Supplementary Results

Characterization of ecCEBPA

To map the entire length of this transcript, we performed primer extension and 5’-3’ RACE on total and poly(A)^− RNAs isolated from HL-60 and U937 cells. That allowed us to identify the TSS of this long ncRNA at -0.8 kb upstream of the canonical CEBPA mRNA TSS in both HL-60 and U937 cell lines. The novel transcript terminated at +3.6 kb downstream from the CEBPA TSS (Supplementary Fig. 1c,d).

To exclude the possibility that this long transcript was an unspliced precursor of the intronless CEBPA gene, we examined the time course of the RNA synthesis and the RNA polymerase type(s) responsible for the initiation of messenger and ecCEBPA. HL-60 cells synchronized by double thymidine block (thymidine arrests cells at the G1/S phase boundary) were analyzed upon release from the block (with ~85% of cells entering the S phase; Supplementary Fig. 1f). Induction of ecCEBPA was observed immediately after the block release, at marked contrast to significantly lower increase of CEBPA mRNA level at the same time points (Supplementary Fig. 1g). Thus, ecCEBPA synthesis precedes the expression of its overlapping mRNA. To determine the polymerase(s) directing ecCEBPA transcription, synchronized HL-60 cells were treated with the RNA Polymerase II (RNAPII) inhibitor DRB, the RNA polymerase III (RNAPIII) inhibitor ML-60218, and α-amanitin, which inhibits RNAPIII only at high concentrations. In the presence of DRB (100 µM), we observed strong down-regulation of mRNA, but no decrease in ecCEBPA expression levels (Supplementary Fig. 1h), suggesting that the ecCEBPA and CEBPA mRNA utilize different polymerases and/or promoters. In contrast, we observed a marked down-regulation of ecCEBPA and RNAPIII-dependent 5S RNA and little effect on CEBPA mRNA
after 24 hour treatment with ML-60218 at low concentrations (25 µM; Supplementary Fig. 1i-j). However, prolonged treatment with or high concentrations of ML-60218 showed downregulation of both ecCEBPA and CEBPA mRNA (Supplementary Fig. 1j-k). Treatment with α-amanitin at 25-100 µg/mL significantly reduced levels of both CEBPA mRNA and ecCEBPA, but not the RNAPIII-dependent U6 snRNA, which was affected only by high (150 µg/mL) concentration of α-amanitin (Supplementary Fig. 1l).

Consistent with dual RNAPII/III regulation, we observed enrichment of RNAPII and RNAPIII in the close proximity to the ecCEBPA TSS by chromatin immunoprecipitation (ChIP) (Supplementary Fig. 1m) and ecCEBPA was immunoprecipitated with antibodies to both RNAPII and RNAPIII by ribonucleoprotein immunoprecipitation (RIP) (Supplementary Fig. 1n). Furthermore, (i) combined treatment with Tobacco Acid Pyrophosphatase (TAP), which removes the cap structure from “capped” RNAs, and 5'-Phosphate-Dependent-Exonuclease (Terminator), which digests RNA that has a 5'-monophosphate end; and (ii) RIP with anti-cap antibody (Anti-m3G-/m7G-cap)1, revealed the presence of a 2,2,7-trimethylguanosine (TMG) modification at the 5’end of ecCEBPA (Supplementary Fig. 1o-p), a common feature of RNAPII- and several RNAPIII-dependent transcripts (e.g., GAPDH and U6 snRNA, respectively2).

**DNMT1 binds to RNA**

REMSA was performed with number of RNA oligonucleotides not related to the ecCEBPA (single stranded [ss] R1 and R3; and double stranded [ds] R13; Supplementary Fig. 3d). We compared DNMT1 binding capacity of mismatched double-stranded (R13) and folded single-stranded (R1) RNAs to double-stranded DNAs (with the same primary sequence as the RNA oligonucleotides, D34 and D12; Supplementary Fig. 3d,e). We observed strong binding for both ds- and ssRNAs in the presence of poly (dI-dC) (non-specific competitor) and dsDNA
oligonucleotides (specific competitor) (Supplementary Fig. 3e). Similarly to the ecCEBPA-
corresponding R2 oligonucleotide, mutation of R1 (mut R1; CpG→UpG), neutral with regard to
the predicted secondary structures according to RNAfold3, did not disrupt the RNA-protein
complexes (Supplementary Fig. 3f). To exclude that the RNA-DNMT1 binary complex was a
case of trivial charge-charge interactions, we performed REMSA in the presence of increasing
concentrations of spermine, a molecule with four positive charges at high density4. Only a
thousand-fold molar excess of spermine began to moderately affect the binding (Supplementary
Fig. 3g), suggesting a strong element of structural recognition (on top of the simple ion
exchange) between DNMT1 and RNA.

Transcription interferes with DNMT1 methylation of DNA

The murine RAW 264.7 cell line was stably transfected with a construct carrying a human
CEBPA genomic segment under T7 promoter control (Supplementary Fig. 4k). After selection,
all tested individual clones demonstrated extensive methylation of the transgene (Supplementary
Fig. 4l,m). To specifically induce transcription of the transgene, the selected clones were
transfected with T7 polymerase-expressing mammalian constructs5 and DNA methylation was
assessed. Remarkably, T7 RNA polymerase-induced transcription led to a pronounced decrease
in DNA methylation of the transgene (Supplementary Fig. 4n-p). In vitro we have observed
inhibition of DNMT1 activity when equimolar concentrations of the protein and RNA
oligonucleotides were employed (Fig. 4g). In vivo we can envisage a similar enzymatic
impairment for transcripts expressed at equal or higher molar concentration than DNMT1 in a
locus specific manner. Our findings demonstrated an inverse correlation between the CEBPA
gene locus methylation and the levels of ecCEBPA in HL-60, U937 and K562 cell lines (Fig.
1d,e; Supplementary Fig. 2p). Among these cell lines, the highest ecCEBPA levels are detected in
the HL-60 cell line, characterized by virtual absence of *CEBPA* gene locus DNA methylation (Supplementary Fig. 2p). Contrariwise the lowest *ecCEBPA* levels (in K562 cell line) are associated with almost completely locus methylation (Supplementary Fig. 2p). Thus, cells expressing lower levels of *ecCEBPA* will not compete out DNMT1 as efficiently as compared to cells expressing higher levels of *ecCEBPA*. The core of DNMT1-RNA interaction resides in the ability of DiRs to remain proximal to their site of action. Therefore RNA levels and DNMT1 protein coexist in a dynamic state and it is conceivable that variations in this “stoichiometry” against or in favor of one of these factors will have opposite results: “protection or no protection” from gene locus DNA methylation.

