Methyl-CpG binding domain 2 (Mbd2) deficiency causes cognitive, social and emotional functional deficits

Elad Lax[1], Sonia Do-Carmo[1], Yehoshua Enuka[2], Niaz Mahmood[3], Shafaat Rabbani[3], Liqing Wang[6], A. Claudio Cuello[1,4,5], Wayne W. Hancock[6], Yosef Yarden[2], Moshe Szyf[1,7 *]

1 Department of Pharmacology and Therapeutics, Faculty of Medicine, McGill University, Montreal, QC, Canada.

2 Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel.

3 Department of Medicine, McGill University Health Center, Montreal, QC, Canada.

4 Department of Anatomy and Cell Biology, McGill University, Montreal, QC, Canada.

5 Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada.

6 Division of Transplant Immunology, Department of Pathology and Laboratory Medicine, and Biesecker Center for Pediatric Liver Diseases, Children's Hospital of Philadelphia and Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

7 Sackler Program for Epigenetics & Psychobiology, McGill University, Montreal, QC, H3G1Y6, Canada.

*Corresponding author

Corresponding author:

Moshe Szyf, Department of Pharmacology and Therapeutics, McGill University, 3655 Sir William Osler Promenade, Montreal, QC, Canada. moshe.szyf@mcgill.ca

Abstract

Methyl-CpG-binding domain 2 (Mbd2) binds methylated DNA and was shown to play an important role in cancer and immunity but its role in brain function is mostly unknown. Here we show that MBD2 deficiency in mice results in deficits in cognitive, social and emotional functions. Chromatin immunoprecipitation (ChIP-seq) revealed that Mbd2 binds to regulatory DNA regions of neuronal genes in the hippocampus. Loss of Mbd2 alters the expression of hundreds of genes with a robust down-regulation of neuronal gene pathways. Genes down-regulated by loss of Mbd2 are highly enriched for ortholog ASD risk genes Further, a genome-wide DNA methylation analysis found an altered DNA methylation pattern in regulatory DNA regions of neuronal specific genes in Mbd2 deficient mice. Taken together, our data suggest Mbd2 regulates hippocampal DNA methylation landscape to control neuronal gene expression and behavior.

Introduction

Epigenetic modifications of the genome are long known to play a crucial role in normal brain function and a wide range of neuropsychological functions and neuropsychological disorders (Sweatt 2016). The most studied epigenetic modification is DNA methylation, the addition of a methyl group to the DNA on a cytosine in a CpG dinucleotide context. DNA methylation in promoters and other regulatory regions suppress gene expression by interfering with transcription factors and transcription machinery binding (Weber, Hellmann et al. 2007, Lister, Pelizzola et al. 2009). An additional mechanism involves recruitment of members of a family of proteins Methylated DNA binding proteins (MeCP2 and MBD1-6) which share a Methyl-CpG Binding Domain (MBD) (Jorgensen and Bird 2002, Du, Luu et al. 2015). MeCp2 and MBD2 were shown to recruit chromatin repressive complexes to genes and thus cause alterations in histone modification and silencing of gene expression (Nan, Campoy et al. 1997, Ng, Zhang et al. 1999, Baubec, Ivanek et al. 2013). These proteins are highly expressed in brain tissues (Hendrich and Bird 1998, Fan and Hutnick 2005).

The most extensively studied MBD protein is MeCP2 since mutations and duplications of this gene cause Rett syndrome (Amir, Van den Veyver et al. 1999, Moretti and Zoghbi 2006). Some studies on the role of MBD1 suggest it has a role in neurodevelopment and neurodevelopmental disorders like autism (Zhao, Ueba et al. 2003). In contrast, little is known about the roles of other MBD proteins in the brain.

Several studies associated MBD2 with neuropsychiatric disorders. Polymorphisms in MBD2 genes were associated with schizophrenia (Xie, Yu et al. 2014). An increased MBD2 DNA binding on the promoters of *BDNF*, *GAD1* and *RELN* genes was observed in post-mortem brains of schizophrenia and bipolar disorder patients (Dong, Ruzicka et al. 2015). One study associated polymorphism of MBD2 gene to rare cases of autism (Li, Yamagata et al. 2005) and several studies found Copy Number Variants at and around the genomic position of the MBD2 gene (18q21.2) in Autism Spectrum Disorder (ASD) patients (for details see SFARI gene project CNV

database: https://gene.sfari.org/database/cnv/). Recently, we found MBD2-bound DNA regions to become hypo-methylated in the prefrontal cortex of depressed- suicide completers (Nagy, Suderman et al. 2015). Further, in an animal model for maternal maltreatment, we found maltreated rat pups to show reduced MBD2 expression in the hippocampus which led to reduced glucocorticoid receptor expression and elevated stress (Weaver, Hellstrom et al. 2014). However, to date the mechanisms by which MBD2 affect gene-expression and ultimately brain function and behaviors are unknown.

In the present study, we directly assessed the role of MBD2 in behavior using a knock-out (KO) mouse model. A comprehensive behavioral battery found cognitive, social and emotional deficits in MBD2 knock-out mice. Since the behavioral abnormalities pointed to hippocampal functions we examined the molecular footprints of MBD2 in the hippocampus. We applied unbiased genome-wide approaches. Using ChIP-seq with an MBD2 specific antibody we mapped the genome binding sites of MBD2 and found that it binds methylated and unmethylated CpGS on and around many neuronal genes in the hippocampus as well as other genomic locations. Loss of MBD2 binding in KO mice led to down-regulation of neuronal-genes. In contrast, genes that are not specific to brain were activated by MBD2 deficiency. MBD2 deficiency led to increased methylation on promoters and enhancers of neuronal genes, suggesting that MBD2 maintains DNA hypomethylation status on neuronal genes and hence hippocampal genome functions. Furthermore, we found that genes down-regulated by loss of Mbd2 are highly enriched for ortholog ASD risk genes implying an important role for Mbd2 in neurodevelopment and neuropsychiatric disorders.

Methods

Animals

The MBD2 KO line was created in the lab of Adrian Bird (Hendrich, Guy et al. 2001). Heterozygous (-/+) mice were provided by Adrian Bird and were bred in McGill University animal facility. Mice were housed in groups (3-5/cage) with 12-h light/dark cycles under conditions of constant temperature (23C) and humidity (50%), and free access to food and water. All behavioral tests were performed on male and female mice. All procedures were

carried out in accordance with the guidelines set down by the Canadian Council of Animal Care and were approved by the Animal Care Committee of McGill University.

Behavioral procedures

Open-Field Test. Mice were placed in an open field (45×45 cm) apparatus with 30-cm-high walls. Their locomotor activity and exploration behaviors were measured for 5 min. The open-field test was also used as a habituation session for the object-recognition and object-location tests.

Object-Location Memory (OLM). Each mouse was placed for 5 min in the open field with two identical objects (A1 and A2) positioned in two adjacent corners. Exploration was defined as sniffing or touching the object with the nose and/or forepaws. The exploratory preference for each object was calculated as time (t) spent in exploring that object expressed as the percentage of total exploration time [(tA2/(tA1+tA2) x100]. For the long-term memory (LTM) test conducted 24h after the training session, the same mouse was allowed to explore the field for 5 min in the presence of the same objects in the same settings with the exception that one of the objects was displaced to a novel location. Visual cues (adhesive tape) were placed on two adjacent walls as described before (Heyward, Gilliam et al. 2016).

Social Interaction Test. The test was performed as described before (Sato, Kasai et al. 2012). This test was selected based on its superior sensitivity over the classical three-chamber social approach test.
The duration of the test was 5 minutes. Sniffing, close following, and allo-grooming were defined as social interaction. No aggressive behaviors were found during the tests.

Spontaneous Alternations in Y Maze. The Y mice was a standard 30 x 6 x 15 cm (Length x Width x Height) grey apparatus. Mice were allowed to freely explore the maze for 5 minutes. An entry to an arm was considered when all four paws were inside the arm. The measures included spontaneous alteration performance (i.e., a successful triad), alternate arm returns (AAR) and same arm returns (SAR). Total number of entries was assessed as well.

Dark-Light-Box. Anxiety-related behavior of the mice was tested in the Dark-Light Box consisting of two chambers connected by an open gate. Mice were placed near the gate in the

light chamber facing the dark chamber. Latency to first entry to the light side, number of exits and total time in the aversive light compartment were recorded for 5 min.

Chromatin-Immunoprecipitation Followed by Sequencing (ChIP-Seq)

For all ex-vivo experiments, mice (8 weeks of age) were sacrificed, the brains were rapidly removed and bi-lateral hippocampi were isolated and flesh frozen for later analysis. For Chromatin-Immunoprecipitation, hippocampi were homogenized in 1 X PBS including 1% formaldehyde, and the homogenates were kept for 10 min at 25°C. Cross-linking reactions were stopped by the addition of glycine (125 mM) for 10 min at 25°C. Fixed chromatin samples were then homogenized in cell lysis solution (PIPES 5 mM (pH 8), KCl 85 mM, NP40 0.5%) and centrifuge for 5 min at 3000 rpm, 4°C. Pellets were resuspended in RIPA-light solution (NaCl 150 mM, SDS 0.3%, Tris-HCl 50 mM (pH 8)) and sonicated using a covaris E220. Sonicated chromatin samples were then centrifuged for 15 min at 14000 rpm, 4°C. Pellets were resuspended in 1 ml of RIPA-light solution. Chromatin samples were pre-cleared with 50 µl of dynabeads protein G (Life Technologies) pre-blocked with BSA and incubated overnight at 4°C with an anti- MBD2 (Epigentek-A1007) antibody. Antibodies and chromatin were then mixed with 100 µl of dynabeads protein G for 3 hours at 4°C. The beads were then washed with RIPA-light solution, washing-solution (Tris-HCl 100 mM (pH 8), LiCl 500 mM). Protein–DNA complexes were eluted from the beads, de-cross-linked, treated with proteinase K and purified. The DNA concentration was determined by fluorometry on the Qubit system (Invitrogen). DNA samples from 8-10 mice per group were pooled, a total of 10–12 ng DNA were used for the preparation of the library. The immunoprecipitated DNA and input DNA were sheared a second time with the Covaris E220 instrument in 53 µl reaction volume (duty factor 10%, Pic Incident Power 175, Cycles per burst 200, time 360 sec) to obtain fragments in the size range of 150 bp followed by purification with AMPure XP beads (×1.8v/v) (Beckman Coulter A63881). Purified DNA was resuspended in 45 µl elution buffer. A library of the chromatin immunoprecipitated DNA fragments was prepared using the Tru Seq DNA Low Throughput Protocol (Illumina). PCR enrichment of

ligation products was performed using the Illumina Primer Cocktail; 15 cycles of PCR were performed for ChIP libraries and 10 cycles for the input. The libraries were purified using AMPure XP beads ×1.0 v/v. Quality of libraries was validated by 260 nm absorbance measurement, quality control on HSdna chip (Agilent Bioanalyzer: size of libraries around 275bp bp) and quantification by Q-PCR with Kappa Library Quantification kit for Illumina Sequencing Platforms (KAPPA Biosystems). The DNA concentration of the different sequencing libraries was from 40 to 500 nM. Clusters (13.5 pM) were generated using TruSeq PE Cluster Kit v3, for cBot protocol, which was followed by 50bp pair--end Sequencing, on an Illumina HiSeq 2000, per the manufacturer recommendations.

Bioinformatic Analysis of the ChIP-Seq Data

Read quality for all next-generation sequencing experiments was assessed using Fastqc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) confirmed high read quality and inconsequential levels of adapter contamination. Reads were aligned to the mouse reference genome (mm9 assembly) using Bowtie (Langmead 2010) with default parameters. Low quality alignments and alignments for read pairs with multiple possible genomic alignments were omitted. Duplicate read pairs were removed. Read count peaks corresponding to likely binding sites were identified for each sample by MACS2 (Liu 2014) with default parameters. Peaks with false discover rates less than 0.05 were selected as high-confidence binding sites. Sites were annotated with their genomic locations relative to nearby genes using HOMER with default parameters. De-novo motif discovery was done with HOMER with background adjusted to CpG context.

