ARTICLES
https://doi.org/10.1038/s41593-018-0270-6
nature neuroscience

Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing

William Renthal[1,2], Lisa D. Boxer[1,2], Sinisa Hrvatin[1], Emmy Li[1], Andrew Silberfeld[1], M. Aurel Nagy[1], Eric C. Griffith[1], Thomas Vierbuchen[1] and Michael E. Greenberg[1*]

In females with X-linked genetic disorders, wild-type and mutant cells coexist within brain tissue because of X-chromosome inactivation, posing challenges for interpreting the effects of X-linked mutant alleles on gene expression. We present a single-nucleus RNA sequencing approach that resolves mosaicism by using single-nucleotide polymorphisms in genes expressed in cis with the X-linked mutation to determine which nuclei express the mutant allele even when the mutant gene is not detected. This approach enables gene expression comparisons between mutant and wild-type cells within the same individual, eliminating variability introduced by comparisons to controls with different genetic backgrounds. We apply this approach to mosaic female mouse models and humans with Rett syndrome, an X-linked neurodevelopmental disorder caused by mutations in the gene encoding the methyl-DNA-binding protein MECP2, and observe that cell-type-specific DNA methylation predicts the degree of gene upregulation in MECP2-mutant neurons. This approach can be broadly applied to study gene expression in mosaic X-linked disorders.

The diversity of cell types in the brain has largely precluded the characterization of cell-type-specific features of neurodevelopmental diseases. For X-linked neurodevelopmental disorders, this cellular heterogeneity poses an additional challenge in females, where random X-chromosome inactivation results in a mixture of wild-type and mutant cells within the brain of the same individual[1].

These challenges are exemplified by Rett syndrome, an X-linked neurodevelopmental disorder predominantly affecting girls and characterized by speech delay, repetitive hand movements, seizures, and autism-like behavior[2]. Rett syndrome is caused by mutations in the *MECP2* gene on the X chromosome, and disease severity is thought to be correlated with the fraction of brain cells expressing the mutant allele after X-inactivation[1,3]. In individuals with Rett syndrome, neural circuits will thus consist of wild-type and mutant cells, raising the possibility that both cell-autonomous and non-cell-autonomous effects contribute to the pathophysiology of Rett syndrome at the cellular and circuit levels. Better understanding of these effects of the *MECP2* mutation will be critical for developing targeted therapeutics, but it has been difficult to distinguish gene expression in *MECP2*-mutant neurons from that of normal neurons within the same brain.

MECP2 encodes a nuclear protein that is enriched in neurons, binds to methylated cytosines broadly across the genome, and has been suggested to act as a transcriptional repressor by recruiting co-repressor complexes (for example, NCOR) to sites of methylated DNA[2,4–7]. Consistent with this finding, we have found, in male mice where all cells express a single allele of *Mecp2*, that when MeCP2 function is disrupted, genes with the highest level of gene-body DNA methylation and MeCP2-binding in wild-type neurons exhibit the largest degree of upregulation in gene expression in *Mecp2*-mutant neurons[8–10]. However, numerous reports have proposed additional functions of MeCP2 at specific loci, including the regulation of mRNA splicing, transcriptional activation, and chromatin structure[2,11–14]. At present, it is not clear whether these effects are due to direct or indirect actions of MeCP2. Notably, since these previous studies of MeCP2 function have mostly focused on male hemizygous animals in which all cells lack functional MeCP2, the extent to which the effects observed in male mice accurately reflect the effects of MeCP2 loss in the mosaic brains of female heterozygous mice or humans with Rett syndrome remains unclear.

The recent development of high-throughput single-cell RNA sequencing (scRNA-seq) technologies has revolutionized gene expression analysis of complex tissues and enabled the characterization of cell-type-specific transcriptional programs in various brain regions in mice and humans[15–17]. While these advances have permitted the identification and characterization of unique cell types within complex tissues, until now it has not been possible, even with scRNA-seq, to reliably distinguish between cells that express the wild-type or mutant allele in mosaic females with X-linked disorders because the sequencing reads generated from single cells rarely include the disease-causing mutations. Here we describe an approach, single-cell single-nucleotide polymorphism (SNP) sequencing (SNP-seq), that reliably determines whether individual cells derived from mosaic murine and postmortem human brain express the wild-type or mutant X-chromosome allele, enabling gene expression profiles of wild-type and mutant cells from the same individual to be distinguished from each other. Using this approach, we found that in the brains of female heterozygous mouse models and humans with Rett syndrome, MECP2 selectively and cell-autonomously repressed the expression of highly methylated genes in a cell-type-specific manner in wild-type but not *MECP2*-mutant neurons. The methods and analyses outlined here for Rett syndrome can be broadly applied to the characterization of gene expression patterns in additional mosaic X-linked disorders such as Fragile X syndrome, CDKL5 deficiency disorder, X-linked intellectual disability, and multiple X-linked genetic causes of autism.

Results

Single-cell SNP sequencing in mouse models of Rett syndrome. Droplet-based high-throughput scRNA-seq methods employ

[1]Department of Neurobiology, Harvard Medical School, Boston, MA, USA. [2]These authors contributed equally: William Renthal, Lisa D. Boxer.
*e-mail: michael_greenberg@hms.harvard.edu

poly-A transcript selection in which the majority of sequence information is restricted to the distal 3′ end of genes, a region that often does not include the disease-causing mutations under investigation[15,16]. Moreover, these methods typically sample a fraction of the total transcripts per cell, which further limits the ability to reliably detect the expressed mutant allele even when the variant of interest lies within the 3′ sequenced region. For the same reason, a failure to detect expression of a given gene in mutant cells is not a reliable way to discriminate between mutant and wild-type cells. However, we reasoned that SNPs that differ between the two X chromosomes and are within genes expressed in cis with the mutant allele might provide a reliable way to determine whether a given cell expresses the mutant or wild-type allele, hereafter defined as the cell's transcriptotype.

To determine the utility of this approach, we first attempted to distinguish between cells expressing wild-type (WT) or mutant alleles in female *Mecp2*$^{+/-}$ mice. These mice were generated by deleting the majority of the *Mecp2* gene (exons 3 and 4), and they recapitulate key features of Rett syndrome[18]. The absence of *Mecp2* expression is not a reliable indicator of a mutant cell, however, both because expression of the *Mecp2* 3′ untranslated region is still detectable at low levels in mutant cells and because scRNA-seq only captures a fraction of genes per cell. Thus, we searched expressed genes for SNPs that were maintained in cis with the mutant *Mecp2* allele during the process of backcrossing the 129P2/OlaHsd strain of mice in which the *Mecp2*-mutant mice were generated. Despite extensive backcrossing (>38 generations at Jackson Labs) of the *Mecp2*-mutant mice with the C57BL/6J strain, we identified four 129P2/OlaHsd-specific SNPs in cis with the *Mecp2*-mutant allele that were present in the expressed 3′ untranslated regions of two genes that are closely linked to *Mecp2* and well sampled in the scRNA-seq datasets (Supplementary Fig. 1).

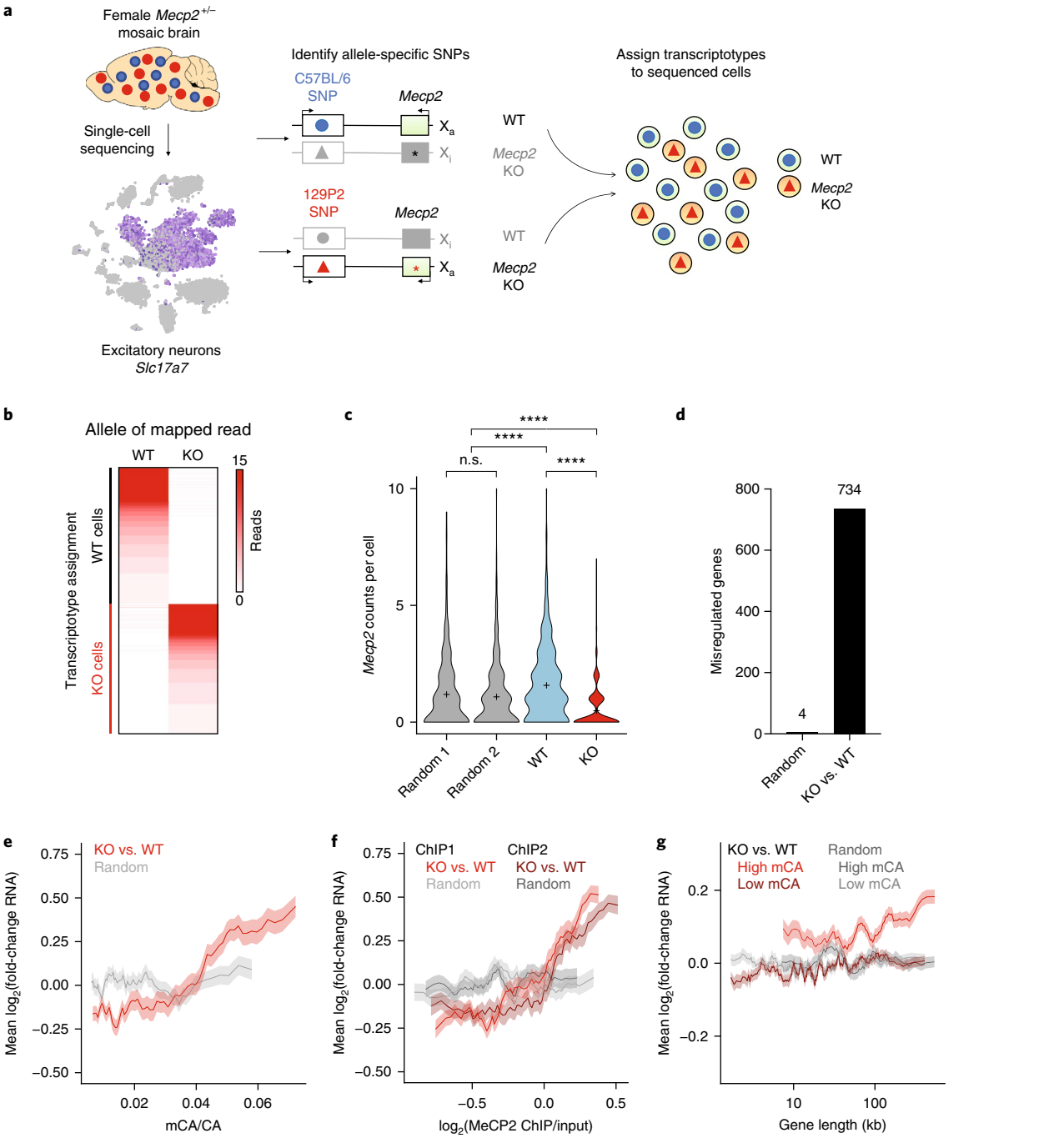
We performed scRNA-seq on visual cortex from 5 adult (12- to 20-week-old) female *Mecp2*$^{+/-}$ mice and obtained 12,451 cells that passed initial quality-control tests. Consistent with data from WT cortex[19], cells from *Mecp2*$^{+/-}$ cortex were clustered into eight major cell types using the Seurat single-cell analysis pipeline[20] (Supplementary Fig. 2a). We focused on excitatory neurons because they have previously been directly implicated in Rett syndrome pathophysiology[21,22] and are the most abundant cell type in our dataset (Fig. 1a). Sequencing reads encompassing the identified strain-specific SNPs allowed 1,289 of 5,761 excitatory neurons to be identified as expressing either the WT or mutant *Mecp2* allele (Fig. 1b and Supplementary Fig. 2b). In support of the SNP-based transcriptotype classification, the resulting *Mecp2*-mutant population of cells exhibited significantly reduced levels of the *Mecp2* transcript relative to WT cells or to groups of excitatory neurons with randomly assigned transcriptotypes (Fig. 1c). Gene expression analysis of the transcriptotyped mutant versus WT cells identified 734 differentially expressed genes (366 that were upregulated, 368 that were downregulated; false-discovery rate < 0.1; Supplementary Table 1). By contrast, only four significantly misregulated genes were identified when cell populations with randomly assigned transcriptotypes were compared (Fig. 1d). These data indicate that we can successfully study gene expression in WT and mutant cells by single-cell SNP-seq, making it possible to address whether MeCP2 function in mosaic females is accurately modeled in male hemizygous mice in which all cells express the mutant form of the protein.