**Supplementary Methods**

**Northern Blotting probes:**

*CEBPA* mRNA: 5’ – CCGCTCCTCCACGCTCTTCTGAGAAAGGGGTTGAAAACACATAGGGACTTG GGCTTGGAACCTAAGGTGTGTTCCCTAGTTGCTACTGAAAGGTGGAGGTGGCTCTCTAGGTCCAGGC CTCTCCCCACCTCCCTCCGCACACACCCACCCACCCAGCTGTAGGCTGCTTTCTTGGGG CGGACTCTACTGCGATGGGGTGGTACCAGTGGCACCAGTGAGGGAGCCCCACCCCGAGTCACACCAG AAAGCTAGTGTTGGTGACTGATGTATACCCCTGTTGGAGAGGAGAGCTTAGAGACTCGGCTGTGGGCGC-3’.

*ecCEBPA*: 5’ – GTCACATTTTGTAATAATACAGCATTTTCCCTGAGCAGGCAATCTCTGACTTTCA TGGAGCTTCATCCATCTCTGAGCCCTCTTAACTACCTAAGGGGTGTATTACTTCTCCACGGCAAG ACAATAAATAGCAGAGGACAAAGGCTTTCCAAATGGAAGTATGTCCTGAGAGCCTGAGGTGCTCTCTT GGAGTGCGAGGTGAGGGTTACGTGAGGCTTGGCTGGCGAGCCCCTTATCTGCCTCAGGGGAGAGGAGGAGGGGCTGAGGCTGAGGTCTTCTATCTGCCGGGGAGGAGGGAGTCCTCTGCCAGGGGTAGGGCTGGCTAACGGCCCTACCCACGAGCATGACCCAG ACGCGACGCTCCAGATTCCACGAGAATCCCTGAGGCTGAGGTGAGCAGCCCTGACTAGAGCAAG TCCAACTCTCTCTCTTCCCTAGACCCTGGCTTACACCCCTACC-3’.

**Primers used for 5’ RACE:** R4 – 5’-AGA GGC GCG CTT GCC TAC AGG TGA-3’; R6 – 5’ CTC GCC ACT GGC GCT GAG GCC TGA-3’; R8 – 5’-GAG TCT GGG TGG CCTG TCA AGT GTC T-3’; R10 – 5’-TTC CGC GCA GAG GTA CCT CTC GGA -3’; R14 – 5’-CGG AGA GGT CCT GAG CGC GGT GGC t-3’; R16 – 5’-GGG GAG GCT GCA GCG CAG AGC
AGC G-3’. **Primers used for 3’ RACE**: **AL21** – 5’-GTC ACA TTT GTA AAT AAT ACA GCA-3’; **AL23** – 5’-CCC TGG CGG CAA TCC TGA CTT TCA-3’; **AL25** – 5’-TCA TGA GCT CTC CAT CCA TCC TGA-3’.

**Primers used for TaqMan real time PCR**: human **CEBPA**: Forward 5’- TCG GTG GAC AAG AAC AG-3’; Reverse 5’-GCA GGC GGT CAT TG -3’; TaqMan Probe 5’-ACA AGG CCA AGC AGC GC-3’; **ecCEBPA**: Forward 5’-GGT TGT TGT CTG TGG GCC AGG TCA-3’; Reverse 5’-AGA GCT CAT GAA AGT CAG GAT TG-3’; TaqMan Probe 5’-AAT AAT AAT ACA GCA TTT TCC CTG GCC G-3’; human **CEBPG** Forward: 5’- GCC TAG AGG AGC AGG TAC AT -3’; Reverse: 5’- GCC TGG GTA TGG ATA ACA CTA -3’; TaqMan Probe: 5’-CGA CAC CAC TCA TGT CAA TGG CC -3’; human **snU6 snRNA**: Forward: 5’- CTC GCT TCG GCA GCA CAT ATA CTA - 3’; Reverse: 5’- AAG CGC TTC ACG AAT TTG CGT -3’; TaqMan Probe 5’- CGA TAC AGA GAA GAT GAC CC -3’; **TP73**: ABI Cat. # Hs01060631_m1; **18S rRNA**: ABI Cat. # 4310893E; human **5S rRNA**: ABI Cat. # Hs02385257_g1.

**Primers used for real-time RT PCR (Sybr)**: human **CEBPA**: Forward: 5’-CCG CTC CTC CAC GCC TGT CCT TAG-3’; Reverse: 5’-GCC CCA CAG CCA GAT CTC TAG GTC-3’; **ecCEBPA**: Forward: 5’-TCA TGA GCT CTC CAT CCA TCC TGA-3’; Reverse: 5’-CTG GCC GAG GGT CCT GCT GGA ATC-3’. **β-actin**: Promega Cat. # G5740; human **CSF3R** (1st peak) Forward: 5’-GTT AGA ATG CAT GTT GGG AAG G-3’; **CSF3R** (1st peak) Reverse 5’-GAGATGGCTGTTTGAGGG-3’; **CSF3R** (2nd peak) Forward: 5’-CAG AAA CGT GTC ATT GGC AAC-3’; **CSF3R** (2nd peak) Reverse 5’-GTC TCC CTA TCC CAC CCT AG-3’; **JUN** (1st peak) Forward: 5’-CGT GGC TCA TGC GGA CTG CTT TCT G-3’; **JUN** (1st peak) Reverse: 5’-CCT GGA AGA GAA AGT GAA GAC C-3’; **JUN** (2nd peak) Forward: 5’-CGT TGG ACT
GA TGA TGA GG-3'; **JUND** (2nd peak) Reverse: 5’-AGC ATG ATG AAG AAG GAC GC-3’; **MAF1** Forward: 5’-GGA CGA GAA AAC AGA AGC AAA G-3’; **MAF1** Reverse: 5’-CAC GGA GCC CAA AGT TAG G-3’; **CEBPE** (1st peak) Forward: 5’-GAG TAC ATG GCA GAG AAC GAG-3’; **CEBPE** (1st peak) Reverse: 5’-CAA CCT CTT CCG CCA GAT T-3’; **CEBPE** (2nd peak) Forward: 5’-GCT ACA ATC CCC TGC AGT AC-3’; **CEBPE** (2nd peak) Reverse: 5’-CGT TCT CAA GGTA AGA GCA GG-3’.

