



Published in final edited form as:

Nature. 2013 November 21; 503(7476): 371–376. doi:10.1038/nature12598.

## DNMT1-interacting RNAs block gene specific DNA methylation

Annalisa Di Ruscio<sup>1,2,3,\*</sup>, Alexander K. Ebralidze<sup>1,2,\*</sup>, Touati Benoukraf<sup>4</sup>, Giovanni Amabile<sup>1,2</sup>, Loyal A. Goff<sup>5,6,7</sup>, Joylon Terragni<sup>9</sup>, Maria Eugenia Figueroa<sup>8</sup>, Lorena Lobo De Figureido Pontes<sup>1,2</sup>, Meritxell Alberich-Jorda<sup>1,2</sup>, Pu Zhang<sup>1</sup>, Mengchu Wu<sup>4</sup>, Francesco D'Alò<sup>3</sup>, Ari Melnick<sup>10</sup>, Giuseppe Leone<sup>3</sup>, Konstantin K. Ebralidze<sup>2</sup>, Sriharsa Pradhan<sup>9</sup>, John L. Rinn<sup>1,2,6,7</sup>, and Daniel G. Tenen<sup>1,4,§</sup>

<sup>1</sup>Harvard Stem Cell Institute, Harvard Medical School, Boston, MA, 02115, USA

<sup>2</sup>Beth Israel Deaconess Medical Center, Boston, MA, 02115, USA

<sup>3</sup>Policlinico A. Gemelli, Catholic University, School of Medicine and Surgery, Rome, Italy

<sup>4</sup>Cancer Science Institute, National University of Singapore, 117599, Singapore

<sup>5</sup>Department of Stem Cell and Regenerative Biology, Harvard University, 7 Divinity Ave, Cambridge, MA, 02138

<sup>6</sup>Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge MA, 02142

<sup>7</sup>Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, 32 Vassar St., Cambridge, MA 02139

<sup>8</sup>University of Michigan, Department of Pathology, Ann Arbor, MI 48109-2200

<sup>9</sup>New England Biolabs, 240 County Road, Ipswich, Massachusetts 01938-2723

<sup>10</sup>Dept of Medicine, Hematology-Oncology, C-620 Weill Cornell Medical College 1300 York Avenue, New York, NY 10021

### Summary

DNA methylation was 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 identified a novel RNA arising from the *CEBPA* gene locus critical in regulating the local DNA methylation profile. This RNA binds to DNMT1 and prevents *CEBPA* gene locus methylation. Deep sequencing of transcripts associated

Users may view, print, copy, download and text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

<sup>§</sup>Contact: Daniel G. Tenen, Center for Life Sciences, 3 Blackfan Circle, Room 437, Boston, MA 02115, Tel: 617-735-2235, FAX: 617-735-2222, [daniel.tenen@nus.edu.sg](mailto:daniel.tenen@nus.edu.sg).

<sup>\*</sup>These two authors equally contributed to the work;

**Author Contribution:** DGT supervised the project; ADR, AKE and DGT conceived and designed the study; ADR, AKE, GA, PZ, MAJ, FDA, SP, LLDFF, JT performed experiments; MW performed sequencing and microarray experiments; TB and LLG analyzed RIP-Seq, RNA-Seq, RRBS, and Microarrays data; MEF and AM performed the MassARRAY experiment and assisted in the analysis; ADR, AKE, GL, KKE, JLR, and DGT wrote the paper.

**Author Information:** Microarray expression and RNA-Seq have been deposited at GEO under accession number GSE32153; RIPseq: GSE32162; RRBS: GSE32168. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to DGT ([daniel.tenen@nus.edu.sg](mailto:daniel.tenen@nus.edu.sg)).

with DNMT1 combined with genome-scale methylation and expression profiling extended 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 disease.

---

## Introduction

DNA methylation is a key epigenetic signature implicated in transcriptional regulation, genomic imprinting, and silencing of repetitive DNA elements<sup>1-2</sup> and it 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<sup>3</sup>. While the bulk of genome is methylated at 70–80% of its CpGs, CpG islands are mostly unmethylated in somatic cells<sup>3-4</sup>. This modification is mediated by the members of the DNA methyltransferases (DNMTs) family conventionally classified as *de novo* (DNMT3a-b) and *maintenance* (DNMT1). In terms of epigenetic inheritance, DNMT1 has the unique ability to identify 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*<sup>5-7</sup>, including the potential involvement of noncoding (nc) RNAs originating within *CEBPA* locus. Recent discoveries of functional ncRNAs have provided new regulatory clues to the control of epigenetic marks. Particularly, long ncRNAs (lncRNAs) have been shown to regulate gene expression by interacting with chromatin modifiers, modulating transcription factor activity and competing for miRNA binding<sup>8-16</sup>. One unexplored aspect of 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* mRNA production. We show that such functional DNMT1-RNA association occurs in numerous gene loci. We thus propose a novel regulatory mechanism of gene methylation governed by RNAs.

## Results

### Characterization of *ecCEBPA*

Noncoding transcripts arising from the promoter and the downstream regions of coding genes can affect expression of the corresponding genes<sup>17-19</sup>. We searched and identified transcripts upstream and downstream of the intronless *CEBPA* gene. Strand specific RT-PCR (data not shown) and Northern blot analysis of RNAs from four leukemic cell lines, probing the region immediately after the *CEBPA* polyadenylation site, revealed the presence

of a major band of ~4.5 kb in HL-60 and U937 (where *CEBPA* is expressed), but not in K562 or Jurkat cell lines (where *CEBPA* is expressed at low or undetectable levels) (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 (ec) CEBPA*”, since it does encompass the entire mRNA sequence in the same-sense orientation (shown by primer extension and 5′/3′RACE; Supplementary Information (SI); Supplementary Fig. 1c, d). qRT-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). Importantly, *ecCEBPA* synthesis precedes the expression of its overlapping mRNA in the S phase (SI and Supplementary Fig. 1f, g) and is regulated by both RNA polymerase II and III (RNAP II and III; SI and Supplementary Fig. h–p) as described for other loci<sup>20–22</sup>.

### ***ecCEBPA* inhibits DNA methylation and facilitates *CEBPA* expression**

To examine the functional role of *ecCEBPA* in regulation of *CEBPA* transcription, we performed both loss and gain-of-function experiments. Knock-down of *ecCEBPA* in U937 cell line (up to a 4-fold decrease) achieved by small hairpin (sh) RNAs 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<sup>6–7,23</sup>. To examine if there was a connection between *ecCEBPA* and methylation of the *CEBPA* locus, we analyzed methylation within the distal promoter (located at –0.8–0.6 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; Supplementary Fig. 2a).

To investigate whether enforced expression of the *ecCEBPA* was sufficient to inhibit DNA methylation, the downstream region of the *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 the *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 the *ecCEBPA* (R1) resulted in greater than 3-fold increase in mRNA expression (Fig. 2d) whereas overexpression of an unrelated region (UR; located 45 kb downstream) and regions immediately outside the *ecCEBPA* boundaries did not affect mRNA levels (Supplementary Fig. 2b–d). Moreover, concomitant decrease of DNA methylation in three tested regions within the *CEBPA* gene, distal promoter, coding region, and 3′UTR, accompanied overexpression of *ecCEBPA* but not of the UR (Fig. 2e; 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 bisulfite sequencing, RRBS<sup>24</sup>) of DNA methylation changes imposed by *ecCEBPA* versus UR 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 neighboring *CEBPG* and distant *TP73* (on chr1p36) genes were achieved after 5-Aza-CR treatment but not after *ecCEBPA* overexpression (Supplementary Fig. 2g–k). Furthermore, RRBS analysis of promoter and 1<sup>st</sup> exon regions revealed that only ~3.3% of the interrogated loci (396 out of 11844) were hypomethylated at levels similar to the *CEBPA* locus (Supplementary Fig. 2l).