Previous reports in male mice indicate that gene bodies of MeCP2-repressed genes are highly methylated, have increased levels of MeCP2-binding, and tend to be long compared to genes that are not repressed by MeCP2[8,10,23–25]. These previous observations of MeCP2 dysfunction provided a molecular signature for assessing the ability of scRNA-seq data to detect relevant gene expression changes in mosaic tissue. Consistent with previous observations, we found that in mosaic female mice the degree of gene upregulation in *Mecp2*-mutant compared to WT excitatory neurons directly correlated with gene body DNA methylation (Pearson's $r = 0.38$), as well as with the length of highly methylated genes (Pearson's $r = 0.10$; Fig. 1e,g). In the brains of mosaic female *Mecp2*$^{+/-}$ mice, we also observed that the degree of gene upregulation in mutant-expressing excitatory neurons directly correlated with increasing levels of gene-body MeCP2-binding in excitatory neurons (chromatin immunoprecipitation (ChIP)1, Pearson's $r = 0.41$; ChIP2, Pearson's $r = 0.31$; Fig. 1f).

Fig. 1 | Single-cell SNP sequencing in a female mouse model of Rett syndrome. a, Flow chart of single-cell SNP sequencing pipeline. scRNA-seq was performed on visual cortex from 5 female *Mecp2*$^{+/-}$ mice followed by graph clustering to identify the group of excitatory neurons (*Slc17a7*$^{+}$). Allele-specific SNPs in genes expressed in cis with the *Mecp2* mutation were identified by variant-calling and then used to assign the corresponding transcriptotype to the individually sequenced cells. Xa, active X chromosome; Xi, inactive X chromosome. **b**, Heatmap of reads per analyzed cell (rows of the heatmap) that map to WT- or knockout (KO)-specific SNPs (columns of the heatmap). **c**, Violin plots of *Mecp2* mRNA counts per cell in cells that were grouped based on their SNP-identified transcriptotype (WT, *Mecp2*$^{+/-}$ WT excitatory neurons; KO, *Mecp2*$^{+/-}$ mutant excitatory neurons; tails represent minima and maxima of data) or by randomly assigned transcriptotypes (Random 1, Random 2). *Mecp2* expression was significantly higher in the WT cells (sampled $n = 593$) compared to KO cells ($n = 593$; Kruskal–Wallis test, $H = 210$, $^{****}P < 0.0001$, + indicates mean) and compared to the populations with randomly assigned transcriptotypes (Random 1, $n = 593$, Random 2, $n = 593$; $^{****}P < 0.0001$). The groups with randomly assigned transcriptotypes had similar levels of *Mecp2* expression ($P > 0.9999$). For the transcriptotyped excitatory neurons, we obtained an average of 7,634 transcripts per cell, representing 3,879 distinct genes. **d**, The number of significantly misregulated genes (false-discovery rate (FDR) < 0.1, monocle2) when comparing gene expression differences between groups of mutant and WT excitatory neurons (KO vs. WT, 734 genes) or two groups of randomly assigned transcriptotypes (Random, 4 genes). **e**, The mean fold-changes of the misregulated genes described in **d** (KO vs. WT, Random) are displayed as a function of excitatory neuron gene body DNA methylation (mCA/CA; KO vs. WT, Pearson's $r = 0.38$; Random, Pearson's $r = 0.04$). The correlation between MeCP2-dependent gene expression and mCA/CA was significantly greater in KO vs. WT than Random (permutation test, $P < 0.001$). **f**, The fold-change of genes in **d** (KO vs. WT, Random) binned by gene-body MeCP2 ChIP enrichment over input. The correlations between MeCP2-dependent gene expression and two MeCP2 ChIP replicates from purified cortical excitatory neurons (ChIP1, Pearson's $r = 0.41$; ChIP2, Pearson's $r = 0.31$) are significantly greater than the correlations observed in the Random controls (Random ChIP 1, Pearson's $r = 0.06$; Random ChIP 2, Pearson's $r = 0.04$; permutation test, $P < 0.001$). **g**, Mean fold-change in gene expression of mutant excitatory neurons (KO) compared to WT excitatory neurons (WT) from *Mecp2*$^{+/-}$ mice, with genes separated into groups of highly methylated genes (normalized expression > 0.1, high mCA, top 25%) or lowly methylated genes (normalized expression > 0.1, low mCA, bottom 66%) and binned by their gene length. The use of all genes with normalized expression > 0.1 provided sufficient gene numbers to separate into groups of high and low mCA. MeCP2-dependent gene expression and gene length were significantly more correlated in KO vs. WT than Random for high mCA genes (KO vs. WT, Pearson's r = 0.10, Random, Pearson's r = 0.00, permutation test $P < 0.001$). The correlations between MeCP2-dependent gene expression and gene length were not statistically different between KO vs. WT and Random for low mCA genes (KO vs. WT, Pearson's r = 0.04, Random, Pearson's r = 0.02, permutation test $P = 0.23$). In **e**–**g**, lines represent mean fold-change in expression for genes binned according to gene length (250 bins, 25-gene step), methylation (100 bins, 10-gene step), or MeCP2 enrichment (100 bins, 10-gene step); ribbon displays s.e.m. of each bin.

MeCP2 binding was characterized by ChIP of MeCP2 in CaMKIIα -positive excitatory neurons isolated from WT male mice using INTACT, a method in which genetically tagged nuclei can be immunopurified[26]. These findings suggest that upregulation of highly methylated genes is a cell-autonomous signature of MeCP2 dysfunction, consistent with the observation that Rett syndrome severity correlates with the number of *Mecp2*-mutant cells[1]. Notably, the differentially expressed genes between mutant and WT excitatory neurons within mosaic female heterozygous mice significantly overlapped with the misregulated genes we identified when comparing excitatory neurons from male *Mecp2*-mutant mice and their WT controls (hypergeometric test, $P = 7.2 \times 10^{-14}$; Supplementary Fig. 3 and Supplementary Tables 2–5). Thus, by resolving mosaicism with single-cell SNP-seq in a female mouse model of Rett syndrome and comparing the patterns of cell-type-specific gene misregulation to those of male mouse models (Supplementary Fig. 4), we have identified a reproducible set of cell-autonomous MeCP2-dependent genes in excitatory neurons.

While our data indicate that the relationships between MeCP2-dependent gene expression and gene-body DNA methylation, MeCP2 occupancy, and gene length are cell-autonomous, it has been difficult to determine whether there are also non-cell-autonomous effects of *Mecp2*-mutant cells on WT cells within the same tissue. Previous attempts to identify such effects have relied on tagged forms of MeCP2 that were not expressed at normal levels[27]. We overcame these challenges by using scRNA-seq to compare WT excitatory neurons (671 cells) from 5 female *Mecp2*$^{+/-}$ mice with WT excitatory neurons from 4 female *Mecp2*$^{+/+}$ control mice (671 sampled



cells). We observed 233 differentially expressed genes (false-discovery rate < 0.1) between these conditions, many of which involve key neuronal processes, such as neuronal activity-dependent gene expression and neurotrophin signaling (Supplementary Table 6). Notably, these differentially expressed genes between WT cells from *Mecp2*$^{+/-}$ and *Mecp2*$^{+/+}$ mice do not appear to be directly repressed by MeCP2 (for example, their degree of gene misregulation does not correlate with the level of gene body DNA methylation (permutation test, $P = 0.55$) or gene length (permutation test, $P = 0.73$); Supplementary Fig. 5). These data suggest that gene expression abnormalities are present in WT cells from *Mecp2*$^{+/-}$ mice and are likely due to indirect effects of neighboring *Mecp2*-mutant cells. This non-cell-autonomous misregulation of gene expression in WT neurons of mosaic individuals with Rett syndrome could, in principle, contribute to disease pathophysiology.

Single-nucleus SNP sequencing of human Rett brain tissue. Given the successful implementation of single-cell SNP-seq in rodent models of Rett syndrome, we reasoned that this method could also be used to characterize MECP2-dependent gene expression changes in postmortem human Rett brain tissue. This approach is potentially powerful because mutant and WT cells of the same age and genetic background can be compared directly in a single experiment, largely eliminating the transcriptional consequences of genetic variation that are introduced when comparing donor samples to unrelated age-matched controls (an especially important advantage in the study of Rett syndrome, where the differences in gene expression are expected to be small in magnitude[2]).

We performed single-nucleus RNA sequencing on occipital cortex from 3 postmortem females with Rett syndrome, each harboring the second-most-common nonsense mutation (c.763 C > T) in a single *MECP2* allele that generates the R255X truncated gene product lacking the MECP2 transcriptional repressor domain (Supplementary Fig. 6). We isolated nuclei for these experiments because nuclei are more reliably extracted than entire cells from postmortem tissue samples and can provide sufficient gene expression information for cell-type classification and analysis[28]. We successfully sequenced a total of 43,558 nuclei, with 30,293 nuclei passing the minimum required threshold of 500 uniquely expressed genes. In line with previous single-cell and single-nucleus RNA-seq experiments[15,16,19], the nuclei analyzed had an average of 2,800 transcripts per nucleus from 1,671 unique genes. Using Seurat[20] and known excitatory neuron and interneuron marker genes[19], the nuclei clustered into a large excitatory neuron population (18,545 nuclei) and multiple distinct interneuron populations (5,952 nuclei total; Fig. 2a). The heterogeneity of cells in the interneuron cluster prompted us to further subdivide this population into their known functional classes by the expression of specific marker genes (for example, *VIP*, *PVALB*, *SST*, or *CCK*; Fig. 2a and Supplementary Fig. 7).

Once each nucleus was assigned to its respective cell type cluster, we next turned to identifying its transcriptotype. Because there were no sequencing reads that included the R255 position of *MECP2*, we reasoned that the large number of SNPs that differ between an individual's two X chromosomes might allow us to identify allele-specific SNPs that are in cis with the mutant *MECP2* locus and therefore expressed only in *MECP2*-mutant neurons. To identify the transcriptotype-specific SNPs in each Rett donor, we took advantage of an MECP2-specific antibody that was raised against a region of the C terminus truncated by the R255X mutation. We used this antibody to separate high-staining (MECP2^{high}) and low-staining (MECP2^{low}) nuclei by fluorescence-activated sorting (Fig. 2b). Sanger sequencing of isolated cDNA from the two populations confirmed that the MECP2^{high} population expressed WT *MECP2* and that the MECP2^{low} population expressed the R255X mutant *MECP2*.

