

1 Domains of methylated CAC and CG target MeCP2 to tune 2 transcription in the brain

3

4 Sabine Lager^{1,9,10}, John C Connelly^{1,10}, Gabriele Schweikert^{1,6,10}, Shaun
5 Webb¹, Jim Selfridge¹, Bernard H Ramsahoye², Miao Yu^{3,4}, Dina DeSousa¹,
6 Christian Seiser⁵, Chuan He^{3,4}, Guido Sanguinetti⁶, Lawrence C Sowers⁷,
7 Malcolm D Walkinshaw⁸ and Adrian Bird¹

8

9 *Affiliations:*

10 ¹ Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh
11 EH9 3BF, UK

12 ² Institute of Genetics and Molecular Medicine, University of Edinburgh,
13 Edinburgh EH4 2XU, UK

14 ³ Department of Chemistry and Institute for Biophysical Dynamics, University
15 of Chicago, Chicago, IL 60637, USA

16 ⁴ Howard Hughes Medical Institute, University of Chicago, Chicago, IL 60637,
17 USA

18 ⁵ Max F. Perutz Laboratories, Medical University of Vienna, 1030 Vienna,
19 Austria

20 ⁶ School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, UK

21 ⁷ Department of Pharmacology and Toxicology, University of Texas Medical
22 Branch, Galveston, TX 77555, USA

23 ⁸ Centre for Translational and Chemical Biology, University of Edinburgh,
24 Edinburgh EH9 3BF, UK

25 ⁹ present address: Unit of Pathology of Laboratory Animals, University of
26 Veterinary Medicine Vienna, 1210 Vienna, Austria

27 ¹⁰ Co-first authors

28

29 Contact Information: a.bird@ed.ac.uk

30

31

32

33 **Summary**

34

35 Mutations in the gene encoding the methyl-CG binding protein MeCP2 cause
 36 neurological disorders including Rett syndrome. The di-nucleotide methyl-CG
 37 (mCG) is the canonical MeCP2 DNA recognition sequence, but additional
 38 targets including non-methylated sequences have been reported. Here we
 39 use brain-specific depletion of DNA methyltransferase to show that DNA
 40 methylation is the primary determinant of MeCP2 binding in mouse brain. *In*
 41 *vitro* and *in vivo* analyses reveal that MeCP2 binding to non-CG methylated
 42 sites in brain is largely confined to the tri-nucleotide sequence mCAC.
 43 Structural modeling suggests that mCG and mCAC may be interchangeable
 44 as minimal structural perturbation of MeCP2 accompanies binding. MeCP2
 45 binding to chromosomal DNA in mouse brain is proportional to mCG + mCAC
 46 density and defines domains within which transcription is sensitive to MeCP2
 47 occupancy. The results suggest that MeCP2 interprets patterns of mCAC and
 48 mCG in the brain to negatively modulate transcription of genes critical for
 49 neuronal function.

50

51 Introduction

52

53 Methylation at the C5 position of cytosine is an epigenetic mark implicated in
 54 gene regulation and disease¹. In mammals, DNA methylation occurs in a CG
 55 di-nucleotide context, but in neuronal cells and embryonic stem cells (ESCs),
 56 mCA is detected at significant levels^{2,3}. Like mCG, mCA is negatively
 57 correlated with transcript abundance, hinting at a repressive function in the
 58 brain^{2,3}. Highest levels of mCA are observed in the human and mouse brain,
 59 where mCA accumulates postnatally during a phase of increased
 60 synaptogenesis³. In mice, the postnatal increase in neuronal mCA coincides
 61 with accumulation of Dnmt3a protein³. All cytosine DNA methyltransferases
 62 exhibit a preference for CG as a substrate, but Dnmt3a is able to methylate
 63 CA, albeit at a lower rate⁴⁻⁷. Brain mCA occurs most frequently in the tri-
 64 nucleotide mCAC^{2,8}, whereas in ESCs mCAG is the preferred sequence
 65 context. The role of mCAC^{2,3} in the developing mammalian brain is yet to be
 66 elucidated.

67

68 A potential mechanism for interpreting the DNA methylation signal is the
 69 recruitment of methyl-CG binding domain (MBD) proteins including MeCP2,
 70 MBD1, MBD2 and MBD4⁹. Of these MeCP2 has attracted considerable
 71 attention as mutations involving the *MECP2* gene cause the X-linked autism
 72 spectrum disorder Rett syndrome¹⁰ and *MECP2* duplication syndrome¹¹. Rett
 73 missense mutations cluster in two domains of MeCP2: the MBD and the
 74 NCoR/SMRT co-repressor interaction domain (NID)^{12,13}. These observations
 75 implicate the loss of binding to methylated DNA and/or the failure to recruit the
 76 NCoR/SMRT repressor complex to DNA as primary causes of Rett syndrome.
 77 MeCP2 has a high affinity *in vivo* and *in vitro* for binding to mCG¹⁴⁻¹⁶, but the
 78 determinants of its targeting to DNA have recently diversified to include mCA,
 79 whose postnatal accumulation is paralleled by an increase in MeCP2
 80 protein^{2,3}. In addition, it has been reported that MeCP2 binds to
 81 hydroxymethylcytosine (hmC), the major oxidized form of mC, which is
 82 abundant in neurons¹⁷. Finally, there is data suggesting that MeCP2 can bind
 83 chromatin in a DNA methylation-independent manner^{15,18,19}.

84 Although the mutational spectrum and biochemical interactions of MeCP2
85 suggest that it behaves as a transcription repressor^{13,20}, changes in the
86 mouse brain transcriptome when the protein is absent involve both up- and
87 down-regulation of genes^{17,21,22}. Accordingly MeCP2 has been proposed to
88 act additionally as an activator of transcription or as a multifunctional hub that
89 effects diverse aspects of cellular metabolism^{12,23,24}. An alternative model
90 proposes that MeCP2 primarily functions by globally modifying the
91 architecture of chromatin via multifaceted interactions with DNA^{16,19}. An
92 inverse correlation between levels of CA methylation and expression of long
93 genes has re-emphasized transcriptional inhibition via MeCP2²⁵. A separate
94 study reported, however, that mCA is enriched within genes that are mis-
95 regulated regardless of the direction of the transcriptional change in response
96 to MeCP2 depletion or excess²⁶. Despite progress, therefore, a consensus
97 view regarding the role of MeCP2 in transcriptional regulation has been
98 elusive.

99

100 Here we define the DNA binding specificity of MeCP2 using *in vitro* and *in vivo*
101 approaches. Supporting the primacy of DNA methylation as a determinant of
102 MeCP2 binding, we find that global reduction of this epigenetic mark in the
103 mouse brain decreases MeCP2 binding. We show for the first time that
104 MeCP2 binding to non-CG methylated sites is primarily restricted to the tri-
105 nucleotides mCAC or hmCAC *in vitro* and *in vivo*. Modeling based on the X-
106 ray structure of the MBD of MeCP2 suggests that mCG, mCAC and hmCAC
107 interact with a common protein conformation and may therefore lead to
108 indistinguishable down-stream biological effects. MeCP2 binding across the
109 adult brain genome reveals long genomic domains of high and low occupancy
110 that match the distribution of mCG + mCAC binding sites. The results have
111 important implications for regulation of gene expression, as we uncover a
112 strong correlation between MeCP2 binding, mCG + mCAC density and the
113 direction of gene mis-regulation when MeCP2 is absent or over-expressed.
114 Our findings shed new light on the binding properties of MeCP2 and implicate
115 MeCP2 as a global modulator of neuronal transcription.

116

117 Results

118

119 ***Global reduction of DNA methylation in brain causes loss of MeCP2*** 120 ***binding***

121 DNA methylation as a determinant of MeCP2 binding has been demonstrated
122 experimentally^{15,16,27}, but its primary role has been questioned²⁸ and the
123 possibility considered that MeCP2 binds to DNA via multiple modalities¹⁸.
124 Depletion of the CG methylation signal specifically in brain would stringently
125 test its importance, but mice in which *Dnmt1* is deleted exclusively in neuronal
126 and glial cells using Nestin–Cre die at birth²⁹. Reinvestigating this conditional
127 mutation strategy, we found that on an outbred background mice survive
128 postnatally for up to two weeks. At this age, neuronal MeCP2 expression
129 levels are low compared to adult brains³⁰ and MeCP2 gains in functional
130 importance only later in development as *Mecp2* knockout (KO) mice are
131 symptom-free at this stage^{31,32}. Nevertheless, *Dnmt1* KO mice represent a
132 unique opportunity to test for causal effects of DNA methylation on MeCP2
133 binding *in vivo*.

134

135 Western blots confirmed that *Dnmt1* was severely reduced from birth to post-
136 natal day 13 in brains expressing Nestin–Cre and the floxed *Dnmt1* allele (Fig.
137 1a). Moreover transcription of intracisternal A particle (*IAP*) elements, which
138 are members of an LTR retrotransposon family in mouse, was greatly up-
139 regulated as reported in *Dnmt1*-deficient mouse embryos³³ (Fig. 1b), but
140 expression of other transposon families was not detectably affected
141 (Supplementary Fig. 1a-b). *Mecp2* -/- brains showed no significant elevation
142 of *IAP* element transcription (Fig. 1b). A previous study¹⁶ found that *IAP*
143 transcription was modestly increased (~ 2-fold) in nuclear RNA from *Mecp2*-
144 null brain, but this effect was not detectable in stable RNA from whole tissue,
145 in agreement with present findings. We analyzed levels of global DNA
146 methylation in wildtype (WT) and *Dnmt1* KO brains by high performance liquid
147 chromatography (HPLC) (Fig. 1c and Supplementary Fig. 1c) and reduced
148 representation bisulfite sequencing³⁴ (RRBS) (Fig. 1d and Supplementary Fig.
149 1d). The unbiased HPLC quantification detected a 38% reduction in CG
150 methylation in *Dnmt1* KO brains whereas the RRBS tested regions, which

151 comprise predominantly promoter regions, showed a 56% reduction.
152 Importantly, MeCP2 was expressed at near normal protein levels in *Dnmt1*
153 KO brains in comparison with age-matched WT brains (Fig. 1e-f and
154 Supplementary Fig. 1e).

155

156 To investigate how local loss of DNA methylation quantitatively affects MeCP2
157 binding by chromatin IP (ChIP), we used primers specific for repetitive *IAP*
158 elements and mouse major satellite sequences (Fig. 1g-h and Supplementary
159 Fig. 1f) plus four inter- and intra-genic single-copy regions from mouse
160 chromosome 3 (Supplementary Fig. 1g-k). Reduced DNA methylation in the
161 conditional mutant brain was confirmed at *IAP* elements and mouse satellite
162 by bisulfite analysis (Fig. 1i-j) and MeCP2 binding was correspondingly lower
163 (Fig. 1g-h). Extending the analysis to unique regions of the genome, we chose
164 four random sites in a relatively gene dense domain (Supplementary Fig. 1g).
165 In all replicates MeCP2 binding to each site was consistently reduced in each
166 *Dnmt1*-deficient brain compared with a matched WT littermate
167 (Supplementary Fig. 1h-k). Furthermore, bisulfite analysis confirmed the
168 reduction of DNA methylation at the nucleotide level in each case
169 (Supplementary Fig. 1l-o). The data for 6 randomly chosen regions of the
170 mouse genome provide strong support for the conclusion that DNA
171 methylation is a major determinant of MeCP2 binding in the mouse brain.

