DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease

Danuta M. Jeziorska[a], Robert J. S. Murray[b,c,1], Marco De Gobbi[a,2], Ricarda Gaentzsch[a,3], David Garrick[a,4], Helena Ayyub[a], Taiping Chen[d], En Li[e], Jelena Telenius[a], Magnus Lynch[a,5], Bryony Graham[a], Andrew J. H. Smith[a,f], Jonathan N. Lund[c], Jim R. Hughes[a], Douglas R. Higgs[a,6], and Cristina Tufarelli[c,6]

[a]Medical Research Council Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Oxford University, Oxford OX3 9DS, United Kingdom; [b]Department of Genetics, University of Leicester, Leicester LE1 7RH, United Kingdom; [c]Division of Medical Sciences and Graduate Entry Medicine, School of Medicine, University of Nottingham, Royal Derby Hospital, Derby DE22 3DT, United Kingdom; [d]Department of Epigenetics and Molecular Carcinogenesis, Division of Basic Science Research, The University of Texas M. D. Anderson Cancer Center, Smithville, TX 78957; [e]China Novartis Institutes for BioMedical Research, Shanghai 201203, China; and [f]Medical Research Council Centre for Regenerative Medicine, The University of Edinburgh, Edinburgh EH16 4UU, United Kingdom

Edited by Adrian P. Bird, University of Edinburgh, Edinburgh, United Kingdom, and approved July 28, 2017 (received for review February 23, 2017)

The human genome contains ∼30,000 CpG islands (CGIs). While CGIs associated with promoters nearly always remain unmethylated, many of the ∼9,000 CGIs lying within gene bodies become methylated during development and differentiation. Both promoter and intragenic CGIs may also become abnormally methylated as a result of genome rearrangements and in malignancy. The epigenetic mechanisms by which some CGIs become methylated but others, in the same cell, remain unmethylated in these situations are poorly understood. Analyzing specific loci and using a genome-wide analysis, we show that transcription running across CGIs, associated with specific chromatin modifications, is required for DNA methyltransferase 3B (DNMT3B)-mediated DNA methylation of many naturally occurring intragenic CGIs. Importantly, we also show that a subgroup of intragenic CGIs is not sensitive to this process of transcription-mediated methylation and that this correlates with their individual intrinsic capacity to initiate transcription in vivo. We propose a general model of how transcription could act as a primary determinant of the patterns of CGI methylation in normal development and differentiation, and in human disease.

CGI transcription | DNA methylation | H3K36me3 | intragenic CGIs | CGI methylation

Methylation of CpG dinucleotides plays a pivotal role in mammalian development and differentiation (1). Most CpG dinucleotides are methylated, with the exception of those within CpG islands (CGIs), which are usually unmethylated (2, 3). Although several mechanisms have been proposed to explain how CGIs normally escape methylation (reviewed in ref. 1), they do not always remain unmethylated. In particular, while most promoter CGIs remain unmethylated, ∼9,000 CGIs within gene bodies (intragenic) are more likely to become methylated (4, 5).

Many intragenic CGIs become methylated during development, and this often coincides with transcription of the gene within which they lie; typical examples are CGIs in imprinted genes and at the X inactivation center (6, 7). In addition, aberrant methylation of CGIs has been reported in inherited and acquired genomic rearrangements when CGIs become abnormally located within the body of another, transcriptionally active gene (8–10). Such examples strongly suggest that aberrant or naturally occurring transcription passing through CGIs is, in some way, linked to their DNA methylation.

To address how transcription running across a CGI may determine whether or not it becomes methylated, we took an experimental approach to study in detail the mechanism by which a CGI located at the α-globin locus becomes methylated when incorporated into a newly formed transcriptional unit causing a human disease [α-thalassemia (10)] (Fig. 1). To relate these findings to endogenous intragenic CGIs, we also analyzed a previously described, naturally occurring CGI associated with the rhomboid 5 homolog 1 (*RHBDF1*), which becomes methylated during normal development in vivo (11) (Fig. 1).

In both cases, we show that methylation of these specific CGIs is associated with transcription traversing the CGI, a reduction of H3K4me3, a gain of H3K36me3, and DNA methyltransferase 3B (DNMT3B)-mediated DNA methylation. At these specific loci, we experimentally determined the order in which these events

Significance

The human genome contains ∼30,000 CpG islands (CGIs), long stretches (0.5–2 kb) of DNA with unusually elevated levels of CpG dinucleotides. Many occur at genes' promoters, and their DNA nearly always remains unmethylated. Conversely, intragenic CGIs are often, but not always, methylated, and thus inactive as internal promoters. The mechanisms underlying these contrasting patterns of CGI methylation are poorly understood. We show that methylation of intragenic CGIs is associated with transcription running across the island. Whether or not a particular intragenic CGI becomes methylated during development depends on its transcriptional activity relative to that of the gene within which it lies. Our findings explain how intragenic CGIs are epigenetically programmed in normal development and in human diseases, including malignancy.

Author contributions: D.M.J., A.J.H.S., J.R.H., D.R.H., and C.T. designed research; D.M.J., R.J.S.M., M.D.G., R.G., D.G., H.A., T.C., J.T., M.L., B.G., and C.T. performed research; T.C., E.L., M.L., A.J.H.S., J.N.L., J.R.H., D.R.H., and C.T. contributed new reagents/analytic tools; D.M.J., R.J.S.M., J.T., J.R.H., and C.T. analyzed data; D.M.J., D.R.H., and C.T. wrote the paper; and E.L., A.J.H.S., J.N.L., D.R.H., and C.T. provided supervision.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: RNA-sequencing, ChIP-sequencing, and 5′-anchored RNA profiling (CapSeq) datasets produced in the course of this study have been submitted to the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (superseries accession no. GSE84355).

[1] Present address: Human Development and Health, Institute of Developmental Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom.

[2] Present address: Dipartimento di Scienze Cliniche e Biologiche, Università degli Studi di Torino, 10043 Orbassano (Torino), Italy.

[3] Present address: Seven Bridges Genomics, London NW1 2RA, United Kingdom.

[4] Present address: INSERM, UMRS-1126, Institut Universitaire d'Hématologie, Université Paris, 75010 Paris, France.

[5] Present address: Centre for Stem Cells and Regenerative Medicine, King's College London, London WC2R 2LS, United Kingdom.

[6] To whom correspondence may be addressed. Email: cristina.tufarelli@nottingham.ac.uk or doug.higgs@imm.ox.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1703087114/-/DCSupplemental.

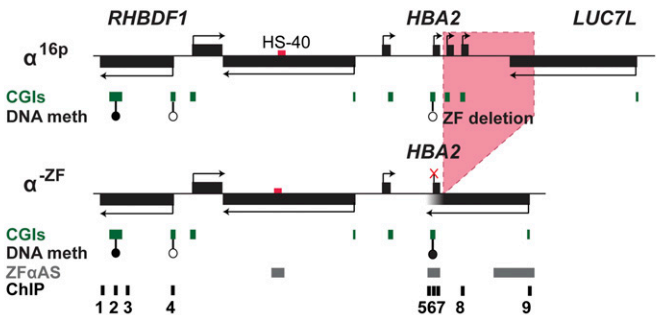

Fig. 1. Short arm of chromosome 16 (chr16) harbors exemplar loci to investigate the mechanism of DNA methylation within transcribed CGIs. Schematic representation of the human chr16 encompassing the *HBA* locus in a normal allele (α^{16p}) and in the ZF-deleted allele (α^{-ZF}). The hypo- and hypermethylated status of CGIs investigated in this study is marked by empty and filled lollipops, respectively. The positions of qPCR assays used for ChIP are shown as numbered black bars (1 = *RHBDF1* intron 17, 2 = *RHBDF1* intragenic CGI, 3 = *RHBDF1* intron 3, 4 = *RHBDF1* promoter CGI, 5 = 5′ *HBA,* 6 = *HBA* CGI, 7 = *HBA* exon 3, 8 = *LUC7L* gene body, 9 = *LUC7L* promoter). Gray rectangles indicate the fragments present in the ZFαAS construct.

lead to establishment of methylation and silencing of the transcribed CGIs. Importantly, we finally show that transcription-mediated DNA methylation is a general mechanism responsible for silencing naturally occurring intragenic CGIs throughout the genome. Interestingly, however, we find that not all CGIs become methylated and silenced when transcribed: CGIs that act as highly active initiators of transcription remain unmethylated. These findings indicate that, in contrast to what is seen at gene body regions, transcription traversing a CGI is not always sufficient to drive its methylation and silencing but also depends on the degree to which such a CGI initiates transcription itself.

Results

Silencing of a CGI Located Within a Newly Formed Transcriptional Unit Is Associated with Enrichment in H3K36me3 and H3K9me3. We have previously shown that silencing of *HBA2* expression on the rearranged allele of an individual (ZF) with α-thalassemia is caused by de novo DNA methylation of the HBA promoter CGI in the presence of antisense transcription from the truncated *LUC7L* gene (10) (Fig. 1). Here, we further characterized the associated epigenetic profile of the rearranged ZF allele. Analyzing EBV-transformed lymphocytes from ZF (EBV-ZF) and a normal individual (EBV-RN), we found no differences in the levels of H3K4me3 (Fig. 2*A*, *Left*). The enrichment for this mark was low at *HBA2*, consistent with the fact that this gene is not expressed in lymphocytes. By contrast, we found an increase in H3K9me3 in EBV-ZF compared with EBV-RN (Fig. 2*A*, *Center*). Given that the *HBA* genes in nonexpressing cells are normally silenced by the polycomb system (H3K27me3) and not by K9 methyltransferases (H3K9me3) (12), it seemed most likely that the enrichment of H3K9me3 only in EBV-ZF originates from the abnormal ZF chromosome. Most notably, we also found enrichment of H3K36me3 at the HBA-CGI and HBA-ex3 regions only in EBV-ZF cells (Fig. 2*A*, *Right*). These observations are consistent with previous data showing antisense transcription occurs across *HBA2* on the ZF chromosome but not on the normal copy of chromosome 16 (10).