Having isolated the two populations from each donor, we next performed total RNA sequencing on both populations and identified 69–75 allele-specific SNPs that were uniquely expressed in MECP2^{high} nuclei (Supplementary Fig. 8 and see Methods). Expression of these allele- and transcriptotype-specific SNPs was then queried in the corresponding single-nucleus RNA-seq dataset from the same donor sample and used to assign the corresponding WT or R255X *MECP2* transcriptotypes (Fig. 2c). Using the allele-specific SNPs identified from each Rett donor, we could assign transcriptotypes to 16,627 nuclei, or 55% of the nuclei assayed (Fig. 2d); the remaining 45% of nuclei were excluded from further analysis. The ratio of WT to mutant nuclei was approximately even across the three donor samples (donor 1 = 49% WT, 51% R255X; donor 2 = 51% WT, 49% R255X; donor 3 = 42% WT, 58% R255X), which suggests that there was not substantial skewing of X-chromosome inactivation and that Rett syndrome in the three donors is likely due to the loss of MECP2 function in approximately 50% of brain cells.

For subsequent analyses, we focused on the excitatory neuron population (*SLC17A7*-expressing, 18,545 cells) and on the most abundant subtype of interneurons in our datasets (*VIP*-expressing, 1,839 cells; Fig. 2a). Notably, the neuronal subtype clusters were similar between WT and mutant cells (Supplementary Fig. 9a), enabling direct comparison of gene expression between WT and mutant cells of the same neuronal subtype. To maximize the number of nuclei and statistical power for cell-type-specific gene expression comparisons, we combined nuclei of the same neuronal subtype and transcriptotype from the three Rett donors. We identified significant gene expression differences between mutant and WT excitatory neurons (3,158 genes, Supplementary Table 7) and VIP interneurons (237 genes, Supplementary Table 8; Fig. 2e). Notably, these findings were dependent on proper transcriptotype assignment, as gene expression analysis between populations of cells that were randomly assigned transcriptotypes consistently recovered ≤10 differentially expressed genes (Fig. 2e). It should be noted that the difference in numbers of significantly misregulated genes between excitatory neurons and VIP interneurons is largely attributable to the greater number of excitatory nuclei sampled with higher transcript coverage, because the numbers of misregulated genes were similar in excitatory and VIP interneurons when equal numbers of nuclei and transcripts were sampled for both cell types (Supplementary Fig. 9b).

Cell-type-specific DNA methylation patterns predict gene misregulation in Rett syndrome. These new human datasets provided the opportunity to determine whether features described in mouse models regarding MECP2-dependent gene expression are also observed in neurons from human individuals with Rett syndrome. It is not known, for example, whether in fact MECP2 in human neurons represses highly methylated long genes in a neuronal-subtype-specific manner, as has been observed in mice[8,10,25]. In mice, DNA methylation in both the CG and CA dinucleotide contexts recruits MeCP2 binding and contributes to MeCP2-dependent gene repression[8,29]. While both CG and non-CG methylation (mCH, which consists of CA, CT, and CC methylation) display cell-type-specific patterns, mCH is more divergent across neuronal cell types[26,30] and, in mice, contributes to cell-type-specific MeCP2-dependent gene repression[31]. To determine whether cell-type-specific patterns of mCH predict the degree of MECP2-dependent gene repression in human females with Rett syndrome, we compared the set of genes that are differentially expressed in human female *MECP2*-mutant-expressing and WT-expressing nuclei with recently published human single-cell methylation data from cerebral cortex[32]. We found that, in humans, the degree of gene misregulation in *MECP2*-mutant compared to WT excitatory neurons and VIP interneurons was directly correlated with the level of gene body mCH in neurons of the respective subtype (excitatory neurons, Pearson's

$r = 0.22$; VIP interneurons, Pearson's $r = 0.18$; Fig. 3a,e). These correlations were dependent on the correct assignment of transcriptotype, as gene expression differences between groups of randomly assigned transcriptotypes did not correlate with gene body mCH for either excitatory neurons (Pearson's $r = -0.01$) or VIP interneurons (Pearson's $r = -0.05$). The relationship between neuronal-subtype-specific mCH and MECP2-dependent gene expression was highly reproducible and could be observed in each of the three donor samples by directly comparing *MECP2*-mutant and WT neurons from the same individual (Supplementary Fig. 10). The direct correlation between MECP2-dependent gene repression and gene-body mCH for each neuronal subtype depended on its subtype-specific DNA methylation patterns, as MECP2-dependent gene repression in excitatory neurons did not correlate with the extent of gene-body mCH from VIP interneurons (Pearson's $r = -0.01$; Fig. 3b), and MECP2-dependent gene repression in VIP interneurons poorly correlated with mCH from excitatory neurons (Pearson's $r = 0.05$; Fig. 3d). Of note, the direct correlation between MECP2-dependent gene repression and DNA methylation was also observed in the CG dinucleotide context in both excitatory neurons and VIP interneurons (Supplementary Fig. 11).

As described above, for highly methylated genes, gene length predicts the degree of gene upregulation in *Mecp2*-mutant mice compared with their WT counterparts[10]. Consistent with this observation, we found that in humans with Rett syndrome the level of gene-body methylation together with gene length predicted the degree of gene upregulation in both MECP2-mutant excitatory and VIP interneuronal nuclei (Fig. 3c,f). We further found that in human females, as in mice, gene length did not positively correlate with MECP2-dependent gene repression for lowly methylated genes, underscoring the importance of accounting for DNA methylation in the analysis of MECP2-dependent gene regulation in humans[8,10]. These findings in human females with Rett syndrome are consistent with our findings in male and female *Mecp2*-mutant mouse models and indicate that MeCP2 acts through an evolutionarily conserved, cell-autonomous mechanism to preferentially repress the expression of highly methylated long genes.

The large number of excitatory neuronal nuclei sequenced from each individual provided sufficient power to study gene expression differences between mutant and WT nuclei of this neuronal subtype within the same individual's brain (Supplementary Tables 9–14), thus eliminating much of the genetic and environmental heterogeneity inherent to previous studies of MECP2-dependent gene expression[33,34]. We were thus able to identify genes that were consistently misregulated in mutant excitatory neurons across all three Rett syndrome donors. This analysis demonstrated a highly significant overlap in affected genes across the three Rett donor samples, identifying 537 genes that were consistently upregulated in *MECP2*-mutant excitatory neurons compared to WT neurons and 395 genes that were reproducibly downregulated (Fig. 4a, Supplementary Fig. 12, and Supplementary Table 15). The upregulated genes had significantly higher levels of gene body methylation than the downregulated genes (Fig. 4b). Genes that control metabolism or regulate neuronal processes such as ion transport or nervous system development were significantly enriched in the set of upregulated or downregulated genes (Fig. 4c,d), and misregulation of these genes may contribute to the metabolic and neuronal deficits observed in Rett syndrome[35]. The ability of single-nucleus SNP-seq to reliably transcriptotype and reproducibly identify gene expression changes between mutant and WT cells within the same individual largely overcomes the previous reliance on age-matched controls for molecular characterization of mosaic X-linked disorders, and will substantially improve our ability to distinguish gene expression differences that are due directly to the mutation under investigation rather than to unrelated genomic variation between cases and controls.

We next sought to identify genes that are controlled by MECP2 in both humans and mice, reasoning that despite the major species differences, the evolutionarily conserved MECP2 targets might provide an opportunity to investigate MECP2 function in mouse models that are relevant to human pathophysiology. To this end, we identified the upregulated or downregulated genes in excitatory neurons across all three Rett syndrome donor samples (537 and 395 genes, respectively) and asked which of these were also significantly misregulated in female *Mecp2*$^{+/-}$ excitatory neurons from mice. We identified 58 evolutionarily conserved genes that were upregulated and 84 genes that were downregulated in MECP2-mutant compared to WT excitatory neurons in both mouse and human (Fig. 4e, Supplementary Fig. 13, and Supplementary Tables 16 and 17). These evolutionarily conserved MECP2-regulated genes represented high-confidence MECP2 targets in excitatory neurons because of their reproducibility across multiple datasets. However, we stress that deeper sequencing would provide greater statistical power and the ability to identify many additional evolutionarily conserved MECP2 targets. We note that the high-confidence evolutionarily conserved genes identified here that were upregulated in *MECP2*-mutant excitatory neurons had significantly higher levels of gene-body DNA methylation than the set of genes that were downregulated in *MECP2*-mutant neurons (Fig. 4f), suggesting that the upregulated gene set may be enriched for direct MECP2 targets. However, it seems likely that the misregulation of both MECP2-repressed and MECP2-activated genes contribute to Rett syndrome pathophysiology, as 25% of the MECP2-repressed genes (enrichment $P = 1.0 \times 10^{-6}$, hypergeometric test) and 13% of the MECP2-activated genes (enrichment $P = 0.02$, hypergeometric test) have been previously shown to be mutated in intellectual disability or autism (see Methods). Many of the MECP2-repressed genes (for example, *AUTS2* and *RBFOX1*) encode transcriptional regulators that are known to control neuronal gene expression[36–38]. The MECP2-repressed genes that encode neuronal ion channels such as *GABRA1* and *SCN1B* are known to cause epilepsy when mutated[39,40], and thus they could contribute to this comorbidity in individuals with Rett syndrome. The evolutionarily conserved genes that were downregulated in *MECP2*-mutant neurons included the neurotrophin *BDNF* and the presynaptic adhesion molecule *NRXN2*, both of which have also been shown to contribute to neurological disorders when mutated[41,42]. Given that the selective disruption of *Mecp2* in excitatory neurons is sufficient to cause Rett-like phenotypes in mice[22], further investigation of evolutionarily conserved MECP2-regulated genes in this cell type could both yield new mechanistic insight into MECP2 function and help characterize the role of these genes in specific aspects of Rett syndrome pathophysiology.

Discussion

Here we present a new experimental approach that leverages the power of single-cell or single-nucleus RNA sequencing and individual genetic variation to simultaneously characterize cell-type-specific gene expression and allele-specific X-chromosome activation status in individual cells within mosaic mouse and human brains. This approach has broad applicability for studying gene expression abnormalities in X-linked neurodevelopmental disorders such as Rett syndrome, Fragile X syndrome, CDKL5 deficiency disorder, X-linked intellectual disability, and multiple X-linked genetic causes of autism (for example, *NLGN3*, *NLGN4*, *SLC6A8*, *PLXNA3*, *DDX3X*, *WDR45*, and *CASK*) in females where mosaicism between WT and mutant cells has hindered previous analyses. This method can be easily adapted (see Methods) to female mouse models of X-linked disorders generated in mixed genetic backgrounds by using strain-specific SNPs to identify the cells expressing the mutant allele. Moreover, this approach is particularly useful for studying mosaic disorders in human samples because the wealth of natural genetic variation across individuals