Furthermore, *ecCEBPA* downregulation by the universal RNAPs inhibitor Actinomycin D and RNAPIII-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; 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*. Similar effect was observed in ML-60218-treated HL-60 cells (Supplementary Fig. 1i; Supplementary Fig. 2o).

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

Collectively, these data highlight the regulatory role of the *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 unveil the mechanism through which it is achieved. Among DNMTs, DNMT1 is the one whose expression and enzymatic activity peaks during S phase<sup>25</sup>. 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, Supplementary Fig. 2n). We therefore asked whether the presence of *ecCEBPA* during S phase led to RNA interference of DNMT1 activity.

To determine if 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; Supplementary Fig. 3a). Analysis of polyA(+)/(–) fractions in DNMT1-RNA precipitates revealed enrichment of *CEBPA* transcripts in the polyA(–) 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 (REMSA). RNA oligonucleotides corresponding to the 5' and 3' parts of the *ecCEBPA* were selected by: (i) ability (R2, R5 and R6) and

inability (R4) to fold into stem-loop structures<sup>26</sup>; (ii) 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<sup>27</sup>, CpG UpG substitutions, neutral with regard of secondary structures, did not affect binding (mutR2; Fig. 3c). In 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 the *ecCEBPA* (single stranded [ss] R1 and R3; and double stranded [ds] R13; Supplementary Fig. 3d), confirmed DNMT1 binding to stem-loop structured RNAs (SI and Supplementary Fig. 3e–f). Importantly, REMSA performed in 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, ssRNA oligonucleotides capable of forming secondary structures (R5) and corresponding unmethylated (umDNA; D5/D6), hemimethylated (hmDNA; D5/D6), and fully methylated (mDNA; D5/D6) dsDNA (Fig. 3a), at a constant molar concentration, were titrated with increasing range of DNMT1 enzyme concentrations using EMSA. RNA formed complexes beginning at <0.013  $\mu\text{M}$  DNMT1, whereas DNA at >0.026  $\mu\text{M}$  DNMT1 (Fig. 3f), with a mean dissociation constant (Kd) for RNA of 0.045 $\mu\text{M}$  and between 0.082 and 0.14 $\mu\text{M}$  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 DNMT1 domain required for the RNA binding, DNMT1-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 (TRD)<sup>28</sup>, 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 TRD led to disruption of DNMT1 enzymatic activity (data not shown) making further refinement of the binding domain not feasible.

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 structures features. Importantly, DNMT1, *via* its catalytic domain, binds with higher affinity to folded RNA than to DNA.

### Transcription interferes with DNMT1 methylation of DNA

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 RNA polymerase promoter (Supplementary Fig. 4a–h), and DNMT1 methylase activity was monitored in 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; Supplementary Fig. 4i–j). In 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 ribooligonucleotides (Fig. 4g). Similarly, T7 RNA polymerase<sup>29</sup>-induced transcription in living cells led to a pronounced decrease in DNA methylation (SI and Supplementary Fig. 4k–p).

Thus, RNA can complex with and impact DNMT enzymatic activity *in vitro*<sup>30–32</sup> 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. cDNA libraries made of RNAs co-immunoprecipitated with anti-DNMT1 antibody (DNMT1 library) and IgG (control library) were tested for *ecCEBPA* enrichment (“quality control”; Supplementary Fig. 5a) and subsequently analyzed by massively parallel sequencing<sup>11</sup>. 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<sup>33</sup> build on RefSeq hg19 (Supplementary Fig. 5b). All DNMT1 peaks were also annotated using the known RNAs databases provided by HOMER<sup>34</sup> (Supplementary Fig. 5c). We focused on genomic regions encompassing the 3 kb upstream to the TSS and downstream to the transcription ending site (TES) of the annotated genes, referred to as “gene loci”. We identified 6,042 gene loci containing one or more peaks from the DNMT1 library (Methods: Data integration).

To confirm that DNMT1 RIP-Seq peaks were associated with actual transcribed elements, RNA-Seq was conducted on polyA (–) HL-60 RNA. 375 million 76-bp paired-end reads were aligned to hg19 using TopHat<sup>235</sup> and assembled using cufflinks<sup>36</sup>. 14,077 (87.02%) of specific DNMT1 peaks overlapped with a transcribed element from the RNA-Seq assembly of the polyA (–) HL-60 RNA fraction. Thus, a vast majority of DNMT1-interacting-RNAs (DiRs) were not polyadenylated.

Additionally, we performed a similar analysis with total HL-60 RNA (300 million 76-bp paired-end reads). 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<sup>24</sup> 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  $\beta$ -scores from -2 kb from the TSS to the end of the first intron). A negative correlation between DNMT1-RNA 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: RRBS-Data integration). 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 (5646 genes) and C (2737) were significantly higher than numbers of genes in clusters A, F, E (2528, 1653, 1146) and G, H, D (584, 930, 582), respectively ( $P$ -value  $< 0.0001$ ). Examples of genes from clusters B and C are presented in Fig. 5c and Supplementary Fig. 5f–h). Further, 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 ( $p$ value  $\leq 0.01$ ) were shared with pyknons-related BP-GO<sup>37</sup>. 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<sup>38</sup>, 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 the *ecCEBPA* does not support a model of trivial titration ('squelching'<sup>39</sup>) of DNMT1 but suggests a *cis* regulatory role of the DiR. We propose a model wherein RNAs contain a locus-selective triplex/quadruplex<sup>40</sup> forming part, the 'anchor', mooring the DNMT1-RNA complex to the locus, and a DNMT1-interacting part, the 'bait', 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, e.g., competing endogenous RNAs (ceRNAs)<sup>15-16,41</sup>. However, unlike ceRNAs, the *ecCEBPA* model also introduces the requirements for both functional and physical co-compartmentalization of the RNA, its parental locus, and DNMT1. Given the DiR's ability to bind DNMT1, it is tempting to hypothesize that DiRs represent a novel class of RNA regulons<sup>42</sup>.

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

### 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 cell line HCT116 hypo and 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<sup>49</sup>. **Cytoplasmic RNA** was isolated with the Paris kit (Ambion) according to the manufacturer's recommendations. Preparation of **nuclear RNAs**: we used a method derived from protocols of nuclei isolation<sup>50</sup>, with minor modifications. Briefly, equal amounts of viable cells (~50 million) were washed with ice-cold PBS supplemented with 5 mM vanadyl complex, 1 mM PMSF and resuspended in the ice-cold lysis buffer: 1x 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 BSA; 0.5% NP-40; freshly



supplemented with protease inhibitors (2 mM leupeptin, add as x400; 2 mM pepstatin, add as x400; 100 mM benzamidine, add as x400; a protease inhibitor cocktail (Roche Applied Science, Cat. No. 1836153), 1 tablet/375  $\mu$ L H<sub>2</sub>O, add as x100; 100 mM PMSF, add as x100); 2 mM vanadyl complex (New England Biolabs); and 20 units/mL RNase inhibitor (RNAguard; Amersham Biosciences). Samples were incubated at 0°C for ~10 minutes and passed through numerous 40 up and down strokes in a Dounce homogenizer (10 with pestle A and 30 with pestle B). The pelleted nuclei were resuspended 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 BSA), freshly supplemented with protease inhibitors and overlaid onto 2 mL “cushions” (200 mL “Cushion” buffer consists of 15 mL ddH<sub>2</sub>O; 15 mL 20x 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 24,400 rpm, 60 minutes, 4°C. The pelleted nuclei were resuspended in 1 mL Storage buffer (1.75 mL ddH<sub>2</sub>O; 2 mL glycerol; 0.2 mL 20x Buffer A), freshly supplemented with protease inhibitors. Nuclear RNAs were extracted as described<sup>50</sup>. All total, cytoplasmic and nuclear RNA samples used in this study were treated with DNase I (10 U of DNase I per 3  $\mu$ g of total RNA; 37°C for one hour; 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 (Biorad; Hercules, CA); PCR conditions: 95°C (10 min), followed by 40 cycles of 95°C (15s) and 60°C (1min) 72°C (1min). 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 sec.) and 60°C (60 sec.). Primers and probes are presented in Supplementary Data 1 (SI). qRT-PCR primer set for the CEBPA mRNA is located in the coding region (black double-headed arrow Fig. 1a) and after the polyA signal for the *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<sup>51</sup>. Southern blot transfer and hybridization with oligonucleotide AL16 were performed as previously reported<sup>51</sup>. Oligonucleotide sequences are shown in Supplementary Data 1 (SI). 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 Data 1 (SI) for primer sequences.

