Supplemental Material

Legrand et al.

Contents:

Supplemental Methods

Supplemental Figures S1-S7

Supplemental Tables S1-S7

Supplemental References

Supplemental Methods

Cell culture. All mouse ES cell lines came from a J1 background. *Dnmt2*$^{-/-}$ cells were provided by En Li (Okano et al. 1998). *Tet1*$^{-/-}$,*Tet2*$^{-/-}$,*Tet3*$^{-/-}$ cells were provided by Rudolf Jaenisch (Dawlaty et al. 2014). ESCs were grown on a primary MEF feeder layer in standard culture medium: knock-out DMEM (Gibco), 15% heat inactivated fetal calf serum (Hyclone), 1% non-essential amino acids (Gibco), 1% penicillin/streptomycin (Gibco), 2nM L-glutamine (Gibco), 0.1mM β-mercaptoethanol, 500 U/mL ESGRO (ESG1106, Merck Millipore). The feeder cells were removed by serial passaging as follows: MEFs and ESCs were harvested and the feeder was allowed to attach to the dishes for 40- 60 min at 37 °C. All cells that had not attached were re-plated into second dishes for an additional 40-60 min. The supernatants (non-attached cells) were then spun down and pellets were snap-frozen in liquid nitrogen. This allowed harvest of pure ESCs without any MEF contaminants, as MEFs attach significantly faster. The purity of the harvest was very high, which was ultimately confirmed by genotyping of *Dnmt2*$^{-/-}$ and *Tet1*$^{-/-}$,*Tet2*$^{-/-}$,*Tet3*$^{-/-}$ ESCs using PCR.

RNA isolation and quality control. RNA extraction was performed according to the TRIzol RNA isolation protocol (Ambion). The nucleic acid concentrations and purity were analyzed on the NanoDrop ND-1000 Spectrophotometer. RNA integrity was measured on a Bioanalyzer (Agilent) or TapeStation (Agilent). Samples were stored at -80 °C.

Library preparation. 30 µg of total RNA was fractionated into a long (<200 nt) and a small RNA fraction (<200 nt) by a modified protocol for small RNA cloning (W. Gu and D. Conte) using *mir*Vana Isolation Kit buffers (Ambion). Briefly, 400 µl of Lysis/Binding Buffer was mixed with 48µl of Homogenate Additive and 80 µl of total RNA. The mix was incubated at RT for exactly 5 min. 1/3rd volume of Ethanol was added and mixed by inversion. This was incubated at RT for

20 min. 10 µg GlycoBlue was added and spun at 21 °C and 2,500 x g for 8 min to pellet the long fraction. The supernatant containing the small fraction was carefully transferred to a fresh tube and precipitated with isopropanol and 15 µg GlycoBlue at -80 °C for at least 10 min. rRNA depletion was carried out on the small fraction and on the half amount of the long fractions according to the RiboMinus Eukaryote System v2 (Ambion) protocol. The other half amount of the long fraction was further processed and sequenced as Ribo+ sample. rRNA-depleted samples were concentrated using ethanol precipitation. Alternatively poly(A)$^+$ RNA was isolated from 500 µg total RNA by two rounds of purification using the polyA Spin mRNA Isolation kit (New England Biolabs). The long fractions and poly(A)$^+$ RNA were further processed with the NEBNext Magnesium RNA Fragmentation Module (New England Biolabs). This was carried out as described in the manual. 3 min of fragmentation at 94 °C has been established to lead to a peak at approximately 250 nt, appropriate for the final 100 bp paired-end sequencing. The fragmented RNA was precipitated using ethanol with 20 µg GlycoBlue at -80 °C for at least 10 min.

TURBO DNase (Ambion) was applied to each sample according to the protocol in a total volume of 20 µl. This volume was immediately applied to the EZ RNA Methylation Kit (Zymo Research), following the manufacturer's manual. As a final step before library preparation, a stepwise RNA end repair was carried out using T4 polynucleotide kinase (TaKaRa). 3'-Dephosphorylation and 5'-phosphorylation (using 1 mM dATP) were carried out at 37°C for 20 min each. RNA was further purified by a phenol/chloroform extraction and ethanol precipitation.

Library preparation was done according to the NEBNext Small RNA Library Prep Set protocol. Following adaptor ligation and cDNA synthesis, cDNA was amplified with 12 cycles of PCR and purified using the QIAquick PCR Purification Kit (Qiagen). The libraries were further size selected on a 6% polyacrylamide gel. Compatible barcodes were selected and samples were pooled in equimolar ratios on multiple lanes on an Illumina HiSeq 2000 platform. A 100 bp paired-end sequencing approach was used.

RNA-seq. Total RNA was DNase treated and rRNA depleted as described above and then submitted to the Genomics and Proteomics Core Facility at the DKFZ for quality control and library preparation using the TruSeq RNA Sample Preparation Kit (Illumina). RNA was sequenced in a 100 bp paired-end approach on an Illumina HiSeq 2000 platform.

