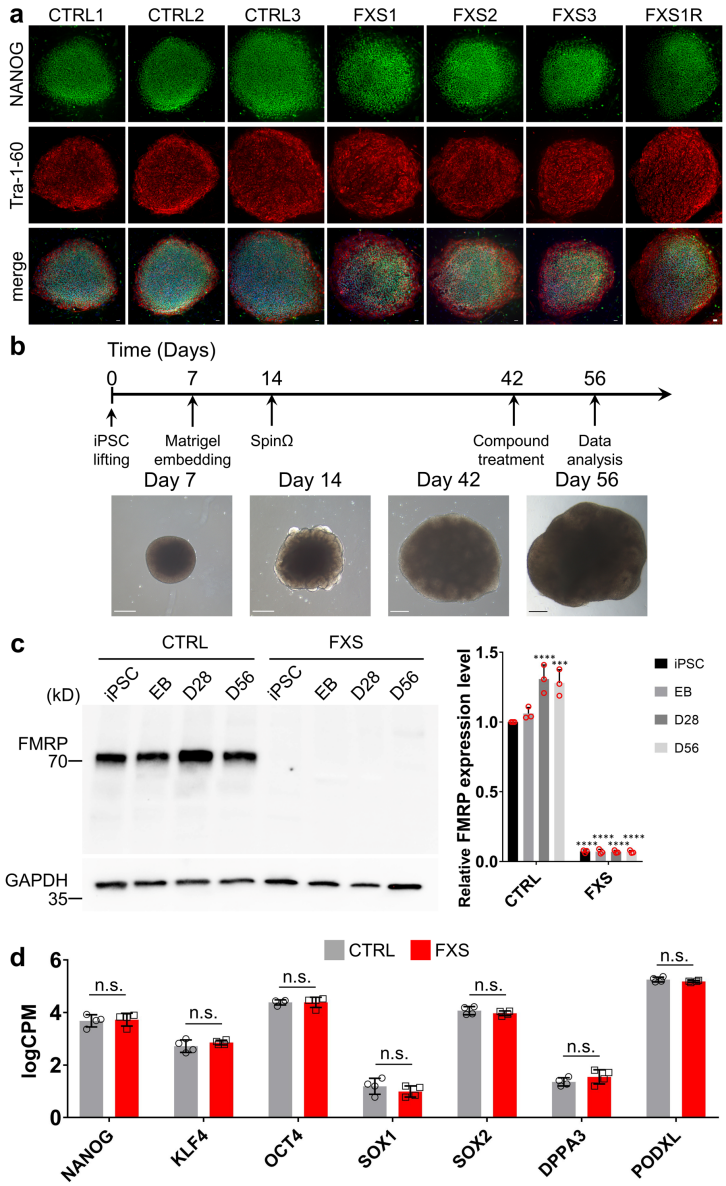
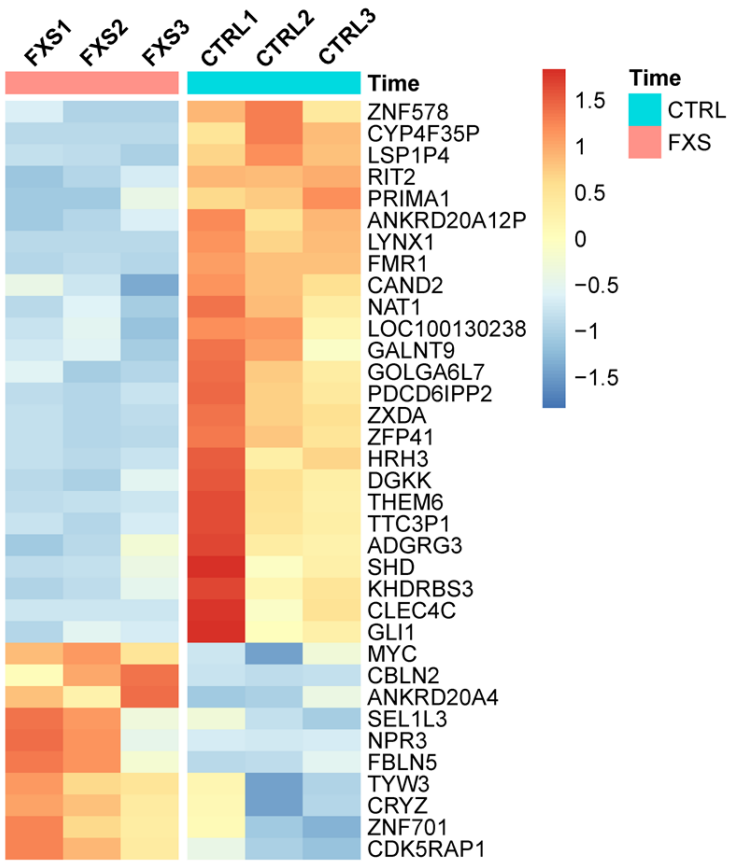
Supplementary Information

