Supplement to:

A newly characterized AT-hook domain in MeCP2 determines clinical course of Rett syndrome and related disorders

Steven Andrew Baker, Lin Chen, Angela Dawn Wilkins, Peng Yu,

Olivier Lichtarge, and Huda Yahya Zoghbi

Inventory

Page Number

Supplementary Figure Legends

Figure S1, related to Figure 1. Characterization of WT;R270X and WT;G273X transgenic lines

(A) (Top) Western blot analysis using whole brain lysate for each of the transgenic lines along with their WT littermates. The MeCP2 antibody against the N-terminus that recognizes both MeCP2-WT and both mutant forms of MeCP2 was used. The transgenic gene product fused with GFP migrates as a band below MeCP2-WT. Lysates were made from 4th generation animals to compare with 2nd generation animals found in Figure 1B. (Bottom) Quantification of the amount of MeCP2-R270X or MeCP2-G273X as a percentage of MeCP2-WT in each transgenic line. Data represent the mean and error bars represent SEM. n = 3 per transgenic line.

(B-F) Double immunofluorescence staining from various brain regions of each transgenic line along with a WT control. The MeCP2 antibody used for immunofluorescence recognizes the C-terminus and therefore is specific for MeCP2-WT. Scale bars represent 50 μm. (B) Cortex (C) Hippocampus (D) Cerebellum (E) Hypothalamus (F) Medulla.

Figure S2, related to Figure 2. WT;R270X and WT;G273X mice are similar to WT mice and the phenotypes of G273X mice are reproduced in independent lines

(A) Representative photographs of WT;R270X and WT;G273X mice at 8 weeks of age.

(B) Kaplan-Meier curves for WT, WT;R270X, and WT;G273X mice. Censored animals are indicated with a black tick mark. n = 54, 23, and 28 for WT, WT;R270X, and WT;G273X, respectively.

(C) Average body weight plotted versus age. Maximum n = 54, 23, and 28 for WT, WT;R270X, and WT;G273X, respectively.

(D) Average brain weights for WT, WT;R270X, and WT;G273X mice at 9 weeks of age. n = 8,

3, and 3 for WT, WT;R270X, and WT;G273X, respectively.

(E) Average cumulative severity scores (see Experimental Procedures) are plotted versus age. Error bars represent SEM. n = 25, 20, 15, 22 for WT, n = 7, 6, 11, 9 for WT;R270X, and n = 8, 6, 7, 10 for WT;G273X at 4-6, 7-9, 9-12, and >12 weeks, respectively.

(F) Western blot analysis using whole brain lysates for each of the transgenic lines. A MeCP2 antibody against the N-terminus that recognizes MeCP2-WT and MeCP2-G273X was used. Mutant MeCP2 fused with GFP migrates as a band below MeCP2-WT. Lysates were made from 2nd generation animals. n = 2 mice per genotype.

(G) Double immunofluorescence staining from the cortex of the transgenic lines along with a WT control. The MeCP2 antibody used for immunofluorescence recognizes the C-terminus and therefore is specific for MeCP2-WT. Scale bars represent 50 μm.

(H) Kaplan-Meier curves for WT, KO, and R270X mice compared with G273X mice (lines B and C). Censored animals are indicated with a black tick mark. * indicates $p<0.05$ for G273X lines B and C compared to KO and R270X, Gehan-Breslow-Wilcoxon test. n = 54, 41, 28, 6 for WT, KO, R270X, and G273X, respectively.

(I) Average body weight plotted versus age. * indicates $p<0.05$ for G273X lines B and C compared to KO and R270X. Maximum n = 54, 41, 28, 6 for WT, KO, R270X, and G273X, respectively.

(J) Average cumulative severity scores (see Experimental Procedures) are plotted versus age. ** indicates $p<0.01$, Mann-Whitney U Test. n = 25, 20, 15 for WT; n = 17, 15, 11 for KO; n = 12, 6, 6 for R270X; and n = 15, 13, 11 for G273X mice (lines B and C) at 4-6, 7-9, and 9-12 weeks, respectively.

(K) Images of symptomatic KO, R270X, and G273X mice demonstrating hind-limb clasping in

all three genotypes.

The WT, KO, and R270X data in this figure are the same as shown in the corresponding main text figures as all mice were analyzed on the same genetic background.

All error bars show SEM.

Figure S3, related to Figure 3. H1 shows similar properties in whole brain nuclei from R270X and G273X mice

(A) ChIP-Seq reads from R270X and G273X mice that were mapped to each repetitive element. IAP refers to the Intracisternal A-Particle, L1 refers to the mouse L1 retrotransposon, Satellite refers to mouse major satellite DNA.

(B) Whole brain nuclei were purified from WT, R270X, or G273X mice and resuspended in buffers comprised of 300mM, 400mM, or 1M NaCl. Western blots using an antibody for histone H1 show the amount extracted under each condition. n = 3 per genotype.

(C) The total amount of H1 from whole brain nuclei extracted in 1M NaCl normalized to H3 is quantified and the average is shown for R270X and G273X. These values are plotted relative to WT.

(D) The average extracted fractions from the experiment described in (B) are plotted for each genotype above the corresponding NaCl concentration. Data are normalized for the amount of H1 extracted in 1M NaCl.

All error bars show SEM.

Figure S4, related to Figure 4. KO, R270X, and G273X show similar transcriptional changes at 9 weeks

Hypothalamic RNA levels were quantified by qPCR for 4 genes that are down-regulated in KO mice (*Bdnf*, *Sst*, *Oprk1*, and *Tac1*) and two genes that are up-regulated (*Mef2c* and *Grin2a*). Transcript levels were normalized to *Gapdh* and represented as the fold expression relative to WT. **** indicates $p<0.0001$, *** $p<0.001$, ** $p<0.01$, * $p<0.05$. Gene expression of 9 week old animals. n = 8, 5, 7, 6 for WT, KO, R270X, and G273X, respectively.
All error bars show SEM.

Figure S5, related to Figure 5. MeCP2 and HMGA1 share high identity and MeCP2 induces oligomerization of NAs

(A) The average sequence identity between MeCP2 and all human proteins binned based on protein length.

(B) The frequency distribution of human proteins (with length similar to HMGA1) based on sequence identity with MeCP2.

(C) The average sequence identity between HMGA1 and all human proteins binned based on length.

(D) The frequency distribution of human proteins (with length similar to MeCP2) based on sequence identity with HMGA1.

(E) Recombinant MeCP2 AT-Hook 2 (amino acids 257-272) was used in an EMSA assay to test binding to a 64-mer double stranded DNA probe. The probe was added at 66.7nM, reacted with the indicated protein concentration, and separated by native PAGE. The AT-Hook 2 domain truncated at R270 (R270X) failed to bind DNA even when highly concentrated. The full AT-Hook 2 domain ending at G273 (G273X) readily forms the indicated complexes with the probe DNA.

(F) Recombinant MeCP2 was mixed with methylated NAs and after the addition of $MgCl_2$ at the indicated concentration the oligomerized material was separated by centrifugation. The NAs that remain in the soluble fraction were run on an agarose gel and visualized by ethidium bromide staining. M stands for the 1Kb plus DNA marker.

(G) Quantification of the experiment described in (F). *** indicates $p<0.001$ and * $p<0.05$, between MeCP2-R270X and MeCP2-G273X, Student's t-test. n = 4.

Figure S6, related to Figure 6. ATRX mislocalization occurs in multiple RTT mouse models

(A) Lower magnification images of the experiment described in Figure 6A. Scale bars represent 15 μm.

(B) Double immunofluorescence for MeCP2 and ATRX in pre-symptomatic *Mecp2* $^{-/+}$ heterozygous female mice (9 weeks old). Within the hippocampus MeCP2 expressing neurons (solid circle) can be distinguished from neurons not expressing MeCP2 (dashed circle). ATRX foci are generally brighter in MeCP2 positive neurons compared to adjacent MeCP2 negative neurons. The top panel shows the same field without the circles. Scale bars represent 10 μm. n = 2 mice.

(C) Double immunofluorescence for MeCP2 and ATRX in symptomatic *MECP2* $^{TG3/+}$ female mice, which are mosaic for MeCP2 overexpression due to X-chromosome inactivation. (Top) Within the hippocampus normal MeCP2 expressing neurons (solid circle) can be distinguished from neurons overexpressing MeCP2 (dashed circle). ATRX foci are generally brighter in MeCP2 overexpressing neurons. Scale bars represent 10 μm. n = 2 mice. (Bottom) Higher magnification image of the WT and TG3 overexpressing neuron circled in the upper panel.

(D) Quantification of the average number of ATRX foci detected per neuron in symptomatic *MECP2* $^{TG3/+}$ heterozygous female mice. n = 4 high-power fields from 2 mice.

(E) Quantification of the average foci intensities in symptomatic *MECP2* $^{TG3/+}$ heterozygous female mice. **** indicates $p<0.0001$. n = 157 WT foci and n = 323 TG3 foci from 2 mice.

Figure S7, related to Figure 7. ATRX mislocalization occurs in unfixed brain nuclei from KO mice

(A) Coimmunoprecipitation of GFP tagged MeCP2 with endogenous ATRX. N2a cells were transfected with the indicated form of MeCP2. Post-lysis, extracted complexes were immunoprecipitated with anti-GFP sera and analyzed by western blot. GAPDH serves as a loading control. MeCP2-WT, MeCP2-R270X, and MeCP2-G273X all interact with ATRX to a similar extent. n = 3 independent experiments.