Filtering of artifacts. (1.) Reads are denoted r_i , where i is in {1,..,N}, N is the number of reads aligning to a specific RNA. Reads r_i containing ≥3 adjacent non-converted cytosines (rRNA and mRNA), or ≥6 adjacent non-converted cytosines (tRNA) were interpreted as non-converted reads resulting from RNA folding or bound proteins and were thus discarded. The presence of NSUN2- and DNMT2-dependent methylation at C38, C48, C49 and C50 in some tRNAs made it necessary to raise this threshold from 3 to 6 in tRNAs. (2.) A non-converted C at position $j1$ in read r_i and mismatches apart from C/T at position $j2$ in read r_i, if they colocalize consistently, indicate the existence of a distinct subset of reads (as caused by RNA editing, single-nucleotide polymorphisms or residual misalignment). We detected such cases with a Fisher's exact test on the following contingency table:

	mismatch	**match**
C	$\sum_i \mathbf{1}_{(r_{i,j1}=C)} \cdot \mathbf{1}_{(r_{i,j2}=mismatch)}$	$\sum_i \mathbf{1}_{(r_{i,j1}=C)} \cdot \mathbf{1}_{(r_{i,j2}=match)}$
T	$\sum_i \mathbf{1}_{(r_{i,j1}=T)} \cdot \mathbf{1}_{(r_{i,j2}=mismatch)}$	$\sum_i \mathbf{1}_{(r_{i,j1}=T)} \cdot \mathbf{1}_{(r_{i,j2}=match)}$

1: indicator function

corrected for multiple testing using Bonferroni's method, and discarded the mismatched subsets of reads upon significance at level $\alpha = 0.05$. We permitted an exception for tRNA(Val), which is known to be subject to RNA editing at position *34*. Cytosines covered by ≥3 filtered reads were kept for further analysis. (3.) Single nucleotide polymorphism (SNP) calling was performed on

RNA-seq data of wild type and *Dnmt2*-deficient samples, to account for possible differences in genetic backgrounds. If a SNP was present at a candidate methylation site, the methylation call was further inspected manually.

Type I error and power simulations. Coverage was sampled from the wild-type empirical distribution. For type I error estimation, non-conversion ratios of simulated Cs were drawn randomly from the aforementioned $\mathcal{P}(\lambda)$ distribution for 1,000,000 sites. This was carried out 6 times with different seeds for random number generation. These sites are denoted as H0 sites since they correspond to the null hypothesis of stochastic non-conversion. The coverage-stratified type I error for each replicate was calculated as the number of significant raw p-values for these sites divided by the number of p-values in each coverage bin. We then calculated the mean and standard error of the type I error from the six different seeds. For statistical power estimation, m=100 true positive C sites were generated, characterized each by a fixed non-conversion ratio in {0.1,0.2,...,1}, and a coverage in {3,5,10,20,30,50,100,250,500,1000}. This led to a total of 10,000 C sites, denoted H1. In order to mirror experimental data in a realistic way, λ was taken equal to 0.0157 (median $\mathcal{P}$ parameter) and affected by a normal-distributed variation term $\mathcal{N}(0,0.001)$ in each replicate. Similarly, the non-conversion ratio was affected by both a normal-distributed biological variation term $\varepsilon_1 \sim \mathcal{N}(0,0.02)$ and an additional $\mathcal{P}$-distributed non-conversion term ε_2 :

$$non\ conversion\ ratio\ =\ x + \varepsilon_1 +\ \varepsilon_2/coverage.$$

The 10,000 H1 sites were complemented with H0 sites to obtain a total of 3,000,000 simulated Cs. The methylation calling procedure was then carried out as described above. In addition, adjustment for multiplicity was performed using the non-conversion ratio-reweighted method as provided in the IHW R package (Ignatiadis et al. 2016). This was repeated 100 times so as to

obtain 10,000 H1 sites per coverage and non-conversion ratio pair. Coverage- and non-conversion ratio- stratified statistical power was obtained by dividing the number of significant H1 sites in each category by 10,000.

454 bisulfite sequencing. Total RNA was extracted and DNase treated as described above. The DNase reaction was immediately applied to the EZ RNA Methylation Kit (Zymo Research), with a maximum of 3 µg input. 200 ng of bisulfite converted RNA was used for the cDNA synthesis, which was carried out with the SuperScript III Reverse Transcriptase (Invitrogen) and 2 pmol of gene-specific primer. The conditions for the reverse transcription were: 10 min RT, 15 min 50 °C, 15 min 70 °C. Amplicons were separated on and isolated from 2% agarose gels and DNA was eluted using the peqGOLD Gel Extraction Kit (PEQLAB). For 454 sequencing, concentrations were determined with the Quant-iT PicoGreen reagent and according to the Roche 454 Junior protocol. The FLUOstar OPTIMA (BMG Labtech) plate reader was used for measurements. The amplicons were then included in the library preparation and sequenced in an equimolar sample pool on a Roche 454 Junior platform. The acquired sequencing reads were analyzed to produce graphical representation in a heat map. PCR primers are provided in Tab. S7.

Logo plot. Sequence frequencies of unconverted cytosines present in at least 1 wild type sample, with non-conversion ratio larger than 0.1 and a p-value<0.01 using the *seqLogo* package (R bioconductor).

Methylation frequency and reproducibility. Frequency was calculated as the proportion of cytosines with coverage higher than 10 in three wild type ES cells replicates and bisulfite non-conversion ratio higher than 0.1. The measure for reproducibility was the standard error.