To determine whether the enrichment of H3K9me3 and H3K36me3 was similarly restricted to the rearranged ZF allele, we analyzed two independently isolated mouse erythroleukemia (MEL) interspecific hybrid cell lines carrying either the normal (MELx16) or rearranged (MELxZF) copy of human chromosome 16 (13). In the erythroid-poised environment of MEL cells, in contrast to the normal hybrids (MELx16), we detected no enrichment of H3K4me3 in MELxZF hybrids (Fig. 2*B*, *Left*), consistent with the ZF chromosome carrying a silenced and DNA-methylated *HBA2* gene (13). However, in MELxZF hybrids (but not MELx16 hybrids), we found enrichment of H3K9me3 (Fig. 2*B*, *Center*) and H3K36me3 (Fig. 2*B*, *Right*) at the HBA-CGI and HBA-ex3, supporting the hypothesis that transcription through the CGI is associated with these chromatin modifications and silencing of the *HBA2* gene on the ZF chromosome.

Transcription Is Required to Establish Epigenetic Silencing of the *HBA2* CGI. Epigenetic silencing of the *HBA2* CGI observed in ZF is faithfully recapitulated in mouse models using a construct [ZFα-antisense (ZFαAS); Fig. 1] in which an α-globin enhancer/*HBA2* cassette is linked to a fragment spanning the *LUC7L* CGI, which, in turn, produces an RNA transcript that runs across the α-globin CGI. In this model, silencing of the α-globin CGI is established before the specification of the three germ layers both in vivo and during in vitro differentiation of mouse embryonic stem (mES) cells (10). Therefore, we used this experimental model to investigate the role of transcription in de novo DNA methylation of the CGI. As a control, we used mES clones harboring the ZFαS construct in which the *LUC7L* transcription is driven away from *HBA2* CGI rather than running across the CGI (10).

To determine if this system established chromatin signatures similar to those observed in the EBV-ZF and MELxZF cell lines, we established chromatin profiles in ZFαAS and ZFαS mES cells and in in vitro-differentiated embryoid bodies (EBs) at day 7. Despite transcription through the *HBA2* CGI in ZFαAS cells, we found no significant differences in the levels of H3K4me3 at the *HBA2* gene in undifferentiated mES cells; however, at day 7, H3K4me3 is enriched at the HBA-CGI and HBA-ex3 in ZFαS

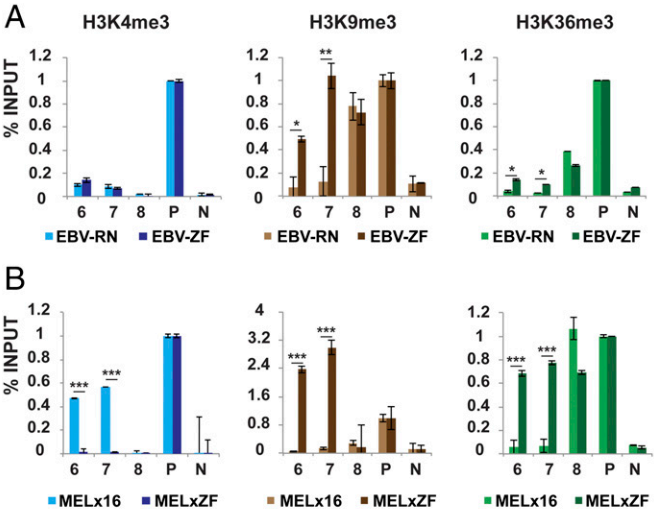

Fig. 2. *HBA2* promoter CGI on the ZF allele is associated with loss of H3K4me3 and enrichment in H3K9me3 and H3K36me3. ChIP was performed with antibodies to H3K4me3, H3K9me3, and H3K36me3 in EBV-transformed lymphocytes from an individual carrying two normal copies of chromosome 16 (EBV-RN) or ZF (EBV-ZF) (*A*), and in MEL hybrids carrying a normal copy (MELx16) or the ZF copy (MELxZF) of human chromosome 16 (*B*). Precipitated DNA fragments were used to amplify regions labeled 6, 7, and 8 in Fig. 1. P, positive control for EBV lymphocytes (*A*; human *ACTB* promoter for H3K4me3, human 18S for H3K9me3, human *ACTB* gene body for H3K36me3) and for MEL hybrids (*B*; mouse *Actb* promoter for H3K4me3, mouse 18S for H3K9me3, mouse *Ccna2* gene body sequence for H3K36me3); N, negative control for EBV lymphocytes (*A*, human *RHBDF1* promoter for all modifications) and for MEL hybrids (*B*, mouse intergenic region for H3K4me3 and H3K36me3; human *RHBDF1* promoter for H3K9me3). *P* values were calculated by *t* test: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

but not ZFαAS EBs (Fig. 3*A*). Consistent with the *HBA2* CGI having a bivalent chromatin signature in mES cells (14–16), we found that ZFαS clones maintained the H3K4me3 and H3K27me3 histone modifications at the *HBA2* CGI in mES cells (Fig. 3*A*, *Left* and *SI Appendix*, Fig. S1*A*, *Left*). By contrast, H3K27me3 was depleted at the *HBA2* CGI in ZFαAS clones, which showed significant enrichment for H3K9me3 upon differentiation (*SI Appendix*, Fig. S1 *A* and *B*). These data show that the loss of H3K4me3 in ZFαAS clones is not dependent on the polycomb (H3K27me3) pathway.

These contrasting profiles in ZFαAS and ZFαS strongly suggest that transcription traversing the α-globin CGI is the key event leading to DNA methylation and silencing. We therefore used a different approach to further test this hypothesis. We made mES cell clones carrying a construct in which the β-globin transcription terminator (17) prevents transcription from the *LUC7L* promoter running through the *HBA2* CGI (ZFαAS-STOP; *SI Appendix*, Fig. S2 *A* and *B*). In ZFαAS-STOP clones, with very low/undetectable transcription through the *HBA2* CGI, this CGI remained unmethylated and the *HBA2* gene was not silenced upon differentiation (*SI Appendix*, Fig. S2 *C* and *D*). Consistent with this, the *HBA2* CGI also remained modified by H3K4me3 rather than acquiring H3K9me3 (*SI Appendix*, Fig. S2*E*). These experiments thus provide complementary evidence supporting the role of transcription in this methylation-mediated silencing.

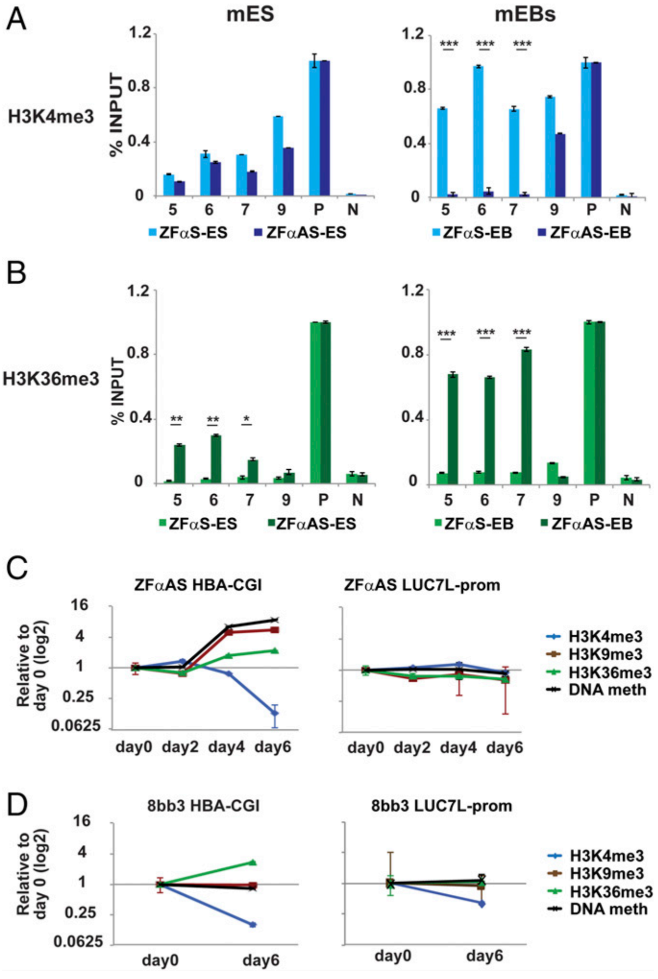

Fig. 3. Transcription-mediated deposition of H3K36me3 precedes recruitment of DNMT3B to *HBA2* in ZF. H3K4me3 (*A*) and H3K36me3 (*B*) ChIP performed in ES cells (*Left*) and EBs (*Right*) from ZFαS and ZFαAS cell lines is shown. Precipitated DNA fragments were used to amplify regions labeled 5, 6, 7, and 9 in Fig. 1. P, positive control (mouse *Ccna2* gene body); N, negative control (mouse intergenic region). *P* values were calculated by *t* test: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Changes in DNA methylation, H3K4me3, H3K36me3, and H3K9me3 during in vitro differentiation of ZFαAS (*C*) and 8bb3 (*D*) cells are shown. Percentages of DNA methylation at the *Eag*1 site within the *HBA2* and *LUC7L* promoter CGIs were calculated as the ratio of methylated to total signal by Southern blotting (*SI Appendix*, Fig. S4), while changes in relative enrichment of histone modifications at HBA-CGI and LUC7L-prom were measured by ChIP. Data points displayed for each modification are the $\log_2$ of the ratios relative to their respective day 0 value.

To prove that transcription running through the *HBA2* CGI, rather than the resulting RNA and its intermediates, is responsible for DNA methylation of the CGI, we performed Northern blots and small RNA-sequencing (RNA-Seq) in our system. We did not detect any *HBA2*-specific small RNAs in undifferentiated or differentiating ZFαAS cells (*SI Appendix*, Fig. S2 *F* and *G*). These observations, together with previous data (10), show that the epigenetic changes that lead to *HBA2* silencing in the ZF context rely on a mechanism that is primarily dependent on transcription *in cis* rather than the RNA transcript, per se, as seen in other systems (18).