Short hairpins RNAs sequences used for downregulation: SC:
5’-ATCTCGCTTGGGCGAGAGTAA-3’; **Sh#1**: 5’-AATAAATAGCAGAGGACAAGG-3’; **Sh#2**: 5’-CAGGCAAGCATAAATAAATAGC-3’; **Sh#3**: 5’-GAAGAGGCTCCAGTGGTAGG-3’.

Primers used to amplify the *ecCEBPA* 3’ and 5’ regions (R1 and R2) and the Unrelated Region (UR) employed for the overexpression experiment: **R1**: forward: 5’-ATG TCG GTG TCT TTT TAA AAC CAG-3’; reverse: 5’-GCT AAG CTT CCA GAG TGT AAA AGG-3’; **UR**: forward: 5’-TA TAC AGC CAT GAA AGA AAC TTAC-3’; reverse: 5’-AGT TTT ACT GTG GTG TGT TTC-3’; **upR**: forward 5’-CCC GTC CAA AAG CGT CAA GC-3’; reverse 5’-AGT CCC AGT GAT TAA TAG CCC AAC-3’; **dwR**: forward 5’ -CCG AGT CCT CCT CCT GGT CCT GAG-3’; reverse 5’-CCC TGC AGG AAG CGG TGG GCC-3’.

Full R1, upR, dwR and UR regions:

**R1**: 5’-ATGTCGGTGTCTTTTTAAACGAAAAGAGCTACTTTCCAAAGGTGTCTGTGGGCCAG GTCACTATTGTAATAAATAAATAGCAGCATATTTCCTGCCGGCAATCTGTACCTTCCAGCTCTCCAT CCATCTTGAGCCCTCTTACCTAAGGGGTGACTTACTTCCCCAGGCAAGCAAATATTAG CAGGAGCAAGGCTCTAAATGGGATATGCTCCAGAGCTGAGCTCCTTTGGGTTCAGGGGA GGGGGTTGGAAGGTTACTGCGCTGAGGCTTTGGGAGGCTCTTACTCTTTACTGCTCCCAGGGAGAGG AGAGGGAGTCCTCTGTGCCAGGTTAGGCTTGCTAAGCAGCCTAGGCTCAAGGAGGCCCTTTTG TGCAGACTTTTGGCCTAACCTACCTAGCTGAGCCCTGCGCCAATCACACACACCCAGGCTCCA GATTCACGCCAAGCCTCCGGCCAGCAGAAGAGGCGCTCCAGTGTTAGGACCCCTCCCAACCTTCTT CTCTTTCCCTAGACCATGTGGCTACACCTTACCCCTGCGAGGCCCAAGGAGCCTGAGAG
AGACCCCTCCATCCAGCCTCCACACCTGCCCAGCCAGCCAGCCACCCCTCACAAGAAGGGGCTGAGG
CAGGCTCTGCAACAGGGGTCTGCTGCTGCCCCACACACTGCTGTTCTGCAAGACCACCACCCAGCATC
CACACTCCCCACTGCAATTACATCCAGACCAGCAGGGGACTGAGTGTGTTGTGTCGCCCACAATCT
CAATGAGCAACCTGTAAGGTCAGGCACTCAGGCCCCACCCCTTTAGCAGCAGTGCAGGCAAGCAAG
GGAGGGACAGGGGCTGAGGCTTCTGGCTCCACACACTGGCTGTTCAAGAACCCACCCCATAC
CACACTCCCCACTGCAATTACATCCAGACCAGCAGGGGACTGAGTGTGTTGTGTCGCCCACAATCT
GGACTATATGTTGAATAGGAACTTGATGTTGAGCCAGATGAAACTGCTCTCTT
TACTGCGATCGTCTGGAAGCTCGGGGCTTCACAGGCTGTTGTTTTTACTGCTGCCCCTCGGCCG
CTCCTGGGCGGTCCTTTCAGGCTCTAAGGCACACTGCTGTTGAGGGCTGGTCTC
CACCCTGTAGGCAAGCGCGCCTCTAGCCGGGACGCAGGCGGCGTCAGGCCTAGCGCCAGTGGCG
AGGGGCGGCGGCGGCGGAGGACACTTGAGGGCTCCCAAGACTCTGAGCTCCGCAAGGCTGAGG
GGGCTGAGGCTCAGGTAAGGCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG
GGTCTTCCCTGTGCGTGCTCGGACCTGGCAGCCGGCGGCCTGGACACCCTCGCTCCCGCCGGGTG
CGCCCAACCTGAAATGGGAGGCCGCGGGGGAGAGAGCGCCTGAGGCTGGGTGGGCTTC
CCGCTCGGAATCTCGGGC CCCAGAGTTAAGTTTGT-3'.

**upR:** 5'-CCACAGGTGAATGCTATATGAATGGGTCCTGCATTTGGCTCAGCAAGCAGCTCTC
CGGCCCTACACTCGGATATTTATTTAATACGTATCTGGTGTTTTTACTCCCTCCGGCTCAAAAAGCGTCA
AGCTAAAAATATCTCTGCTTGCTGGAGAGAACTGGAAATTTTGAACAGTCTTCTTCTTCT
TACTGCGATCGTCTGGAAGCTCGGGGCTTCACAGGCTGTTGTTTTTACTGCTGCCCCTCGGCCG
CTCCTGGGCGGTCCTTTCAGGCTCTAAGGCACACTGCTGTTGAGGGCTGGTCTC
CACCCTGTAGGCAAGCGCGCCTCTAGCCGGGACGCAGGCGGCGTCAGGCCTAGCGCCAGTGGCG
AGGGGCGGCGGCGGCGGAGGACACTTGAGGGCTCCCAAGACTCTGAGCTCCGCAAGGCTGAGG
GGGCTGAGGCTCAGGTAAGGCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG
GGTCTTCCCTGTGCGTGCTCGGACCTGGCAGCCGGCGGCCTGGACACCCTCGCTCCCGCCGGGTG
CGCCCAACCTGAAATGGGAGGCCGCGGGGGAGAGAGCGCCTGAGGCTGGGTGGGCTTC
CCGCTCGGAATCTCGGGC CCCAGAGTTAAGTTTGT-3'.

