Inventory of Supplemental Data

Supplemental Figure S1 – Provides the reader with information regarding the quantification of MeCP2 and histone H4 in brain and liver. These details were requested by the reviewers and have been subsequently inserted. This includes: time-course of MeCP2 accumulation in the brain, quantification of recombinant MeCP2; details of the calculations used in determination of molecular abundance; verification of MeCP2 quantification using an alternative primary antibody; quantification of MeCP2 in liver; quantification of histone H4 in brain.

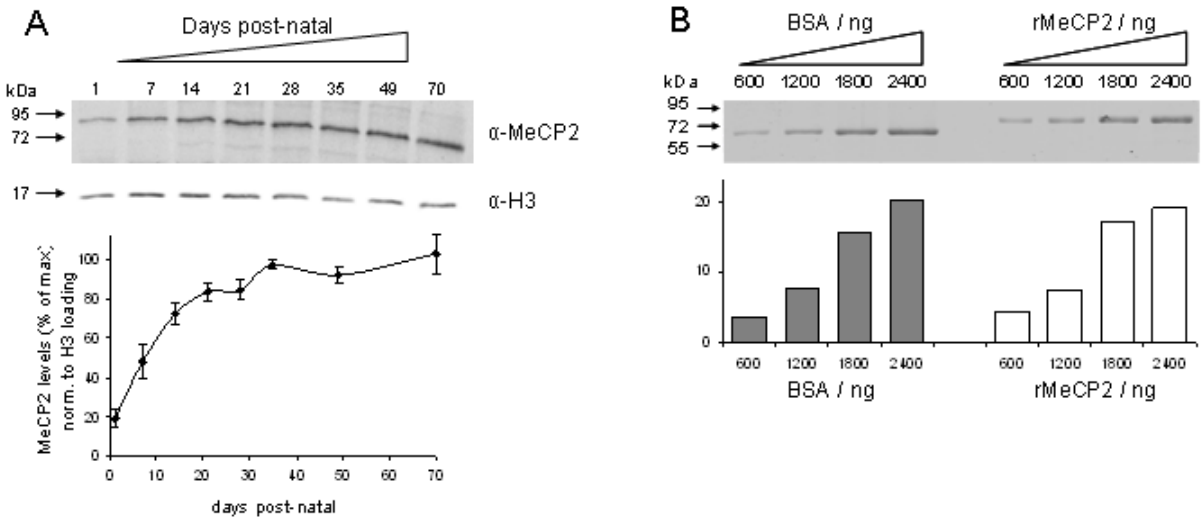
Supplementary Figure S2 – Bioinformatic analysis of the CpG density of mouse CpG islands. This was used in the analysis of MeCP2 ChIP-seq data to show reduced MeCP2 binding over the CpG island portion of the genome.

Supplementary Figure S3 – Histone H3 acetylation profile across the *Xist* locus. This indicates the role of MeCP2 in suppressing acetylation within the bulk genome, supporting the information within the main text.

Supplementary Figure S4 – Supporting information regarding the role of MeCP2 in suppressing spurious transcription from repetitive elements within the bulk chromatin.

