



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## CpG islands influence chromatin structure via the CpG-binding protein Cfp1

**Citation for published version:**

Thomson, JP, Skene, PJ, Selfridge, J, Clouaire, T, Guy, J, Webb, S, Kerr, ARW, Deaton, A, Andrews, R, James, KD, Turner, DJ, Illingworth, R & Bird, A 2010, 'CpG islands influence chromatin structure via the CpG-binding protein Cfp1', *Nature*, vol. 464, no. 7291, pp. 1082-1086. <https://doi.org/10.1038/nature08924>

**Digital Object Identifier (DOI):**

[10.1038/nature08924](https://doi.org/10.1038/nature08924)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Nature

**Publisher Rights Statement:**

Free in PMC.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



Published in final edited form as:

Nature. 2010 April 15; 464(7291): 1082–1086. doi:10.1038/nature08924.

## CpG islands influence chromatin structure via the CpG-binding protein Cfp1

John P. Thomson<sup>1,\*</sup>, Peter J. Skene<sup>1,\*</sup>, Jim Selfridge<sup>1</sup>, Thomas Clouaire<sup>1</sup>, Jacky Guy<sup>1</sup>, Shaun Webb<sup>1</sup>, Alastair R. W. Kerr<sup>1</sup>, Aimée Deaton<sup>1</sup>, Rob Andrews<sup>2</sup>, Keith D. James<sup>2</sup>, Daniel J. Turner<sup>2</sup>, Robert Illingworth<sup>1</sup>, and Adrian Bird<sup>1</sup>

<sup>1</sup>Wellcome Trust Centre for Cell Biology, Michael Swann Building, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, UK.

<sup>2</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

### Abstract

CpG islands (CGIs) are prominent in the mammalian genome owing to their GC-rich base composition and high density of CpG dinucleotides<sup>1,2</sup>. Most human gene promoters are embedded within CGIs that lack DNA methylation and coincide with sites of histone H3 lysine 4 trimethylation (H3K4me3), irrespective of transcriptional activity<sup>3,4</sup>. In spite of these intriguing correlations, the functional significance of non-methylated CGI sequences with respect to chromatin structure and transcription is unknown. By performing a search for proteins that are common to all CGIs, here we show high enrichment for Cfp1, which selectively binds to non-methylated CpGs *in vitro*<sup>5,6</sup>. Chromatin immunoprecipitation of a mono-allelically methylated CGI confirmed that Cfp1 specifically associates with non-methylated CpG sites *in vivo*. High throughput sequencing of Cfp1-bound chromatin identified a notable concordance with non-methylated CGIs and sites of H3K4me3 in the mouse brain. Levels of H3K4me3 at CGIs were markedly reduced in Cfp1-depleted cells, consistent with the finding that Cfp1 associates with the H3K4 methyltransferase Setd1 (refs 7, 8). To test whether non-methylated CpG-dense sequences are sufficient to establish domains of H3K4me3, we analysed artificial CpG clusters that were integrated into the mouse genome. Despite the absence of promoters, the insertions recruited Cfp1 and created new peaks of H3K4me3. The data indicate that a primary function of non-methylated CGIs is to genetically influence the local chromatin modification state by interaction with Cfp1 and perhaps other CpG-binding proteins.

To characterize the chromatin modifications typical of CGIs, we used the methyl-CpG-sensitive restriction endonuclease HinPI (cleavage site GCGC) to release small chromatin fragments from purified brain nuclei, as described previously<sup>9</sup>. As sites for this enzyme in bulk chromatin are rare and generally uncleavable owing to DNA methylation, the released

©2010 Macmillan Publishers Limited. All rights reserved

Correspondence and requests for materials should be addressed to A.B. (a.bird@ed.ac.uk).

\*These authors contributed equally to this work.

**Author Contributions** J.P.T. and P.J.S. performed the ChIP, knockdown and bisulphite analysis. T.C. did preliminary ChIP analysis. J.S. and J.G. generated the MeCP2-eGFP cell line. J.S. performed bisulphite analysis of the TβC44 ES cell line. R.I. prepared samples for sequencing. K.D.J., D.J.T. and R.A. performed the sequencing and mapping. S.W., A.R.W.K., A.D. and R.I. performed the bioinformatic analysis. J.P.T., P.J.S., T.C., R.I. and A.B. wrote the manuscript.

