Published in final edited form as:

Science. 2017 May 05; 356(6337): 503–508. doi:10.1126/science.aag3260.

Integration of CpG-free DNA induces de novo methylation of CpG islands in pluripotent stem cells

Yuta Takahashi^{1,2}, Jun Wu^{1,3}, Keiichiro Suzuki¹, Paloma Martinez-Redondo¹, Mo Li^{1,4}, Hsin-Kai Liao¹, Min-Zu Wu^{1,3}, Reyna Hernández-Benítez^{1,4}, Tomoaki Hishida¹, Maxim Nikolaievich Shokhirev⁵, Concepcion Rodriguez Esteban¹, Ignacio Sancho-Martinez^{1,*}, and Juan Carlos Izpisua Belmonte^{1,†}

¹Gene Expression Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

²Life Science Center, Tsukuba Advanced Research Alliance, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8577, Japan

³Universidad Católica San Antonio de Murcia (UCAM) Campus de los Jerónimos, N° 135 Guadalupe 30107, Murcia, Spain

⁴King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

⁵Integrative Genomics and Bioinformatics Core, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

Abstract

CpG islands (CGIs) are primarily promoter-associated genomic regions and are mostly unmethylated within highly methylated mammalian genomes. The mechanisms by which CGIs are protected from de novo methylation remain elusive. Here we show that insertion of CpG-free DNA into targeted CGIs induces de novo methylation of the entire CGI in human pluripotent stem cells (PSCs). The methylation status is stably maintained even after CpG-free DNA removal, extensive passaging, and differentiation. By targeting the DNA mismatch repair gene *MLH1* CGI, we could generate a PSC model of a cancer-related epimutation. Furthermore, we successfully corrected aberrant imprinting in induced PSCs derived from an Angelman syndrome patient. Our results provide insights into how CpG-free DNA induces de novo CGI methylation and broaden the application of targeted epigenome editing for a better understanding of human development and disease.

[†]Corresponding author. belmonte@salk.edu.

*Present address: Department of Pharmacology (C128), The Genomics Institute of the Novartis Foundation (GNF), 10675 John Jay Hopkins Drive, La Jolla, CA 92037, USA.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/356/6337/503/suppl/DC1

Materials and Methods

Figs. S1 to S18

Table S1

Genome-wide single-base-resolution mammalian methylome studies have revealed that most of the cytosine in CpG dinucleotides is methylated (1–3), presumably during blastocyst to early postimplantation stages by DNA methyltransferase 3A (DNMT3A)– and/or DNMT3B-mediated de novo methylation (4, 5). Once established, the methylation patterns are stably maintained throughout development (2). In contrast, CpG islands (CGIs) remain predominantly unmethylated (3, 5, 6). About 70% of annotated gene promoters are associated with a CGI, making this the most prominent promoter feature (7). Recent reports suggest the existence of protection machinery, composed of proteins harboring the CXXC (Cys-X-X-Cys, where X is any amino acid) zinc finger domain, that safeguards the CpG clusters within a CGI from de novo methylation (8–10). The protein ten-eleven translocation 1 (TET1), which catalyzes the conversion of 5-methylcytosine into 5-hydroxymethylcytosine (11, 12), contains a CXXC domain and localizes to CGIs (8). In addition, CXXC finger protein 1 (CFP1) recruits the Set1 complex to CGIs (10), where this complex catalyzes methylation of histone H3 on lysine 4 (H3K4). Trimethylated H3K4 (H3K4me3) is generally associated with active transcription and potentially inhibits DNA methylation by preventing the recruitment of de novo DNA methyltransferases (13). It has been shown that artificial CpG-rich sequences lacking promoters are sufficient to establish domains of H3K4me3 (14). Nevertheless, it has been suggested that the methylation-blocking machinery alone is insufficient to establish and stably maintain an unmethylated CGI (14). The binding of transcription factors and transcripts themselves likely also plays a part (15, 16). However, it remains unknown how the methylation-blocking machinery cooperates with the active transcription process to establish and maintain the unmethylated status of CGIs.