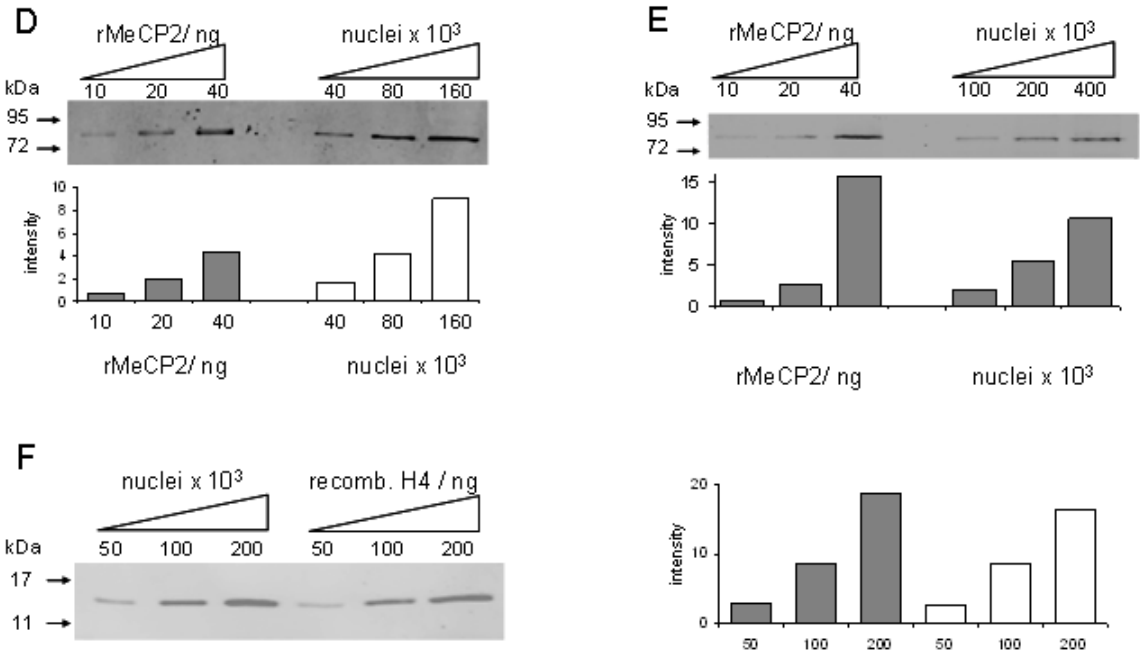
Supplemental Figures and Legends

Supplementary Figure S1



C

Tissue	Sub-population	ng / 1000 nuclei	fmol / nucleus	molecules / nucleus
(A) MeCP2 Brain	Unsorted	0.53	0.010	6 x 10^6
	Glia"	0.22	0.0042	2.5 x 10^6
	Neuron	1.35	0.026	15.5 x 10^6
Liver	-	0.045	0.00086	0.5 x 10^6
(B) H4 Brain	Unsorted	1.2	0.11	64 x 10^6



Supplemental Figure S1 – related to Figure 1

(A) MeCP2 expression in the mouse brain increases dramatically after birth, reaching maximal protein levels at ~5 weeks of age. For each timepoint, three mouse brains were pooled and homogenised. The homogenate was added directly to Laemmli buffer and boiled for 3 min and used directly for gel electrophoresis. Samples were separated by SDS-PAGE and quantitative western blotting was performed for MeCP2 and histone H3 as a loading control using infra-red imaging (Licor Odyssey). Molecular weight markers (kDa) are arrowed. Westerns were quantified and the data represented as % maximal MeCP2 expression, normalised for differences in loading based on H3 quantification; error bars indicate +/- SEM.

(B) The concentration of recombinant MeCP2 was determined by quantification against a bovine serum albumin (BSA) standard. A titration of a known concentration of BSA was compared to recombinant MeCP2 protein by SDS-PAGE and subsequent coomassie staining. The graph depicts densitometry analysis using Licor Odyssey imaging. The positions of molecular weight markers (kDa) are indicated.

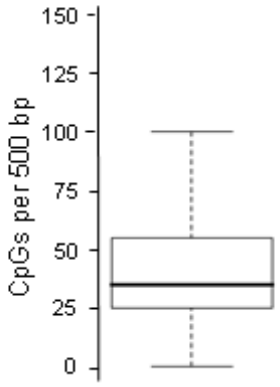
(C) Summary of the quantification of (A) MeCP2 and (B) Histone H4 absolute abundance in brain and liver nuclei. Using infra-red western blotting (Licor Odyssey), a titration of a known concentration of recombinant protein was compared to isolated nuclei (expressed as ng / 1000 nuclei). The number of mols per nucleus was determined using the molecular weight of the protein (MeCP2: 52.3 kDa; H4: 11.4 kDa). From this value the absolute number of molecules per nucleus was calculated using Avogadro's number (6×10^{23} molecules / mol).