(B) Immunofluorescence for GFP within the hippocampus of either R270X mice or G273X mice. Both MeCP2-R270X and MeCP2-G273X remain localized to PCH at both 4 weeks of age (i) and 9 weeks of age (ii). Scale bars represent 10 μm. n = 2 mice per genotype per age.

(C) Immunofluorescence for ATRX in fresh nuclei prepared from WT and KO mice at 9 weeks of age. Nuclei are counterstained with Hoechst. Staining and imaging were performed under identical conditions for WT and KO nuclei. Scale bars represent 10 μm. n = 3 mice per genotype.

(D) Quantification of the average number of ATRX foci detected per neuron in each genotype at 9 weeks of age. **** indicates $p<0.0001$. n = 10 high-power fields from 2 mice per genotype.

(E) Quantification of the average foci intensities for each genotype at 9 weeks of age. **** indicates $p<0.0001$. n = 153 WT foci and n = 66 KO foci from 2 mice per genotype.

All error bars show SEM.

Extended Experimental Procedures

Generation of Vectors.

The full-length human *MECP2* coding sequence was cloned into pENTR using the Gateway Cloning System per the manufacturer's instructions (Invitrogen). The *MECP2* cDNA was then tagged on the C-terminus with GFP by subcloning into pDEST47. Subsequent mutants were generated using the QuikChange XL Site Directed Mutagenesis Kit per the manufacturer's instructions (Stratagene).

The luciferase reporter vector pUAS-hActin-luc2 was generated by insertion of an NdeI site immediately upstream of the hsp70 promoter in pUAS-luc2 by site directed mutagenesis. The hsp70 promoter was then removed by digestion with NdeI and EcoRI. A 615bp fragment of the human *ACTB* promoter was amplified with the following primers:

Forward 5'-ccgccgcatatgactgcctggccactccatgccct-3'
Reverse 5'-ccgccggaattcgtgagccgcctgccccggtcggct-3'

This fragment was digested and inserted into cut pUAS-luc2 to generate pUAS-hActin-luc2.

For the generation of the effector constructs, pCMX-Gal4 was digested with SalI and EcoRV. A full-length wild-type or R111G mutant *MECP2* cDNA was inserted downstream of the Gal4-DNA Binding Domain following PCR and digestion with SalI and EcoRV. Subsequent mutants were generated by site directed mutagenesis.

The bacterial expression vector pGEX-5x3-hMeCP2 was generated by standard cloning of a full-length human *MECP2* cDNA into an empty pGEX-5x3 vector. This resulting vector was further

modified by site directed mutagenesis to delete the N-terminal 256 codons and subsequently introduce the R270X and G273X mutation. The resulting vectors pGEX-5x3-hMeCP2-delN256-R270X and pGEX-5x3-hMeCP2-delN256-G273X code for the expression of an N-terminal GST tag followed by a short linker derived from the vector and then amino acids 257-269 or 257-272 of human MeCP2, for AT-Hook 2-R270X and AT-Hook 2-G273X respectively.

The bacterial expression vector pTXB1-hMeCP2 was generated by standard cloning of a full-length human *MECP2* cDNA into the empty pTXB1 vector using NdeI and EcoRI sites. The frame was adjusted by inserting an alanine codon directly in between the *MECP2* coding sequence and the vector encoded intein by site directed mutagenesis. The vectors pTXB1-hMeCP2-R270X and pTXB1-hMeCP2-G273X were generated by site directed mutagenesis from the original pTXB1-hMeCP2 clone.

All final vectors were checked by sequencing.

PAC Recombineering and Generation of Transgenic Mice.

A PAC containing the full endogenous *MECP2* locus from human, PAC671D9, which had been previously modified by inserting the coding sequence for the Enhanced Green Fluorescent Protein (EGFP) downstream of the *MECP2* coding sequence, was used as starting material. This PAC was electroporated into SW102 cells for recombineering using the protocol described in (http://web.ncifcrf.gov/research/brb/protocol/Protocol3_SW102_galK_v2.pdf). The galK gene was amplified by PCR using the following primer:

galK: Forward 5'-

ggaagcgaaaagctgaggccgaccctcaggccattcccaagaaacggggccctgttgacaattaatcatcggca-3'

Reverse 5’- ggcaccaccccggtgaacagctcctcgcccttgctcaccatggtggcgactcagcactgtcctgctcctt-3’

The resulting product was gel purified and electroporated into induced SW102 cells to replace the portion of *MECP2* exon4 that codes for amino acids 270-486. Recombinants were selected for growth on galactose and streaked onto MacConkey’s agar. Individual Gal+ colonies were then screened for the correct recombination by PCR. A single colony was chosen for the second step of recombination wherein the galK cassette was replaced by either a 6 amino acid poly-serine linker (R270X allele) or the same poly-serine linker preceded by the *MECP2* sequence encoding amino acids 270-272 (G273X allele). The following oligonucleotides were annealed and electroporated into induced cells:

R270X: Forward 5’-

ggaagcgaaaagctgaggccgaccctcaggccattcccaagaaacggggcagtagcagtagcagtagcgtcgccaccatggtgagcaa gggcgaggagctgttcaccggggtggtgcc-3’

Reverse 5’-

ggcaccaccccggtgaacagctcctcgcccttgctcaccatggtggcgacgctactgctactgctactgccccgtttcttgggaatggcctg agggtcggcctcagcttttcgcttcc-3’

G273X: Forward 5’-

ggaagcgaaaagctgaggccgaccctcaggccattcccaagaaacggggccgaaagccgagtagcagtagcagtagcgtcgccacca tggtgagcaagggcgaggagctgttcaccggggtggtgcc-3’

Reverse: 5’-

ggcaccaccccggtgaacagctcctcgcccttgctcaccatggtggcgacgctactgctactgctactcggctttcggccccgtttcttggga

atggcctgagggtcggcctcagcttttcgcttcc-3’

Recombinants were selected for resistance to 2-Deoxy-D-galactose and screened for the correct recombination by PCR. A single colony bearing the desired final construct for each allele was chosen and the region surrounding the recombination events was sequenced. The genomic DNA was liberated from the PAC backbone by NotI digestion followed by pulse field electrophoresis. No other genes were included in the isolated ~99kb fragment. Pronuclear injection using 1 ng/µl was performed to generate the transgenic mice according to standard procedures. All transgenic lines were mated to FVB wild-type mice. Individual transgenic lines were analyzed initially in the second and third generation for expression of MeCP2-R270X or MeCP2-G273X by western blot and immunofluorescence. The two chosen R270X lines and the single G273X lines used in **Figures 1 and 2** of this study were assayed again in the fourth and fifth generation to ensure stable transgene expression. For western blots, extracts were prepared from whole brain by dounce homogenization in Extraction Buffer (100mM Tris-Cl pH 7.4, 2% SDS) followed by lysis for 15 minutes. Insoluble material was removed by centrifugation and the supernatant was mixed with 2x Laemmli Buffer and stored at -20ºC. Lysates were separated by electrophoresis on NuPAGE 4-12% Bis-Tris gels, transferred to PVDF, and detected per standard methods. The following antibodies were used: rabbit anti-serum raised against the MeCP2 N-terminus (1:10000, Zoghbi Lab, #0535), affinity purified rabbit anti-GFP (1:5000, Abcam, ab6556), rabbit anti-H3 (1:100000, Millipore, 07-690), mouse anti-GAPDH 6C5 (1:20000, Advanced Immunochemicals, 2-RGM2), donkey anti-rabbit HRP (1:3000, GE Healthcare, NA934), and donkey anti-mouse HRP (1:100000, Jackson ImmunoResearch Labs, 715-035-150). For immunofluorescence, brains from freshly sacrificed mice were halved and submerged in cold

phosphate buffered saline (PBS) containing 4% paraformaldehyde. The tissue was allowed to fix overnight at 4ºC and then transferred into a 30% sucrose solution for cryoprotection. Cryoprotected brains were immersed in Optimal Cutting Temperature compound (Tissue-Tek) and frozen over dry ice. Sagittal sections were cut in a Leica CM3050S at 50µm and washed in PBS. Washed sections were then blocked in Blocking Solution (2% normal goat serum, 0.3% Triton X-100, PBS) for 1 hour prior to the addition of primary antibody in Blocking Solution. Sections were incubated with primary antibody at 4ºC for 48 hours and then washed four times prior to the addition of secondary antibody in Blocking Solution. Sections were then incubated at 4ºC for 48 hours with the secondary antibody before a final four washes. Stained sections were mounted onto glass slides in Vectashield Mounting Media with DAPI (Vector Laboratories). The following antibodies were used: rabbit anti-MeCP2 D4F3 which recognizes the C-terminus (1:500, Cell Signaling Technology, 3456S), chicken anti-GFP (1:1000, Abcam, ab13970), goat anti-rabbit Alexa 633 (1:1000, Invitrogen, A-21071), goat anti-chicken Alexa 488 (1:1000, Invitrogen, A-11039).

Colocalization Analysis.

High power confocal images (63x lens) of cortical neurons stained for the C-terminus of MeCP2, GFP, and DAPI were analyzed using the JACoP plugin for ImageJ (Bolte and CordeliÈRes, 2006). The average Overlap Coefficient and Manders' Coefficients from all neurons in two high-power fields per transgenic line were calculated using the default settings.

