these critical regions.

In conclusion, we identified and functionally characterized two novel, evolutionarily conserved regulatory elements required for normal expression of *Mecp2*/*MECP2*. Deletion of either regulatory element in mice caused mild neurological dysfunction, highlighting how small changes in MeCP2 levels in either direction disrupt neurological function. More broadly, this study underscores the potential contribution of mutations in regulatory regions to various neuropsychiatric phenotypes and calls for similar studies of regulatory regions in other dosage sensitive genes involved in autism and ID.

Materials and methods

Animals

Mice were housed in an AAALAS-certified level 3 facility on a 14-h light cycle. All CRE knockout mice were generated in the Genetically Engineered Mouse Core at Baylor College of Medicine, backcrossed with wild-type C57BL6/J mice for five generations. Only male offspring were used for analysis. All procedures to maintain and use these mice were approved by the Institutional Animal Care and Use Committee for Baylor College of Medicine and Affiliates.

Generation of the knockout mice of regulatory elements

All CRE deletion mice were generated via CRISPR/Cas9-mediated gene editing. Briefly, two sgRNAs targeting the 5′ and 3′ (left [L] and right [R]) ends of the putative CREs were designed in Benchling and synthesized by IDT. sgRNAs were in vitro transcribed with the MEGAshortscript T7 transcription kit (Invitrogen). Details about injection are available in the Supplemental Material.

Behavioral assays

All data acquisition and analyses were performed by an individual blinded to the genotype. All behavioral studies were performed during the light period. At least 1 d was given between assays for the mice to recover. All the tests were performed as previously described (Chao et al. 2010) with few modifications. A detailed description is in the Supplemental Material.

ATAC-seq

Nuclear isolation was performed according to Mo et al. (2015) with two biological replicates. Nuclei were collected from fresh brain tissues for ATAC-seq. Detailed descriptions are available in the Supplemental Material.

Statistical analysis

Statistical significance was determined using GraphPad Prism software. The number of animals (n) and the specific statistical tests for each experiment are indicated in the figure legends. Sample size for behavioral studies was determined based on previous experience using mice with the same background.

Data and materials

All data needed to evaluate the conclusions in this study are present here and/or the Supplemental Material. ATAC-seq data from different mouse neuronal cells were obtained from Mo et al. (2015) and GEO GSE63137. ATAC-seq data from human brains were obtained from GSE149268. The accession number for the raw and processed data files reported here is GEO GSE152719. Additional data related to this study are available on request.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Y. Sun for mouse genotyping and the Genetically Engineered Mouse Core, Human Stem Cell Core, RNA In Situ Hybridization Core, Neurovisualization Core, Neurobehavioral Core, and Human Neuronal Differentiation Core at the Jan and Dan Duncan Neurological Research Institute. This project was funded by the National Institutes of Health (NIH; 5R01NS057819 to H.Y.Z., and 1F32HD100048-01 to S.S.B.), the American Heart Association (19PRE34410104 to M.C.G.), institutional startup funds (to J.D.W.), the Department of Defense (W81XWH-18-1-0350 to J.D.W.), Canadian Institutes of Health Research (PJT-155922 to J.D.W.), the Howard Hughes Medical Institute (to H.Y.Z.), the Baylor College of Medicine Intellectual and Developmental Disabilities Research Center (NIH 5P50HD103555), the Henry Engel Fund, and the Ziff family fund.

Author contributions: Y.S. and H.Y.Z. designed the project. Y.S. performed the biochemistry, behavioral, and iNeuron experiments. H.P.T. performed behavioral experiments. S.S.B. performed iNeuron experiments. M.C.G. and J.D.W. performed ATAC-seq experiments and conducted bioinformatic analysis. Y.S. wrote the manuscript and prepared the figures, with contributions from all coauthors according to their area of expertise. H.Y.Z. reviewed all of the data and edited the manuscript.

References

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

Boxer LD, Renthal W, Greben AW, Whitwam T, Silberfeld A, Stroud H, Li E, Yang MG, Kinde B, Griffith EC, et al. 2020. MeCP2 represses the rate of transcriptional initiation of highly methylated long genes. *Mol Cell* **77:** 294–309.e9. doi:10.1016/j.molcel.2019.10.032

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10:** 1213–1218. doi:10.1038/nmeth.2688

Chahrour M, Jung SY, Shaw C, Zhou X, Wong ST, Qin J, Zoghbi HY. 2008. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320:** 1224–1229. doi:10.1126/science.1153252

Chao HT, Zoghbi HY. 2012. MeCP2: only 100% will do. *Nat Neurosci* **15:** 176–177. doi:10.1038/nn.3027

Chao HT, Chen H, Samaco RC, Xue M, Chahrour M, Yoo J, Neul JL, Gong S, Lu HC, Heintz N, et al. 2010. Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468:** 263–269. doi:10.1038/nature09582

Collins AL, Levenson JM, Vilaythong AP, Richman R, Armstrong DL, Noebels JL, David Sweatt J, Zoghbi HY. 2004. Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum Mol Genet* **13:** 2679–2689. doi:10.1093/hmg/ddh282

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485:** 376–380. doi:10.1038/nature11082

Gennarino VA, Alcott CE, Chen CA, Chaudhury A, Gillentine MA, Rosenfeld JA, Parikh S, Wheless JW, Roeder ER, Horovitz DD, et al. 2015. NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. *Elife* **4:** e10782. doi:10.7554/eLife.10782

Han K, Holder JL Jr, Schaaf CP, Lu H, Chen H, Kang H, Tang J, Wu Z, Hao S, Cheung SW, et al. 2013. SHANK3 overexpression causes manic-like behaviour with unique pharmacogenetic properties. *Nature* **503:** 72–77. doi:10.1038/nature12630

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38:** 576–589. doi:10.1016/j.molcel.2010.05.004

Kazdoba TM, Leach PT, Yang M, Silverman JL, Solomon M, Crawley JN. 2016. Translational mouse models of autism: advancing toward pharmacological therapeutics. *Curr Top Behav Neurosci* **28:** 1–52.

Liu J, Francke U. 2006. Identification of *cis*-regulatory elements for MECP2 expression. *Hum Mol Genet* **15:** 1769–1782. doi:10.1093/hmg/ddl099

Lombardi LM, Zaghlula M, Sztainberg Y, Baker SA, Klisch TJ, Tang AA, Huang EJ, Zoghbi HY. 2017. An RNA interference screen identifies druggable regulators of MeCP2 stability. *Sci Transl Med* **9:** eaaf7588. doi:10.1126/scitranslmed.aaf7588

Markenscoff-Papadimitriou E, Whalen S, Przytycki P, Thomas R, Binyameen F, Nowakowski TJ, Kriegstein AR, Sanders SJ, State MW, Pollard KS, et al. 2020. A chromatin accessibility atlas of the developing human telencephalon. *Cell* **182:** 754–769.e18. doi:10.1016/j.cell.2020.06.002

Mo A, Mukamel EA, Davis FP, Luo C, Henry GL, Picard S, Urich MA, Nery JR, Sejnowski TJ, Lister R, et al. 2015. Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* **86:** 1369–1384. doi:10.1016/j.neuron.2015.05.018

Moretti P, Bouwknecht JA, Teague R, Paylor R, Zoghbi HY. 2005. Abnormalities of social interactions and home-cage behavior in a mouse model of Rett syndrome. *Hum Mol Genet* **14:** 205–220. doi:10.1093/hmg/ddi016

Nagarajan RP, Hogart AR, Gwye Y, Martin MR, LaSalle JM. 2006. Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation. *Epigenetics* **1:** 172–182. doi:10.4161/epi.1.4.3514

Oz-Levi D, Olender T, Bar-Joseph I, Zhu Y, Marek-Yagel D, Barozzi I, Osterwalder M, Alkelai A, Ruzzo EK, Han Y, et al. 2019. Noncoding deletions reveal a gene that is critical for intestinal function. *Nature* **571:** 107–111. doi:10.1038/s41586-019-1312-2

Ramocki MB, Tavyev YJ, Peters SU. 2010. The *MECP2* duplication syndrome. *Am J Med Genet A* **152A:** 1079–1088. doi:10.1002/ajmg.a.33184

Raveau M, Shimohata A, Amano K, Miyamoto H, Yamakawa K. 2018. DYRK1A-haploinsufficiency in mice causes autistic-like features and febrile seizures. *Neurobiol Dis* **110:** 180–191. doi:10.1016/j.nbd.2017.12.003

Rice AM, McLysaght A. 2017. Dosage-sensitive genes in evolution and disease. *BMC Biol* **15:** 78. doi:10.1186/s12915-017-0418-y