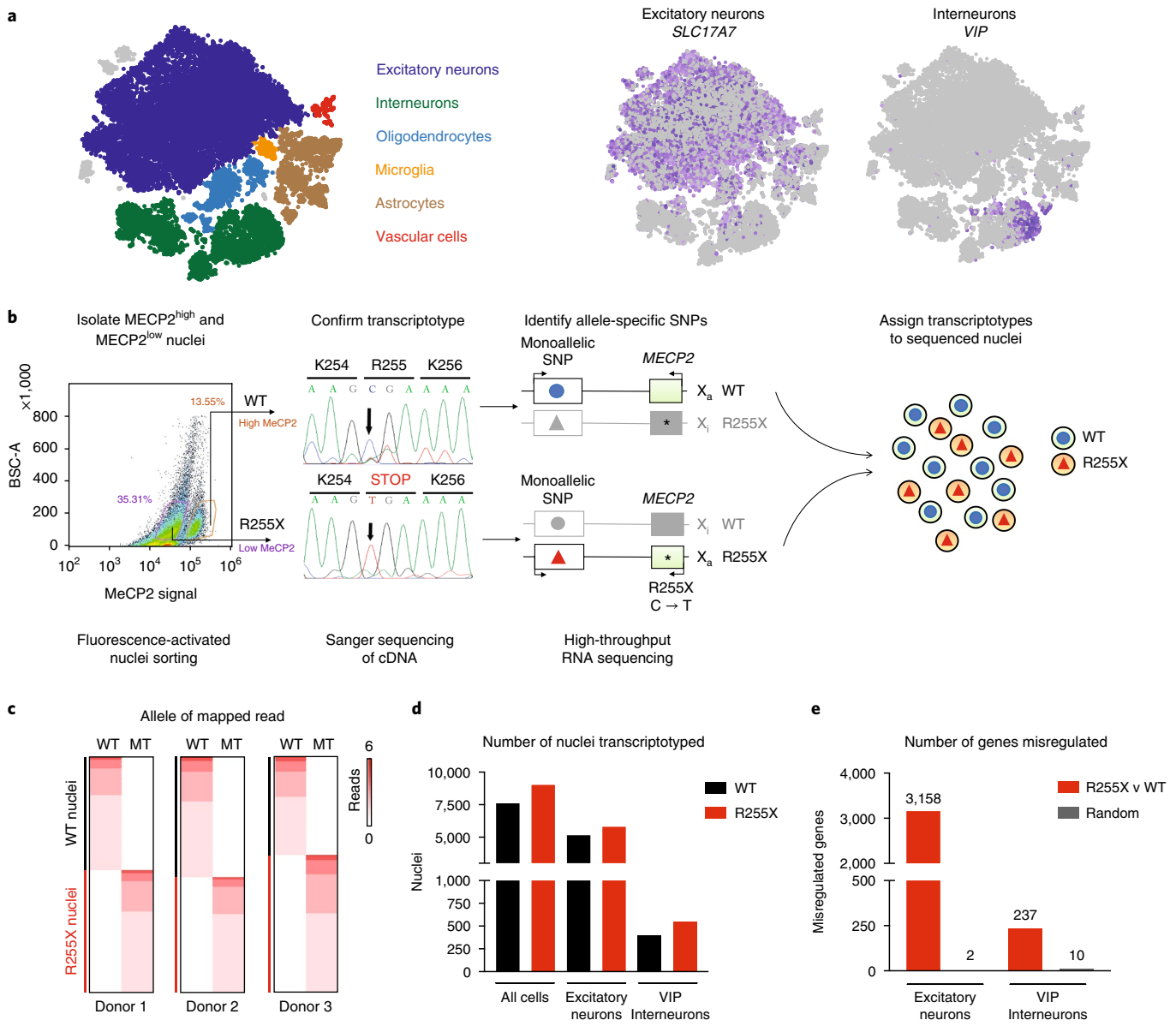


Fig. 2 | Single-nucleus SNP sequencing of human Rett brain tissue. a, Single-nucleus RNA sequencing of occipital cortex from three females with Rett syndrome. Brain cell types were identified by graph-clustering nuclei combined from the three Rett donors. **b**, Flowchart for identifying and assigning allele-specific SNPs for each Rett donor. Single-nuclei suspensions from each Rett donor were sorted based on their level of immunoreactivity to a C-terminal MeCP2 antibody (MECP2^{high} and MECP2^{low}). BSC-A, backscatter. The weak staining observed in MECP2^{low} nuclei represents background immunofluorescence. cDNA from MECP2^{high} and MECP2^{low} nuclei was Sanger sequenced to confirm that the sorted populations expressed the expected *MECP2* allele. Deep high-throughput RNA sequencing of these populations followed by variant calling identified the allele-specific SNPs used to assign transcriptotypes to each nucleus from the single-nucleus RNA sequencing dataset shown in **a**. Xa, active X chromosome; Xi, inactive X chromosome. **c**, Heatmap of reads per cell (rows of the heatmap) that map to WT- or *MECP2*-mutant (MT)-specific SNPs (columns of the heatmap) for each of the three donors. **d**, The number of total nuclei, excitatory neuronal nuclei, and VIP interneuronal nuclei that could be transcriptotyped from the single-nucleus RNA sequencing dataset of Rett donors. **e**, The number of significantly misregulated genes (FDR < 0.01, monocle2, R255X vs. WT; 3,158 genes in excitatory neurons, 237 genes in VIP interneurons) identified when comparing gene expression differences between groups of mutant and WT neurons, or two groups of neurons with randomly assigned transcriptotypes (Random; 2 genes in excitatory neurons, 10 genes in VIP interneurons). The difference in number of misregulated genes between excitatory and inhibitory neurons is largely explained by the number of cells analyzed (Supplementary Fig. 8b). The numbers of excitatory neuronal nuclei and VIP interneuronal nuclei used for differential expression analysis are shown in **d**.

provides many opportunities to identify allele-specific SNPs that are expressed from the same X-chromosome as the mutant allele under investigation[43,44]. Indeed, SNPs have been recently used in conjunction with scRNA-seq data to determine the sample identity of individual cells within a pool of human samples[44] and to study genes that escape X-chromosome inactivation[45].

In addition to validating the single-cell SNP-seq approach, our study provides further insight into important aspects of Rett syndrome pathophysiology and the consequences of MECP2 dysfunction. The inherent X-linked mosaicism in females with Rett syndrome has hampered prior efforts to determine whether genes that are differentially expressed in Rett and age-matched controls are

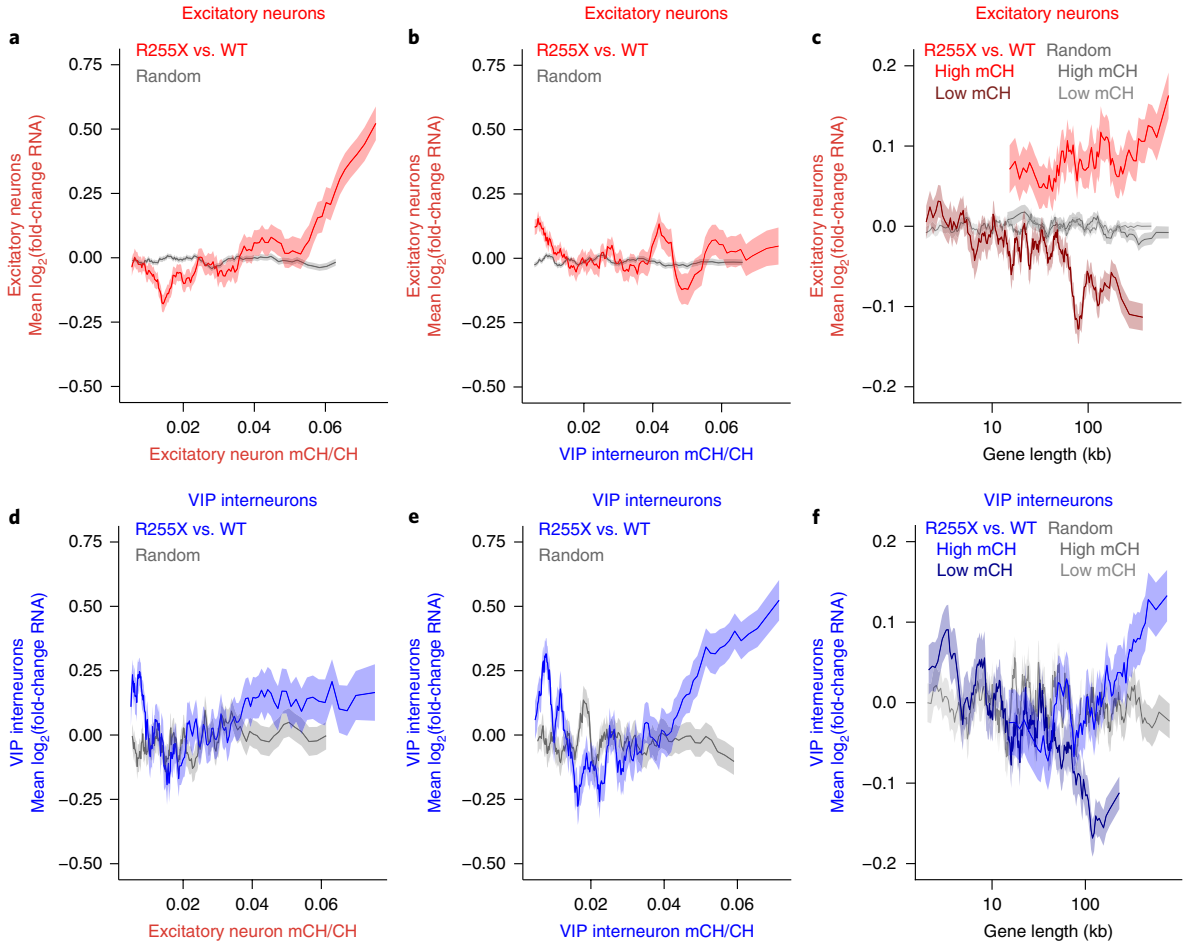


Fig. 3 | Cell-type-specific DNA methylation patterns predict gene misregulation in Rett syndrome. For each graph in **a**,**b** and **d**,**e**, mean fold-change in gene expression of R255X MECP2 nuclei compared to WT nuclei (R255X vs. WT) or of two groups of the respective cell type that were randomly assigned transcriptotypes (Random) is binned according to the fraction of gene-body DNA methylation (mCH/CH). **a**,**b**, Gene expression changes (FDR < 0.01, monocle2) from R255X vs. WT or Random excitatory neurons compared to patterns of DNA methylation from (**a**) excitatory neurons (Pearson's $r = 0.22$) or from (**b**) VIP interneurons (Pearson's $r = -0.01$; 250 bins, 25-gene step). R255X vs. WT is significantly more correlated with excitatory neuron mCH/CH than Random in **a** (permutation test, $P < 0.001$) and significantly more correlated with excitatory mCH/CH patterns than (**b**) mCH/CH patterns from VIP interneurons (R255X vs. WT correlation compared to R255X vs. WT, permutation test, $P < 0.001$). **d**,**e**, Gene expression changes (FDR < 0.25) from R255X vs. WT or Random VIP interneurons are compared to DNA methylation patterns from (**d**) excitatory neurons (Pearson's $r = 0.05$, R255X vs. WT; Pearson's $r = 0.03$, Random) or from (**e**) VIP interneurons (Pearson's $r = 0.18$, R255X vs. WT; Pearson's $r = -0.05$, Random; 50 bins, 5-gene step). In VIP interneurons, R255X vs. WT is significantly more correlated with mCH/CH than Random in **e** (permutation test, $P < 0.001$) but not in **d** (permutation test, $P = 0.71$). In VIP interneurons, the correlation of R255X vs. WT with mCH/CH is significantly greater for mCH/CH patterns from (**e**) VIP interneurons than mCH/CH patterns from (**d**) excitatory neurons (permutation test, $P = 0.008$). **c**,**f**, Mean fold-change in gene expression of R255X vs. WT excitatory neuronal nuclei (**c**) or VIP interneuronal nuclei (**f**) for expressed genes (> 0.1 normalized expression) with high mCH (top 25% mCH/CH) or low mCH (bottom 66% mCH/CH) binned according to gene length (250 bins, 25-gene step). MECP2-dependent gene expression and gene length were significantly more correlated in R255X vs. WT than Random for high mCH/CH genes (R255X vs. WT:,**c**, Pearson's $r = 0.07$; **f**, Pearson's $r = 0.08$; Random: **c**, Pearson's $r = 0.02$; **f**, Pearson's $r = 0.00$; **c**, permutation test, $P = 0.007$; **f**, permutation test, $P < 0.001$) and significantly more anticorrelated for low mCH/CH genes (R255X vs. WT: **c**, Pearson's $r = -0.07$; **f**, Pearson's $r = -0.09$; Random: **c**, Pearson's $r = 0.00$; **f**, Pearson's $r = 0.00$; permutation test for **c** and **f**, $P < 0.001$). Lines represent mean fold-change in expression for genes binned as described; the ribbon is s.e.m. of each bin.

due to the *MECP2* mutation itself or a consequence of genetic and environmental variation between individuals. Our study overcame these limitations and directly assessed the *MECP2*-dependent gene expression changes in the same cell type and genetic background. We found that cell-type-specific patterns of DNA methylation largely predicted the degree of gene upregulation within each subtype of mutant MECP2 (R255X)-expressing neuron from humans with Rett syndrome. Notably, our approach confirmed that the preferential upregulation of highly methylated long genes is a cell-autonomous molecular signature of MECP2 dysfunction, conserved between *MeCP2*-mutant mouse models and humans with Rett syndrome.

