ARTICLE

doi:10.1038/nature12598

DNMT1-interacting RNAs block gene-specific DNA methylation

Annalisa Di Ruscio[1,2,3]*, Alexander K. Ebralidze[1,2]*, Touati Benoukraf[4], Giovanni Amabile[1,2], Loyal A. Goff[5,6,7], Jolyon Terragni[8], Maria Eugenia Figueroa[9], Lorena Lobo De Figueiredo Pontes[1,2], Meritxell Alberich-Jorda[1,2,10], Pu Zhang[1], Mengchu Wu[4], Francesco D'Alò[3], Ari Melnick[11], Giuseppe Leone[3], Konstantin K. Ebralidze[2], Sriharsa Pradhan[8], John L. Rinn[1,2,5,6] & Daniel G. Tenen[1,4]

DNA methylation was first described almost a century ago; however, the rules governing its establishment and maintenance remain elusive. Here we present data demonstrating that active transcription regulates levels of genomic methylation. We identify a novel RNA arising from the *CEBPA* gene locus that is critical in regulating the local DNA methylation profile. This RNA binds to DNMT1 and prevents *CEBPA* gene locus methylation. Deep sequencing of transcripts associated with DNMT1 combined with genome-scale methylation and expression profiling extend the generality of this finding to numerous gene loci. Collectively, these results delineate the nature of DNMT1–RNA interactions and suggest strategies for gene-selective demethylation of therapeutic targets in human diseases.

DNA methylation is a key epigenetic signature implicated in transcriptional regulation, genomic imprinting, and silencing of repetitive DNA elements[1,2] that occurs predominantly within CpG dinucleotides. CpG dinucleotides are underrepresented in the mammalian genome (~1%) and tend to cluster within CpG islands located in the vicinity of the transcription start sites (TSSs) of the majority (~70%) of human protein-coding genes[3]. Although the bulk of genome is methylated at 70–80% of its CpGs, CpG islands are mostly unmethylated in somatic cells[3,4]. This modification is mediated by the members of the DNA methyltransferase (DNMT) family, conventionally classified as *de novo* (DNMT3a and DNMT3b) and maintenance (DNMT1). In terms of epigenetic inheritance, DNMT1 has the unique ability of identifying the hemimethylated portion of newly replicated DNA. This feature may explain how DNMT1-mediated methylation could be an epigenetic mechanism maintaining the status quo. However, it certainly does not explain how DNA methylation is altered, particularly in disease states.

To examine how transcription may regulate the levels of genomic methylation, we investigated methylation dynamics of the well-studied methylation-sensitive gene *CEBPA*[5–7], including the potential involvement of non-coding RNAs (ncRNAs) originating within the *CEBPA* locus. Recent discoveries of functional ncRNAs have provided new regulatory clues to the control of epigenetic marks. In particular, long ncRNAs have been shown to regulate gene expression by interacting with chromatin modifiers, modulating transcription factor activity and competing for microRNA binding[8–16]. One unexplored aspect of the regulation of gene locus DNA methylation was the possible involvement of transcripts encoded within the region.

We identified a functional RNA arising from the *CEBPA* locus, *ecCEBPA*, that regulates *CEBPA* methylation. This RNA interacts with DNMT1, resulting in prevention of *CEBPA* gene methylation and robust *CEBPA* messenger RNA production. We show that such functional DNMT1–RNA association occurs at numerous gene loci. We thus propose a novel regulatory mechanism of gene methylation governed by RNAs.

Characterization of *ecCEBPA*

Non-coding transcripts arising from the promoter and the downstream regions of coding genes can affect the expression of the corresponding genes[17–19]. We searched and identified transcripts upstream and downstream of the intronless *CEBPA* gene. Strand-specific reverse transcriptase PCR (RT–PCR; data not shown) and northern blot analysis of RNAs from four leukaemic cell lines, probing the region immediately after the *CEBPA* polyadenylation site, revealed the presence of a major band of ~4.5 kilobases (kb) in HL-60 and U937 (in which *CEPBA* is expressed), but not in K562 or Jurkat (in which *CEBPA* is expressed at low or undetectable levels), cell lines (Fig. 1a, b). The identified transcript is distinct from the ~2.6 kb signal, detected with a *CEBPA* coding-region probe, and correlates with *CEBPA* mRNA expression. Unlike polyadenylated *CEBPA* mRNA (Fig. 1c), this non-polyadenylated transcript is enriched in the nuclear fraction (Supplementary Fig. 1a, b), suggesting functional roles independent of protein-coding potential.

We termed this nuclear non-polyadenylated *CEBPA* ncRNA *extra-coding CEBPA* (*ecCEBPA*), as it encompasses the entire mRNA sequence in the same-sense orientation (shown by primer extension and 5′ and 3′ rapid amplification of complementary DNA ends (RACE); Supplementary Information and Supplementary Fig. 1c, d). Quantitative (q)RT–PCR analysis confirmed concordant expression between extra-coding and coding transcripts, in both cellular and nuclear RNAs (Fig. 1d, e). Similar correlation was observed in all tested human tissues (Supplementary Fig. 1e). Notably, *ecCEBPA* synthesis precedes the expression of its overlapping mRNA in the S phase (Supplementary Information and Supplementary Fig. 1f, g) and is regulated by both RNA polymerase (RNAP) II and III (Supplementary Information and Supplementary Fig. 1h–p), as described for other loci[20–22].

[1]Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115, USA. [3]Università Cattolica del Sacro Cuore, Institute of Hematology, L.go A. Gemelli 8, Rome 00168, Italy. [4]Cancer Science Institute, National University of Singapore, 117599, Singapore. [5]Department of Stem Cell and Regenerative Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA. [6]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [7]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, Massachusetts 02139, USA. [8]New England Biolabs, 240 County Road, Ipswich, Massachusetts 01938-2723, USA. [9]University of Michigan, Department of Pathology, Ann Arbor, Michigan 48109-2200, USA. [10]Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, 142 20 Prague, Czech Republic. [11]Department of Medicine, Hematology-Oncology, C-620 Weill Cornell Medical College, 1300 York Avenue, New York, New York 10021, USA.
*These authors contributed equally to this work.

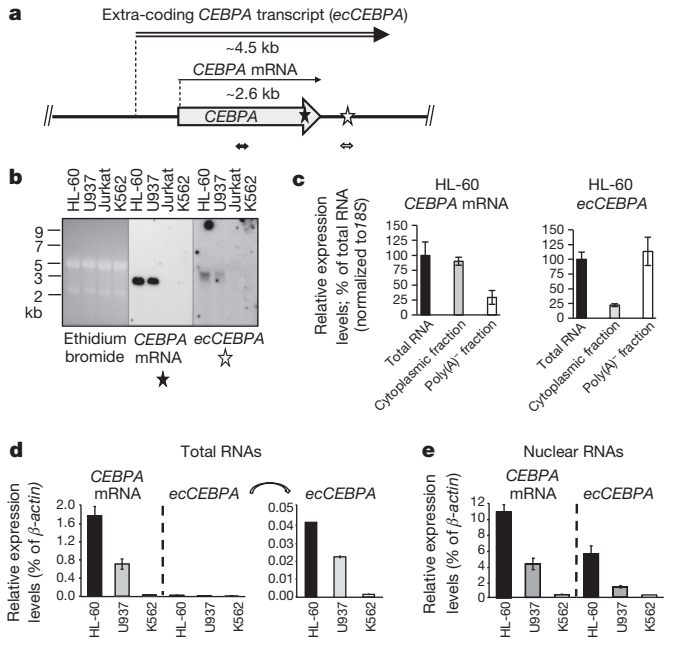

Figure 1 | Characterization of *ecCEBPA*. **a**, Diagram of *CEBPA* transcripts. *CEBPA* mRNA (small black double-headed arrow) and *ecCEBPA* (small white double-headed arrow) qRT–PCR primer sets are located in the coding region and after the poly(A) signal, respectively. **b**, Assessment of transcripts by northern blot hybridization. **c**–**e**, Relative levels of the transcripts in cellular fractions. In panel **d**, *ecCEBPA* levels are shown on different scales. qRT–PCR bars indicate mean ± s.d. ($n = 3$).