A human forebrain organoid model of fragile X syndrome exhibits altered neurogenesis and highlights new treatment strategies

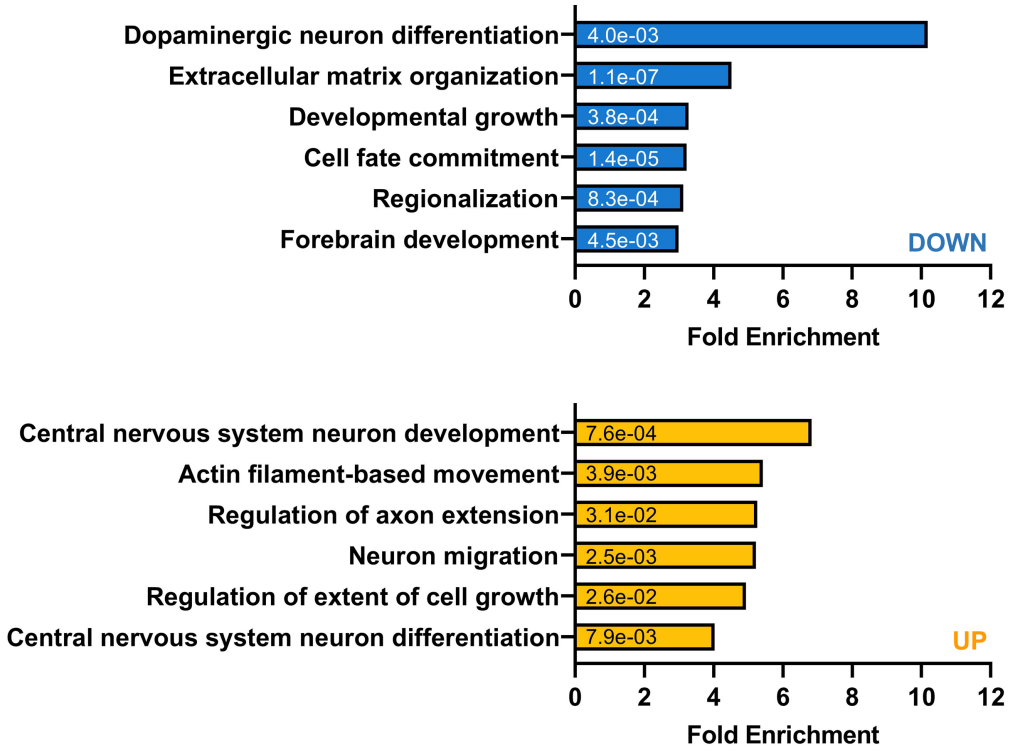
Yunhee Kang[1,9], Ying Zhou[2,9], Yujing Li[1,9], Yanfei Han[2], Jie Xu[3], Weibo Niu[2], Ziyi Li[4], Shiying Liu[5], Hao Feng[5], Wen Huang[6], Ranhui Duan[6], Tianmin Xu[7], Nisha Raj[8], Feiran Zhang[1], Juan Dou[2], Chongchong Xu[2], Hao Wu[9], Gary J Bassell[8], Stephen T Warren[1], Emily G Allen[1], Peng Jin[1,*], and Zhexing Wen[2,8,*]