Targeted CpG-free DNA integration induces stable de novo CGI methylation in human embryonic stem cells (hESCs)

To determine whether the interruption of CpG-rich sequences affects the unmethylated status of CGIs, we introduced CpG-free DNA into targeted CGIs. The CpG-free DNA consisted of a transgene cassette with a promoter and neomycin (neo) resistance gene that were modified to remove all CpG sequences. We first focused on the CGI within the promoter of the DNA mismatch-repair gene *MLH1*. Although *MLH1* CGIs of both alleles are normally unmethylated, heritable DNA methylation of even one allele of *MLH1* CGI, a phenomenon termed epimutation, increases the risk of colorectal cancer (17, 18). We constructed a targeting vector containing 1-kb homology arms on both sides (5′ and 3′) of the CpG-free cassette and delivered it into hESCs together with a plasmid expressing Cas9 nuclease and a single guide RNA (sgRNA) targeting the *MLH1* CGI via electroporation (Fig. 1A). After geneticin (G418) selection, heterozygous homologous recombination (HR) clones, with correct integration of the CpG-free cassette into one allele of the *MLH1* CGI, were obtained (fig. S1, A to C). To analyze the methylation status of CpG sites in the *MLH1* CGI, we performed a Southern blot analysis using the methylation-sensitive restriction enzyme Hae II. Both alleles in untransfected hESCs displayed an unmethylated status. For the heterozygous HR clones, nontargeted alleles were unmethylated, as evidenced by complete digestion by Hae II, whereas alleles containing the CpG-free cassette were methylated (fig. S1C). To further examine the extent of the de novo DNA methylation on the *MLH1* CGI, we

performed target enrichment–genome bisulfite sequencing, which can reveal CpG methylation status at single-base resolution across an 800-kb region around the *MLH1* CGI. In untransfected hESCs, almost the entire *MLH1* CGI was unmethylated, whereas de novo methylation was observed in the *MLH1* CGI of the heterozygous HR clones (Fig. 1B and fig. S2), resulting in ~50% of a colorectal adenocarcinoma cell line (SW48) with homozygous epimutation (fig. S3). Moreover, consistent with Southern blot analysis, allele-specific methylation analysis demonstrated that nontargeted alleles were unmethylated, whereas all alleles containing the CpG-free cassette were methylated (fig. S4). De novo DNA methylation of the *MLH1* CGI did not affect the methylation status of adjacent CGIs (fig. S2). These data indicate that targeted integration of the CpG-free cassette leads to de novo methylation of the entire *MLH1* CGI.

Next, we examined the stability of the de novo CGI methylation conferred by CpG-free DNA integration. We found that Cre-loxP–mediated removal of the CpG-free cassette did not affect the acquired methylation. More than 20 passages after excision, the *MLH1* CGI methylation level, as well as the low expression of *MLH1* mRNA and protein, was stably maintained (Fig. 1B and fig. S1, C to E). In contrast to hESCs, after CpG-free DNA integration, we did not observe de novo methylation of the *MLH1* CGI in human fibroblast cells (HFCs), mesenchymal stem cells (MSCs), and HeLa cells (fig. S5, A to D), likely as a result of the higher level of DNMT3B in hESCs (fig. S5J) (19). Taken together, these data show that integration of the CpG-free cassette through HR induced stable de novo methylation throughout the *MLH1* CGI in hESCs.

The *MLH1* CGI is likely more prone to epigenetic perturbation than other CGIs, as abnormal methylation of this locus has been observed in certain colon cancers. To test whether integration of CpG-free DNA could induce de novo methylation in other CGIs, we targeted the CGIs associated with housekeeping genes *HSP90AB1* and *AARS2*, which encode a heat shock protein and an alanyl-tRNA synthetase, respectively. Consistent with the results from the *MLH1* CGI, we also observed de novo methylation in HR alleles induced by CpG-free cassette integration in both the *HSP90* and *AARS2* CGIs (fig. S6).