RNA structure analysis. RNA secondary structure analysis was performed using mfold (Zucker 2003).

LC-MS/MS analysis. Total RNA was prepared from wildtype mouse ES cells and from *Drosophila* S2 cells cultured in Express Five Serum Free Medium (Gibco). 1 mg of DNase-treated total RNA was sequentially subjected to two rounds of mRNA enrichment using the polyA Spin mRNA Isolation kit (New England Biolabs). Small RNA fractions (<200 nt) were depleted using *mir*Vana Isolation Kit buffers (Ambion) as described above. Additional two rounds of purification were applied using the RiboMinus Eukaryote System v2 (Ambion) protocol. At each step a minimum of 300 ng RNA was collected for LC-MS analysis. An additional 50 ng from each step were fractionated, multiplexed and sequenced on a single line of an Illumina HiSeq 2000 platform, as described above. Resulting reads were quality-controlled and subsequently aligned using Bowtie. For mouse samples, the mRNA reference sequences were the collection of mRNA sequences from NCBI RefSeq (NM-labelled accessions). The tRNA reference sequences were from the genomic tRNA database. The rRNA reference sequences were assembled from Arb-Silva, NCBI (BK000964.3), Ensembl (5S sequences) and NCBI RefSeq (NR-labelled accessions). The mitochondrial RNA sequences were from Ensembl. For *Drosophila* samples, the mRNA and mtRNA reference sequences were the collection of mRNA and mitochondrial RNA sequences from Flybase. tRNA reference sequences were from the genomic tRNA database and the rRNA sequences from Arb-Silva and from NCBI (M21017,M29800). Each best alignment per read was kept and used to calculate the proportion of tRNA, rRNA and mRNA in each sample.

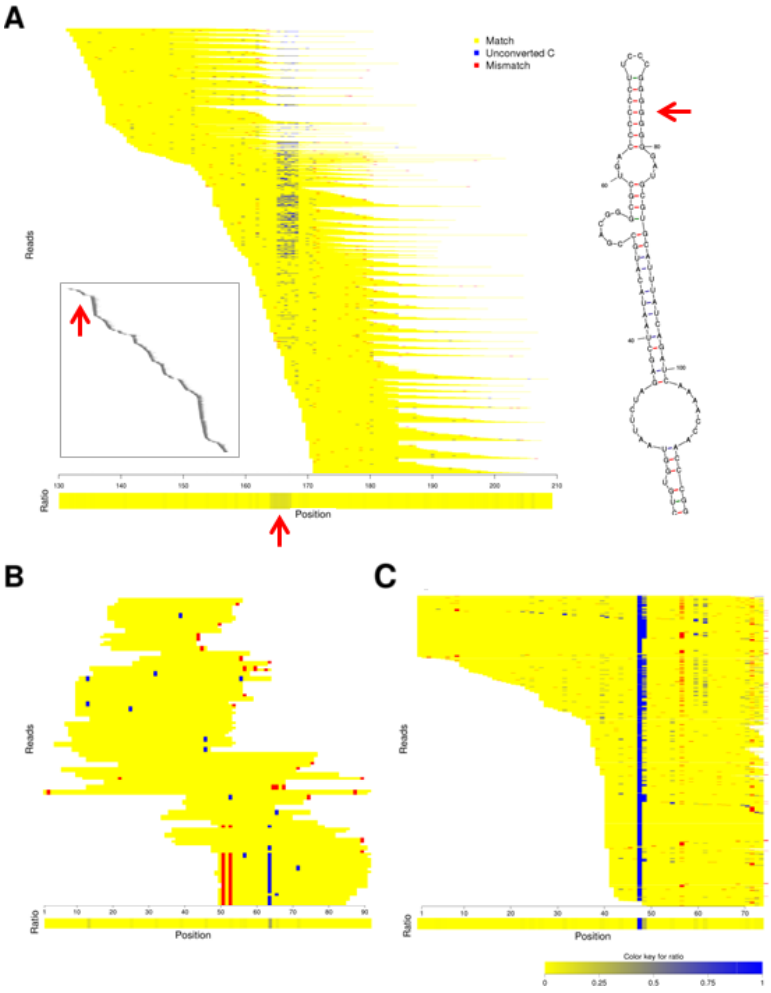
Prior to LC-MS/MS analysis, the RNA samples were digested into nucleosides using the following protocol: 300 ng of each RNA fraction was incubated at 37°C for 2 h in the presence of 0.3 U nuclease P1 (Roche Diagnostics, Mannheim, Germany), 1/10 vol. of 10x nuclease P1

buffer (50 mM NH_4OAc pH 5.3) and 0.1 U snake venom phosphodiesterase (Worthington, Lakewood, USA). Afterwards 1/10 vol. of 10x fast alkaline phosphatase buffer (50 mM NH_4OAc pH 9) and 1 U fast alkaline phosphatase (Fermentas, St. Leon-Roth, Germany) were added, followed by an incubation at 37°C for 1 h. To the obtained nucleoside mix a ^{13}C stable isotope labeled nucleoside mix from *S. cerevisiae* RNA as an internal standard (SIL-IS) was then added to a final concentration of 5 ng/ µL for the sample RNA as well as 5 ng/ µl for the SIL-IS. The injection volume for analysis was 10 µl, triplicates of the samples were measured.