Chromatin Immunoprecipitation.

Mice were sacrificed under anesthesia and brains removed rapidly to a cold glass plate. The

whole brain was minced and snap frozen in liquid nitrogen prior to storage at -80ºC until needed. The minced tissue was then thawed and fixed in 1% formaldehyde in PBS for 15 minutes at room temperature, prior to quenching and washing in 125mM glycine in PBS. An equal packed tissue volume from each animal (100µL) was then homogenized in Lysis Buffer (10mM Tris-Cl pH 7.4, 10mM NaCl, 1.5mM $MgCl_2$, 0.1% Triton X-100, 0.5mM PMSF, Complete Protease Inhibitor, Ambion RNAse Cocktail) by 20 strokes with a tight fitting pestle. Nuclei were collected at 11,000rpm in a table-top centrifuge and resuspended in Micrococcal Nuclease Buffer (10mM Tris-Cl pH 7.4, 10mM NaCl, 3mM $MgCl_2$, 1mM $CaCl_2$, 4% NP40, 0.5mM PMSF, Complete Protease Inhibitor). Nuclei were then sonicated with 6 pulses at 20% duty cycle and 2.7 power. The nuclear suspension was then allowed to lyse for 15 minutes before treating with 60U of Micrococcal Nuclease (Worthington) for 15 minutes at 37ºC. Reactions were stopped by addition of EGTA and SDS and sonicated again for 6 pulses at 20% duty cycle and 2.7 power. The cleaved chromatin was then homogenized by rotation for 30 minutes at 4ºC and cleared by spinning at max speed in a table-top centrifuge for 10 minutes. The resulting soluble chromatin was checked for proper shearing prior to proceeding with the immunoprecipitation. Chromatin sheared to between 100-500bp was used for the remaining steps.

Soluble chromatin was diluted 1:10 in ChIP Dilution Buffer (Millipore) and pre-cleared with Protein A Dynabeads (Invitrogen). An aliquot of pre-cleared chromatin was stored as input and the remaining supernatant was incubated overnight at 4ºC with 3µg of affinity purified rabbit anti-GFP (Abcam, 6556). 50µL of Protein A Dynabeads was then added to each reaction and allowed to bind for >3 hours at 4ºC. Bound complexes were then washed in Low-Salt Immune Complex Wash Buffer (Millipore), High-Salt Immune Complex Wash Buffer (Millipore), LiCl Immune Complex Wash Buffer (Millipore), and 1xTE. Complexes were then eluted twice in 1%

SDS/100mM $NaHCO_3$. Precipitated chromatin and input samples were then reverse cross-linked and treated with proteinase K. DNA was recovered using a Qiagen PCR Purification Kit.
For qPCR analysis, the input and immunoprecipitated chromatin were diluted and stored at -20ºC until needed. Reactions were carried out in duplicate using iTaq SYBER Green Supermix (Bio-Rad) and a CFX96 Real-Time System (Bio-Rad) per the manufacturer's instructions. The following primers were used:

Gapdh: Forwards 5'-ccagctactcgcggctttacgg-3'

Reverse 5'-cctcccgccctgcttatccagt-3'

Afm: Forwards 5'-agacaggctggcctgagagtca-3'

Reverse 5'-ttccaatgcacgcgtctcaccc-3'

Mouse Major Satellite: Forwards 5'-catccacttgacgacttgaaaa-3'

Reverse 5'-gaggtccttcagtgtgcattt-3'

L1 Retrotransposon: Forwards 5'-agaagaaacgggagacagca-3'

Reverse 5'-ctgccgtctactcctcttgg-3'

Sst: Forwards 5'-cattgacaggtacccaactga-3'

Reverse 5'-cagccacataggagcacactt-3'

Crh: Forwards 5'-gtcaccaaggaggcgataccta-3'

Reverse 5’-taaataatagggccctgccaag-3’

Results were averaged for each sample and normalized to input. Data for each genotype were then plotted as the mean ± SEM and analyzed by one-way ANOVA with Bonferroni-Holm post hoc analysis.

ChIP-Seq analysis was performed as above except that final sample preparation of ChIPed DNA was as described in (Klisch et al., 2011). In brief, samples were sent to the Genomic and RNA Profiling Core at Baylor College of Medicine. Libraries were prepared from 10 ng of ChIPed DNA and input control and sequenced on the Illumina Solexa HiSeq according to manufacturer protocols.

Analysis of ChIP-Seq Data.

For the R270X experiment we generated 81,855,910 raw reads of which 46,413,828 were uniquely mapped to the mouse genome. For the G273X experiment we generated 82,698,419 raw reads of which 48,911,608 were uniquely mapped to the mouse genome. Comparison of MeCP2-WT with MeCP2-R270X and MeCP2-G273X was carried out as follows: mapped short reads are counted in a 5kb window shifting with a 1kb step across a 52 Mb region in chromosome 5 as shown in **Figure 3A**. The vertical axis represents the number of reads that fall in the 5kb windows whose positions are indicated by the horizontal axis. For correlation analysis mapped short reads were counted in 100kb bins across the whole genome, with no spacing or overlapping between neighboring bins. Because WT was done in a previous study (Skene et al., 2010), we quantile normalized the bin counts from WT across all the bins with respect to the mean bin count from R270X and G273X across all the bins. The three scatter plots show the

pairwise relations between R270X and WT (quantile normalized), between G273X and WT (quantile normalized), and between R270X and G273X. Spearman's rank correlation coefficients are computed for each pair as shown in the plots.

Salt Extraction from Purified Nuclei.

Mice were sacrificed under anesthesia and the brain quickly removed to a cold glass plate and minced over ice. Tissue was snap frozen in liquid nitrogen and stored at -80ºC until needed. Thawed tissue was dounce homogenized by 20 strokes with a tight pestle in Hypertonic Buffer A (300mM Sucrose, 2mM $MgCl_2$, 3mM $CaCl_2$, 10mM Tris-Cl pH 8.0, 0.1% Triton X-100, 0.5mM DTT, Complete Protease Inhibitor). Cells were allowed to lyse for 5 minutes on ice and then centrifuged at 720 x g for 5 minutes. The nuclei were then washed twice in cold Hypertonic Buffer A and then brought up in Nuclear Extraction Buffer (200mM NaCl, 10mM Tris-Cl pH7.4, 0.5mM DTT, Complete Protease Inhibitor). The nuclear suspension was then aliquoted and 5M NaCl was added to a final concentration of 200mM, 300mM, 400mM or 1M. Nuclei were rotated for 30 minutes at 4ºC and then spun at max speed in a table-top centrifuge for 10 minutes. The supernatant was taken as the nuclear extract and mixed with 2x Laemmli Buffer. Samples were then boiled and western blotting was performed as described above. Antibodies used were as above plus: sheep anti-Histone H1 (1:2000, GeneTex, GTX21938).

Luciferase Repression Assay.

The heterologous repression assay used in this study was designed to recapitulate the same assay used in the original report (Nan et al., 1997). We fused full-length MeCP2 at the N-terminus to the Gal4 DNA binding domain and placed the human beta-actin promoter downstream of a 5X

Gal4 binding site. Instead of using beta-galactosidase activity as a read-out we used luciferase to monitor gene expression from the promoter of both the reporter construct and a control reporter that lacked the 5X Gal4 binding site. N2a cells were transfected with 50ng of pUAS-hActin-luc2 and 10ng of the control vector pRL-TK using Lipofectamine 2000 (Invitrogen) along with the indicated effector constructs within pCMX-Gal4 (150ng). 150ng of pCDNA3.1(+) was used as an empty vector control containing the CMV promoter but no coding sequence. Lysates were prepared from transfected cells 24-36 hours post-transfection and assayed using the Dual Luciferase Reporter Assay System (Promega) per the manufacturer's instructions.

Gene Expression Analysis by RT-qPCR.

Mice were sacrificed under anesthesia and the hypothalamus was quickly dissected from the remaining brain tissue over ice. Tissue samples were homogenized in Trizol Reagent (Invitrogen) and processed per the manufacturer's instructions. Isolated RNA was reverse transcribed using M-MLV (Invitrogen) and random hexamer priming. The resulting cDNA was diluted and stored at -20ºC until needed. Quantitation of gene expression was carried out in duplicate reactions using iTaq SYBER Green Supermix (Bio-Rad) and a CFX96 Real-Time System (Bio-Rad) per the manufacturer's instructions. The following primers were used:

Gapdh: Forwards 5'-ggagattgttgccatcaacga-3'

Reverse 5'-tgaagacaccagtagactccacgac-3'

Bdnf: Forwards 5'-cgccatgcaatttccactatcaataatttaac-3'

Reverse 5'-ctttttcagtcactacttgtcaaagtaaac-3'

Sst: Forwards 5’-cccagactccgtcagtttct-3’

Reverse 5’-gaagttcttgcagccagctt-3’

Tac1: Forwards 5’-cgacagtgaccagatcaaggag-3’

Reverse 5’-gcccattagtccaacaaagg-3’

Oprk1: Forwards 5’-gtcatgtttgtcatcatccgata-3’

Reverse 5’-ggtggcacacagcaatgtag-3’

Mef2c: Forwards 5’-tccatcagccatttcaacaa-3’

Reverse 5’-aactgactgagggcagatgg-3’

Grin2a: Forwards 5’-cagagctcatccccaaagag-3’

Reverse 5’-caagtcacattgaccataaattgg-3’

Results were averaged for each sample and normalized to *Gapdh*. Relative expression levels of the genes of interest were determined by normalizing the fold expression level in each sample to the corresponding WT control. Data are shown as mean ± standard error of mean (SEM) and analyzed by one-way ANOVA with Bonferroni-Holm post hoc analysis.