(D) Verification of the absolute abundance of MeCP2 in total brain nuclei using an alternative antibody (674; Nan et al., 1998). A titration of a known concentration of recombinant MeCP2 and unsorted brain nuclei were compared using quantitative infra-red western blotting. The graph below indicates densitometric analysis of the western blot using Licor Odyssey software. This shows that there are 0.48 ng of MeCP2 per 1000 total brain nuclei, which is equivalent to 5.5×10^6 molecules of MeCP2 per nucleus.

(E) MeCP2 is expressed at much lower amounts in liver nuclei. A titration of a known concentration of recombinant MeCP2 was compared to nuclei isolated from mouse liver by infra-red western blotting. The graph below indicates densitometric analysis using Licor Odyssey. The positions of molecular weight markers (kDa) are indicated.

(F) Quantification of the absolute abundance of histone H4 in total brain nuclei. A titration of a known concentration of recombinant H4 and unsorted brain nuclei were compared using quantitative infra-red western blotting. The graph to the left indicates densitometric analysis of the western blot using Licor Odyssey software. This shows that in total brain nuclei indicates that there are 64×10^6 molecules of H4 per nucleus, which is equivalent to 32×10^6 nucleosomes.

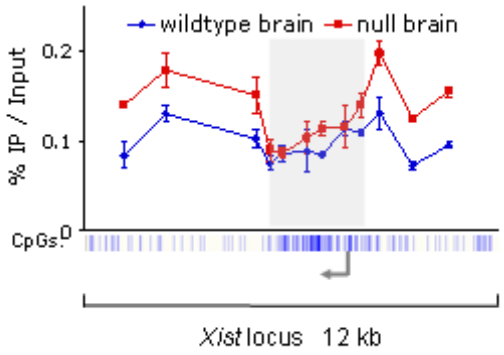
Supplementary Figure S2



Supplemental Figure S2 – related to Figure 4

CpG islands (CGIs) are characterised by having a high a CpG density. Boxplot showing the number of CpGs per 100 bp calculated using the NCBI strict algorithm for CGIs. (www.ncbi.nlm.nih.gov/mapview/static/humansearch.html#cpg)