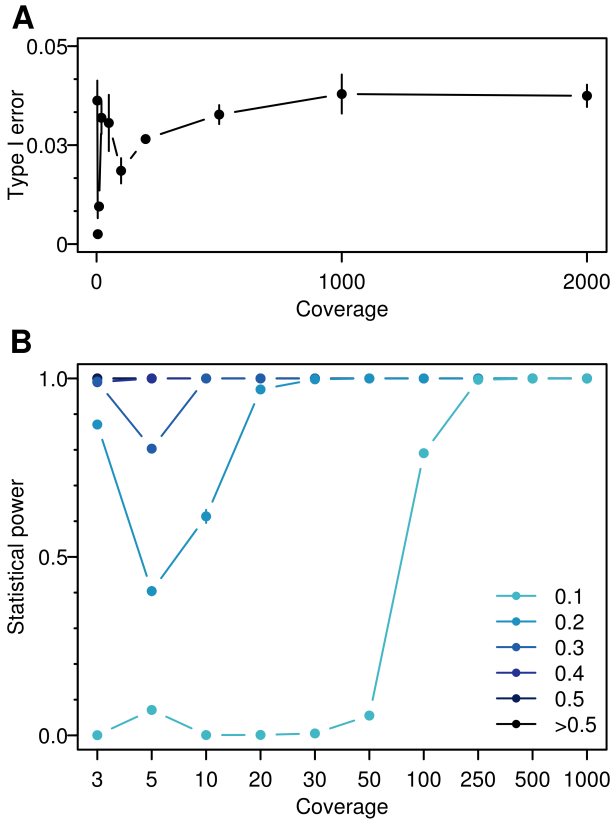
The nucleoside samples were analyzed on an Agilent 1260 series equipped with a diode array detector (DAD) and a Triple Quadrupole mass spectrometer (Agilent 6460). A Synergy Fusion RP column (4 µm particle size, 80 Å pore size, 250 mm length, and 2 mm inner diameter) from Phenomenex (Aschaffenburg, Germany) was used at 35°C. The solvents consisted of 5 mM ammonium acetate buffer adjusted to pH 5.3 using acetic acid (solvent A) and pure acetonitrile (solvent B). The elution (flow rate 0.35 mL/min) started with 100 % solvent A followed by a linear gradient to 8 % solvent B at 10 min. Solvent B was increased further to 40 % over 10 minutes. Over the next 3 minutes Solvent B was decreased back to 0 %. Initial conditions were regenerated by rinsing with 100 % solvent A for additional 7 minutes. For detection of the major nucleosides, the effluent from the column was first measured photometrically at 254 nm by the DAD, and modified nucleosides were then detected by the mass spectrometer equipped with an electrospray ion source (Agilent Jet Stream). ESI parameters were as follows: gas temperature 350°C, gas flow 8 L/min, nebulizer pressure 50 psi, sheath gas temperature 350°C, sheath gas flow 12 L/min, capillary voltage 3500 V. The MS was operated in the positive ion mode using Agilent Mass Hunter software in the DMRM (dynamic multiple reaction monitoring) mode. The monitored mass transitions for hm5C were 274.2 (precursor ion), 142.1 (product ion). The instrument settings were 80 V (fragmentor) anbd 10 eV (collision energy) and the retention time window was 5.2 min. The peak areas were determined by employing Agilent MassHunter Qualitative Analysis Software. In the case of the major nucleosides, peak areas were extracted

from the recorded UV chromatograms in order to avoid saturation of the mass signals. The peak areas of the modified nucleosides were measured by LC-MS/MS.

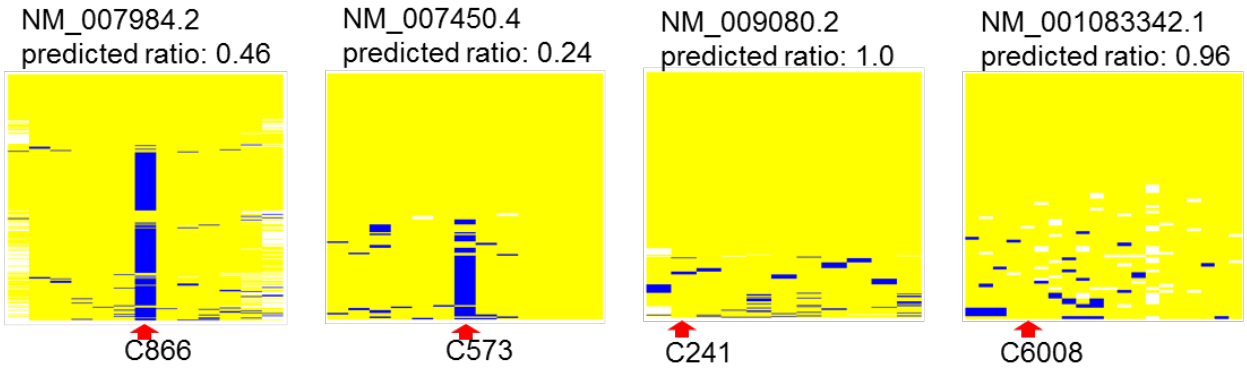
Quantification was performed as described previously (Kellner et al. 2014). In brief, external calibration was used to determine the amount of adenosine by doing a linear calibration of the area under the curve of the UV signal in a range from 2-1000 pmol. The amount of the modified nucleosides was calculated by linear calibration of the SIL-IS in relation to the respective nucleoside. To validate the calibrations, short synthetic RNA oligonucleotides with defined sequences and concentrations, containing the respective modifications, were digested and subsequently measured with the same parameters as the calibrations and the samples. Additionally, an external calibration for hm^5C was conducted in a range of 0.2-1000 fmol to determine the Limit of Detection (LOD). The R^2-values of all calibrations were at least 0.997. From known amounts of adenosine as well as those of the modified nucleosides the respective ratios ratios m5C/A [%], m26A/A [%], m6A/A [%] and m1A/A [%] were calculated.