Epigenetic Silencing of the *HBA2* CGI Occurs During Differentiation. A recent report showed that transcription-mediated deposition of H3K36me3 within gene bodies can form a docking site for the de novo DNMT3B, thus targeting DNA methylation to transcribed regions (19). This may also explain how transcription, per se, may be responsible for silencing of the *HBA2* CGI in ZF via DNA methylation. Consistent with this, we found significant enrichment of H3K36me3 at *HBA2*, particularly when ZFαAS clones were differentiated (Fig. 3*B*). However, in contrast to what was previously seen at gene body regions (19), the presence of H3K36me3 at the *HBA2* CGI in undifferentiated ZFαAS mES cells was not sufficient to mediate methylation of this CGI.

The inducible nature of the ZFαAS mES system enabled us to dissect the timing and order of events leading to this type of epigenetic silencing. We found that H3K4me3 decreases across the *HBA2* locus in ZFαAS between days 2 and 4, with the lowest enrichment seen at day 6 (*SI Appendix*, Fig. S3*A*, *Left*). By contrast, an increase in H3K9me3 and H3K36me3 was observed at all sites in ZFαAS between days 2 and 4 of differentiation (*SI Appendix*, Fig. S3*A*, *Center* and *Right*). This correlates with the acquisition of de novo DNA methylation at the *HBA2* CGI that also occurs between days 2 and 4 of differentiation (10) (*SI Appendix*, Fig. S4). A plot of the values relative to day 0 of the enrichment for the histone modifications as measured by ChIP and of the percentage of (%) methylation at the *Eag*1 site as measured by Southern blotting indicates that *HBA2* silencing is established between days 2 and 4 of ZFαAS differentiation and is associated with a loss of H3K4me3 and a gain of H3K36me3, H3K9me3, and DNA methylation at the abnormally transcribed *HBA2* CGI (Fig. 3*C*, *Left*). By contrast, no significant changes in any chromatin modifications were seen at the *LUC7L* promoter CGI (Fig. 3*C*, *Right*). It must be noted that we have combined in the plots enrichment-based ChIP at and surrounding the 60-bp HBA-CGI region amplified by PCR with the % DNA methylation at the 6-bp *Eag*1 recognition site, and this probably underlies the apparent paradoxical retention of H3K4me3 at the *HBA2* CGI at day 4 when DNA methylation appears to be near maximal (Fig. 3*C*, *Left*). The data from bisulfite sequencing suggest that although the *Eag*1 site is >65% methylated, the region that includes the *HBA2* transcription start site has the higher concentration of unmethylated or lowly methylated CpGs at day 4, consistent with possible H3K4me3 retention at this time point (*SI Appendix*, Fig. S4).

DNMT3B Is Required for Methylation of the *HBA2* CGI. To determine whether DNMT3B is the DNMT responsible for the deposition of DNA methylation at the *HBA2* CGI in ZF, we generated clones stably transfected with the ZFαAS construct in wild-type (J1-1) and Dnmt3b$^{-/-}$ (8bb) mES cells (20). To compare the epigenetic profiles in undifferentiated and differentiated mES cells, we used the cardiomyocyte differentiation system because mES cells lacking DNMTs are most easily differentiated to cardiomyocytes (21) and the dynamics of *HBA2* silencing in ZFαAS cells are maintained under these conditions (*SI Appendix*, Fig. S1 *C* and *D*). Consistent with the data from ZFαAS cells, the *HBA2* CGI became methylated upon differentiation of J1-1 cells (*SI Appendix*, Fig. S3*B*). In contrast to this, we did not detect methylation at HBA-CGI in Dnmt3b$^{-/-}$ cells carrying the ZFαAS construct (8bb2 and 8bb3 in *SI Appendix*, Fig. S3*B*) even though they differentiated normally (*SI Appendix*, Fig. S3*C*). This indicates that DNMT3B is the main methyltransferase responsible for methylation of the *HBA2* CGI in ZF, and is consistent with recently published data showing that DNMT3B methylates actively transcribed regions of the genome (19). We therefore further characterized the epigenetic profiles in Dnmt3b$^{-/-}$ cells and found that the *HBA2* CGI in undifferentiated Dnmt3b$^{-/-}$ cells was enriched in both H3K4me3 and H3K36me3. Upon EB differentiation, there was a decrease in H3K4me3 accompanied by an increase in H3K36me3; importantly, however, this occurred without establishment of DNA methylation at the *Eag*1 site in the *HBA2* CGI (*SI Appendix*, Fig. S3*B*), H3K9me3 at HBA-CGI, or any of the other *HBA2* sites tested (*SI Appendix*, Fig. S3*D*). These observations at the *HBA2* CGI in Dnmt3b$^{-/-}$ cells are highlighted by plotting the ES cell and EB values for enrichment in the different histone marks and % DNA methylation at the *Eag*1 site relative to the ES cell (Fig. 3*D*, *Left*). No significant changes in any chromatin modifications were seen at the *LUC7L* promoter CGI (Fig. 3*D*, *Right*).

These data support an order of events whereby transcription through the H3K4me3-enriched *HBA2* CGI is accompanied by deposition of H3K36me3. Upon differentiation, a decrease in the level of H3K4me3 and an increase in the level of H3K36me3 are followed by DNA methylation and deposition of H3K9me3 at the *HBA2* CGI, which seems to occur at the same time and requires DNMT3B, probably due to the two enzymatic activities being part of the same multiprotein complex (22).

Transcriptional Elongation Through CGIs Has Different Effects on Specific CGIs. To test whether transcription running through a CGI is a general mechanism for differentiation-induced CGI methylation and gene silencing, we analyzed three additional human genes associated with promoter CGIs: the muscle-specific myogenic differentiation 1 (*MYOD1*) gene and the ubiquitously expressed ubiquitin C (*UBC*) and actin B (*ACTB*) genes. As before, transcriptional elongation across each CGI was driven by the *LUC7L* promoter (Fig. 4 *A*, *C*, and *E*). Transcription through each of these genes' CGIs was detected in both mES cells and day 8 EBs (*SI Appendix*, Fig. S5*A*). Transcription traversing the CGIs associated with *MYOD1* and *UBC* was associated with loss of expression (*SI Appendix*, Fig. S5*B*), but, surprisingly, this was not the case for *ACTB* (*SI Appendix*, Fig. S5*B*). Upon differentiation into EBs, the *MYOD1* (*SI Appendix*, Figs. S5*C* and S6) and *UBC* (*SI Appendix*, Figs. S5*D* and S7) CGIs become methylated, with associated loss of H3K4me3 and enrichment of H3K36me3 and H3K9me3 (Fig. 4 *B* and *D*). By contrast, the *ACTB* CGI remained free of methylation (*SI Appendix*, Figs. S5*E* and S8) and showed increased levels of H3K4me3 and decreased levels of H3K36me3 (Fig. 4*F*). Therefore, we can conclude that transcription-mediated silencing is observed at the promoter CGIs of both tissue-specific (*MYOD*) and ubiquitously expressed (*UBC*) genes; however, not all CGIs are susceptible to silencing mediated by transcriptional elongation passing through these elements.

Epigenetic Silencing at a Naturally Occurring Intragenic CGI. We next addressed whether these experimental observations illustrate a general mechanism by which naturally occurring intragenic CGIs become methylated during differentiation and development in vivo.

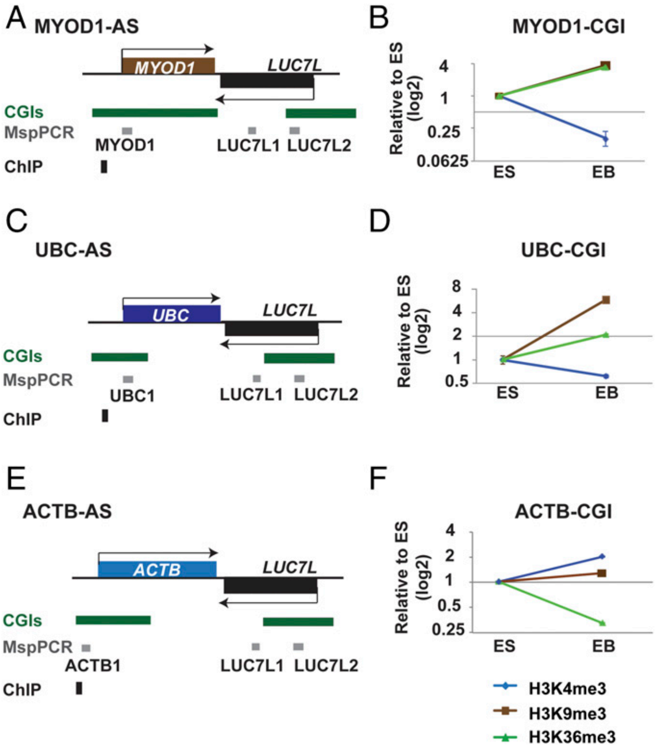

Fig. 4. *LUC7L*-driven transcription can epigenetically silence promoter CGI of the *MYOD* and *UBC* genes but not *ACTB* CGI. A schematic diagram of MYOD1-AS (*A*), UBC-AS (*B*), and ACTB-AS (*C*) constructs is shown, with the position of ChIP primers indicated. Changes in H3K4me3, H3K9me3, and H3K36me3 at MYOD-CGI (*D*), UBC-CGI (*E*), and ACTB-CGI (*F*) in ES cells and day 7 EBs are shown. Data displayed for each modification are the $\log_2$ of the ratios relative to their respective ES cell value.

We initially studied the *RHBDF1* gene, which is located at 16p13.3, ~100 kb from the *HBA* gene cluster (Fig. 1) and encodes a rhomboid protease-like protein (23, 24). *RHBDF1* has a promoter CGI that is unmethylated in all tissues and an intragenic CGI that we previously noted is methylated in all somatic cells (11). Importantly, this intragenic CGI is unmethylated in sperm, where *RHBDF1* is not transcribed, while it is methylated in human embryonic stem (hES) cells, where *RHBDF1* is expressed. Subsequently, in development, the intragenic CGI remains methylated in all somatic cells regardless of the level of *RHBDF1* RNA expression (Fig. 5*A*). This shows that transcription may be required to establish but not to maintain methylation of the *RHBDF1* intragenic CGI during development. This CGI therefore provides an experimental model to analyze intragenic CGI methylation in vivo.