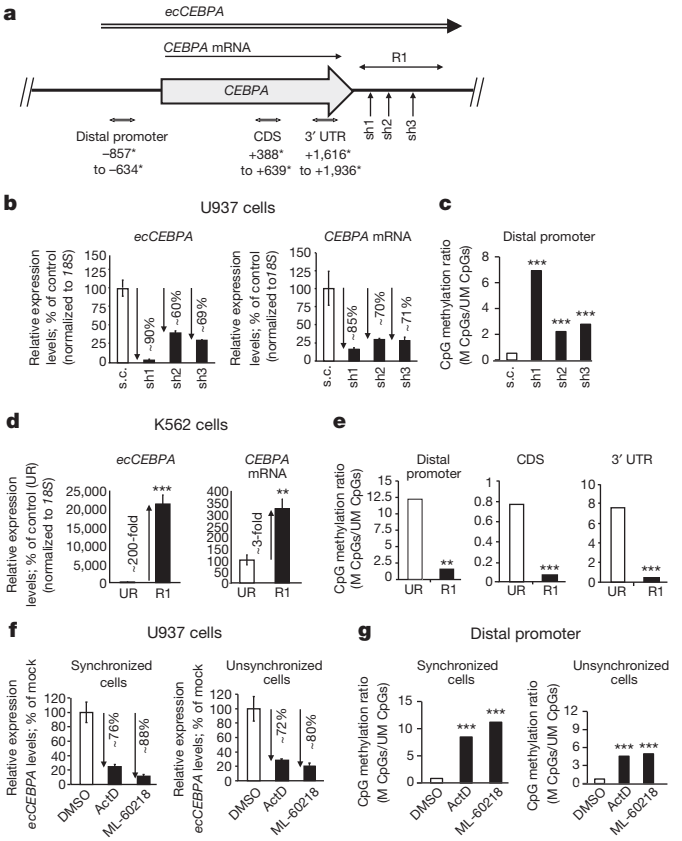

Figure 2 | Loss- and gain-of-function studies demonstrate that *ecCEBPA* maintains *CEBPA* expression by regulating methylation of the *CEBPA* locus. **a**, Diagram indicating the position of target sequences for shRNA constructs (sh1–sh3); the fragment derived from *ecCEBPA* used for overexpression (R1); regions analysed for changes in DNA methylation (distal promoter; coding sequence (CDS) and 3′ untranslated region (UTR)). Asterisks indicate number of base pairs away from the *CEBPA* TSS. **b**, **c**, The results of *ecCEBPA* loss-of-function in *CEBPA*-expressing U937 cells. Effect of *ecCEBPA*-targeting shRNAs on *CEBPA* mRNA levels (qRT–PCR, bars indicate mean ± s.d. (**b**)) and methylation of the *CEBPA* promoter (**c**). DNA methylation changes are shown as the ratios of methylated (M) to unmethylated (UM) CpGs in all clones analysed per each construct ($n = 14$). s.c., scrambled control. **d**, **e**, The results of *ecCEBPA* gain-of-function studies in K562 cells, in which *CEBPA* is methylated and silenced. **d**, Effect of *ecCEBPA* upregulation on *CEBPA* mRNA levels. UR, unrelated region. qRT–PCR, bars indicate mean ± s.d. ($n = 4$). **e**, Effect of *ecCEBPA* upregulation on methylation of the *CEBPA* locus (DNA methylation changes were assessed as described in **c** ($n = 14$ for distal promoter and $n = 6$ for CDS and 3′ UTR)). **f**, **g**, The results of transcription inhibition in U937 cells. **f**, ec*CEBPA* expression levels after treatment with actinomycin D (actD) and ML-60218 in synchronized and unsynchronized cells. qRT–PCR, bars indicate mean ± s.d. **g**, DNA methylation changes after treatment with actinomycin D and ML-60218 in synchronized ($n = 12$) and unsynchronized ($n = 10$) cells (assessed as described in **c**). DMSO, dimethylsulphoxide. Drug concentrations: actinomycin D, 0.8 μm; ML-60218, 150 μm. Duration of treatment was 7 h. All bisulphite sequenced clones were analysed by Fisher's exact test. $^{**}P < 0.01$; $^{***}P < 0.001$.

ecCEBPA blocks methylation and maintains *CEBPA* mRNA

To examine the functional role of *ecCEBPA* in the regulation of *CEBPA* transcription, we performed both loss- and gain-of-function experiments. Knockdown of *ecCEBPA* in a U937 cell line (up to a fourfold decrease) achieved by short hairpin RNAs (shRNAs) targeting *ecCEBPA* (but not *CEBPA* mRNA) led to a decrease of *CEBPA* mRNA expression of similar magnitude (Fig. 2a, b), suggesting that *ecCEBPA* may regulate *CEBPA* expression. Silencing of the *CEBPA* gene can be associated with DNA methylation of the promoter[6,7,23]. To examine whether there was a connection between *ecCEBPA* and methylation of the *CEBPA* locus, we analysed methylation within the distal promoter (located at −0.8 to −0.6 kb from the *CEBPA* TSS; Fig. 2a). Intriguingly, *ecCEBPA* knockdown led to a significant increase in DNA methylation compared to the non-targeting control (Fig. 2c and Supplementary Fig. 2a).

To investigate whether enforced expression of *ecCEBPA* was sufficient to inhibit DNA methylation, the downstream region of *ecCEBPA* (R1; Fig. 2a) was overexpressed in K562 cells expressing *ecCEBPA* and *CEBPA* mRNA at low-to-undetectable levels (Fig. 1b, d). Overexpression of only part of *ecCEBPA* was dictated by the necessity to distinguish the methylation pattern of the endogenous *CEBPA* locus from that of the ectopically expressed construct.

Ectopic expression of *ecCEBPA* R1 resulted in greater-than-threefold increase in mRNA expression (Fig. 2d), whereas overexpression of an unrelated region (located 45 kb downstream) and regions immediately outside of the *ecCEBPA* boundaries did not affect mRNA levels (Supplementary Fig. 2b–d). Moreover, a concomitant decrease of DNA methylation in three tested regions within the *CEBPA* gene, distal promoter, coding region and 3′ untranslated region accompanied overexpression of *ecCEBPA* but not of the unrelated region (Fig. 2e and Supplementary Fig. 2e, f). Interestingly, comparative analysis of DNA methylation changes imposed by *ecCEBPA* overexpression versus the hypomethylating agent 5′-azacytidine (5-aza-CR), together with genome-scale analysis (reduced representation bisulphite sequencing; RRBS[24]) of DNA methylation changes imposed by *ecCEBPA* versus unrelated region overexpression, revealed that *ecCEBPA*-mediated demethylation was relatively selective to the *CEBPA* locus (Supplementary Fig. 2g–l). Indeed, increased mRNA expression and changes in methylation status within the loci of the neighbouring *CEBPG* and distant *TP73* (on chromosome 1p36) genes were achieved after 5-aza-CR treatment but not after *ecCEBPA* overexpression (Supplementary Fig. 2g–k). Furthermore, RRBS analysis of promoter and first exon regions revealed that only ~3.3% of the interrogated loci (396 out of 11,844)

were hypomethylated at levels similar to the *CEBPA* locus (Supplementary Fig. 2l).

Furthermore, *ecCEBPA* downregulation by the universal RNAP inhibitor actinomycin D and RNAP III-specific inhibitor ML-60218 (Supplementary Fig. 2m) led to a corresponding increase in methylation of the *CEBPA* locus, in synchronized and unsynchronized U937 cells (Fig. 2f, g and Supplementary Fig. 2n). Despite comparable decreases in *ecCEBPA* levels in both synchronized and unsynchronized cells (Fig. 2f), DNA methylation increase was more prominent in synchronized cells (Fig. 2g), suggesting a cell-cycle-specific action of the *ecCEBPA*. A similar effect was observed in ML-60218-treated HL-60 cells (Supplementary Figs 1i and 2o).

Consistently, we observed an inverse correlation between the *CEBPA* gene locus methylation and the levels of ec*CEBPA* in HL-60, U937 and K562 cell lines (Supplementary Fig. 2p).

Collectively, these data highlight the regulatory role of *ecCEBPA* in *CEBPA* gene locus methylation, most prominently during the S phase.

DNMT1 binds to RNA with greater affinity than to DNA

The changes in *CEBPA* methylation mediated by *ecCEBPA* prompted us to try to determine the mechanism through which it is achieved. Among DNMTs, it is DNMT1 whose expression and enzymatic activity peaks during S phase[25]. Increased *ecCEBPA* expression occurs during the S phase (Supplementary Fig. 1g), whereas inhibition of *ecCEBPA* during S phase results in a substantial increase of *CEBPA* locus DNA methylation (Fig. 2f, g and Supplementary Fig. 2n). We therefore asked whether the presence of *ecCEBPA* during S phase led to RNA interference of DNMT1 activity.

To determine whether DNMT1 physically associates with *ecCEBPA* we performed RNA immunoprecipitation (RIP) with specific anti-DNMT1 antibody (Supplementary Fig. 3a).We observed *ecCEBPA* enrichment in DNMT1–RNA precipitates, demonstrating a physical interaction between *ecCEBPA* and DNMT1 (Fig. 3a, b and Supplementary Fig. 3a). Analysis of polyadenylated (poly(A)$^+$) and non-polyadenylated (poly(A)$^-$) fractions in DNMT1–RNA precipitates revealed enrichment of *CEBPA* transcripts in the poly(A)$^-$ fraction (Supplementary Fig. 3b, c), suggesting that the major component of *CEBPA* transcripts in DNMT1–RNA precipitates was *ecCEBPA*.

To investigate the molecular properties of RNA–DNMT1 interaction *in vitro*, we performed (RNA) electrophoresis mobility shift assays ((R)EMSAs). RNA oligonucleotides corresponding to the 5′and 3′ parts of *ecCEBPA* were selected by: (1) the ability (R2, R5 and R6) and inability (R4) to fold into stem–loop structures[26]; and (2) the presence (R2, R5 and R6) or absence (R4) of CpG dinucleotides (Fig. 3a). RNA–DNMT1 complex formation was observed with all RNA oligonucleotides able to fold into stem–loop structures (Fig. 3c–e). Unlike DNA[27], CpG to UpG substitutions, neutral with regard of secondary structures, did not affect binding (mutR2; Fig. 3c). By contrast, mutations abrogating RNA folding ability affected RNA–DNMT1 binding (mut R5; Fig. 3e). Analyses extended to a number of RNA oligonucleotides not related to *ecCEBPA* (single-stranded R1 and R3 and double-stranded R13; Supplementary Fig. 3d) confirmed DNMT1 binding to stem–loop-structured RNAs (Supplementary Information and Supplementary Fig. 3e, f). Importantly, REMSAs performed in the presence of increasing concentrations of spermine, a molecule with four positive charges at high density, excluded a case of charge–charge interactions (Supplementary Fig. 3g), supporting a strong element of structural recognition between DNMT1 and RNA.

To determine the relative affinity of DNMT1 for *ecCEBPA* versus DNA, single-stranded RNA oligonucleotides capable of forming secondary structures (R5) and corresponding unmethylated double-stranded DNA (umDNA; D5/D6), hemimethylated double-stranded DNA (hmDNA; D5/D6) and fully methylated double-stranded DNA (mDNA; D5/D6) (Fig. 3a), at a constant molar concentration, were titrated with an increasing range of DNMT1 enzyme concentrations using EMSA. RNA formed complexes beginning at <0.013 μM