### Double Thymidine block (early S-phase block)

HL-60 cells were grown to 70–80% confluence, washed twice with 1xPBS and cultured in DMEM (10% FCS) + 2.5 mM Thymidine for 18 h (first block). Thymidine was washed out with 1xPBS and cells were grown in DMEM (10% FCS). After 8 hours cells were cultured in 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, $\alpha$ -Amanitin and Actinomycin D treatment

After release from double thymidine block, HL-60 cells were treated with 100  $\mu$ M 5,6-Dichlorobenzimidazole 1- $\beta$ -D-ribofuranoside<sup>52</sup> (DRB) (Sigma Aldrich) for 1, 2, and 3 hours. HL-60 cells were treated with 12.5, 25, 50, or 100  $\mu$ M ML-60218<sup>53–54</sup> (Calbiochem®) for 24 hours. HL-60 cells were treated with 5, 23, 50, 75, 100, or 150  $\mu$ g/mL  $\alpha$ -Amanitin (Sigma-Aldrich) for 14 hours. Synchronized and unsynchronized U937 cells were treated with ML-60218 at 100  $\mu$ M and Actinomycin D (Sigma-Aldrich) at 150  $\mu$ M for 7 hours. Total RNA was harvested as described above and expression levels of *CEBPA*, *ecCEBPA*, and 5S were measured by TaqMan qRT-PCR.

### Nuclear Chromatin (ChIP) and RNA immunoprecipitation (RIP)

ChIP was performed as previously described<sup>19</sup>. Fold enrichment was calculated using the formula  $2^{-\Delta Ct} [ChIP/NIS]$ . Antibodies used for ChIP are listed in Supplementary Table 1 (SI).

RIP was performed as described by Ebralidze *et al.*<sup>14</sup>. **Day I:** Cross-linked nuclei were collected as following:  $60 \times 10^6$  HL-60 cells were cross-linked with 1% formaldehyde (formaldehyde solution, freshly made: 50 mM HEPES-KOH; 100 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 11% formaldehyde) for 10 minutes at room temperature. Crosslinking was stopped by adding 1/10<sup>th</sup> volume of 2.66 M Glycine, kept for 5 minutes at room temperature and 10 minutes on ice. Cell pellets were washed twice with ice-cold PBS (freshly supplemented with 1 mM PMSF). Cell pellets were resuspended in cell lysis buffer (volume = 4 mL): 1x 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, Cat. No. 1836153, 1 tablet/375  $\mu$ L H<sub>2</sub>O; add as x100), 1 mM PMSF, and 2 mM vanadyl complex (NEB). Cells were incubated at 0°C for 10–15 minutes and homogenized by Dounce (10 strokes pestle A and 40 strokes pestle B). Nuclei were recovered by centrifugation at 2,000 rpm for 10 minutes at 4°C. Nuclei were resuspended in 3 ml 1x Resuspension Buffer (50 mM HEPES-NaOH, pH 7.4; 10 mM MgCl<sub>2</sub>) supplemented with 1 mM PMSF and 2 mM vanadyl complex. DNase treatment (250 U/ml) was performed for 30 minutes at 37°C, and EDTA (final concentration 20 mM) was added to stop the reaction. Resuspended nuclei were sonicated once for 20s (1 pulse every 3 seconds) at 30% amplitude (Branson Digital Sonifer, Danbury, CT). **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  $\mu$ L magnetic beads (Protein A or G

Magnetic Beads; #S1425S or #S1430S 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: (i) antibody of interest: either DNMT1 antibody (Abcam, cat# ab87656) or anti-cap antibody (Anti-m<sub>3</sub>G-/m<sup>7</sup>G-cap; Synaptic Systems); (ii) preimmune serum: IgG (Sigma-Aldrich); (iii) no antibody, no serum (input). 5 µg antibody or preimmune 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 II.** 200 µL of Protein A coated super-paramagnetic 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 at 56°C overnight. **Day III.** Beads were removed in magnetic field. Phenol (pH 4.3) extraction was performed after addition of NaCl (0.2 M final concentration). EtOH precipitation (in the presence of glycogen); 3 hours at -20°C. The pellet was dissolved in 180 µL H<sub>2</sub>O, heated at 75 °C for 3 minutes, and immediately chilled on ice. Samples were treated with DNase I (250 U/ml) in the presence of RNase inhibitor 300 U/ml in x1 buffer # 2 (NEB) at 37°C for 30 minutes. Phenol (pH 4.3) extraction and EtOH precipitation were repeated. 15. The RNA pellet was dissolved in 50 µL H<sub>2</sub>O.

#### **Tobacco Acid Pyrophosphatase (TAP), 5'-Phosphate-Dependent- Exonuclease (Terminator) treatment**

Equal amounts of RNA harvested from HL-60 cells (as described above) were digested with TAP (Epicentre, Cat. # T81050), Terminator (Epicentre, Cat. # TER51020), 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 resuspended in ddH<sub>2</sub>O. *ecCEBPA*, *CEBPA*, and *18S* expression levels were measured by qRT-PCR using the TaqMan primer sets indicated in Supplementary Data 1 (SI).

#### **Down-regulation of *ecCEBPA***

Three different short hairpin RNAs 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 Data 1. Lentiviral particles were produced as previously described<sup>55</sup>. 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 hours after transfection and concentrated using a Centricon Plus-70 100000 MWCO column (Millipore). Lentiviral transduction was performed in the presence of Hexadimethrine bromide (final concentration 8 µg/ml) in the human myeloid cell line U937. Puromycin (2µg/ml) was added to the cultures two days after infection. Resistant clones were selected and screened for down-regulation of *ecCEBPA* by qRT-PCR.

### Up-regulation of *ecCEBPA*

The 3' *ecCEBPA* region (R1), the upstream (upR) and downstream (dwR) *ecCEBPA* regions, and the unrelated genomic region (UR, see Supplementary Data 1) were cloned into the pBabe retrovirus vector harboring a puromycin selection marker (Addgene; plasmid 1764). Oligonucleotide sequences used to amplify both regions are shown in Supplementary Data 1. K562 cells were transfected with the Amaxa Cell Line Nucleofector® Kit V, Program T-003. Puromycin (2 µg/ml) was added to the cultures 2 days after transfection. Resistant clones were selected and screened for upregulation of *ecCEBPA* and the unrelated region (UR) by Northern Blot Analysis.