ChIP-QPCR

Hippocmpi from 5-6 mice per group as biological replicates for validation of ChIP-seq were subjected to the same ChIP protocol used for the ChIP-Seq. To exclude the possibility of idiosyncrasy of the anti-MBD2 antibody used in the ChIP-seq experiment, we used another antibody for this experiment (Imegenex, IMG147) and IgG antibody (Santa Cruz) served as an additional control. For RNApolii (ser5) and H3K4me1 ChIP experiments antibodies from Abcam

were used (ab5408 and ab8895, respectively). Purified DNA was resuspended in 40ul for QPCR analysis using SYBR green quantitative PCR was performed using the LightCycler® 480 system (Software 3.5, Roche Molecular Biochemicals). To determine the relative enrichment, the 2 - ΔΔCt method was used with normalization to IgG and input data.

RNA-Sequencing and Data Analysis

DNA and RNA from 8-10 mice was isolated and purified with AllPrep-DNA/RNA/miRNA-universal kit (Qiagen) and concentrations were determined by fluorometry on the Qubit system (Invitrogen). Ribosomal RNA was removed using the Ribominus kit (Invitrogen). cDNA libraries (4 per group) were generated and sequenced using an Illumina Hiseq 2500 (100 bp pair-end runs), as instructed by Illumina's RNA-seq protocols. Reads were deduplicated as described above. Reads were aligned to the mouse reference transcriptome (mm9) using the STAR aligner (Dobin, Davis et al. 2013). Differential expression was analyzed using DeSeq2 with an FDR of 0.05 (Love, Huber et al. 2014).

Human Genetic Data Analysis

Human data of Loss-of-Function and damaging missense variants found in autism syndrome disorder (ASD) were extracted from previous study (De Rubeis, He et al. 2014). Data were divided into a subset of top 33 genes which were considered as “ASD risk genes” (FDR<0.1) and a full set of 107 genes which were considered as “likely to affect risk” (FDR<0.3). In addition, data from the SFARI GENE project (gene.sfari.org) were extracted. This database contains gene-list of ASD associated genes based on human and animal model studies. Genes are grouped according to criteria for the strength of evidence for each gene (a total of 970 genes as of 23 November 2017). We excluded 19 genes categorized in category 6 for genes studied in human cohorts and findings suggest against a role in autism, having a total of 951 ASD-associated genes from this database. Next, we assessed the overlap between ASD risk genes from each of the databases to the ortholog mouse genes which were differentially expressed in Mbd2 KO mice. Significance was determined by a hypergeometric test.

Quantitative polymerase chain reaction (qPCR)

RNA was extracted from 5-9 mice per group as biological replicates for validation of RNA-seq cDNA was prepared with random hexamer primers (Invitrogen) and a reverse transcription kit (NEB) according to the manufacturer protocol. GAPDH was used as the reference gene. SYBR green quantitative PCR (qRT-PCR) was performed using the LightCycler® 480 system (Software 3.5, Roche Molecular Biochemicals). To determine the relative concentration of mRNA expression, the 2 -ΔCt method was used.

Capture bisulfite sequencing and DNA methylation mapping

The SeqCap Epi Enrichment System (Roche NimbleGen) for targeted bisulfite sequencing of promoters and enhancers in the mouse genome was used. Mouse target probes (mm9) were custom designed based on H3K4me1 and H3K4me3 signals from mouse public ChIP-seq data. A total of 83276 regions were targeted by the probes which spans 64760111 base pairs. Biotinylated target probes were designed for both strands of bisulfite converted genomic DNA. Bisulfite treated genomic DNA was ligated to methylated next generation sequencing (NGS) adaptors, hybridized to the biotinylated oligonucleotide probes followed by a series of washes of off-target DNA sequences and unbound DNA. Isolated DNA then underwent PCR amplification and sequenced on Illumina Hiseq 2000 with pair-end 50bp reads and a technical repeat with 125bp pair-end reads was performed as well.

Analysis of Differentially methylated cytosines

Sequences were aligned to the mm9 mouse reference genome using Bsmap v2.89 (Xi and Li 2009). Output data were strand-sorted, filtered, and deduplicated with Picard tools. Next, methylation levels and coverage levels were extracted with methratio.py command in Bsmap. Differential methylation cytosines were analyzed with methylKit R package (Akalin, Kormaksson et al. 2012) with FDR threshold of 0.1 and at least 10% difference in methylation levels. Differentially methylated positions were annotated with HOMER (Heinz, Benner et al. 2010).

Statistical Analysis

Comparisons between two groups in behavioral and molecular experiments were done with two-tailed t-test. qPCR results for differentially expressed genes were analyzed as single-group

T-test for fold change over control and one-tailed t-test, based on the a-priori assumption to find differences in the same direction observed by the preceding RNA-seq experiment. For correlation analysis between DNA methylation and gene expression, the Pearson's correlation coefficient was calculated (differential methylation >25% and log2 fold-change gene-expression > 0.27). Other statistical analyses of bioinformatic data are detailed in the results section. Heatmaps were generated with Morpheus (https://software.broadinstitute.org/morpheus/ ; Broad Institute). GO annotations, pathway enrichment analysis and gene-networks were analyzed with Metascape (Tripathi, Pohl et al. 2015) (http://metascape.org).

Results

Mbd2 is required for a range of cognitive, social and emotional behaviors

We investigated whether *mbd2* plays a role in behavioral phenotypes using a battery of behavioral test in *mbd2-/-* mice and control WT mice. There was no gross change in general locomotion in an open-field box (Fig 1A) or in other exploratory behaviors (speed, entries to the center of arena, time near walls; Fig S1). Several forms of memory were explored as well. Short-term working memory was assessed by spontaneous alternations in Y-Maze and was found to be intact in KO mice (Fig 1E). In contrast, *mbd2-/-* showed impaired memory retention in the long-term object-location memory test (Fig1B). *mbd2-/-* mice also exhibited reduced social-interaction time and increased anxiety in the dark-light box (Fig 1C-D). Taken together, these findings suggest that Mbd2 is involved in regulating a range of cognitive, social and emotional behaviors.

Landscape of MBD2 binding in the hippocampus

We focused our analysis on the hippocampus since the behaviors affected are known to involve hippocampal functions (Heyward, Walton et al. 2012, Mineur, Obayemi et al. 2013, Yang, Zou et al. 2013, Tavares, Mendelsohn et al. 2015, Heyward, Gilliam et al. 2016). In addition, other members of the MBD protein family were shown to have an important role in hippocampal function (Zhao, Ueba et al. 2003, Moretti, Levenson et al. 2006, Camarena, Cao et al. 2014, Lu, Ash et al. 2016). MBD2 ChIP-seq experiment from hippocampi of WT mice revealed 2,436 Mbd2

peaks annotated to 1,222 genes. As expected, Mbd2 binds mostly to CpG-containing and GC-rich DNA regions (X^2 = 43.364; df=1, p=4.54E-11 and X^2 = 1129.7; df=1, p=2.2E-16 respectively). Highly-enriched peaks were observed in proximity to transcription start sites (Fig2A, Spearman rho=-0.13, p=2.4E-11, for absolute distance from TSS). Highly-enriched peaks were also correlated with peak length (Spearman rho=-0.489, p=2.2E-16, data not shown). Interestingly, de-novo motif discovery found the transcription factor E2F7 and other E2F family members to be highly enriched in hippocampal Mbd2 peaks (Fig2B and FigS2; p=1E-197). The E2F family has been found before to have an important role in neurogenesis and brain development (Ghanem, Andrusiak et al. 2012, de Bruin, PW et al. 2016). We also compared our data to publicly available ChIP-Seq data of the mouse hippocampus histone-marks (Gjoneska, Pfenning et al. 2015). As expected, Mbd2 bound overall to more repressive histone-marks (mostly H3K9me3) than to active histone marks. However, a subset of the peaks was bound to both active and repressive histone marks implying Mbd2 binds the same DNA regions in different cell populations within the hippocampus potentially participating in chromatin remodeling (Fig2C-D). Mbd2 co-occupancy with histone marks on promoters and gene bodies (mostly exons) was enriched for active histone mark and depleted from repressive histone marks (Fig2E). Pathway analysis by GO enrichment found that Mbd2 binds many neuronal- and brain- related pathways such as: trans-synaptic signaling, synapse organization and behavior (Fig2F). These pathways also showed good clustering as GO networks (FigS3) and protein-protein interactions analysis found forebrain development and axon guidance as the top 2 annotations (log10p=-6.3 and -5.6, respectively) further supporting a role for Mbd2 in neuronal function.

We selected 15 peaks of Mbd2 that were associated with promoters in neuronal-related genes for further validation and for determining the impact of loss of binding on transcription initiation (see Table S1 for genomic coordinates of validated peaks). Quantitative ChIP confirmed binding of Mbd2 to these regions and loss of binding in the *mbd2-/-* mice (Q-ChIP, Fig2G). We then determined whether loss of binding of Mbd2 in promoters or enhancers in *mbd2-/-* mice affects transcription onset. Mbd2 deficiency resulted in reduced transcription onset since occupancy of RNA polymerase phosphorylated on Serine5; RNApolii(ser5)) the form found on promoters upon transcription initiation (Hirose and Ohkuma 2007, Jonkers and Lis

2015) was reduced in *mbd2-/-* mice. Interestingly, this was associated with increased abundance of histone mark (H3K4me1) marking enhancers in *mbd2-/-* (Fig2H-I). H3K4me1 modification marks active as well as inactive enhancers (Heintzman, Hon et al. 2009, Creyghton, Cheng et al. 2010). H3K4me1 peaks that are flanking the TSS were shown to exhibit a peak-valley-peak pattern. The valley usually overlaps with transcription initiation and RNApolII-PS5 peaks (Pundhir, Bagger et al. 2016) reflecting a nucleosome free zone and thus reduced histone presence and reduced signal for the histone marks, which is prerequisite for transcription initiation (McGhee, Wood et al. 1981). In transcriptionally inactive genes H3K4me1 peaks cover the entire TSS forming one continuous peak centered at the TSS and overlapping with the RNApolPS5 peaks. The increase of H3K4me1 concurrently with reduction of RNApolII-PS5 binding at the same position in *mbd2-/-* mice is consistent with inhibition of transcription onset in response to *mbd2-/-*. These data suggest that although overall Mbd2 binding is associated with repressive chromatin marks, Mbd2 is involved in activation of transcription turn on in these neuronal specific promoters.

Mbd2 is required for expression of neuronal specific genes

To further elucidate the role of Mbd2 in gene expression in the hippocampus we delineated the transcriptomes of WT and *mbd2-/-* using RNA-Seq. We found 2907 genes to be differentially expressed (FDR<0.05), of which 1590 genes were up-regulated and 1,317 gene were down-regulated in *mbd2-/-* mice (Fig3A, and FigS4 for validations). Pathway analysis by GO enrichment found a robust down-regulation of neuronal-related pathways such as neuron projection development, trans-synaptic signaling and behaviors (Fig3B) which were highly organized in clusters as further supported by clustering analysis of the GO networks (Fig3C). In strike contrast, the same analyses applied to up-regulated genes revealed mostly homeostasis, metabolism and cell death related pathways with no enrichment of neuronal-related pathways. GO network analysis revealed poor clustering of the differentially expressed genes in these pathways (Fig3D-E).

Next, we determined whether the differentially expressed genes were interacting with Mbd2 by overlapping the Mbd2 ChIPseq and RNAseq results. A significant fraction of the down-

regulated genes had also Mbd2 peaks annotated to them (85 genes; p=1.8E-9; hyper-geometric test) while fewer up-regulated genes (55) had Mbd2 peaks annotated to them (p=0.31; hyper-geometric test). Meta-analysis of the overlapping ChIP-seq and RNA-seq enriched pathways by unsupervised clustering found that the top downregulated pathways in *mbd2-/-* were trans-synaptic signaling, behavior and synaptic organization (Fig3F-G). These findings are consistent with a role for Mbd2 in regulating neuronal-specific gene expression

The impact of *mbd2* depletion on the DNA methylation landscape

MBD proteins are "readers" of DNA methylation and bind specifically to methylated DNA (Du, Luu et al. 2015, Ludwig, Zhang et al. 2016). However, it is possible that they also play role in maintaining DNA methylation/demethylation landscapes as has been previously proposed. Previous studies in cancer cells and in T regulatory cells have suggested that depletion of Mbd2 causes hypermethylation of regulatory regions which in turn results in downregulation of gene expression (Wang, Liu et al. 2013). Therefore, we tested whether Mbd2 deficiency would alter the landscape of DNA methylation in the hippocampus. We mapped with targeted bisulfite sequencing the state of methylation of regulatory DNA regions (promoters and enhancers, see methods) in the hippocampi of *mbd2-/-* and WT mice at a single-base resolution. A global methylation analysis revealed that Mbd2 binding regions defined in WT mice by ChIPseq peaks were hypermethylated in *mbd2-/-* mice (Fig 4A; p<2.2E-16, K-S test). Hypermethylation in response to Mbd2 depletion was observed also in a genome-wide scale analysis which examined DNA outside the regions that bind Mbd2 (p<2.2E-16, K-S test, data not shown). The fact that changes in DNA methylation occur in regions that don't bind Mbd2 suggests that loss of Mbd2 could affect DNA methylation indirectly. Taken together these data suggest that Mbd2 protects sequences from hypermethylation and maintains them in a demethylated state. However, loss of Mbd2 affects DNA methylation in both directions. An analysis of the sequence properties of Mbd2 dependent DNA methylation suggests that Mbd2 deficiency affects low to intermediate methylated CpGs (10-40% methylation) but highly methylated CpGs (>40% methylation) are unaffected as expected if Mbd2 function is to prevent hypermethylation. Partially methylated genes represent a group of genes that are heterogeneously methylated

and possibly heterogeneously expressed across hippocampal nuclei. Mbd2 might be regulating the state of methylation of these genes.