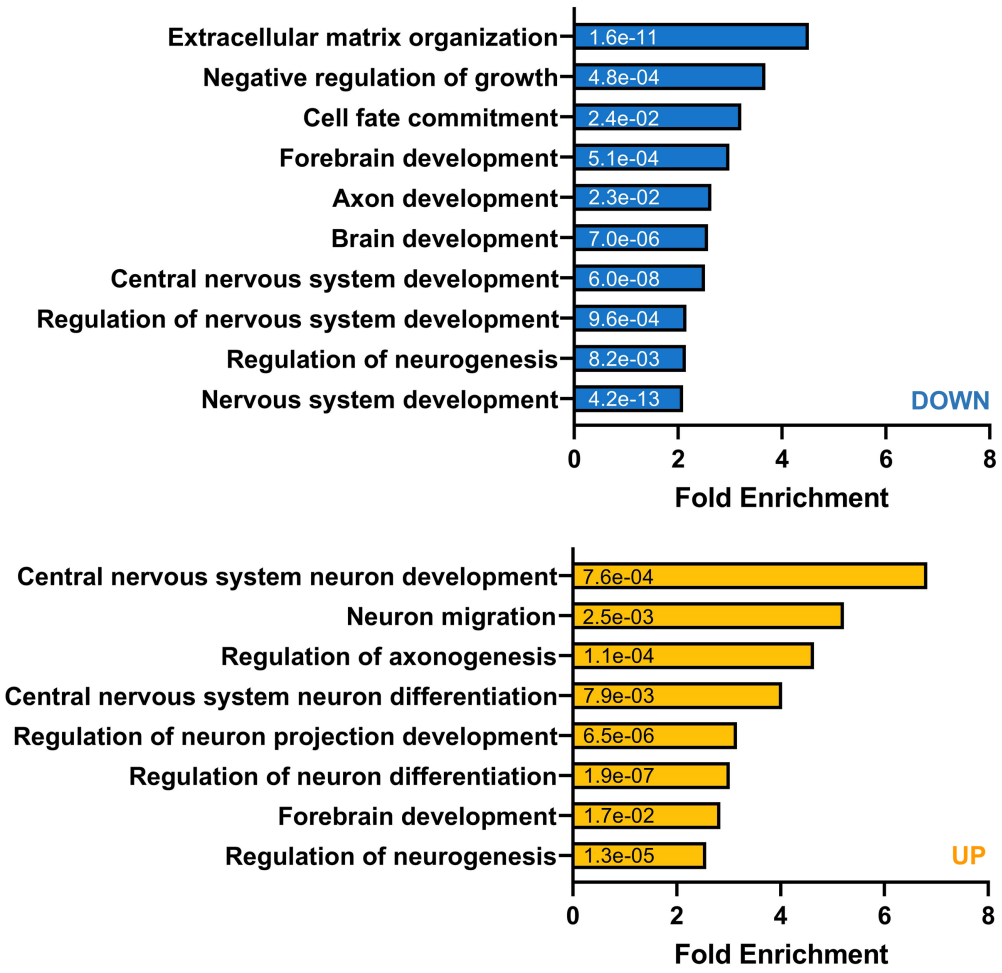
Supplementary Figure 1. Basic characterization of iPSC lines and generation of forebrain organoids. (a) Shown are sample images of immunostaining of pluripotency-associated markers NANOG and Tra-1-60 for different iPSC lines. Scale bar: 20 µm. Each experiment was repeated at least three times independently with similar results. (b) Schematic diagram of forebrain organoid differentiation procedure (top) and sample phase images at different stages (bottom). Scale bars: 500 µm. (c) Sample images (left) and quantification (right) of Western blots are presented for comparing FMRP protein level in iPSCs, EBs, and D28 and D56 forebrain organoids using GAPDH as loading control. Data are presented as mean ± s.d. (n = 3 cultures; ***P = 0.0001, ****P < 0.0001, one-way ANOVA). (d) Gene expression profiling revealed that the control and FXS lines expressed similar levels of genes associated with pluripotency. Data are presented as mean ± s.d. (n = 4 cultures; one-way ANOVA).