The relative contribution of gene length and DNA methylation to MECP2-dependent gene regulation is complex because long genes tend to have a higher level of gene-body methylation compared to shorter genes[10] (Supplementary Fig. 14). Partial correlation analysis was previously used to parse the relative contribution of gene length and DNA methylation to MeCP2-dependent gene regulation in mouse cortical tissue, and it was observed that the total number of gene-body methylcytosine binding sites within a given gene, rather than gene length alone, best predicts MECP2-dependent gene repression[8]. While our scRNA-seq data suggest a role for MECP2 in regulating cell-type-specific gene expression in

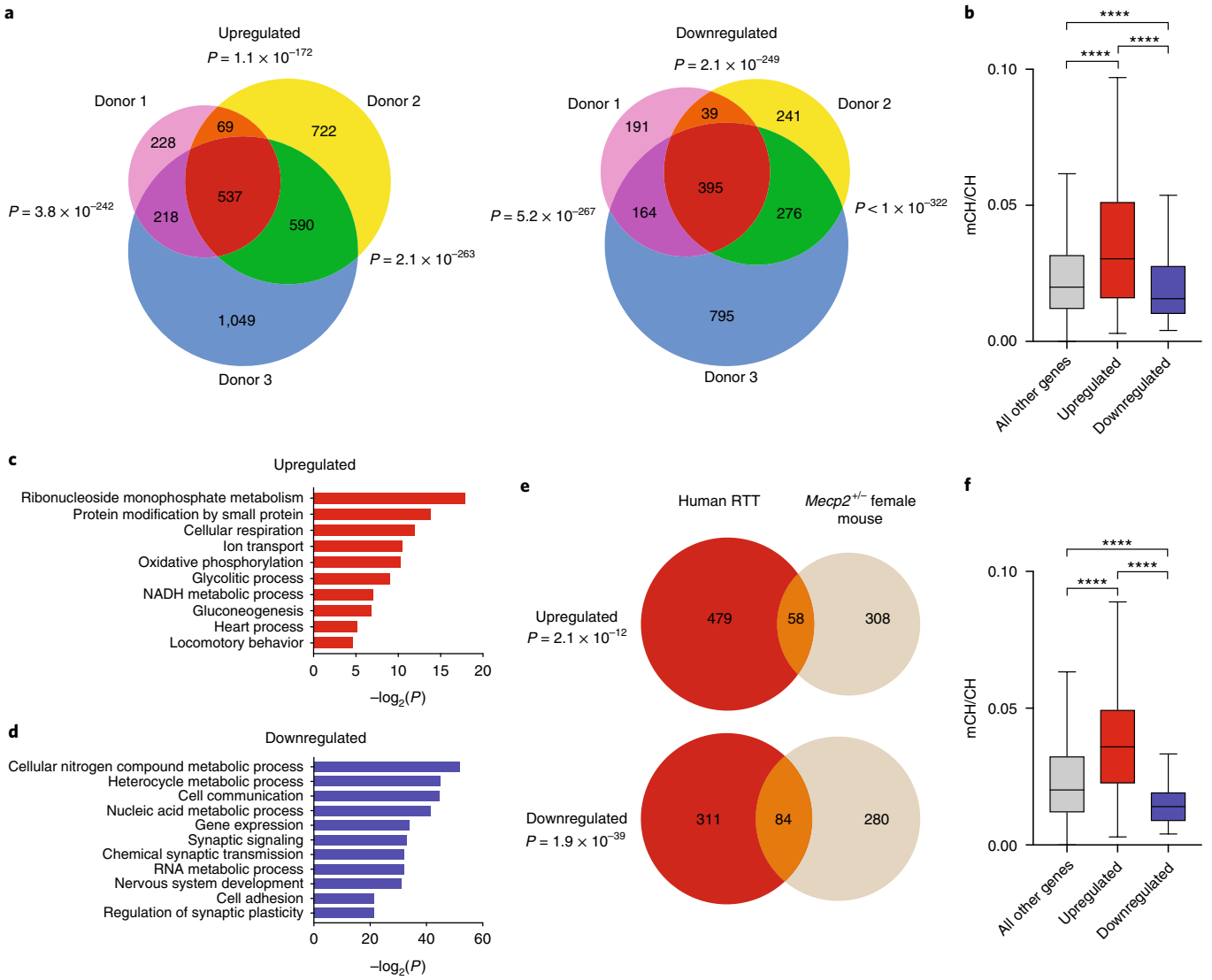


Fig. 4 | Characterization of MECP2-regulated genes in human and mouse. a, Venn diagram of the number of overlapping significantly upregulated (left) or downregulated (right) genes (FDR < 0.1, monocle2) between R255X MECP2-mutant and WT nuclei in excitatory neurons of each donor. *P* values describing the significance of overlap between pairs of upregulated or downregulated gene lists were calculated by hypergeometric testing. **b**, Boxplot of gene-body DNA methylation level (mCH/CH) of the 537 overlapping upregulated genes and 395 overlapping downregulated genes in the 3 donors, as well as all other expressed genes (****$P < 0.0001$, Dunn's test; Kruskal–Wallis test $H_2 = 146.6$). **c**,**d**, Lists of the most highly significant gene ontology terms (Fisher's exact test with FDR) enriched in the 537 overlapping genes that are upregulated (**c**) or the 395 overlapping genes that are downregulated (**d**) between R255X MECP2-mutant and WT excitatory neurons. **e**, Venn diagram of the genes that are commonly upregulated (top, $P = 2.1 \times 10^{-12}$, hypergeometric test) or downregulated (bottom, $P = 1.9 \times 10^{-39}$, hypergeometric test) in mutant *MECP2* compared to WT excitatory neurons in humans and in female heterozygous mice. **f**, Boxplot of the fraction of gene-body DNA methylation (mCH/CH) of the 58 overlapping upregulated genes and 84 overlapping downregulated genes between human and mouse (****$P < 0.0001$, Dunn's test; Kruskal–Wallis test $H_2 = 52.35$). Boxplots show the median (center line), 25th to 75th percentiles (box), and 1.5 × the interquartile range (whiskers).

a DNA methylation-dependent manner, MeCP2-dependent gene expression also correlates with DNA methylation patterns in whole cortical tissue[10,24]. This finding is likely explained both by an averaging effect due to the most abundant cell type driving the observed DNA methylation and gene expression patterns and the presence of commonly methylated regions that would be expected to result in similar MECP2-dependent gene expression across cell types.

The power of single-cell and single-nucleus RNA sequencing to identify MECP2-regulated genes in a given individual with Rett syndrome and in specific cell types enabled the identification of MECP2-repressed genes and MECP2-activated genes that are evolutionarily conserved in both mouse and human excitatory neurons. While deeper single-cell sequencing will provide the statistical power necessary to identify many additional conserved MECP2-regulated genes, the set of genes described here has the potential to provide new insight into Rett syndrome pathophysiology and provides an opportunity to link mechanistic studies of MECP2 function in mouse models to Rett syndrome in humans. Notably, the conserved MECP2-repressed genes had significantly higher levels of gene body DNA methylation than the set of conserved MECP2-activated genes (Fig. 4f). The high levels of DNA methylation within the transcribed region of these MECP2-repressed genes, taken together with abundant evidence that MECP2 binds preferentially to methylcytosines[10,29], suggest that the conserved MECP2-repressed genes are direct targets of MECP2. However, it remains to be determined whether the conserved genes downregulated in the

absence of MECP2 are downregulated due to a secondary change in neurons that occurred as a consequence of the disrupted expression of highly methylated genes or whether these genes were activated directly by MECP2 via a distinct mechanism. It should be noted that a previous report suggested that MeCP2 may regulate long gene expression through a post-transcriptional mechanism, but in this previous study, gene body DNA methylation was not considered[27]. Reanalysis of the data in this study with respect to DNA methylation supports the conclusion that MeCP2 represses gene expression at the level of transcription (Supplementary Fig. 15). Additional studies into the regulation of nuclear or nascent RNA by MeCP2 will likely reveal valuable new insights into MeCP2′s function.

It remains challenging to reconcile the small magnitude of misregulation that occurs for an individual gene when *MECP2* is mutated with the dramatic neurological sequelae of Rett syndrome. It is possible that the deleterious effect of mutating *MECP2* may summate across hundreds to thousands of genes to cause Rett syndrome[10] or that only a small subset of the misregulated genes are responsible for the neurological phenotypes. It is also possible that the kinetics of gene transcription (for example, elongation rates) are altered in the absence of MECP2, which could result in abnormal timing of transcriptional programs in addition to subtle changes in steady-state gene expression[9]. Further study of the proximal mechanisms by which MECP2 regulates gene expression is needed to identify therapeutic approaches for normalizing the diverse gene-expression abnormalities that occur across cell types in Rett syndrome.

Taken together, we have shown that single-cell and single-nucleus SNP-seq enables cell-type-specific characterization of gene expression in mosaic mouse models and post-mortem tissue of human brain donors. Here we have leveraged this approach to glean new insights into Rett syndrome pathophysiology, and in the future, we envision its broad application to the study of additional X-linked disorders in both the brain and other tissues.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41593-018-0270-6.

