Figure S2:

Raw reads quality control and preparation

Adapter and barcode trimming
Low base quality trimming
Clean reads were C-to-T and G-to-A converted

Reference preparation

Combined genome, ERCC
and R-luc sequences
-> C-to-T and G-to-A conversion of
forward and reverse strand

Single annotation of tRNA
Single annotation of rRNA
-> C-to-T conversion of forward strand

Mapping strategy

All reads mapped to Genome,
ERCC and R-luc spike-in

All reads mapped to tRNA
and rRNA sequences
separately

obtain Bam files

-> Uniquely mapping converted reads were replaced by
original unconverted read

bsRNA-seq fractions

Replicates	F1	F2	F3	F4	
B	LibB1	LibB2	LibB3	LibB4	cLibB
C	LibC1	LibC2	LibC3	LibC4	cLibC
E	LibE1	LibE2	LibE3	LibE4	cLibE

Filter application in each library

Combine bsRNA-seq fraction libraries into composite library
3C; Retain converted reads with ≤3 C per read
Site calling at each cytosine position
'S/N90'; Signal to noise ratio filter 0.9

30RC; Minimum coverage of ≥30 reads per site
5C; Minimum frequency of ≥5 non-converted C at each site
80CT; Remove variable site requiring ≥80% C+T at each site

High confidence candidate site selection

10MM; Minimum average non-conversion of ≥10% across replicates

Fraction non-conversion analysis
Deconvolute high confidence sites into fractions

Fraction clustering analysis
Require coverage in 9 out of 12 samples
≥10 read average coverage across respective fraction

F1234 clustering

F234 clustering

-> Mfuzz (*m*=2) clustering aquiring *c*=9 clusters

Differentially methylated sites

Fraction A
non-conversion

Fraction B
non-conversion

etc

≥10 read average coverage across fractions

Logistic regression test
relative non-conversion change ≥10%
q-value ≤0.05

Differentially non-converted sites