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Author Information** High throughput sequencing data has been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE18578. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests.

fraction predominantly contains non-methylated CGIs. Confirming this, further digestion of the deproteinized DNA with HpaII (cleavage site CCGG) specifically collapsed the nucleosomal ladder generated by HinPI, but had little effect on DNA released from bulk chromatin with MseI (cleavage site TTAA; Supplementary Fig. 1)<sup>9</sup>. Western blotting confirmed that non-methylated CGI chromatin is enriched for histone modifications associated with actively transcribed genes (acetylated histone H3, H3K4me3 and H3K4me2) compared with bulk chromatin (Fig. 1a). In contrast, CGI chromatin was depleted for marks not found at active promoters: H3K36me3, H3K9me3, H3K27me3 and H4K20me3 (Fig. 1a). Agreement between these results and genome-wide studies of chromatin modifications<sup>3,4,10,11</sup> indicated that this fraction could be used to identify proteins that preferentially localize to non-methylated CGIs. We first tested CXXC finger protein 1 (Cfp1), which binds to non-methylated CpG dinucleotides *in vitro* by a CXXC zinc finger domain<sup>6,12</sup>. The data showed that Cfp1 is enriched within the CGI fraction of the genome (Fig. 1a). Similarly, Kdm2a, an H3K36 demethylase that also contains a CXXC domain<sup>13</sup>, was enriched in the CGI fraction.

Focusing on Cfp1, we tested its *in vivo* binding specificity by chromatin immunoprecipitation (ChIP) at an endogenous CGI that is present in both methylated and non-methylated states. The *Xist* CGI is mono-allelically methylated in female cells, but fully methylated in males, which only have one X chromosome<sup>14</sup>. ChIP analysis of mouse brain tissue identified a peak of Cfp1 binding over the *Xist* CGI in females, but no peak was present in males, suggesting that Cfp1 exclusively binds to the non-methylated allele (Fig. 1b). To test this more stringently, we used bisulphite sequencing across the *Xist* locus to determine the methylation status of the immunoprecipitated chromatin recovered from females. As expected, input DNA comprised equal numbers of methylated and non-methylated DNA clones. DNA immunoprecipitated by the Cfp1 antibody was almost exclusively non-methylated (96%), however, whereas DNA immunoprecipitated with an antibody against the methyl-CpG-binding protein MeCP2 (refs 15–17) was predominantly methylated (88%; Fig. 1c). We conclude that Cfp1 selectively binds to non-methylated CpGs *in vivo*.

To test whether Cfp1 is concentrated at non-methylated CpGs within CGIs, we analysed the genome-wide distribution of Cfp1 using high-throughput DNA sequencing of immunoprecipitated DNA (ChIP-Seq). Prominent peaks of Cfp1 binding co-localized with non-methylated CGIs (Fig. 2a), 81% of which were Cfp1-associated. Cfp1 has been identified as part of the Setd1 H3K4 methyltransferase complex<sup>8</sup> and ChIP-Seq with H3K4me3 antibodies showed that 93% of Cfp1-bound CGIs also possess this histone modification (Fig. 2b and Supplementary Table 1). Consistent with the possibility that Cfp1 binding is responsible for recruiting the Setd1 complex to these sites, Cfp1-negative non-methylated CGIs (19% of the total) also lack H3K4me3 (Fig. 2b). Despite being rich in non-methylated CpGs, these CGIs are somehow refractory to Cfp1 binding. One potential explanation came from alignment with the published<sup>18</sup> distribution of the polycomb-associated mark H3K27me3 (ref. 19) in mouse brain. More than half (58%) of Cfp1-negative and H3K4me3-negative CGIs contained the H3K27 modification (Fig. 2a, b and Supplementary Fig. 2). In these cases H3K27me3 and polycomb binding may render a CpG island refractory to Cfp1 binding and to H3K4 methylation.

To assess the importance of Cfp1 for the recruitment of H3K4me3, we used stably expressed short hairpin RNAs (shRNAs) directed against *Cfp1* to reduce its level in NIH3T3 cells. Single shRNAs reduced Cfp1 (Supplementary Fig. 3), but a combination of three gave a greater effect (Fig. 3a). Depleted cells showed altered morphology (Fig. 3b) and retarded growth (Fig. 3c). ChIP analysis revealed a loss of Cfp1 binding compared with vector-only transfected cells accompanied by a precipitous drop in levels of H3K4me3 across CGIs at

the brain-derived neurotrophic factor (*Bdnf*),  $\beta$ -actin (*Actb*), *c-Myc* and *Dlx5/6* genes (Fig. 3d). The same results were obtained with clones expressing each of two independent shRNA sequences, ruling out off-target effects of shRNA expression (Supplementary Fig. 3). As a further control, H3K27me3 profiles at the same loci were unaffected by depletion of Cfp1 (Fig. 3d and Supplementary Fig. 3b). The loss of H3K4me3 at six randomly selected CGI promoters in Cfp1-depleted cells argues that this modification is dependent on the presence of Cfp1.