**dwR:** 5'-CCGAGTCCTCCTCCTCCTGCTGCTGAGTCTCCAGGGCTGCTTTCTTTTATGGGACAGAAGAG
GATGAGGAGGAAGGTTTCTTGGGGCGCATGGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG
GGTCTTCCCTGTGCGTGCTCGGACCTGGCAGCCGGCGGCCTGGACACCCTCGCTCCCGCCGGGTG
CGCCCAACCTGAAATGGGAGGCCGCGGGGGAGAGAGCGCCTGAGGCTGGGTGGGCTTC
CCGCTCGGAATCTCGGGC CCCAGAGTTAAGTTTGT-3'.

**UR:** 5'-TATACAGCCATGAAAGAACTTACTTCTCCAAGGCTGCTTTCTGCTGAGACCCTTACAGCC
CAGACACAGATCCAGAATTTAACCAGGAGAAGTTGAGATGAGGCTCACAATATAAAAGTTAATCATG
AATGAGCGAAAAAGCAAGCAACTATATGTCACTGAGGGAGATTTAGTAAGATATTGAACACAA
TTTGAAAGACTACATAATAATAATAGATATTTTACATTTACATAATAATAGATATAATCATATAG
ATATTTAAAAATATAATTTGAGAAGAAAAGAGAAGGTCAGAAGCATATATATAAAATATATATATATAAATAAATATA
ACAAGTATGATTTTCAAAAAATCAAGAAGGTGTTGCTCCTCAAAAATATACCTTTGACTGATATAAAATAGA
AGCTCAGCGAATATAGCAAAATTTTAGTGTCAGTCACAGGAAATAAGAAAATATAAGACATATC

SUPPLEMENTARY INFORMATION

RESEARCH

Di Ruscio, Ebralidze et al. Supplementary Information

WWW.NATURE.COM/NATURE | 7
TGAGGAAATTTCTGACATGCACACACAGATATGAAAACAAAGGAATATATGAGAGTCTAAGAT
TCTTTGGAGATATGAATGAGAAGGCGCAACACATATATAGGAATTCGAGAAGG
TGAAAACCTGAAAACCTTCAAACCTATTATTAGTTGGAGAAATTTCAGACACTAGGGGATTGAAG
TCCTCACTTGAAGATGCACACTGAGTCCTAAGAAGGATAAAATTAAGAACAACACACCACAGT
AAAAACT-3'.

Primers used for COBRA and bisulphite sequencing:

**CEBPA -1.4 kb region-**: forward: 5'-GGT GTT TTT AGT TGT GTT TTT TT-3'; reverse: 5'-AAA CCC TAA AAC CCC TTA-3'; **CEBPA Distal Promoter (-857 to -634 bp)**: forward: 5'-TAG TTT YGT TAG TTT GGG GGG TTT-3'; reverse: 5'-TCT AAT CTC CAA ACT CCT ATA-3'; **CEBPA Distal Promoter (-857 to -384 bp)**: forward: 5'-TAG TTT YGT TAG TTT GGG GGG TTT-3'; reverse: 5'-TCT AAT CTC CAA ACT CCT ATA-3'; **CEBPA coding region**: forward: 5'-AGG TTA AGG YGG TTG TGG GTT TTA-3'; reverse: 5'-CCA ACT ACT TAA CTT CAT CCT ATA-3'; **CEBPA 3'UTR**: forward: 5'-AGG TTY GTG GTA GGA GGA GGG TTT A-3'; reverse: 5'-TAA CCC ACR ACC TAA CTT TCT AAT-3'; **TP73 promoter**: forward: 5'-GTG GGY GGT TTY GTY GGG TTT TGT-3'; reverse: 5'-ACC CCT AAA CRA ATT ATA TAA A-3'; **CEBPG promoter**: forward: 5'-GAA GTG AAT TTT TTA AAA TGA TTT-3'; reverse: 5'-TTT TGT TTT AGT TTA GTT TTA AGT TGG GA-3'.

RNA and DNA oligonucleotides used for RNA and DNA Electromobility Shift Assay (EMSA): **R1** 5'-CCG GGA CGC AGG CGG CGU CAG GC-3'; **mutR1** 5'-CUG GGA UGC AGG UGG UGU CAG GC-3'; **R3** 5'-GCC AGU GGC GAG GGG CGG CGC GG-3'; **R2** 5'-CCC GGG ACG CGG GTC CGG GAC AG-3'; **mutR2** 5'-CCU GGG AUG UGG GTC UGG GAC AG-3'; **R4** 5'-UCA CAC ACA CCG CAG CUC CAG A-3'; **R5** 5'-CUG AGG CCU UGG CGA GGC UUC U-3'; **mut R5** 5'-CAG AGG ACA AGG CGA GGC AAC A-3'; **R6** 5'-UCC AGC AGG ACC CUC GGC CAG C-3'; **umD5** 5'-CTG AGG CCT TGG CGA GGC TTC T-3';
**mD5** 5'-CTG AGG CCT TGG C(CH₃)GA GGC TTC T-3'; **umD6** 5'-AGA AGC CTC GCC AAG GCC TCA G-3'; **mD6** 5'-AGA AGC CTC (CH₃) GCC AAG GCC TCA G-3'.

**Primers used for in vitro transcription/methylation assays:**

Forward: 5’-GGA AGG GCG ATC GGT GCG GGC CTC-3'; reverse (biotinylated): 5’-Biotin-CAG CCC TCG AGG CCC GAA GCC ACC-3'; reverse (not-biotinylated): 5’-CAG CCC TCG AGG CCC GAA GCC ACC-3’.