To determine if transcription is required to establish DNA methylation at the *RHBDF1* intragenic CGI, we used a recombination-mediated cassette exchange (RMCE) system to target sequences of interest, with high efficiency, into a single defined genomic location within the mouse α-globin locus (25) (*SI Appendix*, Fig. S9*A*). This well-characterized locus represents a neutral chromatin environment in mES cells as it is transcriptionally silent and lacks both active (H3K4me3) and repressive (H3K27me3) histone marks (*SI Appendix*, Fig. S9*B*). This region also has no effect on the methylation status of integrated DNA (*SI Appendix*, Fig. S10). Therefore, this system allows us to investigate the contribution of the inserted sequences to their methylation status within an apparently "neutral" chromosomal environment.

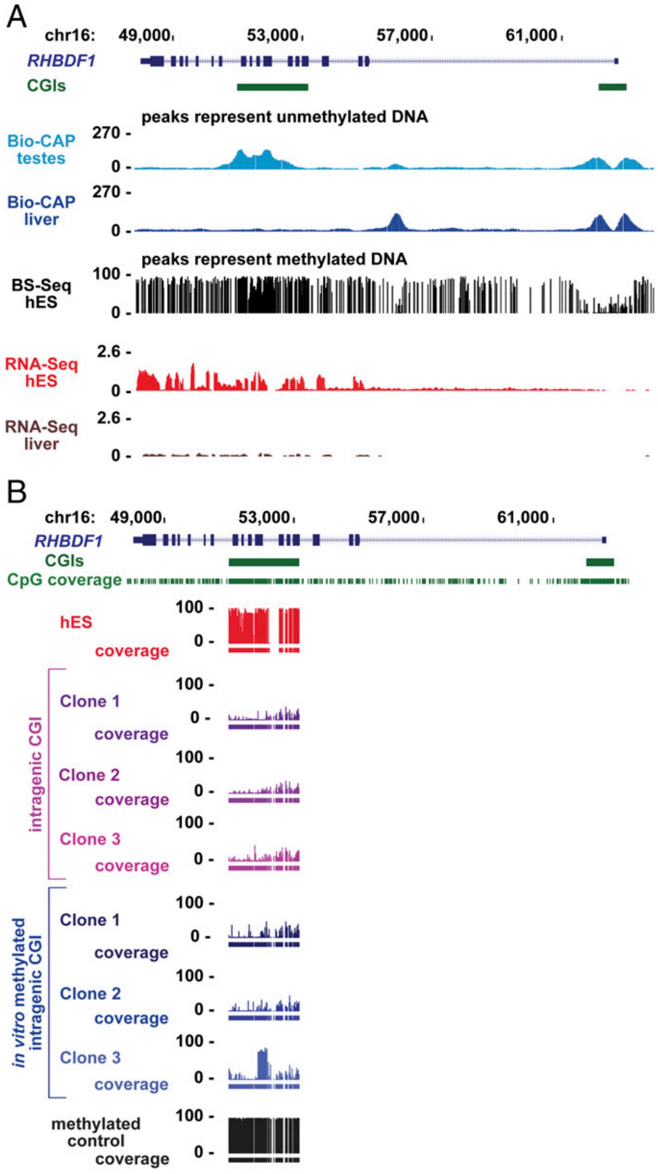

Fig. 5. *RHBDF1*'s intragenic CGI sequence is not sufficient to recapitulate its endogenous methylation status in the RMCE system in mES cells. (*A*) DNA methylation and expression of the human *RHBDF1* gene and the context of the intragenic CGI. The *RHBDF1* gene, located on human chromosome 16, has two CGIs: one at the promoter and one within the gene body. A snapshot of *RHBDF1* from the UCSC Genome Browser is shown, with plotted, publicly available methylation profiles by Bio-CAP-Seq from testes and liver (peaks represent unmethylated DNA), BS-Seq data from hES (peaks represent methylated DNA), and RNA-Seq data from hES and liver. (*B*) DNA methylation status of the intragenic CGI of *RHBDF1* was analyzed by sequencing of bisulfite-converted DNA from hES cells (track in red), three RMCE-derived mES clones with integrated unmethylated (tracks in purple) and in vitro-methylated (tracks in dark blue) sequences of human *RHBDF1* intragenic CGI, and methylated control genomic DNA (track in black). Individual CpG coverage in bisulfite analyses is indicated below each DNA methylation track.

Epigenetic Silencing of the *RHBDF1* Intragenic CGI Depends on Transcription from the Associated Promoter. To determine if the DNA sequence of the *RHBDF1* intragenic CGI alone is sufficient to drive DNA methylation independent of transcription and the normal chromosomal context of the gene, we first integrated a DNA fragment encompassing the *RHBDF1* intragenic CGI alone (Fig. 5) into the RMCE system (*SI Appendix*, Fig. S9*A*). Analysis of targeted clones showed that the intragenic CGI, on its own, remained unmethylated in the RMCE system (intragenic CGI clones 1, 2, and 3 in Fig. 5*B*). Moreover, in two of the three clones tested, this CGI became unmethylated even when its DNA sequence was methylated in vitro before integration (in vitro-methylated intragenic CGI clones 1 and 2 in Fig. 5*B*), consistent with previous studies showing that a high CpG density in mES cells, per se, can create stable nonmethylated CGIs (26, 27). Therefore, the *RHBDF1*'s intragenic CGI sequence alone is not sufficient to recapitulate its endogenous pattern of methylation, supporting the hypothesis that the intragenic context is critical to its methylation.

These findings suggested that transcription originating from the promoter of the *RHBDF1* gene may be required for DNA methylation. We therefore integrated the *RHBDF1* intragenic CGI in the context of the *RHBDF1* gene with (*RHBDF1+P*) or without (*RHBDF1−P*) its promoter into the RMCE system (*SI Appendix*, Fig. S11*A*). We confirmed that *RHBDF1* is expressed in *RHBDF1+P* but not in *RHBDF1−P* mES cells (*SI Appendix*, Fig. S11*B*). In undifferentiated cells, we found that the intragenic CGI is less heavily methylated in both *RHBDF1+P* and *RHBDF1−P* clones than at its normal endogenous locus in hES cells (Fig. 6). This may be simply explained by suboptimal expression of *RHBDF1* in this system. Nevertheless, we observed a small but significant difference in DNA methylation at the 3′ end of the CGI when comparing *RHBDF1+P* and *RHBDF1−P* clones in these undifferentiated cells.

When *RHBDF1+P* and *RHBDF1−P* clones were differentiated into day 7 EBs (Fig. 6), the intragenic *RHBDF1+P* CGI became

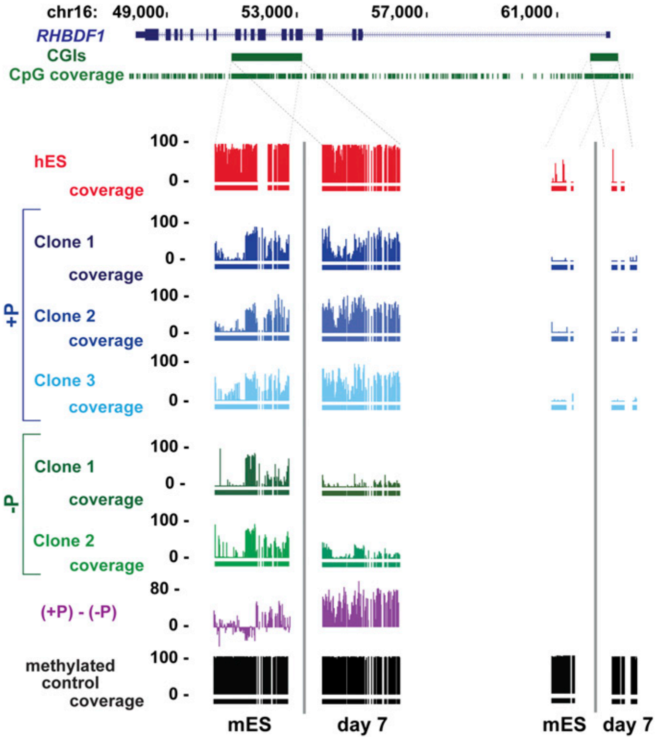

Fig. 6. Upon differentiation, the intragenic CGI becomes hypermethylated in the context of *RHBDF1+P*. The DNA methylation status of the intragenic and promoter CGIs of *RHBDF1* was analyzed by sequencing of bisulfite-converted DNA from hES cells (track in red), three RMCE-derived mES cell/EB day 7 clones with integrated *RHBDF1+P* (tracks in blue) and *RHBDF1−P* (tracks in green), and methylated control genomic DNA (track in black). Individual CpG coverage in bisulfite analyses is indicated below each DNA methylation track. The average DNA methylation difference of individual CpGs within the intragenic CGI between +P and −P RMCE clones is displayed in purple.

heavily methylated. By contrast, the *RHBDF1* intragenic CGI remained largely unmethylated in *RHBDF1−P* clones (Fig. 6). These data show that the context of the CGI within the *RHBDF1* gene alone is not sufficient to drive its methylation; rather, it is transcription from the *RHBDF1* promoter that is critically required. The increased CGI methylation in *RHBDF1+P* clones during in vitro differentiation is associated with significant increases in the level of *RHPDF1+P* nascent transcription (Fig. 7*A*), showing a correlation between level of transcription and methylation of the intragenic CGI.

Using ChIP in *RHBDF1+P* clones we found that the distribution of H3K36me3 within *RHBDF1* follows that typically observed at expressed genes: no enrichment at the transcription start site, with a gradual increase throughout the gene body and maximum levels close to the transcription termination site (Fig. 7*B*). Moreover, the level of H3K36me3 further increases upon in vitro differentiation (Fig. 7*B*). Taken together, these data are consistent with transcription through the intragenic CGI increasing H3K36me3 and consequently recruiting DNMT3B, which then promotes de novo DNA methylation.