172

173 ***MeCP2 recognizes modified di- and tri-nucleotide sequences***

174 The predominant methylated sequence is the di-nucleotide CG, but in adult
175 brain mCA³ and hmCG^{17,35} are implicated as binding partners of MeCP2. A
176 recent study concluded that in addition to mCG, MeCP2 binds both mCA and
177 hmCA²⁵ and confirmed earlier reports that hmCG is a low-affinity binding site
178 for MeCP2^{25,36–38}. In addition MeCP2 has been reported to bind *in vitro* to
179 DNA in which every cytosine was substituted with hmC¹⁷. Given that neurons
180 only accumulate significant levels of both hmCG and mCH (where H is any
181 base except G) from two weeks postnatally³, the *Dnmt1* KO model is unsuited
182 to investigate additional sequence specificities of methylation dependent
183 MeCP2 binding. To comprehensively analyze the DNA sequence
184 determinants of MeCP2 binding *in vitro* we performed EMSA using the 1-205

185 N-terminal domain of MeCP2 that contains the MBD. Surprisingly, the data
186 revealed a further constraint on MeCP2 binding, as the third base following
187 mCA strongly affected MeCP2 binding affinity *in vitro* (Fig. 2a). Probes
188 containing the mCAC tri-nucleotide sequence bound with high affinity to
189 MeCP2, whereas probes containing mCAA, mCAG and mCAT bound much
190 less strongly. This result was confirmed in EMSA experiments using all
191 possible mCXX tri-nucleotide sequences as unlabeled competitors against a
192 labeled mCGG-containing probe (Fig. 2b). Quantification showed that mCAC
193 and, to a lesser extent mCAT, are both effective competitors, but mCAG and
194 mCAA compete no better than non-methylated control DNA (Fig. 2c). All
195 mCGX oligonucleotide duplexes competed strongly indicating that the base
196 following mCG on the 3' side does not have a large effect on binding,
197 although we note that mCGA was reproducibly a weaker competitor than
198 mCGC, mCGG or mCGT. Neither mCCX nor mCTX tri-nucleotides have a
199 significant affinity for MeCP2 *in vitro*.

200

201 As hmCA is reported to bind MeCP2 *in vitro*²⁵, we asked whether the third
202 base is also important for hmC binding. Using hmCXX tri-nucleotides as
203 probes in EMSAs, we found that hmCAC bound with a much higher affinity
204 than hmCAA, hmCAG and hmCAT DNA (Fig. 2d). It is notable that the great
205 majority of hmC in the brain and elsewhere is in the hmCG di-nucleotide, with
206 hmCAC being extremely rare³. The DNA binding specificity of MeCP2 MBD
207 deduced from these *in vitro* experiments is summarized in a matrix of di- and
208 tri-nucleotide sequences that bind to MeCP2 (red lettering, Fig. 2e).

209

210 ***MeCP2 binding specificity in vivo***

211 To determine whether the binding specificities established *in vitro* apply to full-
212 length protein in living cells, we developed a novel assay using transfection
213 followed by ChIP³⁹. Synthetic DNA duplexes containing specific cytosine
214 modifications were transfected into HEK293 cells expressing human MeCP2
215 tagged with GFP (Fig. 3a and Supplementary Fig. 2a-b). Levels of
216 endogenous MeCP2 in these cells are negligible and therefore do not
217 interfere with the assay (Supplementary Fig. 2c). We tested oligonucleotide
218 duplexes containing a single modified cytosine in either a mCG, hmCG,

mCAX or hmCAX context (Fig. 3b). The results showed that mCAC, mCAT, hmCAC and mCG all bound MeCP2 efficiently, whereas hmCG, mCAA and mCAG binding was indistinguishable from background binding to non-methylated DNA (Fig. 3c). The same outcome was seen with different DNA sequences containing three CAC motifs per oligonucleotide, either all unmodified, all methylated or all hydroxymethylated (Fig. 3d-e). These *in vivo* results with full-length protein match those obtained *in vitro* by EMSA, strengthening the evidence that MeCP2 requires specific tri-nucleotide settings to recognize mC or hmC in a non-CG context.

Modeling the MeCP2-DNA interaction

To seek a structural basis for the tri-nucleotide specificity of MeCP2 binding, we asked whether the X-ray structure of the MeCP2 MBD⁴⁰ could suggest why mCAC or mCAT binding is permitted while mCAA, mCAG, mCCX and mCTX are forbidden. Surprisingly, altering the conformation of only one amino acid side chain, R133, while leaving all other coordinates of the protein structure unchanged, could account for the observed interactions. R133 makes essential hydrogen bonds with one guanine base in the mCG complex and also provides salt bridges with a cytosine methyl group (Fig. 4a)⁴⁰. The equivalent guanine residue on the other DNA strand of the mCG dyad is also hydrogen bonded to an arginine: residue R111. Mutations in either R133 or R111 cause Rett syndrome, but despite their related roles, the conformations of R111 and R133 are very different. Whereas the R111 side-chain adopts an extended all-trans conformation that is “pinned” by hydrogen bonds with D121, R133 is relatively unconstrained by surrounding amino acids (Fig. 4a). We therefore asked if R133 could potentially interact with mCAC in the existing X-ray structure⁴⁰. Indeed, by extending the side-chain and tilting the guanidinium group, R133 can make permitted stabilizing hydrogen bonds with the guanine base that is paired with the third cytosine in the mCAC tri-nucleotide (Fig. 4b). An equivalent interaction is also possible with the complementary adenine in mCAT tri-nucleotide (Supplementary Fig. 3a). Importantly, interactions with pyrimidine bases are sterically unfavorable, arguing that mCAA or mCAG are unlikely to interact with the MBD, as is observed experimentally (Supplementary Fig. 3b-c). Modeling of hmCG binding indicates that the

253 presence of the cytosine hydroxyl group would not be accommodated due to
 254 the close proximity of this polar group with the guanidinium group of R111
 255 (Supplementary Fig. 3d). Binding of hmCAC is allowed, however, due to tilting
 256 of the R133 side-chain and the formation of hydrogen bonds with guanine on
 257 the opposite DNA strand (Supplementary Fig. 3e). The presence of a purine
 258 in the third position of hmCAG and hmCAA introduces a clash that is not
 259 observed with hmCAC or hmCAT (data not shown). Thus the observed tri-
 260 nucleotide binding specificity of MeCP2 can theoretically be explained with
 261 minimal perturbation of the established X-ray structure of the MBD-DNA
 262 complex. The minimal change to the structure of the MBD required to
 263 accommodate either mCG or mCAC raises the possibility that the biological
 264 consequences of binding either motif will be the same.

265
 266 A prediction of the modeling is that MeCP2 should bind in only one orientation
 267 to mCAC or mCAT, whereas binding to the symmetrical mCG dyad may occur
 268 in either orientation. The model requires that R111 and D121 interact with the
 269 methyl group of thymine rather than that of 5mC (Fig. 4b). As thymine is
 270 effectively 5-methyluracil (Fig. 4c), we replaced it with uracil in the labeled
 271 probe and performed EMSA analysis (Fig. 4d). Loss of the thymine methyl
 272 group abolished binding to MeCP2, in agreement with the hypothesis that
 273 MeCP2 binding to mCAC is confined to one orientation. The data suggest that
 274 a symmetrical pair of 5-methyl pyrimidines, one of which is mC, offset by one
 275 base pair is an essential pre-requisite for MBD binding to DNA.

276

277 ***DNA sequence specificity of MeCP2 binding in adult mouse brain***

278 To test whether the DNA binding specificities established *in vitro* and in
 279 transfected cells also apply in native tissues, we analyzed MeCP2 ChIP-seq
 280 and whole genome bisulfite (WGBS) datasets derived from adult mouse brain
 281 (references ^{3,25,26} and WGBS from sorted neurons and WGBS from
 282 hypothalamus this study). CG is under-represented in the mouse genome
 283 (~4% of CX), but highly methylated (~80%), whereas CA is the most abundant
 284 CX di-nucleotide (36% of CX), but even in brain only a small fraction of CA is
 285 methylated (<2%) (Fig. 5a-c). Bisulfite analysis of sorted neurons by NeuN
 286 staining¹⁶ confirmed previous reports that CAC is the most methylated tri-

287 nucleotide^{2,8}, being ~12% methylated (Fig. 5c). The finding that the MeCP2
288 tri-nucleotide binding specificity matches the most abundantly methylated
289 non-CG sequence in brain encourages the view that this interaction is
290 biologically relevant.

291

292 Previous ChIP-seq analyses have concluded that the read coverage of
293 MeCP2 chipped samples tracks the density of CG methylation^{15,16,25,26}. Re-
294 analysis of several MeCP2-ChIP data sets for which the antibody used has
295 been rigorously verified, indicates, however, that the profile of Input –
296 specifically, DNA derived from the fragmented chromatin sample used for
297 ChIP – is closely similar to that of MeCP2 (Fig. 5d). Using the published high
298 quality ChIP-seq dataset for hypothalamus²⁶, we fitted a linear model to
299 predict MeCP2 read coverage from Input reads alone and found a coefficient
300 of determination of 0.84, indicating that MeCP2 is almost uniformly distributed
301 across the genome at this resolution (Fig. 5e). These results are in line with a
302 previous report that the number of MeCP2 molecules in mature neurons is
303 sufficient to almost “saturate” mCG sites in the genome¹⁶. Given the similarity
304 between ChIP and Input, we focussed on regions that show deviations from
305 the Input profile regarding enrichment (purple) or depletion (green) of MeCP2
306 (Fig. 5f). First, we investigated genomic regions that are depleted of potential
307 binding sites, e.g. unmethylated CpG islands (CGIs). As the ChIP dataset was
308 derived from mouse hypothalamus, we performed WGBS on three biological
309 replicates of this brain region (see Online Methods). Using ChIP and DNA
310 methylation datasets from the same brain region, we observed a pronounced
311 drop of the $\log_2(\text{MeCP2}/\text{Input})$ signal across CGIs for two different data sets,
312 confirming previous analyses^{16,25,26} (Supplementary Fig. 4a). We next
313 examined regions where the MeCP2 signal was higher than expected by
314 applying the MACS⁴¹ tool to detect summits of MeCP2 binding peaks relative
315 to Input²⁵. As expected, the di-nucleotide mCG showed a sharp peak at
316 MeCP2 ChIP summits in the hypothalamus dataset (Supplementary Fig. 4b-e
317 and ²⁵). In addition, the tri-nucleotide mCAC, but not other mCAX tri-
318 nucleotides, coincided strikingly with MeCP2 peak summits, confirming that
319 mCAC provides a focus for MeCP2 binding (Fig. 5g-h and Supplementary Fig.
320 4f-h). Using random regions as a negative control, we did not detect any

sequence or methylation dependency (Supplementary Fig. 4i). Regarding targeting of mCAT, which bound relatively weakly in EMSA, but strongly in the transfection assay, the ChIP-seq data suggest that this is a relatively low affinity site in native brain (Fig. 5g-h and Supplementary Fig. 4f-h).

To establish MeCP2 binding preferences across the whole genome we adopted a sliding window approach. In the Input sample, both, mCG density and density of unmethylated CGs strongly correlated with coverage revealing a methylation independent CG sequencing bias (data not shown). In contrast, the MeCP2 ChIP coverage was DNA methylation sensitive, with mCG being positively correlated, while unmethylated CG density was anti-correlated (data not shown). Importantly, DNA methylation-sensitivity was also observed in the Input-corrected signal ($\log_2(\text{MeCP2}/\text{Input})$), strongly supporting the view that mCG is targeted by MeCP2 binding (Fig. 5i, Supplementary Fig. 4j, l). This mCG binding preference was independent of the third DNA base (Supplementary Fig. 4n). In agreement with the *in vitro* and *in vivo* results reported above, the density of mCAC correlated strongly with increasing MeCP2 enrichment, whereas a much weaker trend was observed for other methylated tri-nucleotide sequences (Fig. 5j, Supplementary Fig. 4k, m). We found no evidence for MeCP2 binding to hmCG or unmethylated C's in any sequence context (Supplementary Fig. 4j, l, o-r). To complement the sliding window approach, we focused on MeCP2 binding within gene bodies and found once again that the Input-corrected signal is strongly correlated with the density of mCG and mCAC (Fig. 5k). The ChIP data therefore sustain the view that MeCP2 binding is determined by the combined density of mCG and mCAC sites.

Given the high abundance and global distribution of MeCP2 in the neuronal genome, we looked for domains of MeCP2 occupancy that might reflect long-range variation in binding site abundance. To avoid pooling data in arbitrary windows, we used a multiscale representation method (MSR) that identifies patterns of signal enrichment or depletion across scales spanning several orders of magnitude⁴². MSR identified a large number of long domains moderately enriched for MeCP2 extending up to 1 Mb indicating that regions

of high MeCP2 occupancy extend beyond the scale of a single gene (Fig. 5l). These regions share common sequence features, in particular high mCG and mCAC densities (Fig. 5m, Supplementary Fig. 4t-u). We also identified a large number of short regions containing highest GC content (<1 kb) but are strongly depleted in MeCP2 binding. As expected, these regions significantly overlap with CGIs (Supplementary Fig. 4s). Lastly, we found a third group of relatively long regions (10 kb - 1 Mb) which are moderately depleted in MeCP2 binding (MeCP2 enrichment score: <0). These regions are moderately enriched for mCG but lack mCAC (Fig. 5m, Supplementary Fig. 4s-u).