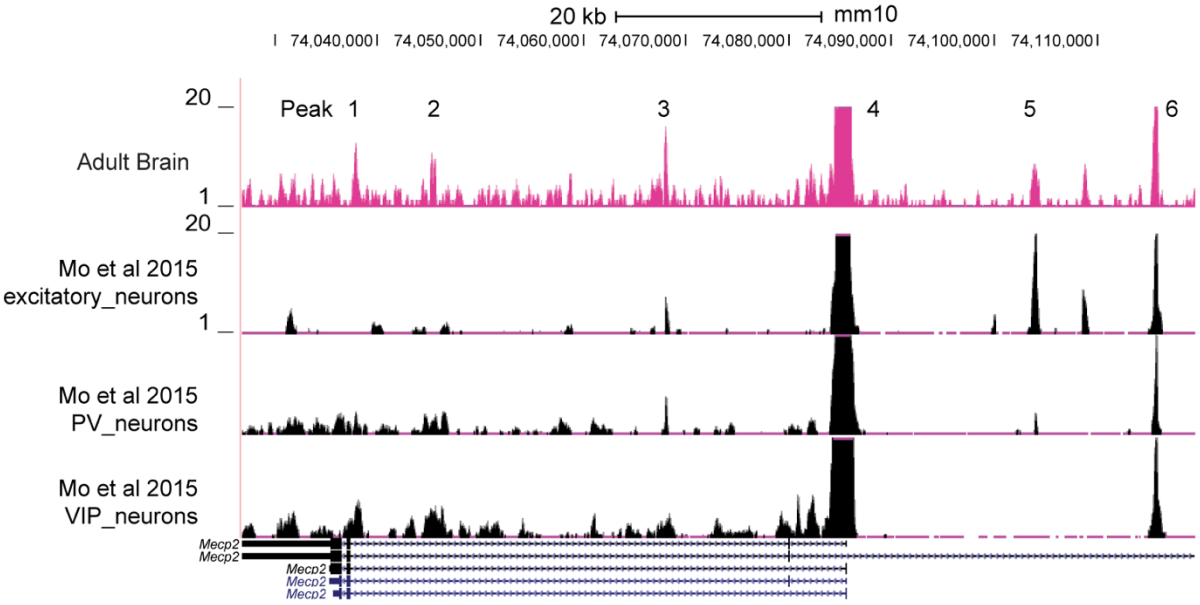
Rocha H, Sampaio M, Rocha R, Fernandes S, Leão M. 2016. MEF2C haploinsufficiency syndrome: report of a new MEF2C mutation and review. *Eur J Med Genet* **59:** 478–482. doi:10.1016/j.ejmg.2016.05.017

Rodrigues DC, Kim DS, Yang G, Zaslavsky K, Ha KC, Mok RS, Ross PJ, Zhao M, Piekna A, Wei W, et al. 2016. MECP2 is post-transcriptionally

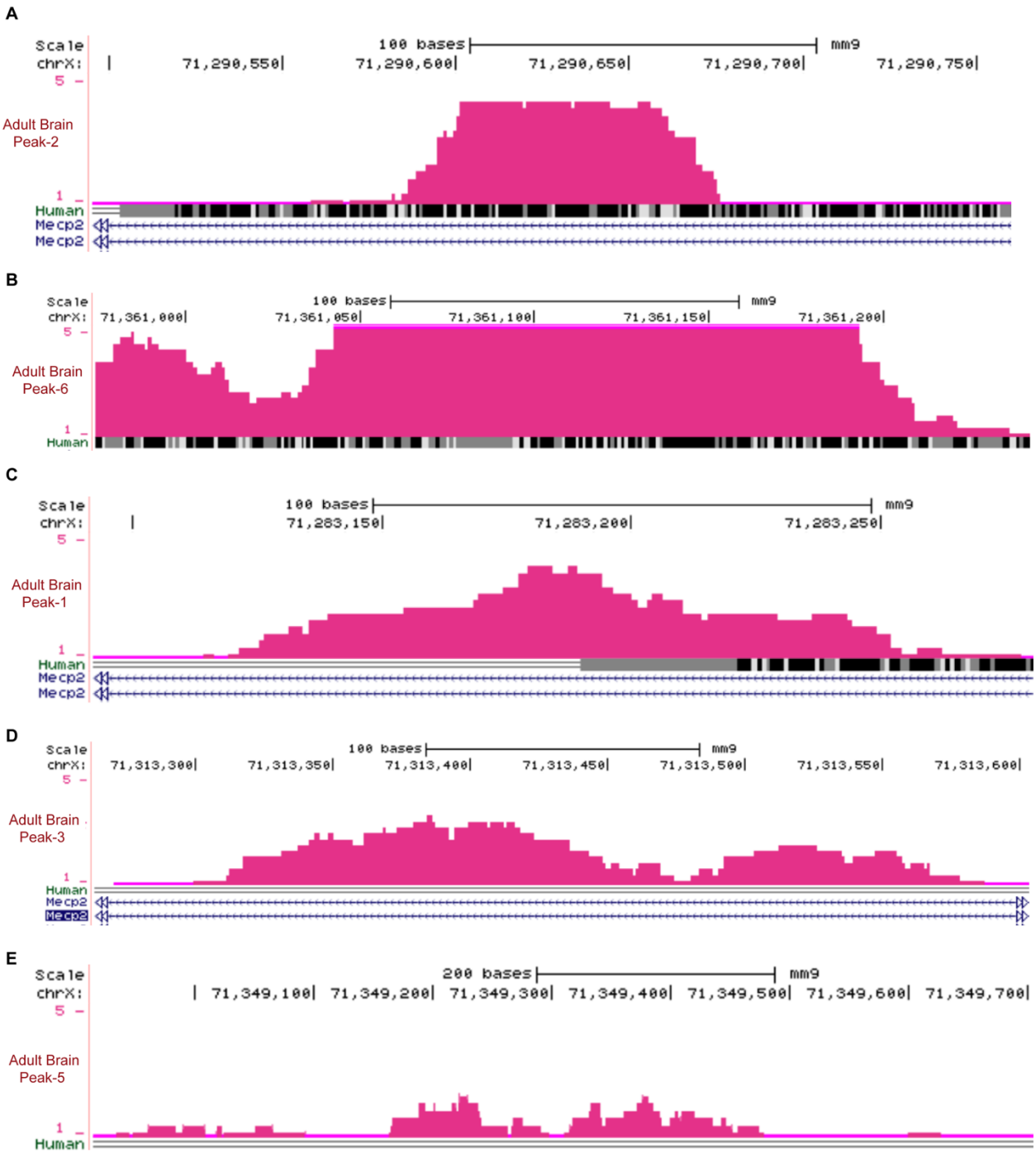
Supplementary Figures

Peak	Deletion size (bp)	Germline Transmitted
1	400	2 females
2	1154	1 male
3	870	1 male
5	859	2 females
6	1440	1 female

Sup. Table. 1. Deletion size and founder information of each CRE knockout mouse line



Sup. Fig. 1. Identification of accessible chromatin regions surrounding, and within, the murine *Mecp2* locus. Comparison between ATAC-seq data from whole adult brain (this study, pink) and previous ATAC-seq data (black) from excitatory neurons, PV neurons and VIP neurons (*Mo et. al.,* 2015). All data are mapped to the *Mus musculus* genome 10 (mm10).



Sup. Fig. 2. Conservation alignment of putative *cis*-regulatory elements of *MECP2* in Mice and Humans

(A) Comparison of sequence conservation for Peak-2 in Mus musculus Genome (mm9) and Human Genome (hg19) as viewed in the UCSC genome browser.

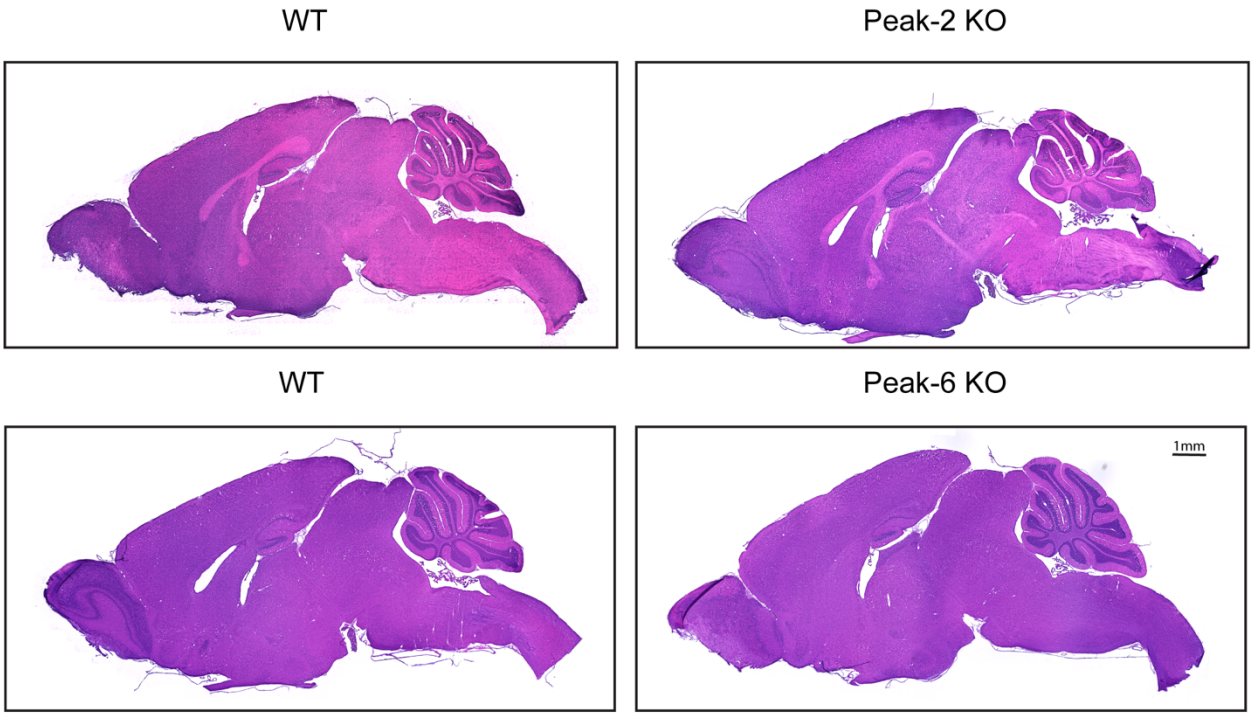
(B) Comparison of sequence conservation for Peak-6 in Mus musculus Genome (mm9) and Human Genome (hg19) as viewed in the UCSC genome browser.