Although Cfp1 binds non-methylated CpGs and seems to be required for H3K4 methylation at CGIs, it is possible that this reflects indirect recruitment of Setd1 by RNA polymerase II, which is present at active CGI promoters. Alignment of ChIP-Seq profiles for Cfp1, H3K4me3 and the unphosphorylated form of RNA polymerase II indeed showed co-localization of all three signals at 86% of all Cfp1-bound CGIs (Supplementary Table 1 and Supplementary Fig. 4). In a small proportion (7%) of cases, however, RNA polymerase II was undetectable, despite the presence of robust peaks of H3K4me3 and Cfp1 (Supplementary Fig. 4). This raised the possibility that RNA polymerase II may not be required and that Cfp1 binding is sufficient to direct H3K4 trimethylation. To test this hypothesis, we used embryonic stem (ES) cell lines in which artificial promoterless CpG-rich DNA sequences had been introduced into the genome at sites that normally lack H3K4me3. The DNA insert in ES line T $\beta$ C44 (ref. 20) comprises a 720-base-pair (bp) enhanced green fluorescent protein (eGFP) coding sequence containing 60 CpGs<sup>21</sup> adjacent to a 600-bp puromycin-resistance gene with 93 CpGs (Fig. 4a). The inserted sequence has the typical CpG density of a CGI, but lacks a promoter. Bisulphite analysis showed that integrated sequence is non-methylated (Fig. 4a). In the targeted cells, prominent domains of Cfp1 and H3K4me3 coincided with the inserted CpG-rich DNA (Fig. 4b). Interestingly, the peaks of H3K4me3 and Cfp1 tracked CpG density as expected if H3K4me3 is determined by this DNA dinucleotide sequence (Fig. 4b, broken line). No peak of RNA polymerase was detected. An independent ES cell line carrying an eGFP insertion on the X chromosome<sup>22</sup> (Fig. 4c) also created a peak of H3K4me3 and Cfp1 (Fig. 4d). In this case, bisulphite sequencing showed that approximately a quarter of the integrated sequences were hypomethylated and the remainder were densely methylated (Fig. 4e, input panel). ChIP-bisulphite analysis demonstrated that Cfp1 and H3K4me3 antibodies significantly enriched the hypomethylated sequences (Fig. 4e). We conclude that clusters of non-methylated CpG are sufficient to recruit Cfp1 and create a peak of H3K4me3 modification, even in the absence of a promoter.

The density of non-methylated CpG is ~50-fold higher in CGIs than in bulk genomic DNA, as CpG in the latter is deficient (20% of expected<sup>23</sup>) and mostly methylated (~70%). It is unclear whether this high CpG density arises as a passive consequence of events at promoters and has no functional significance, or whether it has been selected over evolutionary time because it facilitates transcription (or other DNA-related processes). Our results favour selection, as they indicate that CpG density per se can directly influence histone modification status by the recruitment of the Cfp1 protein and its associated Setd1 histone H3K4 methyltransferase complex. The ability of an exogenous promoter-less CpG-rich insertion to create *de novo* an H3K4me3 focus provides strong support for this notion. An attractive biological rationale for this phenomenon may be simplification of the large mammalian genome by the creation of 'beacons' of H3K4me3 that highlight CGI promoters within the genomic landscape<sup>1</sup>.

Whether CpG clustering is sufficient to create stable non-methylated CGIs is uncertain. There is evidence that H3K4me3 is incompatible with *de novo* methylation as components of the DNA methyltransferase complex (Dnmt3L) are repelled by this modification<sup>24</sup>. In theory, therefore, Cfp1-bound CGIs should be intrinsically stable in the non-methylated

state. Previous studies suggest, however, that transcription also has a role. Maintenance of non-methylated CGIs through the waves of *de novo* methylation in the early embryo depends on promoter function, as point mutations that prevent transcription factor binding without significantly reducing CpG density destroy the immunity of a CpG island to DNA methylation<sup>25,26</sup>. It follows that H3K4 methylation due to CpG clustering may not be sufficient to reliably perpetuate the non-methylated state. Indeed, more than half of cells carrying the promoter-less eGFP insertion at the *Mecp2* locus had acquired dense methylation in ES cells despite the presence of a CpG cluster.

Our data suggest that chromatin modification need not arise secondarily as a result of, for example, transcriptional status, but can be determined genetically due to the sequence characteristics of the underlying DNA. In particular CpG, by virtue of its widely varying local densities and alternative modification states, has the properties of a signalling module that locally influences genome function. As shown here, DNA methylation-free CpG clusters can recruit Cfp1 and probably other CXXC domain proteins. Densely methylated CGIs, on the other hand, attract methyl-CpG-binding proteins, which in turn recruit enzymes that can reinforce repressive histone modifications<sup>17,27,28</sup>. Future studies of proteins that read and interpret CpG signals promise to shed further light on both genetic and epigenetic determinants of chromosome function.

## METHODS

### Release of CGI chromatin

Nuclei were prepared from brains of 4-week-old mice as previously described<sup>31</sup>. Nuclear preparations were digested with a two-fold excess of HinP1 or Mse1 in a buffer containing 50 mM Tris-HCl, pH 8, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.1 mM EGTA and 1 mM β-mercaptoethanol. The released chromatin was retained in the supernatant after centrifugation at 3,800*g* for 5 min and the proteins were precipitated using trichloroacetic acid before western blot analysis.

### Antibodies

Antibodies used are listed in Supplementary Table 2.