As part of this binding site analysis we re-visited an earlier report that MeCP2 binds preferentially to mCG flanked by an AT-rich run of 4-6 base pairs *in vitro*⁴³. To look for this preference in brain, we asked whether isolated mCG and mCAC flanked by a run of 4 or more A or T base pairs within 13 base pairs showed greater MeCP2 ChIP-seq signal than sequences lacking an AT run. In summary, we find no evidence for an effect of AT-flanks on MeCP2 binding site occupancy, suggesting that the *in vitro* preference might not be biologically relevant *in vivo* (Supplementary Fig. 4v-w).

The relationship between MeCP2 occupancy and gene expression

The association of MeCP2 with both methylated sites in the genome and the co-repressor complex NCoR¹³ suggests that the protein can function to inhibit transcription. If so, a relationship would be expected between MeCP2 occupancy and the transcriptional mis-regulation seen when MeCP2 is either depleted by deletion of the gene (KO)³¹ or over-expressed (OE)^{44,45}. Before making use of published datasets for mouse hypothalamus²⁶, we first asked, using HPLC, whether the absence or over-expression of MeCP2 alters total RNA levels. Previous studies using MeCP2-deficient neurons differentiated *in vitro* from mouse ES cells or human iPS cells reported reduced total RNA and transcriptional capacity^{24,46}, but comparable measurements in brains of MeCP2-deficient mice have not been reported. Using a sensitive RNA quantification technique, we observed that total RNA per cell in KO hypothalamus is reproducibly 15% lower than WT (Supplementary Fig. 5a).

389 Over-expression of MeCP2, however, did not significantly affect total RNA. As
 390 the latter is ~98% ribosomal RNA, we asked whether mRNA levels were also
 391 reduced in KO hypothalamus. Quantitative RT-PCR (qPCR) using spiked-in
 392 *Drosophila* cells to control for experimental error and normalized to brain cell
 393 number in each sample (Supplementary Fig. 5b) confirmed that genes
 394 previously reported to be up- or down-regulated in MeCP2-deficient
 395 hypothalamus²⁶ were similarly mis-regulated in our samples (Supplementary
 396 Fig. 5c-f). We then measured the abundance of three housekeeping genes
 397 and found that all were down-regulated by approximately 15%
 398 (Supplementary Fig. 5g-i). These results suggest that total RNA and mRNA
 399 levels are coordinately reduced and we have therefore applied this
 400 normalization to all hypothalamus RNAseq datasets. The mechanisms
 401 responsible for reduced total RNA, and whether this effect is a primary or
 402 secondary consequence of MeCP2 deficiency, are unknown.

403
 404 We next examined gene expression by separating genes whose expression
 405 was increased, unchanged or decreased in response to changing levels of
 406 MeCP2. Normalizing ChIP signals against Input, we found that up-regulated
 407 genes in KO hypothalamus are within domains enriched in MeCP2, whereas
 408 down-regulated genes were relatively depleted (Fig. 6a). Unchanged genes
 409 showed an intermediate level of MeCP2 occupancy. The reciprocal result was
 410 seen in OE hypothalamus, as down-regulated genes had high MeCP2
 411 occupancy, whereas up-regulated and unchanged genes bound relatively less
 412 MeCP2 (Fig. 6d). This relationship, which was not observed in a previous
 413 analysis of this gene expression and ChIP data²⁶, disappeared altogether if
 414 gene body binding of MeCP2 was normalized to binding in gene flanking
 415 regions, as enhanced or depleted binding extended up- and down-stream of
 416 the transcription start and end sites. We also asked whether the increased
 417 MeCP2 occupancy measured by ChIP-seq in hypothalamus correlated with
 418 an elevated level of the two target sequences mCAC and mCG. The
 419 distribution of mCAC strikingly matched the pattern of MeCP2 binding (Fig. 6b,
 420 e), but mCG, which occurs at much higher density, correlated less obviously
 421 (Fig. 6c, f).

422 The strong reciprocal relationship between MeCP2 occupancy and the
 423 direction of gene mis-regulation in KO and OE hypothalamus respectively, is
 424 compatible with the notion that MeCP2 binding is inhibitory to transcription.
 425 Excess MeCP2 preferentially inhibits genes with most binding sites whereas
 426 its depletion preferentially de-represses highly occupied genes. Elevated or
 427 depleted MeCP2 binding extended up- and down-stream of the TSS,
 428 suggesting that these genes are embedded within the extended MeCP2
 429 domains identified by our MSR analysis (Fig. 5l-m). This was confirmed by
 430 mapping MeCP2-enriched and depleted domains onto the genome in relation
 431 to mis-regulated genes (Fig. 6g). Approximately 80% of genes up-regulated in
 432 KO hypothalamus were within or overlapped domains of high MeCP2
 433 occupancy (dark and pale lilac, left pie), whereas only 19% of down-regulated
 434 genes were associated with MeCP2 enrichment (dark and pale lilac, right pie)
 435 (Fig. 6h and Supplementary Fig. 6a-b).

436
 437 In order to comprehensively visualize the relationship between transcription
 438 and MeCP2 occupancy, all reliably detectable transcript levels were plotted
 439 against MeCP2 ChIP signal normalized to Input (Fig. 6i-j and Supplementary
 440 Fig. 6e-f). The 15% global reduction in mRNA and ribosomes meant that in
 441 KO hypothalamus expression of most genes (8,363) goes down (Fig. 6i,
 442 green), but despite this overall trend a small number of genes (403)
 443 significantly increased their expression per cell compared with WT (Fig. 6i,
 444 purple). These genes, which shared high MeCP2 occupancy (Fig. 6i, lower
 445 panel), showed the inverse behavior in OE hypothalamus where they were
 446 down-regulated (Fig. 6j). This inverse relationship is readily apparent in a plot
 447 of gene expression changes in KO/OE hypothalamus (Supplementary Fig. 6g).
 448 De-regulated genes shared similar levels of mCG density, but only the up-
 449 regulated genes displayed increased mCAC (Supplementary Fig. 6h-i). We
 450 noted that the genes up-regulated in KO were significantly longer than
 451 average, in agreement with a report that long genes are preferentially up-
 452 regulated in MeCP2 KO mice²⁵ (Supplementary Fig. 6c) and most were
 453 implicated in brain function by gene ontology analysis (Supplementary Fig.
 454 6d). In addition to this group of genes, a gene set that has been implicated in
 455 a variety of neurological disorders⁴⁷ also showed behavior compatible with

456 repression by MeCP2, as the magnitude of mis-regulation in KO and OE
457 hypothalamus correlated reciprocally with MeCP2 occupancy (Fig. 6k-l).
458

459 Discussion

460

461 ***DNA methylation is the dominant determinant of MeCP2 binding in the*** 462 ***brain***

463 This study explores the determinants of MeCP2 binding to DNA *in vitro* and *in*
464 *vivo* and their relationship to gene expression. It has been shown previously
465 that DNA methylation is the primary determinant of binding in cultured cells by
466 genetic ablation of the DNA methyltransferase Dnmt1^{15,27}. In mature neurons,
467 however, where MeCP2 is extremely abundant¹⁶, MeCP2 binding also tracks
468 DNA methylation¹⁶, but may depend on other features of chromatin. To
469 rigorously test whether DNA methylation is a pre-requisite for MeCP2 binding
470 in brain, we depleted Dnmt1 in the young brain and tested the effect on
471 MeCP2 binding by ChIP. At this early postnatal stage, mCAC is low and mCG
472 is therefore the predominant modified sequence. The results show clearly that
473 a ~40% reduction in DNA methylation causes a commensurate loss of
474 chromatin-bound MeCP2 on randomly selected single copy and repetitive
475 genomic sequences and therefore sustains the view that this modification is
476 quantitatively the major determinant of binding.

477

478 While this paper was under review a study of MeCP2 binding in olfactory bulb
479 of mouse reported contrasting conclusions⁴⁸. The authors found that MeCP2
480 is enriched at non-methylated CGIs and that DNA methylation was apparently
481 a minor determinant of binding. This conclusion conflicts with several
482 previously published MeCP2 ChIP experiments from different laboratories,
483 which found a dramatic drop in MeCP2 binding at CGIs in cultured cells¹⁵,
484 whole mouse brain¹⁶, hypothalamus²⁶, cortex and cerebellum²⁵ coupled with
485 DNA methylation-dependent occupancy of the genome. A potential
486 explanation for the discrepant results in olfactory bulb is that the rules
487 governing MeCP2 binding in this region of the brain differ from those
488 operating in the remainder of the nervous system. Alternatively, there may be
489 a technical issue regarding antibody specificity that has led to inconsistent
490 findings. It is notable that the earlier reports showing DNA methylation
491 dependence were achieved using a diversity of validated antibodies.

492

493 **DNA methylation-dependent recognition sequences of MeCP2**

494 We comprehensively analyzed the modified DNA sequences that determine
 495 MeCP2 binding. To insure the reliability of our conclusions we used three
 496 experimental approaches: *in vitro* EMSA; *in vivo* ChIP in transfected cultured
 497 cells; and *in vivo* ChIP-seq of mouse brain. Three methylated DNA motifs
 498 consistently recruited MeCP2: mCG, mCAC and hmCAC. Interestingly, mCAC
 499 is the predominant methylated non-CG sequence in brain, comprising 15 –
 500 30% of all methylated cytosine in sorted mouse neurons, probably due to the
 501 action of the *de novo* DNA methyltransferase Dnmt3a^{2,4}. The
 502 hydroxymethylated sequence hmCAC, on the other hand, is reportedly
 503 extremely rare perhaps due to the preference of Tet enzymes for mCG as a
 504 substrate³. Given the inability of MeCP2 to bind hmCG and the apparent
 505 extreme rarity of hmCAC, it seems unlikely that hmC is a major contributor to
 506 the biological role of MeCP2.

507
 508 Our modeling provides a structural explanation for the sequence specificity of
 509 the interaction between MeCP2 and DNA. We previously observed that the
 510 replacement of a mC at a methylated CG di-nucleotide with T, forming a T:G
 511 mispair, had a negligible effect upon the binding affinity of MeCP2³⁶,
 512 indicating that hydrogen bonding with the mC/T is not essential and that the
 513 interaction is flexible enough to accommodate the T:G wobble geometry. Here
 514 we observed that in duplex DNA pyrimidine-methyl groups can be provided by
 515 either the mC or T, as demonstrated by the finding that the replacement of T
 516 with U, which lacks the T methyl group, results in loss of MeCP2 binding.
 517 Remarkably the observed specificity for mC in either mCG or mCAC
 518 sequence contexts can be rationalized with minimal changes to the
 519 conformation of the MBD that was established by X-ray crystallography⁴⁰.
 520 Only the configuration of the side-chain of amino acid arginine 133 needs to
 521 be adjusted to account for both permitted and non-permitted interactions. The
 522 model has potentially important biological consequences, as it suggests that
 523 binding to mCAC or mCG is structurally very similar, making it likely that
 524 MeCP2 binding to any of these sequences will lead to the same outcome
 525 down-stream. In agreement with this scenario, experiments comparing
 526 reporters methylated at mCA or mCG suggest that both modified sequences

lead to transcriptional repression². Missense mutations that cause Rett syndrome predominantly affect the MBD or the NID, suggesting that the primary role of MeCP2 is to bridge DNA sequences with the NCoR/SMRT co-repressor complexes¹³. We propose that mCG and mCAC, function identically to facilitate this bridging process.

MeCP2 binding profiles and gene expression

The high frequency of mCG and mCAC binding sites for MeCP2 throughout the neuronal genome (one per ~100 bp) poses problems for conventional ChIP-seq analysis. In most studies binding sites of, for example, transcription factors are widely spaced compared with the size of chromatin-derived DNA fragments generated by ChIP protocols (200-500 bp). Therefore, only a minority of DNA fragments is expected to be immunoprecipitated and discrete peaks are recovered. In the case of MeCP2, most DNA fragments of ~200-500 bp contain binding sites (CGIs being a conspicuous exception) leading to a relatively uniform recovery of genomic DNA that resembles Input. The difference between Input and ChIP signal, which is the measure of MeCP2 density, provides an undulating continuum in which peaks are broad.