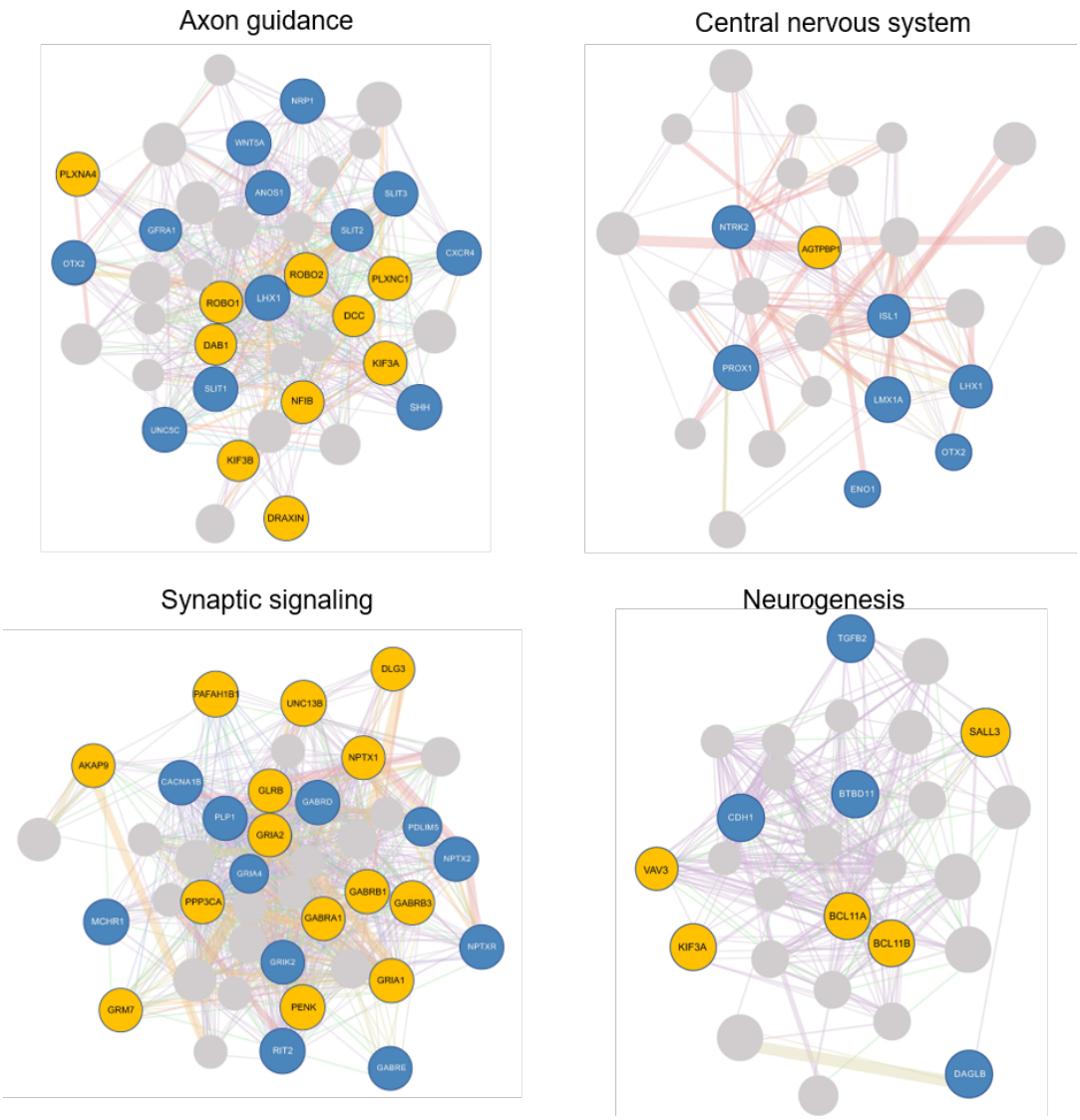
Supplementary Figure 2. Basic characterization of consistency across D56 forebrain organoids. Shown are heatmap of DEGs of control and FXS human forebrain organoids used in the study. The heatmap shows relative consistency within the group of control and FXS human forebrain organoids.