Received: 2 April 2018; Accepted: 25 September 2018;
Published online: 19 November 2018

References

1. Amir, R. E. et al. Influence of mutation type and X chromosome inactivation on Rett syndrome phenotypes. *Ann. Neurol.* **47**, 670–679 (2000).
2. Lyst, M. J. & Bird, A. Rett syndrome: a complex disorder with simple roots. *Nat. Rev. Genet.* **16**, 261–275 (2015).
3. Amir, R. E. et al. Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185–188 (1999).
4. Ebert, D. H. et al. Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. *Nature* **499**, 341–345 (2013).
5. Lewis, J. D. et al. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**, 905–914 (1992).
6. Lyst, M. J. et al. Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16**, 898–902 (2013).
7. Skene, P. J. et al. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell* **37**, 457–468 (2010).
8. Kinde, B., Wu, D. Y., Greenberg, M. E. & Gabel, H. W. DNA methylation in the gene body influences MeCP2-mediated gene repression. *Proc. Natl. Acad. Sci. USA* **113**, 15114–15119 (2016).
9. Kinde, B., Gabel, H. W., Gilbert, C. S., Griffith, E. C. & Greenberg, M. E. Reading the unique DNA methylation landscape of the brain: non-CpG methylation, hydroxymethylation, and MeCP2. *Proc. Natl. Acad. Sci. USA* **112**, 6800–6806 (2015).
10. Gabel, H. W. et al. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522**, 89–93 (2015).
11. Brero, A. et al. Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J. Cell. Biol.* **169**, 733–743 (2005).
12. Chahrour, M. et al. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320**, 1224–1229 (2008).
13. Cheng, T. L. et al. MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex. *Dev. Cell* **28**, 547–560 (2014).
14. Young, J. I. et al. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc. Natl. Acad. Sci. USA* **102**, 17551–17558 (2005).
15. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
16. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
17. Zeisel, A. et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
18. Guy, J., Hendrich, B., Holmes, M., Martin, J. E. & Bird, A. A mouse *Mecp2*-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* **27**, 322–326 (2001).
19. Hrvatin, S. et al. Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* **21**, 120–129 (2018).
20. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
21. Nelson, E. D., Kavalali, E. T. & Monteggia, L. M. MeCP2-dependent transcriptional repression regulates excitatory neurotransmission. *Curr. Biol.* **16**, 710–716 (2006).
22. Zhang, W., Peterson, M., Beyer, B., Frankel, W. N. & Zhang, Z. W. Loss of MeCP2 from forebrain excitatory neurons leads to cortical hyperexcitation and seizures. *J. Neurosci.* **34**, 2754–2763 (2014).
23. Sugino, K. et al. Cell-type-specific repression by methyl-CpG-binding protein 2 is biased toward long genes. *J. Neurosci.* **34**, 12877–12883 (2014).
24. Guo, J. U. et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* **17**, 215–222 (2014).
25. Chen, L. et al. MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proc. Natl. Acad. Sci. USA* **112**, 5509–5514 (2015).
26. Mo, A. et al. Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* **86**, 1369–1384 (2015).
27. Johnson, B. S. et al. Biotin tagging of MeCP2 in mice reveals contextual insights into the Rett syndrome transcriptome. *Nat. Med.* **23**, 1203–1214 (2017).
28. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
29. Lagger, S. et al. MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. *PLoS Genet.* **13**, e1006793 (2017).
30. Lister, R. et al. Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
31. Stroud, H. et al. Early-life gene expression in neurons modulates lasting epigenetic states. *Cell* **171**, 1151–1164.e16 (2017).
32. Luo, C. et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).
33. Deng, V. et al. *FXYD1* is an MeCP2 target gene overexpressed in the brains of Rett syndrome patients and *Mecp2*-null mice. *Hum. Mol. Genet.* **16**, 640–650 (2007).
34. Lin, P. et al. Transcriptome analysis of human brain tissue identifies reduced expression of complement complex C1Q genes in Rett syndrome. *BMC Genomics* **17**, 427 (2016).
35. Kyle, S. M., Vashi, N. & Justice, M. J. Rett syndrome: a neurological disorder with metabolic components. *Open Biol.* **8**, 170216 (2018).
36. 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* **80**, 129–134 (2002).
37. Fogel, B. L. et al. RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* **21**, 4171–4186 (2012).
38. Gao, Z. et al. An AUTS2-Polycomb complex activates gene expression in the CNS. *Nature* **516**, 349–354 (2014).
39. Cossette, P. et al. Mutation of *GABRA1* in an autosomal dominant form of juvenile myoclonic epilepsy. *Nat. Genet.* **31**, 184–189 (2002).
40. Wallace, R. H. et al. Febrile seizures and generalized epilepsy associated with a mutation in the Na^+-channel beta1 subunit gene *SCN1B*. *Nat. Genet.* **19**, 366–370 (1998).

41. Egan, M. F. et al. The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell* **112**, 257–269 (2003).
42. Gauthier, J. et al. Truncating mutations in *NRXN2* and *NRXN1* in autism spectrum disorders and schizophrenia. *Hum. Genet.* **130**, 563–573 (2011).
43. Auton, A. et al. Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
44. Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).
45. Tukiainen, T. et al. Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).

Acknowledgements

D. Harmin assisted with data processing and scripting. H. Gabel provided thoughtful comments on the manuscript, and A. Ratner provided technical assistance. The single-cell methylation data was graciously provided by C. Luo and J. Ecker. We thank the Rett Syndrome Research Trust for support of this work, along with NIH K08NS101064 (WR), F32NS101739 (LDB), and R01NS048276 (MEG). We are also grateful to donor families and the Harvard Brain Bank for providing tissue from brain donors with Rett syndrome.

Author contributions

W.R. and L.D.B. conceived, designed, and performed all experiments and analyzed all data. S.H. developed the cell dissociation protocol. E.L. designed and performed gene-specific library preparations. S.H., A.S., and T.V. designed and performed FACS experiments. M.A.N. mapped sequencing data. W.R., L.D.B., E.C.G., and M.E.G. wrote the manuscript. M.E.G. advised on all aspects of the study.

Competing interests

M.E.G. is on the Board of Directors and holds equity in Allergan, plc.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-018-0270-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.E.G.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Mice. All animal experiments were approved by the National Institutes of Health and the Harvard Medical School Institutional Animal Care and Use Committee and were conducted in compliance with the relevant ethical regulations. Male and female *Mecp2*-knockout mice and their wild-type controls were obtained from Jackson Labs (Stock No. 003890). This line was originally generated by Adrian Bird[18]. Mice were housed under a standard 12-h light cycle before being placed in constant darkness for 7 d before they were killed for analysis. *Mecp2*-mutant mice all demonstrated decreased locomotor activity at time of analysis; male mice were 8 weeks old and female mice were 12–20 weeks old. Mice of the respective genotype, age, and sex were randomly selected for inclusion in the study.

Brain tissue samples from donors with Rett syndrome. Postmortem cortical tissue (visual cortex, BA17) was obtained from the National Institutes of Health NeuroBioBank and Harvard Brain Bank with approval from the coordinating foundation, Rettsyndrome.org. The study was conducted in compliance with relevant consent and ethical considerations. Work was approved by Harvard Medical School and is compliant with all ethical regulations. Rett donor samples were genotyped by the NeuroBioBank/Harvard Brain Bank and were confirmed by Sanger sequencing.

Single-cell isolation from male and female mouse cortex. Single-cell suspensions from adult male and female visual cortex were prepared as described[19]. Briefly, mice were euthanized with isoflurane and perfused with an ice-cold choline solution. Visual cortices were dissected, chopped into 300 µm fragments, and dissociated with papain (Worthington). Cells were then triturated into a single-cell suspension and collected by gradient centrifugation.

Single-nuclei isolation from human postmortem cortex. Single nuclei suspensions from postmortem human occipital cortex were collected as described previously[26] with minor modifications. Cortical tissue was removed from dry ice and placed directly into a dounce with homogenization buffer (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM tricine-KOH, pH 7.8, 1 mM DTT, 0.15 mM spermine, 0.5 mM spermidine, protease inhibitors, 5 µg/mL actinomycin, and 0.04% BSA). After ten strokes with the tight pestle, a 5% IGEPAL (Sigma) solution was added to a final concentration of 0.32% and five additional strokes with the tight pestle were performed. The tissue homogenate was then passed through a 40-µm filter, and diluted 1:1 with OptiPrep and layered onto an OptiPrep gradient as described previously[26]. After ultracentrifugation, nuclei were collected between the 30 and 40% OptiPrep layers, confirmed to be single nuclei, and diluted to 80,000 nuclei/mL for inDrops. All buffers and gradient solutions for nuclei extraction contained RNAsin (Promega) and 0.04% BSA.

Nuclei sorting and RNA sequencing. Cortical tissue from each Rett donor was dounce homogenized in buffer HB (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM tricine-KOH pH 7.8, 1 mM DTT, 0.15 mM spermine, 0.5 mM spermidine, and protease inhibitors). A 5% IGEPAL solution was added to a final concentration of 0.16% followed by five additional dounce strokes, and then the lysate was filtered through a 40 µm strainer. Nuclei were pelleted by centrifuging at 500 *g* for 5 min at 4 °C and washed once with PBS with 1% BSA. To stain nuclei for sorting, nuclei were incubated with a C-terminal MeCP2 antibody[46] at 1:500 for 1 h at 4 °C, washed once with wash buffer (PBS with 1% BSA and 0.16% IGEPAL), incubated with a goat anti-rabbit 647 secondary antibody (Life Technologies, cat# A21244) at 1:500 for 30 min at 4 °C, then washed once with wash buffer. All washes were performed by centrifuging at 500 *g* for 5 min at 4 °C. Nuclei were then resuspended in PBS with 1% BSA and sorted on a Sony SH800Z Cell Sorter (100 µm nozzle, default laser settings). Nuclei were sorted into Trizol LS (Invitrogen), and total RNA was chloroform extracted and purified with the Qiagen RNeasy Micro Kit with on-column DNase treatment. For Sanger sequencing of the *MECP2* R255X mutation, cDNA was generated with the SuperScript III First-strand Synthesis System (Invitrogen). The *MECP2* R255X region was amplified with Q5 Hot Start High-Fidelity Master Mix (NEB) with the following primers: *MECP2* R255X forward: AAGATGCCTTTTCAAACTTCG and *MECP2* R255X reverse: CCCAGGGCTCTTACAGGTCT, and Sanger sequencing was performed with the *MECP2* R255X reverse primer at the DF/HCC DNA Sequencing Facility. To identify monoallelic SNPs in the two populations of nuclei, total RNA-seq libraries were generated with the NEBNext Ultra Directional Library Prep Kit with rRNA depletion. Libraries were sequenced on an Illumina NextSeq 500 with 85-bp single-end reads. Reads were mapped to the hg38 genome with Tophat2.

Single-cell/single-nucleus RNA sequencing (inDrops). Single-cell or single-nuclei suspensions were encapsulated into droplets and lysed, and the RNA within each droplet was reverse-transcribed using a unique nucleotide barcode as described previously[15]. Cell or nuclei encapsulation was performed in a blinded fashion. Approximately 3,000 cells were processed per library and sequenced on an Illumina NextSeq 500 to achieve at least five reads on average per unique molecular index (typically about 500 million reads per 30,000 droplets collected by inDrops). Transcripts were processed and mapped using a previously described pipeline[15]. Briefly, a custom transcriptome was built from Ensembl GRCh38 (GRCm38.85 annotation) and GRCm38 (GRCm38.84 annotation) with the referenced pipeline.

Quality control for cell or nuclei inclusion. Cells or nuclei with more than 500 unique genes detected per cell were included for further consideration. Cells or nuclei with more than 15,000 unique molecular identifiers detected were omitted to minimize inclusion of data that represented the common barcoding of two or more cells.

Cell-type identification by dimensionality reduction. We used the R package Seurat[20] to cluster cells based on similar gene expression profiles. The raw counts obtained from the mapping pipeline described above were log-normalized and scaled to 10,000 transcripts per cell. Variable genes were identified by the MeanVariablePlot() function with the following parameters: x.low.cutoff = 0.0125, x.high.cutoff = 3, y.cutoff = 0.5. Principal component analysis was then performed, and the top 30 principal components were used for the FindClusters() function (*k*NN clustering) and RunTSNE function (for *t*-distributed stochastic neighbor embedding). Clusters with fewer than 100 cells were omitted from further analysis. Classifications of cell types were determined by visualizing known marker gene expression within each identified cluster. Excitatory neurons were marked by the expression of vesicular glutamate transporter 1 (*Slc17a7*) and calcium/calmodulin-dependent protein kinase II alpha (*Camk2a*). Interneurons were marked by glutamate decarboxylase 1 (*Gad1*), and were further separated into three major subtypes by the expression of parvalbumin (*Pvalb*), vasoactive intestinal peptide (*Vip*), or somatostatin (*Sst*). Astrocytes were marked by the expression of aldolase dehydrogenase (*Aldoc*), oligodendrocytes by the expression of *Olig1*, microglia by the expression of *Cx3cr1*, and endothelial cells by the expression of *Cldn5*. Cells expressing significant levels of two or more of the above marker genes were considered doublets and discarded from further analysis.