(C) Comparison of sequence conservation for Peak-1 in Mus musculus Genome (mm9) and Human Genome (hg19) as viewed in the UCSC genome browser.

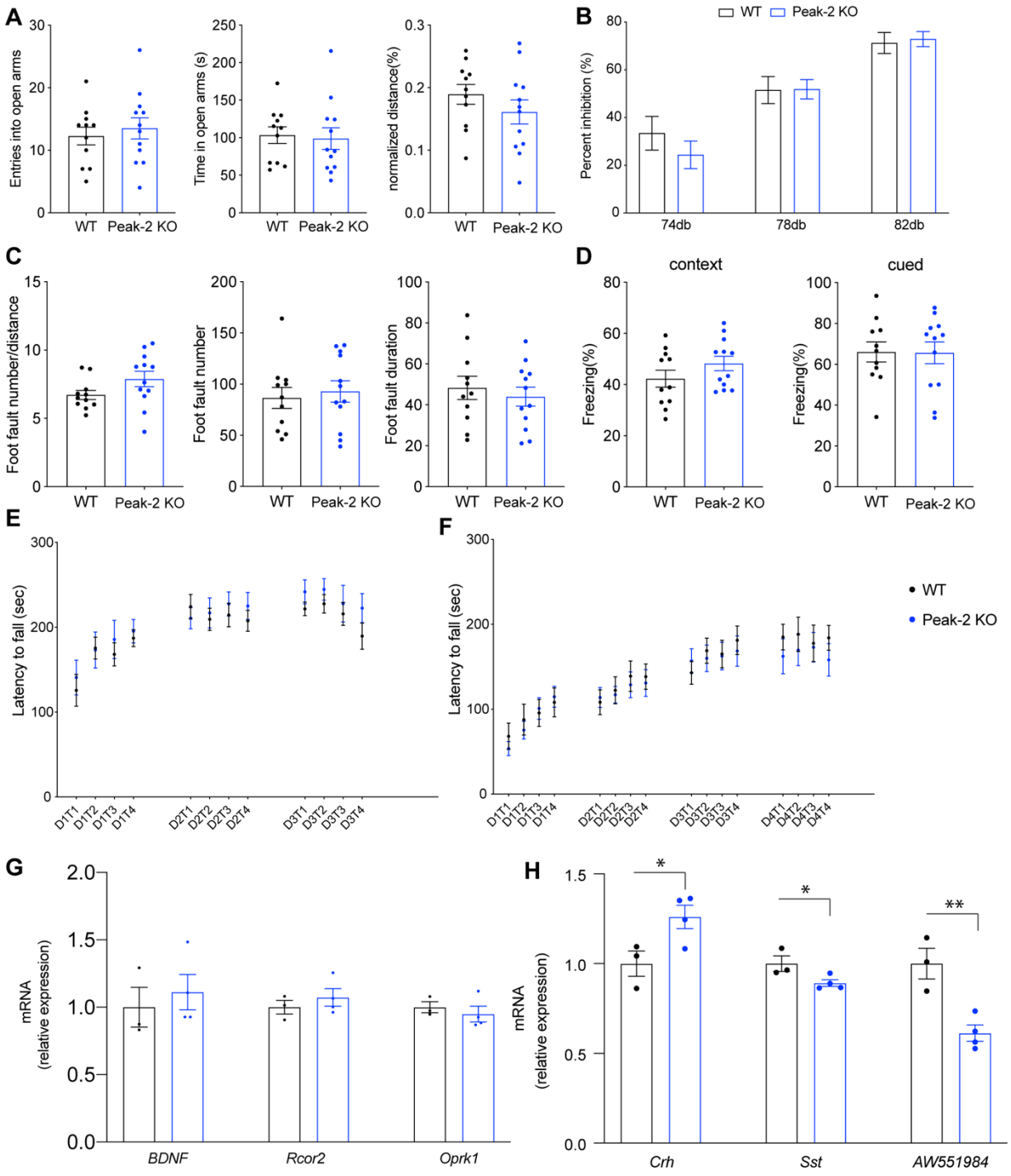
(D) Comparison of sequence conservation for Peak-3 in Mus musculus Genome (mm9) and Human Genome (hg19) as viewed in the UCSC genome browser.

(E) Comparison of sequence conservation for Peak-5 in Mus musculus Genome (mm9) and Human Genome (hg19) as viewed in the UCSC genome browser.

For all panels, the pink peaks (top of each panel) correspond to the number of reads mapped to the murine genome for each respective peak (the y-axis is shown on the far left). The grey scale color bar on the human sequence track (labelled in green text at the bottom left of each panel) indicates sequence conservation, with a darker hue (from white to black) indicating greater sequence conservation. The scale represents probability of conservation between 0-1. A blue line means no conserved sequence is detected in the human genome.



Sup. Fig. 3. Cresyl Violet staining does not identify any histological abnormalities in Peak-2 KO and Peak-6 KO mice compared to their littermate controls.



Sup. Fig. 4. Behavioral characterization of Peak-2$^{KO/y}$

(A) Peak-2$^{KO/y}$ mice do not show anxiety-like phenotype in Elevated Plus Maze at 9 weeks of age (n=12)

(B) Peak-2$^{KO/y}$ mice do not show sensorimotor gating deficits in Pre-pulse Inhibition assay at 24 weeks of age (n=12)

(C) Peak-2$^{KO/y}$ mice do not show motor incoordination in foot slip assay at 26 weeks of age (n=12)

(D) Peak-2$^{KO/y}$ mice do not show learning and memory deficits in fear conditioning assay at 30 weeks of age (n=12)

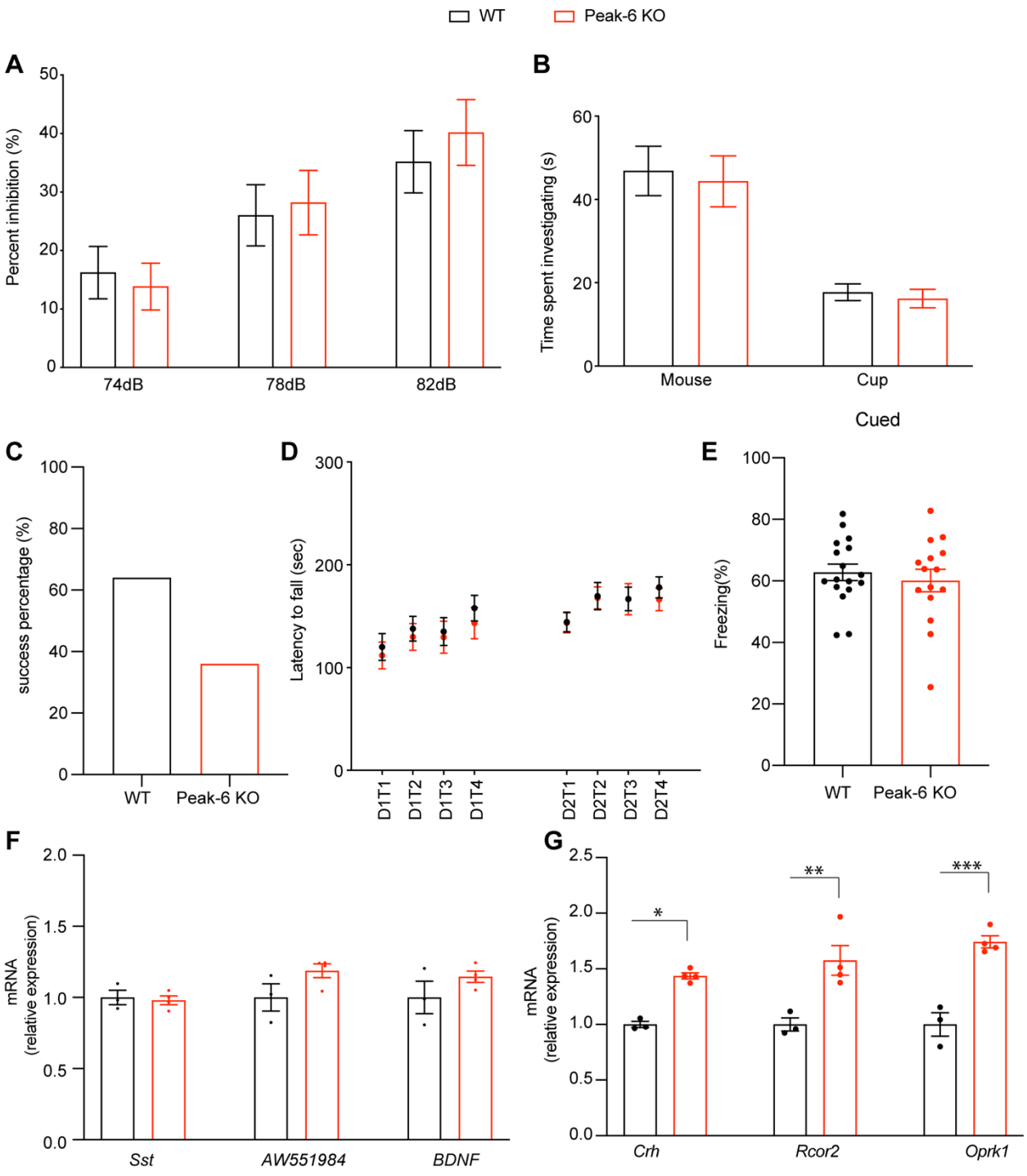
(E) Peak-2$^{KO/y}$ mice have normal motor coordination and learning in Rotarod assay at 10 weeks of age (n=12)