We also analyzed differential methylation at single CpG resolution. At this resolution differential methylation analysis (at least 10% difference) revealed 361 differentially methylated CpGs, with 151 hypo-methylated CpGs (annotated to 117 genes) and 210 hyper-methylated CpGs (annotated to 127 genes) (differential methylation data ($p<0.001$) can be found at Table S2). This finding also supports a significant overall hypermethylation in *Mbd2-/-* mice hippocampus ($p=0.0022$, binomial test). For pathway analysis, we applied a more lenient significance cut-off for our data ($p<0.001$) resulting in 3005 differentially methylated CpGS (1519 hypo-methylated and 1486 hyper-methylated). Next, we focused the analysis on CpGs located on promoters (between -1000bp to + 1000bp from TSS). We found 494 hypo-methylated CpGs located in 460 gene-promoters and 476 hyper-methylated CpGs located in 434 gene promoters (Fig4B). Pathway analysis for differentially methylated gene-promoters revealed adrenergic receptor signaling and sodium- and metal- ion transports as the top three hyper-methylated pathways (-log10p values: 4.86, 4.16 and 3.42, respectively). In contrast, the top hypo-methylated pathways were not specifically related to neurons or brain functions, although neuronal ion channel clustering and positive regulation of sodium ion transport pathways were enriched (-log10p values: 3.289 and 3.280, respectively) (Fig4C-D). Overall, this analysis suggests an overall enrichment of hyper-methylated genes related to neuronal system in *mbd2-/-* mice hippocampi.

To assess the relation between promoter DNA methylation and gene-expression we calculated a methylation-expression correlation. We found a significant linear inverse correlation between promoter DNA methylation and gene expression (Fig4E), supporting a role for DNA methylation in gene-repression in the hioppocampus.

Genes whose expression is reduced in *mbd2-/-* are associated with ASD risk genes

Other MBD proteins were associated with ASD (Zhao, Ueba et al. 2003, Moretti and Zoghbi 2006, Cukier, Lee et al. 2012, Camarena, Cao et al. 2014, Lu, Ash et al. 2016). The presentation

of cognitive deficits in *mbd2-/-* mice described above resembles some of the observed symptoms of ASD. CNVs that were shown to associate with ASD include those that map onto genomic regions of the *Mbd2* gene. We further explored whether differentially expressed genes due to loss of Mbd2 in the mouse hippocampus are associated with ortholog human "risk genes". Comparative analyses of our RNA-seq data with that of two human ASD databases ((De Rubeis, He et al. 2014)) and SFARI-gene (see methods), found a robust overlap between ASD risk genes and down-regulated genes in Mbd2 KO mice. Eight (24.24%) out of 33 top loss-of-function ASD risk-genes (FDR<0.1) and 17 genes (15.88%) out of 107 ASD risk-genes (FDR<0.3) found by (De Rubeis, He et al. 2014) were down-regulated in *mbd2-/-* (p=0.0014 and p=0.00095 respectively, Fig5A-B and table S3). In contrast, only 1 gene from the 33 gene-list and 5 genes from the 107 gene-list were up-regulated in *mbd2-/-* (p=0.942 and p=0.947, respectively). Next, we compared our RNA-seq results to the SFARI-gene list of ASD-associated genes which contains all known human genes associated with autism spectrum disorders. Here again, out of 951 human genes associated with ASD, 125 (13.14%) were down-regulated in *mbd2-/-* (p=8.8E-13, Fig6) while only 44 were up-regulated in *mbd2-/-* (4.62%, p=0.99, Fig6 and table S3). These cross-species comparative analyses suggest that Mbd2 might serve as an up-stream activator to many ASD-associated genes as has previously been shown for liver, breast and prostate cancer genes in cancer cell lines (Shukeir, Pakneshan et al. 2006, Stefanska, Suderman et al. 2013, Cheishvili, Chik et al. 2014) NGFIA in hippocampal cells (Weaver, Hellstrom et al. 2014)and *Foxp3* gene (Wang, Liu et al. 2013) in regulatory T-cells.

Discussion

There is growing evidence for the crucial role of epigenetic mechanisms in neuropsychological disorders and CNS function. In this study, we explore the role of Mbd2, a methylated-CpG binding protein, in gene-expression and brain function. Mbd2, like other MBD proteins, serves as a "reader" of the epigenome (Du, Luu et al. 2015) translating DNA methylation marks into gene-expression regulation mechanisms. We found that loss of Mbd2 resulted in several forms

of behavioral deficits: cognitive, social and emotional, implying a rather broad role for this protein in normal brain function.

Mbd2 regulates gene expression by binding to methylated CpGs in DNA (Berger and Bird 2005). Mbd2 was previously shown to be involved in both suppression of promoters through recruitment of transcriptional repressors, activation of promoters in cancer cells and the hippocampus through recruitment of transcriptional activators (Shukeir, Pakneshan et al. 2006, Stefanska, Suderman et al. 2013, Cheishvili, Chik et al. 2014, Weaver, Hellstrom et al. 2014) such as CBP and NGFIA ((Angrisano, Lembo et al. 2006, Stefanska, Suderman et al. 2013, Weaver, Hellstrom et al. 2014) as well as targeting DNA demethylation (Detich, Theberge et al. 2002, Cui and Irudayaraj 2015)(Stefanska; Detich, Theberge et al. 2002) possibly through recruitment of demethylating activities such as Tet2 (Wang, Liu et al. 2013) . Since the effects of Mbd2 deficiency on behavior were broad and relating to hippocampal function (Heyward, Walton et al. 2012, Mineur, Obayemi et al. 2013, Yang, Zou et al. 2013, Tavares, Mendelsohn et al. 2015, Heyward, Gilliam et al. 2016) , we used genome wide approaches to define the landscape of MBd2 binding in normal animals and then determined how its deficiency affects the landscapes of DNA methylation and gene expression using *mbd2-/-* mice. Consistent with an important role in brain function detected by the behavioral assays presented here, pathway analyses revealed a highly clustered and networked enrichment of genes relating to cognitive functions and brain development such as trans-synaptic signaling and synapse organization and behavior. Consistent with previous data we found evidence for the bimodal function of Mbd2. Mbd2 deficiency resulted in both gene activation and repression as well as hypermethylation and hypomethylation. Interestingly, the neuronal specific functions seem to be mainly repressed and hypermethylated by Mbd2 deficiency, suggesting a mostly activating role for Mbd2 on neuronal specific genes. This is supported by the observation that Mbd2 localizes to promoters.

To further examine the role of Mbd2 in transcriptional activation/repression, we examined selected several promoters and enhancers that bind Mbd2 and examined the consequences of

Mbd2 deficiency on transcription initiation. The majority of the 15 genes examined had reduced RNApolII-PS5 occupancy and increased H3K4me1 coverage of the transcription initiation region indicating reduced transcription initiation upon Mbd2 deficiency.

Previous studies suggested that MBD2 might be involved in psychiatric disorders (Li, Yamagata et al. 2005, Xie, Yu et al. 2014, Dong, Ruzicka et al. 2015), our data offers a possible mechanism. Interestingly, several of the genes that we found here to be Mbd2 dependent were associated with ASD in GWAS studies. Indeed, there is a significant enrichment of ASD associated genes in the list of genes that are repressed by Mbd2 deficiency. The behavioral phenotypes that we characterize here for Mbd2 deficient mice are also seen in people with ASD. Our data points to a potential role for Mbd2 in ASD.

MBD proteins are generally considered to have a repressive role on gene-expression by interacting with chromatin modification inactivating complexes (Nan, Cross et al. 1998). Therefore, it is somewhat surprising that we found an overall repressive effect on neuronal specific genes upon *mbd2-/-* deficiency. However, more recent evidence suggest other MBD proteins are also involved in gene-expression in more than one way. For example, hypothalamic MeCP2 dysfunction led to robust changes in gene-expression with 85% of the genes to be activated by MECP2, possibly mediated by interaction of the transcriptional activator CREB1 and MECP2 in binding gene promoters and regulatory regions (Chahrour, Jung et al. 2008). Similarly, others have showed that in regulatory T-cells (Treg) MeCP2 promotes the expression of Fopx3, a key regulator of Treg function, by collaborating with CREB1 (Li, Jiang et al. 2014). Neuronal MBD1 was also found to have a mixed effect on gene-expression in the hippocampus with more than half of the genes down-regulated in Mbd1-/- mice, many of them associated with neurogenesis (Jobe, Gao et al. 2017).

Taken together, our findings are in line with evidence from other studies showing the importance of MBD proteins in neuropsychological disorders. Our findings provide evidence that Mbd2 is a regulator of neuronal gene-expression, brain physiology and behavior.

Acknowledgements

The work was funded by a grant (PSR-SIIRI) from the Ministère du Développement économique, de l'Innovation, Quebec Government. EL was funded by the Richard and Edith Strauss Canada Fund post-doctoral fellowship. The authors wish to thank Dr. David Cheishvili for critical discussion. Capture and hybridization was performed by the Institut de recherches cliniques de Montréal (Montreal, Canada; licensed by Roche Nimblegen). Next-generation sequencing was performed in Genome-Quebec (Montreal, Canada; licensed by Illumina).

References

Akalin, A., M. Kormaksson, S. Li, F. E. Garrett-Bakelman, M. E. Figueroa, A. Melnick and C. E. Mason (2012). "methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles." Genome Biol **13**(10): R87.

Amir, R. E., I. B. Van den Veyver, M. Wan, C. Q. Tran, U. Francke and H. Y. Zoghbi (1999). "Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2." Nat Genet **23**(2): 185-188.

Angrisano, T., F. Lembo, R. Pero, F. Natale, A. Fusco, V. E. Avvedimento, C. B. Bruni and L. Chiariotti (2006). "TACC3 mediates the association of MBD2 with histone acetyltransferases and relieves transcriptional repression of methylated promoters." Nucleic Acids Res **34**(1): 364-372.

Baubec, T., R. Ivanek, F. Lienert and D. Schubeler (2013). "Methylation-dependent and -independent genomic targeting principles of the MBD protein family." Cell **153**(2): 480-492.

Berger, J. and A. Bird (2005). "Role of MBD2 in gene regulation and tumorigenesis." Biochem Soc Trans **33**(Pt 6): 1537-1540.

Camarena, V., L. Cao, C. Abad, A. Abrams, Y. Toledo, K. Araki, M. Araki, K. Walz and J. I. Young (2014). "Disruption of Mbd5 in mice causes neuronal functional deficits and neurobehavioral abnormalities consistent with 2q23.1 microdeletion syndrome." EMBO Mol Med **6**(8): 1003-1015.

Chahrour, M., S. Y. Jung, C. Shaw, X. Zhou, S. T. Wong, J. Qin and H. Y. Zoghbi (2008). "MeCP2, a key contributor to neurological disease, activates and represses transcription." Science **320**(5880): 1224-1229.

Cheishvili, D., F. Chik, C. C. Li, B. Bhattacharya, M. Suderman, A. Arakelian, M. Hallett, S. A. Rabbani and M. Szyf (2014). "Synergistic effects of combined DNA methyltransferase inhibition and MBD2 depletion on breast cancer cells; MBD2 depletion blocks 5-aza-2'-deoxycytidine-triggered invasiveness." Carcinogenesis **35**(11): 2436-2446.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young and R. Jaenisch (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.