Epigenetic correction of abnormal imprinting of *SNURF/SNRPN* CGI in Angelman syndrome induced PSCs (iPSCs)

To explore the possibility of using targeted CGI methylation to develop therapeutics for epigenetic disorders, we focused on the imprinting disease Angelman syndrome (AS), which is characterized by severe intellectual disability, speech impairment, and frequent laughter. AS is mainly caused by genetic alterations at the imprinting locus on chromosome 15q11-q13, which result in neuron-specific depletion of ubiquitin-protein ligase E3A (UBE3A) (20, 21). The CGI of the *SNURF/SNRPN* gene promoter, located ~400 kb upstream of *UBE3A*, normally exhibits allele-specific methylation on the maternal allele; the paternal allele is unmethylated and its transcription generates a *UBE3A* antisense RNA (a long noncoding RNA) that silences paternal *UBE3A* expression. Thus, UBE3A is normally expressed only from the maternal allele (22–24). One cause of AS is chromosome 15 paternal uniparental disomy (UPD), in which both copies of chromosome 15 are of paternal origin. In neurons

from these individuals, *UBE3A* antisense RNA is expressed from both alleles, thereby depleting UBE3A (Fig. 2A) (25).

To determine whether targeted CGI methylation can correct abnormal imprinting of paternal UPD, we generated iPSC lines using B lymphocytes derived from an AS patient with chromosome 15 paternal UPD and a control normal, or wild-type (WT), individual. No obvious differences in iPSC characteristics were observed between WT and AS iPSCs (fig. S7). It has been reported that, unlike some imprinting loci such as *H19*, *IGF2*, and *MEG3*, the *SNURF/SNRPN* CGI is highly stable and insensitive to perturbation in human PSCs (26–28). Consistent with this, we also found that the methylation patterns of the *SNURF/SNRPN* CGI in both WT and AS iPSCs were not affected by the reprogramming process and extensive passaging (Fig. 2C and fig. S8C). We next introduced the CpG-free cassette into the *SNURF/SNRPN* CGI of AS iPSCs (Fig. 2B). Correctly targeted AS iPSC HR clones were identified (fig. S8, A to C). Southern blot analysis using the methylation-sensitive restriction enzyme Bss HII revealed that nontargeted alleles remained unmethylated, whereas all eight targeted alleles found in six HR clones (four heterozygous, AS HR1-4; two homozygous, AS HR5 and HR6) exhibited de novo methylation (fig. S8C). Target enrichment–genome bisulfite sequencing showed that the entire *SNURF/SNRPN* CGI was unmethylated in AS iPSCs and became fully methylated in a homozygous HR clone (HR5) and that the methylation levels were restored to the WT level (~50%) in heterozygous HR clones (HR1 and HR2) (Fig. 2C). Allele-specific methylation analysis confirmed that de novo methylation only occurred on the targeted alleles (fig. S9). In AS iPSCs, aberrant DNA methylation patterns were also observed in distant genomic locations upstream of the target site (fig. S10). Notably, aberrant methylation at these distal regions (100 kb upstream) was also corrected in HR clones. In contrast, the methylation pattern of the *UBE3A* CGI, which lies downstream of the targeted CGI, was unaffected (fig. S10). These data support the notion that the *SNURF/SNRPN* CGI may serve as an imprinting center to regulate the imprinted expression of many genes from its upstream region (29). Taken together, our results demonstrate that de novo CGI methylation induced by integration of the CpG-free cassette is able to correct aberrant DNA methylation in AS iPSCs with paternal UPD.