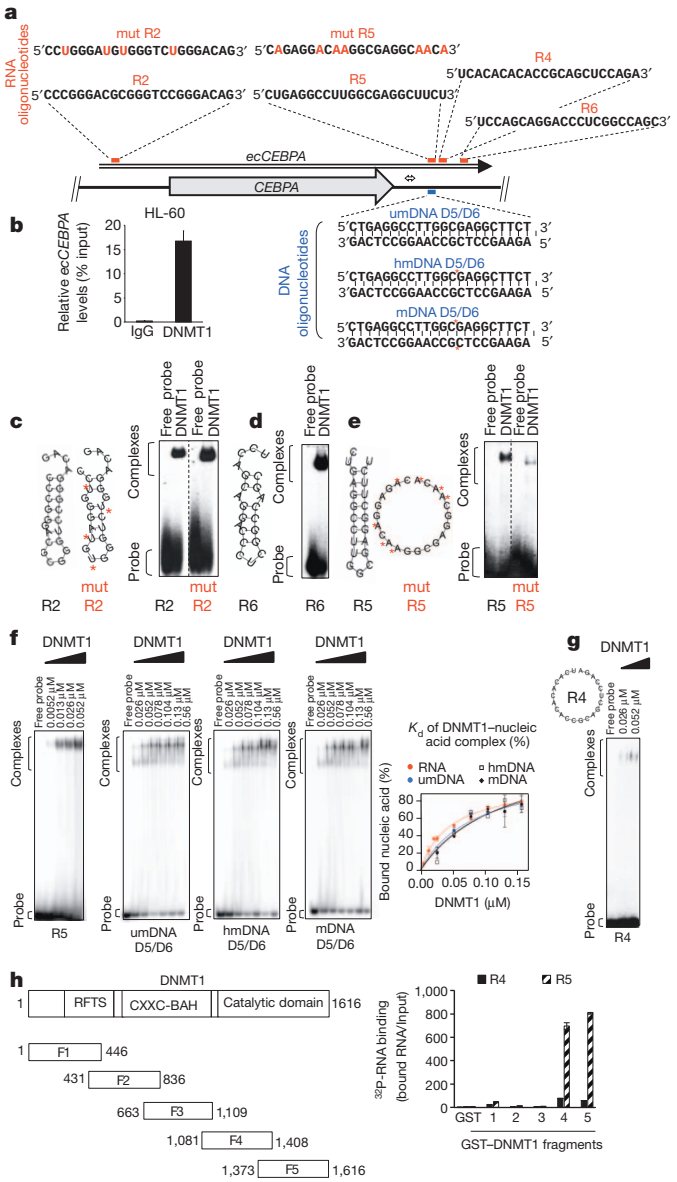

Figure 3 | *ecCEBPA*–DNMT1 interactions: DNMT1 binds to RNA with greater affinity than to DNA. **a**, Diagram showing position of qRT–PCR primers used in RIP (double-headed arrow) and RNA and DNA oligonucleotides used in EMSA and REMSA. Asterisks indicate position of methylated cytosines. **b**, *ecCEBPA* is immunoprecipitated with anti-DNMT1 antibody. qRT–PCR, bars indicate mean ± s.d. **c**, RNA–DNMT1 binding is not affected by the absence of CpG dinucleotides (right). Left and middle: RNA oligonucleotide R2 and its mutated form mut R2 (asterisks indicate cytosines substituted to uridines), both able to form stem–loop structures. **d**, RNA oligonucleotides able to form stem–loop structure bind DNMT1 (R6). **e**, R5 RNA oligonucleotide forming stem–loop structure (R5) has a greater DNMT1 affinity compared to mut R5, which is unable to fold into stem–loop (taken in equimolar amounts) at the same DNMT1 concentration. **f**, Left four panels, REMSA and EMSA performed with the fixed concentration of single-stranded RNA and double-stranded DNA oligonucleotides (1 nM) and increasing concentrations of DNMT1 protein. Right, nonlinear regression analysis of bound RNA/DNA versus DNMT1 concentrations. Error bars indicate s.d. from two independent experiments. **g**, REMSA showing that RNA oligonucleotide R4, which is unable to form stem–loop structure, displays lower DNMT1 affinity as compared to R5 (**f**, left) at the same DNMT1 concentrations. **h**, Left, schematic diagram showing the DNMT1 domains and the GST–DNMT1-isolated fragments (F1–F5). BAH, bromo-adjacent homology; RFTS, replication foci targeting sequence. Right, GST–DNMT1 pulldown assay demonstrating binding of the folded RNA oligonucleotide R5 to the catalytic domain of DNMT1.

DNMT1 and DNA at >0.026 μM DNMT1 (Fig. 3f), with a mean dissociation constant (K_d) for RNA of 0.045 μM and between 0.082 and 0.14 μM for DNA, indicating that RNA has a stronger affinity for the enzyme than DNA (Fig. 3f). Consistently, RNA unable to fold into stem–loop structures (R4; Fig. 3g) did not display the same affinity for DNMT1 as 'folded' RNA (R5; Fig. 3f, g), demonstrating that RNA secondary structure represents an essential feature of RNA–DNMT1 complex formation.

Finally, to assess which DNMT1 domain is required for the RNA binding, DNMT1–glutathione *S*-transferase (GST)-purified domains (Fig. 3h) were incubated with RNA oligonucleotides able or unable to fold into stem–loop structures (R5 and R4, respectively; Fig. 3e, g). The catalytic domain, including the target recognition domain[28] shared by both fragments F4 and F5, selectively bound the 'folded' RNA oligonucleotide (Fig. 3h). Next, we deleted the DNMT1 region including the sequence overlapping F4 and F5. Unfortunately, even minimal removal of the target recognition domain led to disruption of DNMT1 enzymatic activity (data not shown), making further refinement of the binding domain unfeasible.

Collectively, these data indicate that RNA can associate with DNMT1. This interaction is not contingent upon the presence of CpG dinucleotides, is not a trivial ion pairing, and is dependent upon certain RNA secondary-structure features. Importantly, DNMT1, through its catalytic domain, binds with higher affinity to folded RNA than to DNA.

Transcription interferes with DNMT1 activity

To examine whether newly synthesized transcripts could interfere with the ability of DNMT1 to methylate hmDNA, we performed a combined *in vitro* transcription–DNA methylation assay. A hmDNA segment (bottom-strand methylated) was engineered downstream of the T7 RNAP promoter (Supplementary Fig. 4a–h) and DNMT1 methylase activity was monitored in the presence and absence of transcription (Fig. 4a–d). In the absence of polymerase there was, as expected, increased DNA methylation of the upper strand (Fig. 4d–f and Supplementary Fig. 4i–j). By contrast, no changes in DNA methylation were observed in the presence of both polymerases and DNMT1 (Fig. 4c–f and Supplementary Fig. 4i–j). Standard *in vitro* DNA methylation assays confirmed the enzymatic impairment of DNMT1 mediated by ribo-oligonucleotides (Fig. 4g). Similarly, T7 RNA polymerase[29]-induced transcription in living cells led to a pronounced decrease in DNA methylation (Supplementary Information and Supplementary Fig. 4k–p).

Thus, RNA can complex with and affect DNMT enzymatic activity *in vitro*[30–32] and in living cells. These findings suggest that RNAs arising from methylation-sensitive genes and their promoters can regulate expression of the corresponding genes by interfering with DNA methylation.

RNA inhibition of DNA methylation is a global effect

Our observations suggested an inverse correlation between RNA–DNMT1 complexes and methylation of the *CEBPA* locus. Therefore, we sought to explore the extent of DNMT1–RNA association in other genomic loci with respect to DNA methylation and gene expression profiles. Complementary DNA libraries made of RNAs coimmunoprecipitated with anti-DNMT1 antibody (DNMT1 library) and IgG (control library) were tested for *ecCEBPA* enrichment ('quality control'; Supplementary Fig. 5a) and subsequently analysed by massively parallel sequencing[11]. Using 76-base paired-end sequencing, we produced a total of 30.25 and 26.95 million pair reads for DNMT1 and control libraries, respectively (detailed analysis described in Methods). All significant DNMT1 peaks (a total of 16,186; $P < 0.0001$; false discovery rate of 7.5%) were annotated with CEAS[33] build on RefSeq hg19 (Supplementary Fig. 5b). All DNMT1 peaks were also annotated using the known RNAs databases provided by HOMER[34] (Supplementary Fig. 5c). We focused on genomic regions encompassing the 3 kb upstream of the TSS and downstream to the transcription ending site of the annotated

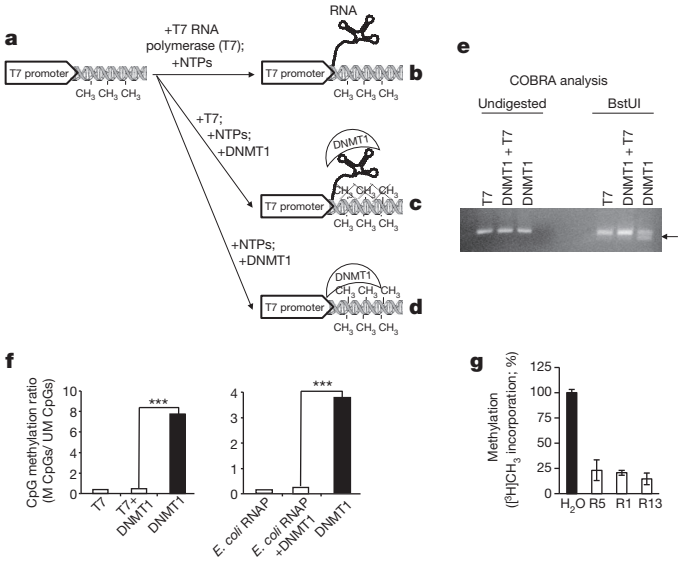

Figure 4 | Transcription impedes DNA methylation. **a–d**, Diagram showing the parallel *in vitro* transcription–methylation assays performed on a hemimethylated template containing the T7 promoter (Supplementary Fig. 4) with and without combinations of RNA polymerase, DNMT1, or both. NTPs, nucleotide triphosphates. **e**, DNMT1 exerts enzymatic activity only in the absence of transcription. Combined bisulphite restriction analysis assay (COBRA) analysis of methylation patterns acquired in reactions shown in **b–d**. **f**, DNA methylation changes are shown as the ratios of methylated to unmethylated CpGs in all clones analysed per construct ($n = 5$). The same effect was observed with two different RNA polymerases: T7 and sigma-saturated (σ70)-holoenzyme (*Escherichia coli* RNAP). DNA methylation changes were analysed by Fisher's exact test (***$P < 0.001$). **g**, *In vitro* DNMT1 assay demonstrating DNMT1 enzymatic impairment by RNA oligonucleotides. The assay was performed using *ecCEBPA*-related and -unrelated RNA oligonucleotides. Sequences and position of the ribo-oligonucleotides are shown in Fig. 3a and Supplementary Fig. 3. Error bars indicate mean ± s.d. ($n = 2$).

genes, referred to as 'gene loci'. We identified 6,042 gene loci containing one or more peaks from the DNMT1 library (Methods).