**Table 1. Antibodies and Primer Sequences used for Chromatin Immunoprecipitation qPCR**

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>anti-RNAPII</td>
<td>clone CTD4118 (Millipore)</td>
</tr>
<tr>
<td>anti-RNAPIII</td>
<td>clone 1900 (generous gift from R. White)</td>
</tr>
<tr>
<td>IgG</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>GenDes-F⁶</td>
<td>5’-GGC TAA TCC TCT ATG GGA GTC TGT C-3’</td>
</tr>
<tr>
<td>GenDes-R⁶</td>
<td>5’-CCA GGT GCT CAA GGT CAA CAT C-3’</td>
</tr>
<tr>
<td>HU6-F⁶</td>
<td>5’-CGC TTC GGC AGC ACA TAT AC-3’</td>
</tr>
<tr>
<td>HU6-R⁶</td>
<td>5’-AAA AAT ATG GAA CGC TTCA CG-3’</td>
</tr>
<tr>
<td>-4.1-F</td>
<td>5’-CAC ATT TCA GCT TCA GAT AAA GTC-3’</td>
</tr>
<tr>
<td>-4.1-R</td>
<td>5’-CTG ATT CTG GAG TCA ATC ACA GCA-3’</td>
</tr>
<tr>
<td>-1.8-F</td>
<td>5’-AGG TCT CCC TCC ACA GGT GAA TGC-3’</td>
</tr>
<tr>
<td>-1.8-R</td>
<td>5’-AGT CCC AGT GAT TAA TAG CCC AAC-3’</td>
</tr>
<tr>
<td>-1.2-F</td>
<td>5’-AAG AAT GCG AGG GAC GCA CGT GGC-3’</td>
</tr>
<tr>
<td>-1.2-R</td>
<td>5’-CCG CCG GGC TGC CAG GTC CGA GCA-3’</td>
</tr>
<tr>
<td>-0.9-F</td>
<td>5’-TCC CCG TGC TCG CCC CGG CGT CCA-3’</td>
</tr>
<tr>
<td>-0.9-R</td>
<td>5’-CGG AGA GGT CCT GAG CGC GGT GGC-3’</td>
</tr>
<tr>
<td>-0.2-F</td>
<td>5’-CCG CCC CGG AGC CGC TGG GCG GCA-3’</td>
</tr>
<tr>
<td>-0.2-R</td>
<td>5’-GGG CGC GCC CGG GTA GTC AAA GTC-3’</td>
</tr>
</tbody>
</table>
Supplementary References


**Site of mRNA polyadenylation**

<table>
<thead>
<tr>
<th>HL-60 cells</th>
<th>U937 cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL16</td>
<td>AL25</td>
</tr>
</tbody>
</table>

**ncRNA TSS in HL-60 and U937 cells**

Blue capital letters correspond to 5'3' ends of transcripts in HL-60 cells; Red capital letters correspond to 5'3' ends of transcripts in U937 cells; Black capital letters correspond to 5'3' ends of transcripts in both HL-60 and U937 cells. Arrows indicate the longest detected transcripts. Boxed sequences indicate highly conserved homology regions.

**Supplementary Figure 1 (part 1)**
Supplementary Figure 1 (part 2)

**e** CEBPA mRNA

**f** Cell Cycle phase

**g** CEBPA mRNA

**h** CEBPA mRNA

**i** ecCEBPA

**j** ecCEBPA

Relative expression levels; % of mock (normalized to 18S)

HL-60 + ML-60218 (25 µM, 24h)
Supplementary Figure 1 (part 3)
Supplementary Figure 1. Characterization of ecCEBP A. a, Diagram of coding and noncoding transcription within CEBPA locus. The CEBPA mRNA probe is located in the 3’UTR (black star), and the extra-coding probe is located after the poly(A) signal outside the coding sequence (white star). b, Visualization of noncoding CEBPA transcripts by Northern blot hybridization performed on nuclear RNA fractions of three different cell lines (left panel). Two middle panels demonstrate enrichment of noncoding CEBPA transcripts in nuclear poly(A)− fraction. Right panel illustrates extent of poly(A)− fraction purity. c, Primer extension performed on total cellular and cellular poly(A)−/− RNAs of HL-60 cell line followed by Southern blot hybridization (black star indicates the hybridization probe). The AL16 oligonucleotide is located in the coding region of the CEBPA gene. Primer extension reveals ~3 kb band in the poly(A)− fraction, marking the major TSS for noncoding transcripts at ~1 kb upstream of the mRNA TSS. d, ecCEBP A is initiated and terminated in the vicinity of highly conserved homology regions. Upper panel: Strategy of mapping noncoding CEBPA transcripts by 5′, 3′ RACE in two cell lines, HL-60 and U937. R10, R14, R16, AL21, AL23, AL25 are primers used in RACE. Two bottom panels present position of 5′(TSS) and 3′(TES) ends of noncoding and coding transcripts in both cell lines (capital bold letters indicate TSS and TES). e, CEBPA mRNA and ecCEBP A levels assessed by qPCR show concordant expression in human tissues. Bars indicate mean ± s.d. f, Synchronization of HL-60 cells by double thymidine block. Treatment with thymidine (2.5 mM) arrests the cells in G1/S phase border. FACS analysis shows synchronization after the treatment. g, Levels of coding and ecRNA immediately after release from double thymidine block. Induction of ecCEBP A preceded and surpassed expression of CEBPA mRNA (qRT-PCR; Bars indicate mean ± s.d.). h, DRB treatment did not reduce ecCEBP A levels upon release cells from double thymidine block; i, Treatment with the RNAPIII inhibitor ML-60218 significantly reduced levels of ecCEBP A and RNAPIII-transcribed 5S rRNA and only moderately affected CEBPA mRNA at 24 hours (P=n.s.; qRT-PCR; Bars indicate mean ± s.d.; n=3). Statistical analysis performed by paired T test (* P<0.05; ** P<0.01; *** P<0.001). j, Effect of RNAPIII inhibitor ML-60218 (25 µM) at different time points. Downregulation of ecCEBP A preceded downregulation of CEBPA mRNA (qRT-PCR; Bars indicate mean ± s.d.). k, Effect of different concentrations of RNAP III inhibitor ML-60218 (12.5-100 µM). Treatment with high drug concentrations (50 and 100 µM) significantly reduced levels of ecCEBP A and RNAP III-transcribed 5S rRNA and affected CEBPA mRNA. l, Effect of different concentrations of RNAP II/III inhibitor α-amanitin (5-150 µg/mL). Treatment with α-amanitin at 25-100 µg/mL significantly reduced levels of both CEBPA mRNA and ecCEBP A, but not RNAPIII-dependent U6 snRNA, which was affected only by high (150 µg/mL) concentration of α-amanitin.(qRT-PCR; Bars are mean ± s.d; n=3). m, ChIPpcr analysis of the CEBPA promoter regions. Upper panel: Diagram showing transcription within the CEBPA locus and position of used primer sets (double-headed arrow). G.D.: gene desert. Right panel: control experiment demonstrating specificity of the RNAPIII antibodies. Two bottom panels: qPCR analysis of immunoprecipitates. Peaks of RNAP II and III enrichments coincide with the ecCEBP A TSSs .(qRT-PCR; Bars are mean ± s.d). n, ecCEBP A is immunoprecipitated with antibodies to both RNAPIII (clone 1900; generous gift from R. White) and RNAPII (clone CTD4118; Millipore). Bars are mean ± s.d. o, Tobacco Acid Pyrophosphatase (TAP) removes the cap structure from ecCEBP A. Upper panel: Outline of the experiment: equal amounts ofHL-60 RNA was treated with: a) Tobacco Acid Pyrophosphatase (TAP; which generates "decapped" 5´-monophosphorylated terminus on the capped RNA molecules); b) 5´-Phosphate-Dependent Exonuclease (Terminator; which digests RNA that has a 5´-monophosphate end); and c) sequential digestion with TAP and Terminator. Bottom panels: 18S is digested by Terminator without prior TAP treatment. ecCEBP A becomes sensitive to Terminator only after removal of the cap structure with TAP (similar to CEBPA mRNA and U1 snRNA, not shown). p, ecCEBP A and CEBPA mRNA are specifically immunoprecipitated with anti-cap antibody (Anti-m3G-/-m7G-cap; Synaptic Systems). Bottom panels: control immunoprecipitation of capped GAPDH and no enrichment for capless 18S. Bars are mean ± s.d.
Supplementary Figure 2 (part 1)