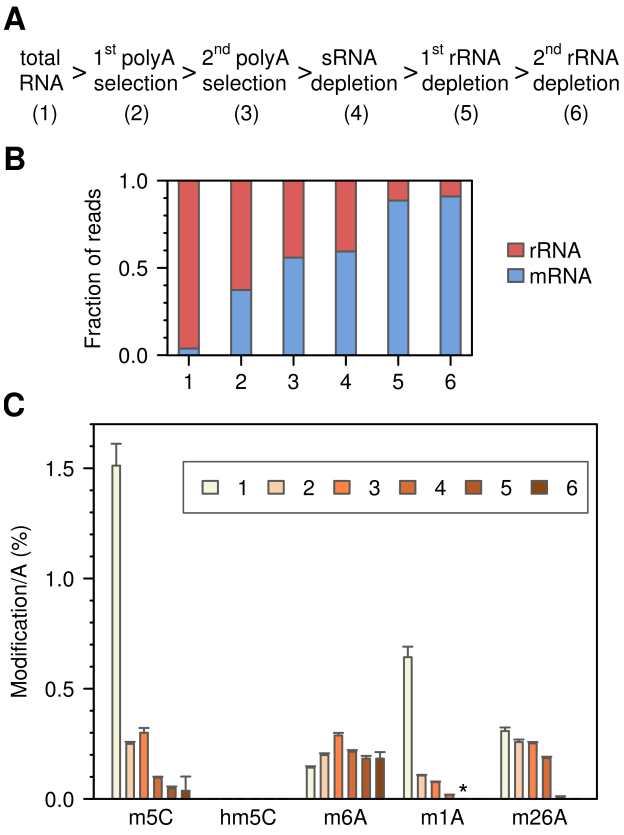
Supplemental Figure S1. Filtering of artifacts. (A) Detection of unconverted reads in 18S rRNA corresponding to stable secondary structure. (B and C) Colocalization of a mismatch with a non-converted C in tRNA-Tyr (B) or with a converted C in tRNA-Gln (C).



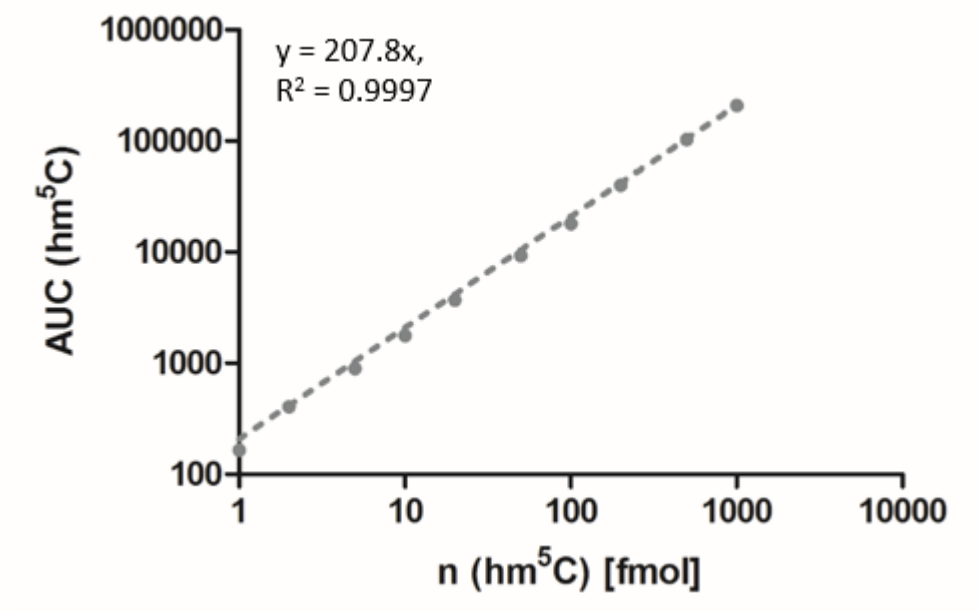
Supplemental Figure S2. Statistical characteristics of the WTBS pipeline from simulations. (A) Type I error estimates. (B) Statistical power estimates using IHW adjustment for multiple testing.