To evaluate the stability of the de novo methylation at the *SNURF/SNRPN* CGI, we removed the integrated CpG-free cassette (neo-ex clones). These heterozygous neo-ex clones showed the same level of methylation as WT iPSCs, which remained unchanged even after 30 passages (Fig. 2C and fig. S8D). In addition, we found that *SNURF/SNRPN* mRNA levels were reduced to half in neo-ex heterozygous HR clones and completely repressed in a neo-ex homozygous HR clone (Fig. 2D). Finally, in AS iPSCs, the *SNURF/SNRPN* CGI was typically bound by abnormally high levels of CXXC proteins (TET1 and CFP1) and exhibited elevated H3K4me3 levels, which were restored to WT levels in the heterozygous AS HR2 neo-ex clone (Fig. 2, E and F).

Correction of abnormal imprinting restores UBE3A expression in AS neurons

Several lines of evidence suggest that AS symptomatology is associated with alterations in a variety of neuronal types, including CA1 pyramidal neurons (30), Purkinje neurons (31, 32), and dopaminergic (DA) neurons (33). To determine whether the induced de novo methylation of the *SNURF/SNRPN* CGI in AS iPSCs restores UBE3A expression in affected neurons, we differentiated iPSC lines into DA neurons. Neuronal cultures derived from different iPSC lines were stained for tyrosine hydroxylase (TH) and neuron-specific class III β -tubulin (TUB1) to label DA neurons. Consistent with a previous report (28), no obvious differences in differentiation ability and derived DA neuron morphology were observed between AS iPSCs (AS DA) and WT iPSCs (WT DA), as well as the neo-ex HR iPSCs (HR neo-ex DA) (Fig. 3A and fig. S11). Notably, the DNA methylation pattern was stably maintained after differentiation of the neo-ex HR clones (Fig. 3B). Quantitative reverse transcription polymerase chain reaction (RT-PCR) analysis revealed that the elevated expression of the *UBE3A* antisense RNA found in AS DA neurons was reduced to WT levels in the HR neo-ex DA neurons (Fig. 3C). Moreover, UBE3A protein expression, which is absent in AS DA neurons, was restored to WT levels in both HR neo-ex DA neuronal cultures (Fig. 3D). We also determined the levels of α -synuclein protein, which is subjected to UBE3A-mediated degradation (34), in differentiated DA neurons. Consistently, elevated levels of α -synuclein protein, typically found in AS DA neurons, were reduced to WT levels in HR neo-ex DA neurons (Fig. 3E). Thus, these data indicate that correction of aberrant imprinting methylation by integration of a CpG-free cassette remains stable upon differentiation and restores UBE3A protein levels in AS DA neurons by silencing the transcription of the *UBE3A* antisense RNA.

CpG-free content is obligatory for inducing de novo methylation

Next, we sought to determine the mechanism underlying CpG-free cassette-induced de novo CGI methylation. First, we ruled out the possibility that a DNA double-strand break (fig. S12) or HR process alone (fig. S13) could induce de novo methylation. Further, HR clones, in which CpG-containing DNA was integrated into targeted CGIs, did not exhibit de novo methylation of the *SNURF/SNRPN* CGI and *MLH1* CGI (fig. S14). To determine the minimal requirements that induce de novo methylation, we generated HR clones using targeting vectors containing a series of truncated CpG-free cassettes without drug selection (Fig. 4A). We first generated a targeting vector containing the neo-S/MAR fragment, including the CpG-free neomycin resistance gene, poly-adenylation signal (pAn), and scaffold/matrix-associated region (S/MAR), with a partial deletion of the N terminus from the full-length CpG-free cassette. Similar to what was observed with the full-length cassette, the integration of neo-S/MAR also induced de novo methylation at the *SNURF/SNRPN* CGI in AS iPSCs (Fig. 4B and fig. S15A). Next, we found that de novo methylation was induced by the integration of neo- β -globin (bGlo) pAn and S/MAR, but not by sv40 pAn (Fig. 4, B and D, and fig. S15, A and C). Unlike the whole neo-bGlo pAn, its three fragments did not induce de novo methylation (Fig. 4C and fig. S15B). Furthermore, we generated a scrambled neo-bGlo pAn fragment devoid of CpG sequence (scrambled neo)