A striking feature of the *Mecp2* KO hypothalamus is the reduced level of total RNA, matching reports in cultured mouse and human neurons^{24,46}. The mechanism responsible is unknown, but one possibility is that MeCP2 is a direct global activator of transcription^{23,24}. Arguing against this possibility, we found that most KO down-regulated genes lie in domains of low MeCP2 occupancy. Also, it might be expected that two-fold over-expression of an activator would lead to increased levels of RNA compared to WT, but this is not observed (Supplementary Fig. 5a). An alternative explanation is that reduced RNA reflects reduced cell size, perhaps as a secondary consequence of sub-optimal neuronal gene expression. In this case the change in total RNA and the relative mis-regulation of genes within the RNA population may be separate phenomena. Given the close relationship between gene mis-regulation, MeCP2 binding site density and MeCP2 occupancy, we favor the view that the effects of MeCP2 concentration on the balance of neuronal gene expression are primary, whereas the downward

561 shift in total RNA is a secondary effect. To test this rigorously it will be
562 important to track down the origins of the total RNA deficiency.

563

564 By re-analyzing ChIP-seq and RNA-seq datasets from hypothalamus of WT
565 mice, *Mecp2*-null mice and mice over-expressing MeCP2 we were unable to
566 confirm reports that MeCP2 is more highly bound to the transcription units of
567 mis-expressed genes regardless of up- or down-regulation²⁶. Instead we
568 found that MeCP2 is bound at higher levels within and surrounding genes that
569 are up-regulated when MeCP2 is missing, or down-regulated by MeCP2 over-
570 expression. These findings fit well with the evidence that MeCP2 may link
571 methylated DNA with the NCoR/SMRT corepressor¹³ and they endorse the
572 conclusions of Gabel and colleagues²⁵ who found repression of long genes by
573 MeCP2. Over-expression would be expected to increase repression of the
574 most binding site-rich genes whereas MeCP2 depletion would preferentially
575 relieve their inhibition, as is observed. Although we emphasize the correlation
576 with binding site density per unit of DNA length, it is possible that the absolute
577 number of binding sites per gene also contributes to the MeCP2 response.
578 Further work is required to disentangle the roles of these related variables.

579

580 A notable feature of the MeCP2 binding site and occupancy profiles is that
581 enriched or depleted regions are larger than genes. Mis-regulated genes
582 belong to large domains that are rich in mCG and mCAC with relatively
583 homogeneous MeCP2 binding levels extending up- and down-stream of the
584 transcription unit. Within one domain, however, not all genes respond in the
585 same way to changes in MeCP2 abundance, perhaps due to additional
586 transcriptional regulatory mechanisms. The tri-nucleotide mCAC, despite its
587 lower abundance compared with relatively uniformly distributed mCG,
588 correlates strongly with MeCP2 binding and transcriptional mis-regulation in
589 response to altered levels of MeCP2. The accumulation of mCAC during
590 development of the brain adds numerically to the number of MeCP2 binding
591 targets in the genome, but it also changes the distribution of binding sites. Our
592 analysis agrees with previous suggestions that mCA is biologically important
593 for MeCP2 function in relation to transcription²⁵.

594

595 Superficially, mCG correlates less well than mCAC with changes in gene
596 expression. This effect may be exaggerated by the difference in their densities,
597 however. Mis-regulated genes have on average 0.5 extra mCAC per 1000 bp
598 when compared with those changing in the opposite direction. As the density
599 of mCG in the genome is higher than mCAC (~15 per kb), a comparable
600 incremental difference in mCG density would represent <4% of the total and
601 would not therefore be reliably measurable. These considerations may mean
602 that the small differences in gene expression in response to changing
603 amounts of MeCP2 are accompanied by equivalent subtle differences in the
604 densities of both mCG and mCAC. Which bound MeCP2 could mediate
605 repression of this kind? It is possible that only MeCP2 occupancy within a
606 gene is relevant to regulation, the flanking methylation playing no role.
607 Alternatively, the regulatory influence of MeCP2 may operate over large
608 domains encompassing many genes. Future research will attempt to
609 distinguish these possibilities.

610

611 ***Subtle effects on transcription of many genes***

612 While there is no direct evidence that aberrant gene expression is the
613 proximal cause of Rett syndrome or MeCP2 over-expression syndrome, it is
614 noteworthy that thousands of genes, including many implicated in human
615 neuronal disorders, are sensitive to altered levels of MeCP2. Mild mis-
616 regulation on this scale may destabilize neuronal function²⁵. It is worth
617 recalling that Rett neurons, though sub-optimal, are viable for many decades.
618 In this sense the biological defect can be seen as mild, despite the profound
619 effects on higher functions of the brain. The challenge now is to determine
620 how brain function might be affected by a multitude of small discrepancies in
621 gene expression. Overall, the results presented here sustain a coherent view
622 of MeCP2 function: namely that MeCP2 binding at mCG and mCAC sites
623 determines the magnitude of a repressive effect on transcription that is
624 exacerbated by MeCP2 excess and relieved by MeCP2-deficiency. With the
625 benefit of a comprehensive list of MeCP2 target sequences at the molecular
626 level, the predictions of this model can be experimentally tested, clarifying
627 further the role of MeCP2 in regulating transcription in the brain.

628 Online Methods

629

630 Animal care and transgenic mouse lines

631 All mouse studies were approved by the Austrian Animal Care Committee and
632 were licensed under the UK Animals (Scientific Procedures) Act 1986 and
633 conducted in accordance with guidelines for use and care of laboratory
634 animals. To delete *Dnmt1* in the nervous system, mice with a floxed *Dnmt1*
635 allele⁴⁹ were mated with Nestin-Cre mice⁵⁰. The *Dnmt1* Nestin-Cre strain was
636 bred in a mixed genetic background C57BL/6J x 129SV. Male *Mecp2*^{STOP/y}
637 and corresponding WT littermates⁵¹ and male *Mecp2* -/y and WT littermates³¹
638 were used as Western blotting, Real Time PCR and ChIP controls. C57Bl6
639 male WT 10 week old mice were used for FACS sorting experiments and
640 consecutive WGBS and TAB-seq.

641

642 Normalization of mRNA

643 Hypothalamus was isolated from 6 week old male WT and *Mecp2* -/y mice in
644 5 replicates. RNA and DNA were co-isolated with the AllPrep DNA/RNA Mini
645 kit (*Qiagen*) according to manufacturer's instructions with some modifications.
646 In short, tissue was homogenized in 1ml RLT buffer (spiked with 3 x 10⁶
647 *Drosophila* S2 cells/10ml RLT buffer) and centrifuged in a *Qias shredder*
648 column (*Qiagen*) for 2 minutes at full speed. The eluted RNA was next
649 subjected to treatment with the DNA-free DNA removal kit (*Ambion*) and
650 reverse transcribed with the iScript cDNA synthesis kit (*BioRad*). Real Time
651 quantitative PCR was performed on cDNA and DNA with *Drosophila* and
652 mouse specific mRNA and genomic DNA primers. For analysis, mouse mRNA
653 was normalized to *Drosophila* RNA and analogous mouse DNA was
654 normalized to *Drosophila* DNA. In the final step, corrected mRNA levels were
655 normalized to corrected DNA values. Primer sequences can be found in
656 Supplementary Table 3.

657

658 Mouse brain nuclei isolation and FACS sorting

659 Brain nuclei isolation and consecutive FACS sorting according to NeuN
660 expression was performed as described previously¹⁶. 10 week old WT Bl6
661 male mice were used and 4 brains were pooled for each replicate.

662 **Nuclear protein extracts and Western blotting**

663 Whole brain protein extracts were prepared as previously described with
664 modifications¹³. After homogenization in NE10 buffer and addition of 250 units
665 Benzonase (*Sigma*), equal amounts of 2x SDS Loading Buffer were added
666 and extracts boiled for 3 minutes. Western blotting was performed according
667 to standard procedures. Antibodies used for Western: DNMT1 1:1000 (gift
668 from Dr. Amir Eden); MeCP2 M6818 (C-terminal) 1:1000 (*Sigma*); MeCP2
669 M7443 (N-terminal) 1:1000 (*Sigma*); β -Actin A5316 (*Sigma*) 1:5000; Gapdh
670 D16H11 (*Cell Signaling*) 1:5000.

671

672 **Verification of modified oligonucleotides**

673 Dot blots of modified oligonucleotides and control DNA (Methylated standard
674 kit, *Active Motif*) were generated with Bio-Dot® Microfiltration Apparatus
675 (*BioRad*) using manufacturer's recommendations. Oligonucleotides and
676 control DNA were denaturated by the addition of [0.4M] NaOH, [10mM] EDTA
677 in a total volume of 50 μ l and boiled for 10 minutes. DNA was neutralized by
678 addition of an equal volume of ice-cold 2M ammonium acetate. Control DNA
679 and oligonucleotides were spotted in duplicate serial dilutions (Control DNA:
680 50ng, oligonucleotides: 10 μ M starting concentration). Nitrocellulose
681 membrane was UV auto-crosslinked and then blocked for 30 minutes in 5%
682 non-fat dried milk powder, 0.05% *Tween* 20/ 1x TBS. Primary antibodies were
683 incubated for 45 minutes at room temperature (5hmC: 1:10.000 *Active Motif*;
684 5mC: 1:500 *Eurogentech*). Secondary *LI-COR* antibodies were incubated in
685 the dark for 30 minutes (donkey anti mouse IRDye 800Cw; donkey anti rabbit
686 IRDye 680; *LI-COR*). Membranes were scanned with a *LI-COR* Odyssey
687 instrument.

688

689 **Protein expression, purification and EMSA**

690 Protein was prepared as described³⁹. When examining MeCP2 [1-205]
691 specificity, DNA sequence was varied at the tri-nucleotide indicated in bold
692 (See Supplementary Table 1). The primary cytosine of this tri-nucleotide was
693 either non-methylated, methylated, or hydroxymethylated. All oligonucleotides
694 were annealed to their complement, ³²P-labelled and electrophoretic mobility
695 shift assays performed on ice for 30 min using conditions described

696 previously⁴³. In competition assays to assess tri-nucleotide-binding
697 preferences of MeCP2 [1-205] a parent 58 bp *Bdnf*-probe, containing the
698 centrally methylated sequence mCGG, was ³²P-labelled and co-incubated
699 with a 2000-fold excess of cold-competitor DNA bearing one of the sequences
700 described in Fig. 2b-c. Bound complexes were resolved as described above
701 and levels of competition visualized by Phosphorimager analysis and ImageJ
702 quantification. These experiments were performed in triplicate.

703

704 **Preparation of total nucleic acid for estimation of RNA versus DNA** 705 **quantity**

706 Dissected hypothalamus tissue was homogenized in lysis buffer (10mM Tris
707 HCl [pH 7.4], 0.5% SDS, 100mM EDTA, 300µg/ml proteinase K) and
708 incubated at 50°C for 2 hours. Total nucleic acid was recovered from the
709 completely lysed sample by ethanol precipitation in 2 volumes of 100%
710 ethanol at room temperature (for 30 minutes), and pelleting by centrifugation.
711 The pellet was washed once in 2 volumes of 70% ethanol, and the nucleic
712 acid pellet was resuspended in hydrolysis buffer containing 1x DNase I buffer
713 (*NEB*), 1mM zinc sulphate, DNase I (*NEB*) and Nuclease P1 (*Sigma*). After 4
714 hours the sample was mixed thoroughly and digested for a further 8 hours.
715 After 12 hours at 37°C, the sample was heated to 92°C for 3 minutes and
716 cooled on ice. Two volumes of 30mM sodium acetate, 1mM zinc sulphate [pH
717 5.2] were added plus additional Nuclease P1 and the nucleic acids were
718 further digested to deoxyribonucleotide and ribonucleotide 5'
719 monophosphates for a further 24 hours at 37°C. The samples were then
720 subjected to HPLC as set out below.