Cui, Y. and J. Irudayaraj (2015). "Dissecting the behavior and function of MBD3 in DNA methylation homeostasis by single-molecule spectroscopy and microscopy." Nucleic Acids Res **43**(6): 3046-3055.

Cukier, H. N., J. M. Lee, D. Ma, J. I. Young, V. Mayo, B. L. Butler, S. S. Ramsook, J. A. Rantus, A. J. Abrams, P. L. Whitehead, H. H. Wright, R. K. Abramson, J. L. Haines, M. L. Cuccaro, M. A. Pericak-Vance and J. R. Gilbert (2012). "The expanding role of MBD genes in autism: identification of a MECP2 duplication and novel alterations in MBD5, MBD6, and SETDB1." Autism Res **5**(6): 385-397.

de Bruin, A., A. C. PW, B. C. Kirchmaier, M. Mokry, E. Iich, E. Nirmala, K. H. Liang, D. V. AM, K. T. Scholman, M. J. Groot Koerkamp, F. C. Holstege, E. Cuppen, S. Schulte-Merker and W. J. Bakker (2016). "Genome-wide analysis reveals NRP1 as a direct HIF1alpha-E2F7 target in the regulation of motorneuron guidance in vivo." Nucleic Acids Res **44**(8): 3549-3566.

De Rubeis, S., X. He, A. P. Goldberg, C. S. Poultney, K. Samocha, A. E. Cicek, Y. Kou, L. Liu, M. Fromer, S. Walker, T. Singh, L. Klei, J. Kosmicki, F. Shih-Chen, B. Aleksic, M. Biscaldi, P. F. Bolton, J. M. Brownfeld, J. Cai, N. G. Campbell, A. Carracedo, M. H. Chahrour, A. G. Chiocchetti, H. Coon, E. L. Crawford, S. R. Curran, G. Dawson, E. Duketis, B. A. Fernandez, L. Gallagher, E. Geller, S. J. Guter, R. S. Hill, J. Ionita-Laza, P. Jimenz Gonzalez, H. Kilpinen, S. M. Klauck, A. Kolevzon, I. Lee, I. Lei, J. Lei, T. Lehtimaki, C. F. Lin, A. Ma'ayan, C. R. Marshall, A. L. McInnes, B. Neale, M. J. Owen, N. Ozaki, M. Parellada, J. R. Parr, S. Purcell, K. Puura, D. Rajagopalan, K. Rehnstrom, A. Reichenberg, A. Sabo, M. Sachse, S. J. Sanders, C. Schafer, M. Schulte-Ruther, D. Skuse, C. Stevens, P. Szatmari, K. Tammimies, O. Valladares, A. Voran, W. Li-San, L. A. Weiss, A. J. Willsey, T. W. Yu, R. K. Yuen, D. D. D. Study, A. Homozygosity Mapping Collaborative for, U. K. Consortium, E. H. Cook, C. M. Freitag, M. Gill, C. M. Hultman, T. Lehner, A. Palotie, G. D. Schellenberg, P. Sklar, M. W. State, J. S. Sutcliffe, C. A. Walsh, S. W. Scherer, M. E. Zwick, J. C. Barett, D. J. Cutler, K. Roeder, B. Devlin, M. J. Daly and J. D. Buxbaum (2014). "Synaptic, transcriptional and chromatin genes disrupted in autism." Nature **515**(7526): 209-215.

Detich, N., J. Theberge and M. Szyf (2002). "Promoter-specific activation and demethylation by MBD2/demethylase." J Biol Chem **277**(39): 35791-35794.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras (2013). "STAR: ultrafast universal RNA-seq aligner." Bioinformatics **29**(1): 15-21.

Dong, E., W. B. Ruzicka, D. R. Grayson and A. Guidotti (2015). "DNA-methyltransferase1 (DNMT1) binding to CpG rich GABAergic and BDNF promoters is increased in the brain of schizophrenia and bipolar disorder patients." Schizophr Res **167**(1-3): 35-41.

Du, Q., P. L. Luu, C. Stirzaker and S. J. Clark (2015). "Methyl-CpG-binding domain proteins: readers of the epigenome." Epigenomics **7**(6): 1051-1073.

Fan, G. and L. Hutnick (2005). "Methyl-CpG binding proteins in the nervous system." Cell Res **15**(4): 255-261.

Ghanem, N., M. G. Andrusiak, D. Svoboda, S. M. Al Lafi, L. M. Julian, K. A. McClellan, Y. De Repentigny, R. Kothary, M. Ekker, A. Blais, D. S. Park and R. S. Slack (2012). "The Rb/E2F pathway modulates neurogenesis through direct regulation of the Dlx1/Dlx2 bigene cluster." J Neurosci **32**(24): 8219-8230.

Gjoneska, E., A. R. Pfenning, H. Mathys, G. Quon, A. Kundaje, L. H. Tsai and M. Kellis (2015). "Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease." Nature **518**(7539): 365-369.

Heintzman, N. D., G. C. Hon, R. D. Hawkins, P. Kheradpour, A. Stark, L. F. Harp, Z. Ye, L. K. Lee, R. K. Stuart, C. W. Ching, K. A. Ching, J. E. Antosiewicz-Bourget, H. Liu, X. Zhang, R. D. Green, V. V. Lobanenkov, R. Stewart, J. A. Thomson, G. E. Crawford, M. Kellis and B. Ren (2009). "Histone

modifications at human enhancers reflect global cell-type-specific gene expression." Nature **459**(7243): 108-112.
Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." Mol Cell **38**(4): 576-589.
Hendrich, B. and A. Bird (1998). "Identification and characterization of a family of mammalian methyl-CpG binding proteins." Mol Cell Biol **18**(11): 6538-6547.
Hendrich, B., J. Guy, B. Ramsahoye, V. A. Wilson and A. Bird (2001). "Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development." Genes Dev **15**(6): 710-723.
Heyward, F. D., D. Gilliam, M. A. Coleman, C. F. Gavin, J. Wang, G. Kaas, R. Trieu, J. Lewis, J. Moulden and J. D. Sweatt (2016). "Obesity Weighs down Memory through a Mechanism Involving the Neuroepigenetic Dysregulation of Sirt1." J Neurosci **36**(4): 1324-1335.
Heyward, F. D., R. G. Walton, M. S. Carle, M. A. Coleman, W. T. Garvey and J. D. Sweatt (2012). "Adult mice maintained on a high-fat diet exhibit object location memory deficits and reduced hippocampal SIRT1 gene expression." Neurobiol Learn Mem **98**(1): 25-32.
Hirose, Y. and Y. Ohkuma (2007). "Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression." J Biochem **141**(5): 601-608.
Jobe, E. M., Y. Gao, B. E. Eisinger, J. K. Mladucky, C. C. Giuliani, L. E. Kelnhofer and X. Zhao (2017). "Methyl-CpG-Binding Protein MBD1 Regulates Neuronal Lineage Commitment through Maintaining Adult Neural Stem Cell Identity." J Neurosci **37**(3): 523-536.
Jonkers, I. and J. T. Lis (2015). "Getting up to speed with transcription elongation by RNA polymerase II." Nat Rev Mol Cell Biol **16**(3): 167-177.
Jorgensen, H. F. and A. Bird (2002). "MeCP2 and other methyl-CpG binding proteins." Ment Retard Dev Disabil Res Rev **8**(2): 87-93.
Langmead, B. (2010). "Aligning short sequencing reads with Bowtie." Curr Protoc Bioinformatics **Chapter 11**: Unit 11 17.
Li, C., S. Jiang, S. Q. Liu, E. Lykken, L. T. Zhao, J. Sevilla, B. Zhu and Q. J. Li (2014). "MeCP2 enforces Foxp3 expression to promote regulatory T cells' resilience to inflammation." Proc Natl Acad Sci U S A **111**(27): E2807-2816.
Li, H., T. Yamagata, M. Mori, A. Yasuhara and M. Y. Momoi (2005). "Mutation analysis of methyl-CpG binding protein family genes in autistic patients." Brain Dev **27**(5): 321-325.
Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren and J. R. Ecker (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences." Nature **462**(7271): 315-322.
Liu, T. (2014). "Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells." Methods Mol Biol **1150**: 81-95.
Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.
Lu, H., R. T. Ash, L. He, S. E. Kee, W. Wang, D. Yu, S. Hao, X. Meng, K. Ure, A. Ito-Ishida, B. Tang, Y. Sun, D. Ji, J. Tang, B. R. Arenkiel, S. M. Smirnakis and H. Y. Zoghbi (2016). "Loss and Gain of MeCP2 Cause Similar Hippocampal Circuit Dysfunction that Is Rescued by Deep Brain Stimulation in a Rett Syndrome Mouse Model." Neuron **91**(4): 739-747.
Ludwig, A. K., P. Zhang and M. C. Cardoso (2016). "Modifiers and Readers of DNA Modifications and Their Impact on Genome Structure, Expression, and Stability in Disease." Front Genet **7**: 115.
McGhee, J. D., W. I. Wood, M. Dolan, J. D. Engel and G. Felsenfeld (1981). "A 200 base pair region at the 5' end of the chicken adult beta-globin gene is accessible to nuclease digestion." Cell **27**(1 Pt 2): 45-55.

Mineur, Y. S., A. Obayemi, M. B. Wigestrand, G. M. Fote, C. A. Calarco, A. M. Li and M. R. Picciotto (2013). "Cholinergic signaling in the hippocampus regulates social stress resilience and anxiety- and depression-like behavior." Proc Natl Acad Sci U S A **110**(9): 3573-3578.
Moretti, P., J. M. Levenson, F. Battaglia, R. Atkinson, R. Teague, B. Antalffy, D. Armstrong, O. Arancio, J. D. Sweatt and H. Y. Zoghbi (2006). "Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome." J Neurosci **26**(1): 319-327.
Moretti, P. and H. Y. Zoghbi (2006). "MeCP2 dysfunction in Rett syndrome and related disorders." Curr Opin Genet Dev **16**(3): 276-281.
Nagy, C., M. Suderman, J. Yang, M. Szyf, N. Mechawar, C. Ernst and G. Turecki (2015). "Astrocytic abnormalities and global DNA methylation patterns in depression and suicide." Mol Psychiatry **20**(3): 320-328.
Nan, X., F. J. Campoy and A. Bird (1997). "MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin." Cell **88**(4): 471-481.
Nan, X., S. Cross and A. Bird (1998). "Gene silencing by methyl-CpG-binding proteins." Novartis Found Symp **214**: 6-16; discussion 16-21, 46-50.
Ng, H. H., Y. Zhang, B. Hendrich, C. A. Johnson, B. M. Turner, H. Erdjument-Bromage, P. Tempst, D. Reinberg and A. Bird (1999). "MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex." Nat Genet **23**(1): 58-61.
Pundhir, S., F. O. Bagger, F. B. Lauridsen, N. Rapin and B. T. Porse (2016). "Peak-valley-peak pattern of histone modifications delineates active regulatory elements and their directionality." Nucleic Acids Res **44**(9): 4037-4051.
Sato, A., S. Kasai, T. Kobayashi, Y. Takamatsu, O. Hino, K. Ikeda and M. Mizuguchi (2012). "Rapamycin reverses impaired social interaction in mouse models of tuberous sclerosis complex." Nat Commun **3**: 1292.
Shukeir, N., P. Pakneshan, G. Chen, M. Szyf and S. A. Rabbani (2006). "Alteration of the methylation status of tumor-promoting genes decreases prostate cancer cell invasiveness and tumorigenesis in vitro and in vivo." Cancer Res **66**(18): 9202-9210.
Stefanska, B., M. Suderman, Z. Machnes, B. Bhattacharyya, M. Hallett and M. Szyf (2013). "Transcription onset of genes critical in liver carcinogenesis is epigenetically regulated by methylated DNA-binding protein MBD2." Carcinogenesis **34**(12): 2738-2749.
Sweatt, J. D. (2016). "Neural plasticity and behavior - sixty years of conceptual advances." J Neurochem **139 Suppl 2**: 179-199.
Tavares, R. M., A. Mendelsohn, Y. Grossman, C. H. Williams, M. Shapiro, Y. Trope and D. Schiller (2015). "A Map for Social Navigation in the Human Brain." Neuron **87**(1): 231-243.
Tripathi, S., M. O. Pohl, Y. Zhou, A. Rodriguez-Frandsen, G. Wang, D. A. Stein, H. M. Moulton, P. DeJesus, J. Che, L. C. Mulder, E. Yanguez, D. Andenmatten, L. Pache, B. Manicassamy, R. A. Albrecht, M. G. Gonzalez, Q. Nguyen, A. Brass, S. Elledge, M. White, S. Shapira, N. Hacohen, A. Karlas, T. F. Meyer, M. Shales, A. Gatorano, J. R. Johnson, G. Jang, T. Johnson, E. Verschueren, D. Sanders, N. Krogan, M. Shaw, R. Konig, S. Stertz, A. Garcia-Sastre and S. K. Chanda (2015). "Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding." Cell Host Microbe **18**(6): 723-735.
Wang, L., Y. Liu, R. Han, U. H. Beier, R. M. Thomas, A. D. Wells and W. W. Hancock (2013). "Mbd2 promotes foxp3 demethylation and T-regulatory-cell function." Mol Cell Biol **33**(20): 4106-4115.
Weaver, I. C., I. C. Hellstrom, S. E. Brown, S. D. Andrews, S. Dymov, J. Diorio, T. Y. Zhang, M. Szyf and M. J. Meaney (2014). "The methylated-DNA binding protein MBD2 enhances NGFI-A (egr-1)-mediated transcriptional activation of the glucocorticoid receptor." Philos Trans R Soc Lond B Biol Sci **369**(1652).
Weber, M., I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan and D. Schubeler (2007). "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome." Nat Genet **39**(4): 457-466.