### Bisulfite Treatment, Combined Bisulfite Restriction Analysis Assay (COBRA) and Bisulfite Sequencing

The methylation profile of the *CEBPA* gene locus was performed by bisulfite sequencing as previously described<sup>56</sup>. Briefly, 1 µg of genomic DNA was bisulfite converted by using the EZ DNA Methylation kit (Zymo Research; Orange, CA). Primers and PCR conditions for bisulfite sequencing and combined bisulfite restriction analysis assay (COBRA) are summarized in Supplementary Data 1. 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 bisulfite sequencing, PCR products were gel-purified (Qiagen) and cloned into the pGEM®-T Easy Vector System (Promega). Sequencing results were analyzed using BiQ analyzer software<sup>57</sup>. Samples with conversion rate <90% and sequences identity <70% as well as clonal variants were excluded from our analysis<sup>55</sup>. The minimum number of clones for each sequenced condition was 6. Primer sequences are shown in Supplementary Data 1 (SI).

### 5' Azacytidine (5Aza-CR) treatment<sup>58</sup>

K562 cells were treated with 5Aza-CR (Sigma Aldrich) according to the manufacturer's instructions. Medium was refreshed every 48 hours. RNA (for RT-PCR) and genomic DNA (for bisulfite sequencing) were isolated after 7 days of treatment.

### MassARRAY

Quantitative DNA methylation analysis using the MassARRAY technique was performed by Sequenom, Inc., as previously described<sup>43</sup>. Briefly, 1 µg of genomic DNA was converted with sodium bisulfite using the EZ DNA methylation kit (Zymo Research, Orange, California), 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 percentage. Primer sequences are provided in Supplementary Excel File 1.

### Transfection of human DNMT1-HA tag construct and western blot analysis

The human DNMT1-HA tagged cloned into the expression vector pCDNA3.1 was a kind gift from Prof. Steven Baylin. HEK293T cells were transfected using Lipofectamine 2000, and 2 days later harvested for western blot analysis. Single cell suspensions were lysed with modified RIPA buffer, and whole cell lysates separated on 6 % SDS-PAGE gels. Immunoblots were stained with the following primary antibodies: DNMT1 (1:5000; Abcam, cat# ab87656) and HSP90 (1:2000, BD Bioscience). All secondary antibodies were horseradish peroxidase (HRP) conjugated (Santa Cruz) and diluted 1:5000 for rabbit-HRP, and 1:3000 for mouse-HRP. Western Blot analysis for the HCT116 hypo and wild type cell line was similarly performed.

### T7 polymerase-induced transcription

The T7 expression system is based on technology developed at Brookhaven National Laboratory under contract with the U.S. 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, Upton, New York 11973, telephone 631-344-7134. Maps of T7 polymerase constructs are presented in Dunn *et al.*,<sup>59</sup>. Briefly, 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, USA). After selection in G418, individual clones were transfected with T7 polymerase-expressing mammalian constructs and were tested by COBRA for genomic methylation.

### Electrophoretic gel mobility shift assays (EMSAs) and Kd determination

DNA and RNA oligonucleotides (15 pmol) were end-labeled with [ $\gamma$ -<sup>32</sup>P] ATP (Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs). Reactions were incubated at 37°C for 1h and then passed through G-25 spin columns (GE Healthcare) according to the manufacturer's instructions to remove unincorporated radioactivity. Labeled samples were gel-purified on 10% polyacrylamide gels. Binding reactions were carried out in 10  $\mu$ L volumes in the following buffer: 5 mM Tris pH 7.4, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 3% v/v glycerol, 100 mM NaCl. Various amounts (0.021–0.156  $\mu$ M) of purified DNMT1 protein (BPS Bioscience Inc.) were incubated with 1.1 nM of <sup>32</sup>P-labeled dsDNA and ss/ds RNAs. In the competitive assay, a fixed amount of protein and increasing amounts of competitors (dsDNA 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 (0.5xTBE; 4 °C; ~3h at 140 V). Gels were dried and exposed to X-ray film and/or PhosphorImager screens. Quantitation was done with ImageQuant software. For affinity assays, the percent shifted species was determined as follows: the migration of the labeled DNA in this reaction was defined as zero percent shifted and the ratio of the PhosphorImager 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 non-linear regression analysis performed with Prism 4.0a software. RNA and DNA oligonucleotides used in EMSA are shown in Figure 3a and Supplementary Figure 3d and listed in Supplementary Data 1.

### **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 previously described<sup>60</sup>. Protein concentrations for the recombinant GST and GST-DNMT1 fragments were determined by gel electrophoresis and subsequent Coomassie staining and densitometry. <sup>32</sup>P end-labeling of *ecCEBPA* oligonucleotides was carried out at 37°C for 30 min in a total volume of 50 µl. The reaction contained 50 pmoles of *ecCEBPA*, adenosine 5'-Triphosphate, [ $\gamma$ -<sup>32</sup>P] (specific activity 3000 Ci/mmol, Perkin Elmer), and 20 units of T4 kinase (New England Biolabs) mixed in assay buffer (70 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 5 mM Dithiothreitol, pH 7.6). The labeled *ecCEBPA*-<sup>32</sup>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 5µl *ecCEBPA*-<sup>32</sup>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 Dithiothreitol, 1 mM EDTA, 5 % Glycerol, pH 7.8). The sepharose beads were then washed 4 times in phosphate buffered solution and placed in 3 ml of scintillation fluid and bound <sup>32</sup>P was measured for 1 minute. All measurements were normalized to <sup>32</sup>P readings for the corresponding input <sup>32</sup>P-*ecCEBPA*.

### ***in vitro* Transcription-Methylation Assay *in vitro***

transcription-methylation assays were performed on hemimethylated DNA (hmDNA; described in Supplementary Information, legend to Supplementary Figure 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 and 1.25 mM MgCl<sub>2</sub>, including the "DNMT1 only" reaction. This pre-determined concentration of Mg<sup>2+</sup> 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-<sup>32</sup>H]methionine (AdoMet) (specific activity 18 Ci/mmol, Perkin Elmer), 200ng of substrate DNA, recombinant DNMT1 enzyme (25 pmoles), and *ecCEBPA* (25 pmoles) mixed in assay buffer (50 mM Tris-HCl, 1 mM Dithiothreitol, 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, catalog no. 3658-325). These membranes were processed by washes in 3×1 ml of 0.2 M ammonium bicarbonate, 3×1 ml of water, and 3×1 ml of ethanol. Processed membranes were air-dried, placed in 3ml of

scintillation fluid and tritium incorporation was measured for 1 min. Background subtraction (no DNA substrate) was performed for all experimental sample counts.

### RNA Immunoprecipitation Sequencing (RIPseq)

Total RNA immunoprecipitated with DNMT1 antibody (Abcam, Cat. # ab87656) or IgG (Sigma Aldrich) was processed for sequencing as described by Mortazavi *et al*<sup>61</sup> with some modifications. Double stranded cDNA was synthesized using the Just cDNA Double-Stranded cDNA Synthesis Kit (Agilent Technology, Santa Clara, CA) according to the manufacturer's instructions. Illumina sequencing libraries were constructed from these cDNA using a ChIP-Seq sample preparation kit (cat# IP-102-1001, Illumina, San Diego, CA) 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, Carlsbad). DNA fragments of 175 – 200 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 2x76 bp reads. Paired-End reads were trimmed to 50 bp and aligned to the reference genome hg19 using BWA<sup>62</sup> 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 coarse 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 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. 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<sup>33</sup> build on RefSeq hg19. All DNMT1 RIP-Seq peaks were also annotated using the HOMER pipeline (version 4.2)<sup>34</sup> which provide a comprehensive RNAs database (coding and non-coding, including miRNA, snoRNA, rRNA, snRNA, tRNA, *etc.*).

A peak was considered belonging to a gene if located in the gene body or 3 kb up- or downstream the gene (gene loci). Altogether, 6042 gene loci were covered by a least 1 significant RIPseq peak.