721

722 **HPLC estimation of 5mC**

723 Measurement of genomic cytosine methylation in mouse brains was as
724 described⁵². After DNA extraction residual RNA was removed by enzymatic
725 hydrolysis (6-hour incubation with RNase A and RNase T1) followed by DNA
726 precipitation in 3 volumes of ethanol. This procedure was repeated once.
727 Purified DNA was digested with DNase I (*NEB*) for 12 hours in the
728 recommended buffer after which the sample was heated (92°C for 3 minutes)
729 to denature any remaining double stranded DNA. After cooling on ice 2

730 volumes of 30mM sodium acetate, 1mM zinc sulphate [pH 5.2] was added
 731 and the DNA was further digested to nucleotide 5' monophosphates with
 732 Nuclease P1 (*Sigma*) for 12 hours. The quantification of 5-methylcytosine in
 733 genomic DNA was by isocratic high performance reverse phase liquid
 734 chromatography as previously⁵², with the following alterations. A Dionex UM
 735 3000 HPLC system was used complete with a column chiller, C18 column
 736 (250mm x 4.6mm 5 um APEX ODS, #4M25310, *Grace Discovery Sciences*),
 737 and column guard (*Phomenex*, #AJ0-7596). The mobile phase was 50mM
 738 (monobasic) ammonium phosphate [pH4.1]. The column was chilled to 8°C to
 739 improve peak separation. Deoxyribonucleotides (dNMPs) were detected over
 740 a 70 minute run time using a Dionex 3000 multiple wavelength detector.

741

742 **HPLC nucleotide quantifications**

743 UV absorbance was recorded at 276 nm (dCMP, elution time 9.4 minutes),
 744 282 nm (5mdCMP, elution time 17 minutes), 268 nm (dTMP, elution time 21.9
 745 minutes), 260nm AMP and dAMP (elution times 27 minutes and 62.47
 746 minutes) and 254 nm (GMP and dGMP, elution times 11.1 minutes and 29.7
 747 minutes). Extinction coefficients used in nucleotide quantifications were
 748 dCMP, 8.86×10^3 ; 5mdCMP 9.0×10^3 ; dTMP, dGMP/GMP 12.16×10^3 ;
 749 dAMP/AMP 15.04×10^3 . Quantifications were calculated from the area under
 750 each peak estimated using Chromeleon software using the respective
 751 extinction coefficients.

752

753 **Transfection ChIP assay**

754 HPLC purified oligonucleotides and corresponding antisense oligonucleotides
 755 were purchased from *biomers.net*. Some oligonucleotides containing 5hmC
 756 were synthesized and characterized as described previously⁵³. All
 757 oligonucleotide sequences used are listed in Supplementary Table 1. Equal
 758 amounts of sense and antisense oligonucleotide stocks (100µM) were mixed
 759 with 10x Ligation Buffer (*NEB*) in 50µl volumes. Oligonucleotide mix was
 760 boiled in a water bath for 8 minutes and cooled to room temperature.
 761 Annealed oligonucleotides were cleaned up with MSB Spin PCRapace
 762 cleanup kit (*Invitex*) and diluted to 10µM stocks for transfections. HEK293FT
 763 cells (1.5×10^6) were transfected overnight with 0.5µg of full length MeCP2

764 tagged at the N-terminus with EGFP using Lipofectamine 2000
 765 (*Lifetechnologies*) according to manufacturer's instructions. After assessment
 766 of transfection efficiency (described in ³⁹), the medium was changed and
 767 replaced with annealed unmodified, methylated or hydroxymethylated
 768 oligonucleotides [100nM final concentration] using TransIT Oligofect reagent
 769 (*Mirus*) for 4 hours. Cells were washed with PBS and harvested by scraping.
 770 Cells were then crosslinked with formaldehyde to 1% final concentration for 5
 771 minutes at room temperature and quenched by the addition of glycine to a
 772 final concentration of 0.125M for 5 minutes followed by another two washes in
 773 PBS. Cell pellets were flash frozen in liquid nitrogen and stored at -80°C or
 774 directly used for chromatin isolation and consecutive ChIP with 4µg MeCP2
 775 M6818 antibody (*Sigma*). Primer sequences can be found in Supplementary
 776 Table 2.

777

778 **Mouse brain chromatin preparation**

779 Frozen or fresh mouse brains were homogenized in PBS supplemented with
 780 Protease Inhibitor Cocktails (*Roche*) and crosslinked with formaldehyde to 1%
 781 final concentration for 10 minutes at room temperature. After quenching the
 782 crosslink with 0.125M glycine, brains were washed in ice cold PBS twice and
 783 soluble chromatin preparation was performed.

784

785 **Chromatin preparation and ChIP**

786 Soluble chromatin preparation and ChIP assays were carried out as described
 787 previously⁵⁴ with some modification. In short, chromatin was sonicated using
 788 a Twin Bioruptor (*Diagenode*) 30sec on/off for 15 cycles at 4°C. Equal
 789 amounts of chromatin were used for IP with 4µg MeCP2 M6818 antibody
 790 (*Sigma*) and incubated overnight. Protein-antibody complexes were bound to
 791 magnetic protein G beads (*Lifetechnologies*) for 4-5 hours and washed with
 792 standard IP wash buffers for 10 minutes at 4°C. The crosslink was reversed
 793 by addition of 0.05 volume of 4M NaCl overnight at 65°C. After proteinase K
 794 digestion, DNA was recovered by phenol-chloroform-isoamylalcohol
 795 extraction and dissolved in 200µl H₂O. Real Time PCR of diluted ChIP DNA
 796 and corresponding Input DNA was performed on LightCycler (*Roche*). All
 797 primer sequences used for ChIP are listed in Supplementary Table 2.

798 **Structural modeling**

799 Modeling was based on the X-ray structure of MeCP2 (PDB code 3C2I) using
800 the program COOT⁵⁵. Atomic coordinates for DNA bases were generated
801 using the ‘mutate’ option. To optimize hydrogen-bonded and van der Waals
802 contacts between protein and different base pair sequences, the conformation
803 of the side chain of R133 was adjusted manually (all other atoms in the
804 structure were left unchanged). The potential role of water molecules in the
805 recognition of different base-pair sequences by MeCP2 was examined by
806 placing a water molecule in the highly conserved and most probable sites of
807 hydration in the major groove of B-DNA as described⁵⁶. All figures were
808 prepared using the graphics program PyMol (*DeLano Scientific, San Carlos,*
809 *CA*).

810

811 **Locus-specific bisulfite sequencing**

812 Bisulfite treatment of genomic DNA of P11 WT and *Dnmt1* KO whole brain
813 triplicates and DNA sequencing was carried out as previously described⁵⁷ with
814 some modifications. Genomic DNA was isolated with the DNeasy Blood &
815 Tissue kit (*Qiagen*) and bisulfite treated with the EpiTect bisulfite kit (*Qiagen*).
816 Bisulfite Primers were designed with MethPrimer software⁵⁸. Sequences were
817 analyzed with the BISMA online tool⁵⁹.

818

819 **Reduced representation bisulfite sequencing**

820 Genomic DNA from whole brain of 5 day old *Dnmt1* WT and KO littermate
821 replicates was isolated using the DNeasy Blood & Tissue Kit (*Qiagen*). For
822 RRBS³⁴, the “Reduced Representation Bisulfite Sequencing for Methylation
823 Analysis” protocol (www.support.illumina.com) was followed and TruSeq® v2
824 DNA sample preparation kit (*Illumina*) was used. Cleaned up libraries were
825 validated on a Bioanalyzer High Sensitivity DNA Chip (*Agilent*) and 100 bp
826 paired-end sequencing was performed on Illumina HiSeq 2000 platform
827 (Wellcome Trust Sanger Institute, Hinxton, UK).

828

829 **Whole genome bisulfite and TAB sequencing**

830 Genomic DNA of three replicates of 10 week old male Bl/6 WT dissected
831 hypothalamus samples was prepared with the DNeasy Blood & Tissue Kit

(Qiagen) and 0.5% unmethylated λ DNA (Promega) was spiked in. Equal amounts of genomic DNA were bisulfite converted with the EZ DNA Methylation Gold Kit (Zymo Research) and libraries were prepared with the TruSeq DNA methylation kit (Illumina) according to manufacturer's instructions. WGBS and TAB sequencing from NeuN positive sorted neuronal nuclei was as described previously⁶⁰. For TAB treatment, half the DNA was glycosylated, TET oxidized and spiked with control DNA. The other half was left untreated and spiked with unmethylated λ DNA (Promega). NGS libraries were prepared with TruSeq DNA Sample Preparation Kit (Illumina) according to manufacturer's instructions. After size selection, all libraries were bisulfite treated with EpiTect Bisulfite Kit (Qiagen) and amplified with Pfu Turbo Cx Polymerase (Stratagene) for 7 PCR cycles. Cleaned-up libraries were validated on a Bioanalyzer High Sensitivity DNA Chip (Agilent) and 100 bp paired-end sequencing performed on an Illumina HiSeq 2000 platform (Wellcome Trust Sanger Institute, Hinxton, UK).

847

848 **Bioinformatics and statistical analyses**

849 *Genome build and annotations*

850 All data were aligned to the mouse NCBI 37 (mm9) assembly. Gene
851 annotations were obtained from version 67 of the Ensembl database.

852 *ChIP-seq*

853 MeCP2 Chip and Input datasets were downloaded from the Gene Expression
854 Omnibus (GEO) under the accession numbers GSE66868²⁶ and GSE60062²⁵.
855 Raw sequencing reads were first trimmed and filtered using Trimmomatic
856 v0.32⁶¹ then aligned with bwa using the samse algorithm. Alignments were
857 then filtered to remove reads classed as duplicates, non-unique or those that
858 fell in the blacklisted regions outlined by the Encode project⁶². Alignments
859 were converted to bigWig files using the deepTools package for genome wide
860 visualization and analysis. Read counts were normalized to RPKM to account
861 for differences in library size. The bigwigCompare tool was used to calculate a
862 log2 ratio of MeCP2 ChIP/Input signal across the genome.

863

864

865 *Whole genome bisulfite sequencing of mouse hypothalamus*

866 Quality trimming, filtering and adapter removal were performed by
867 trimmomatic v0.32 prior to alignment. The software package Bismark v0.15
868 was used to map reads to the mm9 genome, remove duplicate alignments
869 and extract methylation calls⁶³.

870 *BS-seq and TAB-seq*

871 Processed bisulfite and TAB-seq datasets containing aligned methylation calls
872 were obtained from the GEO records GSM1173786_allC.MethylC-
873 Seq_mm_fc_male_7wk_neun_pos and GSM1173795_allC.TAB-
874 Seq_mm_fc_6wk respectively. Percentage (%) DNA methylation at a given
875 site i corresponds to the ratio of mC basecalls for that site to the count of all
876 reads mapping to that site ($m^i = mC/C$). Context-specific % mean
877 methylation for a given region (i.e. bin or gene) was defined as $\hat{m}_{CX} = \frac{\sum_i m_{CX}^i}{N_{CX}}$,
878 where N_{CX} is the number of C's within the region, occurring in context CX . In
879 addition, context-specific methylation density was defined as $\frac{N_{CX} \cdot \hat{m}_{CX}}{L}$, where L
880 is the length of the region.

881 *RNA-seq*

882 Raw reads from mouse hypothalamus were downloaded from the Gene
883 Expression Omnibus (GEO) under the accession no. GSE66870²⁶.
884 Trimmomatic v0.32 was used to remove adaptor contamination and to trim
885 low quality reads. Reads were mapped to the genome using STAR v 2.4.2a⁶⁴.
886 Alignments were then filtered to remove non-unique and blacklisted reads.
887 HTseq-count v0.6.0 was used to quantify read counts over gene exons in the
888 union mode.

889 *Correlation between MeCP2 ChIP-Seq and Input read counts (Fig. 5e-f)*

890 For the MeCP2 ChIP and Input sample of the Chen dataset²⁶ we computed
891 the number of reads that mapped to 10 kb windows covering the entire
892 genome. For this we have shifted the reads by 134 bp which corresponds
893 to the estimated fragment length determined by MACS. A Kernel density
894 estimate was applied. We have applied linear regression (MeCP2 ~ Input)
895 and analysed the residuals r_i for each bin. Windows with residuals exceeding,

896 a threshold are considered enriched or depleted respectively. The threshold

897 was chosen to be $2 \cdot \sqrt{\sum_i r_i^2 / N}$.

898 *Correlation between log2(MeCP2/Input) and DNA methylation in genomic*
899 *windows (Fig. 5i-j and Supplementary Figs. 4j-r)*

900 We calculated read coverage values and context-dependent methylation
901 densities for 1 kb windows across the genome. Regions were ordered by their
902 log2(MeCP2 Chip/Input) signal and mean methylation densities were
903 calculated for groups of 1000 regions. Datasets used are ²⁶ and
904 hypothalamus WGBS (this study).