To confirm that DNMT1 RIP-seq peaks were associated with actual transcribed elements, RNA-seq was conducted on poly(A)$^-$ HL-60 RNA. In total, 375 million 76-bp paired-end reads were aligned to hg19 using TopHat2[35] and assembled using Cufflinks[36], and 14,077 (87.02%) of the specific DNMT1 peaks overlapped with a transcribed element from the RNA-seq assembly of the poly(A)$^-$ HL-60 RNA fraction. Thus, the vast majority of DNMT1-interacting RNAs (DiRs) were not polyadenylated.

In addition, we performed a similar analysis with total HL-60 RNA (300 million 76-bp paired-end reads). In total, 14,497 specific DNMT1 peaks (89.61%) were found to overlap with transcripts from the total HL-60 RNA-seq assembly. A merged assembly of the two RNA libraries validated a total of 15,238 (94.20%) DNMT1 RIP-seq peaks (Fig. 5a). These findings confirmed the existence of DiRs on a genome-wide level. Next, we assessed the linkage between genomic loci giving rise to DiRs, levels of genomic methylation by RRBS[24] and expression of the corresponding nearby genes by microarray analysis, performed on HL-60 cells. Within all 15,806 RRBS-covered loci, 10,973 loci were not covered by DNMT1-specific peaks (DNMT1-unbound group) and 4,833 loci were covered by DNMT1-specific peaks (DNMT1-bound group). These 4,833 loci represent the majority (79.99%) of all 6,042 gene loci identified by DNMT1 RIP-seq (Supplementary Fig. 5d).

Within DNMT1-bound and -unbound groups, genes were stratified according to expression and methylation levels (the latter computed as the mean of all CpG β-scores from −2 kb from the TSS to the end of the first intron). A negative correlation between DNMT1–RNA

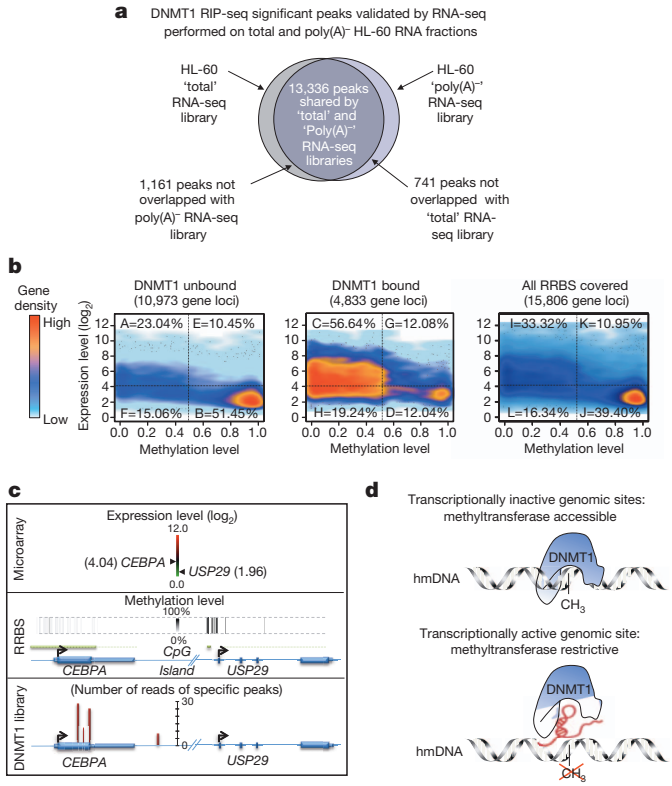

Figure 5 | Genome-wide alignment of DNMT1-bound and -unbound transcripts, DNA methylation and gene expression. **a**, Two-way Venn diagram showing DNMT1-specific peaks overlapping with transcribed elements identified in HL-60 total and poly(A)$^+$-depleted RNA-seq libraries. **b**, Cloud plots representing genes within DNMT1-unbound, -bound and all RRBS-covered groups stratified by DNA methylation and expression levels. All genes are presented in Supplementary Data 2. **c**, Examples of genes from the C (*CEBPA*) and B (*USP29*) clusters. Peaks are visualized using the SSIRs (site identification from short sequence reads)[48]. **d**, Model of DNMT1 sequestration. Top, DNMT1 can access transcriptionally inactive hemimethylated genomic regions. Bottom, DNMT1 cannot access transcriptionally active hemimethylated genomic regions.

association with gene locus methylation status was observed (Supplementary Fig. 5e).

Next, we clustered genes within both groups according to levels of expression and methylation. We defined genes as 'expressed' or 'low or not expressed' if the $\log_2$ score was above or below 4, respectively; and 'hypomethylated' or 'methylated' if the mean of all CpG scores was below or above 50%, respectively (Methods). This approach allowed us to identify four clusters within DNMT1-unbound, DNMT1-bound, and all RRBS-covered groups (Fig. 5b). Hypomethylated and expressed genes appeared to be predominant in the DNMT1-bound group (cluster C), accounting for 56.64%, whereas hypermethylated and low or unexpressed genes represented the 51.45% in the DNMT1-unbound group (cluster B). Moreover, the numbers of genes in clusters B (5,646 genes) and C (2,737 genes) were significantly higher than numbers of genes in clusters A, F, E (2,528, 1653, 1,146) and G, H, D (584, 930, 582), respectively ($P < 0.0001$). Examples of genes from clusters B and C are presented in Fig. 5c and Supplementary Fig. 5f–h). Furthermore, genes from cluster C belonged to a multiplicity of biological processes, indicating the diversity of DiRs (Supplementary Fig. 6a). Interestingly, ~60% of these Biological Process Gene Ontology (BP-GO) terms ($P \leq 0.01$) were shared with pyknon (non-random pattern of repeated elements)-related BP-GO[37]. This overlap is 71-fold higher than expected. Moreover, among all DNMT1 RIP-seq peaks, 46% carry at least one pyknon (Supplementary Fig. 6b) suggesting a potential relation between DiRs and pyknons.

Grouping of genes in clusters A, F, E, G, H and D could result from technical limitations of RRBS, contingent upon the genomic location of the restriction sites and the DNA library size selection[38], or these genes may be governed by yet another mechanism of transcriptional control.

In conclusion, we have generated the first comprehensive map cross-referencing DiRs to DNA methylation and gene expression. These data demonstrate that RNA–DNMT1 association is widespread and might modulate genomic DNA methylation (Fig. 5d).

Discussion

This study explores the role of a new class of RNAs: DNMT1-interacting RNAs. Using the *CEBPA* gene as a model, we show that mRNA transcription is accompanied by the production of an additional RNA species, *ecCEBPA*. In every instance studied, DNA methylation levels are inversely correlated with *ecCEBPA* levels, and the extent of DNA methylation is determined by the absence or presence of *ecCEBPA* (Fig. 2).

We demonstrate that *ecCEBPA* associates with DNMT1 and establish a functional link between *ecCEBPA* and *CEBPA* expression as through RNA–DNMT1 association.

We show that RNAs capable of adopting stem–loop structures exhibit the potential to associate with DNMT1, suggesting that the basis of this preferential interaction is recognition of RNA secondary structure (Fig. 3).

Importantly, we demonstrate that this type of RNA–DNMT1 association is not restricted to the *CEBPA* gene locus. We have globally identified RNA species associated with DNMT1 and their relationship to DNA methylation and gene expression. These alignments defined a large set of expressed unmethylated genes and a complementary set of silent methylated genes that could possibly be induced following expression of the DiR.

Our findings suggest a model of site-specific DNMT1 sequestration in which RNAs act as a shield, halting DNMT1 and thus modulating DNA genomic methylation at their site of transcription (Fig. 5d). Indeed, the loss of *CEBPA* locus methylation following overexpression of *ecCEBPA* does not support a model of trivial titration ('squelching'[39]) of DNMT1 but suggests a *cis*-regulatory role of the DiR. We propose a model wherein RNAs contain a locus-selective triplex/quadruplex[40]-forming part, the 'anchor', mooring the DNMT1–RNA complex to the locus, and a DNMT1-interacting part, the 'bait', a stem–loop-like-forming sequence serving to lure the DNMT1 into association. DNMT1 sequestration by RNA does resemble a competing mechanism described for other regulatory RNAs, for example, competing endogenous RNAs[15,16,41]. However, unlike competing endogenous RNAs, the *ecCEBPA* model also introduces the requirements for both functional and physical co-compartmentalization of the RNA, its parental locus, and DNMT1. Given the ability of DiRs to bind DNMT1, it is tempting to suggest that DiRs represent a novel class of RNA regulons[42].

Taken together, these data support the hypothesis that RNA participates in the establishment of genomic methylation patterns by interacting with DNMT1 and pave the way for the site-specific adjustments of aberrant DNA methylation.

METHODS SUMMARY

RIP-seq. Double-stranded cDNA from total RNA immunoprecipitated with DNMT1 antibody (Abcam) or IgG (Sigma Aldrich) was synthesized using the Just cDNA Double-Stranded cDNA Synthesis Kit (Agilent Technology) according to the manufacturer's instructions. cDNA libraries were paired-end sequenced on an Illumina GA IIx.

MassARRAY and RRBS. MassARRAY and RRBS were performed as described[43,44]. For RRBS, sequenced reads were mapped to the reference genome hg19 using RRBSmap[45] allowing two mismatches and gene loci were scored by Genome Bisulfite Sequencing Analyzer (GBSA)[46]. Differentially methylated domains were computed using the R/MethylKit package[47].

RNA-seq. Total and non-polyadenylated RNA were depleted of ribosomal RNA with a Ribo-ZeroTM Magnetic Gold Kit. Double-stranded cDNA libraries were

constructed using ScriptSeq v2 RNA-Seq Library Preparation Kit. Libraries were sequenced (1 per lane) on a Hi-Seq-2000 Illumina instrument.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 November 2011; accepted 21 August 2013.
Published online 9 October 2013.

1. Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nature Rev. Genet.* **3,** 415–428 (2002).
2. Robertson, K. D. DNA methylation, methyltransferases, and cancer. *Oncogene* **20,** 3139–3155 (2001).
3. Illingworth, R. S. *et al.* Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* **6,** e1001134 (2010).
4. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA* **103,** 1412–1417 (2006).
5. Tenen, D. G. Disruption of differentiation in human cancer: AML shows the way. *Nature Rev. Cancer* **3,** 89–101 (2003).
6. Tada, Y. *et al.* Epigenetic modulation of tumor suppressor CCAAT/enhancer binding protein α activity in lung cancer. *J. Natl. Cancer Inst.* **98,** 396–406 (2006).
7. Hackanson, B. *et al.* Epigenetic modification of CCAAT/enhancer binding protein α expression in acute myeloid leukemia. *Cancer Res.* **68,** 3142–3151 (2008).
8. Gupta, R. A. *et al.* Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* **464,** 1071–1076 (2010).
9. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142,** 409–419 (2010).
10. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129,** 1311–1323 (2007).
11. Zhao, J. *et al.* Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **40,** 939–953 (2010).
12. Nagano, T. *et al.* The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322,** 1717–1720 (2008).
13. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482,** 339–346 (2012).
14. Ebralidze, A., Wang, Y., Petkova, V., Ebralidse, K. & Junghans, R. P. RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. *Science* **303,** 383–387 (2004).
15. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465,** 1033–1038 (2010).
16. Tay, Y. *et al.* Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* **147,** 344–357 (2011).
17. Ebisuya, M., Yamamoto, T., Nakajima, M. & Nishida, E. Ripples from neighbouring transcription. *Nature Cell Biol.* **10,** 1106–1113 (2008).
18. Preker, P. *et al.* RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322,** 1851–1854 (2008).
19. Ebralidze, A. K. *et al. PU.1* expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared *cis*-regulatory element. *Genes Dev.* **22,** 2085–2092 (2008).
20. Oler, A. J. *et al.* Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nature Struct. Mol. Biol.* **17,** 620–628 (2010).
21. Listerman, I., Bledau, A. S., Grishina, I. & Neugebauer, K. M. Extragenic accumulation of RNA polymerase II enhances transcription by RNA polymerase III. *PLoS Genet.* **3,** e212 (2007).
22. Raha, D. *et al.* Close association of RNA polymerase II and many transcription factors with Pol III genes. *Proc. Natl Acad. Sci. USA* **107,** 3639–3644 (2010).
23. Wouters, B. J. *et al.* Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced *CEBPA* and mutations in *NOTCH1*. *Blood* **110,** 3706–3714 (2007).
24. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33,** 5868–5877 (2005).
25. Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71,** 865–873 (1992).
26. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31,** 3429–3431 (2003).
27. Pradhan, S. & Esteve, P. O. Allosteric activator domain of maintenance human DNA (cytosine-5) methyltransferase and its role in methylation spreading. *Biochemistry* **42,** 5321–5332 (2003).
28. Song, J., Rechkoblit, O., Bestor, T. H. & Patel, D. J. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331,** 1036–1040 (2011).
29. Fuerst, T. R., Niles, E. G., Studier, F. W. & Moss, B. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc. Natl Acad. Sci. USA* **83,** 8122–8126 (1986).
30. Svedružić, Z. M. Mammalian cytosine DNA methyltransferase Dnmt1: enzymatic mechanism, novel mechanism-based inhibitors, and RNA-directed DNA methylation. *Curr. Med. Chem.* **15,** 92–106 (2008).
31. Bolden, A., Ward, C., Siedlecki, J. A. & Weissbach, A. DNA methylation. Inhibition of *de novo* and maintenance methylation *in vitro* by RNA and synthetic polynucleotides. *J. Biol. Chem.* **259,** 12437–12443 (1984).
32. Bolden, A. H., Nalin, C. M., Ward, C. A., Poonian, M. S. & Weissbach, A. Primary DNA sequence determines sites of maintenance and *de novo* methylation by mammalian DNA methyltransferases. *Mol. Cell. Biol.* **6,** 1135–1140 (1986).
33. Shin, H., Liu, T., Manrai, A. K. & Liu, X. S. CEAS: *cis*-regulatory element annotation system. *Bioinformatics* **25,** 2605–2606 (2009).
34. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–589 (2010).
35. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
36. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7,** 562–578 (2012).
37. Tsirigos, A. & Rigoutsos, I. Human and mouse introns are linked to the same processes and functions through each genome's most frequent non-conserved motifs. *Nucleic Acids Res.* **36,** 3484–3493 (2008).
38. Bock, C. *et al.* Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nature Biotechnol.* **28,** 1106–1114 (2010).
39. Gill, G. & Ptashne, M. Negative effect of the transcriptional activator GAL4. *Nature* **334,** 721–724 (1988).
40. Frank-Kamenetskii, M. D. & Mirkin, S. M. Triplex DNA structures. *Annu. Rev. Biochem.* **64,** 65–95 (1995).
41. Karreth, F. A. *et al. In vivo* identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell* **147,** 382–395 (2011).
42. Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nature Rev. Genet.* **8,** 533–543 (2007).
43. Figueroa, M. E. *et al.* DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17,** 13–27 (2010).
44. Smith, Z. D., Gu, H., Bock, C., Gnirke, A. & Meissner, A. High-throughput bisulfite sequencing in mammalian genomes. *Methods* **48,** 226–232 (2009).
45. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10,** 232 (2009).
46. Benoukraf, T., Wongphayak, S., Hadi, L. H., Wu, M. & Soong, R. GBSA: a comprehensive software for analysing whole genome bisulfite sequencing data. *Nucleic Acids Res.* **41,** e55 (2013).
47. Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13,** R87 (2012).
48. Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of *in vivo* protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.* **36,** 5221–5231 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by grants CA118316, CA66996 and HL56745 from the National Institutes of Health (NIH) to D.G.T., the Italian Foundation for Cancer Research (FIRC) 'Leonino Fontana e Maria Lionello' fellowship, the NIH T32 HL007917-11A1 and the Società Italiana di Ematologia Sperimentale (SIES) 'Dr.Tito Bastianello' fellowship to A.D.R.; FAMRI CIA (103063) grant to A.K.E.; Fondazione Roma 'Progetto cellule staminali' to G.L. and F.D'A.; the American Italian Cancer Foundation Fellowship (AICF) to G.A.; Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)—grant no. 2011/11822-6—to L.L.D.F.P.; the National Research Foundation and the Singapore Ministry of Education under its Centres of Excellence initiative to M.W. and T.B.; the MSMT Navrat grant LK21307 to M.A.-J. S.P. and J.T. were supported by New England Biolabs. We thank R. White, D. Johnson, M. Frank-Kamenetskii, S. M. Mirkin, D. Gautheret, B. Tazon-Vega, C. Bonifer, C. Bock, M. T. Voso, J. J. Dunn (deceased) and R. A. M. Fouchier for helpful advice and reagents; I. Rigoutsos for providing the latest released pyknons database; all the members of the Tenen Laboratory; P. Tan and T. S. Ting from the Genome Institute of Singapore; R. Soong from the Cancer Science Institute Translational Interface; J. LaVecchio and G. Buruzula from the Harvard Stem Cell Institute/Joslin Diabetes Center flow cytometry facility; and F. Hyde from Epicentre-Illumina. This research is supported by the Singapore Ministry of Health's National Medical Research Council under its Singapore Translational Research (STaR) Investigator Award (D.G.T.).

Author Contributions D.G.T. supervised the project; A.D.R., A.K.E. and D.G.T. conceived and designed the study; A.D.R., A.K.E., G.A., P.Z., M.A.-J., F.D'A., S.P., L.L.D.F.P. and J.T. performed experiments; M.W. performed sequencing and microarray experiments; T.B. and L.A.G. analysed RIP-seq, RNA-seq, RRBS and microarray data; M.E.F. and A.M. performed the MassARRAY experiment and assisted in the analysis; A.D.R., A.K.E., G.L., K.K.E., J.L.R. and D.G.T. wrote the paper.

Author Information Sequencing and microarray data sets are available for download at Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/projects/geo/) under the following accession numbers: microarray expression, GSE32153; RIP-seq, GSE32162; RRBS, GSE32168; and RNA-seq, GSE41279. The accession number for the project is GSE32260. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.G.T. (daniel.tenen@nus.edu.sg).

METHODS

Cell culture. All cell lines were obtained from ATCC and grown according to the manufacturer's instructions in the absence of antibiotics. The DNMT1 hypomorphic HCT116 cell line and its wild-type counterpart HCT116 were grown in McCoy's 5A modified medium supplemented with 10% FCS.

RNA isolation and northern blot analysis. Total RNA isolation, electrophoresis, transfer and hybridization were carried out as described[49]. Cytoplasmic RNA was isolated with the Paris kit (Ambion) according to the manufacturer's recommendations. For the preparation of nuclear RNAs we used a method derived from protocols of nuclei isolation[50], with minor modifications. In brief, equal amounts of viable cells (~50 million) were washed with ice-cold PBS supplemented with 5 mM vanadyl complex, 1 mM phenylmethylsulphonyl fluoride (PMSF) and re-suspended in the ice-cold lysis buffer: 1 × buffer A (10 mM HEPES-NaOH, pH 7.6, 25 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 1 mM EDTA, 2 mM Na butyrate), 1.25 M sucrose, 10% glycerol, 5 mg ml^{-1} BSA, 0.5% NP-40, freshly supplemented with protease inhibitors (2 mM leupeptin, add as ×400; 2 mM pepstatin, add as ×400; 100 mM benzamidine, add as ×400; a protease inhibitor cocktail (Roche Applied Science), 1 tablet per 375 µl H_2O, add as ×100; 100 mM PMSF, add as ×100), 2 mM vanadyl complex (New England Biolabs) and 20 units per ml RNase inhibitor (RNAguard; Amersham Biosciences). Samples were incubated at 0 °C for ~10 min and passed through 40 up-and-down strokes in a Dounce homogenizer (10 with pestle A and 30 with pestle B). The pelleted nuclei were re-suspended in 0.5 ml lysis buffer and diluted with 2.25 ml dilution buffer (2.13 ml 'cushion' buffer plus 0.12 ml 0.1 g ml^{-1} BSA), freshly supplemented with protease inhibitors and overlaid onto 2 ml 'cushions' (20 ml cushion buffer consists of 15 ml double-deionized (dd)H_2O, 15 ml 20× buffer A, 30 ml glycerol, 240 ml 2.5 M sucrose; freshly supplemented with protease inhibitors) into one SW 55 Ti tube and centrifuged at 100,000*g* for 60 min at 4 °C. The pelleted nuclei were re-suspended in 1 ml storage buffer (1.75 ml ddH_2O, 2 ml glycerol, 0.2 ml 20× buffer A), freshly supplemented with protease inhibitors. Nuclear RNAs were extracted as described[50]. All total, cytoplasmic and nuclear RNA samples used in this study were treated with DNase I (10 U of DNase I per 3 µg of total RNA; 37 °C for 1 h; in the presence of RNase inhibitor). After DNase I treatment, RNA samples were extracted with acidic phenol (pH 4.3) to eliminate any remaining traces of DNA. Polyadenylated and non-polyadenylated RNA fractions were selected with the MicroPoly(A) Purist purification kit (Ambion). cDNA syntheses were performed with Random Primers (Invitrogen) with Transcriptor Reverse Transcriptase (Roche Applied Science) according to the manufacturer's recommendation. cDNA was purified with a High Pure PCR Product Purification Kit (Roche Applied Science).

qRT–PCR. SYBR green reactions were performed using iQ Sybr Green supermix (Bio-Rad); PCR conditions: 95 °C (10 min) followed by 40 cycles of 95 °C (15 s) and 60 °C (1 min) and 72 °C (1 min). TaqMan analysis was performed using Hotstart Probe One-step qRT–PCR master mix (USB); PCR conditions: 50 °C (10 min), 95 °C (2 min), followed by 40 cycles of 95 °C (15 s) and 60 °C (60 s). Primers and probes are presented in Supplementary Methods. qRT–PCR primer set for the *CEBPA* mRNA is located in the coding region (black double-headed arrow in Fig. 1a) and after the poly(A) signal for *ecCEBPA* (white double-headed arrow Fig. 1a).