## RNA Sequencing (RNAseq)

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 one hour 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 ribosomal RNA with Ribo-Zero™ 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/lane) on a Hi-Seq-2000 Illumina instrument. Raw read sequences were deposited in GEO (Accession Number GSE32153).  $2.96 \times 10^8$  reads from the HL-60 and  $3.75 \times 10^8$  76-bp paired end reads from the HL-60 polyA(+)-depleted RNA were aligned to the human genome (hg) 19 (UCSC release) using Tophat2<sup>63</sup>. Aligned reads were assembled into individual full-length transcripts using Cufflinks v2.0.2<sup>63</sup> and a merged assembly was created from the two assemblies and additionally, all level 1 and 2 transcripts from the gencode v11 catalog<sup>64</sup> using cuffmerge<sup>36</sup>. To confirm the transcription of the significant DNMT1 RIP-Seq peaks, we overlapped the peak intervals with the RIP-Seq assemblies using the bedtools<sup>60</sup> intersect Bed utility.

## Reduced Representation Bisulfite Sequencing (RRBS)

RRBS was performed as described<sup>44</sup>. Briefly, 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<sup>+</sup>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'-methyl-cytosine instead of cytosine were ligated to size-selected MspI fragments. Adapter-ligated fragments were bisulfite-treated using the EZ DNA Methylation kit (Zymo Research, Orange, CA). 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<sup>45</sup> allowing 2 mismatches. Reads from replicates were merged and processed as previously described<sup>46</sup>. We considered only CpG located in regions with a depth of coverage greater than 3 reads. The  $\beta$ -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  $\beta$ -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 3 sequenced CpG in the promoter or less than 3 sequenced CpG in the first exon-intron were excluded.

For RRBS in R1 and UR overexpressing cells bisulfite sequenced UR -R1 genomes were binned at 100 bp intervals using the R-Bioconductor/methylKit “*tileMethylCounts*” function (<http://code.google.com/p/methylkit>)<sup>47</sup>. 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 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 1<sup>st</sup> exon was overlapped by at least one significant differentially methylated bins. In total, 11844 promoter/1<sup>st</sup> exon regions were analyzed.

### RNA Expression profiling

RNA isolated from HL-60 cells was employed for sample amplification and labeling using the Whole Transcriptome assay reagent kits from Affymetrix. 10  $\mu$ g of labeled 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, Santa Clara, CA, USA) was used to measure the fluorescence intensity emitted by the labeled target. Raw data processing was performed using the Affymetrix GeneChip® Operating Software (GCOS). Microarrays were RMA normalized using ‘*affy*’, an R-Bioconductor library<sup>65</sup>. *CEBPA* expression was used as a threshold to define expressing ( $\log_2$  score above 4) and not expressing ( $\log_2$  score below 4) genes for further analysis.

### Gene Ontology and Pyknons comparison

GO analysis was performed with DAVID<sup>66</sup>. 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](https://cm.jefferson.edu/tools_and_downloads/pyknons.html)).

### Data integration

We used the RefSeq 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 40857 RefSeq Ids was reduced to 23250 transcript Ids. Then, we annotated all RIP-Seq peaks against the gene loci which include exonic, intronic, and UTR regions plus 3 kb upstream of the TSS and 3 kb downstream of the Transcription End Site (TES) regions. We identified 6042 gene loci with DNMT1-RIPseq peaks and 17208 gene loci without DNMT1-RIPseq peaks. Finally, we focused our study on gene loci covered by the RRBS. We identified 4833 gene loci with DNMT1-RIPseq peaks and covered by RRBS and 10973 gene loci without DNMT1-RIPseq 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 analyzed by bisulfite sequencing were calculated using the Fisher's exact test (GraphPad Prism Software Inc.). Methylation changes assessed by MassARRAY were calculated using a student t-test (GraphPad Prism Software Inc.). The statistical evaluation of DNMT1-RNA interaction *versus* expression and methylation was estimated using the student 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* < 0.05 were considered statistically significant. The mean ± s.d. of two or more replicates is reported.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

This work was supported by grants CA118316, CA66996, and HL56745 from the National Institutes of Health to DGT, the Italian Foundation for Cancer Research (F.I.R.C.) "Leonino Fontana e Maria Lionello" fellowship, the NIH T32 HL007917-11A1 and the to ADR; FAMRI CIA (103063) grant to AKE; Fondazione Roma "Progetto cellule staminali" to GL and FDA; the American Italian Cancer Foundation Fellowship (A.I.C.F.) to GA and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) - grant No. 2011/11822-6 to LLDFFP; the National Research Foundation and the Singapore Ministry of Education under its Centres of Excellence initiative to MW and TB. We thank Robert White, Debbie Johnson, Maxim Frank-Kamenetskii, Sergei M. Mirkin., Barbara Tazon-Vega, Constanze Bonifer, Christoph Bock, John J. Dunn (postmortem), Ron A.M. Fouchier for helpful advice and reagents; Isidore Rigoutsos for providing us the latest released pyknons' database; all the members of the Tenen Lab; Patrick Tan, Tay Su Ting from Genome Institute of Singapore; Richie Soong from the Cancer Science Institute Translational Interface; Joyce LaVecchio and Girijesh Buruzula from the Harvard Stem Cell Institute/Joslin Diabetes Center flow cytometry facility; and Fred 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 (DGT).