905 *MeCP2 summit analysis (Fig. 5g-h and Supplementary Figs. 4b-i)*

906 Summits of MeCP2 ChIP enrichment over Input were defined using MACS
907 and the method described in ²⁵ for both MeCP2 datasets (^{25,26}). We used the
908 Bioconductor package *seqplots* v1.4.0 to plot methylation density across the
909 summits for various cytosine contexts. Mean methylation densities were
910 calculated for 20 bp windows across the summits extended by 4 kb on either
911 side. The ChIP-seq dataset from cortex was combined with bisulfite analysis
912 from ³ (Fig. 5g and Supplementary Figs. 4i, l-m) and the hypothalamus ChIP-
913 seq dataset ²⁶ was correlated with bisulfite data from hypothalamus WGBS
914 (this study) (Fig. 5h and Supplementary Figs. 4b-h and n-r) or bisulfite data
915 from ³ (Supplementary Figs. 4j-k).

916 *Rolling mean plots for genic regions (Fig. 5k)*

917 Genes were sorted according to their MeCP2/Input enrichment and rolling
918 means of methylation density were applied over subsets of 400 genes with a
919 step of 80 genes. Datasets used are ²⁶ and hypothalamus WGBS (this study).

920 *MSR (Fig. 5 l-m and Supplementary Figs. 4s-u)*

921 We used the MSR tool to find domains of MeCP2 enrichment and depletion
922 relative to the Input sample⁴². As background we used a mappability map for
923 the m9 genome (parameters: L=45, P-value threshold: 1e-6). For each
924 significantly enriched or depleted segment we determined the methylation
925 densities and averaged over segments with same scale and enrichment score.
926 Datasets used are ²⁶ and hypothalamus WGBS (this study).

927 *Differential expression analysis (Fig. 6)*

We used DeSeq2 v1.8.1⁶⁵ to determine mis-regulated genes in KO and OE. To account for the observed 15% reduction of total mRNA in the KO samples, we first used the DESeq function estimateSizeFactors to normalize the data sets and subsequently multiplied the obtained size factors for the KO samples by 1.15. An adjusted p-value threshold of <0.05 was used to determine up- or down-regulated genes. For an exemplary group of genes for which little change is observed between conditions we used $p_{adj} > 0.5$ with a $\log_2\text{FoldChange} < 0.01$. We further filtered out protein-coding genes with constitutively low expression (TPM <5 in all samples) and genes with the lowest Input coverage. We retained 12,510 protein-coding genes, of which 403 showed higher, and 8363 lower expression in KO vs WT, respectively. In addition, we found 1372 to have higher and 2719 to have lower expression in OE relative to WT. Dataset used is ²⁶.

Aggregated signal plots across gene features (Fig. 6a-f and Supplementary Fig. 4a)

We used *seqplots* to generate aggregate plots at transcription start sites of protein coding genes and annotated CGIs. Mean values for $\log_2(\text{MeCP2 ChIP/Input})$ and methylation densities were calculated over 1 kb windows for each set of genes and 100 bp windows for CGIs (Supplementary Fig. 4a). Datasets used are ²⁶ and hypothalamus WGBS (this study).

Author Contributions

SL, JCC and GSch performed or directed experiments and analyses in the study. SL designed and performed *Dnmt1* KO mouse experiments and BHR did HPLC. JCC and MDW designed and performed EMSA analyses and structural modeling. SL and LS designed and performed *in vivo* transfection assays. SL, JCC and DD performed bisulfite sequencing and fragment cloning. SL, CH and MY designed and performed WGBS and TAB sequencing. GSch, SW and GSi performed bioinformatics and statistical analyses. JS and CS provided mouse reagents. LS provided 5hmC-containing oligonucleotides and contributed ideas. AB, SL, JCC and GSch wrote the manuscript, SL prepared the figures and all co-authors passed comment. AB advised on all aspects of the study.

962 **Acknowledgements**

963 We are grateful to Kashyap Chhatbar for bioinformatics assistance and to
 964 Michael Greenberg, Gail Mandel and their group members for valuable Input.
 965 We also thank Justyna Cholewa-Waclaw, Rebekah Tillotson, Ruth Shah and
 966 Martha Koerner for useful discussions and Astrid Hagelkruys, Alan McClure
 967 and Martin Waterfall for technical assistance. We are grateful to Amir Eden for
 968 the DNMT1 antibody and the Patrick Heun lab for providing *Drosophila* S2
 969 cells and primers. The work was supported by a Consortium Grant from The
 970 Rett Syndrome Research Trust and by a Wellcome Trust programme grant
 971 (091580), Wellcome Trust Centre Core Grant (092076), European Research
 972 Council (MLCS 306999) (GSI) and Austrian Science Fund (P25807) (CS). SL
 973 and GSch held EMBO long-term fellowships. GSch was also funded by a
 974 Marie Curie Postdoc fellowship and SL is currently a fellow of the Veterinary
 975 Medicine University of Vienna Postdoc program. LS was supported in part by
 976 grant CA184097 from the National Cancer Institute, National Institutes of
 977 Health.

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

Figure Legends

Figure 1 DNA methylation is the primary determinant of MeCP2 binding *in vivo*. (a) Western blot of whole brain from WT and *Dnmt1* conditional KO mice at different age points (E = embryonic day, P = postnatal day) with antibodies against Dnmt1 and β -Actin. (b) Real Time PCR of *IAP* expression in WT and *Mecp2* $-/-$ male (8 weeks old) and WT and *Dnmt1* KO (P11) triplicate whole brain cDNAs normalized to *Cyclophilin A*. *Mecp2* WT and *Dnmt1* WT data points were set to 1. (c) HPLC analysis of % total C methylation in 5 samples of P12 and P5 WT and *Dnmt1* KO whole brain. WT methylation levels set to 100%. Individual replicates at two age points are shown in Supplementary Fig. 1c. (d) RRBS of WT and *Dnmt1* KO brain duplicates at P5. UCSC browser screenshot of chromosome 5. Scale bar = 50 Mb. Grey levels indicate methylation levels between 0 (white) and 100% (black). (e) Western blot of WT, *Mecp2* KO adult and WT, *Dnmt1* KO whole brain at P5 and P11 with antibodies against MeCP2 and Gapdh. (f) Quantification of triplicate WT and *Dnmt1* KO whole brain at P5 and P11. Western blot for quantification is shown in Supplementary Fig. 1e. (g-h) ChIP of WT and *Mecp2* KO adult and WT and *Dnmt1* KO replicates at *IAP* (g) and major satellite repeats (h) at P12. Results are presented as % of input. All error bars represent $\pm$ SD. P values were calculated with Students unpaired one-tailed t-test (g-h) or Students unpaired two-tailed t-test (b-c, f): ns $p>0.05$; * $p<0.05$; ** $p<0.01$, *** $p<0.001$. (i-j) Bisulfite sequencing of *IAP* (i) and major satellite repeats (j) in WT and *Dnmt1* KO genomic DNA from whole brain at P12. Percentage of methylation is indicated. Rows correspond to individual sequenced DNA strands.

Figure 2 MeCP2 binds mCAC and hmCAC *in vitro*. (a) EMSA using no protein (-) or varying amounts of MeCP2 [1-205] with a probe (Supplementary Table 1) containing a centrally methylated C in a CAX context. Gap indicates separate gels. (b) EMSA using MeCP2 [1-205] or no added protein (-) in the presence of excess unlabeled unmethylated competitor DNA (CGG) or methylated competitor DNA (mCXX). Labeled probe contains mCGG. Red denotes strongest competition. (c) Quantification of (b). Three individual experiments were averaged. Red denotes most significant competition. Significance was calculated in relation to unmethylated CGG (grey bar). Error

bars represent $\pm$ SD. Students unpaired t-test: * $p < 0.05$; ** $p < 0.01$. (d) EMSA using no protein (-) or varying amounts of MeCP2 [1-205] with a probe containing a centrally hydroxymethylated C in a CAX triplet context. * : free probe; - : bound probe. (e) Summary of MeCP2 binding motifs *in vitro*. M = 5-methylcytosine, H = 5-hydroxymethylcytosine. Bright red: strong binding; pale red: weaker binding; grey: not tested.

1036

Figure 3 Full-length MeCP2 binds mCAC and hmCAC *in vivo*. (a) Schematic of *in vivo* transfection assay in HEK293FT cells. (b) Differentially modified oligonucleotide derived from the mouse *Bdnf* locus. Light grey: T3 and M13-20 adapters; red circles: mC; blue circles: hmC. See also Supplementary Table 1. (c) Real Time PCR of *in vivo* transfection assay in triplicate where WT MeCP2-GFP was co-transfected with oligonucleotide as described in (b). (d) Differentially modified oligonucleotide at three CAC sites. See also Supplementary Table 1. Light grey: T3 and M13-20 adapters; red circles: mC; blue circles: hmC. (e) Real Time PCR of *in vivo* transfection assay in triplicates where WT MeCP2-GFP was co-transfected with oligonucleotides as described in (d). Real Time PCR results are presented as % of Input (red bars: mC; blue bars: hmC, white bars: unmethylated; grey bar: mCG oligonucleotide transfected without prior transfection of MeCP2-GFP as a background control). Error bars represent $\pm$ SD. Significance was calculated in relation to unmethylated oligonucleotide transfections (white bars). Students unpaired t-test: ns $p > 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

1053

Figure 4 MeCP2 tri-nucleotide recognition may depend on flexibility of the R133 side chain and requires the methyl group of thymine. (a) X-ray structure of MeCP2 [77-167] interaction with the mCG di-nucleotide in double-strand DNA⁴⁰. Critical amino acids R111, R133 and D121 are shown in pink. Green sphere: methyl group; grey sphere: water molecule; green dotted lines (from R111 and D121): favorable interactions; grey dotted lines: hydrogen bonds. (b) Modeling of the MeCP2 [77-167] interaction with double-strand DNA containing mCAC. Key as in (a) and black dotted lines: modeled interactions. (c) Structure of 5-methylcytosine, thymine and uracil. Note that thymine and

1063 uracil are distinguishable by a methyl group on position 5 of the pyrimidine
1064 ring. (d) EMSA using no (-) or varying amounts of MeCP2 [1-205] to assess
1065 the influence of the methyl group of thymine on binding. * : free probe; - :
1066 bound probe.

1067

1068 **Figure 5** DNA methylation and MeCP2 binding in the mouse brain. (a) Pie
1069 chart showing CX frequencies in the mouse genome (dataset WGBS of sorted
1070 neurons from this study). (b) Pie chart showing modified (mCG and hmCG)
1071 CX frequencies in brain neuronal nuclei as determined by WGBS (dataset
1072 WGBS of sorted neurons from this study). (c) Mean methylation levels (%) of
1073 CXX in brain neuronal nuclei based on WGBS of sorted neurons (dataset
1074 from this study). (d) IGV browser screenshot of mCAC, mCG, MeCP2 ChIP-
1075 seq and corresponding Input DNA based on sequence reads. Data represent
1076 a 100 Mb region of chromosome 6 (datasets from ^{3,26}). (e) Correlation
1077 between sequence coverage of MeCP2 ChIP-seq and corresponding Input
1078 DNA in 10 kb windows. Fitting a linear model (MeCP2 ~ Input) yields a
1079 coefficient of determination of 0.84 (dataset from ²⁶). (f) As for (e) but
1080 highlighting relative depletion (4.1% of windows; green) and relative
1081 enrichment (1.7% of windows; purple). When these outliers are excluded,
1082 90% of the variability of the MeCP2 signal in the remaining binned windows is
1083 explained by sequence bias (Input coverage) (dataset from ²⁶). (g-h)
1084 Enrichment of mCAC at summits of MeCP2 binding (datasets from ^{3,25} (g) and
1085 ²⁶ and hypothalamus WGBS from this study (h)). (i) Relationship between
1086 mCG and unmethylated CG density/kb and MeCP2 occupancy corrected for
1087 Input in genome-wide 1 kb windows (datasets from ²⁶ and hypothalamus
1088 WGBS from this study). (j) Relationship between mCAX and mCTC density/kb
1089 and MeCP2 occupancy corrected for Input in genome-wide 1 kb windows
1090 (datasets from ²⁶ and hypothalamus WGBS from this study). (k) Methylation
1091 density as a function of log₂(MeCP2 ChIP/Input) in protein-coding genes. The
1092 density of tick marks on the x-axis represents the distribution of genes with
1093 respect to MeCP2 binding (datasets from ²⁶ and hypothalamus WGBS from
1094 this study). (l-m) Domains of MeCP2 enrichment and depletion as determined
1095 by MSR (datasets from ²⁶ and hypothalamus WGBS from this study). Grey

1096 dotted line indicates zero. (l) Heatmap showing number of segments binned
1097 by their scale (x-axis) and MeCP2 enrichment scores (y-axis). X-axis shows
1098 the median lengths of segments found with a given scale. Positive scores on
1099 the y-axis indicate MeCP2 enrichment; negative scores indicate MeCP2
1100 depletion. Plot is colored according to number of segments. (m) Heatmap
1101 showing number of segments binned by scale and scored as in (l). Plot is
1102 colored according to combined mCG + mCAC density.