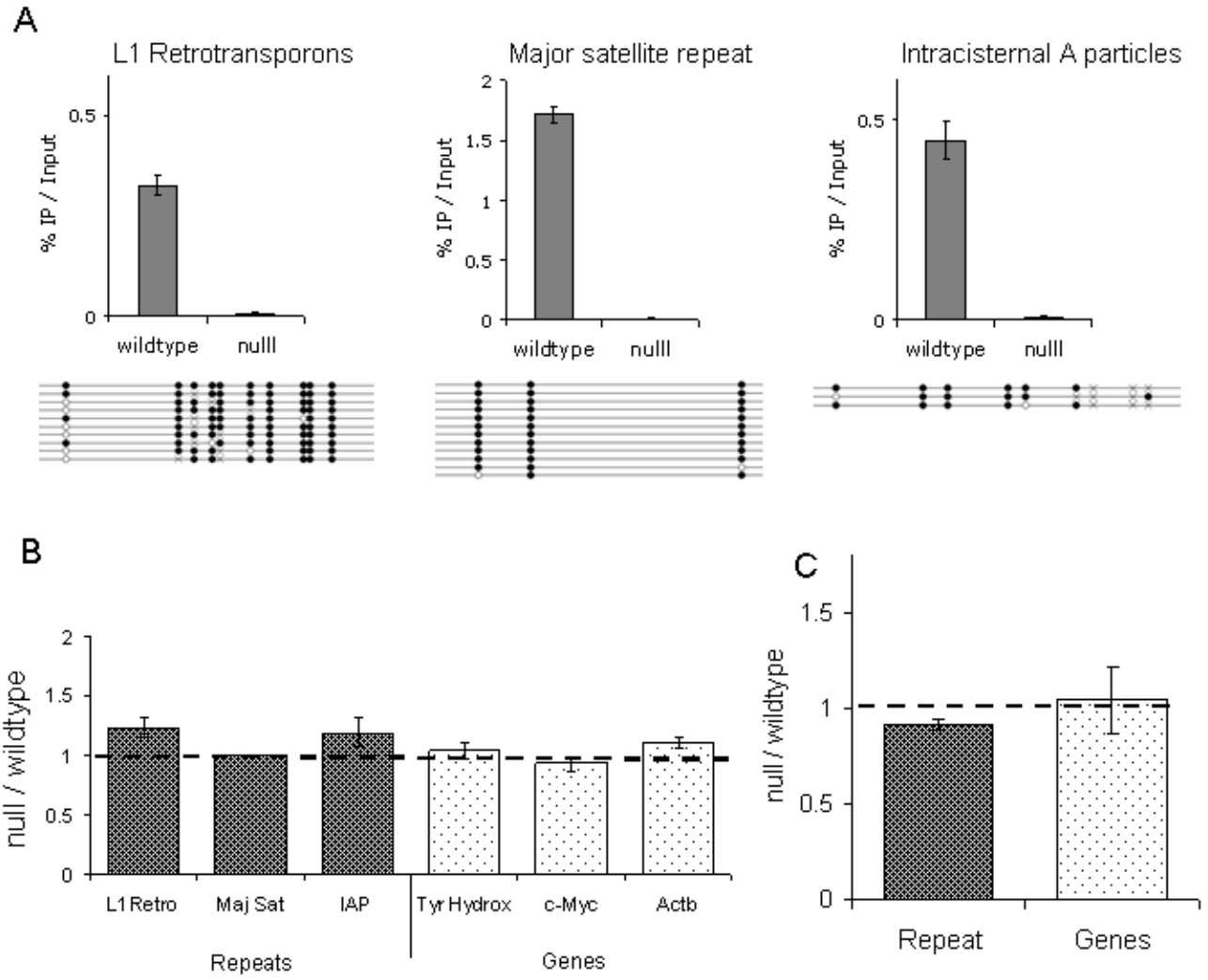
Supplementary Figure S3



Supplemental Figure S3 – related to Figure 5

MeCP2 impacts the histone H3 acetylation levels in the bulk genome flanking the *Xist* CpG island but does not affect the *Xist* promoter region itself. H3Ac ChIP profile across the promoter region of *Xist* showing wildtype brain (blue) and *Mecp2*-null brain (red). The transcription start site is indicated by an arrow. The blue vertical lines below the graph indicate CpG sites.

Supplementary Figure S4



D

Class	Locus	Average expression level relative to GAPDH		Standard Deviation	
		Wildtype	*Mecp2*-null	Wildtype	*Mecp2*-null
Repeats	L1 Retrotransposons	3.42	5.02	1.65	2.15
	Major Satellite	0.0662	0.113	0.0318	0.0378
	Intracisternal A Particles	0.481	0.724	0.217	0.250
Genes	*Tyrosine Hydroxylase*	0.00397	0.00345	0.00222	0.00125
	c-Myc	0.00233	0.00250	0.000987	0.000893
	Actin B	0.529	0.595	0.306	0.309

Supplemental Figure S4 – related to Figure 6

(A) Repetitive elements are distributed throughout the genome and are largely methylated and bound by MeCP2. MeCP2 ChIP followed by quantitative PCR using wildtype mouse brain shows significant binding to 3 classes of repeats as indicated. *MeCP2*-null brain was used as a negative control. Bisulphite sequencing of mouse genomic DNA indicates that these repeats are largely methylated. Each line represents a single clone. Open and filled circles indicate non-methylated and methylated CpG sites, respectively. Crosses indicate uncharacterised CpG sites

(B) Using RNA extracted from a whole brain there is no difference in expression of repetitive elements between wildtype and *Mecp2*-null mice. RNA was extracted from whole mouse brain and cDNA prepared, with and without reverse transcriptase. Quantitative PCR was used to determine levels of expression of repetitive regions and gene regions. The data was normalised to GAPDH and shown as a ratio between *Mecp2*-null and wildtype mice. The horizontal line marks no change between wildtype and *Mecp2*-null mice.