### Chromatin immunoprecipitation and bisulphite sequencing

ChIP on brain tissue was performed as described<sup>17</sup> using antibodies as shown in Supplementary Table 2. Most ChIP-qPCR profiles were replicated using independent Cfp1 antibodies. Illumina linkers were ligated in-house and Solexa sequencing was carried out using Illumina 2G Solexa sequencers using two replicate lanes per biological sample. ChIP-Seq was analysed using custom bioinformatic tools generated in-house (see Supplementary Table 3 for the parameters used). ChIP using formaldehyde crosslinked NIH3T3 cells was performed as previously described<sup>32</sup>. Bisulphite sequencing was performed as described<sup>29</sup>. Real-time PCR was carried out using Quantace Sensimix Plus using a Biorad iCycler according to the manufacturer's instructions (primer sequences are available on request).

### Generation of stable Cfp1-knockdown cells

NIH3T3 cells were transfected using lipofectamine reagent (Invitrogen) with three independent pSuper vectors containing short hairpin constructs directed against *Cfp1* (Oligoengine) or vector alone. Target sequences were as follows: target 986, 5'-GAAGGUGAA GCACGUGAAG-3'; target 1250, 5'-CAGCCAACCGAAUCUAUGA-3'; and target 1920, 5'-CUUCACCAAACGAUCCAAC-3'. Stable clones were selected for puromycin resistance. A combination of the three shRNAs reduced Cfp1 more robustly and was therefore used for the data in Fig. 3. Individual shRNAs gave comparable results by

western and ChIP (see Supplementary Fig. 3). RNA was extracted using Tri reagent (Sigma) and complementary DNA was prepared using reverse transcriptase (Promega). Expression levels were determined using real-time PCR analysis (primer sequences available on request).

## ES cell lines

ES cell line T $\beta$ C44 was generated by homologous recombination as described<sup>20</sup>. A *Mecp2-eGFP* knock-in targeting vector was constructed by sequential cloning of 5' (5.3 kb) and 3' (1.9 kb) regions of *Mecp2* homology into peGFP-N1 (Clontech). A PGK-Neo cassette flanked by *loxP* sites was added to enable selection of transfected cells. Gene targeting was carried out in the ES cell line E14 TG2a to generate an insertion into the *Mecp2* gene transcription unit at the junction between the open reading frame and the 3' untranslated region. This construct was initially designed to create a MeCP2-eGFP fusion protein after transcription and translation. Cells were grown on gelatinized dishes in the presence of recombinant human LIF in Glasgow MEM (Invitrogen) supplemented with 10% FBS (Gibco), 1 $\times$  MEM non-essential amino acids, sodium pyruvate (1 mM) and  $\beta$ -mercaptoethanol (50  $\mu$ M; all Invitrogen). ES cells ( $5 \times 10^7$  cells) were transfected with linearized targeting vector (250  $\mu$ g DNA in 0.8 ml HEPES buffered saline) by electroporation (800 V, 3  $\mu$ F, BioRad Gene Pulser) and plated at  $5 \times 10^6$  cells per dish. Correctly targeted clones were first identified by PCR specific for homologous recombination. The integrity of the targeted locus was confirmed by Southern blot analyses. A single positive clone was transiently transfected with pCAGGS-CRE<sup>33</sup> for the Cre-mediated deletion of the selectable marker and a recombinant clone was then used for this study.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We are grateful to D. Skalnik for the gift of a Cfp1 antibody, I. Chambers for the T $\beta$ C44 ES cell line, R. Klose for discussions, E. Sheridan for testing the DNA sequencing protocol, and K. Auger and J. Parkhill for coordinating the DNA sequencing. We also thank R. Ekiert and J. Connelly for comments on the manuscript. This work was funded by a Cancer Research UK studentship to J.P.T. and by grants from the Wellcome Trust, the Medical Research Council and the European Union 'Epigenome' Network of Excellence.