(F) Peak-2$^{KO/y}$ mice have normal motor coordination and learning in Rotarod assay at 10 months of age (n=12)

(G) A subset of genes misregulated in MDS mice do not show differential expression in Peak-2 $^{KO/y}$ compared to wild-type mice at 12 weeks of age (n=3)

(H) Peak-2$^{KO/y}$ mice show RTT-like abnormal gene expression measured with RT-qPCR in cortex at 12 weeks of age (n=3)

Data are presented as mean ± s.e.m. All data were analyzed by two-tailed *t*-test, except figure E, F were analyzed by two-way ANOVA with repeated measures and figure G, H were analyzed by one-way ANOVA followed by Tukey post hoc test. *$P < 0.05$; ** $P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$. n.s. not significant.



Sup. Fig. 5. Behavioral characterization of Peak-6 KO

(A) Peak-$6^{KO/y}$ mice do not show sensorimotor gating deficits in Pre-pulse Inhibition assay at 24 weeks of age (n=13-15)

(B) Peak-6$^{KO/y}$ mice have normal social interaction in three-chamber assay at 24 weeks of age (n=13-15)

(C) Peak-6$^{KO/y}$ mice do not have social dominance deficit in Tube test at 30 weeks of age (n=13-15)

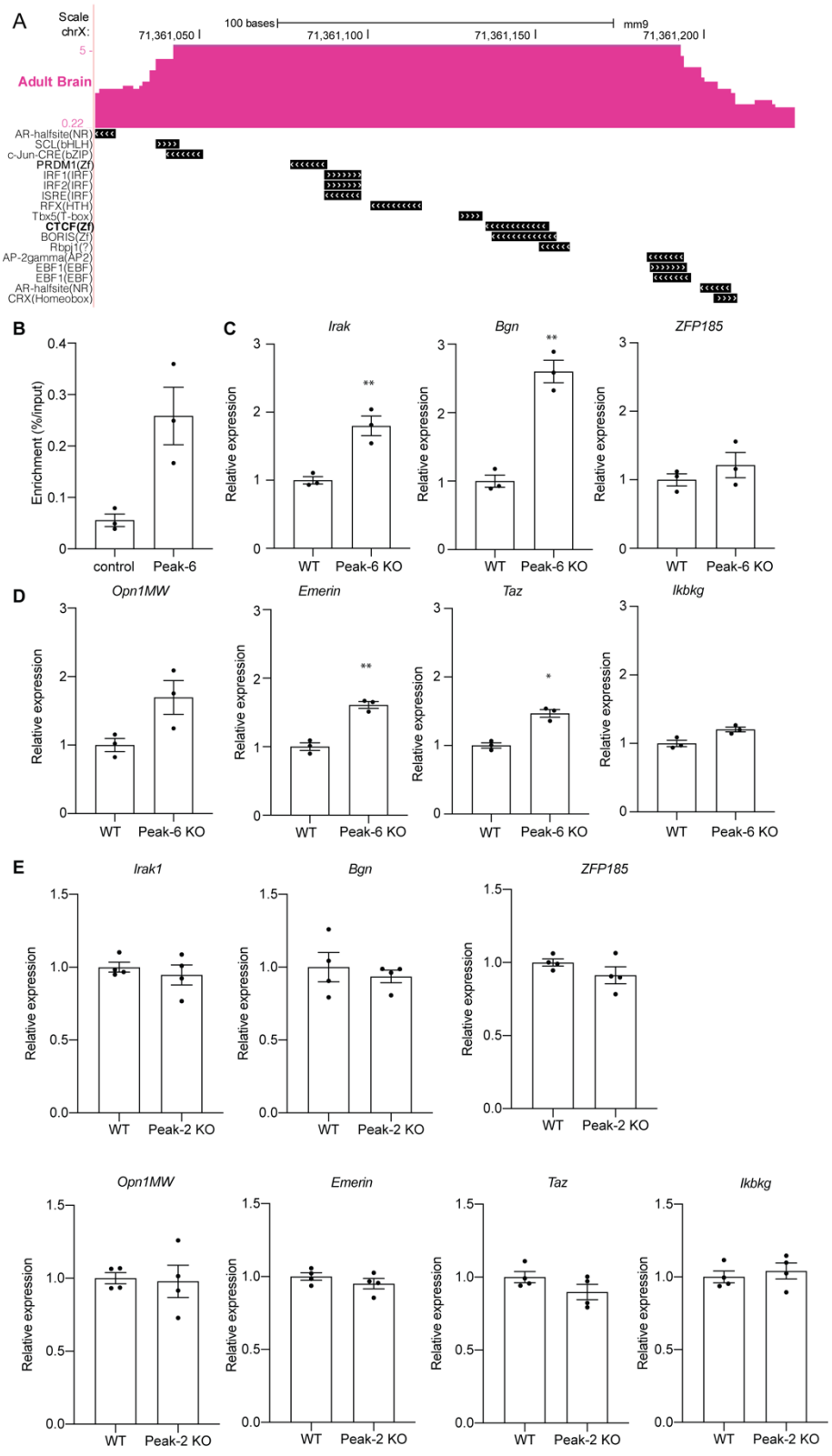
(D) Peak-6$^{KO/y}$ mice have normal motor coordination and learning in Rotarod assay at 30 weeks of age (n=13-15)

(E) Peak-6$^{KO/y}$ mice show normal cued learning using Fear Conditioning test at 36 weeks of age (n=13-15)

(F) A subset of genes misregulated in MDS mice did not show differential expression in Peak-6 $^{KO/y}$ compared to wild-type mice at 12 weeks of age (n=3)

(G) Peak-6$^{KO/y}$ mice show MDS-like abnormal gene expression measured with RT-qPCR in cortex at 12 weeks of age (n=3-4)

Data are presented as mean ± s.e.m. All data were analyzed by two-tailed *t*-test, except figure D were analyzed by two-way ANOVA with repeated measures and figure F, G were analyzed by one-way ANOVA followed by Tukey post hoc test. *$P < 0.05$; ** $P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.



Sup. Fig. 6. Analysis of putative transcription factors that bind to the cis-regulatory elements

(A) Transcription factor DNA binding site motifs present in the Peak 6 region were identified using HOMER. Black rectangles display the region where the predicted motifs occur within the Peak 6 sequence.

(B) ChIP on CTCF followed by qPCR on Peak 6 using frontal cortex of 8-week-old wildtype mice (n=3)

(C) qPCR shows the gene expression downstream of *Mecp2* in 10-week-old Peak 6-KO mice (n=3)