**a**
U937
Distal promoter (-857 to -634 bp from CEBPA TSS)

- Unmethylated (UM) CpGs
- Methylated (M) CpGs

**b**
K562 cells

c

**d**

<table>
<thead>
<tr>
<th>Relative expression levels; (normalized to 18S,% UR)</th>
<th>ecCEBPA</th>
<th>CEBPA mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative expression levels; (normalized to 18S,% UR)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1st 2nd 3rd 4th</td>
<td>1st 2nd 3rd 4th</td>
<td></td>
</tr>
<tr>
<td>5th 6th 7th 8th</td>
<td>5th 6th 7th 8th</td>
<td></td>
</tr>
<tr>
<td>9th 10th 11th 12th</td>
<td>9th 10th 11th 12th</td>
<td></td>
</tr>
</tbody>
</table>

**e**
K562
Distal promoter I (-857 to -634 bp from CEBPA TSS)

**f**

Supplementary Figure 2 (part 1)
**Supplementary Figure 2 (part 2)**

### MassARRAY Target regions

**CpG Islands**

![MassARRAY Target regions](image)

**Scale**
- Chr19:
- MassARRAY Target regions

**UR**
- R1
- K562
- HL-60
- K562 Mock
- K562 Aza

**3'UTR**
- -69 kb
- -0.7 kb
- +32 kb
- +29 kb
- +20 kb
- +9 kb
- +6 kb
- +1.4 kb
- +7 kb
- +14 kb
- +35 kb
- +40 kb
- +45 kb
- +69 kb
- +69.5 kb
- +70 kb

**Proximal promoter**
- -8.5 kb
- -40 kb
- -45 kb
- -47 kb
- -69 kb
- -69.5 kb
- -70 kb

**Distal promoter**
- (-857 to -634 bp from CEBPA TSS)

**CEBPA**
- CEBPA gene locus

**CEBPG**
- CEBPG gene locus

**Methylation Delta**
- Δ [Average (Methylation UR) – Average (Methylation R1)]

**Methylation Ratio**
- (M CpGs/UM CpGs)

**Relative expression levels, % of mock (normalized to 18S)**
- CEBPA mRNA

**K562**
- CEBPG promoter region
- (-571 bp to -312 bp from CEBPG TSS)

**HL-60**
- CEBPA mRNA

**K562 Mock**
- CEBPG mRNA

**K562 Aza**
- CEBPG mRNA

**UR**
- R1
- K562
- HL-60
- K562 Mock
- K562 Aza

**Unmethylated (UM) CpGs**

**Methylated (M) CpGs**

**Cel-Size**
- Chr19:
- MassARRAY Target regions

**Supplementary Information**
**Supplementary Figure 2 (part 3)**

**k** K562 TP73 promoter region (-214 bp to -59 bp from TP73 TSS)

<table>
<thead>
<tr>
<th>Unmethylated (UM) CpGs</th>
<th>Methylated (M) CpGs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**l**

| CEBPA | Bin 1: promoter  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pvalue = 5.44e-22</td>
</tr>
<tr>
<td></td>
<td>qvalue = 3.09e-10</td>
</tr>
<tr>
<td></td>
<td>Meth.diff = -57.14%</td>
</tr>
</tbody>
</table>
| Bin 2: 1st exon  
|       | pvalue = 7.83e-12 |
|       | qvalue = 1.53e-10 |
|       | Meth.diff = -91.67% |

*Total number of gene loci covered by a least by 1 bin

**m**

<table>
<thead>
<tr>
<th>U937 Cells</th>
<th>Thymidine 2.5 mM; 18 h</th>
<th>Release</th>
<th>Thymidine 2.5 mM; 18 h</th>
<th>Release</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2x wash</td>
<td>Release</td>
<td>2x wash</td>
<td>RNA/DNA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DMSO</th>
<th>Actinomycin D 0.8 µM</th>
<th>ML-60218 150 µM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Synchronized U937 cells</th>
<th>Unsynchronized U937 cells</th>
</tr>
</thead>
</table>

**n**

<table>
<thead>
<tr>
<th>DNA ladder</th>
<th>Digested</th>
<th>Undigested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**o**

<table>
<thead>
<tr>
<th>ecCEBPA</th>
<th>CEBPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**p**

<table>
<thead>
<tr>
<th>HL-60 + DMSO/ML-60218 (24h)</th>
<th>DNA ladder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Undigested</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>25 µM ML-60218</th>
<th>100 µM ML-60218</th>
</tr>
</thead>
</table>