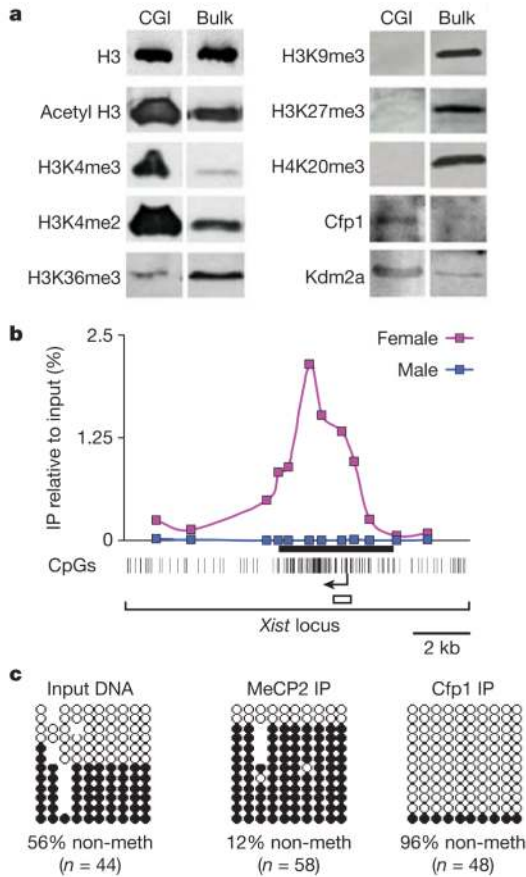
## References

1. Bird AP. CpG-rich islands and the function of DNA methylation. *Nature*. 1986; 321:209–213. [PubMed: 2423876]
2. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nature Rev. Genet.* 2008; 9:465–476. [PubMed: 18463664]
3. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*. 2007; 130:77–88. [PubMed: 17632057]
4. Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006; 125:315–326. [PubMed: 16630819]
5. Lee JH, Skalnik DG. CpG-binding protein is a nuclear matrix- and euchromatin-associated protein localized to nuclear speckles containing human trithorax. Identification of nuclear matrix targeting signals. *J. Biol. Chem.* 2002; 277:42259–42267. [PubMed: 12200428]
6. Shin Voo K, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA

- methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol. Cell. Biol.* 2000; 20:2108–2121. [PubMed: 10688657]
7. Lee JH, Skalnik DG. CpG-binding protein (CXXC fingerprotein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem.* 2005; 280:41725–41731. [PubMed: 16253997]
  8. Lee JH, Tate CM, You JS, Skalnik DG. Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *J. Biol. Chem.* 2007; 282:13419–13428. [PubMed: 17355966]
  9. Tazi J, Bird A. Alternative chromatin structure at CpG islands. *Cell.* 1990; 60:909–920. [PubMed: 2317863]
  10. Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature.* 2007; 448:553–560. [PubMed: 17603471]
  11. Barski A, et al. High-resolution profiling of histone methylations in the human genome. *Cell.* 2007; 129:823–837. [PubMed: 17512414]
  12. Lee JH, Voo KS, Skalnik DG. Identification and characterization of the DNA binding domain of CpG-binding protein. *J. Biol. Chem.* 2001; 276:44669–44676. [PubMed: 11572867]
  13. Tsukada Y, et al. Histone demethylation by a family of JmjC domain-containing proteins. *Nature.* 2006; 439:811–816. [PubMed: 16362057]
  14. McDonald LE, Paterson CA, Kay GF. Bisulfite genomic sequencing-derived methylation profile of the *Xist* gene throughout early mouse development. *Genomics.* 1998; 54:379–386. [PubMed: 9878240]
  15. Lewis JD, et al. Purification, sequence and cellular localisation of a novel chromosomal protein that binds to methylated DNA. *Cell.* 1992; 69:905–914. [PubMed: 1606614]
  16. Nan X, Tate P, Li E, Bird A. DNA methylation specifies chromosomal localization of MeCP2. *Mol. Cell. Biol.* 1996; 16:414–421. [PubMed: 8524323]
  17. Skene PJ, et al. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell.* 2010; 37:457–468. [PubMed: 20188665]
  18. Mikkelsen TS, et al. Dissecting direct reprogramming through integrative genomic analysis. *Nature.* 2008; 454:49–55. [PubMed: 18509334]
  19. Schwartz YB, Pirrotta V. Polycomb complexes and epigenetic states. *Curr. Opin. Cell Biol.* 2008; 20:266–273. [PubMed: 18439810]
  20. Chambers I, et al. Nanog safeguards pluripotency and mediates germline development. *Nature.* 2007; 450:1230–1234. [PubMed: 18097409]
  21. Yang TT, Cheng L, Kain SR. Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res.* 1996; 24:4592–4593. [PubMed: 8948654]
  22. Guy J, Gan J, Selfridge J, Cobb S, Bird A. Reversal of neurological defects in a mouse model of Rett syndrome. *Science.* 2007; 315:1143–1147. [PubMed: 17289941]
  23. Waterston RH, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 2002; 420:520–562. [PubMed: 12466850]
  24. Ooi SK, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature.* 2007; 448:714–717. [PubMed: 17687327]
  25. Brandeis M, et al. Sp1 elements protect a CpG island from *de novo* methylation. *Nature.* 1994; 371:435–438. [PubMed: 8090226]
  26. Macleod D, Charlton J, Mullins J, Bird AP. Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes Dev.* 1994; 8:2282–2292. [PubMed: 7958895]
  27. Lorincz MC, Dickerson DR, Schmitt M, Groudine M. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nature Struct. Mol. Biol.* 2004; 11:1068–1075. [PubMed: 15467727]
  28. Lorincz MC, Schubeler D, Groudine M. Methylation-mediated proviral silencing is associated with MeCP2 recruitment and localized histone H3 deacetylation. *Mol. Cell. Biol.* 2001; 21:7913–7922. [PubMed: 11689684]