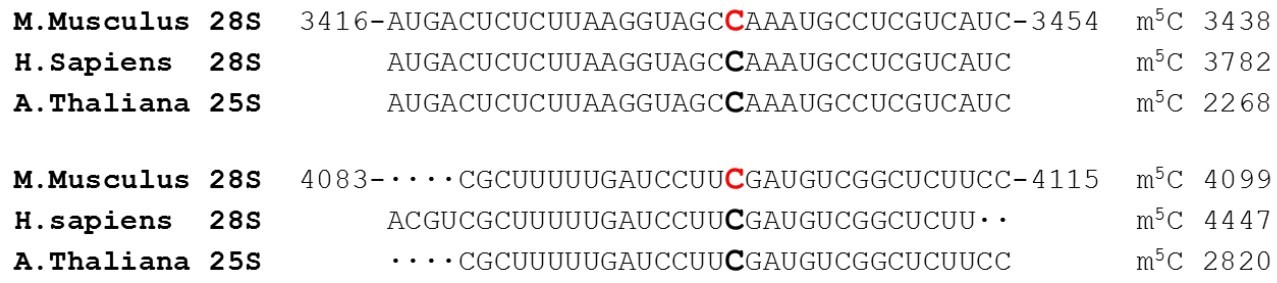
Supplemental Figure S3. Representative 454 bisulfite sequencing of candidate mRNA methylation sites. Each column indicates a cytosine residue, each row represents a sequencing read. Converted cytosines are shown in yellow, unconverted cytosines are shown in blue. Complete results for 454 bisulfite sequencing are shown in Supplemental Table S4.



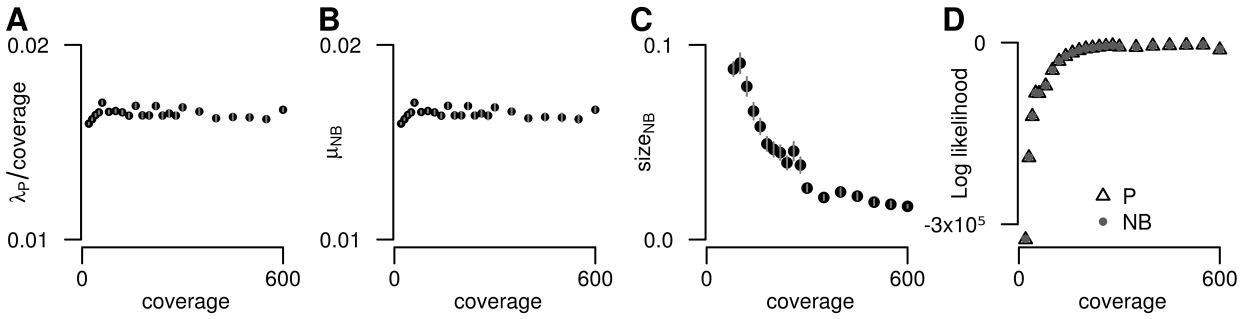
Supplemental Figure S4. LC-MS/MS analysis of RNA samples from *Drosophila* S2 cells that were subjected to multi-step mRNA enrichment. (A) Basic outline of the mRNA enrichment protocol. (B) Relative amounts of mRNA and rRNA, as determined by RNA-seq. The proportion of tRNA reads was <1 % for all samples. (C) Modification analysis of m5C, hm5C, m26A, m6A and m1A content, relative to A content. Asterisk: not determined.



Supplemental Figure S5. Detection limit for hm5C. External calibration of ion count response to a range of 1-1000 fmol hm^5C (x-axis). The y-axis shows the areas under the curve (AUC) of the mass peak counts for the hm^5C-specific mass-transition from 274.2 m/z to 142.1 m/z. The limit of detection is 1 fmol with a signal-to-noise ratio of 3.7.



Supplemental Figure S6. Evolutionary conservation of rRNA methylation. Mouse ES cell 28S rRNA is methylated at C3438 and C4099, similarly to human 28S and A. thaliana 25S methylation.



Supplemental Figure S7. Poisson and Negative Binomial fits to non-conversion distribution in one wild type long fraction replicate, with non-conversion ratio <0.3. Parameters are given as a function of coverage. (*A*). Poisson parameter divided by coverage. (*B*) and (*C*). Negative binomial parameters. (*D*). Log-likelihood. Error bars in A, B and C represent the standard deviation.

Supplemental Table S1. Known sources of bisulfite sequencing artifacts.

Source	Artifact	References
RNA	Protection due to RNA structure	(Warnecke et al. 2002; Edelheit et al. 2013; Hussain et al. 2013)
	Protection by other modifications than m5C (m3C, Cm, m5Cm, ac4Cm,m4Cm, f5Cm, m_2^4Cm,ac4C, hm5C)	(Schaefer et al. 2009; Huang et al. 2010; Behm-Ansmant et al. 2011; Schaefer 2015)
	Protection by proteins	(Warnecke et al. 2002)
Bisulfite treatment	Randomly distributed incomplete deamination	(Warnecke et al. 2002; Schaefer et al. 2009)
Sequencing	Sequencing errors	
	PCR bias	(Warnecke et al. 1997; Schaefer et al. 2009; Edelheit et al. 2013)
Alignment	Identical sequences	(Squires et al. 2012)
	tRNA-like sequences inside mRNA sequence	(Hussain et al. 2013)
	Sequence variation, multiple splice-forms	(Barturen et al. 2013; Rieder et al. 2016)
	Mapping ambiguity due to repetitive RNA sequences and removal of C bases	(Laird 2010)

Supplemental Table S2. Libraries and sequencing.

