Supporting Information

Genome-wide DNA methylation signatures are determined by DNMT3A/B sequence preferences

Shi-Qing Mao\textsuperscript{1,4}, Sergio Martínez Cuesta\textsuperscript{1,2,4}, David Tannahill\textsuperscript{1}, Shankar Balasubramanian\textsuperscript{1,2,3}\textsuperscript{*}

\textsuperscript{1} Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Cambridge, UK
\textsuperscript{2} Department of Chemistry, University of Cambridge, Cambridge, UK
\textsuperscript{3} School of Clinical Medicine, University of Cambridge, Cambridge, UK
\textsuperscript{4} These authors contributed equally: Shi-Qing Mao, Sergio Martínez Cuesta

\textsuperscript{*} To whom correspondence should be addressed. (email: sb10031@cam.ac.uk)
**Figure S1.** *E.coli* genomic DNA provides good representation of all possible base combinations flanking cytosine.

Observed ...NCGN... (A) or ...NCAN... (B) or ...NCN... (C) sequences in the human genome (3 billion bp, dark gray), *E. coli* genome (4.6 million bp, red) and lambda DNA (48,502 base pairs, light gray) are shown as percentages with respect to all possible combinations.
**Figure S2**

**Figure S2.** Boxplots of methylation level in NCN context after incubation with DNMT3A (A) or DNMT3B (B) for 30, 120 and 240 min, or M.SssI (C) for 10, 30 and 240 min, ranked by median methylation level according to the samples treated for 240 min.
Figure S3

**Figure S3.** Boxplots of methylation level in NCGN context after incubation with DNMT3A (**A**) or DNMT3B (**B**) for 30, 120 and 240 min, or M.SssI (**C**) for 10, 30 and 240 min, ranked by median methylation level according to the samples treated for 240 min.
Figure S4

Figure S4. Boxplots of methylation levels in NNCGNN context after incubation with DNMT3A (A) or DNMT3B (B) for 30, 120 and 240 min, or M.SssI (C) for 10, 30 and 240 min, ranked by median methylation level according to the samples treated for 240 min. The top and bottom ten 6-mers in the 240 min samples are listed on the right. The flanking bases are colour-shaded to reveal the sequence preferences. N refers to the number of times that a 6-mer is found in the E.coli genome.
Figure S5

Figure S5. Sequence logos of the top 1000 methylated 10-mer CG-sequences after incubation with DNMT3A (A) and DNMT3B (B) for 30, 120 or 240 min. The far-right panels show the coverage of all mapped 10-mers with more than 10 reads. Sequence logo of the top 1000 methylated 10-mer CA-sequences after incubation with DNMT3A (C) and DNMT3B (D) for 120 and 240 min.
Figure S6.

**Figure S6.** Sequence logos of the top 1000 methylated 10-mer after 240 min incubation with DNMT3A or DNMT3B at (A, B) CT sequences, (C, D) CC sequences.
Figure S7. Sequence logos of the symmetrically heavy methylated (> 60% on both strands) 10-mer CG-sequences after incubation with DNMT3A (Left) and DNMT3B (Right) for 240 min.
**Figure S8**

**Figure S8.** Sequence logos of the top 1000 methylated 10-mer CG-sequences after incubation with DNMT3A/L (A) and DNMT3B/L (B) for 120 min.
Figure S9. (A) Expression of DNMT3A and DNMT3B in mouse J1 and human HUES64 cells. Gene expression in J1 were obtained from GSE100957, and expressions in HUES64 were from ENCODE experiments ENCSR738PHQ and ENCSR348EFG. (B) Sequence context of CA methylation of mouse and human WT and DNMT3-knockout stem cells. (C) Sequence context of CG methylation in mouse and human WT and DNMT3-knockout stem cells. (D) The ratio between CGC and CGG methylation in mouse and human WT and DNMT3-knockout stem cells.
Table S1

| Species | GEO id      | Sample description                                                      |
|---------|-------------|------------------------------------------------------------------------|
| Mouse   | GSM1382253  | Dnmt1,3a,3b-triple-KO mouse J1 ES cells expressing DNMT3A2             |
|         | GSM1382256  | Dnmt1,3a,3b-triple-KO mouse J1 ES cells expressing DNMT3B1             |
|         | GSM748786   | WT mouse J1 ES cells                                                   |
|         | GSM2339912  | WT mouse embryonic stem cells                                          |
|         | GSM2339913  | Dnmt3a,3b-double-KO mouse embryonic stem cells                        |
|         | GSM2339914  | Dnmt3a,3b-double-KO mouse embryonic stem cells expressing Dnmt3a full-length enzyme |
|         | GSM2420248  | Dnmt3a,3b-double-KO mouse embryonic stem cells expressing Dnmt3a catalytic domain alone |
| Human   | GSM1112840  | WT human HUES64 cells replicate 1                                      |
|         | GSM1112841  | WT human HUES64 cells replicate 2                                      |
|         | GSM1545002  | DNMT3A KO human HUES64 cells (early passage)                           |
|         | GSM1545005  | DNMT3A KO human HUES64 cells (late passage)                            |
|         | GSM1545003  | DNMT3B KO human HUES64 cells (early passage)                           |
|         | GSM1545006  | DNMT3B KO human HUES64 cells (late passage)                            |
|         | GSM1545004  | DNMT3A/3B double KO human HUES64 cells (early passage)                |
|         | GSM1545007  | DNMT3A/3B double KO human HUES64 cells (late passage)                  |