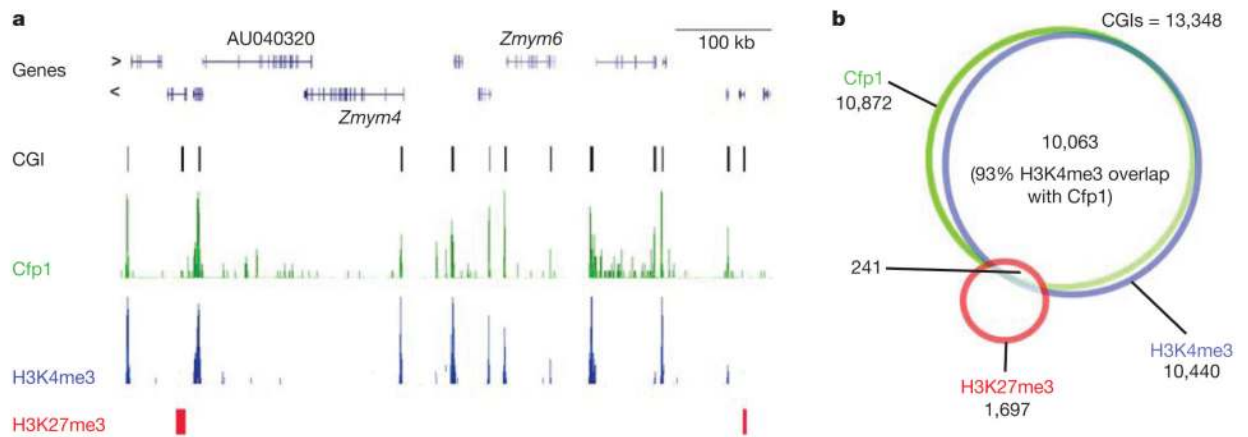
29. Illingworth R, et al. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* 2008; 6:e22. [PubMed: 18232738]
30. Meissner A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature.* 2008; 454:766–770. [PubMed: 18600261]
31. Klose RJ, Bird AP. MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. *J. Biol. Chem.* 2004; 279:46490–46496. [PubMed: 15322089]
32. Schmiedeberg L, Skene P, Deaton A, Bird A. A temporal threshold for formaldehyde crosslinking and fixation. *PLoS ONE.* 2009; 4:e4636. [PubMed: 19247482]
33. Araki K, Araki M, Yamamura K. Targeted integration of DNA using mutant lox sites in embryonic stem cells. *Nucleic Acids Res.* 1997; 25:868–872. [PubMed: 9016639]





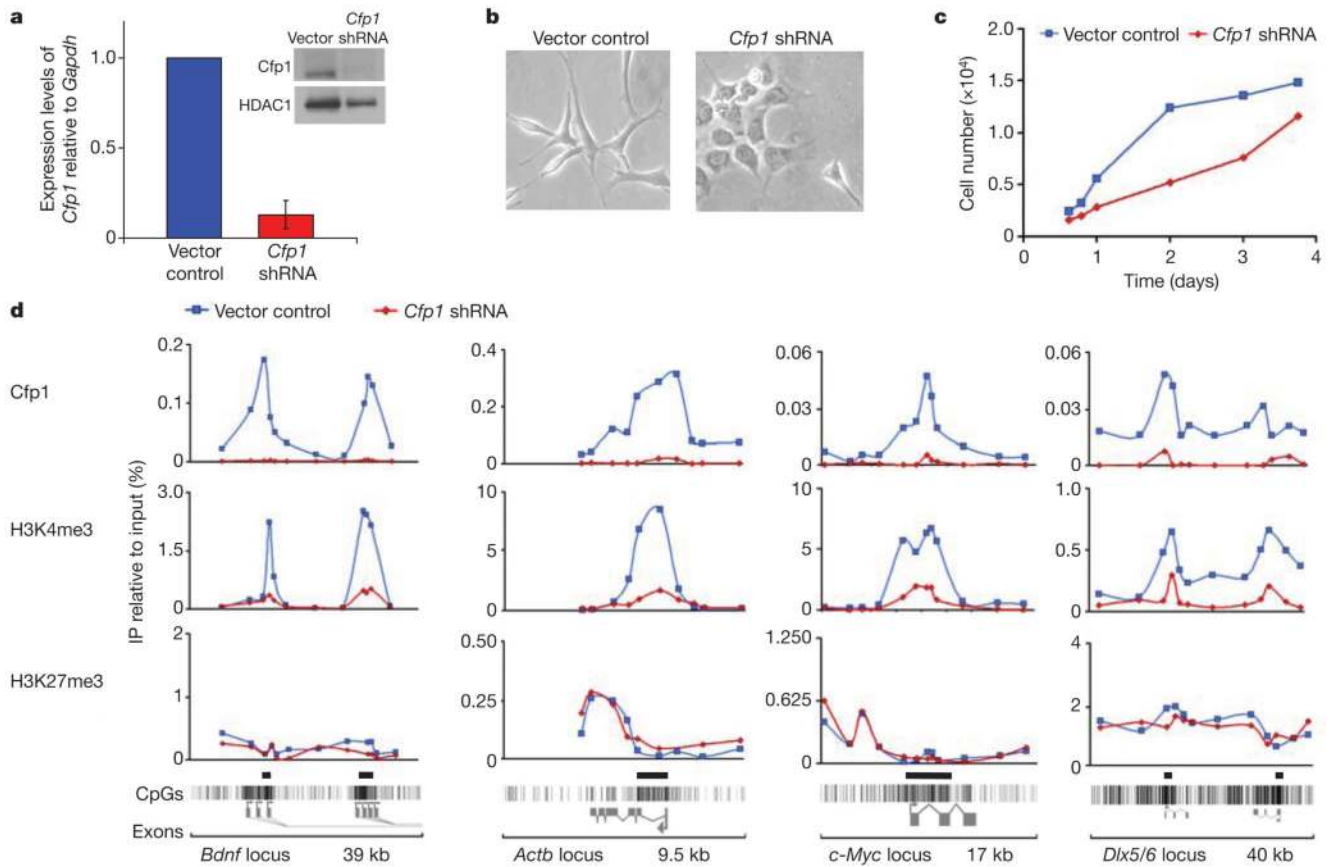
**Figure 1. Cfp1 is enriched in non-methylated CpG island chromatin**