Sample	Description	Number of sequenced reads	Remaining reads after filtering	Conversion efficiency
Wt1S	Wild type, short fraction, sample 1	26,240,036	1,685,313	n.d.
Wt2S	Wild type, short fraction, sample 2	26,637,671	1,980,537	n.d.
Wt3S	Wild type, short fraction, sample 3	23,465,734	1,664,161	n.d.
Wt1L	Wild type, long fraction Ribo-, sample 1	52,782,688	7,241,309	98.3%
Wt2L	Wild type, long fraction Ribo-, sample 2	84,435,380	13,562,752	98.5%
Wt3L	Wild type, long fraction Ribo-, sample 3	20,509,939	3,612,678	98.4%
Wt1R	Wild type, long fraction Ribo+, sample 1	51,212,028	3,639,069	98.4%
Wt2R	Wild type, long fraction Ribo+, sample 2	48,122,614	3,547,129	98.6%
D1S	Dnmt2$^{-/-}$, short fraction, sample 1	19,989,042	1,916,744	n.d.
D2S	Dnmt2$^{-/-}$, short fraction, sample 2	19,198,431	1,669,271	n.d.
D3S	Dnmt2$^{-/-}$, short fraction, sample 3	41,006,670	3,497,429	n.d.
D1L	Dnmt2$^{-/-}$, long fraction Ribo-, sample 1	67,977,380	9,052,515	98.2%
D2L	Dnmt2$^{-/-}$, long fraction Ribo-, sample 2	57,960,844	8,588,002	98.2%
D3L	Dnmt2$^{-/-}$, long fraction Ribo-, sample 3	47,836,895	9,839,063	98.5%
D1R	Dnmt2$^{-/-}$, long fraction Ribo+, sample 1	47,668,762	3,237,946	98.6%
D2R	Dnmt2$^{-/-}$, long fraction Ribo+, sample 2	50,572,980	3,540,972	98.6%
T1	Tet TKO, mRNA, sample 1	41,108,655	12,569,101	98.5%
T2	Tet TKO, mRNA, sample 2	39,857,668	12,551,897	98.5%
T3	Tet TKO, mRNA, sample 3	35,574,019	11,833,390	98.6%

Conversion efficiency in rRNA is calculated on all cytosines except known methylated cytosines (2 in 28S rRNA and 2 in mitochondrial 12S rRNA). (n.d.) Not determined.

Supplemental Table S3. Parameter of the Poisson distribution divided by coverage.

	Wild type median (95% CI)	Dnmt2$^{-/-}$ median (95% CI)	Tet$^{-/-}$ median (95% CI)
replicate 1	0.0164 (0.0164 ; 0.0166)	0.0185 (0.0184 ; 0.0186)	0.0152 (0.0151 ; 0.0153)
replicate 2	0.0149 (0.0147 ; 0.0149)	0.0184 (0.0183 ; 0.0184)	0.0145 (0.0143 ; 0.0145)
replicate 3	0.0157 (0.0156 , 0.0159)	0.0147 (0.0146 ; 0.0148)	0.0140 (0.0139 ; 0.0140)
Median (3 replicates)	0.016 (0.0149 ; 0.0164)	0.018 (0.0147 ; 0.0185)	0.014 (0.0140 ; 0.0152)

Supplemental Table S4. Re-analysis of candidate mRNA methylation sites by 454 bisulfite sequencing.

mRNA	(cytosine-5) position	WTBS			454 BS
		median	standard error	p-value	
NM_008143.3	*186*	0.17	0.01	$< 2.7 \times 10^{-274}$	0.19
NM_007918.3	*567*	0.37	0.03	2.70×10^{-274}	0.35
NM_007984.2	*866*	0.46	0.03	1.49×10^{-169}	0.60
NM_007531.2	*326*	0.25	0.01	2.44×10^{-168}	0.27
NM_007475.5	*416*	0.07	<0.01	7.74×10^{-104}	0.09
NM_007475.5	*368*	0.09	0.01	1.57×10^{-102}	0.10
NM_008231.4	*1104*	0.41	0.02	1.99×10^{-90}	0.45
NM_001199350.1	*15*	1.00	<0.01	7.82×10^{-87}	0.02 (n.m.)
NM_007450.4	*573*	0.24	0.03	6.42×10^{-53}	0.38
NM_008210.4	*79*	0.15	0.03	8.16×10^{-52}	0.09
**NM_008695.2*	*521*	0.83	0.08	3.85×10^{-68}	0.85
**NM_008503.5*	*703*	0.97	0.03	7.62×10^{-50}	0.03 (n.m.)
**NM_009080.2*	*241*	1.00	<0.01	7.12×10^{-53}	0.01 (n.m.)
**NM_001083342.1*	*6008*	0.96	0.04	5.87×10^{-16}	0.00 (n.m.)

The table shows the 10 candidate sites with the smallest p-values, present in the three wild type biological replicates, and 4 candidate sites present in two wild type replicates with both methylation ratio close to 1 and significant p-value (marked with *). The results obtained with the independent bisulfite sequencing approach (454 BS) are shown in the rightmost column. (n.m.) Not methylated.