Primer extension and 5′/3′ RACE. cDNA from the HL-60 cell line was synthesized as described above and run in alkaline conditions[51]. Southern blot transfer and hybridization with oligonucleotide AL16 were performed as reported previously[51]. Oligonucleotide sequences are shown in Supplementary Methods. 5′/3′ RACE was performed on two myeloid cell lines—HL-60 and U937—using the Exact START Eukaryotic mRNA 5′- & 3′-RACE Kit (Epicentre) according to the manufacturer's instructions. See Supplementary Methods for primer sequences.

Double thymidine block (early S phase block). HL-60 cells were grown to 70–80% confluence, washed twice with 1 × PBS and cultured in DMEM (10% FCS) plus 2.5 mM thymidine for 18 h (first block). Thymidine was washed out with 1 × PBS and cells were grown in DMEM (10% FCS). After 8 h cells were cultured in the presence of thymidine for 18 h (second block) and then released as described. Synchrony was monitored by flow cytometry analysis (propidium iodide staining) using a LSRII flow cytometer (BD Biosciences) at the Harvard Stem Cell Institute/Beth Israel Deaconess Medical Center flow cytometry facility.

DRB, ML-60218, α-amanitin and actinomycin D treatment. After release from double thymidine block, HL-60 cells were treated with 100 µM 5,6-dichlorobenzimidazole 1-β-D-ribofuranoside[52] (DRB; Sigma-Aldrich) for 1, 2 and 3 h. HL-60 cells were treated with 12.5, 25, 50 or 100 µM ML-60218 (refs 53, 54; Calbiochem) for 24 h. HL-60 cells were treated with 5, 23, 50, 75, 100 or 150 µg ml^{-1} α-amanitin (Sigma-Aldrich) for 14 h. Synchronized and unsynchronized U937 cells were treated with ML-60218 at 100 µM and actinomycin D (Sigma-Aldrich) at 150 µM for 7 h. Total RNA was collected as described above and expression levels of *CEBPA*, *ecCEBPA* and 5S were measured by TaqMan qRT–PCR.

Nuclear chromatin immunoprecipitation and RIP. Chromatin immunoprecipitation (ChIP) was performed as described previously[19]. Fold enrichment was calculated using the formula $2^{(-\Delta\Delta Ct(ChIP/non\text{-}immune\ serum))}$. Antibodies used for ChIP are listed in Supplementary Table 1.

RIP was performed as described in ref. 14. Day 1: crosslinked nuclei were collected as follows: 60 × 10^6 HL-60 cells were crosslinked with 1% formaldehyde (HCHO; formaldehyde solution, freshly made: 50 mM HEPES-KOH, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 11% formaldehyde) for 10 min at room temperature (21 °C). Crosslinking was stopped by adding one-tenth the volume of 2.66 M glycine, kept for 5 min at room temperature and 10 min on ice. Cell pellets were washed twice with ice-cold PBS (freshly supplemented with 1 mM PMSF). Cell pellets were re-suspended in cell lysis buffer (volume 4 ml): 1 × buffer (10 mM Tris, pH 7.4, 10 mM NaCl, 0.5% NP-40, freshly supplemented with protease inhibitors (protease inhibitors cocktail: Roche Applied Science, 1 tablet per 375 µl H_2O; add as ×100), 1 mM PMSF and 2 mM vanadyl complex (NEB). Cells were incubated at 0 °C for 10–15 min and homogenized by Dounce (10 strokes with pestle A and 40 strokes with pestle B). Nuclei were recovered by centrifugation at 2,000 r.p.m. for 10 min at 4 °C. Nuclei were re-suspended in 3 ml 1 × re-suspension buffer (50 mM HEPES-NaOH, pH 7.4, 10 mM $MgCl_2$) supplemented with 1 mM PMSF and 2 mM vanadyl complex. DNase treatment (250 U ml^{-1}) was performed for 30 min at 37 °C, and EDTA (final concentration 20 mM) was added to stop the reaction. Re-suspended nuclei were sonicated once for 20 s (1 pulse every 3 s) at 30% amplitude (Branson Digital Sonifer). Immunoprecipitation for RIP was performed as follows: before preclearing, the sample was adjusted to 1% Triton X-100, 0.1% sodium deoxycholate, 0.01% SDS, 140 mM NaCl, protease inhibitors, 2 mM vanadyl complex and 1 mM PMSF to facilitate solubilization. Preclearing step: ~50 µl magnetic beads (Protein A or G Magnetic Beads, NEB) were added to the sample and incubation was carried out for 1 h on a rocking platform at 4 °C. Beads were removed in a magnetic field. The sample was divided into three aliquots: (1) antibody of interest: either DNMT1 antibody (Abcam) or anti-cap antibody (Anti-m_3G-/m^7G-cap; Synaptic Systems); (2) pre-immune serum: IgG (Sigma-Aldrich); (3) no antibody, no serum (input). 5 µg antibody or pre-immune serum was added to the respective aliquot and incubation performed on a rocking platform overnight at 4 °C. Input was stored at −20 °C after addition of SDS to 2% final concentration. Day 2: 200 µl of protein A-coated superparamagnetic beads (enough to bind 8 µg IgG) were added to the samples and incubated on a rocking platform for 1 h at 4 °C. Six washes were performed with immunoprecipitation buffer (150 mM NaCl, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 1 mM EGTA, pH 8.0, 1% Triton X-100, 0.5% NP-40 freshly supplemented with 0. 2 mM vanadyl complex and 0.2 mM PMSF) in a magnetic field. Proteinase K treatment to release DNA/RNA into solution and to reverse HCHO crosslinking was performed in 200 µl of: 100 mM Tris-HCl, pH 7.4; 0.5% SDS for the immunoprecipitated samples and in parallel for the input; proteinase K, 500 µg ml^{-1} at 56 °C overnight. Day 3: beads were removed in magnetic field. Phenol (pH 4.3) extraction was performed after addition of NaCl (0.2 M final concentration). Ethanol precipitation (in the presence of glycogen): 3 h at −20 °C. The pellet was dissolved in 180 µl H_2O, heated at 75 °C for 3 min and immediately chilled on ice. Samples were treated with DNase I (250 U ml^{-1}) in the presence of RNase inhibitor 300 U ml^{-1} in ×1 buffer no. 2 (NEB) at 37 °C for 30 min. Phenol (pH 4.3) extraction and ethanol precipitation were repeated. The RNA pellet was dissolved in 50 µl H_2O.

Tobacco acid pyrophosphatase and 5′-phosphate-dependent-exonuclease (Terminator) treatment. Equal amounts of RNA collected from HL-60 cells (as described above) were digested with tobacco acid pyrophosphatase (TAP; Epicentre), Terminator (Epicentre) or no enzyme according to the manufacturer's instructions. RNA was re-extracted in presence of glycogen (Ambion) with acidic phenol (pH 4.3), precipitated with ethanol and re-suspended in ddH_2O.

ecCEBPA, *CEBPA* and *18S* expression levels were measured by qRT–PCR using the TaqMan primer sets indicated in Supplementary Methods.

Downregulation of *ecCEBPA*. Three different shRNAs targeting human *ecCEBPA* and scrambled control were designed according to Dharmacon software and cloned into the lentiviral vector pLKO.1 (Sigma-Aldrich), which has a puromycin selection marker. shRNA sequences are shown in Supplementary Methods. Lentiviral particles were produced as described previously[55]. HEK293T cells were co-transfected with either empty vector or the pLKO-shRNA vector and Gag-Pol and Env constructs using Lipofectamine TM 2000 (Invitrogen) according to the manufacturer's recommendation. Virus-containing supernatants were collected 48 and 72 h after transfection and concentrated using a Centricon Plus-70 molecular weight cut-off column (Millipore). Lentiviral transduction was performed in the presence of hexadimethrine bromide (final concentration 8 µg ml^{-1}) in the human myeloid cell line U937. Puromycin (2 µg ml^{-1}) was added to the cultures 2 days after infection. Resistant clones were selected and screened for downregulation of *ecCEBPA* by qRT–PCR.

Upregulation of *ecCEBPA*. The 3′ *ecCEBPA* region (R1), the upstream and downstream *ecCEBPA* regions, and the unrelated genomic region (UR, see Supplementary Methods) were cloned into the pBabe retrovirus vector harbouring a puromycin selection marker (Addgene; plasmid 1764). Oligonucleotide sequences used to amplify both regions are shown in Supplementary Methods. K562 cells were transfected with the Amaxa Cell Line Nucleofector Kit V, Program T-003. Puromycin (2 μg ml^{-1}) was added to the cultures 2 days after transfection. Resistant clones were selected and screened for upregulation of *ecCEBPA* and the UR by northern blot analysis.