Microarray Analysis.

Mice were sacrificed under anesthesia and the hippocampi were quickly dissected from the

remaining brain tissue over ice. Tissue samples were frozen by submersion in liquid nitrogen and stored at -80ºC. Frozen tissue was homogenized in PureZOL Reagent (Bio-Rad) and processed using an Aurum Total RNA Fatty and Fibrous Tissue Kit (Bio-Rad) per the manufacturer's instructions. Isolated RNA was eluted in water and submitted to the Genomic and RNA Profiling Core at Baylor College of Medicine. After polyA selection, gene expression was measured using GeneChip Mouse Gene 1.0 ST Arrays (Affymetrix).

The gene array data were processed by the Affymetrix Power Tools. Raw data were normalized with RMA normalization method. The gene annotation of the probesets on the array were obtained from Affymetrix. The significance of the Genotype effect was tested with ANOVA by comparing a model with both Genotype effect and Probeset effect and a model with Probeset effect only. By borrowing information between genes, variances were stabilized with an empirical Bayes method. When comparing gene expression between WT mice and KO, R270X, or G273X mice we determined significantly altered genes to be those with a $p_{adj}<0.05$ (FDR corrected) and an absolute $\log_2$(fold change) greater than 0.1. Genes that were significantly altered between mutants were determined by $p_{adj}<0.05$ without a threshold for fold change differences.

Pairwise Alignment of MeCP2 and HMGA1.

In order to determine a possible ancestral relationship between MeCP2 and the HMGA family, we performed a pairwise alignment between the C-terminus of MeCP2 (amino acids 163-486) and the HMGA family member HMGA1. The sequences were aligned with the PROMALES3D server (default settings) (Pei et al., 2007). To test the significance of the alignment, we utilized PairwiseStatSig method (Ankit Agrawal, 2011). The algorithm estimates the pairwise statistical

significance of a local alignment of two protein sequences independent of any database information.

Proteome Wide Comparison of MeCP2 and HMGA1 Similarity.

To further test the significance of the pairwise alignment, we aligned individual human protein sequences with each MeCP2 and HMGA1. The human set was constructed of the reviewed human protein sequences (nearly 20000) from Uniprot database (Consortium, 2012). The pairwise alignments were constructed individually with the MUSCLE algorithm (Edgar, 2004). The sequence identity was calculated ignoring gaps between the two sequences. In the end, two sets of pairwise alignments are constructed to assess the average similarity between a random human protein with either MeCP2 or HMGA1.

When aligning sequences of proteins with large differences in length, we would expect some increased sequence identity because longer protein sequences have more information. In order to test this assumption, the proteins were binned based on protein length (sets of 1000 protein) and the average sequence identity was calculated. **Figure S5A** shows the average sequence identity between MeCP2 and all human proteins as a function of protein length. For proteins with a comparable length to HMGA1 (107 amino acids), the average is near 26% (st.dev 4%). At 44%, the sequence identity between HMGA1 and MeCP2 is over three standard deviations from average. This analysis was repeated for HMGA1 versus all human proteins (**Figure S5C**). At the length of MeCP2, the average sequence identity between HMGA1 and all human proteins is near 30% (st.dev 4%). Again this confirms that 44% sequence identity is statistically significant and the relationship between MeCP2 and HMGA1 is symmetrical.

Next we examined the similarity in sequence between MeCP2 and proteins within 10% the length of HMGA1. These proteins were binned based on sequence identity. **Figure S5B** shows the count distribution of the relationship between MeCP2 and human proteins. The relationship between MeCP2 and HMGA1 is highly significant because when comparing proteins of similar length it is one of the few with a sequence identity over 40%. A second HMG group protein HMGA2 was also shown to have high identity (41%). This analysis was repeated (**Figure S5D**) with the HMGA1 and human proteins within 10% of the length of MeCP2 (total of 2411 protein). MeCP2 was found to be one of the few proteins with identity higher then 40%. Because we can see a predictable relationship between the average sequence identity and the length of the protein, we can develop a simple z-score that will subtract the noise from the sequence identity of proteins with different lengths. We can measure the non-randomness of the similarity in MeCP2 to all other human proteins

$$\frac{SI - \langle SI \rangle}{\sigma}$$

where *SI* is the actual sequence identity between a protein and MeCP2. <*SI*> is the random sequence identity expected from proteins with a similar length and σ is the standard deviation of the same sample. HMGA1 was found to be the highest scoring with z_{SI}=5.31. Proteins in the MBD family were also ranked highly (z-score > 4.0).

Electrophoretic Mobility Shift Assay.

A recombinant form of the human MeCP2 AT-Hook 2 domain was generated by expression in *E. coli* BL-21(DE3) cells following 1mM IPTG induction of either pGEX-5x3-hMeCP2-delN256-R270X or pGEX-5x3-hMeCP2-delN256-G273X. Recombinant proteins were purified on Glutathione Sepharose 4B beads (GE Healthcare) and eluted into Elution Buffer (50mM Tris-Cl

pH 8.0, 10mM glutathione) prior to dialysis into PBS. We checked the purity of recovered protein by Coomassie staining and then stored protein at -80ºC until needed.

A 64 base pair probe containing a single homopolymeric $(A•T)_5$ stretch was generated by annealing two single stranded DNA oligonucleotides at 95ºC and allowing them to cool slowly to room temperature. The sequences were:

T5: Forward 5'- ggactccaggtccaggaccgcgtttttcgcgcgcacggcgcgggaggtccagctgtccacctcc-3'

T5: Reverse 5'- ggaggtggacagctggacctcccgcgccgtgcgcgcgaaaaacgcggtcctggacctggagtcc-3'

For EMSA reactions recombinant protein at the concentration indicated in the figure was mixed in a 30µL reaction with 66.7nM of probe, 0.1mg/mL BSA, and EMSA Buffer (10mM Tris-Cl pH 7.4, 50mM KCl, 0.5mM $MgCl_2$, 0.1mM EDTA, 5% glycerol) for 30 minutes at 25ºC. Reactions were then loaded onto a pre-electrophoresed 6% acrylamide/bis (37.5:1) gel in 0.5xTBE and run at 100V at 4ºC. Gels were then stained with ethidium bromide and reaction products were visualized under ultraviolet illumination.

The apparent Kd for AT-Hook 2-G273X was calculated by plotting the densitometric data on a double reciprocal plot and fitting a linear regression line.

Native ChIP-qPCR.

Unfixed chromatin was prepared from whole brains as described (Fan et al., 2005). Briefly, whole brain tissue was washed in cold PBS and resuspended in RSB buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl2, 0.2% NP-40, protease inhibitors). Nuclei pellets were digested with micrococcal nuclease (MNase) in the presence of 1 mM CaCl2. Reactions were

terminated by adding 10 mM Na-EDTA. Soluble chromatin solutions were collected and used for subsequent experiments as described above for cross-linked ChIP.

Nucleosomal Array Reconstitution.

NAs were generated using a HindIII-XbaI fragment of the pUC19-601x12 plasmid described in (Nikitina et al., 2007). After CpG methylation with M.SssI (NEB), 12.5µg of purified 601x12 sequence was mixed with chicken erythrocyte histones (GenWay Biotech), at molar ratios ranging from 0.5 to 4 histone octamers per 601 repeat, in salt-dialysis tubing containing 2M NaCl, 10mM Tris-Cl pH 7.4, and 1mM EDTA. The reaction was dialyzed progressively into 1M, 0.8M, 0.6M, and finally twice into 2.5mM NaCl containing 10mM Tris-Cl pH 7.4 and 1mM EDTA before being aliquoted and stored at 4ºC prior to use. Proper loading of nucleosomes onto the DNA template was monitored by performing electrophoresis in a 0.8% agarose gel in 0.2xTBE. Reactions containing reconstituted arrays just prior to saturation were used for experiments.

Nucleosomal Array Electrophoretic Mobility Shift Assay.

A recombinant human MeCP2 was generated by expression in *E. coli* BL-21(DE3) cells following 0.5mM IPTG induction of pTXB1-hMeCP2, pTXB1-hMeCP2-R270X, or pTXB1-hMeCP2-G273X. Recombinant proteins were purified using the IMPACT kit (NEB), per the manufacturer's recommendations prior to being dialyzed into 0.33xPBS. We checked the purity of recovered protein by Coomassie staining and then stored protein at -80ºC until needed.
For EMSA reactions, recombinant MeCP2 at 0, 10, 20, or 40nM was mixed in a 20µL reaction with CpG-methylated nucleosomal arrays (20nM per nucleosome repeat), 0.1mg/mL BSA, and

EMSA Buffer (10mM Tris-Cl pH 7.4, 50mM KCl, 0.5mM $MgCl_2$, 0.1mM EDTA, 5% glycerol) for 30 minutes at 25ºC. Reactions were then loaded onto a 0.8% agarose gel in 0.2xTBE and run at 100V at room temperature. Gels were then stained with ethidium bromide and reaction products were visualized under ultraviolet illumination.