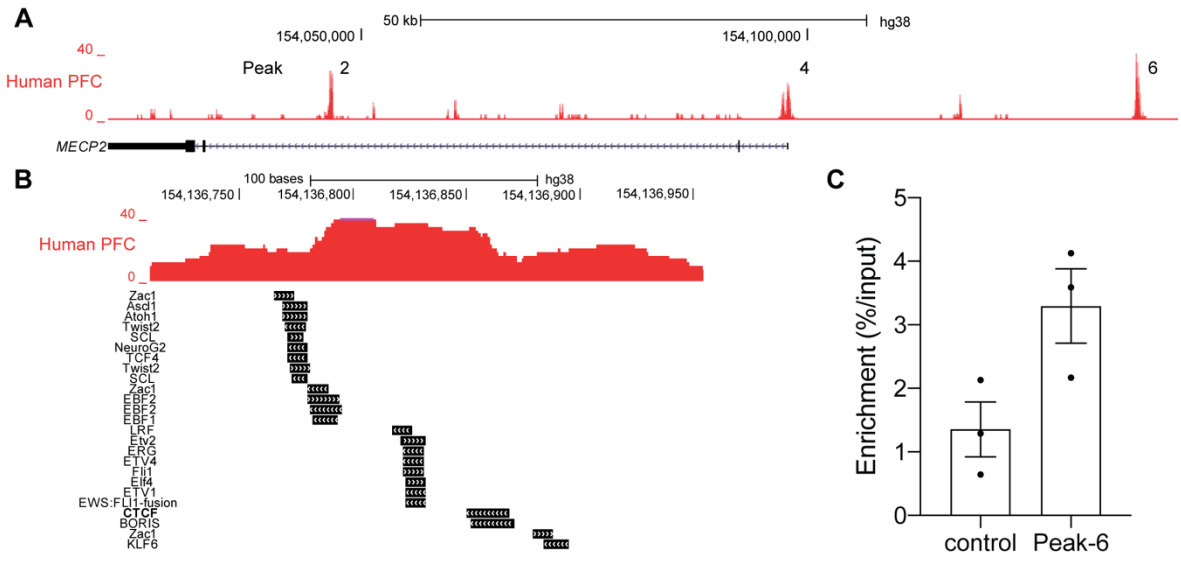
(D) qPCR shows the gene expression upstream of Peak 6 in 10-week-old Peak 6-KO mice (n=3)

(E) qPCR shows the expression of genes surrounding *Mecp2* in 10-week-old Peak 2-KO mice (n=4)

Data are presented as mean ± s.e.m. All data were analyzed by two-tailed *t*-test. **P* < 0.05; ** *P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

Target region	gRNA-L	gRNA-R
Peak-2	chrX:153,317,193-153,317,212	chrX:153,316,423-153,316,442
Peak-6	chrX:153,402,591-153,402,610	chrX:153,402,097-153,402,116

Sup. Table 2 Coordinates of gRNA used in editing human iPSCs

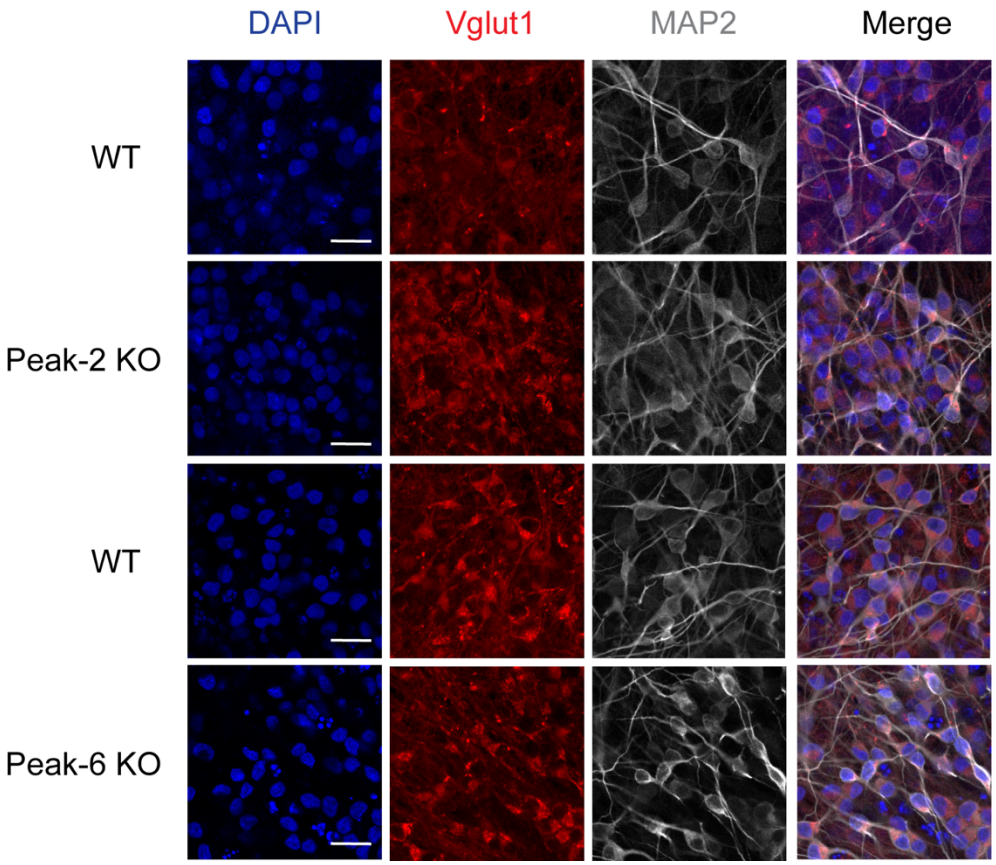


Sup. Fig. 7. ATAC-seq and motif analysis in human prefrontal cortex

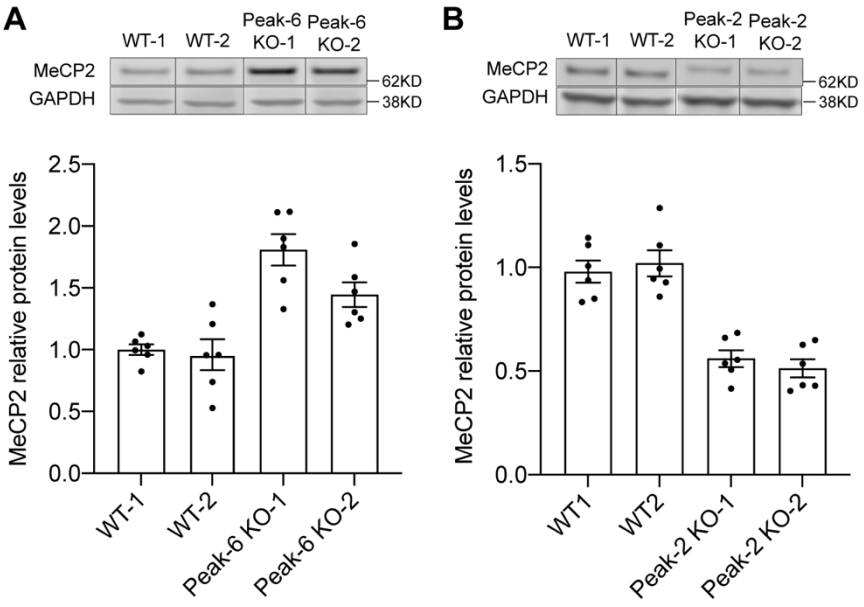
(A) ATAC-seq reads mapped to the Human Genome (hg38) are shown from nuclei isolated from the human prefrontal cortex in a previously published dataset (GSE149268). Exons are denoted by solid black rectangles, while introns are between exons. The 3' UTR is represented by a thinner but longer block relative to the coding exons. Note that locus is transcribed in the opposite direction relative to the image shown (e.g. the promoter is on the right side and the 3' UTR is on the left). Peaks 2 and 6 were indicated the same as the ones in the mouse (Figure 1), whereas Peak 4 spans the proximal promoter.

(B) Motif of putative binding factors in the Peak 6 region in the human genome were identified using HOMER. Black rectangles display the region where the predicted motifs occur within the human Peak 6 sequence.

(C) ChIP on CTCF followed by qPCR on Peak-6 using iNeurons induced from iPSCs (n=3)



Sup. Fig. 8. Vglut1 immunofluorescence staining in human iNeurons confirms NGN2 overexpression generates glutamatergic neurons. Scale bar is 20μm.

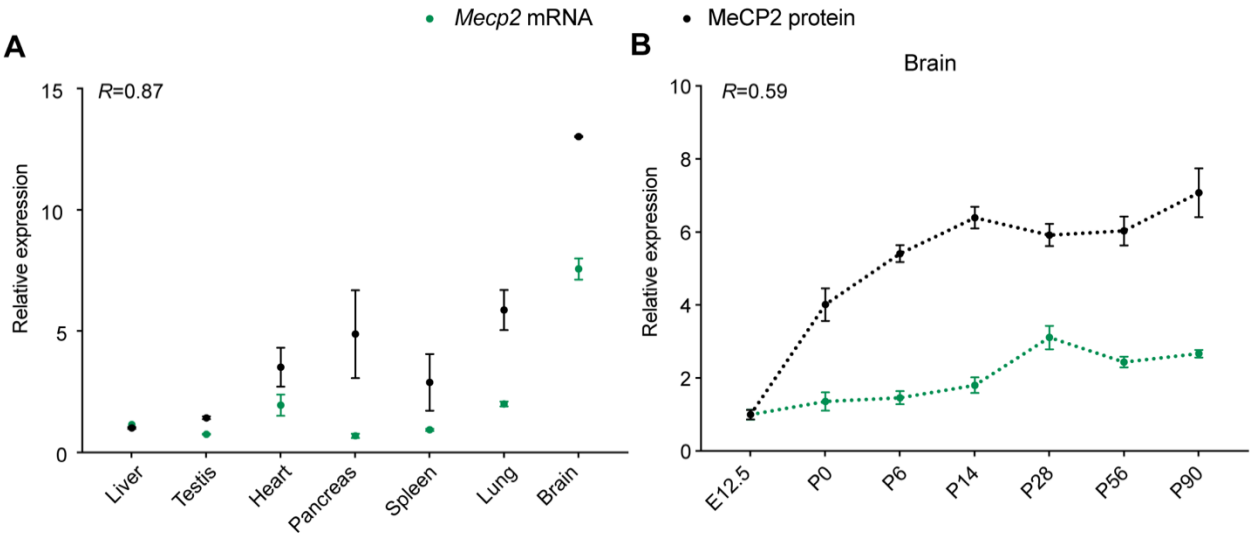


Sup. Fig. 9. Deletion in conserved *cis*-regulatory elements of *MECP2* affect the MeCP2 protein levels in human iPSC-derived neurons from another independent individual