The Intragenic CGI Acts as a Promoter in the Absence of a Transcriptionally Active Promoter of *RHBDF1*. Since the *RHBDF1* intragenic CGI is predominantly unmethylated in *RHBDF1−P* EBs, we investigated whether this is associated with changes in its epigenetic signature. We showed that the low level of DNA methylation in *RHBDF1−P* EBs is associated with a gain in the H3K4me3 modification characteristic of active promoters (*SI Appendix*, Fig. S12*A*). Of particular note, we observed an atypical distribution of H3K36me3 throughout the *RHBDF1* gene in *RHBDF1−P* compared with *RHBDF1+P* clones. Interestingly, in *RHBDF1−P*, there is now a small dip in enrichment of H3K36me3 at the intragenic CGI in contrast to the surrounding amplicons (*SI Appendix*, Fig. S12*B*). This is consistent with the pattern normally seen at promoter CGIs (28). It therefore seemed possible that, in the absence of the *RHBDF1* promoter CGI, the intragenic CGI in *RHBDF1* now acts itself as a promoter since it is no longer silenced by RNA transcripts running through this element.

To examine this, we generated a map of transcription initiation sites in *RHBDF1+P*- and *RHBDF1−P*-derived EBs using a modified version of the CapSeq method (29) and found that the CGI in *RHBDF1−P* EBs gives rise to multiple transcription initiation sites (*SI Appendix*, Fig. S12*C*). To investigate this further and obtain a quantitative measurement of transcription initiation, we carried out a strand-specific 5′ RACE assay. This confirmed that in EBs, the CGI in *RHBDF1−P* clones produced twofold higher levels of nascent sense transcripts than the CGI in *RHBDF1+P* clones (*SI Appendix*, Fig. S12*D*). These data show that, in the absence of transcription running through its sequence, the intragenic CGI of *RHBDF1* acts as a typical promoter CGI.

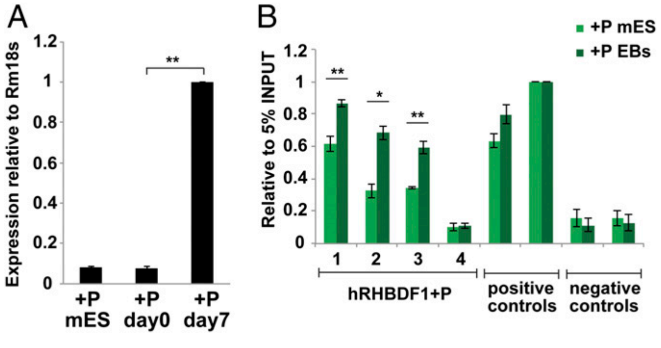

Fig. 7. Increase in the *RHBDF1+P* intragenic CGI DNA methylation upon in vitro differentiation is associated with higher gene expression and H3K36me3 enrichment. (*A*) *RHBDF1+P* nascent transcription level was measured by real-time RT-PCR in mES cells on days 0 and 7 of in vitro differentiation. The data were normalized to Rm18s and displayed as the average ± SD of three independent experiments. A paired *t* test was used to calculate statistical significance: $^{**}P < 0.01$. (*B*) ChIP was performed using antibodies to H3K36me3, H3, or a control nonspecific IgG in *RHBDF1+P* mES cells and day 7 EBs. The precipitated DNA fragments were used to amplify regions labeled 1, 2, 3, and 4 in Fig. 1. The data were normalized to INPUT and H3 level, and are displayed as the average ± SD of three independent experiments relative to the highest value. A paired *t* test was used to calculate statistical significance: $^{*}P < 0.05$, $^{**}P < 0.01$. The gene body sequences of mouse *Nprl3* and *Ccna2* served as positive controls, and two intergenic sites served as negative controls.

Global Analysis of Intragenic CGIs. Having demonstrated experimentally that transcription across a CGI and its modification by H3K36me3 is correlated with DNA methylation at the two specific intragenic CGIs described above, we asked whether this is a general phenomenon leading to methylation of intragenic CGIs throughout the genome. Based on UCSC Genome Browser annotations, we first classified 27,718 human CGIs into 13,478 promoter, 8,795 intragenic, and 5,445 intergenic islands. The promoter CGIs were further divided into 10,229 truly promoter and 2,836 alternative promoter CGIs. The latter act as a promoter CGI for one isoform of a gene and as an intragenic CGI of another isoform of the same gene (Fig. 8*A*).

The majority of promoter (~95%) and alternative promoter CGIs (~85%) are unmethylated (≤45% methylation). Their hypomethylated state is associated with H3K4me3 enrichment and lack of H3K36me3 (*SI Appendix*, Fig. S13). In contrast, the majority (~65%) of intragenic CGIs are DNA-methylated (≥55% methylation). The hypermethylated state of these CGIs is associated with traversing transcription elongation, H3K36me3 enrichment, and lack of H3K4me3 and H3K27me3 (Fig. 8 *B* and *C* and examples in *SI Appendix*, Fig. S14). The remaining intragenic CGIs are not methylated but appear to be silenced by a polycomb-mediated mechanism as they are neither transcribed nor enriched in H3K36me3 but are marked by a bivalent H3K4me3 and H3K27me3 chromatin signature (Fig. 8*C*).

We next analyzed alternative promoter CGIs and found that ~14% of these are DNA-methylated. Methylation of these CGIs is associated with traversing transcription elongation and H3K36me3 enrichment (*SI Appendix*, Fig. S13 *B* and *D*), suggesting that they are also methylated in a transcription-dependent mechanism. Of interest, this is in contrast to methylated promoter CGIs that are neither transcribed nor marked by H3K36me3, H3K4me3, and H3K27me3 (*SI Appendix*, Fig. S13 *A* and *C*), suggesting an alternative mechanism of methylation that is independent of transcription.

The gene body CGIs are smaller, with an average size of 513 ± 494 bp compared with the promoter CGIs, with an average size of 995 ± 725 bp. However, the % CpG and % GC within the identified classes of CGIs are similar (the difference between gene body CGIs and the other classes is less than 1.68% and less than 0.65%, respectively) (*SI Appendix*, Fig. S15). We did not detect any known motif enrichment within the gene body CGIs compared with all CGIs or the promoter CGIs only.

Therefore, our data indicate that when CGIs form a part of a transcriptional unit in early development, a large proportion of such islands lose H3K4me3, are modified by H3K36me3, and become DNA-methylated and silenced as development progresses. This applies to both transcribed alternative promoters and intragenic CGIs.

The Methylation Status of Intragenic CGIs Traversed by Transcription Elongation Depends on Their Intrinsic Activity as Promoters. Here, we have shown that whether or not a specific alternative promoter or intergenic CGI becomes methylated depends on its chromosomal context and, in particular, on whether or not transcription passes through the CGI at some point. Although many CGIs become methylated as transcription traverses them, this is not always the case, as illustrated by the ubiquitously expressed *ACTB* CGI

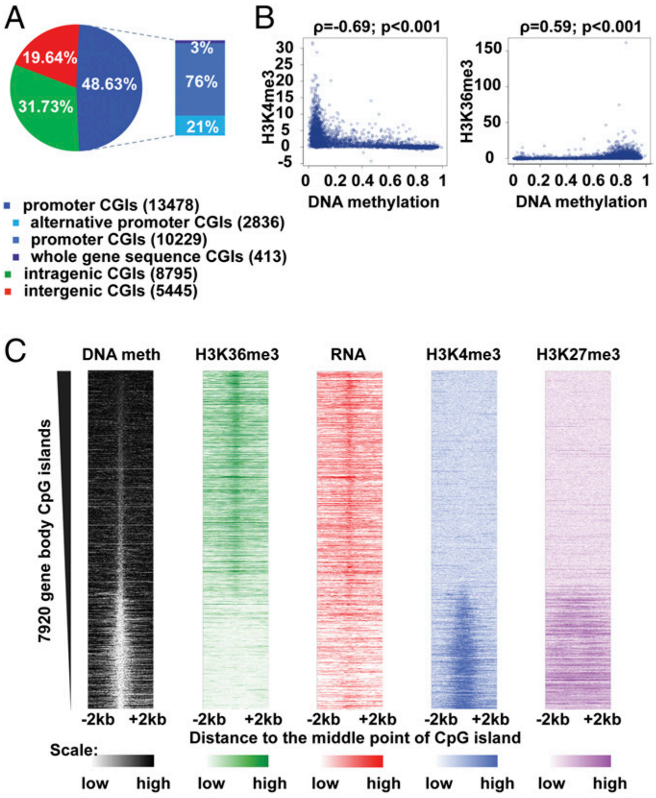

Fig. 8. Transcribed intragenic CGIs are enriched in H3K36me3 and hypermethylated in hES cells genome-wide. (*A*) Classification of CGIs into promoter, intragenic, and intergenic based on their genomic location. The promoter CGIs were subclassified into promoter only, alternative promoter CGIs (CGI that acts as a promoter CGI for one isoform of a gene and as an intragenic CGI for another isoform of the same gene), and CGIs that cover the whole gene sequence. (*B*) High DNA methylation level at intragenic CGIs strongly correlates with H3K36me3 and anticorrelates with H3K4me3 in hES cells. The Spearman correlation coefficient is shown. (*C*) Hypermethylated intragenic CGIs are overlapped by transcripts and enriched in H3K36me3 in hES cells genome-wide. Each line in the pileup plot represents one CGI (±2 kb from the middle of the CGI) sorted in descending order based on the DNA methylation levels (in black). Plots from hES cells show the read distribution of H3K36me3 (green), RNA expression (red), H3K4me3 (blue), and H3K27me3 (purple), while the DNA methylation status of individual CpGs from BS-Seq data is plotted in black.

that remains unmethylated when experimentally placed in the transcription unit of the *LUC7L* gene. So, what determines which intragenic CGIs become methylated and which remain unmethylated?