Xi, Y. and W. Li (2009). "BSMAP: whole genome bisulfite sequence MAPping program." BMC Bioinformatics **10**: 232.
Xie, B., Y. Yu, X. Meng, Q. Yu, J. Shi, H. Sang and C. Kou (2014). "Genetic association study between methyl-CpG-binding domain genes and schizophrenia among Chinese family trios." Psychiatr Genet **24**(5): 221-224.
Yang, L., B. Zou, X. Xiong, C. Pascual, J. Xie, A. Malik, J. Xie, T. Sakurai and X. S. Xie (2013). "Hypocretin/orexin neurons contribute to hippocampus-dependent social memory and synaptic plasticity in mice." J Neurosci **33**(12): 5275-5284.
Zhao, X., T. Ueba, B. R. Christie, B. Barkho, M. J. McConnell, K. Nakashima, E. S. Lein, B. D. Eadie, A. R. Willhoite, A. R. Muotri, R. G. Summers, J. Chun, K. F. Lee and F. H. Gage (2003). "Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function." Proc Natl Acad Sci U S A **100**(11): 6777-6782.

Figure Legends

Figure 1. Behavioral analysis of *mbd2-/-* mice A. Locomotion in Open-field box was assessed for 5 min. n=12-14/group. B. Exploration time during object-location memory training (left) and discrimination ratio in object-location memory test (right). n=11-13/group. C. Social interaction, mice were introduced to a novel mouse for 5 min and interaction time was recorded. n=12-17/group. D. Time in the light compartment of the Dark-Light Box (left) and number of entries to the light side (right). n=7-8/group. E. Y-maze spontaneous alteration were not affected by MBD2-/- (right), exploration is expressed as number of arm entries (left). n=13-17/group. *p<0.05, **p<0.01.

Figure 2. A. Histogram of MBD2 binding peaks around Transcription start sites (TSS, left) and position of peaks Vis-a Vis distance to TSS and their peak fold enrichment (right). B. HOMER-based De-novo motif discovery found E2F7 consensus binding site as highly enriched in MBD2 binding peaks. C. Heatmap of the overlap between MBD2 peaks and the following histone marks in the hippocampus: H3K4me1, H3K4me3, H3k2me3, H3K27ac, H39Kme3, H3K36me3 and H4K20me1. D. Analysis of the number of M2 peaks located on histone marks. E. Co-occupancy enrichment analysis of Mbd2 and histone mark. F. Pathway analysis enrichment of Mbd2 binding peaks G. Q-chip validation of Mbd2 peaks (n=3/group) p<0.001. H. RNApol2 (ser5) and H3K4me1 I. Q-ChIP for peaks Mbd2 (n=5-6 pool of mice analyzed by triplicate/group). *p<0.05, ***p<0.001.

Figure 3. A. An unsupervised clustered heatmap of differentially expressed genes in *mbd2-/-* and *mbd2+/+* mice. B-E. Pathway analysis enrichment (B, D) and gene-network (C, E) of genes down-regulated (B, C) and up-regulated (D, E) genes. F. unsupervised clustering of gene pathways of Mbd2 binding and differentially expressed genes. G. Circus plot depicting the relation between gene represented in CHIP-seq and RNA-Seq experiments, showing stronger connection between Mbd2 peaks and down-regulated genes in comparison with up-regulated genes.

Figure 4. A. A cumulative distribution of methylated levels of regulatory DNA regions bound by MBD2 demonstrating a global hyper-methylation in MBD2 KO hippocampus. B. A table summarizing the number of differentially methylated CpGs and genes under the different analytical conditions. C-D. Pathway analysis of hypo- (C) and hyper- (D) methylated gene promoters. E. Correlation between differential promoter DNA-methylation and gene-expression in *mbd2-/-* hippocampi.

Figure 5. An overlap analysis revealed a significant intersection between human ASD high risk genes (Loss of Function; FDR<0.1) and down-regulated (Top) but not up-regulated (Bottom) genes in *mbd2-/-* hippocampi. B. Similar results were obtained for the expanded ASD risk gene list (Loss of Function; FDR<0.3) and down-regulated (Top) but not up-regulated (Bottom) genes.

Figure 6. An overlap analysis revealed a significant intersection between human ASD high risk genes based on the SFARI Gene project and down-regulated (Top) but not up-regulated (Bottom) genes.

Figure S1. Average speed (left), entries to center (middle) and time near walls (right) in the open-field box tests. $p>0.05$ in all cases.

Figure S2. Top E2F De-novo motif discovery sequences found in HOMER analysis.

Figure S3. Gene-network analysis of Mbd2 binding peaks.

Figure S4. A. A snip from genome browser showing RNA-seq read alignments for the first 3 exons of the MBD2 gene. *mbd2-/-* mice show expression of the first exon only, as expected by the design of this mouse line. B. QPCR of 9 neuronal genes (Grin2b, Dclk2, Chl1, Apba2, Kcnd2, Clstn1, Gria3a, Jarid2, Gria1) shows reduced expression in *mbd2-/-* mice in agreement with the RNA-Seq (n=5-9/group; *p<0.05).

Fig.1

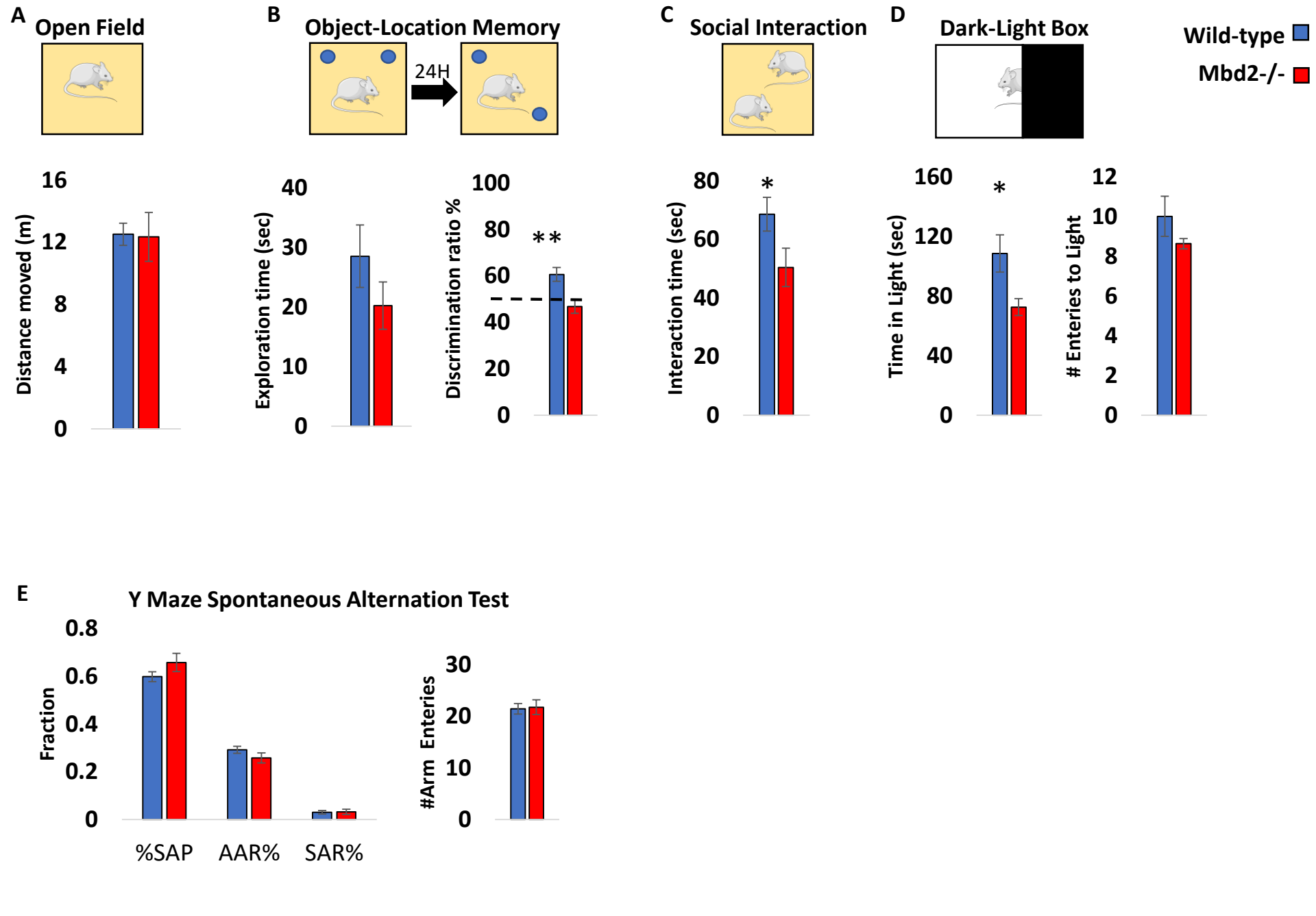

A
Open Field
B
Object-Location Memory
24H
C
Social Interaction
D
Dark-Light Box
Wild-type
Mbd2-/-
Distance moved (m)
16
12
8
4
0
Exploration time (sec)
40
30
20
10
0
Discrimination ratio %
100
80
60
40
20
0
**
Interaction time (sec)
80
60
40
20
0
*
Time in Light (sec)
160
120
80
40
0
*
Enteries to Light
12
10
8
6
4
2
0
E
Y Maze Spontaneous Alternation Test
Fraction
0.8
0.6
0.4
0.2
0
%SAP
AAR%
SAR%
#Arm Enteries
30
20
10
0