Nucleosomal Array Oligomerization Assay.

Recombinant MeCP2 was generated as above (see Nucleosomal Array Electrophoretic Mobility Shift Assay).

In a 20µL reaction MeCP2-WT, MeCP2-R270X, or MeCP2-G273X at 80nM was mixed with CpG-methylated nucleosomal arrays (20nM per nucleosome repeat), in buffer containing 3mM Tris-Cl pH 7.4 and 0, 2.5, 5, 10, or 20mM $MgCl_2$. The reaction was allowed to equilibrate at room temperature for 30 minutes prior to incubation on ice for 20 minutes. The reactions were then spun at 13,200 rpm in a table-top centrifuge at 4ºC. The supernatant was then separated and either directly loaded onto a 1% agarose gel in 1xTBE containing 0.5µg/mL ethidium bromide (**Figures S5F and S5G**) or adjusted to 1% SDS and 0.2mg/mL proteinase K. Proteinase K digestion was carried out at 55ºC for 1 hour. The digested samples were then loaded onto a 1% agarose gel in 1xTBE containing 0.5µg/mL ethidium bromide (**Figures 5F and 5G**). The resulting DNA from was then visualized under ultraviolet illumination.

***In Vivo* Immunofluorescence.**

For immunofluorescence staining of ATRX and MeCP2, brains from freshly sacrificed mice were halved and submerged in cold phosphate buffered saline (PBS) containing 4% paraformaldehyde. The tissue was allowed to fix overnight at 4ºC and then transferred into a

30% sucrose solution for cryoprotection. Cryoprotected brains were immersed in Optimal Cutting Temperature compound (Tissue-Tek) and frozen over dry ice. Tissue was then stored at -80ºC until needed. Sagittal sections were cut in a Leica CM3050S at 50µm and washed in PBS. Washed sections were then blocked in Blocking Solution (2% normal goat serum, 0.3% Triton X-100, PBS) for 1 hour prior to the addition of primary antibody in Blocking Solution. Sections were incubated with primary antibody at 4ºC for 48 hours and then washed four times prior to the addition of secondary antibody in Blocking Solution. Sections were then incubated at 4ºC for 48 hours with the secondary antibody before a final four washes. Stained sections were mounted onto glass slides in Vectashield Mounting Media with DAPI (Vector Laboratories). The following antibodies were used: mouse anti-MeCP2 Men-8 which recognizes the C-terminus (1:200, Sigma-Aldrich, M7443), rabbit anti-ATRX (1:200, Abcam, ab97508), goat anti-mouse Alexa 488 (1:1000, Invitrogen, A11029), goat anti-rabbit Alexa 633 (1:1000, Invitrogen, A-21071).

Staining was performed on mice from three age groups: 4, 7, and 9 weeks old. Staining and imaging within each age group was always carried out in batches that included at least 1 WT and 1 KO control. Imaging was performed with a Leica SP5 Confocal Microscope using identical settings for each age group batch. Image data was processed using identical settings for each age group with ImageJ software and the Foci Picker3D plugin to identify foci and quantify the corresponding average intensity of each focus. The number of foci per cell was calculated per high-power-field (63x lens) and averaged across 3 fields per animal for a total of 6 hpf per genotype per age. The data were plotted as mean ± SEM for each genotype. The average intensity of each focus as calculated by Foci Picker3D was plotted for 4 week old animals as the mean ± SEM of all foci identified per genotype normalized to WT.

Coimmunoprecipitation.

N2a cells were cultured in DMEM +10% FBS and pen/strep at 37ºC in 5% CO_2. On the day prior to transfection cells were split into 6-well plates. Cells were transfected with 1µg of pDest-hMeCP2-GFP, pDest-hMeCP2-GFP-R270X, pDest-hMeCP2-GFP-G273X, or pCDNA3.1(+) using Lipofectamine 2000 (Invitrogen) per the manufacturer's instructions. After 36 to 48 hours cells were harvested in PBS and the cell pellet was lysed High Salt Cell Lysis Buffer (0.5% NP-40, 20mM Tris-Cl pH 8.0, 500mM NaCl, 1mM EDTA, protease inhibitors). Complexes were allowed to extract for 20 minutes at 4ºC and then insoluble material was pelleted at 13,200 rpm in a table top centrifuge. The resulting supernatant was then diluted 1:3 with 10mM Tris-Cl pH 7.4. After saving a sample for western analysis the remaining input was mixed with 1µg affinity purified rabbit anti-GFP (Abcam, ab65566) bound to 25µL of Protein A Dynabeads (Invitrogen) and rotated at 4ºC for 1-3 hours. The beads were then washed four times and bound complexes eluted by boiling in 2x Laemmli buffer. Western analysis was performed as described above. The following antibodies were used: rabbit anti-ATRX (1:1000, Santa Cruz, H300), mouse anti-GFP (1:2000, Neuromab, N86/8), mouse anti-GAPDH 6C5 (1:20000, Advanced Immunochemicals, 2-RGM2), mouse anti-rabbit HRP (1:3000, Jackson ImmunoResearch Labs, 211-032-171), and donkey anti-mouse HRP (1:100000, Jackson ImmunoResearch Labs, 715-035-150).

Immunofluorescence on Fresh Nuclei.

A WT and KO mouse were sacrificed simultaneously under anesthesia and organs were quickly washed in PBS and removed to a cold glass plate over ice. Tissue was then processed immediately without freezing or storage. One hemi-brain or the equivalent tissue volume of each

peripheral organ was dounce homogenized by 20 strokes with a tight pestle in Hypertonic Buffer A (300mM Sucrose, 2mM $MgCl_2$, 3mM $CaCl_2$, 10mM Tris-Cl pH 8.0, 0.1% Triton X-100, 0.5mM DTT, Complete Protease Inhibitor). Cells were allowed to lyse for 5 minutes on ice and then centrifuged at 720 x g for 5 minutes. The nuclei were then washed twice in cold Hypertonic Buffer A and brought up in cold PBS with Complete Protease Inhibitor and 0.5mM DTT. Fresh purified nuclei were then incubated with rabbit anti-ATRX (1:250, Abcam, ab97508) for 2 hours at 4ºC, washed 3 times in PBS and then incubated with goat anti-rabbit Alexa 633 (1:1000, Invitrogen, A-21071) for 1 hour at 4ºC. Stained nuclei were then washed 3 times in cold PBS and aliquoted onto Lab-Tek Chambered Coverglass (Nunc). Nuclei were then treated with Hoechst 33342 at 5ug/mL for 20 minutes at 25ºC and imaged using a Leica SP5 Confocal Microscope. Image data was processed using ImageJ software. The 95% confidence intervals for the percentage of nuclei containing ATRX foci were calculated using the Agresti-Coull method under the "binom" package for R.

Animals

WT;R270X and WT;G273X transgenic mice were maintained on a pure FVB background. *Mecp2* KO mice, carrying *Mecp2*$^{tm1.1Bird}$ allele (Guy et al., 2001) were backcrossed and maintained on a pure 129SvEvTac background. Male WT;R270X or WT;G273X FVB mice were crossed to *Mecp2* KO 129SvEvTac females and the resulting F1 FVB;129SvEv male progeny were used for further analysis.

Phenotypic Analyses of Mice

Mice were evaluated for each of six phenotypic categories as previously described (Guy et al.,

2007). In brief, mice were scored from 0-2 for each of six categories: tremor, gait abnormalities, immobility, hind-limb clasping, breathing, and general condition, where 0 represents the absence of symptoms and 2 represents severe symptoms. These scores were then summed for a cumulative severity score ranging from 0-12.

Supplementary References

Ankit Agrawal, X.H. (2011). Pairwise Statistical Significance of Local Sequence Alignment Using Sequence-Specific and Position-Specific Substitution Matrices. IEEE/ACM Transactions on Computational Biology and Bioinformatics *8*, 194-205.

Bolte, S., and CordeliÈRes, F.P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. Journal of Microscopy *224*, 213-232.

Consortium, T.U. (2012). Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Research *40*, D71-D75.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research *32*, 1792-1797.

Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultchi, A.I. (2005). Histone H1 Depletion in Mammals Alters Global Chromatin Structure but Causes Specific Changes in Gene Regulation. Cell *123*, 1199-1212.

Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of Neurological Defects in a Mouse Model of Rett Syndrome. Science *315*, 1143-1147.

Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet *27*, 322-326.