Bisulphite treatment, COBRA and bisulphite sequencing. The methylation profile of the *CEBPA* gene locus was performed by bisulphite sequencing as described previously[56]. In brief, 1 μg of genomic DNA was bisulphite-converted by using the EZ DNA Methylation kit (Zymo Research). Primers and PCR conditions for bisulphite sequencing and COBRA are summarized in Supplementary Methods. For COBRA, PCR products were gel-purified and incubated with BstUI at 60 °C for 3 h. The digested DNA was then separated on a 3.5% agarose gel and stained with ethidium bromide. For bisulphite sequencing, PCR products were gel-purified (Qiagen) and cloned into the pGEM-T Easy Vector System (Promega). Sequencing results were analysed using BiQ analyser software[57]. Samples with conversion rate $<$ 90% and sequences identity $<$ 70% as well as clonal variants were excluded from our analysis[55]. The minimum number of clones for each sequenced condition was ≥6. Primer sequences are shown in Supplementary Methods.

5-aza-CR treatment. K562 cells were treated with 5-aza-CR[58] (Sigma-Aldrich) according to the manufacturer's instructions. Medium was refreshed every 48 h. RNA (for RT–PCR) and genomic DNA (for bisulphite sequencing) were isolated after 7 days of treatment.

MassARRAY. Quantitative DNA methylation analysis using the MassARRAY technique was performed by Sequenom as described previously[43]. In brief, 1 μg of genomic DNA was converted with sodium bisulphite using the EZ DNA methylation kit (Zymo Research), PCR amplified, *in vitro* transcribed and then cleaved by RNase A. The samples were then quantitatively tested for their DNA methylation status using matrix-assisted laser desorption ionization-time of flight mass spectrometry. The samples were desalted and spotted on a 384-pad SpectroCHIP (Sequenom) using a MassARRAY nanodispenser (Samsung), followed by spectral acquisition on a MassARRAY Analyzer Compact MALDI-TOF MS (Sequenom). The resulting methylation calls were obtained using the EpiTyper software v1.0 (Sequenom) to generate quantitative results for each CpG site or an aggregate of multiple CpG sites. The methylation levels of aggregated multiple CpGs were calculated as the mean of each CpGs methylation value and presented as a percentage. Primer sequences are provided in Supplementary Data 1.

Transfection of human DNMT1–haemagglutinin tag construct and western blot analysis. The human DNMT1–haemagglutinin-tagged cloned into the expression vector pCDNA3.1 was a kind gift from S. Baylin. HEK293T cells were transfected using Lipofectamine 2000, and 2 days later collected for western blot analysis. Single-cell suspensions were lysed with modified radioimmunoprecipitation assay buffer, and whole-cell lysates separated on 6% SDS–PAGE gels. Immunoblots were stained with the following primary antibodies: DNMT1 (1:5,000, Abcam) and HSP90 (1:2,000, BD Biosciences). All secondary antibodies were horseradish peroxidase (HRP)-conjugated (Santa Cruz) and diluted 1:5,000 for rabbit-HRP, and 1:3,000 for mouse-HRP. Western blot analysis for the HCT116 hypomorphic and wild-type cell lines were similarly performed.

T7 polymerase-induced transcription. The T7 expression system is based on technology developed at Brookhaven National Laboratory under contract with the US Department of Energy and is the subject of patents and pending applications. Full information may be obtained from the Office of Intellectual Property and Sponsored Research, Brookhaven National Laboratory. Maps of T7 polymerase constructs are presented in ref. 59. In brief, the murine RAW 264.7 cell line was stably transfected with a construct carrying the human genomic segment under T7 promoter control (derived from pBlueScript plasmid; Agilent). After selection in G418, individual clones were transfected with T7 polymerase-expressing mammalian constructs and were tested by COBRA for genomic methylation.

EMSAs and K_d determination. DNA and RNA oligonucleotides (15 pmol) were end-labelled with [γ-^{32}P]ATP (Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs). Reactions were incubated at 37 °C for 1 h and then passed through G-25 spin columns (GE Healthcare) according to the manufacturer's instructions to remove unincorporated radioactivity. Labelled samples were gel-purified on 10% polyacrylamide gels. Binding reactions were carried out in 10-μl volumes in the following buffer: 5 mM Tris, pH 7.4, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 3% (v/v) glycerol, 100 mM NaCl. Various amounts (0.021–0.156 μM) of purified DNMT1 protein (BPS Bioscience) were incubated with 1.1 nM of ^{32}P-labelled double-stranded DNA and single/double-stranded RNAs. In the competitive assay, a fixed amount of protein and increasing amounts of competitors (double-stranded DNA or poly(dI-dC)) were used. All reactions were assembled on ice and incubated at room temperature. Samples were separated on 6% native polyacrylamide gels (0.5 × Tris/Borate/EDTA (TBE); 4 °C; ~3 h at 140 V). Gels were dried and exposed to X-ray film and/or PhosphoroImager screens. Quantification was done with ImageQuant software. For affinity assays, the per cent shifted species was determined as follows: the migration of the labelled DNA in this reaction was defined as zero per cent shifted and the ratio of the PhosphoroImager counts in the area of the lane above this band to the total counts in the lane was defined as background and subtracted from all other lanes. This band represented total input. Subsequent lanes containing DNMT1–nucleic acid complexes were treated identically, and the percentage complex formation was calculated as follows: [% bound complex = (1 − ((unbound − background)/(input − background))]. All experiments contained a control reaction lacking DNMT1. The percentage complex formation was plotted as a function of DNMT1 concentration using nonlinear regression analysis performed with Prism 4.0a software. RNA and DNA oligonucleotides used in EMSA are shown in Fig. 3a and Supplementary Fig. 3d and listed in Supplementary Methods.

***ecCEBPA* binding to GST–DNMT1 fragments.** GST and GST–DNMT1 fragments were expressed and purified by glutathione sepharose affinity beads (GE Healthcare Life Sciences) as described previously[60]. Protein concentrations for the recombinant GST and GST–DNMT1 fragments were determined by gel electrophoresis and subsequent Coomassie staining and densitometry. ^{32}P end-labelling of *ecCEBPA* oligonucleotides was carried out at 37 °C for 30 min in a total volume of 50 μl. The reaction contained 50 pmol *ecCEBPA*, adenosine 5′-triphosphate, [γ-^{32}P] (specific activity 3,000 Ci mmol^{-1}, Perkin Elmer) and 20 U of T4 kinase (New England Biolabs) mixed in assay buffer (70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM DTT, pH 7.6). The labelled *ecCEBPA*-^{32}P was purified with illustraMicroSpin G-25 Columns (GE Healthcare Life Sciences) according to manufacturer's specifications. Equal amounts of the GST and GST–DNMT1 fragments were mixed with 5 μl *ecCEBPA*-^{32}P, in duplicate, and incubated at 37 °C for 10 min in a total volume of 25 μl of reaction buffer (50 mM Tris-HCl, 1 mM DTT, 1 mM EDTA, 5% glycerol, pH 7.8). The sepharose beads were then washed four times in phosphate-buffered solution and placed in 3 ml of scintillation fluid and bound ^{32}P was measured for 1 min. All measurements were normalized to ^{32}P readings for the corresponding input ^{32}P-*ecCEBPA*.

***In vitro* transcription–methylation assay.** *In vitro* transcription–methylation assays were performed on hmDNA (described in Supplementary Information, legend to Supplementary Fig. 4) in the presence or absence of 5 U of human DNMT1 enzyme (New England Biolabs) and 5 U of T7 RNA polymerase (Promega) or 5 U of *E. coli* RNA polymerase sigma-saturated holoenzyme (Epicentre). Reactions were performed in DNMT1 buffer according to the manufacturer's recommendations supplemented with rNTPs (ribonucleotide triphosphates) and 1.25 mM $MgCl_2$, including the 'DNMT1 only' reaction. This predetermined concentration of Mg^{2+} cations is high enough to sustain activity of RNA polymerases and low enough not to inhibit DNMT1 activity.

DNMT1 *in vitro* methylation assay. DNMT1 enzymatic assays were carried out in duplicate at 37 °C for 30 min in a total volume of 25 μl. The reaction contained *S*-adenosyl-L-[methyl-^{3}H]methionine (AdoMet) (specific activity 18 Ci mmol^{-1}, Perkin Elmer), 200 ng of substrate DNA, recombinant DNMT1 enzyme (2.5 pmol) and *ecCEBPA* (2.5 pmol) mixed in assay buffer (50 mM Tris-HCl, 1 mM DTT, 1 mM EDTA, 5% glycerol, pH 7.8). Methyltransferase reactions were snap-frozen in an ethanol-dry ice bath. The entire reaction volume (25 μl) was spotted on 2.5-cm DE81 membranes (GE Healthcare). These membranes were processed by washes in 3 × 1 ml of 0.2 M ammonium bicarbonate, 3 × 1 ml water and 3 × 1 ml ethanol. Processed membranes were air-dried, placed in 3 ml of scintillation fluid and tritium incorporation was measured for 1 min. Background subtraction (no DNA substrate) was performed for all experimental sample counts.

RIP-seq. Total RNA immunoprecipitated with DNMT1 antibody (Abcam) or IgG (Sigma-Aldrich) was processed for sequencing as described in ref. 61 with some modifications. Double-stranded cDNA was synthesized using the Just cDNA Double-Stranded cDNA Synthesis Kit (Agilent Technology) according to the manufacturer's instructions. Illumina sequencing libraries were constructed from these cDNA using a ChIP-seq sample preparation kit (Illumina) with minor modifications. Illumina paired-end adaptor and PCR primers were used to replace the single-read adaptor and primers in the kit. Constructed libraries were subjected to a final size-selection step on 10% Novex TBE gels (Invitrogen). DNA fragments of 175–200 base pairs (bp) were excised from a SYBR-green-stained gel. DNA was recovered from the gel and quantified following Illumina's qPCR quantification protocol. Paired-end sequencing of these libraries was then performed on an Illumina GA IIx to achieve 2 × 76-bp reads. Paired-end reads were trimmed to 50 bp and aligned to the reference genome hg19 using BWA[62] with the following parameters: bwa aln -o 1 -l 25 -k 2; bwa sampe -o 200. To estimate a normalization factor (alpha) between the immunoprecipitations, the genome was divided into course bins (10 kb) and reads were counted for DNMT1 RIP and IgG control in

each bin. A linear regression was fitted across all non-zero bins and the slope of the regression was used as a scaling factor (alpha) to normalize the RIP and control libraries. To identify distinct regions specifically bound by DNMT1, all downstream analyses were conducted on a set of regions derived by aggregating overlapping DNMT1 RIP reads into contiguous intervals. Each DNMT1 interval was tested for significance by comparing the number of reads within the interval with the number of reads in the same region of the IgG control, multiplied by the previously estimated scaling factor, alpha (exact binomial test, $P = 0.5$). Multiple tests were corrected by Benjamini–Hochberg. In total, 16,186 intervals (representing the start and end boundaries of contiguous, overlapping reads) were determined to be significantly enriched in the DNMT1 RIP as compared to the IgG control ($P < 0.0001$; $q < 0.0001$). A false discovery rate of 7.5% was determined by determining the number of significantly enriched intervals in the IgG immunoprecipitate using DNMT1 as a control. Significantly enriched DNMT1 intervals have a mean length of 347 bp and a median of 67 reads per interval. Every peak represents an interval with a 'height' value: the sum of all reads within an interval. All peaks were annotated with CEAS[33] build on RefSeq hg19. All DNMT1 RIP-seq peaks were also annotated using the HOMER pipeline (version 4.2)[34] which provide a comprehensive RNAs database (coding and non-coding, including miRNA, small nucleolar RNA, ribosomal RNA, small nuclear RNA, transfer RNA, etc.).