Fig.2

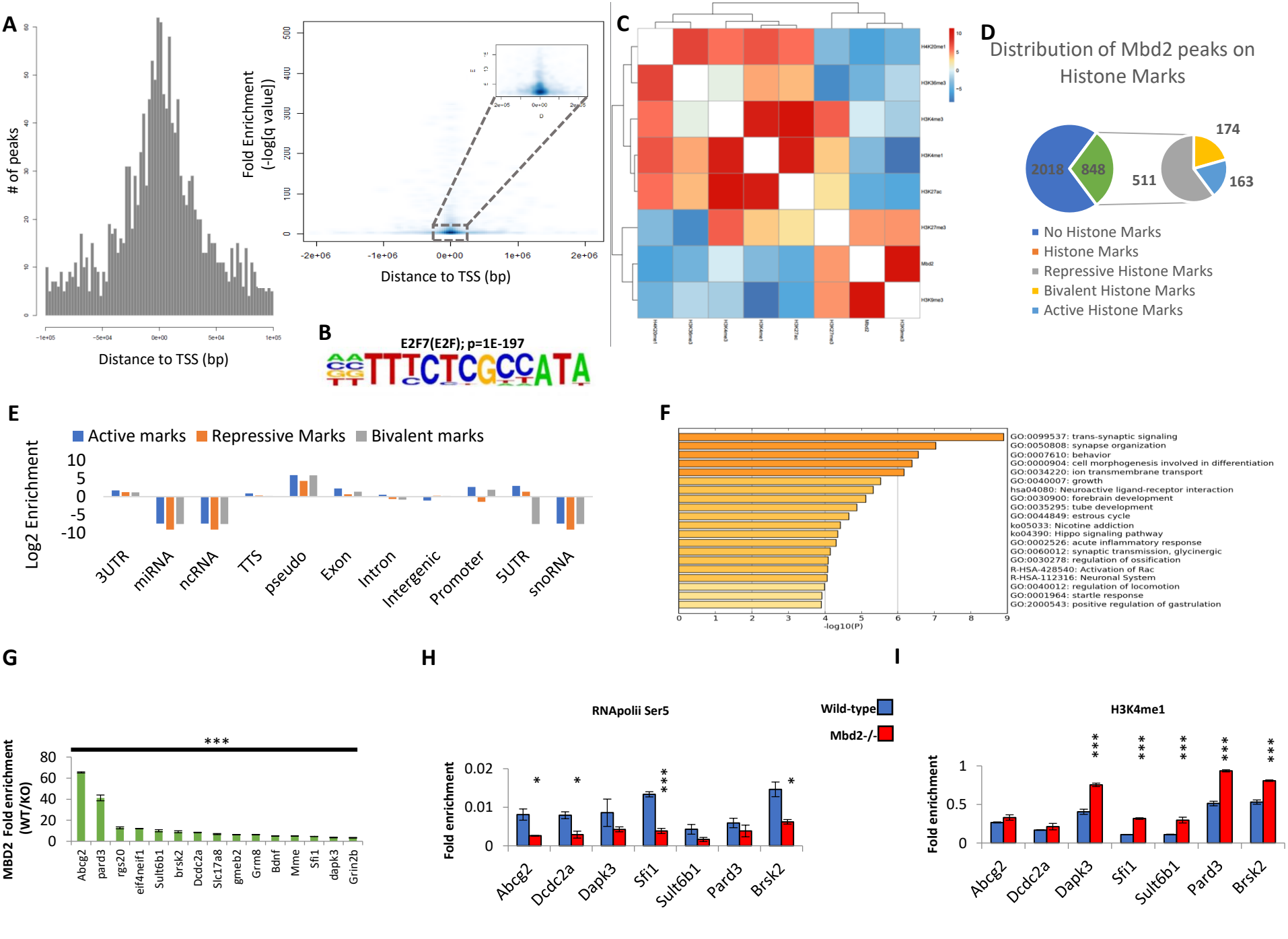

A
# of peaks
Distance to TSS (bp)
Fold Enrichment (-log[q value])
Distance to TSS (bp)
B
E2F7(E2F); p=1E-197
C
H4K20me1
H3K36me3
H3K4me3
H3K4me1
H3K27ac
H3K27me3
Mbd2
H3K9me3
D
Distribution of Mbd2 peaks on Histone Marks
2018
848
511
174
163
No Histone Marks
Histone Marks
Repressive Histone Marks
Bivalent Histone Marks
Active Histone Marks
E
Active marks
Repressive Marks
Bivalent marks
Log2 Enrichment
3UTR
miRNA
ncRNA
TTS
pseudo
Exon
Intron
Intergenic
Promoter
5UTR
snoRNA
F
GO:0099537: trans-synaptic signaling
GO:0050808: synapse organization
GO:0007610: behavior
GO:0000904: cell morphogenesis involved in differentiation
GO:0034220: ion transmembrane transport
GO:0040007: growth
hsa04080: Neuroactive ligand-receptor interaction
GO:0030900: forebrain development
GO:0035295: tube development
GO:0044849: estrous cycle
ko05033: Nicotine addiction
ko04390: Hippo signaling pathway
GO:0002526: acute inflammatory response
GO:0060012: synaptic transmission, glycinergic
GO:0030278: regulation of ossification
R-HSA-428540: Activation of Rac
R-HSA-112316: Neuronal System
GO:0040012: regulation of locomotion
GO:0001964: startle response
GO:2000543: positive regulation of gastrulation
-log10(P)
G
MBD2 Fold enrichment (WT/KO)
***
Abcg2
pard3
rgs20
eif4neif1
Sult6b1
brsk2
Dcdc2a
Slc17a8
gmeb2
Grm8
Bdnf
Mme
Sfi1
dapk3
Grin2b
H
RNApolii Ser5
Wild-type
Mbd2-/-
Fold enrichment
Abcg2
Dcdc2a
Dapk3
Sfi1
Sult6b1
Pard3
Brsk2
I
H3K4me1
Fold enrichment
Abcg2
Dcdc2a
Dapk3
Sfi1
Sult6b1
Pard3
Brsk2

Fig.3

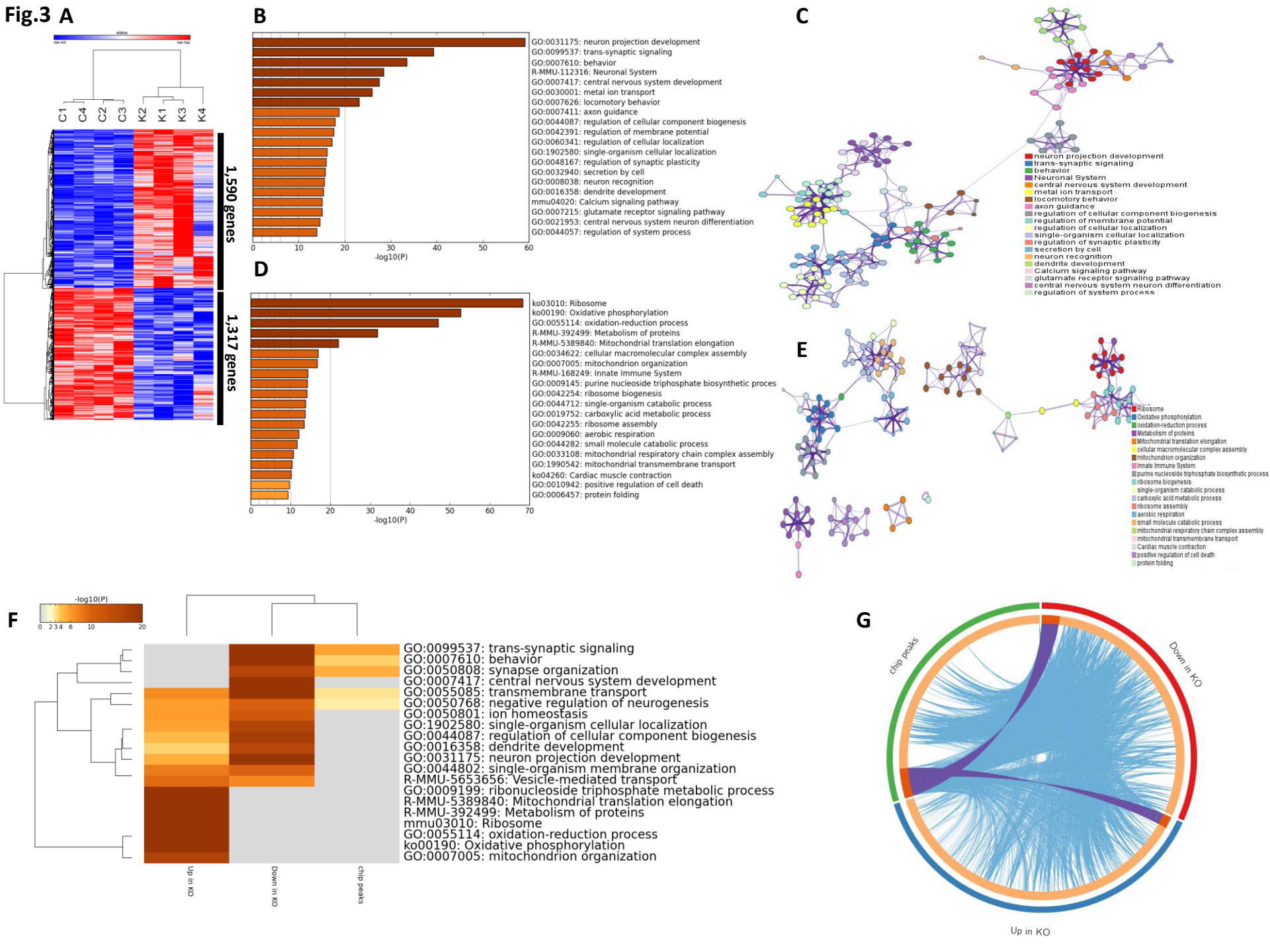
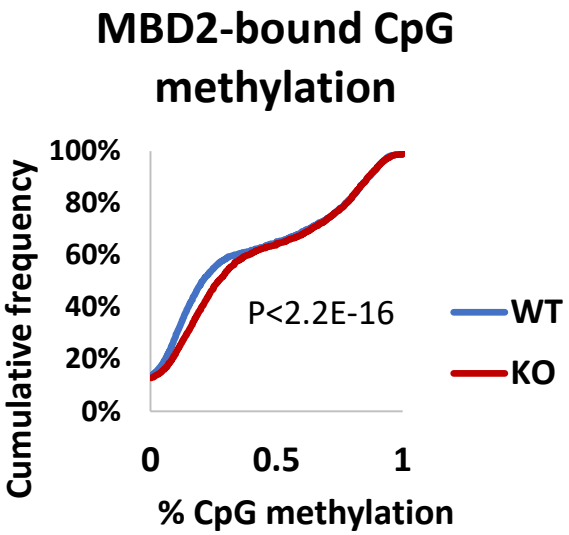
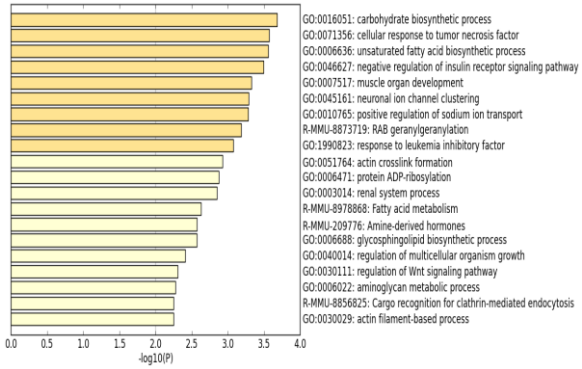

A
C1
C4
C2
C3
K2
K1
K3
K4
1,590 genes
1,317 genes
B
GO:0031175: neuron projection development
GO:0099537: trans-synaptic signaling
GO:0007610: behavior
R-MMU-112316: Neuronal System
GO:0007417: central nervous system development
GO:0030001: metal ion transport
GO:0007626: locomotory behavior
GO:0007411: axon guidance
GO:0044087: regulation of cellular component biogenesis
GO:0042391: regulation of membrane potential
GO:0060341: regulation of cellular localization
GO:1902580: single-organism cellular localization
GO:0048167: regulation of synaptic plasticity
GO:0032940: secretion by cell
GO:0008038: neuron recognition
GO:0016358: dendrite development
mmu04020: Calcium signaling pathway
GO:0007215: glutamate receptor signaling pathway
GO:0021953: central nervous system neuron differentiation
GO:0044057: regulation of system process
0
10
20
30
40
50
60
-log10(P)
C
neuron projection development
trans-synaptic signaling
behavior
Neuronal System
central nervous system development
metal ion transport
locomotory behavior
axon guidance
regulation of cellular component biogenesis
regulation of membrane potential
regulation of cellular localization
single-organism cellular localization
regulation of synaptic plasticity
secretion by cell
neuron recognition
dendrite development
Calcium signaling pathway
glutamate receptor signaling pathway
central nervous system neuron differentiation
regulation of system process
D
ko03010: Ribosome
ko00190: Oxidative phosphorylation
GO:0055114: oxidation-reduction process
R-MMU-392499: Metabolism of proteins
R-MMU-5389840: Mitochondrial translation elongation
GO:0034622: cellular macromolecular complex assembly
GO:0007005: mitochondrion organization
R-MMU-168249: Innate Immune System
GO:0009145: purine nucleoside triphosphate biosynthetic proces
GO:0042254: ribosome biogenesis
GO:0044712: single-organism catabolic process
GO:0019752: carboxylic acid metabolic process
GO:0042255: ribosome assembly
GO:0009060: aerobic respiration
GO:0044282: small molecule catabolic process
GO:0033108: mitochondrial respiratory chain complex assembly
GO:1990542: mitochondrial transmembrane transport
ko04260: Cardiac muscle contraction
GO:0010942: positive regulation of cell death
GO:0006457: protein folding
0
10
20
30
40
50
60
70
-log10(P)
E
Ribosome
Oxidative phosphorylation
oxidation-reduction process
Metabolism of proteins
Mitochondrial translation elongation
cellular macromolecular complex assembly
mitochondrion organization
Innate Immune System
purine nucleoside triphosphate biosynthetic process
ribosome biogenesis
single-organism catabolic process
carboxylic acid metabolic process
ribosome assembly
aerobic respiration
small molecule catabolic process
mitochondrial respiratory chain complex assembly
mitochondrial transmembrane transport
Cardiac muscle contraction
positive regulation of cell death
protein folding
F
-log10(P)
0 2 3 4 6 10 20
GO:0099537: trans-synaptic signaling
GO:0007610: behavior
GO:0050808: synapse organization
GO:0007417: central nervous system development
GO:0055085: transmembrane transport
GO:0050768: negative regulation of neurogenesis
GO:0050801: ion homeostasis
GO:1902580: single-organism cellular localization
GO:0044087: regulation of cellular component biogenesis
GO:0016358: dendrite development
GO:0031175: neuron projection development
GO:0044802: single-organism membrane organization
R-MMU-5653656: Vesicle-mediated transport
GO:0009199: ribonucleoside triphosphate metabolic process
R-MMU-5389840: Mitochondrial translation elongation
R-MMU-392499: Metabolism of proteins
mmu03010: Ribosome
GO:0055114: oxidation-reduction process
ko00190: Oxidative phosphorylation
GO:0007005: mitochondrion organization
Up in KO
Down in KO
chip peaks
G
chip peaks
Down in KO
Up in KO