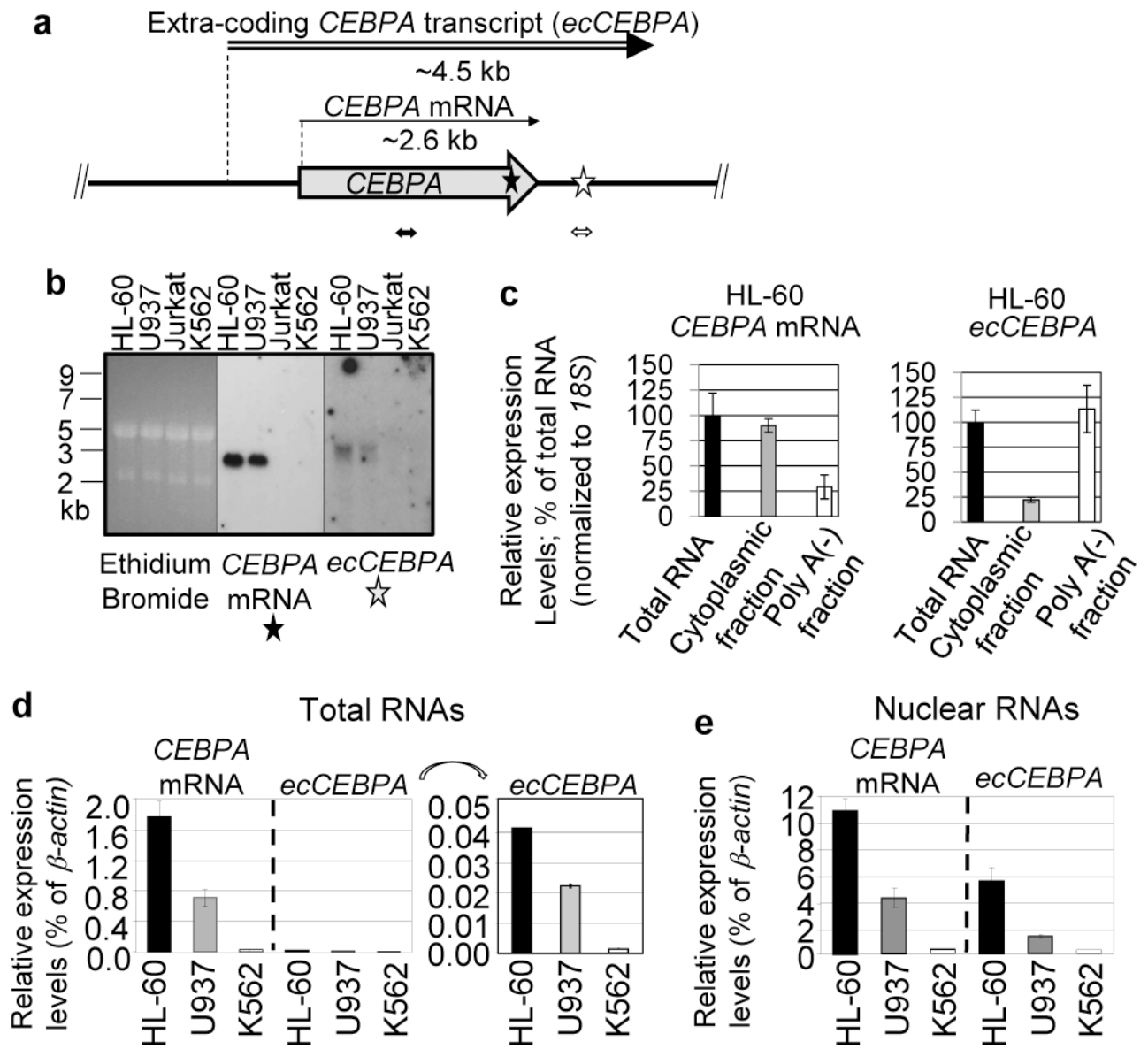
### References

1. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet.* 2002; 3:415–428. [PubMed: 12042769]
2. Robertson KD. DNA methylation, methyltransferases, and cancer. *Oncogene.* 2001; 20:3139–3155. [PubMed: 11420731]
3. Illingworth RS, et al. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* 2010; 6
4. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A.* 2006; 103:1412–1417. [PubMed: 16432200]

5. Tenen DG. Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer*. 2003; 3:89–101. [PubMed: 12563308]
6. Tada Y, et al. Epigenetic modulation of tumor suppressor CCAAT/enhancer binding protein alpha activity in lung cancer. *J Natl Cancer Inst*. 2006; 98:396–406. [PubMed: 16537832]
7. Hackanson B, et al. Epigenetic modification of CCAAT/enhancer binding protein alpha expression in acute myeloid leukemia. *Cancer Res*. 2008; 68:3142–3151. [PubMed: 18451139]
8. Gupta RA, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010; 464:1071–1076. [PubMed: 20393566]
9. Huarte M, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. 2010; 142:409–419. [PubMed: 20673990]
10. Rinn JL, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*. 2007; 129:1311–1323. [PubMed: 17604720]
11. Zhao J, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell*. 2010; 40:939–953. [PubMed: 21172659]
12. Nagano T, et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science*. 2008; 322:1717–1720. [PubMed: 18988810]
13. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature*. 2012; 482:339–346. [PubMed: 22337053]
14. Ebralidze A, Wang Y, Petkova V, Ebralidze K, Junghans RP. RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. *Science*. 2004; 303:383–387. [PubMed: 14657503]
15. Polisenio L, et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*. 2010; 465:1033–1038. [PubMed: 20577206]
16. Tay Y, et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell*. 2011; 147:344–357. [PubMed: 22000013]
17. Ebisuya M, Yamamoto T, Nakajima M, Nishida E. Ripples from neighbouring transcription. *Nat Cell Biol*. 2008; 10:1106–1113. [PubMed: 19160492]
18. Preker P, et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science*. 2008; 322:1851–1854. [PubMed: 19056938]
19. Ebralidze AK, 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*. 2008; 22:2085–2092. [PubMed: 18676813]
20. Oler AJ, et al. Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nat Struct Mol Biol*. 2010; 17:620–628. [PubMed: 20418882]
21. Listerman I, Bledau AS, Grishina I, Neugebauer KM. Extragenic accumulation of RNA polymerase II enhances transcription by RNA polymerase III. *PLoS Genet*. 2007; 3:e212. [PubMed: 18039033]
22. Raha D, et al. Close association of RNA polymerase II and many transcription factors with Pol III genes. *Proc Natl Acad Sci U S A*. 2010; 107:3639–3644. [PubMed: 20139302]
23. Wouters BJ, et al. Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood*. 2007; 110:3706–3714. [PubMed: 17671232]
24. Meissner A, et al. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res*. 2005; 33:5868–5877. [PubMed: 16224102]
25. Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*. 1992; 71:865–873. [PubMed: 1423634]
26. Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res*. 2003; 31:3429–3431. [PubMed: 12824340]
27. Pradhan S, Esteve PO. Allosteric activator domain of maintenance human DNA (cytosine-5) methyltransferase and its role in methylation spreading. *Biochemistry*. 2003; 42:5321–5332. [PubMed: 12731873]

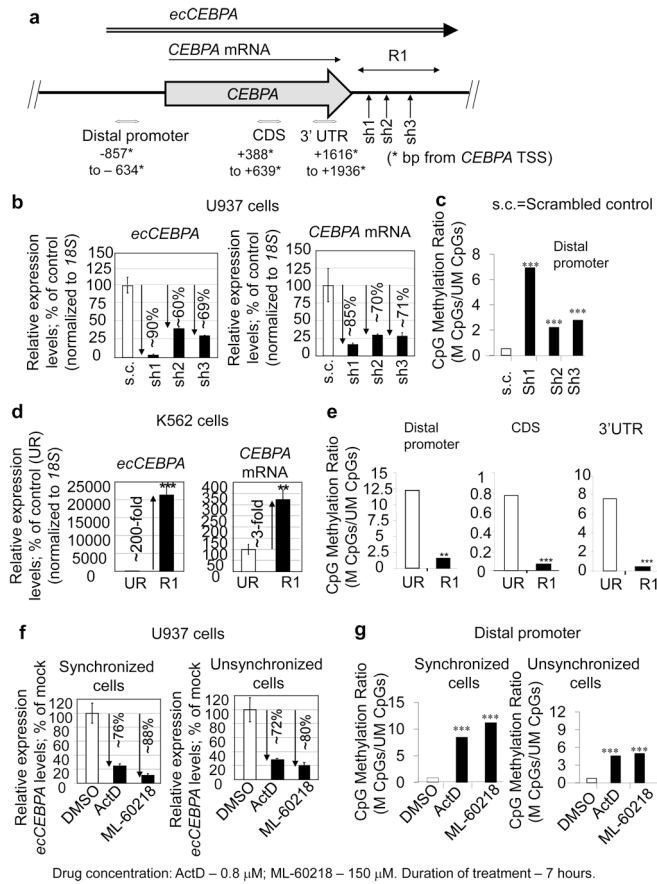
28. Song J, Rechkoblit O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science*. 2011; 331:1036–1040. [PubMed: 21163962]
29. Fuerst TR, Niles EG, Studier FW, Moss B. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc Natl Acad Sci U S A*. 1986; 83:8122–8126. [PubMed: 3095828]
30. Svedruzic ZM. Mammalian cytosine DNA methyltransferase Dnmt1: enzymatic mechanism, novel mechanism-based inhibitors, and RNA-directed DNA methylation. *Curr Med Chem*. 2008; 15:92–106. [PubMed: 18220765]
31. Bolden A, Ward C, Siedlecki JA, Weissbach A. DNA methylation. Inhibition of de novo and maintenance methylation in vitro by RNA and synthetic polynucleotides. *J Biol Chem*. 1984; 259:12437–12443. [PubMed: 6208188]
32. Bolden AH, Nalin CM, Ward CA, Poonian MS, Weissbach A. Primary DNA sequence determines sites of maintenance and de novo methylation by mammalian DNA methyltransferases. *Mol Cell Biol*. 1986; 6:1135–1140. [PubMed: 3023872]
33. Shin H, Liu T, Manrai AK, Liu XS. CEAS: cis-regulatory element annotation system. *Bioinformatics*. 2009; 25:2605–2606. [PubMed: 19689956]
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*. 2010; 38:576–589. [PubMed: 20513432]
35. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009; 25:1105–1111. [PubMed: 19289445]
36. Trapnell C, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012; 7:562–578. [PubMed: 22383036]
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*. 2008; 36:3484–3493. [PubMed: 18450818]
38. Bock C, et al. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol*. 2010; 28:1106–1114. [PubMed: 20852634]
39. Gill G, Ptashne M. Negative effect of the transcriptional activator GAL4. *Nature*. 1988; 334:721–724. [PubMed: 3412449]
40. Frank-Kamenetskii MD, Mirkin SM. Triplex DNA structures. *Annu Rev Biochem*. 1995; 64:65–95. [PubMed: 7574496]
41. Karreth FA, et al. In vivo identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell*. 2011; 147:382–395. [PubMed: 22000016]
42. Keene JD. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet*. 2007; 8:533–543. [PubMed: 17572691]
43. Figueroa ME, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell*. 2010; 17:13–27. [PubMed: 20060365]
44. Smith ZD, Gu H, Bock C, Gnirke A, Meissner A. High-throughput bisulfite sequencing in mammalian genomes. *Methods*. 2009; 48:226–232. [PubMed: 19442738]
45. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics*. 2009; 10:232. [PubMed: 19635165]
46. Benoukraf T, Wongphayak S, Hadi LH, Wu M, Soong R. GBSA: a comprehensive software for analysing whole genome bisulfite sequencing data. *Nucleic Acids Res*. 2013; 41:e55. [PubMed: 23268441]
47. Akalin A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol*. 2012; 13:R87. [PubMed: 23034086]
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*. 2008; 36:5221–5231. [PubMed: 18684996]
49. Maniatis, T.; Fritsch, EF.; Sambrook, J. *Molecular cloning: a laboratory Manual*. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, New York: 1982.