General approach to single-cell/nucleus SNP sequencing. There are four general strategies for identifying SNPs in genes expressed in cis with the mutant or wild-type form of a gene. (i) Identify cells that have transcripts covering the mutated genomic region of interest. Because of low per-cell sequencing coverage, it is rare for an individual cell to have coverage of this precise genomic region to directly determine its transcriptotype. Therefore, the few definitively mutant and wild-type cells can be used to search for genomic variation in the expressed X-chromosome genes between mutant and wild-type cells. This provides a set of allele-specific SNPs that can be used in addition to the gene of interest itself to increase the likelihood that a given cell can be transcriptotyped. (ii) Long-read DNA sequencing to directly confirm which SNPs are in cis with the wild-type and mutant gene of interest. This approach would start by identifying SNPs in the single-cell RNA sequencing dataset (for example, half of the reads mapping to the reference nucleotide and the other half mapping to an alternate nucleotide) and perform long-read DNA sequencing (for example, Pacific Biosciences, Oxford Nanopore) to directly confirm which neighboring SNPs are in cis. Once the allele containing a SNP is confirmed to be expressed in cis with either the wild-type or mutant allele of interest, this SNP can be used in turn to identify additional allele-specific SNPs as described in approach i. (iii) Identify SNPs that are in cis with the mutant allele by sequencing members of the donor's family. For example, if the mutation is inherited, DNA sequencing of the X-chromosome of each parent can provide the set of allele-specific SNPs that are unique to the wild-type or mutant alleles. This approach has been employed to catalog X-chromosome inactivation status of human cells[45] and is the approach we used for the analysis of the *Mecp2*-mutant mouse. (iv) Separate wild-type and mutant cells from an individual sample and perform deep RNA sequencing to identify the set of expressed SNPs that are unique to the wild-type or mutant population of cells. This is the approach we used to transcriptotype cells from the human Rett syndrome brain donors.

An additional consideration when implementing single-cell SNP sequencing to study X-linked disorders is that some X-linked disease-causing genes escape X-chromosome inactivation (for example, *IQSEC2*). In these cases, both the mutant and wild-type allele will be expressed in each cell. Therefore, it is important to assess the X-inactivation status of the gene under investigation to ensure it is not biallelically expressed in the cell types of interest. While there is a report that a small percentage of neuroprogenitor cells express *Mecp2* biallelically[47], this event is exceedingly rare in postnatal mouse brain tissue and has not been observed in humans despite an in-depth genome-wide search for X-inactivation escape genes[45].

After transcriptotypes are assigned, it should be noted that, while the wild-type and mutant cells have the same genetic background, there are also allele-specific X-chromosome SNPs expressed in cis with the mutant or wild-type gene. Thus, it is important to have multiple donors with the same mutation to confirm the gene expression differences observed are not secondary to differences in X-chromosome SNPs between mutant and wild-type cells. In our data, these X-chromosome SNPs did not contribute substantially to the gene expression changes observed between mutant and wild-type cells because the three individuals had similar patterns of gene misregulation despite having unique sets of X-chromosome SNPs.

Single-cell SNP sequencing in mosaic female brain tissue. To transcriptotype cells from mosaic female *Mecp2*$^{+/-}$ mutant mice, we first identified SNPs that

were consistently inherited with the mutant *Mecp2* allele. Because this line has been inbred (backcrossed > 38 generations), sequencing offspring from previous litters was equivalent to sequencing the parents directly. For the same reason, however, we also expected that the only retained SNPs from the 129/OlaHsd strain in which the mutant *Mecp2* allele was made would be closely linked to the *Mecp2* locus itself. Indeed, variant calling (Freebayes using default settings, discussed further below) on scRNA-seq data from either *Mecp2* WT or KO male hemizygous mice identified four SNPs within 2 MB of the *Mecp2* locus that were confirmed by manually browsing the RNA sequencing tracks (Integrative Genomics Viewer, Broad Institute; see Supplementary Fig. 1). All male *Mecp2*-knockout mice across two separate generations (WT1–3 and WT4–6 were from separate generations) contained the same SNPs, indicating that these SNPs can be used as a reliable marker of the mutant allele. Given the small number of SNPs that were identified, we attempted to maximize their detection by modifying the standard inDrops single-cell library preparation. Specifically, half of the amplified RNA was processed according to the published protocol using random hexamers and universal primers for PCR amplification, and the other half was reverse-transcribed and then PCR-amplified with gene-specific primers for each allele-specific SNP (see primer sequences below).

To identify the set of expressed SNPs that were unique to wild-type or mutant nuclei from postmortem human Rett syndrome brain donor samples, we first separated wild-type and mutant cells by FACS (described above) and performed deep total RNA-seq on the separate populations. Unlike the highly backcrossed *Mecp2*-mutant mice, there was a wealth of genomic variability that could be used to transcriptotype cells once the variants were confirmed to be expressed in cis selectively with the mutant or wild-type *MECP2* allele. After performing RNA-seq on sorted wild-type ($MECP2^{high}$) and mutant ($MECP2^{low}$) populations of cells, we performed X-chromosome variant calling on these datasets using Freebayes version 1.1.0-4[48] with default parameters. The genomic locations of SNPs with a Freebayes score > 10 were used to extract reads (SAMtools version 1.2) from the mapped RNA-seq data of wild-type and mutant *MECP2* populations. Based on the reference genome, each SNP sequence was assigned as the 'reference' sequence, the 'alternate' sequence, or 'other'. For both the sorted wild-type and mutant samples, the fractions of reference, alternate, or other reads that covered each SNP were calculated. If a gene was expressed from both alleles, approximately 50% of the reads sequenced would be expected to map to each allele. Because X-inactivation typically results in monoallelic expression, for a given cell, most of the sequencing reads (allowing for some sequencing error and/or sorting error) would be expected to map to a single allele. We thus considered the expression of a SNP to be allele-specific if ≥ 85% of wild-type ($MECP2^{high}$) reads and < 85% of the mutant ($MECP2^{low}$) reads encompassing this region contained the same sequence variant (for example, reference or alternate). We used the $MECP2^{high}$ population as the primary filter for monoallelic SNP expression because the $MECP2^{low}$ population, while mainly defined by the low background immunofluorescence signal from MECP2-mutant cells (Sanger sequencing confirmed the cell population expressed the mutant allele), could also contain small numbers of wild-type cells that have background levels of fluorescence because they express low levels of MECP2 (for example, non-neuronal cells). The identification of allele-specific SNPs with these parameters was supported by the observation that for a given SNP, an average of 98% of the reads from wild-type cells ($MECP2^{high}$) mapped to the same SNP (for example, reference) and an average of 76% of the reads from the mutant ($MECP2^{low}$) cells mapped to the alternative SNP (for example, alternate). The allele-specific SNPs identified from each donor sample (Donor 1 = 69 SNPs, Donor 2 = 69 SNPs, Donor 3 = 75 SNPs) were then used to mark the X chromosome alleles that were expressed in cis with either the wild-type or mutant allele of *MECP2*. Custom R scripts were written to process BAM output files from the inDrops mapping pipeline or total RNAseq mapping pipeline for the identification of allele-specific SNPs.

Assignment of transcriptotype to individual cells. After the identification of the allele-specific SNPs that were expressed in cis with either the wild-type or mutant allele, we next used this information to assign transcriptotypes to the individually sequenced cells. To do this, we used SAMtools to identify the sequencing reads within the single-cell or single-nucleus RNA sequencing datasets that contained both the allele-specific SNPs identified above and a unique cell barcode. We then grouped the reads from each cell or nucleus and assigned the *MECP2* transcriptotype corresponding to the profile of allele-specific SNPs expressed. Specifically, a transcriptotype was assigned if ≥ 85% of the reads covering an individual allele-specific SNP mapped to the same allele (for example, ≥ 85% of the reads were reference) and ≥ 80% of the total SNPs covered in each cell were concordant with the same transcriptotype (for example, ≥ 80% of the SNPs covered in a cell were expressed in cis with the R255X allele of *MECP2*). Some cells or nuclei only had one or two reads mapping to an allele-specific SNP, which increases the chance of an incorrect transcriptotype call. After estimating that the mean error rate for transcriptotype assignments was only 0.5% in female $Mecp2^{+/-}$ mice and 4.6% in human Rett donors, we chose to include these cells in the differential gene expression analysis to maximize the number of cells and resulting statistical power. The estimated transcriptotype error rate for a given cell with only one or two reads encompassing allele-specific SNPs was determined as the percent of genotype discordant reads in cells with at least three reads (defined as cells with confident transcriptotypes). The mean estimated transcriptotype error was then calculated by averaging the error rates for each cell in the dataset. The lower estimated error rate in mouse was accomplished by deeper sequencing of the allele-specific SNPs using gene-specific library preparations, an approach that can be used to further improve the confidence of transcriptotype calls in any dataset. Custom R scripts were written to process BAM output files from the inDrops mapping pipeline for assigning transcriptotypes to specific cells based on allele-specific SNPs.

Gene-specific primers for enriching SNP coverage in $Mecp2^{+/-}$ mouse single-cell libraries. Reverse transcription:

rs13468851: TGTATGTCGGACTTGATGTACT
rs13468852: TTTACAGTATTCTTTCTACATGGA
rs31144974: GATTAACTGTAACAACGATCACAAC
rs29035084: GGTTTCAAAGTACCCAGCATAAAT
PCR:
rs13468851: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNCT TGCTCTGTCAAGCTCTTTGC
rs13468852: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNN GATTACATCCGACACGTCTGC
rs31144974: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNN GCATGTTGGATTAGATTGTC
rs29035084: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNCA GCAGAGGTGGCTGAACTT

Differential gene expression analysis. We used Monocle 2 to identify genes differentially expressed between wild-type and mutant cells[49]. The single-cell and single-nucleus RNA-seq data were modeled by a negative binomial distribution, consistent with the expression profiles of our data. Differential expression analysis was conducted independently for each cell type by aggregating the gene counts from the population of cells or nuclei within a given cell type (median counts per gene in human excitatory neuron cluster = 2,029), which provided sufficient coverage of expressed genes for differential expression analysis between mutant and wild-type cells. Certain analyses, where described, required combining cells or nuclei of the same cell-type from multiple mice or human donor samples (for example, Fig. 3 because of limited inhibitory neuron populations). Otherwise, differential expression was performed between mutant and wild-type cells from each human donor sample individually (for example, excitatory neurons in Fig. 4). A gene was included for differential expression analysis if its minimum expression was ≥ 0.1 and it was detected in at least 100 cells or nuclei. Significantly misregulated genes were identified by the FDR cutoff described for the specific analysis and number of cells studied. Differential expression analysis of randomly transcriptotyped cell populations resulted in few, if any, significantly misregulated genes. Thus, the uncorrected *P* values were ranked from smallest to largest and the number of genes selected for analysis was determined by the corresponding number of significantly misregulated genes identified in the respective mutant-to-WT comparison. To generate randomly transcriptotyped groups of cells, the sample function in R was used to randomly select the same number of cells from each individual (without respect to transcriptotype) as was used for the SNP-seq-based transcriptotype analyses. Differential expression analysis was performed on two groups of randomly transcriptotyped cells (the same cell could not exist in both randomly generated lists). If analyses combined cells that were transcriptotyped from multiple individual donor samples (for example, Figs 2 and 3), the corresponding number of cells was first randomly sampled on a per-individual basis and then combined to form the control group. To ensure that the randomly sampled groups in each figure were representative of the entire dataset, differential expression was performed on three independent pairs of randomly sampled cells.