Taking *ACTB* as an example, it is more highly expressed in ZFαAS cells than the *HBA2, UBC*, and *MYOD* genes (*SI Appendix*, Figs. S5 and S16): Could it be that naturally occurring promoter CGIs that strongly initiate transcription in vivo may be refractory to silencing mediated by elongating transcription passing through them? We addressed this by identifying and analyzing 434 pairs of promoters linked to alternative promoter CGIs within the same gene (*SI Appendix*, Fig. S17). As a measure of transcriptional initiation strength, we focused on the level of preinitiation RNA polymerase II (Pol2) binding. We found that Pol2 binding correlates positively with unmethylated and H3K4me3-enriched promoters ($\rho = -0.33$, $P < 0.001$ and $\rho = 0.64$, $P < 0.001$, respectively) and alternative promoter CGIs ($\rho = -0.46$, $P < 0.001$ and $\rho = 0.73$, $P < 0.001$, respectively). By contrast, it correlates negatively with the presence of transcription traversing the CGI and H3K36me3 enrichment (promoter CGI: $\rho = -0.14$, $P < 0.001$ and alternative promoter CGIs: $\rho = -0.28$, $P < 0.001$) (*SI Appendix*, Fig. S13). We calculated the differences in Pol2 binding between promoter CGIs. We then linked alternative promoter CGIs in each pair and sorted by the calculated difference in Pol2 binding the level of DNA methylation; RNA expression; and enrichment in H3K36me3, Pol2, H3K4me3, and H3K27me3 at the two linked promoters (*SI Appendix*, Fig. S17). We found that the promoter CGI is hypomethylated regardless of gene expression. This is in contrast to the alternative promoter CGI, which is more likely methylated, H3K36me3-enriched, and transcribed through when the level of the Pol2 is higher at the upstream promoter CGI (*SI Appendix*, Figs. S17 and S18, group 1). Importantly, however, alternative promoter CGIs with higher Pol2 binding than on the paired promoter CGIs retain an unmethylated state, H3K4me3 enrichment, and lack H3K36me3 (*SI Appendix*, Figs. S17 and S18, group 3).

Therefore, transcription is not sufficient to drive DNA methylation of a transcribed alternative promoter CGI, and whether or not the latter becomes methylated depends on the activity of the promoter CGI relative to that of the promoter CGI driving transcription through it. Alternative promoter CGIs that are active (i.e., have a higher enrichment in Pol2 than the paired upstream promoter CGI) are protected from transcription-mediated methylation. Put simply, there appears to be a "duel" between competing CGIs within a transcription unit. These findings are consistent with observations made at the individual loci tested in mES cells and provide an explanation for why some intragenic CGIs, such as *ACTB* CGI, are able to escape transcription-mediated silencing and methylation.

Discussion

This study shows how transcription of CGIs, either as normal components of transcriptional units or when incorporated into abnormal transcriptional units as a result of chromosomal rearrangements, frequently leads to their methylation. We have shown that transcription through a CGI leads to the modification of its histones by H3K36me3 and results in DNMT3B-dependent DNA methylation and modification by H3K9me3. Given that H3K36me3 has been shown to act as a docking site for DNMT3B (19), it can be implied that, together, these modifications repress the default activity of intragenic CGIs to act as transcriptional promoters. This is also consistent with other studies supporting a role for transcription in regulating the activity of alternative promoters (30, 31). However, future work will be needed to determine if the chromatin marks are functionally involved in transcription-mediated silencing of CGIs (e.g., by deleting histone methyltransferases, such as SetD2 or the PWWP domain of DNMT3B).

Our findings are in agreement with recently published observations showing that de novo methylation of gene bodies requires transcription (19). However, we find two major differences when we compare the behavior of transcribed CGIs with that of gene body regions. First, in contrast to surrounding gene body sequences, CGIs that are traversed by transcription elongation remain unmethylated in naive mES cells and only become de novo-methylated upon mES differentiation to EBs (Figs. 3 and 6). A similar developmentally dependent establishment of CGI methylation in the presence of transcription has been observed in specialized situations, such as imprinting and X-inactivation (32, 33). This is probably due to the fact that in mES cells, CGIs are normally bound by CpG binding factors like CFP1, a component of the Set1 complex (34) that mediates trimethylation of H3K4 (26), a mark known to prevent recruitment of de novo DNMTs (35). Consistent with our findings, in mature oocytes, DNA methylation has been reported to be present preferentially at CGIs located within active transcription units (36), to be regulated by transcription (37), and to occur following loss in H3K4me3 and increase in H3K36me3 (38). Moreover, a new report has now shown that transient transcription in the early embryo leads to de novo methylation by altering histone profiles with functional consequences in later adult phenotypes (39), further supporting the observed correlation between histone marks and transcription-mediated de novo methylation of intragenic CGIs.

Second, unlike gene bodies, not all CGIs are susceptible to transcription-mediated DNA methylation (Fig. 4 and *SI Appendix*, Fig. S17), indicating that transcription, per se, is not sufficient to drive methylation of intragenic CGIs. Our data show that the relative strengths of transcription initiation from the intragenic CGI and the promoter of the gene within which it lies also play an important role in determining the status of the intragenic CGI. An intragenic CGI, which itself acts as a strong initiator of transcription, will not necessarily be silenced by elongating transcripts from a linked, but weaker, CGI promoter.

The data presented here, together with previously published evidence (19, 26, 40), support a model by which DNA methylation is established at intragenic CGIs (Fig. 9). Early during development, both promoter and intragenic CGIs are unmethylated and H3K4me3-enriched. The presence of H3K4me3 prevents methylation of the CGIs despite the presence of low levels of H3K36me3 (Fig. 9, *Upper*). Upon differentiation, at susceptible intragenic CGIs, loss of H3K4me3 is correlated with an increased level of H3K36me3 deposited by the elongating Pol2/SETD2 complex. DNMT3B in a complex with H3K9 methyltransferase is then recruited to intragenic CGIs via H3K36me3 and leads to their silencing via de novo DNA methylation and trimethylation of H3K9 (22) (Fig. 9, *Lower Right*). However, at highly active CGIs, H3K4me3 is retained and is associated with loss of H3K36me3, resulting in the CGI remaining free of methylation (Fig. 9, *Lower Left*).

This model highlights a mechanism whereby the act of transcription is primarily responsible for intragenic CGI methylation. Nevertheless, the context and nature of the CGI itself are both important in controlling whether a specific intragenic CGI is likely to become methylated. This model may explain other instances of aberrant CGI methylation observed in inherited diseases, in which transcription has been implicated (9, 41). In addition, the model is compatible with some instances of aberrant CGI methylation observed in acquired diseases, such as cancer. Indeed, silencing of the p15 tumor suppressor in leukemia has also been associated with up-regulation of antisense RNA transcription and consequent promoter CGI hypermethylation correlated with a loss of positive H3K4me3 and a gain of repressive H3K9me3 histone marks (42). Moreover, we have reported that in sporadic cases of colorectal cancer, transcription through the metastasis suppressor gene *TFPI-2* driven by an aberrantly active long interspersed element-1 promoter is associated with epigenetic silencing of its promoter CGI (43). Finally, based on the model, it is feasible that the switch to maintenance of methylation at intragenic CGIs in somatic cells could depend on the H3K9me3-mediated recruitment of the UHRF1/DNMT1/proliferating cell nuclear antigen complex (44), providing a potential way in which the *RHBDF1*'s intragenic CGI remains methylated in tissues in which *RHBDF1* is not expressed.

In summary, we have shown that the methylation status of a large proportion of the 30,000 CGIs is primarily determined by transcription. Silencing of intragenic CGIs by transcription-mediated DNA methylation will reduce transcriptional noise from CGIs that would otherwise initiate transcription from within the gene. Although this applies to the majority of intragenic CGIs, a proportion of intragenic CGIs with a strong propensity for transcriptional initiation themselves are not silenced despite being transcribed. Presumably a proportion of these remain active to subserve important biological functions. Finally, transcription-mediated DNA methylation and silencing clearly represents an epigenetic mutational process in a wide range of acquired diseases, including cancer, although its contribution to such diseases is, as yet, poorly documented.

Materials and Methods

Constructs. The ZFαAS-STOP construct was generated by blunt-end cloning the β-globin terminator from the βΔ5-7 plasmid (17) into a unique restriction site at the junction between *HBA2* and *LUC7L* fragments in the pZFαAS plasmid (10). The MYOD-AS, UBC-AS, and ACTB-AS constructs were assembled by cloning the PCR-amplified full length of *MYOD1*, *UBC*, and *ACTB* into the pZERO-LUC7L plasmid in the place of *HBA2* gene. The assembled vectors were assessed by restriction digest and sequencing. Primers used are listed in *SI Appendix*, Table S1.

Cell Culture. EBV-transformed lymphocytes, MEL cells, and mES clones were cultured and differentiated as described (10). Stable mES clones carrying the ZFαAS-STOP, MYOD1-AS, UBC-AS, and ACTB-AS clones were obtained by coelectroporation of 1.2 × 10^7 ES cells with 50 μg of purified construct and 2 μg of a 1.8-kb fragment of the pPNT plasmid containing the geneticin