50. Blobel G, Potter VR. Nuclei from rat liver: isolation method that combines purity with high yield. *Science*. 1966; 154:1662–1665. [PubMed: 5924199]
51. Sambrook, J.; Russell, DW. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, N.Y: 2001.
52. Mitchell JA, Fraser P. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev*. 2008; 22:20–25. [PubMed: 18172162]
53. Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A. The expanding RNA polymerase III transcriptome. *Trends Genet*. 2007; 23:614–622. [PubMed: 17977614]
54. Pagano A, et al. New small nuclear RNA gene-like transcriptional units as sources of regulatory transcripts. *PLoS Genet*. 2007; 3:e1. [PubMed: 17274687]
55. Stewart SA, et al. Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA*. 2003; 9:493–501. [PubMed: 12649500]
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 U S A*. 1992; 89:1827–1831. [PubMed: 1542678]
57. Bock C, et al. BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics*. 2005; 21:4067–4068. [PubMed: 16141249]
58. Christman JK. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene*. 2002; 21:5483–5495. [PubMed: 12154409]
59. Dunn JJ, Krippel B, Bernstein KE, Westphal H, Studier FW. Targeting bacteriophage T7 RNA polymerase to the mammalian cell nucleus. *Gene*. 1988; 68:259–266. [PubMed: 3065146]
60. Esteve PO, et al. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev*. 2006; 20:3089–3103. [PubMed: 17085482]
61. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008; 5:621–628. [PubMed: 18516045]
62. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. [PubMed: 19451168]
63. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010; 28:511–515. [PubMed: 20436464]
64. Harrow J, et al. GENCODE: producing a reference annotation for ENCODE. *Genome Biol*. 2006; 7(Suppl 1):S41–9.
65. Irizarry RA, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003; 4:249–264. [PubMed: 12925520]
66. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26:841–842. [PubMed: 20110278]



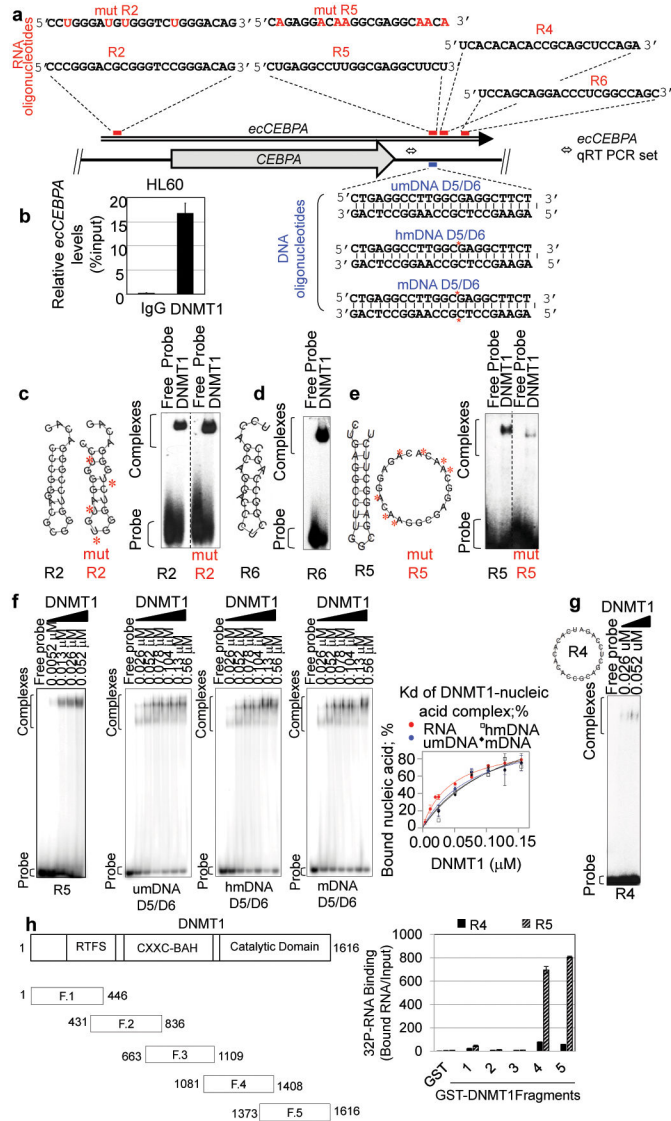
**Figure 1. Characterization of the *ecCEBPA***

**a**, Diagram of *CEBPA* transcripts; **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  $\pm$  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 indicates: position of target sequences for shRNA constructs (sh1–3); the fragment derived from the *ecCEBPA* employed for overexpression (R1) regions analyzed for changes in DNA methylation (distal promoter; coding sequence, CDS; and 3'UTR); **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  $\pm$  s.d. **(b)** and methylation of the *CEBPA* promoter **(c)**. DNA methylation changes are shown as the ratios of methylated to unmethylated CpGs in all clones analyzed per each construct (n=14); **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  $\pm$  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, 3'UTR); **f–g**, The results of transcription inhibition in U937 cells. **f**, *ecCEBPA* expression levels after treatment with Actinomycin D and ML-60218 in synchronized and unsynchronized cells. qRT-PCR, bars indicate mean  $\pm$  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 **e**). All bisulfite sequenced clones were analyzed by Fisher's exact test. \**P*<0.05; \*\**P*<0.01; \*\*\**P*<0.001.