**a**, Western blot analysis of non-methylated CGI and bulk chromatin released from purified nuclei using antibodies against selected histone modifications and CpG-binding proteins (normalized to histone H3 levels). An antibody against histone H3 served as a loading control. Modifications associated with transcriptional activity, including H3 acetylated on amino acids K9 and K14, H3K4me3 and H3K4me2 were enriched in CGI chromatin, whereas the elongation and silencing marks H3K36me3, H3K9me3, H3K27me3 and H4K20me3 were depleted. The CXXC domain proteins Cfp1 and Kdm2a also showed enrichment within the CGI fraction. **b**, Cfp1 ChIP assayed by qPCR across the X-linked mouse *Xist* locus. Vertical strokes beneath the plot represent CpGs within the locus and the black bar above demarcates the CGI. The open box below the CpG map shows the region amplified for bisulphite analysis. IP, immunoprecipitation. **c**, Bisulphite analysis of input chromatin (female brain) and chromatin immunoprecipitated with Cfp1 antibodies and control MeCP2 antibodies. Twelve representative clones are shown from the total number sequenced (number in brackets). Solid and open circles represent methylated and non-methylated CpGs, respectively. Uncharacterized CpGs are represented as gaps.



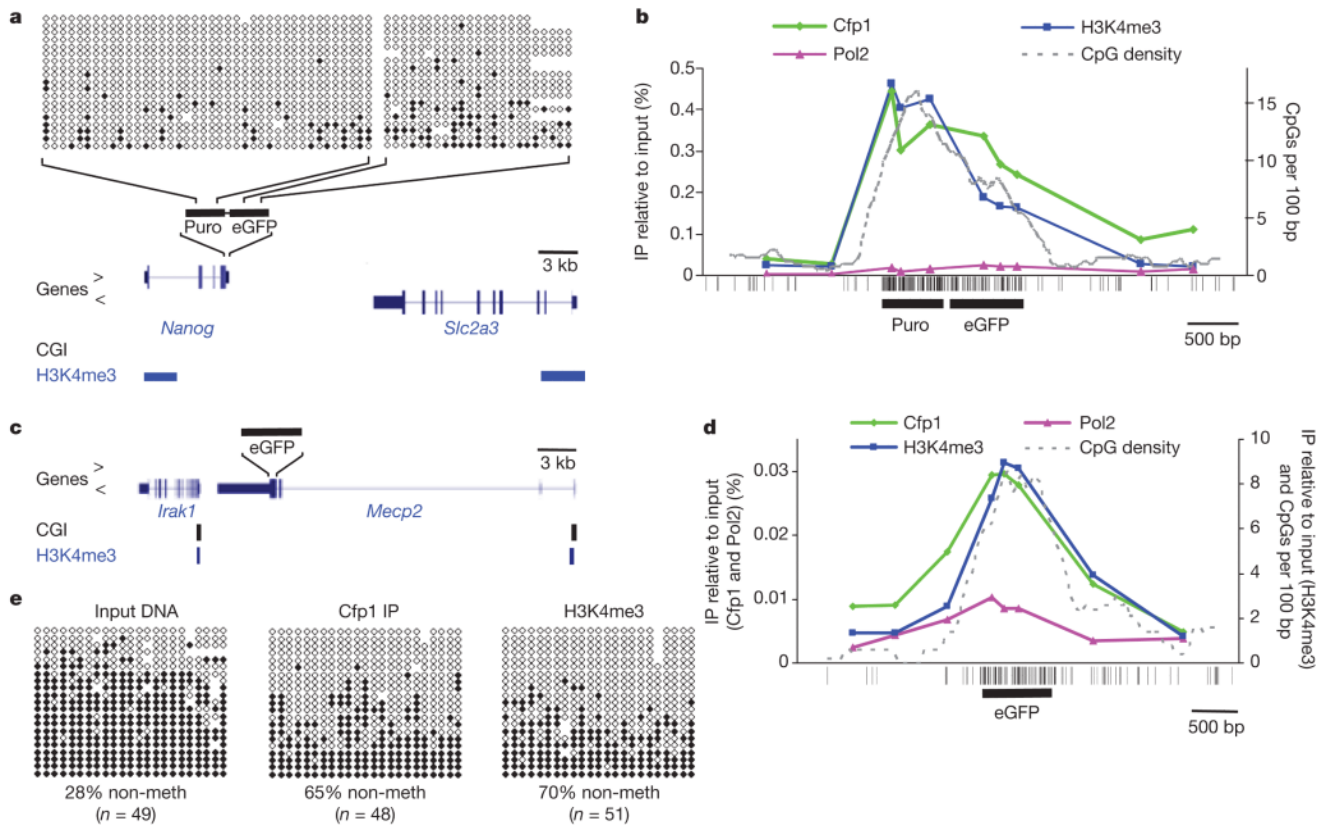
**Figure 2. Genome-wide ChIP sequencing shows a tight association between Cfp1 and H3K4me3 at CGIs**