Correlations of MeCP2-dependent gene expression with DNA methylation, MeCP2 ChIP, and gene length. To generate the smooth-line correlation plots, genes were sorted by their gene length, DNA methylation, or MeCP2 ChIP signal, and a sliding window was defined by the indicated bins and step sizes for each analysis. The bins and step sizes were adjusted to the size of the gene list. The $\log_2$(fold-change) for each bin was averaged and plotted with the standard error for each bin. The gene length for a gene was obtained from RefSeq annotation (gene end–gene start). Cell-type-specific mouse DNA methylation data were obtained from ref. [26]. Gene-body-level cell-type-specific human DNA methylation data were obtained from ref. [32] and averaged across all cells within the indicated cell type. Excitatory neuron-specific MeCP2 ChIP-seq data was generated as described below.

Gene ontology analysis and cell-type-specific enrichment analysis. Gene ontology analysis was performed at geneontology.org using the PANTHER overrepresentation test (Fisher's exact test with FDR). All expressed genes (normalized expression > 0.1 in both mutant and wild-type cells of the corresponding cell type) for the respective comparison were used as the background lists. Gene ontology biological processes were reported with redundant/overlapping pathways only displayed once. Single-cell mRNA

sequencing data from > 160,000 cells and 39 distinct cell types were obtained from www.mousebrain.org[50]. For each gene, the normalized gene expression counts were averaged across all cells of the same cell type. The mean gene expression level within each cell type was then row-normalized using the Morpheus heatmap tool (https://software.broadinstitute.org/morpheus). Enrichment statistics were calculated by the following formula: the number of times per 1,000 iterations the mean expression of randomly sampled cells was greater than or equal to the cell type of interest/1,000 iterations.

INTACT nuclei isolation and MeCP2 ChIP-seq. CaMKIIa-Cre mice were crossed with mice that express the SUN1-sfGFP-MYC protein in the nuclear membrane in a Cre-dependent manner, and SUN1-GFP-expressing nuclei were isolated from the forebrain of 8-week-old male SUN1-GFP;CaMKIIa-Cre mice as previously described[26]. Nuclei were immunoprecipitated with a GFP antibody (Fisher G10362) and Protein G Dynabeads (Invitrogen). Nuclei were cross-linked in 1% formaldehyde in PBS for 10 min at room temperature (20–23 °C), quenched with 125 mM glycine for 5 min, and washed twice with PBS. Nuclei were then resuspended in LB3 buffer (10 mM Tris, pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium-deoxycholate, 0.5% *N*-lauroylsarcosine, and protease inhibitors), and sonicated in a Diagenode Bioruptor. Insoluble material and beads were removed by spinning at 16,000 *g* for 10 min at 4 °C, and Triton X-100 was added to soluble chromatin at a final concentration of 1%. Chromatin was precleared for 2 h with Protein A Dynabeads, then incubated with Protein A Dynabeads conjugated to an MeCP2 antibody[46] overnight at 4 °C. Beads were washed twice with low-salt buffer (20 mM Tris, pH 8, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), twice with high-salt buffer (20 mM Tris, pH 8, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), twice with LiCl wash buffer (10 mM Tris, pH 8, 1 mM EDTA, 1% NP-40, 250 mM LiCl, 1% sodium deoxycholate), and once with TE buffer (50 mM Tris, pH 8, 10 mM EDTA) at 4 °C. Chromatin was eluted off the beads by incubating in TE buffer with 1% SDS at 65 °C for 1 h, and cross-links were reversed by incubating overnight at 65 °C. Chromatin was treated with RNase A for 30 min at 37 °C and Proteinase K for 2 h at 55 °C. DNA was phenolchloroform-extracted and purified with the Qiagen PCR purification kit. Libraries were generated using the NuGEN Ovation Ultralow System V2 following the manufacturer's instructions. Libraries were sequenced on an Illumina NextSeq 500 with 85-bp single-end reads. Reads were mapped to the mm10 genome with Bowtie2, and PCR duplicates were removed using SAMtools rmdup. Mapped reads from MeCP2 ChIP and input were randomly downsampled to the same number of reads. Bedtools map was used to count ChIP and input reads mapped to gene bodies for comparison to gene expression.

Reanalysis of published RNA sequencing data. Gene read-count tables for male 6-week-old WT and R106W excitatory neuron nuclear RNA-seq and female 18-week-old $R106W_{WT}$ and $R106W_{MUT}$ excitatory neuron nuclear RNA-seq were downloaded from GEO (GSE83474). Differential expression analysis was performed with the R package edgeR. FDR < 0.1 was used to identify differentially expressed genes. For comparison to DNA methylation, excitatory neuron mCA[26] was mapped to the gene body locations in the Johnson et al. counts tables using bedtools map.

Overlap with autism and intellectual disability genes. Rett syndrome gene lists were compared to the autism genes list at SFARI (https://gene.sfari.org/) and to the intellectual disability gene lists at the University of Colorado (http://gfuncpathdb.ucdenver.edu/iddrc/iddrc/home.php). Enrichment statistics were calculated using the hypergeometric test in R 3.3.2.

Statistical analysis. Enrichment statistics of pairwise comparisons between two gene lists was calculated using the hypergeometric test as calculated in R 3.3.2. Pearson correlations between gene expression and DNA methylation, gene length, or MeCP2 ChIP were compared by permutation. *P* values for these comparisons were estimated by calculating: (number of events where $|corr1^{permutation} - corr2^{permutation}| > |corr1^{observed} - corr2^{observed}|$)/1,000 permutations. Kruskal–Wallis tests and Mann–Whitney tests were performed using Prism v7. No statistical methods were used to predetermine sample sizes, but our samples sizes are similar to or larger than those reported in previous publications[19,28].

Accession codes. Data has been deposited at GEO: GSE113673.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability. Custom R scripts can be made available upon reasonable request.

Data availability

All sequencing data reported in this study have been deposited in the NCBI Gene Expression Omnibus under accession GSE113673.

References

46. Chen, W. G. et al. Derepression of *BDNF* transcription involves calcium-dependent phosphorylation of MeCP2. *Science* **302**, 885–889 (2003).
47. Gendrel, A. V. et al. Developmental dynamics and disease potential of random monoallelic gene expression. *Dev. Cell* **28**, 366–380 (2014).
48. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at *arXiv* https://arxiv.org/abs/1207.3907 (2012).
49. Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
50. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018).

nature research

Corresponding author(s): Michael Greenberg

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	SH800Z Cell Sorter Software version 2.1.3
Data analysis	TopHat2 v2.1.0 was used to align total RNAseq reads. Freebayes version 1.1.0-4 was used to identify single nucleotide variants. Samtools 1.2 was used to extract reads containing variants. Prism v7 was used for certain statistical analyses. Single-cell clustering was performed with Seurat and differential expression was performed with Monocle 2. Custom scripts were written in R version 3.3.3 to count the number of reads that contain allele-specific single nucleotide polymorphisms in the single-cell RNA sequencing data. These are available upon request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All sequencing data reported in this study have been deposited in the NCBI Gene Expression Omnibus under accession GSE113673

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine sample size but our samples sizes are similar or larger to those reported in previous publications.
Data exclusions	Cells or nuclei with greater than 500 unique genes detected per cell were included for further consideration. Cells or nuclei with greater than 15,000 unique molecular identifiers detected were omitted to minimize inclusion of data that represented the common barcoding of two or more cells.
Replication	We observed the same findings within each rodent and human individual studied.
Randomization	mice of the correct genotype, age, and sex were randomly selected for inclusion in the study. To generate randomly transcriptotyped groups of cells, the sample function in R version 3.2.2
Blinding	Cell encapsulation was blinded. Analysis was conducted using automated scripts, but knowledge of sample IDs to write the code.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	MECP2 C-terminal antibody (made in-house); goat anti-rabbit 647 secondary antibody (Life Technologies) cat#A21244
Validation	Chen, et al. Science 2003, Cohen et al. Neuron 2011. Signal is absent in MeCP2 KO tissue.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mecp2tm1.1Bird/J mutant mice and their wild-type controls were obtained from Jackson labs (Stock No. 003890). Male mice were 8 weeks old, female mice were 12-20 weeks old.
Wild animals	This study did not use wild animals
Field-collected samples	This study did not use field-collected samples

Human research participants

Policy information about studies involving human research participants

Population characteristics	Female Rett syndrome brain donors were 8-24 years old. All donors had the R255X mutation in MECP2.
Recruitment	Rett syndrome donor brain tissue was obtained from the NIH NeuroBioBank/Harvard Brain Bank in coordination with Rettsyndrome.org

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	GEO accession GSE113673
Files in database submission	Raw Samples: CamkIIa_cortex_MeCP2_ChIP_rep1, CamkIIa_cortex_input_rep1, CamkIIa_cortex_MeCP2_ChIP_rep2, CamkIIa_cortex_input_rep2 Processed: GSE113903_CamkIIa_mecp2_chip_counts.txt.gz
Genome browser session (e.g. UCSC)	N/A

Methodology

Replicates	2 biological replicates of MeCP2 ChIP and input from CamkIIa-positive nuclei from mouse cortex.
Sequencing depth	MeCP2 ChIP rep 1: 59108980 total reads; 32580048 mapped, de-duplicated reads Input rep 1: 62068745 total reads; 44909766 mapped, de-duplicated reads MeCP2 ChIP rep 2: 55833090 total reads; 27752219 mapped, de-duplicated reads Input rep 2: 59495391 total reads; 40515255 mapped, de-duplicated reads Read length: 85bp single-end
Antibodies	MeCP2 antibody from Chen WG et al. Science 2003
Peak calling parameters	Read mapping: Used Bowtie2 (version 2.2.4) to map to mm10 genome using default parameters Peak calling: N/A; Mapped ChIP and input reads were counted in mm10 gene body locations using Bedtools map (version 2.23.0)
Data quality	N/A; did not call peaks; see MeCP2 gene body densities in processed counts tables.
Software	Trimmomatic (version 0.33) was used to trim reads Bowtie2 (version 2.2.4) was used to map reads to mm10 genome Samtools (version 0.1.19) was used to remove duplicate reads UCSC-tools was used to extend reads to 250bp Bedtools (version 2.23.0) was used to count mapped ChIP and input reads in mm10 gene body locations

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	To stain nuclei for sorting, nuclei were incubated with a C-terminal MeCP2 antibody at 1:500 for 1 hour at 4°C, washed once with Wash buffer (PBS with 1% BSA and 0.16% IGEPAL), incubated with a goat anti-rabbit 647 secondary antibody (Life Technologies) at 1:500 for 30 min at 4°C, then washed once with Wash buffer. All washes were performed by centrifuging at 500g 5 min at 4°C. Nuclei were then resuspended in PBS with 1% BSA.
Instrument	Sony SH800Z Cell Sorter
Software	SH800 Software
Cell population abundance	The MECP2_high population was 13.55% and MECP2_low population was 35.31%. We confirmed purity through Sanger sequencing and total RNA sequencing of these cell populations.
Gating strategy	DAPI-positive singlets were gated on fluorescence intensity of MeCP2 staining to define high and low populations. A no primary antibody control was used to confirm specificity of staining.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.