(A) MeCP2 protein expression is increased in Peak-6 KO iNeurons. Quantification of MeCP2 protein expression by Western blot normalized to GAPDH loading control (n=6)

(B) MeCP2 protein expression is reduced in Peak-2 KO iNeurons. Quantification of MeCP2 protein expression by Western blot normalized to GAPDH loading control (n=6)

Two clones per line were generated and tested. Representative Western blot of MeCP2 levels in CRE deletion iNeurons with GAPDH displayed as loading control. Solid vertical line indicates a spliced region of the single gel image to remove unneeded lanes



Sup. Fig. 10. Spatial and temporal regulation of MeCP2 RNA and protein

(A) *Mecp2* mRNA and MeCP2 protein level in seven different tissues from adult mice were measured by RT-qPCR or Western blot, respectively (n=5). Pearson correlation between mRNA expression and protein level is shown in the upper left.

(B) *Mecp2* mRNA and MeCP2 protein level in the brain at seven different developmental time points were measured by RT-qPCR or Western blot, respectively (n=4). Pearson correlation between mRNA expression and protein level is shown in the upper left.

Supplementary Methods

Generation of the knockout mice of regulatory elements

On the day of injection, Cas9 protein (PNA Bio) and sgRNAs were injected (pronuclear) into ova from wild-type C57BL6/J female mice and transferred into oviducts of pseudopregnant females. Founder mice (Table S1) were sequence-verified and backcrossed at least five times prior to experimentation to eliminate potential off-target mutations.

sgRNAs were designed as following:

Peak	gRNA-1	gRNA-2
Peak-1	chrX:71,283,071-71,283,090	chrX:71,283,484-71,283,503
Peak-2	chrX:71,290,301-71,290,320	chrX:71,291,475-71,291,494
Peak-3	chrX:71,313,274-71,313,293	chrX:71,314,167-71,314,186
Peak-5	chrX:71,349,107-71,349,126	chrX:71,349,991-71,350,010
Peak-6	chrX:71,360,308-71,360,327	chrX:71,361,755-71,361,774

Brain lysates preparation and Western Blot

Cortex from the mice were dissected and homogenized in cold lysis buffer (20mM Tris-HCl pH=8.0, 180mM NaCl, 0.5% NP-40, 1mM EDTA, 2% SDS, Complete Protease inhibitor, Roche). Lysates were rotated at room temperature for 20 min and then centrifugated at maximum speed for 20 min at room temperature. The supernatant was then mixed with 1x NuPAGE sample buffer, heated for 10 min at 95°C and run on a NuPAGE 4-12% Bis-Tris gradient gel with MES SDS running buffer (NuPAGE, Carlsbad, CA). Separated proteins were transferred to a nitrocellulose membrane using NuPAGE transfer buffer for 2.5hr at 4°C at 120V. The membrane was then blocked with 5% BSA in TBS with 2% Tween-20 (TBST) for 1 hr and incubated with primary antibody diluted in blocking buffer overnight at 4°C. After three washes with TBST, the membrane was incubated with secondary antibody provided by LiCOR in Odyssey Blocking Buffer (0.5x) for 1-2hr at room temperature, followed by three washes with TBST.

Images were acquired on a LiCOR CLx Imager. The antibodies used were MeCP2 (D4F3) Rabbit mAb #3456 (1:1,000; Cell Signaling) and mouse anti-GAPDH (1:10,000; Abcam ab8245).

Gene expression analysis by qRT-PCR

Total RNA was extracted from different regions of the adult mouse brain using a Qiagen RNeasy Mini kit. 3 µg of total RNA was used to synthesize cDNA by M-MLV reverse transcriptase (Life Technologies, Carlsbad, CA). qRT-PCR was performed in a CFX96 Real-Time System (Bio-Rad) using PowerUp SYBR Green Master Mix (ThermoFisher). Sense and antisense primers spanned multiple exons, and the RNA was treated with DNase prior to reverse transcription to avoid false-positive results caused by genomic DNA contamination. The specificity of the amplification products was verified by melting curve analysis. All qRT-PCR reactions were conducted in technical triplicates, with a minimum of 3 biological replicates, and the results were averaged for each sample, normalized to S16 levels, and analyzed using the comparative ΔΔCt method.

Chromatin Immunoprecipitation

ChIP was performed as previously described(Ito-Ishida et al. 2018) with minor modifications. Frozen tissue was cross-linked by incubating in 1% PFA (Fisher, # 28906) in PBS for 10 min at room temperature. After quenching the fixation with 125 mM glycine in ice-cold PBS, the samples were homogenized in hypotonic solutions (10 mM Tris (pH 7.5), 0.5% Igepal CA-630, 10 mM NaCl, 30 mM $MgCl_2$, 1:1000 RNase cocktail (Ambion, #2286), 1 mM PMSF, and 1x CPI). Nuclei were collected at 3,000 rpm in a table-top centrifuge and resuspended in nuclei lysis buffer containing 50 mM Tris (pH 8), 10 mM EDTA, 1% SDS, 1 mM PMSF, and 1x CPI. The sample was incubated in the nuclei lysis buffer for 15 min on ice, and sonicated by Bioruptor Pico (Diagenode) for 10 cycles (30 sec off/ 30 sec on) to obtain DNA fragments of 100-500 bp. The chromatin samples were flash frozen in liquid nitrogen and stored at -80 ºC until needed. To perform ChIP, chromatin sample was first diluted based on their DNA concentration, further diluted in ChIP dilution buffer (0.01% SDS, 1.1% Triton-X, 1.2 mM EDTA, 16.7

mM Tris (pH 8.1), and 167 mM NaCl), and was prerecleared with Protein A Dynabeads (Invitrogen) for 1 hour. Approximately 5 μg of chromatin was incubated overnight at 4 °C with anti-MeCP2 antibody as suggested (Abcam, #ab2828). An aliquot of 10% of the precipitated chromatin was stored as input. The next day, 40 μl of Protein A Dynabeads were added, and the sample was rotated for 3 hours at 4 °C. The beads were then washed in 700 μl each of low-salt buffer (0.1% SDS, 1% Triton-X, 2 mM EDTA, 20 mM Tris (pH 8.1), and 150 mM NaCl), high-salt buffer (0.1% SDS, 1% Triton-X, 2 mM EDTA, 20 mM Tris (pH 8.1) and 500 mM NaCl), LiCl wash buffer (250 mM LiCl, 1% Igepal-CA630, 1 mM EDTA and 10 mM Tris (pH 8.1)) and twice with TE+NaCl buffer (10 mM Tris HCl, 1 mM EDTA and 50 mM NaCl), for 5 min each at room temperature. The beads were eluted twice in 250 μl of elution buffer (1%SDS, 100 mM $NaHCO_3$) for 15 min each at 65 °C. Precipitated chromatin and input samples were reverse crosslinked and treated with proteinase K. ChIP-DNA and input-DNA were recovered using PCR Purification Kit (Qiagen), and were used for qPCR. Following primers were used for ChIP-qPCR experiments:

Mouse:

control: forward 5'- *cctatatatttgatggtagcaaaaggg*-3', reverse 5'- *actccataaattatttgacatcccc*-3'.

Peak-6: forward 5'- *tctcctttgtgcaaatgcag*-3', reverse 5'- *aatttacaattttcaagaacccagg*-3'.

Human:

Control: forward 5'- *tttggactgcacagatattct*-3', reverse 5'-*aggtactgagggttaggact*-3'.

Peak-6: forward 5'- *gccttctgtctgccttagca*-3', reverse 5'-*aggatccacccttcaggac*-3'.

Peak-6 primers amplify the region with the prediction of CTCF binding motif on Peak-6. Control primers amplify the region in the middle of MeCP2 intron 2 without any ATAC-seq peaks.

Enrichment over input (%) was plotted in the graphs. Statistical analyses were performed using delta Ct values.

ATAC-seq

Briefly, fresh total brain tissue was harvested on ice and immediately dissected and minced into 1mm by 1mm portions with curved scissors prior to dounce homogenization (on ice) in hypotonic lysis buffer. Nuclei were isolated via density gradient centrifugation with optiprep density gradient medium (Sigma, D1556). Nuclei were collected from the 30-40% interface. Approximately 50,000 nuclei from murine tissues were used as input for ATAC-seq. ATAC-seq libraries for murine cells were generated as previously described (Buenrostro et al. 2013). Paired-end 2x75 bp sequencing was performed on an Illumina Nextseq 500 instrument. Reads were mapped to the mouse genome (mm9) using Bowtie2 with default paired-end settings (Langmead et al. 2009). Next, all non-nuclear reads, and improperly paired reads were discarded. Duplicated reads were then removed with picard MarkDuplicates. Peak calling was carried out with MACS2 (callpeak--nomodel--broad).