(C) Elevated expression of repetitive elements is only apparent in mature mice, where MeCP2 is highly abundant. Nuclei were prepared from wildtype and *Mecp2*-null E18.5 embryonic brains. RNA was extracted and cDNA was prepared, with and without reverse transcriptase. Quantitative PCR was used to determine the expression levels of various regions. The data was normalised to GAPDH and grouped as repetitive elements and gene regions. Data shown as a ratio between *Mecp2*-null and wildtype nuclei; the horizontal line marks no change between wildtype and *Mecp2*-null mice

(D) Raw expression levels of repetitive and genic regions. Nuclei were prepared from wildtype and *Mecp2*-null brains. RNA was extracted and cDNA was prepared, with and without reverse transcriptase. Quantitative PCR was used to determine the expression levels of various regions. The data was normalised to GAPDH as shown here, standard deviations are included. The ratio of *Mecp2*-null to wildtype was then determined and shown in Fig 6.

Detailed Supplemental Protocols

List of Antibodies Used

Antibody	Source; Cat no.	Application	Notes
Anti-MeCP2	Sigma-Aldrich; Mec-168	WB	
Anti-MeCP2	Upstate; 07-013	FACS	
Anti-MeCP2	Self-made rabbit polyclonal; 674	WB; ChIP	(Nan et al., 1998)
Anti-H4	Abcam; ab7311	WB	
Anti-H3	Abcam; ab1791	WB	
Anti-NeuN	Chemicon, MAB377	WB; FACS	
Anti-H3Ac	Millipore; 06-599	WB; ChIP	
Anti-Histone H1	Abcam; ab1938	WB	
Infra-red anti-rabbit 680 IgGs	Licor; 32223	WB	
Infra-red anti-mouse 680 IgGs	Licor; 32212	WB	
Infra-red anti-goat 680 IgGs	Licor; 32214	WB	
Peroxidase anti- rabbit IgGs	GE Healthcare; NA934	WB	
Peroxidase anti-mouse IgGs	GE Healthcare; NA931	WB	
Peroxidase anti-goat IgGs	Sigma-Aldrich; A-3415	WB	
Alexa-Fluor 488 anti-rabbit	Invitrogen; A11008	FACS	
Biotinylated anti-mouse	Invitrogen; B2763	FACS	
Alexa-Fluor 647 streptavidin conjugate	Invitrogen; A11008	FACS	

WB: Western blotting

FACS: Fluorescence activated cell sorting

ChIP: Chromatin immunoprecipitation

List of Recombinant Proteins

Protein	Source; Cat no.	Notes
Histone H4	New England Biolabs; M2504S	
MeCP2	Gift from R. J. Klose	Full length untagged wildtype MeCP2 (NM_004992.3) was expressed and purified using a baculovirus expression system.

Fluorescence activated cell sorting

Nuclei were thawed, pelleted (600 g, 5 min, 4 °C) and resuspended in 1 ml of 1 x PBS, supplemented with 10 mM sodium butyrate and complete protease inhibitors (Roche). NeuN antibody was added at 1:500 dilution, MeCP2 antibody was added at 1:200 dilution. Nuclei were incubated at 4 °C for 2 h rotating on a wheel. Nuclei were washed with 3 x 1 ml supplemented PBS for 5 min and then incubated with secondary antibodies at 1:1000 dilution for 1 h at 4 °C rotating on a wheel. Nuclei were washed as before and then used for FACS. Recovered nuclei were pelleted and lysed in protein loading buffer.