and found that the scrambled neo also induced de novo methylation (Fig. 4E). Taken together, these results suggest that the CpG-free content, rather than specific sequence(s), of integrated DNA is important to induce de novo CGI methylation. To further confirm that CpG-free content is essential to induce de novo methylation, we introduced different numbers of CpG sequences in S/MAR (S/MAR +CpG10 and +CpG20). Although the CpG-free S/MAR fragment induced de novo methylation, the integration of the S/MAR fragment with CpG sequences did not induce de novo methylation (Fig. 4D and fig. S15C). CGIs are typically associated with high GC content. We synthesized a CpG-free scrambled CGI, with ± 600 -bp sequences centered on the insertion site of the *SNURF/SNRPN* CGI, which was scrambled and stripped of all CpG sequences while maintaining the same GC rate. We found that de novo CGI methylation also occurred after targeted integration of the scrambled CGI (Fig. 4F). Taken together, our results demonstrate that the integrated DNA fragment needs to be CpG-free to induce de novo methylation in targeted CGIs.

Both interruption of the methylation-blocking machinery and repression of transcriptional activity are required for de novo CGI methylation

Transcripts (15) and transcription factors (16) also play important roles in ensuring that CGIs are unmethylated. To examine whether transcription may affect the methylation status of the *SNURF/SNRPN* CGI, we established cell lines in which the *SNURF/SNRPN* promoter region, including relevant transcription factor binding sites, was deleted from one allele in AS iPSCs (Fig. 5A and fig. S16A). For the allele lacking the *SNURF/SNRPN* promoter, *SNURF/SNRPN* transcription was repressed (fig. S16B), and de novo methylation of the *SNURF/SNRPN* CGI was not observed (Fig. 5B). This suggests that repression of transcription alone is not sufficient to induce de novo CGI methylation.

To assess the importance of the integration site in the induction of de novo methylation, we introduced the same CpG-free cassette into different sites within the *SNURF/SNRPN* CGI. When the CpG-free cassette was integrated near the transcription starting site (TSS) (site 2) (Fig. 5A), de novo DNA methylation was induced in the entire CGI (Fig. 5C and fig. S17A), similar to that observed with site 1 (Fig. 2). Notably, integration into site 3, which is located 530 bp downstream of the TSS, successfully induced de novo methylation downstream of the insertion site, whereas only marginal de novo methylation was observed in the upstream region, including the *SNURF/SNRPN* promoter and exon 1 (Fig. 5C and fig. S17A). This may be attributed to the open chromatin structure upstream of site 3 integration maintained by an active transcriptional activity (fig. S17, B and C). Thus, these results suggest that both transcriptional repression and CpG-free DNA integration are essential to induce de novo CGI methylation.

To investigate whether disruption of the methylation-blocking machinery is the cause or consequence of the observed de novo CGI methylation, we generated AS iPSCs lacking the de novo methyltransferase DNMT3B (fig. S18A). In DNMT3B knockout AS iPSCs, de novo methylation of the *SNURF/SNRPN* CGI was not observed (fig. S18, B and C), even after the introduction of CpG-free DNA into site 1. Nonetheless, integration of the CpG-free cassette into this CGI repressed *SNURF/SNRPN* transcription (fig. S18D) and led to a pronounced

reduction in H3K4me3 levels on the CGI (fig. S18E). Collectively, these results indicate that CpG-free DNA integrated near the TSS interrupts the methylation-blocking machinery and represses transcriptional activity. To further investigate whether both are required for de novo methylation, we inserted the CpG-free cassette into site 3 of the allele with the *SNURF/SNRPN* promoter region removed. As described above, neither promoter deletion nor site 3 integration of the CpG-free cassette alone could induce de novo methylation in the entire CGI. However, when combined, de novo methylation was induced in the whole *SNURF/SNRPN* CGI (Fig. 5C). Taken together, both interruption of the methylation-blocking machinery and repression of transcriptional activity are necessary and sufficient to trigger de novo CGI methylation in PSCs.