Generation of M10 and M16 iPS cells

To initiate reprogramming, primary human male dermal fibroblasts from a 10-year-old and a 16-year-old healthy donor (Coriell Institute, GM09503) were infected with non-integrating Sendai viruses expressing OCT4, SOX2, KLF4, and C-MYC (Thermo Fisher, CytoTune-iPS 2.0). Clonal human induced pluripotent cell (hiPSC) colonies were manually picked and expanded under feeder-free conditions using hESC-qualified Matrigel (Corning) and mTeSR1 medium (Stemcell Technologies). “M10c5” and “M16c4” have shown to have a normal male karyotype at passage 6 and subsequently used for genome editing.

Peak-2 and Peak-6 Deletions iPS cells

iPSCs were adapted to single-cell passaging using Accutase and StemFlex medium (Thermo Fisher Scientific) prior to nucleofection. gRNAs flanking the human CRE regions were designed and ordered from Synthego. On the day of nucleofection, iPSCs were pre-treated with 10 μM Y-27632 for 30 minutes. Cells were dissociated to single cell suspensions using Accutase™ and 300,000 cells were nucleofected with two ribonucleoprotein complexes (RNPs) each consisting of 200 pmol sgRNA (Synthego)

complexed to 40 pmol Cas9 protein (Aldevron, *Spy*Fi™ Cas9 Nuclease) using the P3 Primary Cell 4D-Nucleofector™ X Kit S (Lonza; Program CA-137). Nucleofected cells were seeded onto a Matrigel-coated plate in StemFlex medium supplemented with 10 μM Y-27632 for the first 24 hours. Colonies were manually picked and screened for homozygous deletion of the regulatory elements by genomic PCR.

gRNA sequences:

Target region	gRNA-L	gRNA-R
Peak-2	CTGGTGGGACAAAAATTGTG	GAGGGATTGAACCAGGACAG
Peak-6	AGCCCCAGAACAGCTCCCAG	TTTCCTCCAGTATATCAAGG

The following primers are used for genotyping:

Target region	Primer-F	Primer-R
Peak-2	TGTGGCACTTGTTCCATTTC	GAAAAGCAGCAGCAACAGAA
Peak-6	TTCACTGTAAGCGGAAGGAGA	CTGGGTTCAGGCAGGAGTAG

Lentivirus production

pLV-TetO-hNGN2-mCherry-BSD and pLV-TetO-hNGN2-eGFP-BSD plasmids were used for production of lentiviral vectors with inducible expression of the human NGN2(Crutcher et al. 2019). HEK-293FTs were transfected with jetPRIME transfection reagent (Polyplus-transfection, 114–15), following the manufacturer's recommended protocol, to initiate virus production. Efficient generation of pseudo-typed viral particles was ensured with the second-generation lentiviral packaging plasmids - psPAX2 (Addgene, #12260) and pMD2.g (Addgene, #12259), both gifts from Dr. Didier Trono. We collected lentivirus supernatant at 24- and 48-hours post-transfection and concentrated it overnight at 4°C according to the Lenti-X Concentrator (Takara, 631232) protocol.

Neuron differentiation

The iPSCs were plated in 12-well coated with Matrigel (150,000 cells / well), and first infected with lentivirus packaged with rtTA with hygromycin selection. Cells were then selected in 0.2 mg/mL hygromycin until uninfected control cells were absent. Then rtTA-positive lines were infected with lentivirus packaged with NGN2-EGFP with blasticidin selection (10-20 μg/mL). For biochemical experiments, cells were plated in 24-wells with 100,000 cells per well and induced into neurons using neural-induction medium (DMEM/F12:Neurobasal (1:1) supplemented with 1x B27, 1x N2 and 2nM Glutamax) with 2 mg/mL Doxycycline for four days, with fresh media changes of neural-induction medium + Doxycycline each day and neural-differentiation medium with selection (Neuobasal with 1x B27, 2 mM Glutamax, 20 ng/ml BDNF, 10 ng/ml GDNF, 10 ng/ml NT-3, 100 μM db-cAMP and 200 μM ascorbix acid, 10 mg/mL blasticidin) afterwards. The 50% media for the neurons was replaced with fresh neural-differentiation medium every 3-4 days. The neurons were harvested 60 days after differentiation.

Cresyl Violet Staining

Brains from 6-week-old mice were perfused with 4% formaldehyde and harvested. Sagittal slices were made with frozen tissues. Rehydrate slides in 100% ETOH through alcohols gradients (100%-95%-75%) to distilled water (2 mins each). Then stain CV 5 mins (Solution of 0.1 % Cresyl Violet: 0.1 g Cresyl Violet; 100 ml ddH2O; 0.3 ml glacial acetic acid; 0.0205 g sodium acetate. Adjust pH 3.5). Then dehydrate slides in alcohol gradients (75%-95%-100%-100%) to xylene (2 mins each). Finally, the slides are mounted using Cytoseal 60 (VWR). The images were scanned with Zeiss Axio Scan.Z1 at 20x.

Behavioral assays

Open field test

After habituation in the test room (150 lx, 60 dB white noise), mice were placed in the center of an open arena (40 × 40 × 30 cm), and their behavior was tracked by laser photobeam breaks for 30 min. General locomotor activity was automatically analyzed using AccuScan Fusion software (Omnitech) by counting

the number of times mice break the laser beams (activity counts). In addition, rearing activity, the time spent in the center of the arena, entries to the center and distance travelled were analyzed.

Elevated plus maze

After habituation in the test room (700 lx, 60 dB white noise), mice were placed in the center part of the maze facing one of the two open arms. Mouse behavior was video-tracked for 10 min, and the time mice spent in the open arms and the entries to the open arms, as well as the distance travelled in the open arms, were recorded and analyzed using ANY-maze system (Stoelting).

Three-Chamber test

The three-chamber apparatus consists of a clear Plexiglas box (24.75 × 16.75 × 8.75) with removable partitions that separate the box into three chambers. In both left and right chambers a cylindrical wire cup was placed with the open side down. Age and gender-matched C57Bl/6 mice were used as novel partners. Two days before the test, the novel partner mice were habituated to the wire cups (3 inches diameter by 4 inches in height) for 1h per day. After habituation in the test room (700 lx, 60 dB white noise), mice were placed in the central chamber and allowed to explore the 3 chambers for 10 min (habituation phase). Next, a novel partner mouse was placed into a wire cup in either the left or the right chamber. An inanimate object was placed as control in the wire cup of the opposite chamber. The location of the novel mouse was randomized between left and right chambers across subjects to control for side preference. The mouse tested was allowed to explore again for an additional 10 min. The time spent investigating the novel partner (defined by rearing, sniffing or pawing at the wire cup) and the time spent investigating the inanimate object were measured manually.

Tube test

The tube test was performed according to a modified protocol from the previous literature(Spencer et al. 2005; Wang et al. 2019). We used transparent plexiglass tubes with 30.5 cm length/3 cm inner diameter for male mice. Mice were habituated to walk through the tube ten sessions per day for two consecutive days before testing. On the day of testing, mice with different genotypes (same age and gender) were put at the end of the tube and released simultaneously. The mouse that completely retreated first from the tube

within the first 5 mins of the test was defined as the loser, and the other as the winner. In very rare cases, when no mice retreated within 5 mins, the tests were repeated. Each mouse was tested against up to four different mice of a different genotype with a 30 min inter-trial interval. The tubes were cleaned with 75% ethanol between trials. A two-tailed binomial test was used to determine the significance of test score between mice.

PPI

The acoustic prepulse inhibition task consists of presenting the animal with two closely paired sound pulses: a prepulse at +0 dB, +4 dB (74 dB), +8 dB (78 dB), +12 dB (82 dB) and over background followed 100 ms later by a pulse of 120 dB. The amount of startle the pulse induces in the animal is recorded using a startle chamber for mice which records activity for 65 ms after the pulse. The maximum amplitude recorded over the 65 ms is recorded and compared using an ANOVA (genotype) at each prepulse level.

Parallel Rod Test

Mice were habituated in the test room for 30 min. Each mouse was placed in a parallel rod chamber consisting of a plexiglass box with a floor of parallel-positioned rods and allowed to move freely for 10 min. Movement was recorded by a suspended digital camera, while footslips were recorded using ANY-maze software (Stoelting Co.). At the completion of the test, mice were removed to their original home cage. Total footslips were normalized to the distance traveled for data analysis.