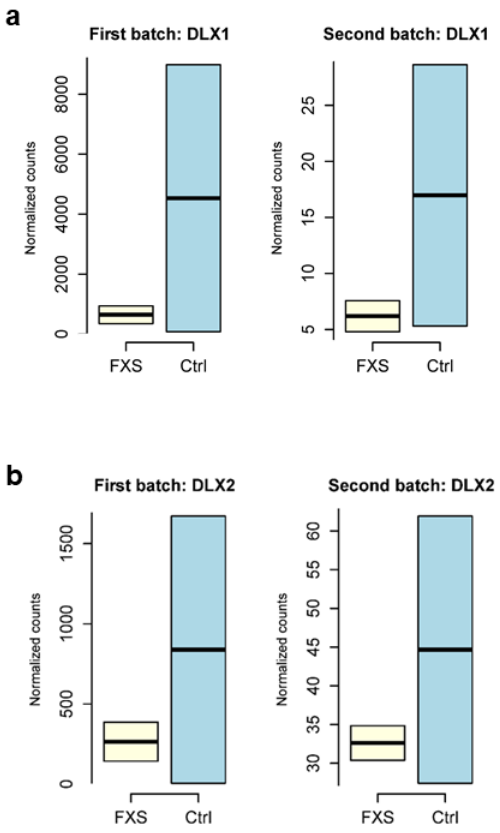
Supplementary Figure 3. The PANTHER overrepresentation test on the overlapped genes between all forebrain organoids and human fetal tissues shows enrichment in distinct pathways. Yellow represents the up-regulated genes and blue represents the down-regulated genes. The DEGs of the overlapped genes between all forebrain organoids and human fetal tissues are enriched in forebrain development, regulation of axon extension and regulation of neurogenesis. The numbers on the bars indicate the two-sided p-values by Fisher's exact test. The p values have been adjusted for multiple testing using Bonferroni correction.



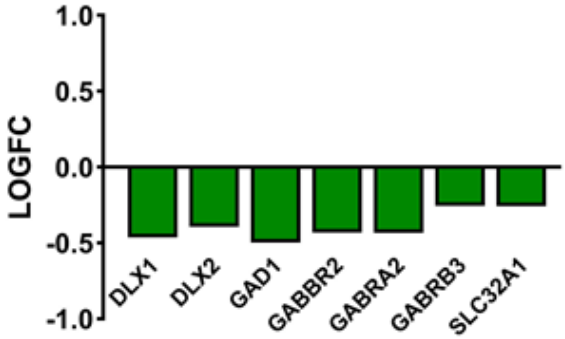
Supplementary Figure 4. The PANTHER overrepresentation tests for DEGs identified from D56 forebrain organoid RNA-seq are shown. The upregulated genes in FXS show specific pathway enrichment in cell fate commitment, forebrain development, axon development and regulation of neurogenesis. DE genes with higher expression in FXS than in control organoids are concentrated in central nervous system neuron development, neuron migration, axonogenesis, neuron differentiation, forebrain development and regulation of neurogenesis. Yellow represents up-regulated genes and blue represents down-regulated genes. The numbers on the bars indicate the two-sided p-values by Fisher's exact test. PANTHER analysis of all overlapped DE genes is shown in Figure 4d. The p values have been adjusted for multiple testing using Bonferroni correction.