In conclusion, here we demonstrate that targeted integration of CpG-free DNA in human PSCs induces de novo CGI methylation that remains stable after extensive passaging and differentiation. Our approach has enabled the establishment of a PSC epimutation model and allows for the correction of aberrant DNA methylation associated with an imprinting disease. Our results offer insights into the mechanism by which CGIs are protected from de novo methylation in PSCs and open new avenues for epigenome editing. Collectively, the observations reported here may facilitate a better understanding of the fundamental principles underlying DNA methylation and deepen our knowledge of how epigenetic dysregulation leads to alterations in human physiology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank M. C. Ku and T. Nguyen for next-generation sequencing; I. Dubova and K. McIntyre for teratoma analysis; D. O'Keefe for critical reading of the manuscript; C. J. Chang, A. Goebel, E. Aizawa, N. Y. Kim, Y. Hishida, R. D. Soligalla, and M. Morales Valencia for experimental help; M. Schwarz and P. Schwarz for administrative help; and A. Fukamizu for scientific discussion. Y.T. was partially supported by Astellas Foundation, Uehara Memorial Foundation, and Mochida Memorial Foundation research fellowships. K.S. and M.L. were supported by a California Institute for Regenerative Medicine (CIRM) Training Grant fellowship. P.M.-R. was partially supported by the Fundación Alfonso Martín Escudero. This work was supported by the Razavi Newman Integrative Genomics and Bioinformatics Core Facility and the Next Generation Sequencing (NGS) Core Facility at the Salk Institute, with funding from NIH–National Cancer Institute (NCI) Cancer Center Support Grant (CCSG) no. P30 014195, the Chapman Foundation, and The Leona M. and Harry B. Helmsley Charitable Trust. Work in the laboratory of J.C.I.B. was supported by UCAM (iPSC work), The G. Harold and Leila Y. Mathers Charitable Foundation, The Moxie Foundation, and The Leona M. and Harry B. Helmsley Charitable Trust.

REFERENCES AND NOTES

1. Lister R, et al. *Nature*. 2009; 462:315–322. [PubMed: 19829295]
2. Lister R, et al. *Science*. 2013; 341:1237905. [PubMed: 23828890]
3. Gifford CA, et al. *Cell*. 2013; 153:1149–1163. [PubMed: 23664763]
4. Okano M, Bell DW, Haber DA, Li E. *Cell*. 1999; 99:247–257. [PubMed: 10555141]
5. Guo H, et al. *Nature*. 2014; 511:606–610. [PubMed: 25079557]
6. Xie W, et al. *Cell*. 2013; 153:1134–1148. [PubMed: 23664764]
7. Saxonov S, Berg P, Brutlag DL. *Proc Natl Acad Sci USA*. 2006; 103:1412–1417. [PubMed: 16432200]
8. Wu H, et al. *Nature*. 2011; 473:389–393. [PubMed: 21451524]