Fig.4

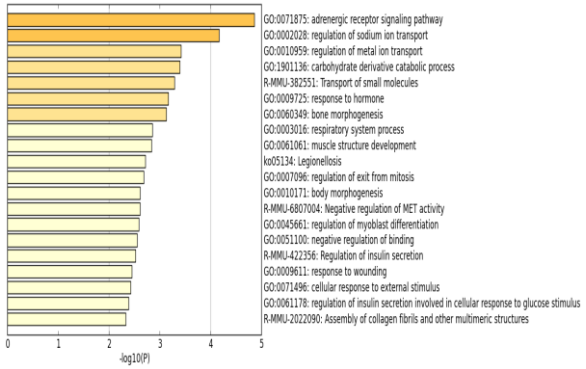
A



C



D



B

	Hypo-methylated CpGs	Hyper-methylated CpGs
All CpGs (FDR<0.05)	151 (117 genes)	210 (127 genes)
All CpGs (p<0.001)	1519 (1330 genes)	1486 (1214 genes)
promoters (p<0.001)	494 (460 genes)	476 (434 genes)

E

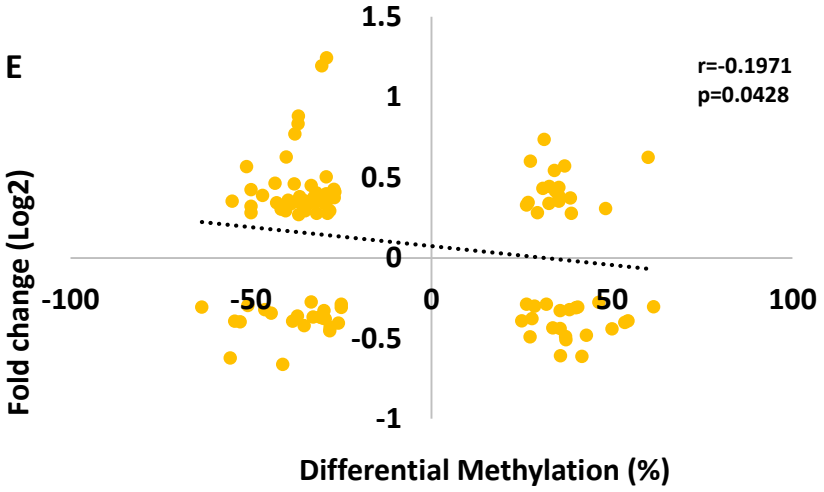


Fig.5

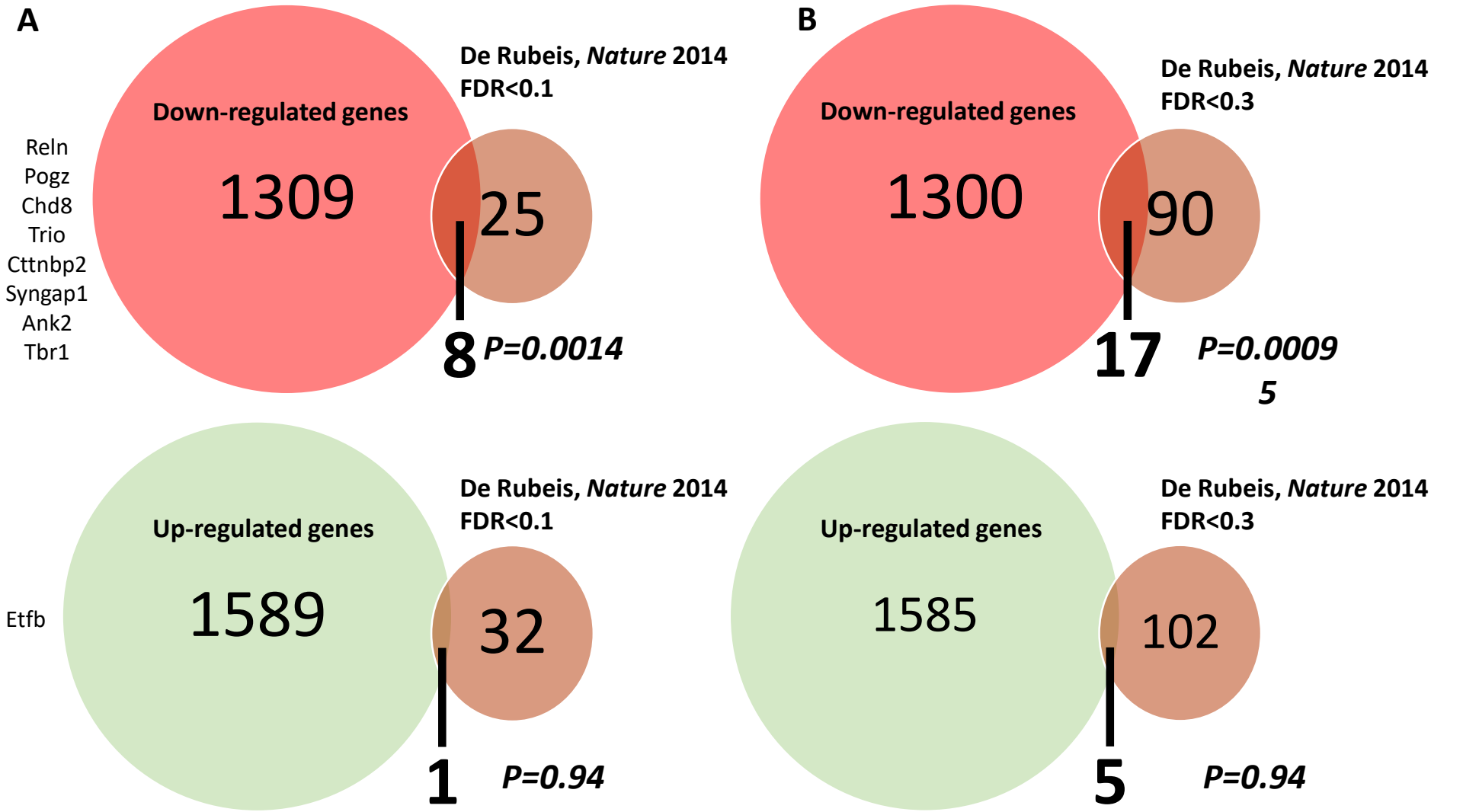

A
Down-regulated genes
1309
25
8
P=0.0014
De Rubeis, Nature 2014
FDR<0.1
Reln
Pogz
Chd8
Trio
Cttnbp2
Syngap1
Ank2
Tbr1
B
Down-regulated genes
1300
90
17
P=0.00095
De Rubeis, Nature 2014
FDR<0.3
Up-regulated genes
1589
32
1
P=0.94
De Rubeis, Nature 2014
FDR<0.1
Etfb
Up-regulated genes
1585
102
5
P=0.94
De Rubeis, Nature 2014
FDR<0.3

Fig.6

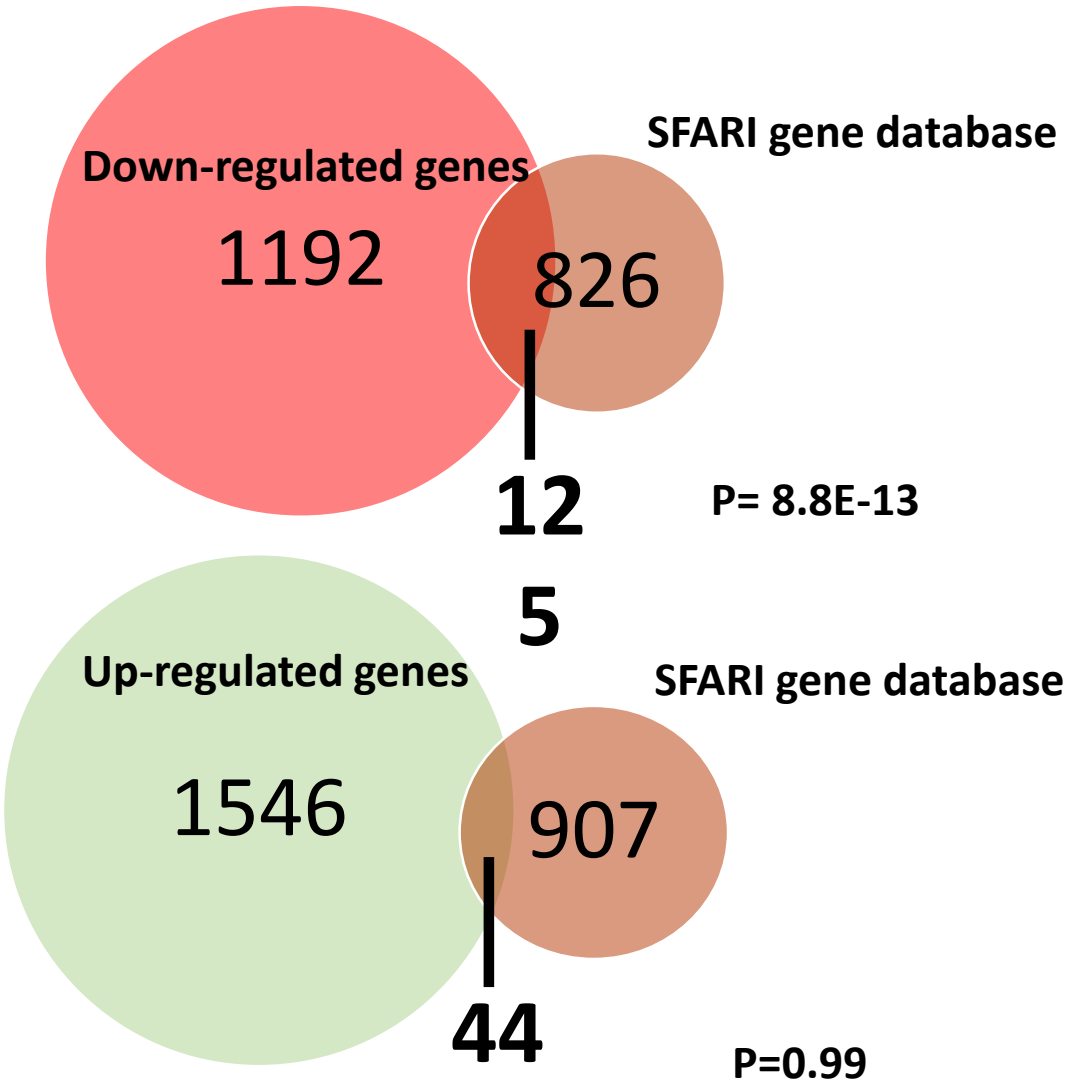

Down-regulated genes
1192
826
SFARI gene database
12
P= 8.8E-13
5
Up-regulated genes
1546
907
SFARI gene database
44
P=0.99