**a**, Typical Cfp1 ChIP-Seq profiles from whole mouse brain. For comparison, we also carried out H3K4me3 ChIP-Seq. The data were aligned with non-methylated CGIs mapped in mouse brain using a CXXC affinity column<sup>29</sup>. The panel shows a typical region of the genome from chromosome 4 (nucleotides 126,333,759–127,054,849) demonstrating the coincidence of Cfp1 and H3K4me3 peaks with CGIs. A subset of genes is labelled (RefSeq). Two CGIs that lack H3K4me3 and Cfp1 coincide with sites of H3K27me3 binding (red rectangles; data of ref. 30 for mouse brain). **b**, Venn diagram showing strong overlap between H3K4me3 and Cfp1 peaks in mouse brain chromatin, but minimal overlap with H3K27me3.



**Figure 3. Depletion of Cfp1 results in reduced H3K4 trimethylation levels at CpG islands**

**a**, Expression of three short hairpin RNAs in NIH3T3 fibroblasts reduced *Cfp1* messenger RNA levels to 15% compared with vector-only transfected control cells. Expression of *Cfp1* relative to *Gapdh* in control cells is set to 1. The inset shows reduction of Cfp1 by western blotting. Error bars indicate s.d. ( $n = 3$ ) **b**, **c**, Gross morphology (**b**) and growth rate (**c**) of Cfp1-depleted versus vector-only transformant cells. Cells were plated at low density and monitored at the times shown using a haemocytometer. Original magnification,  $\times 200$ . **d**, ChIP qPCR using Cfp1, H3K4me3 and H3K27me3 antibodies at selected loci in vector-only control and Cfp1-depleted NIH3T3 cells. The results were replicated with an independent clone expressing the same shRNA combination (data not shown) and with each of two individual shRNA constructs (see Supplementary Fig. 3).



**Figure 4. Artificial promoterless CpG-rich sequences recruit Cfp1 and generate new H3K4me3 peaks in mouse ES cells**

**a**, The TC $\beta$ 44 ES cell line carries adjacent promoterless eGFP and bacterial puromycin-resistance sequences (black bars) inserted together within the 3' untranslated region of the *Nanog* gene<sup>20</sup>. The positions of CGIs and H3K4me3 peaks<sup>4</sup> at this locus in wild-type ES cells are shown below the map. DNA methylation within the insertion was determined by bisulphite sequencing of 306-bp (eGFP) and 275-bp (Puro) segments of the inserted sequence. **b**, ChIP qPCR across the region containing the insertion using antibodies against Cfp1, RNA polymerase II (Pol2) and H3K4me3. The dotted line plots CpG density in a 500-bp window with a 100-bp step size. Vertical strokes below the graph mark CpG sites within and surrounding the insertion. **c**, A second ES cell line carried an eGFP coding sequence (black bar) inserted into the 3' untranslated region of the X-linked *Mecn2* gene. CGIs and H3K4me3 peaks<sup>4</sup> at this locus in wild-type ES cells are shown below. **d**, ChIP qPCR across the region containing the insertion (black bar) using antibodies against Cfp1, RNA polymerase II and H3K4me3. **e**, Bisulphite sequence analysis determined the methylation status of input and DNA immunoprecipitated by the Cfp1 and H3K4me3 antibodies. The percentage of non-methylated CpGs is shown below each panel. H3K4me3 ChIP data in **b** and **d** used different commercial antibodies with differing affinities (see Supplementary Table 2).