9. Dawlaty MM, et al. *Dev Cell*. 2014; 29:102–111. [PubMed: 24735881]
10. Clouaire T, et al. *Genes Dev*. 2012; 26:1714–1728. [PubMed: 22855832]
11. Tahiliani M, et al. *Science*. 2009; 324:930–935. [PubMed: 19372391]
12. Ito S, et al. *Nature*. 2010; 466:1129–1133. [PubMed: 20639862]
13. Ooi SKT, et al. *Nature*. 2007; 448:714–717. [PubMed: 17687327]
14. Thomson JP, et al. *Nature*. 2010; 464:1082–1086. [PubMed: 20393567]
15. Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. *Mol Cell*. 2012; 45:814–825. [PubMed: 22387027]
16. Brandeis M, et al. *Nature*. 1994; 371:435–438. [PubMed: 8090226]
17. Weisenberger DJ, et al. *Nat Genet*. 2006; 38:787–793. [PubMed: 16804544]
18. Hitchins M, et al. *Gastroenterology*. 2005; 129:1392–1399. [PubMed: 16285940]
19. Liao J, et al. *Nat Genet*. 2015; 47:469–478. [PubMed: 25822089]
20. Lossie AC, et al. *J Med Genet*. 2001; 38:834–845. [PubMed: 11748306]
21. Kishino T, Lalande M, Wagstaff J. *Nat Genet*. 1997; 15:70–73. [PubMed: 8988171]
22. Rougeulle C, Cardoso C, Fontés M, Colleaux L, Lalande M. *Nat Genet*. 1998; 19:15–16. [PubMed: 9590281]
23. Herzing LBK, Cook EH Jr, Ledbetter DH. *Hum Mol Genet*. 2002; 11:1707–1718. [PubMed: 12095913]
24. Runte M, et al. *Hum Genet*. 2004; 114:553–561. [PubMed: 15014980]
25. Malcolm S, et al. *Lancet*. 1991; 337:694–697. [PubMed: 1672177]
26. Rugg-Gunn PJ, Ferguson-Smith AC, Pedersen RA. *Nat Genet*. 2005; 37:585–587. [PubMed: 15864307]
27. Rugg-Gunn PJ, Ferguson-Smith AC, Pedersen RA. *Hum Mol Genet*. 2007; 16:R243–R251. [PubMed: 17911167]
28. Chamberlain SJ, et al. *Proc Natl Acad Sci USA*. 2010; 107:17668–17673. [PubMed: 20876107]
29. Cassidy SB, Schwartz S, Miller JL, Driscoll DJ. *Genet Med*. 2012; 14:10–26. [PubMed: 22237428]
30. Jiang YH, et al. *Neuron*. 1998; 21:799–811. [PubMed: 9808466]
31. Mercer AA, Palarz KJ, Tabatadze N, Woolley CS, Raman IM. *eLife*. 2016; 5:230.
32. Egawa K, et al. *Sci Transl Med*. 2012; 4:163ra157.
33. Harbord M. *J Clin Neurosci*. 2001; 8:421–422. [PubMed: 11535008]
34. Mulherkar SA, Sharma J, Jana NR. *J Neurochem*. 2009; 110:1955–1964. [PubMed: 19645749]

Fig. 1. De novo *MLH1* CGI methylation induced by targeted integration of CpG-free cassette in hESCs

(A) Schematic of HR with the donor CpG-free vector targeting the *MLH1* gene locus. Open box containing “1” represents exon 1. Green boxes represent the CGI defined by the University of California, Santa Cruz (UCSC) Genome Browser. (B) Target enrichment–genome bisulfite sequencing analysis showing methylation status of *MLH1* CGI in hESCs. One vertical bar represents methylation status at a single CpG. Me+ and Me– represent methylated and unmethylated status, respectively. P20, passage 20; *EPM2AIP1*, gene encoding EPM2A-interacting protein 1. See also fig. S2, showing the whole region.