Fig.S1

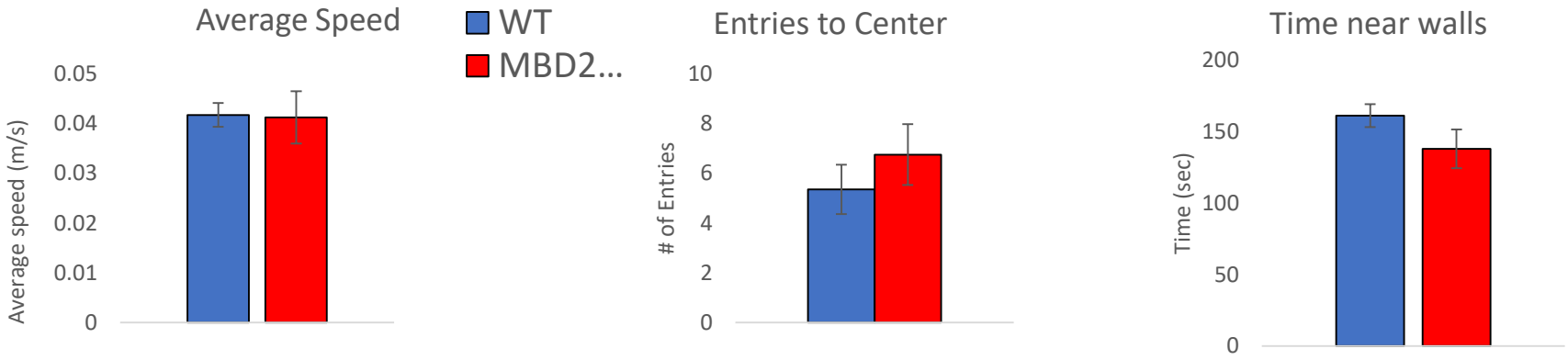
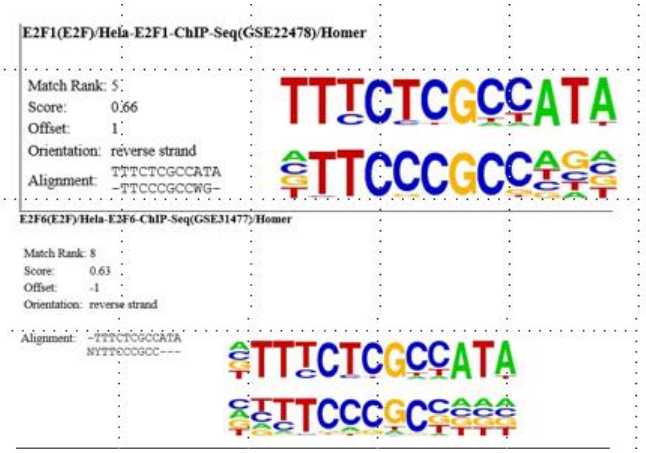

Average Speed
WT
MBD2...
Average speed (m/s)
0.05
0.04
0.03
0.02
0.01
0
Entries to Center
# of Entries
10
8
6
4
2
0
Time near walls
Time (sec)
200
150
100
50
0

Fig.S2


E2F1(E2F)/Hela-E2F1-ChIP-Seq(GSE22478)/Homer
Match Rank: 5
Score: 0.66
Offset: 1
Orientation: reverse strand
Alignment: TTTCTCGCCATA
-TTCCCGCCWG-
E2F6(E2F)/Hela-E2F6-ChIP-Seq(GSE31477)/Homer
Match Rank: 8
Score: 0.63
Offset: -1
Orientation: reverse strand
Alignment: -TTTCTCGCCATA
NTTTCCCGCC---

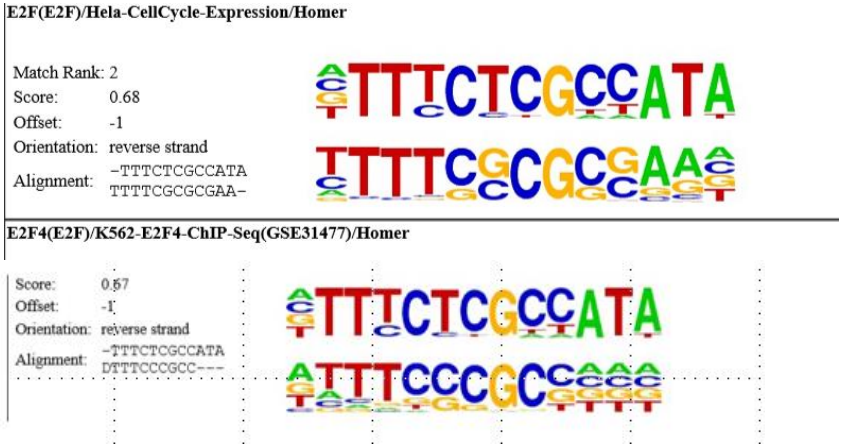

E2F(E2F)/Hela-CellCycle-Expression/Homer
Match Rank: 2
Score: 0.68
Offset: -1
Orientation: reverse strand
Alignment: -TTTCTCGCCATA
TTTTCGCGCGAA-
E2F4(E2F)/K562-E2F4-ChIP-Seq(GSE31477)/Homer
Score: 0.67
Offset: -1
Orientation: reverse strand
Alignment: -TTTCTCGCCATA
DTTTCCCGCC---

Fig.S3

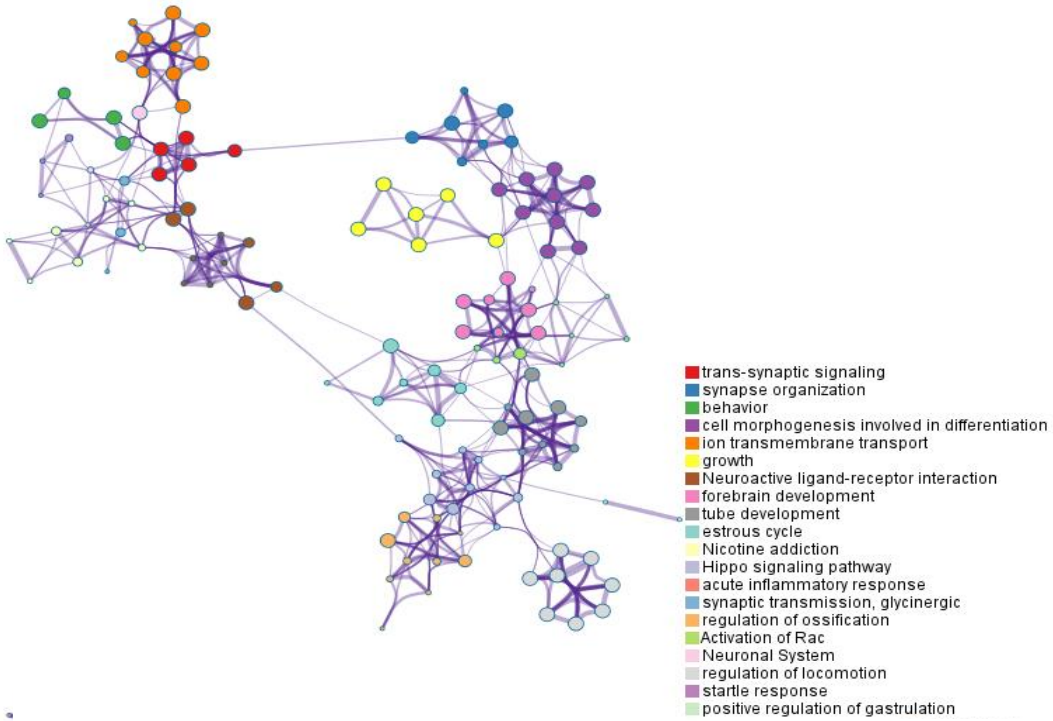
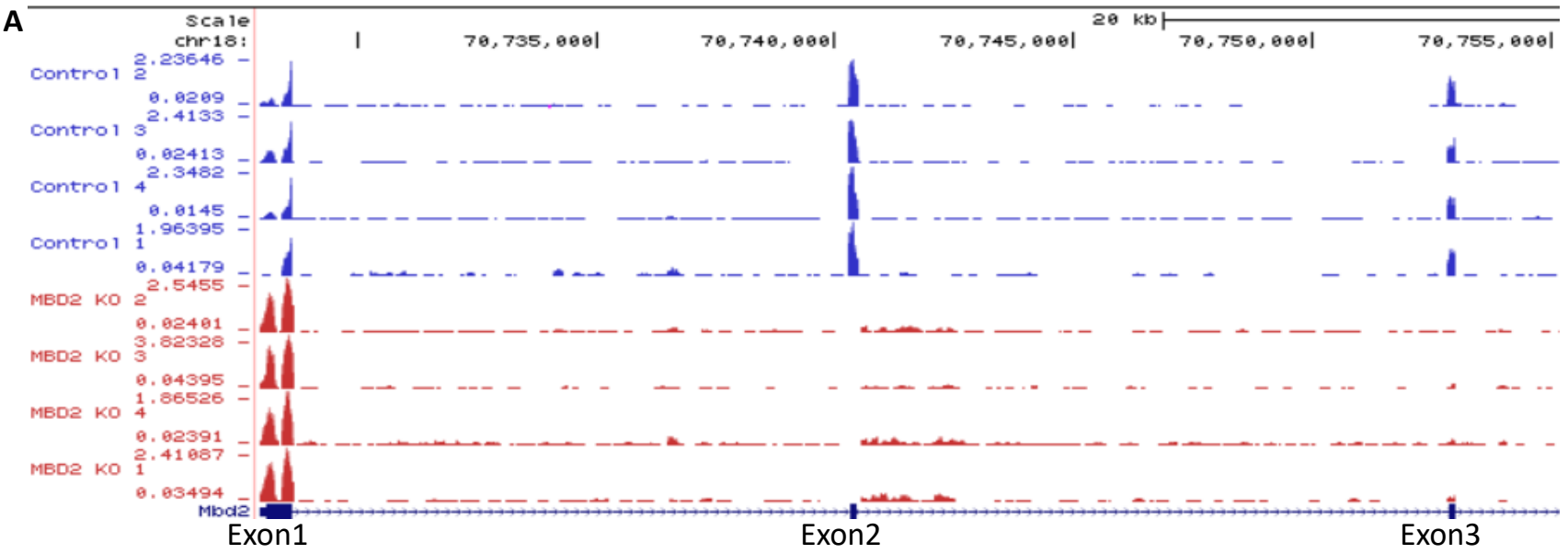
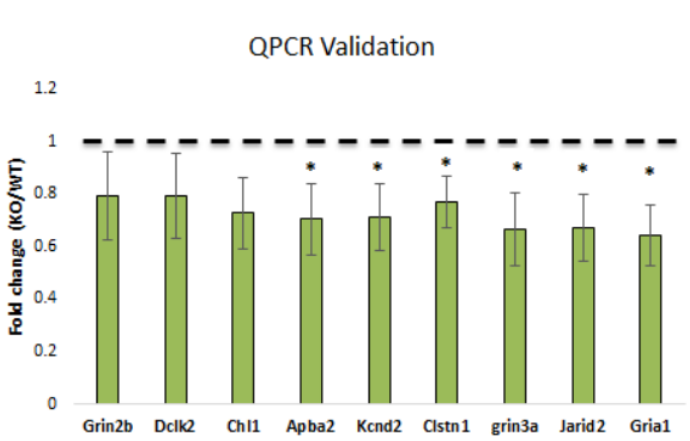

trans-synaptic signaling
synapse organization
behavior
cell morphogenesis involved in differentiation
ion transmembrane transport
growth
Neuroactive ligand-receptor interaction
forebrain development
tube development
estrous cycle
Nicotine addiction
Hippo signaling pathway
acute inflammatory response
synaptic transmission, glycinergic
regulation of ossification
Activation of Rac
Neuronal System
regulation of locomotion
startle response
positive regulation of gastrulation

Fig.S4


A
Scale
chr18:
20 kb
70,735,000
70,740,000
70,745,000
70,750,000
70,755,000
Control 2
Control 3
Control 4
Control 1
MBD2 KO 2
MBD2 KO 3
MBD2 KO 4
MBD2 KO 1
Mbd2
Exon1
Exon2
Exon3


B
QPCR Validation
Fold change (KO/WT)
1.2
1
0.8
0.6
0.4
0.2
0
Grin2b
Dclk2
Chl1
Apba2
Kcnd2
Clstn1
grin3a
Jarid2
Gria1