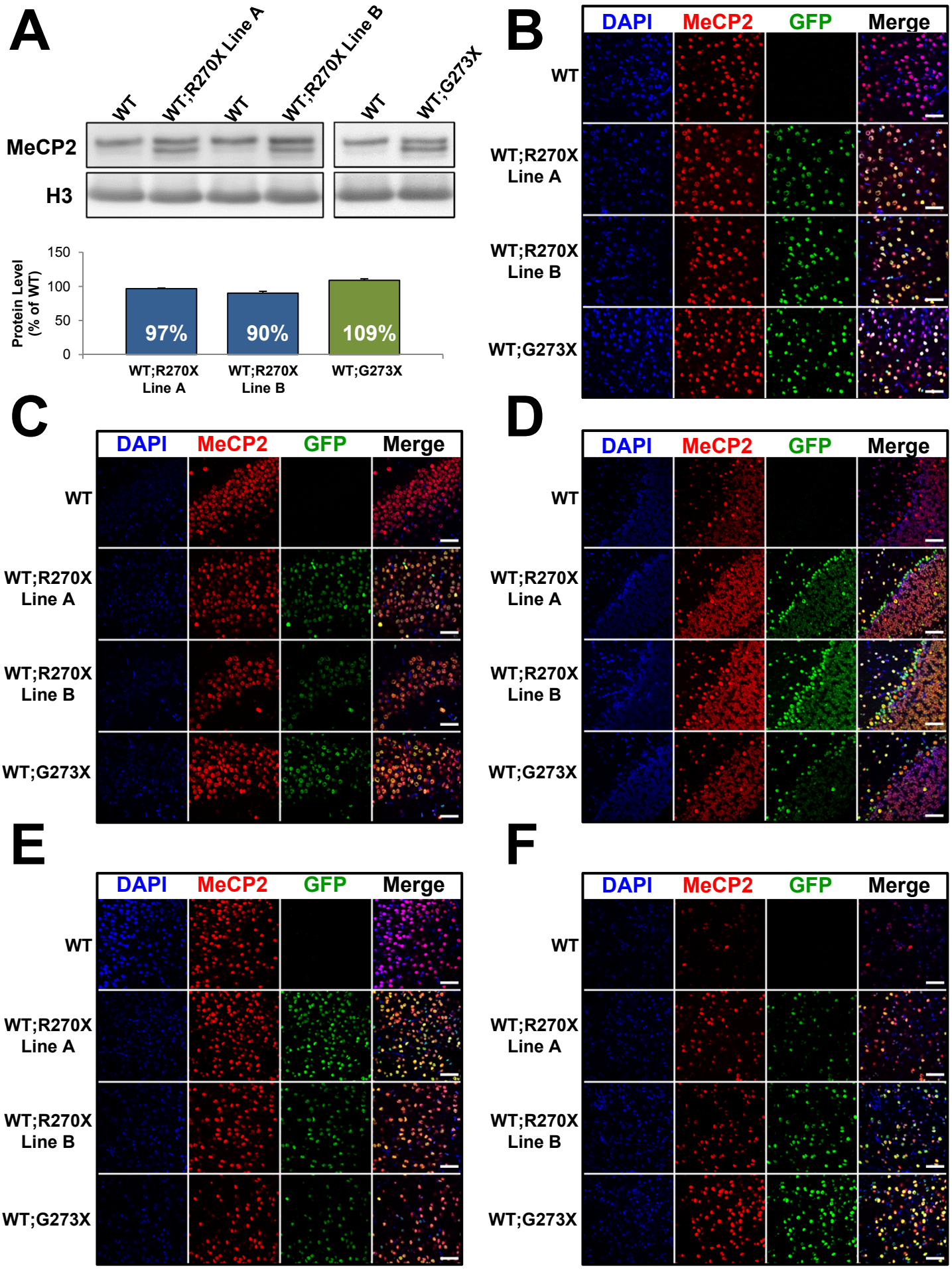
Klisch, T.J., Xi, Y., Flora, A., Wang, L., Li, W., and Zoghbi, H.Y. (2011). In vivo Atoh1 targetome reveals how a proneural transcription factor regulates cerebellar development. Proceedings of the National Academy of Sciences *108*, 3288-3293.

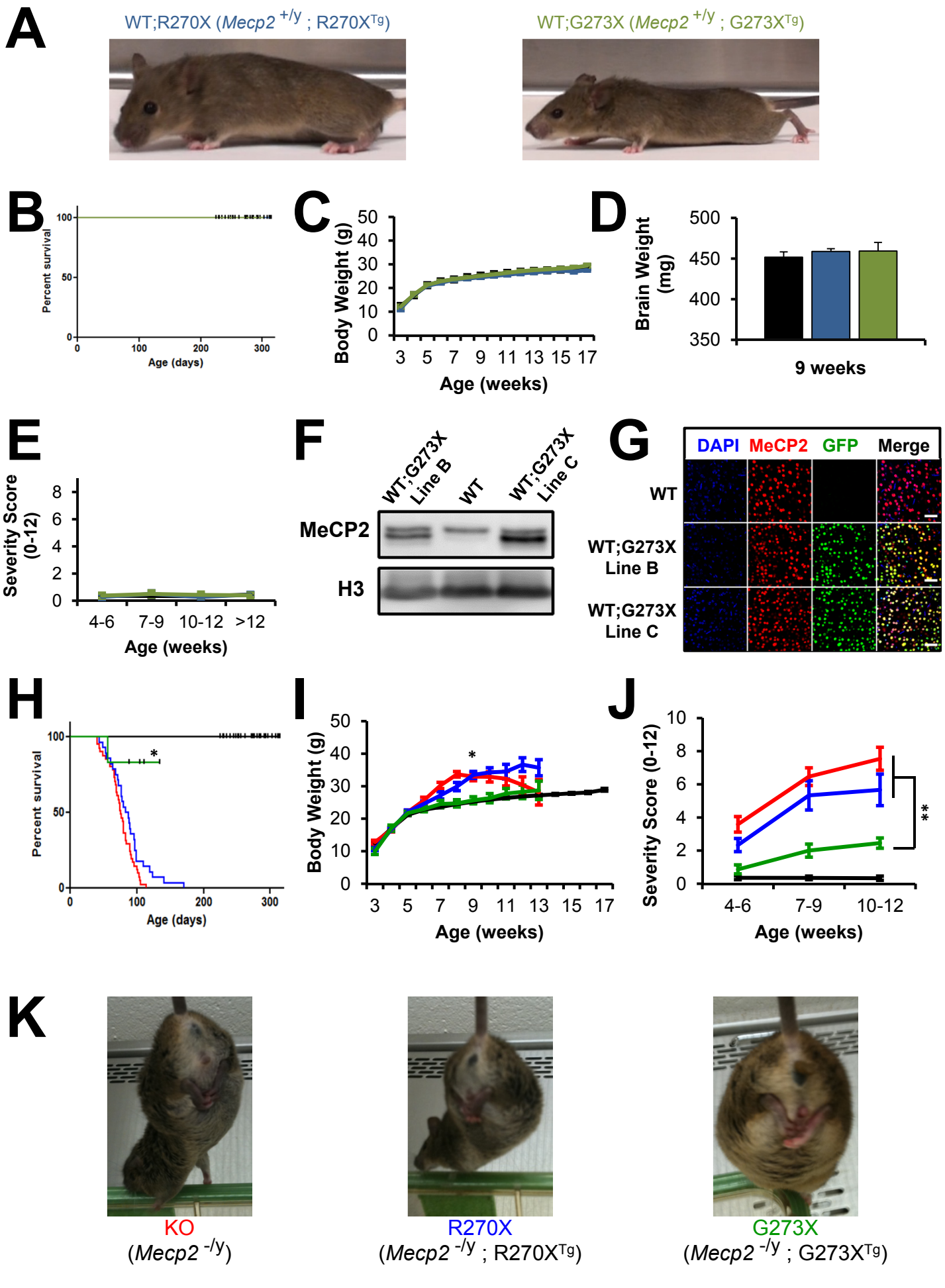
Nikitina, T., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., Grigoryev, S.A., and Woodcock, C.L. (2007). MeCP2-Chromatin Interactions Include the Formation of Chromatosome-like Structures and Are Altered in Mutations Causing Rett Syndrome. Journal of Biological Chemistry *282*, 28237-28245.

Pei, J., Kim, B.-H., Tang, M., and Grishin, N.V. (2007). PROMALS web server for accurate multiple protein sequence alignments. Nucleic Acids Research *35*, W649-W652.