Fig. 2. Targeted integration of the CpG-free cassette corrects the aberrant DNA methylation in AS iPSCs

(A) Schematic of allele-specific methylation status and gene expression on the imprinting locus chr15q11-q12 in neurons of a normal individual and an AS patient with paternal UPD. P and M represent paternal and maternal allele, respectively. Filled and open circles represent methylated and unmethylated CpGs on the CGI, respectively. The dashed line represents a long noncoding RNA. (B) Schematic of HR with CpG-free donor vector targeting the *SNURF/SNRPN* gene locus. TALEN, transcription activator-like effector nuclease. (C) Target enrichment-genome bisulfite sequencing analysis showing methylation status of *SNURF/SNRPN* locus in iPSCs. See also fig. S10, showing the whole region. (D) Quantitative RT-PCR analysis of *SNURF/SNRPN* mRNA expression normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA in iPSCs. Error bars indicate $\pm$ SEM of three independent experiments. Statistical significance was determined by one-way

analysis of variance (ANOVA) followed by Bonferroni's multiple-comparison test (versus AS iPSCs; ** $P < 0.001$; *** $P < 0.0001$). (E) Chromatin immunoprecipitation (ChIP)–sequencing data for H3K4me3 across *SNURF/SNRPN* locus in iPSCs. (F) ChIP–quantitative PCR (qPCR) analysis of CFP1 and TET1 in iPSCs using primer sets targeting de novo–methylated locus inside (primer sets 1 to 3) or outside (primer set 4) the CGI shown in (E). Error bars indicate $\pm$ SEM from triplicate experiments.

Fig. 3. UBE3A expression is recovered in neurons derived from corrected AS iPSCs
(A) Representative immunofluorescence images of neurons derived from WT, AS, and AS neo-ex HR iPSCs for neuron-specific marker class III β -tubulin (TUJ1, red) and the dopaminergic neuron-specific marker tyrosine hydroxylase (TH, green). Nuclei were counterstained blue with 4',6-diamidino-2-phenylindole (DAPI). Scale bars, 50 μ m. **(B)** Bisulfite sequencing analysis of *SNURF/SNRPN* CGI in DA neurons. The primers for bisulfite PCR are shown in fig. S8A. Percentage of DNA methylation and *P* value (versus WT and versus AS) are shown at the bottom. Filled and open circles represent methylated and unmethylated CpGs, respectively. Statistical significance was determined by unpaired Student's *t* test. **(C)** Quantitative RT-PCR analysis of *UBE3A* antisense RNA expression normalized to GAPDH mRNA in DA neurons. Error bars indicate $\pm$ SEM of three independent experiments. Statistical significance was determined by one-way ANOVA followed by Bonferroni's multiple-comparison test (versus AS DA, $**P < 0.005$). **(D and E)** Western blot analysis of total cell lysates from iPSCs and DA neurons with anti-UBE3A monoclonal and polyclonal antibodies, an anti- β -actin antibody (D), and an anti- α -synuclein antibody (E).

Fig. 4. CpG-free content of the integrated DNA is essential to induce de novo CGI methylation (A) Schematic of HR with a series of truncated CpG-free cassettes targeting the *SNURF/SNRPN* locus and the outcomes of de novo-methylation induction after integration of each fragment. Filled and open circles represent methylated and unmethylated CGIs, respectively. In S/MAR +CpG fragments, the dashed vertical lines represent the added CpG dinucleotides. CMV enh, human cytomegalovirus enhancer; EF1 prom, elongation factor 1 promoter. (B to F) Bisulfite sequencing analysis of the HR clones. The primers for bisulfite PCR are shown in fig. S8A.

Fig. 5. Both integration of CpG-free DNA and repression of transcription are required and sufficient to trigger de novo CGI methylation

(A) Schematic of CpG-free cassette–integration and promoter-deletion locations in the *SNURF/SNRPN* gene locus. The primers for bisulfite PCR of promoter-deleted clones are shown as filled arrowheads. F, forward; R, reverse. (B) Bisulfite sequencing analysis of the *SNURF/SNRPN* CGI in promoter-deleted clones. The primers for bisulfite PCR are shown in Fig. 5A. (C) Target enrichment–genome bisulfite sequencing analysis showing methylation status of the *SNURF/SNRPN* CGI in iPSCs. Arrows point to CpG-free cassette integration sites. Black bars with white letters a to c indicate the regions of ChIP-qPCR shown in fig. S17C.