**Supplementary Figure 2 (part 3)**
Supplementary Figure 2. Downregulation and Overexpression of ecCEBPA leads to increased and decreased DNA methylation of the CEBPA locus. 
a, Lollipop representation of bisulfite sequenced clones in U937 ecCEBPA knockdown cells. b, Diagram indicating the regions employed for overexpression: ecCEBPA3’ region (R1), immediately upstream (upR), downstream (dwR) ecCEBPA regions and unrelated region (UR) located ~45 kb downstream to CEBPA gene (double-headed arrows). Regions of CEBPA gene analyzed for changes in DNA methylation (Distal Promoter, CDS and 3’ UTR; white double-headed arrows). c, Northern blot analysis of untransfected K562 and K562 stably transfected with constructs R1, UR, and EV (empty vector) demonstrating uniformity of overexpression levels of R1 and UR constructs. Two identical filters were hybridized with the probes corresponding to R1 and UR regions. 32P-labeled probes were generated with the same specific activity and filters were exposed to the X-ray film for the same time. d, Effect of ecCEBPA and regions neighboring ecCEBPA on CEBPA mRNA expression (qRT-PCR; Bars indicate mean ± s.d.); e, Effect of ecCEBPA upregulation in K562 on methylation of the CEBPA locus and lollipop representation of bisulfite sequenced clones; f, COBRA analysis of the coding and 3’ UTR regions of CEBPA gene. White arrows indicate incompletely digested PCR products of bisulfite-converted genomic DNAs isolated from cell stably transfected with R1 and UR constructs. g, Comparison of DNA methylation changes imposed by ecCEBPA overexpression and 5-Aza-CR treatment by MassARRAY analysis. Upper part: Diagram showing position of MassARRAY target regions; CEBPA and CEBPG genes, CpG islands. Middle part: Heatmap representing the methylation level of individual CpGs for overexpression samples (UR, R1); two control cell lines (K562, HL-60); and mock /5-Aza-CR-treated cells. Lower part: Snapshots of regions corresponding to CEBPA and CEBPG genes. h, Methylation changes (Methylation Delta) induced by ectopic ecCEBPA expression are significant only within CEBPA locus. The region +6 kb includes the R1 fragment and was excluded from the statistical analysis. i, j, k, Comparative analysis of CEBPA, CEBPG and TP73 expression and DNA methylation changes following 5-Aza-CR treatment and ecCEBPA overexpression in K562 cells. Shown are: Lollipop representation of bisulfite sequenced clones; DNA methylation changes shown as the ratios of methylated to unmethylated CpGs in all clones analyzed; qRT-PCR (Bars indicate mean ± s.d.); l, RRBS analysis of R1- versus UR-expressing K562 cells. Venn diagram presenting the number of gene loci hypomethylated, at levels similar to CEBPA, in R1- versus UR-expressing K562 cells. The statistical comparison has been done using the R/MethylKit package using a q value <0.01 and a methylation difference > 50%. The right box indicates the two differentially methylated bins in the CEBPA promoter/1st exon. m, Schematic diagram showing synchronization of U937 cells by double thymidine block followed by treatment with RNA Polymerase Inhibitors. Upon release from double thymidine block, cells were treated with DMSO; Actinomycin D (RNA Polymerase I, II and III Inhibitor), 0.8 µM; and ML-60218 (RNAPIII Inhibitor), 150 µM. After 7 hours, cells were ficolled and RNA and DNA were collected. n, Blocking transcription during the S phase led to significant increase in DNA methylation of CEBPA distal promoter. Lollipop representation of bisulfite sequenced clones. o, Left Panel: Schema of COBRA design; Right Panel: COBRA analysis of CEBPA distal promoter. HL-60 cells were treated with ML-60218. p, COBRA analysis showing CEBPA promoter methylation pattern in HL-60, U937 and K562 cells (schema of COBRA is shown in o). All bisulfite sequenced clones were analyzed by Fisher’s exact test. MassARRAY data were analyzed by Student’s T test. (* P<0.05; ** P<0.01; *** P<0.001).
Supplementary Figure 3 (Part 1)
Supplementary Figure 3. RNA–DNMT1 interaction. a, ecCEBP A is specifically immunoprecipitated with anti-hDNMT1. Panels from left to right: (1) Western blot of DNMT1 hypomorph cell line “HCT116 hypo” and wild type counterpart HCT116. (2) DNMT1 RIP qRT-PCR on HCT116 and HCT116 hypo, bars indicate mean ± s.d. (3) Western blot of HEK 293T cells transfected with a vector expressing DNMT1-HA tagged protein (hDNMT1-HA) or empty vector (EV); western blot analysis confirmed DNMT1-HA overexpression; short (1 sec.) and long (30 sec.) exposures are presented (panel 3). (4) DNMT1 RIP qRT-PCR of HEK 293T cells expressing DNMT1-HA tagged protein. IP with both DNMT1 and HA antibodies demonstrates comparable ecCEBP A precipitates in DNMT1-HA-expressing cells. (5) DNMT1 RIP qRT-PCR performed on U937 cells. qRT-PCR, bars indicate mean ± s.d. b, Diagram of coding and “extra-coding” transcription within the CEBP A locus. The TaqMan probe set (black double-headed arrow) detects both CEBP A transcripts (polyadenylated CEBP A mRNA and polyA (-) ecCEBP A). c, Non-polyadenylated CEBP A transcripts are enriched in DNMT1-RNA precipitates. Left panel: Schematic diagram of the procedure; Middle panel: extent of poly(A)⁺/− fractions purity: 18S is not polyadenylated; Right panel: enrichment of non-polyadenylated ecCEBP A in DNMT1-RNA precipitate (qRT-PCR, bars indicate mean ± s.d.). d, Diagram showing sequences and positions of ecCEBP A unrelated RNA oligonucleotides. On the top, an example of a RNA sequence, located 5’ upstream to the ecCEBP A, with 5’ and 3’ sequences able to fold in a stem and loop-like structure; single stranded (ss) RNA oligonucleotides (R1 and R3) corresponding to 5’ and 3’ sequences of the RNA shown on top; the double-stranded (ds) RNA oligonucleotide (R13) represents an imperfect duplex between ss RNA oligonucleotides R1 and R3; dsDNA oligonucleotides corresponding to the same sequences as RNA oligonucleotides, D12 and D34. e, RNA and DNA EMSA performed with the increasing amounts of competitors: polydI-dC and dsDNA. Left and middle panels illustrate DNMT1 binding to ds RNA and DNA oligonucleotides (R13 and D34, D12, respectively); Right panel illustrates DNMT1 binding to folded single-stranded RNA (R1). f, Upper panel: secondary structures: ss R1 and its mutated form (mutR1; four cytosines within CpG dinucleotides were substituted with uridines). Both R1 and mutR1 oligonucleotides are able to form stem and loop-like structures. Asterisks indicate C to U substitutions. Bottom panels: Left, DNMT1 binding to folded ss RNA is not affected by the absence of CpG dinucleotides; Right, original gel, the areas outlined by dotted red lines indicate the parts shown in the left panel. g, REMSA performed in the presence of increasing concentration of spermine. No drastic changes in the RNA-DNMT1 binding were observed.
Mix mDNA with 10-fold molar excess of not-biotinylated umDNA

Check for methylation efficiency

Restriction enzyme digest

Capture biotin labeled DNAs on streptavidin magnetic beads

Uncut BstUI

Position of primers for COBRA

Unmethylated (UM) CpGs
Methylated (M) CpGs

E. coli RNA pol
E. coli RNA pol +DNMT1
DNMT1
Supplementary Figure 4 (Part 2)

**Human chr. 19**

4.3kb 2.8kb 2.4kb

hCEBPA

9.5 kb fragment was cloned into pBluescript II plasmid and stably co-transfected with Neomycin-carrying vector into murine RAW267.1 cell line

Transfection, G418 selection, PCR and Northern blot analysis

Individual clones carrying human sequence

COBRA analysis demonstrating extensive methylation of the transgenic sequences in clones ## 1 and 2.