A peak was considered as belonging to a gene if located in the gene body or 3 kb up- or downstream the gene (gene loci). Altogether, 6,042 gene loci were covered by a least one significant RIP-seq peak.

RNA-seq. Total RNA was extracted with TRI Reagent (MRC). RNA samples were treated with 10 U of DNase I (Roche Applied Science) per 3 μg of total RNA at 37 °C for 1 h in the presence of RNaseA inhibitor. Non-polyadenylated RNA fractions were selected with the MicroPoly(A) Purist purification kit (Ambion). Total and non-polyadenylated RNA were depleted of rRNA with Ribo-ZeroTM Magnetic Gold Kit (Epicentre). Double-stranded cDNA libraries were constructed using ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicentre) followed by duplex specific nuclease (DSN) normalization (Evrogen). DSN-treated libraries were subjected to final size selection in 3% agarose gel. 250–500-bp fragments were excised and recovered using the Qiaquick Gel Extraction Kit (Qiagen). Libraries were sequenced (1 per lane) on a Hi-Seq-2000 Illumina instrument. Raw read sequences were deposited in GEO (accession number GSE32153). 2.96×10^8 reads from the HL-60 and 3.75×10^8 76-bp paired-end reads from the HL-60 poly(A)$^+$-depleted RNA were aligned to the human genome hg19 (UCSC release) using Tophat2 (ref. 63). Aligned reads were assembled into individual full-length transcripts using Cufflinks v2.0.2 (ref. 63) and a merged assembly was created from the two assemblies and additionally, all level-1 and -2 transcripts from the Gencode v11 catalogue[64] using Cuffmerge[36]. To confirm the transcription of the significant DNMT1 RIP-seq peaks, we overlapped the peak intervals with the RIP-seq assemblies using the BEDtools[60] intersect BED utility.

RRBS. RRBS was performed as described[44]. In brief, high-quality genomic DNA was isolated from the myeloid cell line HL-60. DNA was digested with MspI (NEB), a methylation-insensitive enzyme that cuts C^CGG. Digested DNA was size selected on a 4% NuSieve 3:1 agarose gel (Lonza). For each sample, two slices containing DNA fragments of 40–120 bp and 120–220 bp, respectively, were excised from the unstained preparative portion of the gel. These two size fractions were kept apart throughout the procedure and mixed 1:2 for the final sequencing. Pre-annealed Illumina adaptors containing 5′-methylcytosine instead of cytosine were ligated to size-selected MspI fragments. Adaptor-ligated fragments were bisulphite-treated using the EZ DNA Methylation kit (Zymo Research). The products were PCR amplified, size selected and sequenced on the Illumina GAIIx at a reading length of 36 bp. Sequencing reads were mapped to the reference genome hg19 using RRBSmap[45] allowing two mismatches. Reads from replicates were merged and processed as described previously[46]. We considered only CpG located in regions with a depth of coverage greater than three reads. The β-score of CpG methylation in a given position is the ratio of methylated CpGs within the total number of CpGs through all reads. The level of gene methylation is the mean of all CpG β-scores within −2 kb from the TSS to the end of first intron; for intronless genes, the entire gene body was considered. Genes with less than three sequenced CpGs in the promoter or less than three sequenced CpGs in the first exon–intron were excluded.

For RRBS in R1- and the UR-overexpressing cells bisulphite sequenced UR–R1 genomes were binned at 100-bp intervals using the R-Bioconductor/methylKit '*tileMethylCounts*' function (http://code.google.com/p/methylkit)[47]. The level of differential methylation was computed by comparing all sequenced CpG sites within the overlapping bins between the two samples (R1 and UR). The significant differentially methylated bins were obtained using the Fisher's test from the R-Bioconductor/methylKit '*calculateDiffMeth*' function (q value < 0.01 and methylation difference $\geq 50\%$). A gene was considered differentially methylated if the region including the promoter (−2 kb from the TSS) and first exon was overlapped by at least one significant differentially methylated bin. In total, 11,844 promoter/first exon regions were analysed.

RNA expression profiling. RNA isolated from HL-60 cells was used for sample amplification and labelling using the Whole Transcriptome assay reagent kits from Affymetrix. 10 μg of labelled RNA was hybridized on Affymetrix GeneChip Human Gene 1.0 ST array. Hybridization, washing, staining and scanning were carried out as recommended by the manufacturer. Each hybridization was performed in triplicate. Washes and staining were performed through the Fluidics Station 400 and the GeneChip Scanner 3000 (Affymetrix) was used to measure the fluorescence intensity emitted by the labelled target. Raw data processing was performed using the Affymetrix GeneChip Operating Software (GCOS). Microarrays were RMA normalized using 'affy', an R-Bioconductor library[65]. *CEBPA* expression was used as a threshold to define expressing ($\log_2$ score above 4) and non-expressing ($\log_2$ score below 4) genes for further analysis.

GO and pyknons comparison. GO analysis was performed with DAVID[66]. We focused our analysis on biological process annotations. GO enrichment was scored using the Benjamini–Hochberg-corrected *P* value. DNMT1 RIP-seq-specific peaks were compared to the human pyknons database released in January 2013 (https://cm.jefferson.edu/tools_and_downloads/pyknons.html).

Data integration. We used the Ref-seq transcripts database built on hg19 (UCSC release) as a genome annotation reference for Rip-seq, RRBS and microarray expression experiments. We selected only the longest transcripts. Accordingly, the number of 40,857 RefSeq IDs was reduced to 23,250 transcript IDs. Then, we annotated all RIP-seq peaks against the gene loci, which includes exonic, intronic and UTR regions plus 3 kb upstream of the TSS and 3 kb downstream of the transcription end site regions. We identified 6,042 gene loci with DNMT1 RIP-seq peaks and 17,208 gene loci without DNMT1 RIP-seq peaks. Finally, we focused our study on gene loci covered by the RRBS. We identified 4,833 gene loci with DNMT1 RIP-seq peaks and covered by RRBS and 10,973 gene loci without DNMT1 RIP-seq peaks and covered by RRBS. We plotted genes within each group against expression and methylation profile. Using *CEBPA* levels of expression as a cut-off threshold, we defined genes as 'expressed' or 'low or not expressed' if the $\log_2$ score was above or below 4, respectively, and as 'hypomethylated' and 'methylated' genes with mean of all CpG scores below and above 50%, respectively. We identified by this approach four clusters in each group. In DNMT1-unbound group clusters: A (expressed, hypomethylated genes; 23.04%), B (low or not expressed, methylated genes; 51.45%), E (expressed, methylated genes; 10.45%) and F (low or not expressed, hypomethylated genes; 15.06%). In DNMT1-bound group clusters: C (expressed, hypomethylated genes; 56.64%), D (low or not expressed, methylated genes; 12.04%), G (expressed, methylated genes; 12.08%), H (low or not expressed, hypomethylated genes; 19.24%). In all RRBS-covered group clusters: I (expressed, hypomethylated genes; 33.32%), J (low or not expressed, methylated genes; 39.40%), K (expressed, methylated genes; 10.95%) and L (not or low expressed, hypomethylated genes; 16.34%).

Statistical analysis. Methylation changes of clones analysed by bisulphite sequencing were calculated using the Fisher's exact test (GraphPad Prism Software). Methylation changes assessed by MassARRAY were calculated using a student's *t*-test (GraphPad Prism Software). The statistical evaluation of DNMT1–RNA interaction versus expression and methylation was estimated using the student's *t*-test (box-plots; Supplementary Fig. 5d). The overrepresentation of genes in clusters B and C following our hypothesis against those which did not, was assessed using a 2-sample proportion test (Fig. 5b). *P* values for *t*-test and 2-sample proportion test—'t.test' and 'prop.test', respectively—were calculated by the R functions (http://www.r-project.org). Values of $P \leq 0.05$ were considered statistically significant. The mean ± s.d. of two or more replicates is reported.

49. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1982).
50. Blobel, G. & Potter, V. R. Nuclei from rat liver: isolation method that combines purity with high yield. *Science* **154,** 1662–1665 (1966).
51. Sambrook, J. & Russell, D. W. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2001).
52. Mitchell, J. A. & Fraser, P. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev.* **22,** 20–25 (2008).
53. Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M. & Pagano, A. The expanding RNA polymerase III transcriptome. *Trends Genet.* **23,** 614–622 (2007).
54. Pagano, A. *et al.* New small nuclear RNA gene-like transcriptional units as sources of regulatory transcripts. *PLoS Genet.* **3,** e1 (2007).
55. Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9,** 493–501 (2003).
56. Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA* **89,** 1827–1831 (1992).
57. Bock, C. *et al.* BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* **21,** 4067–4068 (2005).

58. Christman, J. K. 5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21,** 5483–5495 (2002).
59. Dunn, J. J., Krippl, B., Bernstein, K. E., Westphal, H. & Studier, F. W. Targeting bacteriophage T7 RNA polymerase to the mammalian cell nucleus. *Gene* **68,** 259–266 (1988).
60. Estève, P. O. *et al.* Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* **20,** 3089–3103 (2006).
61. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5,** 621–628 (2008).
62. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
63. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28,** 511–515 (2010).
64. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7** (suppl. 1), S4.1–9 (2006).
65. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4,** 249–264 (2003).
66. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).