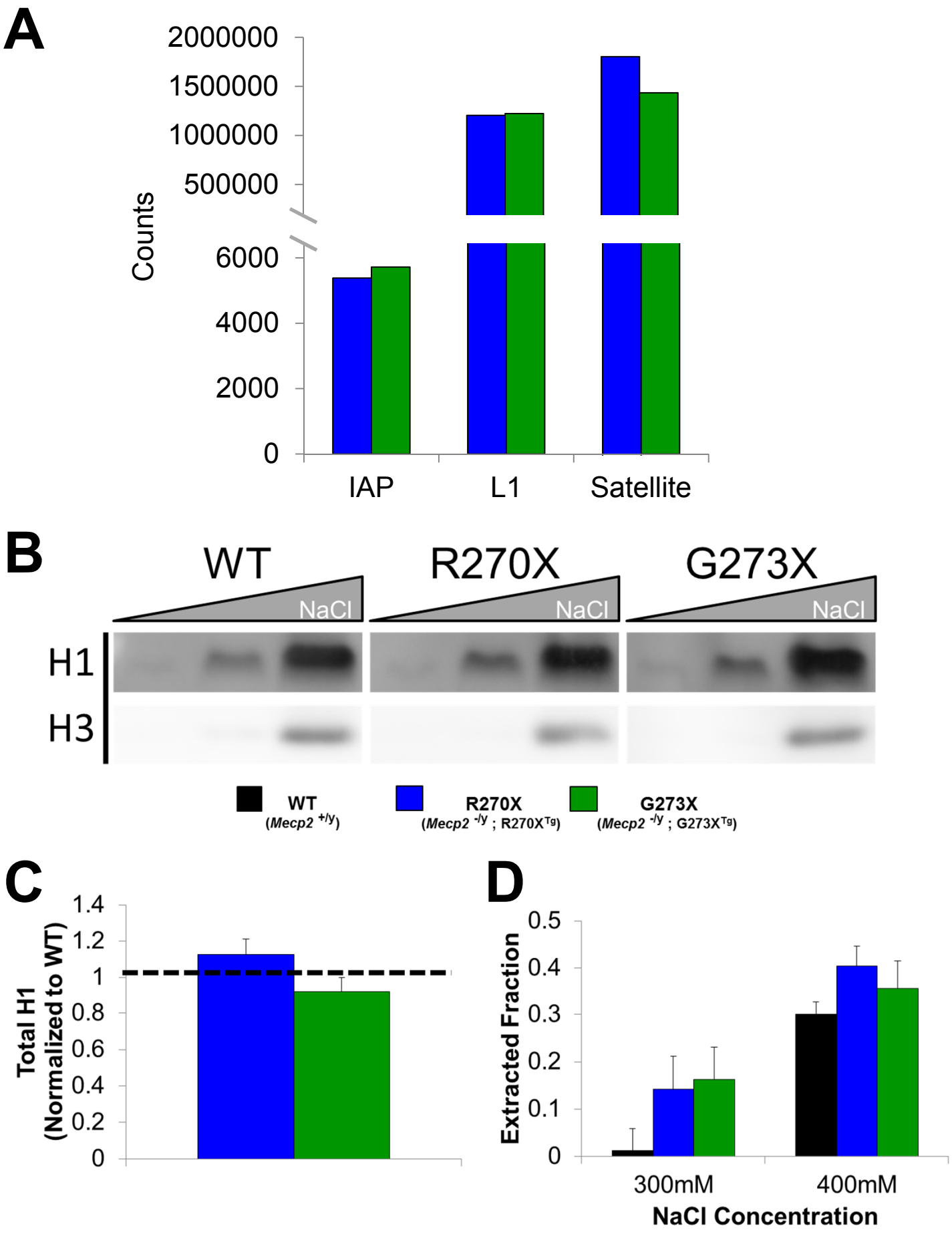
Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Andrews, R., and Bird, A.P. (2010). Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. Molecular Cell *37*, 457-468.

Suppl. Fig 1.pdf

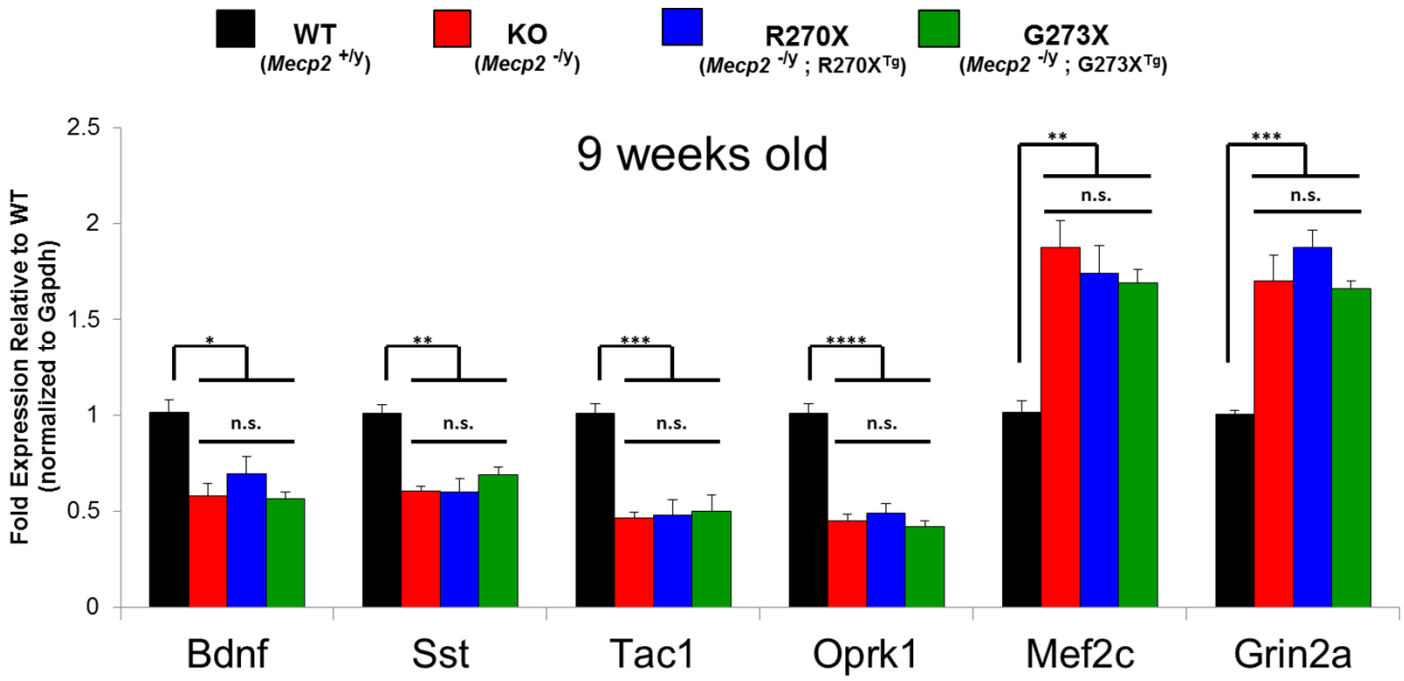



A
WT;R270X (Mecp2 +/y ; R270XTg)
WT;G273X (Mecp2 +/y ; G273XTg)
B
Percent survival
Age (days)
C
Body Weight (g)
Age (weeks)
D
Brain Weight (mg)
9 weeks
E
Severity Score (0-12)
4-6
7-9
10-12
>12
Age (weeks)
F
WT;G273X Line B
WT
WT;G273X Line C
MeCP2
H3
G
DAPI
MeCP2
GFP
Merge
WT
WT;G273X Line B
WT;G273X Line C
H
Percent survival
Age (days)
I
Body Weight (g)
Age (weeks)
J
Severity Score (0-12)
Age (weeks)
K
KO
(Mecp2 -/y)
R270X
(Mecp2 -/y ; R270XTg)
G273X
(Mecp2 -/y ; G273XTg)