Chromatin Immunoprecipitation Protocol

Throughout the protocol buffers were supplemented with complete protease inhibitors (Roche Applied Science) and 10 mM sodium butyrate. Frozen mouse tissues were thawed quickly in 2.5 ml 1 x PBS and homogenised using a 5 ml dounce (Braun). Homogenate was pelleted (1000 g, 5 min, 4 °C) and then crosslinked in 10 ml PBS containing 1% formaldehyde for 10 min at room temperature. Crosslinking was quenched by adding glycine to a final concentration of 125 mM. Homogenate was washed in ice-cold PBS, then resuspended in 1.4 ml lysis buffer (50 mM Tris-HCl pH 8.1, 1% SDS, 10 mM EDTA) and incubated on ice for 10 min. For each immunoprecipitation, 140 µl (~10 x 10^6 cells) was taken, and 1260 µl dilution buffer added (20 mM Tris-HCl pH 8.1, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 3 mM $CaCl_2$, 10 mM $MgCl_2$). Chromatin was incubated with 3 U *micrococcal* nuclease (Fermentas) at room temperature for 17 min and stopped by adding EDTA and EGTA to a final concentration of 10 mM and 20 mM respectively. Chromatin was further fragmented by sonication for 3 min (Branson digital sonifier). Overall, chromatin was fragmented with an

average size of 500 bp. Precipitated debris was pelleted by centrifugation (16000 g, 10 min, 4 °C) and the chromatin pre-cleared with 50 µl washed protein A sepharose (GE Healthcare).

The supernatant was used for immunoprecipitations (1200 µl) and input (200 µl). Antibodies were added to the supernatant (7 µl MeCP2 674 antibody; 5 µl H3Ac Millipore) and rotated at 4 °C overnight. Precipitated debris was pelleted and the supernatant incubated with 25 µl protein A sepharose for 1 h at 4 C. The beads were then washed for 4 min at room temperature using 1 ml of ice-cold buffers as follows: once in buffer 1 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 8.1); four times in buffer 2 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl, 20 mM Tris-HCl pH 8.1); once in buffer 3 (250 mM LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.1); and three times in TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA). Immunocomplexes were then eluted with 200 µl of extraction buffer (1% SDS, 100 mM $NaHCO_3$), and cross-links were reversed by adding 5 M NaCl to a final concentration of 300 mM, followed by incubation at 65 °C overnight. DNA was extracted by proteinase K treatment, phenol extraction and ethanol precipitation. Isolated DNA from both immunoprecipitations and inputs were either resuspended in 300 µl 0.1 x TE for real time PCR analysis (5 µl / PCR) or 25 µl for bisulfite treatment.

Real time PCR was carried out using Quantace Sensimix Plus with a Roche Lightcycler according to manufacturer's instructions (primer sequences available on request). Cycle threshold (Ct) values were determined by the second differential maximum method as calculated by the Roche Lightcycler software. % IP / Input was calculated using the comparative Ct method and factoring in the 6-fold difference in volume of sheared chromatin supernatant used for input and immunoprecipitation and represented as a percentage. Real time amplicons had an average spacing of 1 kb.

Preparation of ChIP DNA for Solexa Sequencing

Six independent MeCP2 immunoprecipitations were performed from a single mouse brain then subsequently pooled and end repaired by incubation at 20 °C for 30 min with 3 U of T4

DNA Polymerase (NEB), 10 U of Polynucleotide Kinase (NEB), 2 U DNA Polymerase I Large (Klenow) fragment (NEB), 1x T4 DNA ligase reaction buffer (NEB) and 400 nM dNTPs. The enzymes were then heat inactivated at 75 C for 20 min, after which the DNA was ethanol precipitated. A tail of 'A' bases was added to the 3' ends of the DNA by incubation with 5 U Klenow Fragment (3'-5' exo-; NEB), 200nM dATP and 1x buffer 2 (NEB) at 37 °C for 30 min. The enzymes were heat inactivated and cleaned up as before. Ilumina paired end adaptors were then ligated to the processed ChIP DNA by incubation with 300 U of T4 DNA ligase (NEB), 1x T4 DNA ligase buffer (NEB), 7.5% PEG-6000 and 2 pmol of annealed Ilumina adaptors for 3 h at room temperature. Ligated DNA was purified using MinElute PCR columns (Qiagen) and eluted in 10 µl water.