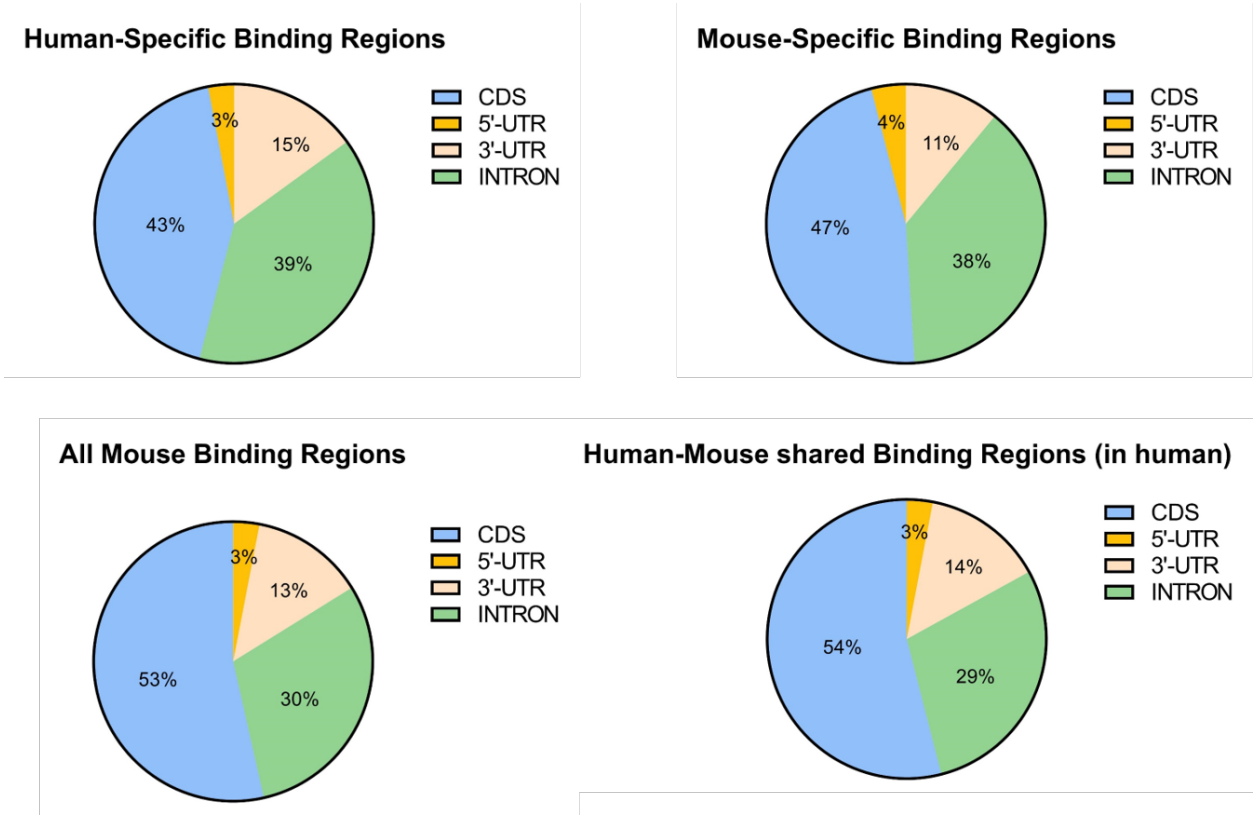
Supplementary Figure 5. Interactome plots of forebrain development on various developmental aspects are represented. Significant DE genes found in day 56 forebrain organoid RNA-seq results are shown and highlighted: central nervous system neuron differentiation (GO:0021953, P=8.7E-06), axon guidance (GO:0007411, P=1.2E-05), forebrain development (GO:0030900, P=1.6E-08), neurogenesis (GO:0022008, P=9.7E-21), synaptic signaling (GO:0099536, P=6.3E-03). Yellow represents up-regulated genes and blue represents down-regulated genes.