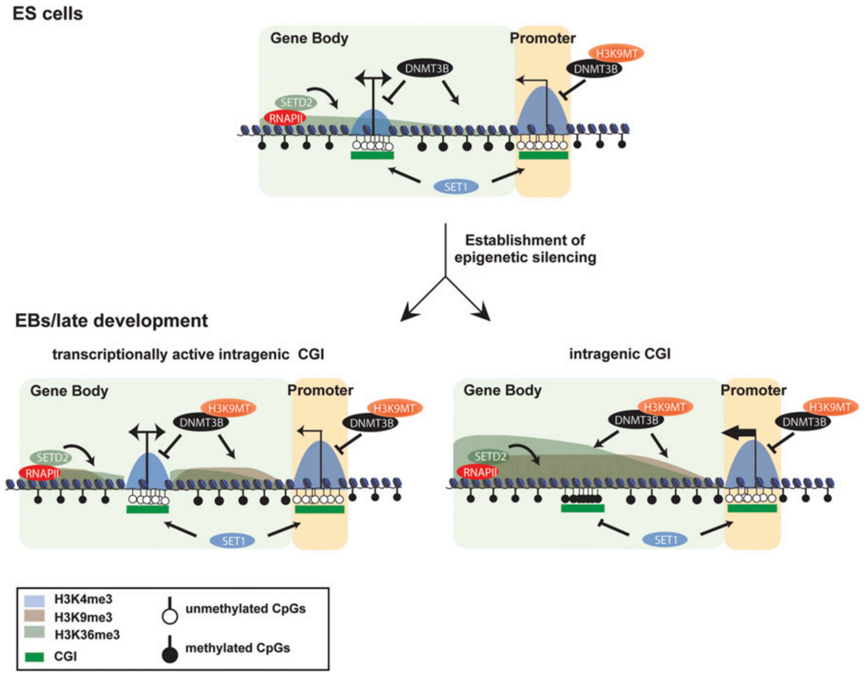

Fig. 9. Model for establishment of epigenetic silencing at intragenic CGIs. (*Upper*) In ES cells, CGIs found within the body of transcribed genes are free of methylation and marked by H3K4me3, despite the presence of a low level of H3K36me3. The hypomethylated state of CGIs is maintained through binding of ZF-CxxC domain-containing proteins, including SETD1; histone H3K4 methyltransferase (H3K4MT) complex (26); and KDM2A, an H3K36 demethylase enzyme (58). H3K4me3 is known to inhibit DNMT3B binding to CGIs (35). Therefore, in mES cells, DNA methylation and H3K9me3 are found only within the non-CGI part of the transcribed gene body (19, 40), but are omitted from the intragenic CGI. (*Lower Right*) Establishment of epigenetic silencing at the intragenic CGI occurs upon developmental progression or in vitro differentiation of ES cells to EBs. It is associated with loss of H3K4me3 at the intragenic CGI and a further increase in H3K36me3 probably mediated by raised levels of transcription (thick bent arrow). DNMT3B in complex with an H3K9 methyltransferase (H3K9MT) is now able to bind to the intragenic CGI, leading to establishment of DNA methylation and H3K9me3 within it. (*Lower Left*) In contrast, transcriptionally active intragenic CGIs with an expression level higher than the upstream promoter are refractory to DNA methylation and epigenetic silencing. These CGIs remain H3K4me3-enriched and transcriptionally active. (*Upper* and *Lower Left*) Double arrow at the intragenic CGI in indicates that the direction of transcription from the intragenic CGI is not important.

(G418) resistance gene. G418-resistant colonies were then selected and analyzed by PCR and Southern blotting for construct incorporation. Stable Dnmt3b$^{-/-}$ mES clones carrying the ZFαAS construct were similarly obtained, except that a fragment carrying the blasticidin resistance gene was used for cotransfection. Sequential selection, culture, and differentiation were carried out according to published protocols (21, 45).

E14-TG2a.IV mES cells with an RMCE cassette (*frt/Hprt*$^{-\Delta 3'}$*/loxP/MC1neo/lox511*) in the mouse α-globin locus (25) were cultured and differentiated as previously described (25, 46, 47).

H1 hES cells were expanded in mTeSR1 medium (Stem Cell Technologies) on Matrigel.

Recombinase-Mediated Cassette Exchange. Test DNA sequences were amplified from human genomic DNA and cloned into the multiple cloning sites (MCSs) in the RMCE donor vector (*loxP/Hprt*$^{-\Delta 5'}$*/frt/MCS/lox511*) (25). Primers used are listed in *SI Appendix*, Table S2. The in vitro-methylated version of the *RHBDF1*'s intragenic CGI was obtained by incubating the modified RMCE donor vector with *M.SssI* enzyme in the presence of *S*-adenosyl-L-methionine. The efficiency of the reaction was assessed by digesting the purified vector with DNA methylation-sensitive and -insensitive restriction enzymes. The targeting of the RMCE donor vector to E14-TG2a.IV mES cells with an RMCE cassette in the mouse α-globin locus was performed as described by Lynch et al. (25) and verified by Southern blot analysis.

Recombineering Strategy Called Bacterial Artificial Chromosome Shaving. The RMCE donor vectors containing the gene sequence of *RHBDF1*, including its promoter (*RHBDF1+P*; chr16:47,861–63,210, hg18) and minus its promoter (*RHBDF1−P*; chr16:47,911–60,819, hg18), were engineered using a λ-Red–mediated recombineering strategy called bacterial artificial chromosome (BAC) shaving. The human BAC clone, CTD3077J14 (Life Technologies), was modified by replacing BAC sequence between the 3′ end of the *RHDBF1+P/−P* gene sequence and the BAC origin of replication with the elements of the RMCE donor vector and a selection cassette, PGK-Tn903hyg-*loxP/Hprt*$^{-\Delta 5'}$*/frt/*. Next, the sequence upstream of the *RHBDF1* gene up to the BAC origin of replication was replaced with the cassette *lox511-PGK-Tn903neo*. Finally, an additional recombineering step was carried out to retrieve the Tn903hyg-*loxP/Hprt*$^{-\Delta 5'}$*/frt/RHDBF1+P/−P/lox511-PGK-Tn903neo* sequence from the shaved BAC into a minimal p15 plasmid vector by gap repair. Individual recombineering steps were assessed by antibiotic selection, PCR amplification to sequence the recombined junctions and determine BAC coverage, and restriction digestions.

ChIP and ChIP-Sequencing. ChIP was performed with the Millipore ChIP Assay Kit (17-295; Millipore). Briefly, 1×10^7 cells were cross-linked with 1% formaldehyde for 10 min. Chromatin was prepared according to the Millipore instructions and sonicated to an average fragment size of 500–1,000 bp using a Diagenode Bioruptor. Fragmented chromatin was immunoprecipitated with antibodies raised against H3K4me3 (ab8580; Abcam), H3K36me3 (ab9050; Abcam), H3K9me3 (kindly donated by Thomas Jenuwein, Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany; 17-625, Millipore; and pAb-056-050, Diagenode), H3K27me3 (17-622; Millipore and pAb-069-050; Diagenode), H3 (ab1791; Abcam), and IgG (X0903, Dako; used as a control). Precipitated DNA was analyzed by real-time PCR using SYBR Green chemistry except for the amplicon along the HBA locus (probes 5–9 in Fig. 1*A*), for which TaqMan chemistry was used. Primers used are listed in *SI Appendix*, Table S3.

H3K36me3-immunoprecipitated chromatin from H1 hES cells was prepared as described above. ChiP-sequencing (ChIP-Seq) libraries were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (E7370S; New England BioLabs) and sequenced on a NextSeq 500 sequencing system. The ChIP-Seq data were analyzed using an in-house pipeline that automatically assesses the quality of data, trims adaptors, merges short reads, maps reads to the reference genome (hg19), filters data for duplicates, and performs peak calling. The data were visualized in the UCSC Genome Browser.

RT-PCR. RNA was isolated using TRI Reagent (T9424; Sigma–Aldrich) and subsequently DNase-treated using a DNA-free DNA Removal Kit (AM1906; Ambion). The expression analysis of constructs in the ZF model was performed using the Roche Expand reverse transcriptase kit (11785826001) following the manufacturer's guidelines. Gene-specific primers or random primers (C118A; Promega) were used depending on whether strand-specific analysis was required. Expression of strand-specific RNA was performed using end-point PCR, while expression of spliced sense genes was analyzed by real-time PCR using SYBR Green chemistry and expressed relative to mouse *Aprt*. For the experiments on *RHBDF1*, first-strand cDNA was synthesized from total RNA using a SuperScript III First-Strand Synthesis SuperMix for qRT-PCR kit (11752-050; Invitrogen) and quantified by real-time PCR using SYBR Green chemistry. The gene expression was normalized to mouse Rm18s. Primer sequences are listed in *SI Appendix*, Tables S4 and S5.

Bisulfite Sequencing. Genomic DNA was bisulfite-converted using the EZ DNA Methylation-Gold kit (D5005; Zymo Research). Enzymatically methylated human male genomic DNA was used as a methylated control (S41821; Millipore). Several nested primer sets were designed to cover the region of interest (*SI Appendix*, Table S6). The amplified bisulfite-modified DNA was purified to remove the primer dimers. Samples were pooled together, and sequence libraries were generated using a NEBNext Ultra DNA Library Prep Kit for Illumina (E7370S; New England BioLabs), to sequence on MiSeq (Illumina). The bisulfite sequencing data were aligned to hg18 and mm9 using Bismark (48) and SeqMonk tools, and then visualized in the UCSC Genome Browser.

For the ZFαAS time course, amplified products were cloned, sequenced, and analyzed as described by Vafadar-Isfahani et al. (49).

CapSeq. Total RNA was extracted using TRI Reagent (T9424; Sigma–Aldrich) and then fragmented using the NEBNext Magnesium RNA Fragmentation Module (New England BioLabs). Fragmented RNA was cleaned up using the RiboMinus Concentration Module (Life Technologies), and the 3′ adapters were ligated using a modified version of the NEBNext Small RNA Library Prep Set for Illumina (Multiplex Compatible) (New England BioLabs). Samples were subjected to a second round of cleanup using the RiboMinus Concentration Module, and then subsequently treated with calf intestinal phosphatase and tobacco acid phosphatase. The reaction was cleaned up using the RiboMinus Concentration Module as before, and annealing of the small RNA reverse transcription primer and ligation of the 5′ adapter were performed. Reverse transcription was carried out using the SuperScript III First-Strand Synthesis system, followed by library amplification, and was cleaned up using a QIAquick PCR Purification kit (28104; Qiagen). Samples were sequenced on the NextSeq 500 sequencing system. CapSeq data were aligned to build hg19 using STAR and visualized in the USCS Genome Browser.

5′ RACE. A 5′ RACE assay was performed on total DNase-treated RNA using the 5′ RACE System for Rapid Amplification of cDNA Ends (18374-058; Invitrogen). Briefly, first-strand cDNA was synthesized using strand-specific primers to *hRHBDF1*, and *mGapdh* was used as an internal control. Actinomycin D was added to prevent second-strand cDNA synthesis. The cDNA was then cleaned up using a S.N.A.P. column (Invitrogen), TdT tailing was performed, and the dC-tailed cDNA was amplified using nested strand-specific *hRHBDF1* and *mGapdh* primers in combination with the provided abridged anchor primer. Following amplification, the cDNA was quantified by real-time PCR using SYBR Green chemistry. Gene expression was normalized to mouse *Gapdh*. Primer sequences are listed in *SI Appendix*, Table S7.