**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; RNA and DNA oligonucleotides used in EMSA and REMSA. Asterisks indicate position of methylated cytosines; umDNA, hmDNA, and mDNA refer to unmethylated, hemimethylated, and methylated DNA probes, respectively; **b**, *ecCEBPA* is immunoprecipitated with anti-DNMT1 antibody. qRT-PCR, bars indicate mean  $\pm$  s.d.; **c**, RNA- DNMT1 binding is not affected by the absence of CpG dinucleotides (right panel). Left and middle panels: RNA oligonucleotide R2 and its mutated form R2 (asterisks indicate cytosines substituted into uridines), both able to form stem-loop-structures; **d**, RNA oligonucleotide 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, 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



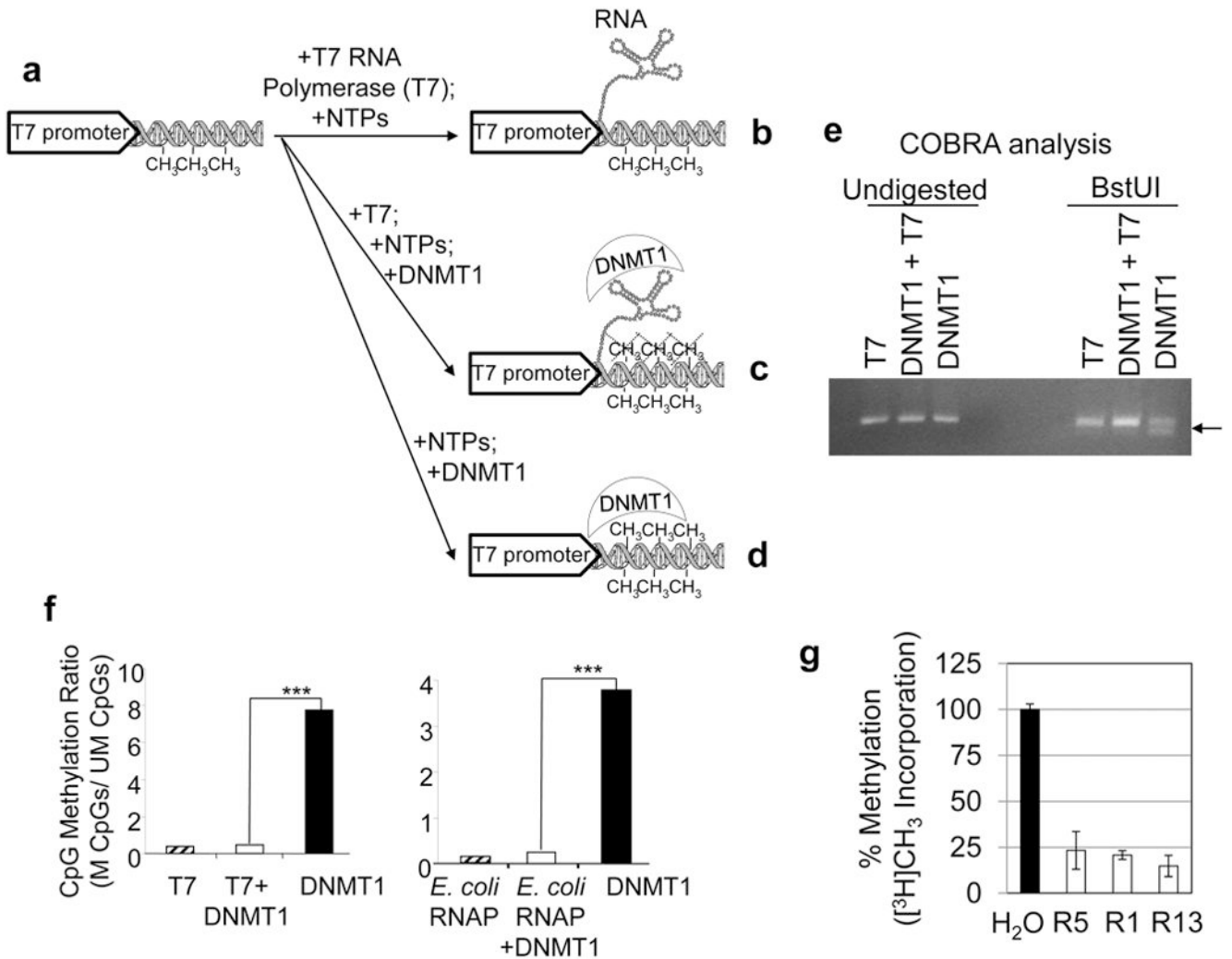
concentration of ssRNA and dsDNA oligonucleotides (1 nM) and increasing concentrations of DNMT1 protein; Right panel: Nonlinear regression analysis of bound RNA/DNA *versus* DNMT1 concentrations. Error bars indicate s.d. from two independent experiments; **g**, REMSA showing that RNA oligonucleotide R4, which is unable to form stem-loop structure, displays lower DNMT1 affinity as compared to R5 (Fig. 3f left panel) at the same DNMT1 concentrations; **h**, Left panel: Schematic diagram showing the DNMT1 domains and the GST-DNMT1 isolated fragments (F1–F5); Right panel: GST-DNMT1 pull down assay demonstrating binding of the folded RNA oligonucleotide R5 to the catalytic domain of DNMT1.

Author Manuscript

Author Manuscript

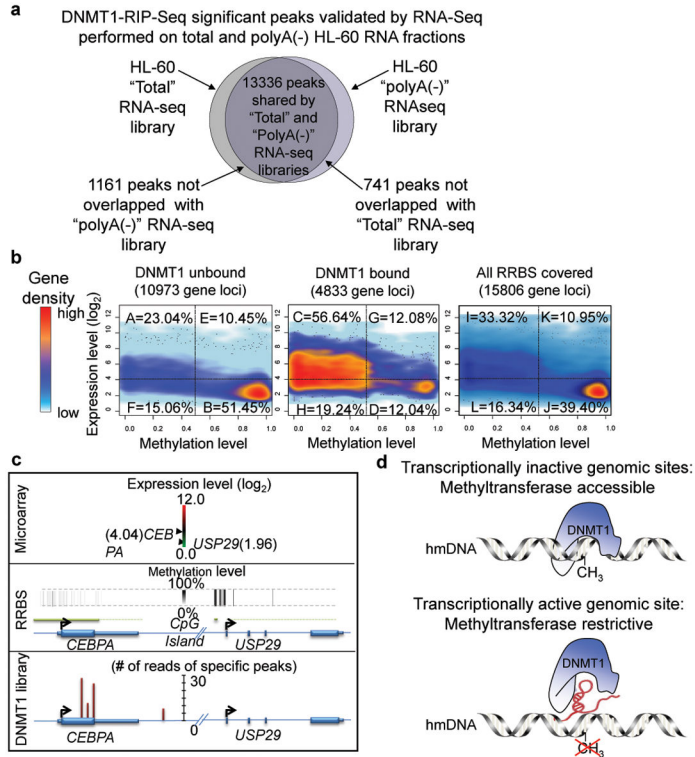
Author Manuscript

Author Manuscript



#### 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; **e**, DNMT1 exerts enzymatic activity only in the absence of transcription. COBRA analysis of methylation patterns acquired in reactions shown in **b-d**; **f**, DNA methylation changes as are shown as the ratios of methylated to unmethylated CpGs in all clones analyzed per each construct (n=5). The same effect was observed with two different RNA polymerases: T7 and Sigma-Saturated ( $\sigma 70$ )-Holoenzyme (*E. coli* RNA Polymerase). DNA methylation changes were analyzed by Fisher's exact test (\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ); **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 ribooligonucleotides are shown on Fig. 3a and Supplementary Fig. 3. Error bars indicate mean  $\pm$  s.d. (n=2).



**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<sup>+</sup>)-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 Excel File 2. **c**, Examples of genes from the C (*CEBPA*) and B (*USP29*) clusters. Peaks are visualized using the SSIRs<sup>48</sup>. **d**, Model of DNMT1 sequestration. Upper panel: DNMT1 can access transcriptionally inactive hemimethylated genomic regions. Lower panel: DNMT1 cannot access transcriptionally active hemimethylated genomic regions.