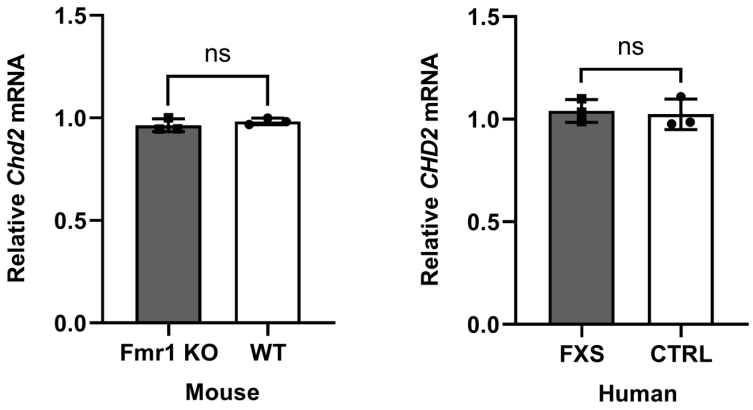
Supplementary Figure 6. Expression change trend of DLX1/2 in control and FXS human forebrain organoid RNA-seq. Expression change trend of DLX1 (a) and DLX2 (b). (n = 3 organoid cultures). In bulk RNA-seq, the trend shows that FXS human forebrain organoids have decreased DLX1/2 compared to controls. The middle line of the box indicates the median value. The top and bottom lines correspond to the maxima and minima.



Supplementary Figure 7. The differential expression of GABAergic system molecules in FXS human forebrain organoids in single cell RNA-seq. Several representative inhibitory synaptic system markers such as DLX1/2, GAD1, GABBR2, GABRA2, GABRB3, and SLC32A1 are down-regulated in FXS human forebrain organoids single cell RNA-seq inhibitory neuron cluster, C7. The genes are presented with gene expression proportions between control and FXS, log fold-change value, and p-values from two-sample proportional test. Two-sided p-values are generated from the likelihood ratio test (LRT) using a generalized linear model in Monocle3. Both original p-values and multiple testing adjusted q-values are reported. Q-values were adopted to control for the positive false discovery rate (pFDR). (N=3 single-cell RNAseq of 3 independent culture sets, Average LogFC=0.43, $P < 6.27e-09$)



Supplementary Figure 8. Most of the FMRP binding sites on the FMRP target mRNA species identified using eCLIP are in CDS and introns. Binding region analysis was performed by counting mapped reads identified by eCLIP. Pie charts of specific subsets of transcripts such as human specific, mouse specific, all mouse, or human-mouse shared genes are shown. The binding sites of FMRP on its target RNAs are mapped to 4 regions (CDS, 5'-UTR, 3'-UTR and intron) and the percentage of binding to each region on target mRNA is plotted in the pie chart.



Supplementary Figure 9. No significant difference in CHD2 mRNA was seen in either mouse embryonic cortex or human forebrain organoids in the presence or absence of FMRP by qRT-PCR. Data are presented as mean ± s.e.m. (N=3, ns, two-tailed unpaired t test).

Supplementary Table 1: iPSC lines

Supplementary Table 2: Cortical layer marker expression
The results were obtained by comparing the FXS versus control organoids at day 56 (N = 2 FXS organoids versus 2 controls) using R package DESeq2. The default statistical test for differential analysis in DESeq2 is used to obtain p values. The p values are two sided and have been adjusted for multiple testing by Benjamini-Hochberg procedure.

Supplementary Table 3: Organoid_DEGs
Differentially expressed genes of organoid samples at different stages. "Summary" provides a summarization of number of differentially expressed genes detected from organoid samples at day 28, 56, or 84, as well as the overlaps between DEGs at different dates. "D28", "D56" and "D84" include results of analyzing FXS organoids versus controls at day 28, 56 and 84 (N = 2 FXS organoids versus 2 controls at each time point) using R package DESeq2. The default statistical test for differential analysis in DESeq2 is used to obtain p values. Adjusted p value < 0.20 is used as threshold for DE genes. P values are two-sided and have been adjusted for multiple testing by Benjamini-Hochberg procedure.

Supplementary Table 4: FetalBrain_DEGs
Differentially expressed genes of human fetal brain samples (N = 2 FXS fetal brains versus 2 controls) using R package DESeq2. The default statistical test for differential analysis in DESeq2 is used to obtain p values. Adjusted p value < 0.20 is used as threshold for DE genes. P values are two-sided and have been adjusted for multiple testing by Benjamini-Hochberg procedure.

Supplementary Table 5: Summary of RNA-seq analysis with various analysis packages
For bulk RNA-seq analyses, either salmon or STAR package was used to perform alignments and differentially express (DE) genes were called using DESeq2 or edgeR.