1103

1104 **Figure 6** Genes show global down-regulation in KO, but MeCP2 enriched
1105 genes are up-regulated upon MeCP2 depletion and down-regulated when
1106 MeCP2 is overexpressed. Datasets ²⁶ and hypothalamus WGBS from this
1107 study. (a, d) Aggregate WT MeCP2 occupancy plotted 100 kb up- and down-
1108 stream of the transcription start site (TSS) of genes that are up-regulated
1109 (purple), down-regulated (green) or unchanged (black) in (a) *Mecp2* KO or (d)
1110 *Mecp2* OE hypothalamus. (b, e) Methylation density per kb of CAC and (c, f)
1111 CG plotted 100 kb up- and down-stream of the TSS using the same gene sets
1112 as in (a) and (d) respectively. (g) IGV screenshot of chromosome 5
1113 (chr5:95,751,179-110,732,516) showing unpruned MSR regions of scale 30,
1114 corresponding to a median segment length of 270 kb and 240 kb for MeCP2
1115 enriched and MeCP2 depleted segments. Datasets ²⁶ and hypothalamus
1116 WGBS from this study. (h) Pie charts showing percentage of up- (left) and
1117 down-regulated (right) genes and their overlap with MeCP2 enriched (purple)
1118 or depleted (green) domains. Dataset ²⁶ and hypothalamus WGBS from this
1119 study. (i) Transcriptional changes resulting from MeCP2 deficiency as a
1120 function of MeCP2 occupancy. Genes that are up-regulated in KO (purple)
1121 have high MeCP2 occupancy and show the strongest down-regulation in OE
1122 (j). FC = fold change. Histogram depicting the correlation between MeCP2
1123 occupancy and % of genes that are up- (purple) or down-regulated (green).
1124 Dataset ²⁶. (k-l) MeCP2 occupancy and expression changes are well
1125 correlated at genes previously implicated in neurological diseases⁴⁷:
1126 Spearman correlation coefficients of 0.67 for KO vs WT (k) and -0.5 for OE vs
1127 WT (l) Dataset ²⁶.

1128

1129

References

1. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-26 (1992).
2. Guo, J. U. *et al.* Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* **17**, 215-22 (2014).
3. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
4. Ramsahoye, B. H. *et al.* Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5237-42 (2000).
5. Aoki, A. *et al.* Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* **29**, 3506-12 (2001).
6. Gowher, H. & Jeltsch, A. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *J. Mol. Biol.* **309**, 1201-8 (2001).
7. He, Y. & Ecker, J. R. Non-CG Methylation in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **16**, 55-77 (2015).
8. Xie, W. *et al.* Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816-31 (2012).
9. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18**, 6538-47 (1998).
10. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185-8 (1999).
11. del Gaudio, D. *et al.* Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet. Med.* **8**, 784-92 (2006).
12. Lyst, M. J. & Bird, A. Rett syndrome: a complex disorder with simple roots. *Nat. Rev. Genet.* **16**, 261-75 (2015).
13. Lyst, M. J. *et al.* Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16**, 898-902 (2013).
14. Lewis, J. D. *et al.* Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**, 905-14 (1992).
15. Baubec, T., Ivánek, R., Lienert, F. & Schübeler, D. Methylation-dependent and

- 1165 -independent genomic targeting principles of the MBD protein family. *Cell* **153**,
1166 480-92 (2013).
- 1167 16. Skene, P. J. *et al.* Neuronal MeCP2 is expressed at near histone-octamer
1168 levels and globally alters the chromatin state. *Mol. Cell* **37**, 457-68 (2010).
- 1169 17. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to
1170 5hmC enriched within active genes and accessible chromatin in the nervous
1171 system. *Cell* **151**, 1417-30 (2012).
- 1172 18. Shin, J., Ming, G.-L. & Song, H. By hook or by crook: multifaceted DNA-
1173 binding properties of MeCP2. *Cell* **152**, 940-2 (2013).
- 1174 19. Baker, S. A. *et al.* An AT-hook domain in MeCP2 determines the clinical
1175 course of Rett syndrome and related disorders. *Cell* **152**, 984-96 (2013).
- 1176 20. Nan, X., Campoy, F. J. & Bird, A. MeCP2 is a transcriptional repressor with
1177 abundant binding sites in genomic chromatin. *Cell* **88**, 471-81 (1997).
- 1178 21. Tudor, M., Akbarian, S., Chen, R. Z. & Jaenisch, R. Transcriptional profiling of
1179 a mouse model for Rett syndrome reveals subtle transcriptional changes in the
1180 brain. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 15536-41 (2002).
- 1181 22. Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C. A. & Zoghbi, H. Y.
1182 Mouse models of MeCP2 disorders share gene expression changes in the
1183 cerebellum and hypothalamus. *Hum. Mol. Genet.* **18**, 2431-42 (2009).
- 1184 23. Chahrour, M. *et al.* MeCP2, a key contributor to neurological disease,
1185 activates and represses transcription. *Science* **320**, 1224-9 (2008).
- 1186 24. Li, Y. *et al.* Global transcriptional and translational repression in human-
1187 embryonic-stem-cell-derived Rett syndrome neurons. *Cell Stem Cell* **13**,
1188 446-58 (2013).
- 1189 25. Gabel, H. W. *et al.* Disruption of DNA-methylation-dependent long gene
1190 repression in Rett syndrome. *Nature* **522**, 89-93 (2015).
- 1191 26. Chen, L. *et al.* MeCP2 binds to non-CG methylated DNA as neurons mature,
1192 influencing transcription and the timing of onset for Rett syndrome. *Proc. Natl.*
1193 *Acad. Sci. U. S. A.* **112**, 5509-14 (2015).
- 1194 27. Nan, X., Tate, P., Li, E. & Bird, A. DNA methylation specifies chromosomal
1195 localization of MeCP2. *Mol. Cell. Biol.* **16**, 414-21 (1996).
- 1196 28. Dhasarathy, A. & Wade, P. A. The MBD protein family-reading an epigenetic
1197 mark? *Mutat. Res.* **647**, 39-43 (2008).
- 1198 29. Fan, G. *et al.* DNA hypomethylation perturbs the function and survival of CNS
1199 neurons in postnatal animals. *J. Neurosci.* **21**, 788-97 (2001).

- 1200 30. Mullaney, B. ., Johnston, M. . & Blue, M. . Developmental expression of
1201 methyl-CpG binding protein 2 is dynamically regulated in the rodent brain.
1202 *Neuroscience* **123**, 939-949 (2004).
- 1203 31. Guy, J., Hendrich, B., Holmes, M., Martin, J. E. & Bird, A. A mouse Mecp2-null
1204 mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.*
1205 **27**, 322-6 (2001).
- 1206 32. Chen, R. Z., Akbarian, S., Tudor, M. & Jaenisch, R. Deficiency of methyl-CpG
1207 binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat.*
1208 *Genet.* **27**, 327-31 (2001).
- 1209 33. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous
1210 retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20**, 116-7
1211 (1998).
- 1212 34. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and
1213 differentiated cells. *Nature* **454**, 766-70 (2008).
- 1214 35. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is
1215 present in Purkinje neurons and the brain. *Science* **324**, 929-30 (2009).
- 1216 36. Valinluck, V. *et al.* Oxidative damage to methyl-CpG sequences inhibits the
1217 binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding
1218 protein 2 (MeCP2). *Nucleic Acids Res.* **32**, 4100-8 (2004).
- 1219 37. Khrapunov, S. *et al.* Unusual characteristics of the DNA binding domain of
1220 epigenetic regulatory protein MeCP2 determine its binding specificity.
1221 *Biochemistry* **53**, 3379-91 (2014).
- 1222 38. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and
1223 erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40**, 4841-9 (2012).
- 1224 39. Brown, K. *et al.* The molecular basis of variable phenotypic severity among
1225 common missense mutations causing Rett syndrome. *Hum. Mol. Genet.* 1-42
1226 (2015). doi:10.1093/hmg/ddv496
- 1227 40. Ho, K. L. *et al.* MeCP2 binding to DNA depends upon hydration at methyl-CpG.
1228 *Mol. Cell* **29**, 525-31 (2008).
- 1229 41. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**,
1230 R137 (2008).
- 1231 42. Knijnenburg, T. A. *et al.* Multiscale representation of genomic signals. *Nat.*
1232 *Methods* **11**, 689-94 (2014).
- 1233 43. Klose, R. J. *et al.* DNA binding selectivity of MeCP2 due to a requirement for
1234 A/T sequences adjacent to methyl-CpG. *Mol. Cell* **19**, 667-78 (2005).

- 1235 44. Luikenhuis, S., Giacometti, E., Beard, C. F. & Jaenisch, R. Expression of
1236 MeCP2 in postmitotic neurons rescues Rett syndrome in mice. *Proc. Natl.*
1237 *Acad. Sci. U. S. A.* **101**, 6033-8 (2004).
- 1238 45. Collins, A. L. *et al.* Mild overexpression of MeCP2 causes a progressive
1239 neurological disorder in mice. *Hum. Mol. Genet.* **13**, 2679-89 (2004).
- 1240 46. Yazdani, M. *et al.* Disease modeling using embryonic stem cells: MeCP2
1241 regulates nuclear size and RNA synthesis in neurons. *Stem Cells* **30**, 2128-39
1242 (2012).
- 1243 47. Zhu, X., Need, A. C., Petrovski, S. & Goldstein, D. B. One gene, many
1244 neuropsychiatric disorders: lessons from Mendelian diseases. *Nat. Neurosci.*
1245 **17**, 773-81 (2014).
- 1246 48. Rube, H. T. *et al.* Sequence features accurately predict genome-wide MeCP2
1247 binding in vivo. *Nat. Commun.* **7**, 11025 (2016).
- 1248 49. Jackson-Grusby, L. *et al.* Loss of genomic methylation causes p53-dependent
1249 apoptosis and epigenetic deregulation. *Nat. Genet.* **27**, 31-9 (2001).
- 1250 50. Tronche, F. *et al.* Disruption of the glucocorticoid receptor gene in the nervous
1251 system results in reduced anxiety. *Nat. Genet.* **23**, 99-103 (1999).
- 1252 51. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of neurological
1253 defects in a mouse model of Rett syndrome. *Science* **315**, 1143-7 (2007).
- 1254 52. Ramsahoye, B. H. Measurement of genome wide DNA methylation by
1255 reversed-phase high-performance liquid chromatography. *Methods* **27**, 156-61
1256 (2002).
- 1257 53. Tardy-Planechaud, S., Fujimoto, J., Lin, S. S. & Sowers, L. C. Solid phase
1258 synthesis and restriction endonuclease cleavage of oligodeoxynucleotides
1259 containing 5-(hydroxymethyl)-cytosine. *Nucleic Acids Res.* **25**, 553-9 (1997).
- 1260 54. Hauser, C., Schuettengruber, B., Bartl, S., Lagger, G. & Seiser, C. Activation
1261 of the mouse histone deacetylase 1 gene by cooperative histone
1262 phosphorylation and acetylation. *Mol. Cell. Biol.* **22**, 7820-30 (2002).
- 1263 55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics.
1264 *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126-32 (2004).
- 1265 56. Schneider, B. & Berman, H. M. Hydration of the DNA bases is local. *Biophys.*
1266 *J.* **69**, 2661-9 (1995).
- 1267 57. Illingworth, R. *et al.* A novel CpG island set identifies tissue-specific
1268 methylation at developmental gene loci. *PLoS Biol.* **6**, e22 (2008).
- 1269 58. Li, L.-C. & Dahiya, R. MethPrimer: designing primers for methylation PCRs.