Small RNA Northern Blots. Total RNA was run on a polyacrylamide gel, transferred to a nylon membrane, and hybridized with oligonucleotide probes to detect small RNA species of *HBA2* in accordance with published techniques (50). An oligonucleotide complementary to microRNA miR16 (5′ CGCCAATATTTACGTGCTGCTA 3′) was used as a positive control, as previously reported, to be expressed at low levels in mES cells and higher levels in EBs (51, 52).

Small RNA-Seq. Total RNA was extracted using a Qiagen AllPrep kit using the conditions indicated for enrichment in small RNAs. Libraries were prepared according to the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England BioLabs) and sequenced on a HiSeq system. The small RNA-Seq data were trimmed with an in-house Perl script to 25-base-long reads to remove sequencing adaptors. Reads were mapped to both mm9 and hg19 genomes using STAR (allowing each pair to map a maximum of two times to the genome). Read pairs per million mapped read pairs were calculated within the elements of the ZFαAS constructs [HS-40 enhancer (chr16:163249–164691), *HBA2* gene (chr16:222845–223709), part of *LUC7L* gene (chr16:277024–282124)], and miR16 as a positive control (chr14:62250772–62250794 and chr3:68813839–68813861) using samtools. Reads that mapped to hg19 and mm9 within the ZFαAS constructs were filtered out using an in-house Perl script (only a few double-mapped reads were reported). The data were expressed as reads per million normalized to mouse mapped reads and visualized in UCSC using bedtools and ucsctools.

Bioinformatics Analyses. Annotated CGIs (n = 27,718, UCSC definition, hg19) were classified into promoter, intragenic, and intergenic based on their genomic location. CGIs were classified as promoter-associated (n = 13,478) if they overlap TSS ± 100 bp of USCS-annotated genes. CGIs were classified as intragenic-associated (n = 8,795) if they overlap a gene sequence lacking promoter CGI ± 100 bp and also were not already categorized as promoter CGI, while the remaining CGIs were cataloged as intergenic-associated (n = 5,445). Sequentially, the promoter CGIs were classified into (*i*) promoter CGIs that overlap the whole gene sequence (n = 413), (*ii*) alternative promoter CGIs that overlap TSS ± 100 bp and gene body sequence (n = 2,836), and (*iii*) promoter-only CGIs (n = 10,229). Data from each step of the analysis were visually inspected using MIG. Heat map plots for promoter and intragenic CGIs in hES H1 cells were generated for CGIs with >60% coverage using datasets either publicly available [bisulfite sequencing (BS-Seq) (GSM429322) (53), RNA-Seq (GSM958733) (54), H3K4me3 (GSM733657), H3K27me3 (GSM733748), Pol2 MMS-126R (GSM803366), input control for H3K4me3 and H3K27me3 (GSM733770), Pol2 (GSM803364)] or generated in-house (H3K36me3). The input was subtracted from the ChIP data. CGIs were sorted in descending order based on the DNA methylation level or differential Pol2 binding within ±2 kb from the middle of CGI, and enrichment for H3K36me3, RNA expression, H3K4me3, and H3K27me3 reads was displayed using the in-house Perl script and visualized in R. Other publically available datasets used to determine *RHBDF1* DNA methylation levels and expression were as follows: Bio-CAP-Seq: testes (GSM1064675), liver (GSM1064676), input control (GSM1064674) (55); BS-Seq: hES (GSM429322); and RNA-Seq: hES (GSM958733), liver (GSM752705) (56). RNA-Seq, ChIP-Seq, and CAP-Seq datasets produced in the course of this study have been submitted to the Gene Expression Omnibus with superseries accession no. GSE84355.

Motif enrichment was conducted using Homer v4.9 (57).

Statistical Analyses. Statistical significance was calculated by a two-tailed t test using Graphpad Prism 6, R, or Excel on data obtained by analyzing in triplicate a minimum of two biological replicates.

ACKNOWLEDGMENTS. We thank Rob Klose and Skirmantas Kriaucionis for critical reading of the manuscript, Thomas Jenuwein for H3K9me3 antibodies, Sue Butler and Jackie Sloane Stanley for tissue culture, Nick Proudfoot for donating the β-globin terminator, Nicki Gray for help with drawing the model, and Felix Krueger for help with analysis of the bisulfite sequencing data. This work was funded by the Medical Research Council (MRC), R.J.S.M. was supported by an MRC Capacity Building Studentship, and C.T. was supported by a Royal Society Dorothy Hodgkin Fellowship.

1. Smith ZD, Meissner A (2013) DNA methylation: Roles in mammalian development. *Nat Rev Genet* 14:204–220.
2. Deaton AM, et al. (2011) Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res* 21:1074–1086.
3. Jones PA (2012) Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13:484–492.
4. Zeng J, Nagrajan HK, Yi SV (2014) Fundamental diversity of human CpG islands at multiple biological levels. *Epigenetics* 9:483–491.
5. Illingworth R, et al. (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol* 6:e22.
6. Edwards CA, Ferguson-Smith AC (2007) Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol* 19:281–289.
7. Reik W (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447:425–432.
8. Jones PA, Baylin SB (2007) The epigenomics of cancer. *Cell* 128:683–692.
9. Ligtenberg MJ, et al. (2009) Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3′ exons of TACSTD1. *Nat Genet* 41:112–117.
10. Tufarelli C, et al. (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet* 34:157–165.
11. Vyas P, et al. (1992) Cis-acting sequences regulating expression of the human alpha-globin cluster lie within constitutively open chromatin. *Cell* 69:781–793.
12. Garrick D, et al. (2008) The role of the polycomb complex in silencing alpha-globin gene expression in nonerythroid cells. *Blood* 112:3889–3899.
13. Barbour VM, et al. (2000) Alpha-thalassemia resulting from a negative chromosomal position effect. *Blood* 96:800–807.
14. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130:77–88.
15. Ku M, et al. (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* 4:e1000242.
16. Pan G, et al. (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1:299–312.
17. Dye MJ, Proudfoot NJ (2001) Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell* 105:669–681.
18. Holoch D, Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* 16:71–84.
19. Baubec T, et al. (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* 520:243–247.
20. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99:247–257.
21. Jackson M, et al. (2004) Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol Cell Biol* 24:8862–8871.
22. Rose NR, Klose RJ (2014) Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim Biophys Acta* 1839:1362–1372.
23. Daniels RJ, et al. (2001) Sequence, structure and pathology of the fully annotated terminal 2 Mb of the short arm of human chromosome 16. *Hum Mol Genet* 10:339–352.
24. Nakagawa T, et al. (2005) Characterization of a human rhomboid homolog, p100hRho/RHBDF1, which interacts with TGF-alpha family ligands. *Dev Dyn* 233:1315–1331.
25. Lynch MD, et al. (2012) An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J* 31:317–329.
26. Thomson JP, et al. (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464:1082–1086.
27. Quante T, Bird A (2016) Do short, frequent DNA sequence motifs mould the epigenome? *Nat Rev Mol Cell Biol* 17:257–262.
28. Barski A, et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837.
29. Gu W, et al. (2012) CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. elegans piRNA precursors. *Cell* 151:1488–1500.
30. Maunakea AK, et al. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466:253–257.
31. Neri F, et al. (2017) Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 543:72–77.
32. Latos PA, et al. (2012) Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338:1469–1472.
33. Ohhata T, Hoki Y, Sasaki H, Sado T (2008) Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. *Development* 135:227–235.
34. Lee JH, Skalnik DG (2005) CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J Biol Chem* 280:41725–41731.
35. Ooi SK, et al. (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448:714–717.
36. Smallwood SA, et al. (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* 43:811–814.
37. Veselovska L, et al. (2015) Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol* 16:209.
38. Stewart KR, et al. (2015) Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. *Genes Dev* 29:2449–2462.
39. Greenberg MV, et al. (2017) Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. *Nat Genet* 49:110–118.
40. Vakoc CR, Mandat SA, Olenchock BA, Blobel GA (2005) Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol Cell* 19:381–391.
41. Venkatachalam R, et al. (2010) Germline epigenetic silencing of the tumor suppressor gene PTPRJ in early-onset familial colorectal cancer. *Gastroenterology* 139:2221–2224.
42. Yu W, et al. (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 451:202–206.
43. Cruickshanks HA, et al. (2013) Expression of a large LINE-1-driven antisense RNA is linked to epigenetic silencing of the metastasis suppressor gene TFPI-2 in cancer. *Nucleic Acids Res* 41:6857–6869.
44. Rothbart SB, et al. (2012) Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat Struct Mol Biol* 19:1155–1160.
45. Chen T, Ueda Y, Dodge JE, Wang Z, Li E (2003) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* 23:5594–5605.
46. Keller G, Kennedy M, Papayannopoulou T, Wiles MV (1993) Hematopoietic commitment during embryonic stem cell differentiation in culture. *Mol Cell Biol* 13:473–486.
47. Wallace HA, et al. (2007) Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell* 128:197–209.
48. Krueger F, Andrews SR (2011) Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27:1571–1572.
49. Vafadar-Isfahani N, et al. (2017) Decoupling of DNA methylation and activity of intergenic LINE-1 promoters in colorectal cancer. *Epigenetics* 12:465–475.
50. Pall GS, Codony-Servat C, Byrne J, Ritchie L, Hamilton A (2007) Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res* 35:e60.
51. Houbaviy HB, Murray MF, Sharp PA (2003) Embryonic stem cell-specific MicroRNAs. *Dev Cell* 5:351–358.
52. Tang F, Hajkova P, Barton SC, Lao K, Surani MA (2006) MicroRNA expression profiling of single whole embryonic stem cells. *Nucleic Acids Res* 34:e9.
53. Lister R, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–322.
54. Djebali S, et al. (2012) Landscape of transcription in human cells. *Nature* 489:101–108.
55. Long HK, et al. (2013) Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *Elife* 2:e00348.
56. Brawand D, et al. (2011) The evolution of gene expression levels in mammalian organs. *Nature* 478:343–348.
57. Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589.
58. Blackledge NP, et al. (2010) CpG islands recruit a histone H3 lysine 36 demethylase. *Mol Cell* 38:179–190.