Fear conditioning

A delayed fear conditioning protocol was employed to evaluate hippocampus-dependent contextual fear memory and hippocampus-independent cue fear memory. On day 0 animals were trained in a mouse fear conditioning chamber with a grid floor that can deliver an electric shock (Med Associates, Inc.). This enclosure was located in a sound-attenuating box that contained a digital camera, a loudspeaker and a house light. Each mouse was initially placed in the chamber and left undisturbed for 2 min, after which a tone (30 s, 5 kHz, 80 dB) coincided with a scrambled foot shock (2 s, 0.7 mA). The tone/foot-shock stimuli were repeated after 2 min. The mouse was then returned to its home cage. The context test was

assessed in 24 hours. The mice were placed in exactly the same environment and observed for 5 min. The cued fear test was assessed one hour after the context test. The mice were placed in a novel environment for 3 min, followed by a 3 min tone. Mouse behavior was recorded and scored automatically by Freeze Frame (Actimetrics). Freezing, defined as the absence of all movement except for respiration, was scored only if the animal was immobile for at least 1 s. The percentage of time spent freezing during the tests serves as an index of fear memory. Cued fear memory was the subtraction of freezing time between the tone phase and the no-tone phase.

Accelerating rotarod test

After habituation in the test room (700 lx, 60 dB white noise), motor coordination was measured using an accelerating rotarod apparatus (Ugo Basile). Mice were tested for four consecutive days, four trials each, with an interval of at least 30 min between trials to rest. Each trial lasted for a maximum of 10 min, and the rod accelerated from 4 to 40 r.p.m. in the first 5 min. The time that it took for each mouse to fall from the rod (latency to fall) was recorded.

Generation of the knockout mice of regulatory elements

Target region	sgRNA-L	sgRNA-R
Peak1	AATAGGGCAAAGCCAGAAGA	CAAAACTAAACCAAACTGGA
Peak2	ACATGCACAAACCCAAACAT	TGATAAGTGTGATTCATGAG
Peak3	AATGTACTTGTACACAACGC	CAGGTATACGAAAAAGACAA
Peak5	ACAGATTTCGAGCTGCAGAG	AGACAAATACTGACACAATG
Peak6	AAAAGGATATGGGATACTAG	GGTCCCCAATAGGGTCACCA

The following primers were used to distinguish the knockout mice:

Target region	Primer-F	Primer-R
Peak1	TGGGCTTGCCACATGACAA	AGGCAGGTGTCCTTGGATAT
Peak2	GGAGGAGGAATGGGATGAGG	ATTTCCTGTCTGTCCGGCTT

Peak3	TCCACCTAAAATGTCAATGGCA	TGATCAAAGGCAGGATAGAAGC
Peak5	GGGGTGGGACTCTTAGTGTT	ATGGACAAAGAAAATGCGGTC
Peak6	TCACTGAACCTAGAGCCCTC	CCCACGAACTCAAGAAGAACA

Gene expression analysis by qRT-PCR

The following primers were used in the qRT-PCR reactions:

Gene	Primer-F	Primer-R
MECP2	TATTTGATCAATCCCCAGGG	CTCCCTCTCCCAGTTACCGT
GAPDH	CGACCACTTTGTCAAGCTCA	TTACTCCTTGGAGGCCATGT
S16	AGGAGCGATTTGCTGGTGTGG	GCTACCAGGGCCTTTGAGATG
AW551984	CATAAGAGATCCAGTGGCAC	AGTTTAGGGTTGCAGACAC
Oprk1	CGATAGTCCTTGGAGGCACC	GGACTGGGATCACAAAGGCA
Crh	GGAGAAACTCAGAGCCCAAGTA	GTTAGGGGCGCTCTCTTCTCC
Tsc22d3	ACTGGATAACAGTGCCTCC	TCAGGTGGTTCTTCACGAG
Gdf11	TAAGCGCTACAAGGCCAAC	CAGGGATCTTGCCGTAGATAA
Sst	GCAGGAAAGGAGCTGCTGA	CTTGCTTGAGCGCTGTCTCT
Mc4r	GAGTCTTTGTTGTCTGCTGG	GGGCAAGAGATGTAGAACAG
Rcor2	ACCCGAAGTCGAACTAGTG	CTAGTTCATCACTGTCTTCTTTGTC
Bdnf	GCGGCAGATAAAAAGACTGC	TCAGTTGGCCTTTGGATACC
Fdf11	TAGCCTGATCCGACAGAAGC	GGCAGAACAGTTTGGTGACG

All primers are mouse-specific, except the *MECP2* primer is common to mouse and human, and *GAPDH* is human specific.

References

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213-1218.

Crutcher E, Pal R, Naini F, Zhang P, Laugsch M, Kim J, Bajic A, Schaaf CP. 2019. mTOR and autophagy pathways are dysregulated in murine and human models of Schaaf-Yang syndrome. *Sci Rep* **9**: 15935.

Ito-Ishida A, Yamalanchili HK, Shao Y, Baker SA, Heckman LD, Lavery LA, Kim JY, Lombardi LM, Sun Y, Liu Z et al. 2018. Genome-wide distribution of linker histone H1.0 is independent of MeCP2. *Nat Neurosci* **21**: 794-798.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Spencer CM, Alekseyenko O, Serysheva E, Yuva-Paylor LA, Paylor R. 2005. Altered anxiety-related and social behaviors in the Fmr1 knockout mouse model of fragile X syndrome. *Genes Brain Behav* **4**: 420-430.

Wang L, Pang K, Han K, Adamski CJ, Wang W, He L, Lai JK, Bondar VV, Duman JG, Richman R et al. 2019. An autism-linked missense mutation in SHANK3 reveals the modularity of Shank3 function. *Mol Psychiatry*.

regulated during human neurodevelopment by combinatorial action of RNA-binding proteins and miRNAs. *Cell Rep* **17:** 720–734. doi:10.1016/j.celrep.2016.09.049

Samaco RC, Fryer JD, Ren J, Fyffe S, Chao HT, Sun Y, Greer JJ, Zoghbi HY, Neul JL. 2008. A partial loss of function allele of methyl-CpG-binding protein 2 predicts a human neurodevelopmental syndrome. *Hum Mol Genet* **17:** 1718–1727. doi:10.1093/hmg/ddn062

Samaco RC, Mandel-Brehm C, McGraw CM, Shaw CA, McGill BE, Zoghbi HY. 2012. *Crh* and *Oprm1* mediate anxiety-related behavior and social approach in a mouse model of *MECP2* duplication syndrome. *Nat Genet* **44:** 206–211. doi:10.1038/ng.1066

Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, Peng M, Collins R, Grove J, Klei L, et al. 2020. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180:** 568–584.e23. doi:10.1016/j.cell.2019.12.036

Schanze I, Bunt J, Lim JWC, Schanze D, Dean RJ, Alders M, Blanchet P, Attié-Bitach T, Berland S, Boogert S, et al. 2018. NFIB haploinsufficiency Is associated with intellectual disability and macrocephaly. *Am J Hum Genet* **103:** 752–768. doi:10.1016/j.ajhg.2018.10.006

Schnabel F, Smogavec M, Funke R, Pauli S, Burfeind P, Bartels I. 2018. Down syndrome phenotype in a boy with a mosaic microduplication of chromosome 21q22. *Mol Cytogenet* **11:** 62. doi:10.1186/s13039-018-0410-4

Soldner F, Stelzer Y, Shivalila CS, Abraham BJ, Latourelle JC, Barrasa MI, Goldmann J, Myers RH, Young RA, Jaenisch R. 2016. Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. *Nature* **533:** 95–99. doi:10.1038/nature17939

Spencer CM, Alekseyenko O, Serysheva E, Yuva-Paylor LA, Paylor R. 2005. Altered anxiety-related and social behaviors in the *Fmr1* knockout mouse model of fragile X syndrome. *Genes Brain Behav* **4:** 420–430. doi:10.1111/j.1601-183X.2005.00123.x

Swanberg SE, Nagarajan RP, Peddada S, Yasui DH, LaSalle JM. 2009. Reciprocal co-regulation of EGR2 and MECP2 is disrupted in Rett syndrome and autism. *Hum Mol Genet* **18:** 525–534. doi:10.1093/hmg/ddn380

Sztainberg Y, Chen HM, Swann JW, Hao S, Tang B, Wu Z, Tang J, Wan YW, Liu Z, Rigo F, et al. 2015. Reversal of phenotypes in MECP2 duplication mice using genetic rescue or antisense oligonucleotides. *Nature* **528:** 123–126. doi:10.1038/nature16159

Telenti A, Pierce LC, Biggs WH, di Iulio J, Wong EH, Fabani MM, Kirkness EF, Moustafa A, Shah N, Xie C, et al. 2016. Deep sequencing of 10,000 human genomes. *Proc Natl Acad Sci* **113:** 11901–11906. doi:10.1073/pnas.1613365113

Van Esch H, Bauters M, Ignatius J, Jansen M, Raynaud M, Hollanders K, Lugtenberg D, Bienvenu T, Jensen LR, Gécz J, et al. 2005. Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am J Hum Genet* **77:** 442–453. doi:10.1086/444549

Yagasaki Y, Miyoshi G, Miyata M. 2018. Experience-dependent MeCP2 expression in the excitatory cells of mouse visual thalamus. *PLoS One* **13:** e0198268. doi:10.1371/journal.pone.0198268

Zhang Y, Pak C, Han Y, Ahlenius H, Zhang Z, Chanda S, Marro S, Patzke C, Acuna C, Covy J, et al. 2013. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78:** 785–798. doi:10.1016/j.neuron.2013.05.029