Supplementary Table 6: Seurat cluster marker
List of marker genes for fourteen clusters using Seurat. After scRNA-seq data normalization and case-control integration, we identified 14 cell clusters and their corresponding marker genes. Each row describes one gene, with columns describing the characteristics of this gene among FXS and control groups. Characteristics include proportions of cells that has gene expressed, log fold-change values, p-values from the proportional test, combined meta p-values across conditions. Each tab represents one cluster. The last column "minimal_p-val" measures each marker's significance for being a cluster-specific marker gene. Here, multiple Fisher's p-values resulted from the test were combined using meta-analysis strategy. Genes were sorted by minimal p-values across FXS and control conditions. Multiple testing adjustments are not utilized here as they do not alter the gene ordering pattern.

Supplementary Table 7: Seurat cluster DE
List of cluster-specific differentially expressed (DE) genes between FXS and CTRL. FXS and CTRL single cells were integrated for cluster identification in previous step, then DE genes were identified within each cluster, contrasting FXS versus CTRL. As a result, each cluster has its own DE gene list. Statistically significant DE genes are displayed, with gene expression proportions among two groups, log fold-change value, p-values from two-sample proportional

test. Each tab represents one cluster. Multiple testing adjust has been conducted on p-values to reflect the FDR.

Supplementary Table 8: Seurat cluster Ontology
GO ontology analysis of 14 Seurat clusters. The GO analysis of each cluster was done with PANTHER Overrepresentation Test with BONFERRONI correction.

Supplementary Table 9: PI3K pathway DEGs in clusters
Two genes, CCND2 and FOXO3 related with PI3K pathway among statistically significant DE genes are presented with gene expression proportions among two groups, log fold-change value, and p-values from two-sample proportional test. Multiple testing adjust has been conducted on p-values to reflect the FDR.

Supplementary Table 10: pseudotime trajectory cluster marker
List of marker genes for fourteen clusters using Monocle3. For pseudo-time trajectory analysis purpose, we identified cell clusters based on Minimal Spanning Tree (MST) and their corresponding pseudo-time cluster marker genes. Each row represents one gene, and the pseudo-time cluster membership information is shown in column "cell_group". The characteristics and significance are shown in rest columns. Two-sided p-values are generated from the likelihood ratio test (LRT) using a generalized linear model in Monocle3. Both original p-values and multiple testing adjusted q-values are reported. Q-values were adopted to control for the positive false discovery rate (pFDR).

Supplementary Table 11: eCLIP_targets
FMRP targets from eCLIP-seq experiment. FMRP human targets are identified from 3 human organoids IP samples versus 1 Input and 1 IgG sample. FMRP mouse targets are identified from 3 mouse fetal brain IP samples versus 1 Input and 1 IgG sample. Mouse targets have been mapped to human symbols. "Both_Human_Mouse_FMRP_eCLIP_target" is the joint set of human and mouse FMRP targets. "Human_FMRP_eCLIP_target" is all the human FMRP targets. "Mouse_FMRP_eCLIP_target" is all the mouse FMRP targets.
"Human_specific_FMRP_eCLIP_target" is targets identified in human but not in mouse.
"Mouse_specific_FMRP_eCLIP_target" is targets identified in mouse but not in human.
"Human_Mouse_shared_FMRP_targets" is the targets shared by human and mouse experiments.

Supplementary Table 12: Overlaps_Organoids_CHD2
Differentially expressed genes overlapped by organoid experiment and by CHD2-KO mice RNA-seq as well as CHIP-seq experiment. "DE overlap ORG vs Chd2 mice" is the overlapped differentially expressed genes identified in organoid day 56 (N = 2 FXS organoids versus 2 controls) and CHD2-KO mice reported in Kim et al. Neuron (2018). "DE overlaps vs CHD2 chip" is the overlaps of DEG in organoid day 56, CHD2-KO mice experiment and CHD2 binding sites identified by CHIP-seq dataset downloaded from https://www.encodeproject.org/search/?type=Experiment&target.uuid=b683cf04-2ee7-469e-8124-c072c5df49e3&status=released&status=submitted&status=in+progress&limit=all.