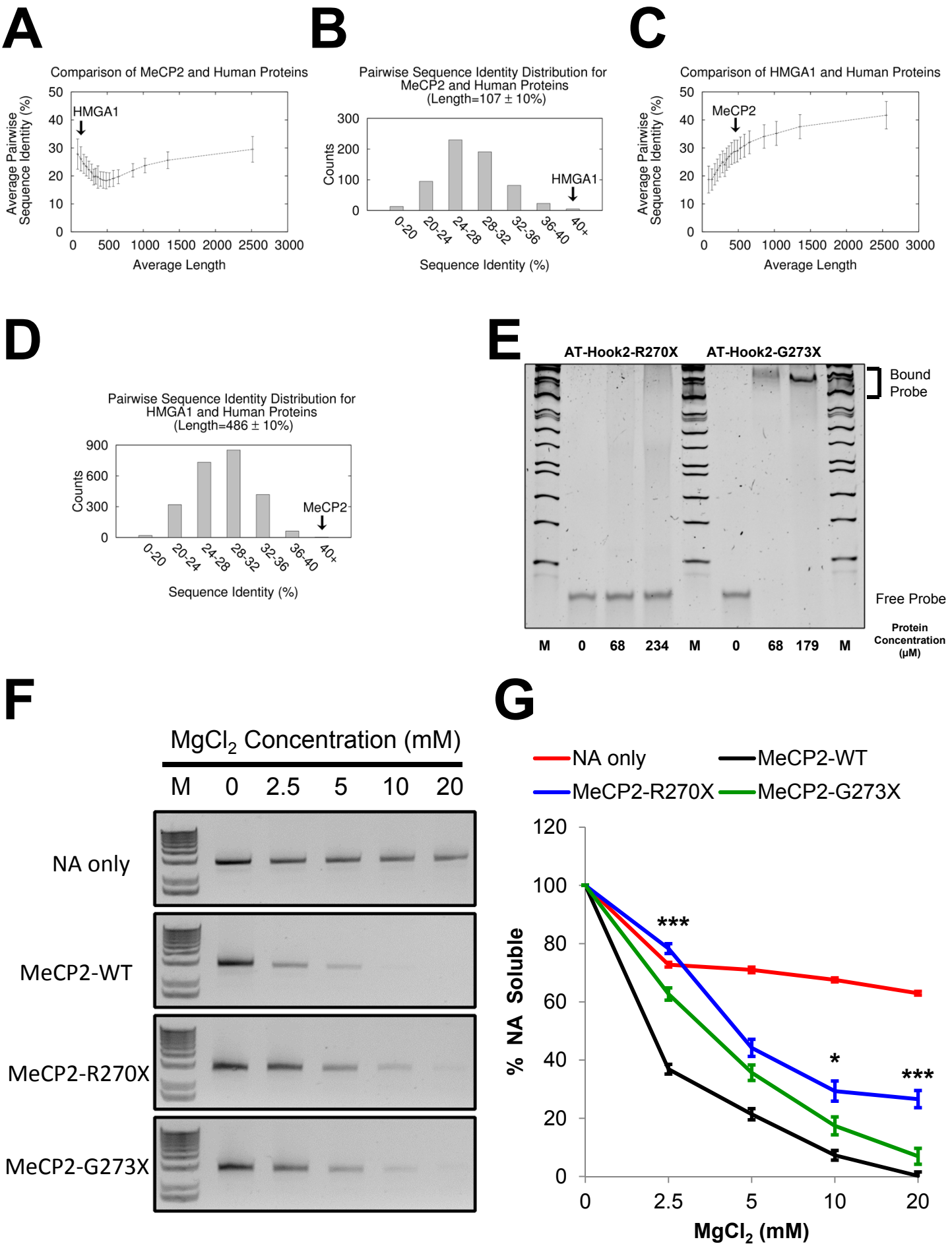
Suppl. Fig 3.pdf



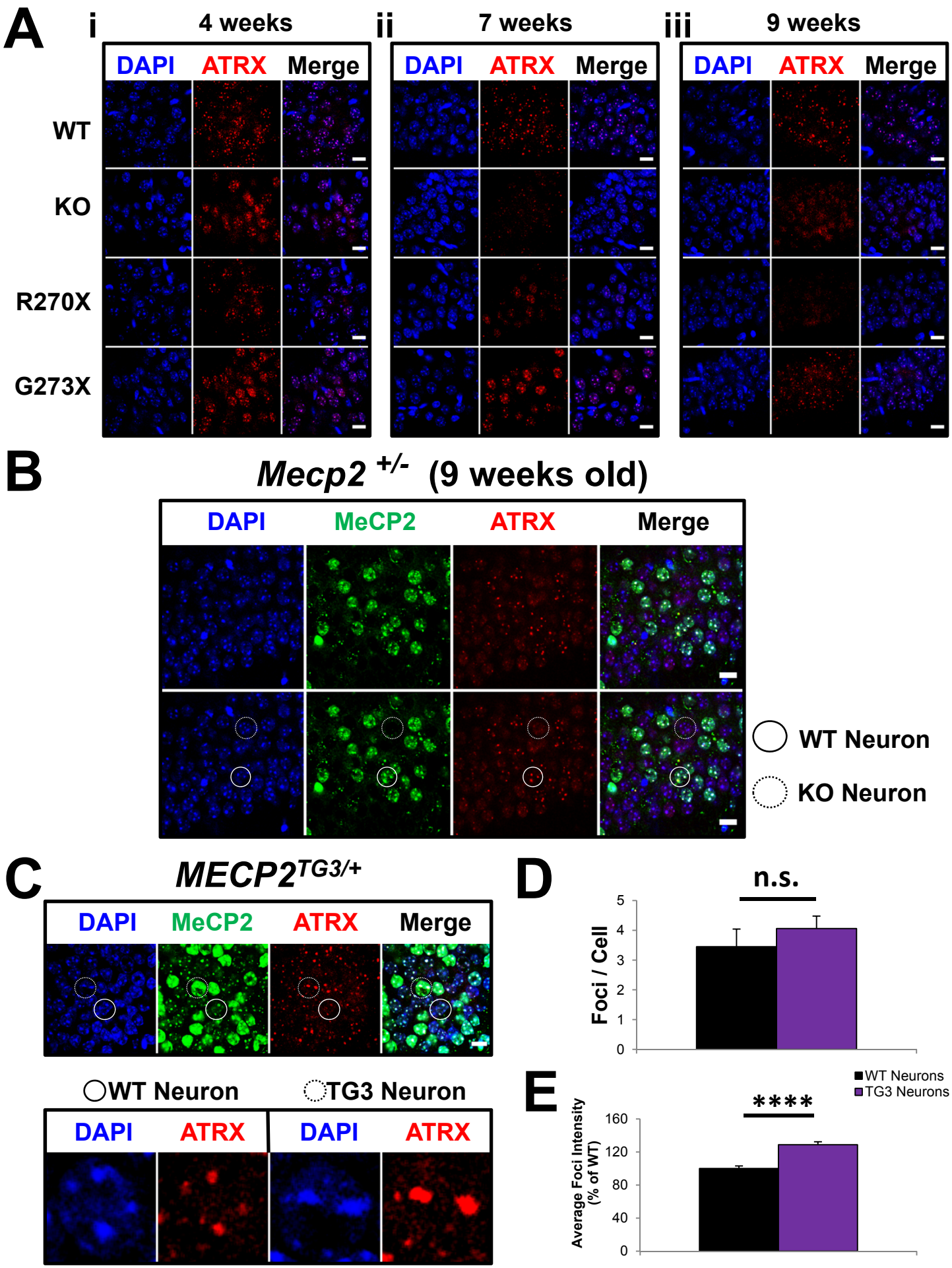
Suppl. Fig 4.pdf

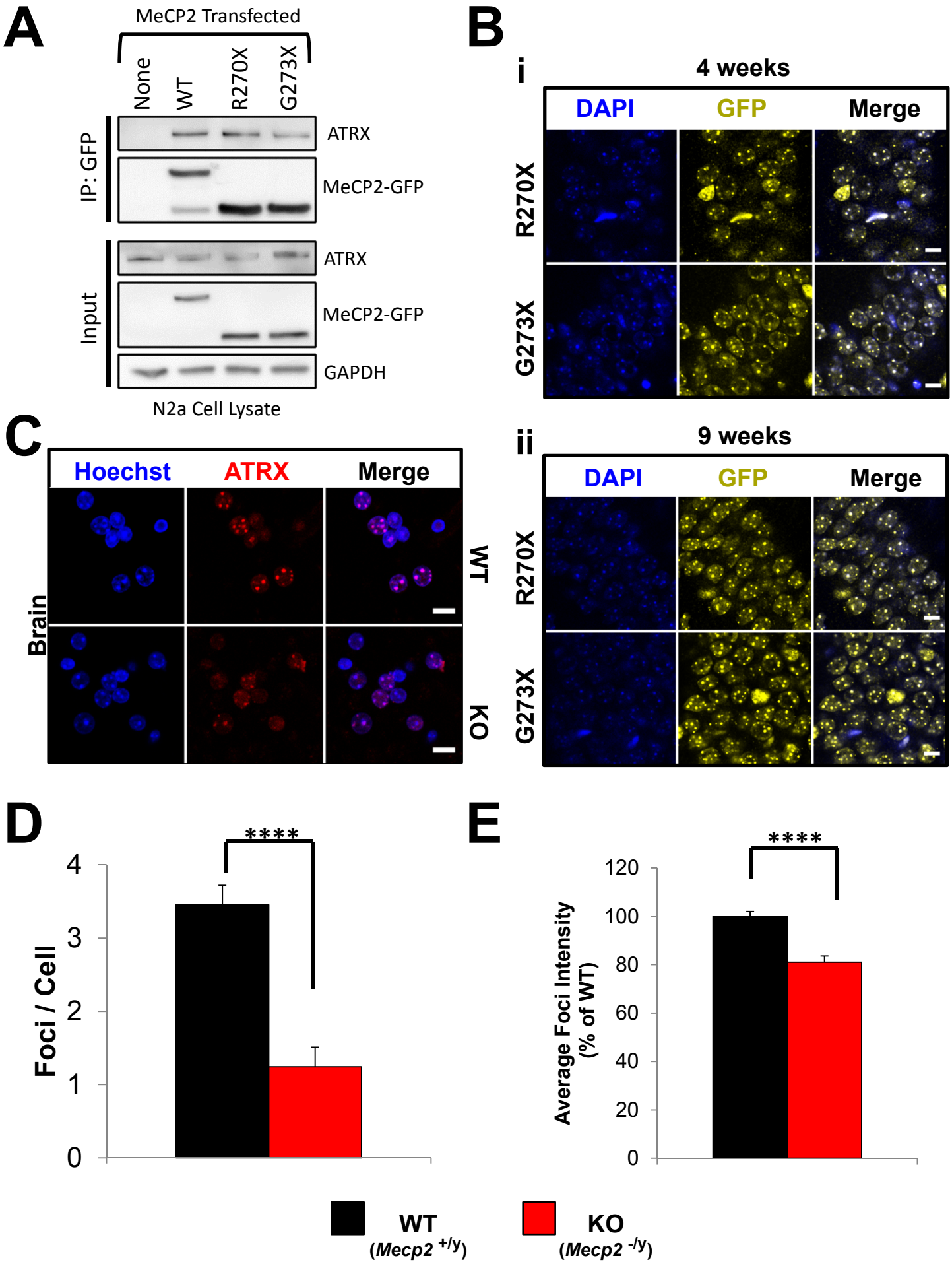


Suppl. Fig 5.pdf



Suppl. Fig 6.pdf




A
MeCP2 Transfected
None
WT
R270X
G273X
IP: GFP
ATRX
MeCP2-GFP
Input
ATRX
MeCP2-GFP
GAPDH
N2a Cell Lysate
B
i
4 weeks
DAPI
GFP
Merge
R270X
G273X
ii
9 weeks
DAPI
GFP
Merge
R270X
G273X
C
Hoechst
ATRX
Merge
Brain
WT
KO
D
****
Foci / Cell
0
1
2
3
4
E
****
Average Foci Intensity (% of WT)
0
20
40
60
80
100
120
WT (Mecp2 +/y)
KO (Mecp2 -/y)