Individual clones carrying human sequence were transfected with T7-expressing mammalian constructs (at 0 and 24 hours)

Induction of the T7-driven transcripts

qRT PCR and COBRA

qRT PCR with T7-specific primer set

qRT PCR with human-specific primer set

Supplementary Figure 4 (Part 2)
Supplementary Figure 4. Generation of hemimethylated DNA (a-j) and transgenic cell lines (k-p). Diagram illustrating the generation of hemimethylated DNA (hmDNA) for the *in vitro* transcription-methylation assay presented in Figure 4. a and b, The 5’ upstream region of the *CEBPA* locus was amplified using a forward primer carrying the T7 promoter sequence and biotinylated (arrow-head B) or non-biotinylated reverse primers containing a unique BstXI restriction site ( unmethylated DNA; umDNA). c, Biotinylated umDNA was *in vitro* methylated with M.SssI (NEB). d, Efficiency of methylation was assessed by restriction digestion of the methylated DNA (mDNA) with methylation insensitive and sensitive restriction enzymes MspI and HpaII, respectively. e, 10x molar excess of non-biotinylated umDNA was mixed with biotinylated mDNA. The mixture was denatured (100°C; 5 minutes); quickly chilled to 70°C and re-annealed by slow chilling down to 4°C. f, The biotinylated DNAs, a ~10:1 mixture of hmDNAs and mDNAs, were captured with streptavidin magnetic beads. g, The beads were removed following BstXI restriction digestion. h, COBRA analysis was performed to assess the purity of the hmDNA. The primers for bisulfite converted DNA were designed against the upper ( unmethylated) strand, hence BstUI digestion of hmDNA and umDNA should present identical restriction patterns. Presence of trace amount of BstUI digestion fragments in hmDNA lane (marked with asterisk) reflects ~<10% of the mDNA in the mixture (panels e and f). i and j, Lollipop representation of bisulfite sequenced clones. k, Human *CEBPA* construct used to generate transgenic cell lines. l, Schema of COBRA design. m, COBRA analysis of the transgenic sequences. Similar results were obtained for all tested individual clones (data not shown). n, Diagram showing outline of the experiment in living cells. o, Levels of T7 polymerase and T7 polymerase-induced transcripts. (qRT-PCR, bars indicate mean ± s.d.). p, DNA methylation changes after induction of the transcription assessed by COBRA.
Supplementary Figure 5 (part 1)

**Supplementary Information**

**Panel a**
- Relative ecRNA levels (normalized to 18S).

**Panel b**
- Distribution of gene loci.
  - 17208 gene loci without DNMT1-RIP peaks.
  - 6042 gene loci with DNMT1-RIP peaks.
  - 4833 gene loci covered by RRBS.
  - 10973 gene loci covered by RRBS.

**Panel c**
- Analysis flowchart.
  - 40857 Refseq transcripts (hg19 – from UCSC).
  - 23250 unique genes (longest Refseq isoform).

**Panel d**
- Gene loci with DNMT1-RIP peaks.
  - 17208 gene loci without DNMT1-RIP peaks.
  - 6042 gene loci with DNMT1-RIP peaks.

**Panel e**
- Expression level and Methylation level.
  - DNMT1 unbound and bound.

---

**Analysis flowchart**

1. **40857** Refseq transcripts (hg19 – from UCSC)
2. **23250** unique genes (longest Refseq isoform)
3. **17208** gene loci without DNMT1-RIP peaks
   - (Gene locus definition: 3kb upstream + gene body + 3kb downstream)
4. **6042** gene loci with DNMT1-RIP peaks
5. **10973** gene loci covered by RRBS
   - (Group: DNMT1 unbound)
6. **4833** gene loci covered by RRBS
   - (Group: DNMT1 bound)
Examples of gene loci from Cluster C
g Examples of gene loci from Cluster B

Supplementary Figure 5 (part 3)
Supplementary Figure 5. RNA-Immunoprecipitation sequencing (RIP-seq) analysis. a, Quality control test of RIP-seq libraries: ecCEBP4 was enriched in cDNA library made of RNAs that co-immunoprecipitated with anti-hDNMT1 antibody (qRT-PCR Bars are mean ± s.d.). b, Genomic localization of DNMT1 RIP-seq specific peaks throughout the RefSeq hg19 annotation. c, Pie chart showing distribution of DNMT1 RIP-seq specific peaks across the families of repetitive elements and small ncRNAs. d, Flow-chart of RIP-seq analysis. Schematic outline of the key steps applied for the comparative RIP-seq and RRBS analyses. e, Global comparison of gene expression and methylation levels between ‘DNMT1-unbound’ and ‘DNMT1-bound’ groups. The statistical significance of the difference between both groups is assessed by unpaired two-tailed student's t-test (t.test function from R using the default parameters), comparing the expression (left boxplot) and methylation (right boxplot) values of the 10973 DNMT1 unbound loci versus the 4833 DNMT1 bound loci. f and g, Additional examples of the gene loci from clusters C and B (main Figure 5b), respectively. Peaks are visualized using SSIRs. h, Validation of the identified RIP-seq peaks (the black double-headed arrows in f indicate location of qRT-PCR primer sets. Bars are mean ± s.d.).
Supplementary Figure 6. Biological Process Gene ontology (BP-GO) analysis of genes within the cluster C. a, Enrichment for the first 20 GO’s biological process annotations. We measured the GO term enrichment using a Benjamini-Hochberg corrected P-values. Here, for a convenient visualization, we plotted the -log2 (P-value). Other GO’s biological processes are presented in Supplementary Data 3. b, Bar Plot showing the percentage of DNMT1 RIP-Seq Peaks overlapped by at least one pyknon.