1270 *Bioinformatics* **18**, 1427-31 (2002).

1271 59. Rohde, C. *et al.* BISMA - Fast and accurate bisulfite sequencing data analysis
1272 of individual clones from unique and repetitive sequences. *BMC Bioinformatics*
1273 **11**, 230 (2010).

1274 60. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the
1275 mammalian genome. *Cell* **149**, 1368-80 (2012).

1276 61. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for
1277 Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014).

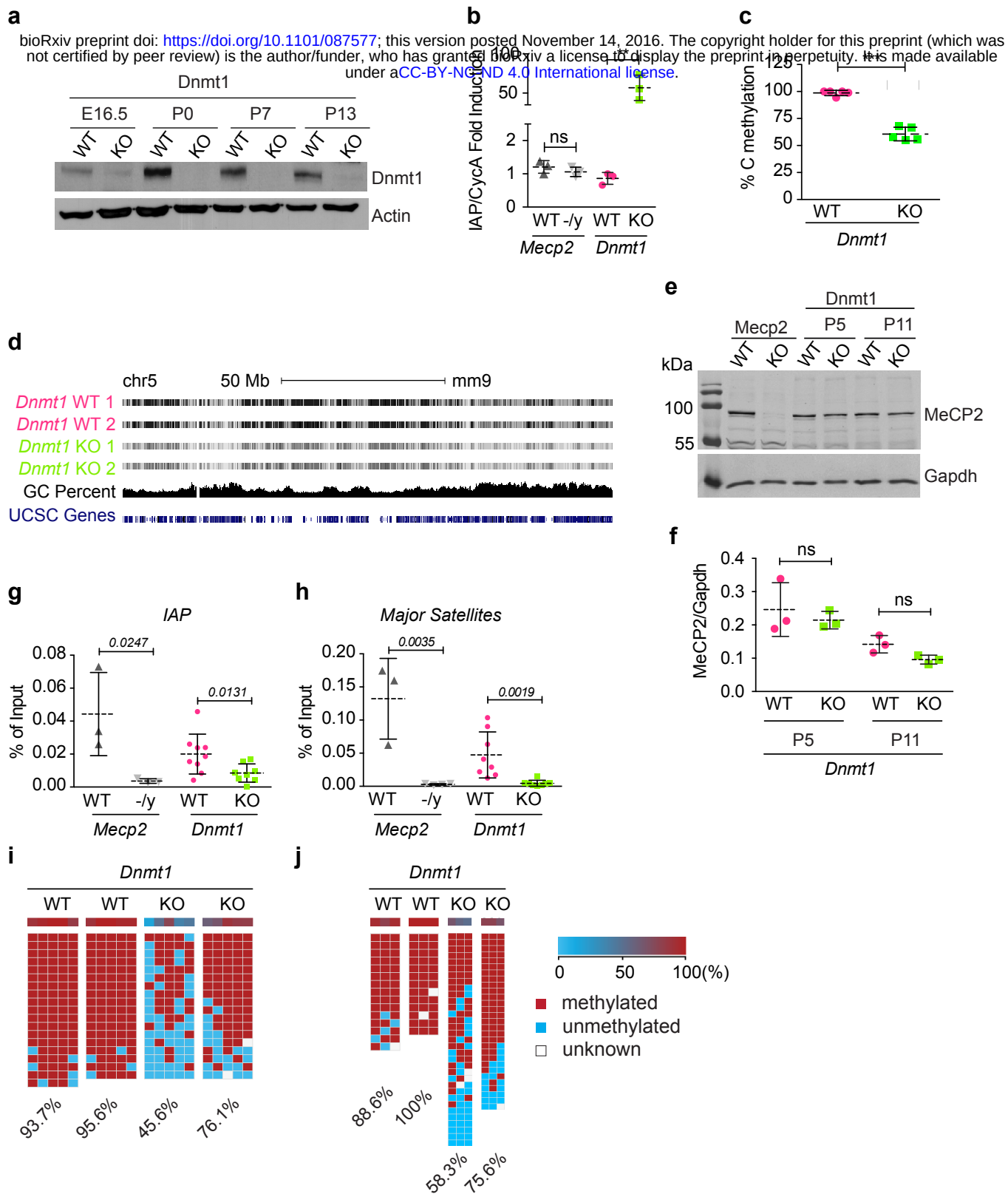
1278 62. Bernstein, B. E. *et al.* An integrated encyclopedia of DNA elements in the
1279 human genome. *Nature* **489**, 57-74 (2012).

1280 63. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller
1281 for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-2 (2011).

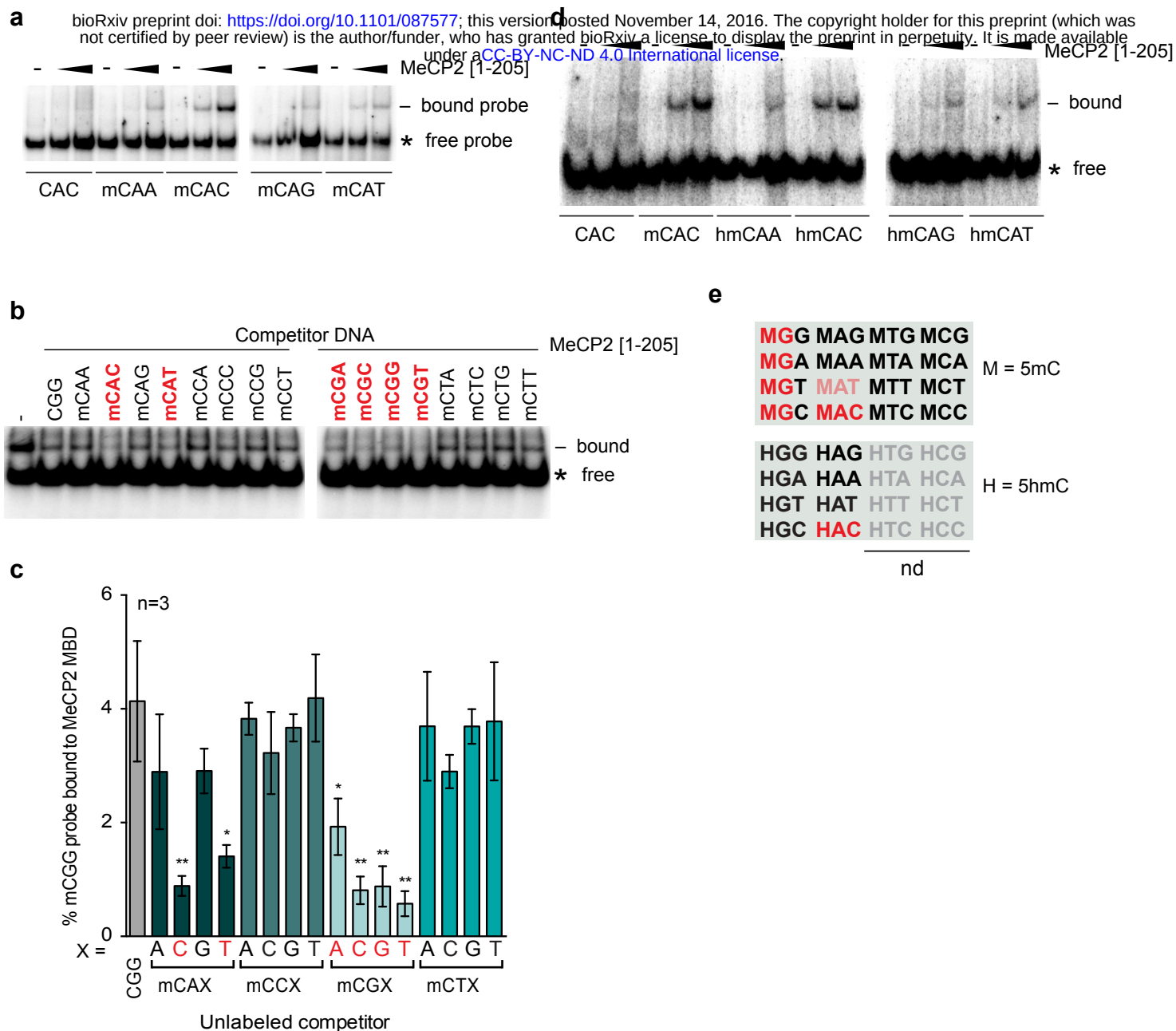
1282 64. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**,
1283 15-21 (2013).

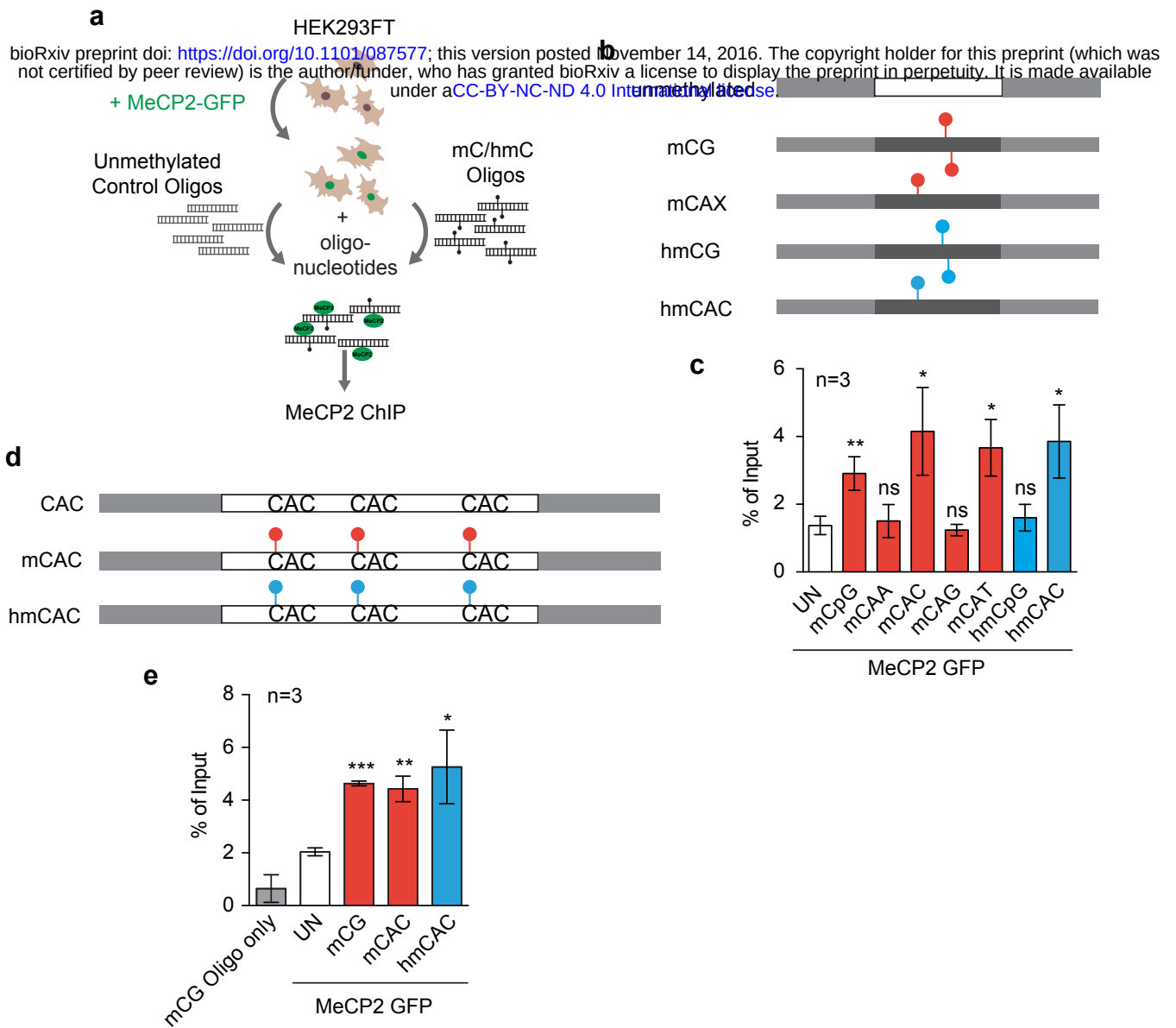
1284 65. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
1285 dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

1286



Lagger et al – Figure 1





Lagger et al – Figure 3