Methyl-specific DNA chromatography

75 µg of mouse cerebellum genomic DNA (C57-Bl6) was sonicated to an average length of 400 bp (Branson digital sonifier). Fragmented DNA was subsequently phosphorylated and end repaired by incubation at 20 °C for 30 min with 30 U of T4 DNA Polymerase (NEB), 100 U of Polynucleotide Kinase (NEB), 15 U DNA Polymerase I Large (Klenow) fragment (NEB), 400 nM dNTPs and 1x T4 DNA ligase reaction buffer (NEB). The enzymes were then heat inactivated at 75 °C for 20 min after which the repaired DNA was ethanol precipitated. A tail of 'A' bases was added to the 3' ends of the DNA by incubating at 37 °C for 30 min with 30 U Klenow Fragment (3' – 5' exo-; NEB), 200 nM dATP and 1x buffer 2 (NEB). The enzymes were heat inactivated and cleaned up as before. Ilumina Paired end adaptors were ligated to the processed DNA by incubation with 3000 U of T4 DNA ligase (NEB), 1x T4 DNA ligase buffer (NEB), 7.5% PEG-6000 and 2.8 nmol of annealed Ilumina adaptors for three hours at room temperature. The ligated DNA was separated into two 35 µg aliquots and purified independently to enrich for sequences containing methylated CpGs as previously described (Illingworth et al., 2008). DNA tightly retained by the methyl-affinity matrix (NaCl >700 mM) was to passed over the affinity column for a second time increase binding specificity. The volume of purified DNA was reduced to 200 µl using an Amicon Ultra 30 kDa centrifugal filter (30,000 NMWL columns; Millipore) and subsequently ethanol precipitated.

Library Preparation and Illumina Solexa Sequencing

Ligated DNA was amplified by 10-12 cycles of PCR with primers complementary to the adaptor sequences and Phusion 2x premix (Finnzymes). The DNA was purified using QIAquick PCR Purification columns (Qiagen) and library fragments of >200 bp (insert plus adaptor and PCR primer sequences) were isolated by gel extraction. The purified DNA was captured on an Illumina flow cell for cluster generation. Libraries were sequenced on the Genome Analyzer following the standard Illumina protocol to generate 37 bp reads using a Solexa sequencer. Single-end sequence reads were mapped to the mouse genome (NCBIm37) using MAQ (http://maq.sourceforge.net/). Reads with a mapping score greater or equal to 30 where retained. For MeCP2 ChIP-seq seven independent lanes of sequence was generated and the results from these were combined.

Analysis of High Throughput Sequencing

CpG density calculations were performed on the ensEMBL repeat masked version of the mouse NCBI build 37 genome. Sliding window analysis was employed to count observed CpGs, ignoring all windows containing greater than 50% masked repeats. The same sliding window parameters have been used to produce a measure of MeCP2 hits per window across the mouse genome, according to mapped Solexa sequenced reads. Genome wide levels of methylation have been traced by further applying the sliding window to mouse cerebellum MBD-seq data. An inner join, on relative sets of data, pairs windows with identical co-ordinates to form a relationship between observed CpG, MeCP2 reads and MBD reads across the mappable genome. For visualisation genomic intervals and values have been translated in to wiggle files and displayed via the Integrated Genome Browser.

Comparing MeCP2 Distribution Across Methylated CGIs

The methylated CpG rich fraction of the mouse genome was defined using MBD-seq data for cerebellum as outlined above. Methylated CGIs were identified by intersecting a repeat masked NCBI CGI-strict dataset with regions of high MBD-seq enrichment (>4 reads per base that span at least 90 bp and with a maximum gap of 250 bp). Sliding window analysis was

employed to count MeCP2 reads across 5 kb domains centred upon the midpoint of each methylated CGI (500 bp window with a 100 bp slide).