Supplemental Table S5. WTBS analysis of tRNA methylation patterns. tRNAs were selected (significant p-value and non-conversion ratio >0.7) among known (cytosine-5) tRNA methylation marks for M. musculus in the Modomics database (Machnicka et al. 2013).

tRNA	(cytosine-5) position	Median m5C in the three replicates	Standard Error
Arg ACG	49	0.76	0.024
Gln CUG	49	0.97	0.001
Gln UUG	49	0.95	0.016
Ile AAU	48	1.00	0.022
Lys CUU	48	0.98	0.006
Met CAU	48	0.93	0.004
Phe GAA	49	0.95	0.008
Val AAC	38	0.94	0.003
	48	0.93	0.001
	49	0.92	0.002

Supplemental Table S6. Overview of WTBS methylation calling in other publications.

	Alignment	Filters	m5C discovery	Replicates
Squires et al., 2012	SOCS-Bprog (for SOLiD-generated sequences and including custom reference).	≥10 reads; 0 to 4 mismatches; Positive Control tRNA-Asp; Negative Control Renilla Luciferase mRNA in vitro-transcribed.	Non-conversion ratio ≥ 20%.	n.d.
Amort et al., 2013	n.d.	n.d.	Non-conversion ratio ≥ 15%.	3 (n.d.)
Edelheit et al., 2013	Novoalign; Genome from NCBI, unique copy of each tRNA,unique copy of 5S, 16S and 23S rRNA.	≥5 reads; Unconverted reads (>3C) removed; Alignment Quality>20; Unaligned reads trimmed and realigned; Multiply aligned reads randomly assigned to synonym regions; Only one copy of duplicate reads (mRNA).	Comparison to full conversion (Fisher test); Joint methylation levels in case of gene variants.	n.d.
Hussain et al., 2013	Bismark; Bisulfite-converted transcripts (human: Ensembl v68, mouse: Ensembl v67).	≥ 10 reads; Unconverted reads (>30%C) removed.	Non-conversion ratio ≥ 20%.	4 or more
Burgess et al., 2015	n.d.	≥ 5 reads.	Non-conversion ratio ≥ 20%.	n.d.
Rieder et al., 2016	meRanT: Bowtie2 with custom NCBI RefSeq reference; or: meRanG: STAR or Tophat2 with bisulfite-converted NCBI RefSeq.	Custom read coverage, base quality and non-conversion rate; Limit on number of identical reads; Exclusion of C-biased read ends.	Fisher test (methylation test); Binomial test (methylation rate) depending on bisulfite conversion rate.	n.d.

(n.d.) Not disclosed.

Supplemental Table S7. PCR primers for 454 bisulfite sequencing.

NM_008143.3F	ATGGGTAATATAGATTGTTATTATATTG
NM_008143.3R	CTTCCACATAATAATAATCTTATCTCAAAACAC
NM_007918.3F	GATATTGGGTGTGGGTTGGTTATTTAG
NM_007918.3R	CACAAAATTCAAAACAAAACTAACACC
NM_007984.2F	GTTATTAAGGTGGGTAAAGATGAG
NM_007984.2R	CACATTCCTCTCATTAACCACC
NM_007531.2F	GTTTATTGTGGAAGGTGGTTATAGAG
NM_007531.2R	CCTTCAACCAAAATCATATCC
NM_007475.5F*	GGAGAATAATTTAGTTTTGGAG
NM_007475.5R*	CCTTATTAACCAACAACATATCC
NM_008231.4F	TGTTTGGGTGTTATTGGGGAAATTGG
NM_008231.4R	CCAAAACACTCTCCAAAAATAAACTT
NM_001199350.1F	GTTTATTTTTATATTAAGAAGGTTGGG
NM_001199350.1R	CCTCAAAATCTACCTAAACACACACC
NM_007450.4F	GGGATTTTTTTAGTGAGAATTTAATGGG
NM_007450.4R	CCCTAATAAAAACCCTTCAAACCATCCAAC
NM_008210.4F	AATTGTGTTTGTAGTTTTTGTTGTGTTG
NM_008210.4R	CTAAAACTCAACAAACAAAAAACAACAC
NM_008695.2F	GATATTAGTTATAGTAGGGGTTGGA
NM_008695.2R	CTCCTCATAAACACCCACATACTCCC
NM_008503.5F	GTTTTTAGAGGTATTGGTATTGTTTTTG
NM_008503.5R	TATAACAATCATCTATACCAACCATC
NM_009080.2F	TTGGAAGGATGATGAAGTTTAGG
NM_009080.2R	TAAATAACATACTTCTTCCTATACAC
NM_001083342.1F	GATAGGGTTTTTTTGTGTAGTTTTGG
NM_001083342.1R	CACAATAACACTTACCTTTAATCCCAACAC

*amplicon covers two sites.

Supplemental References

Barturen G, Rueda A, Oliver JL, Hackenberg M. 2013. MethylExtract: High-Quality methylation maps and SNV calling from whole genome bisulfite sequencing data. *F1000Res* **2**: 217.

Behm-Ansmant I, Helm M, Motorin Y. 2011. Use of specific chemical reagents for detection of modified nucleotides in RNA. *J Nucleic Acids* **2011**: 408053.

Dawlaty MM, Breiling A, Le T, Barrasa MI, Raddatz G, Gao Q, Powell BE, Cheng AW, Faull KF, Lyko F et al. 2014. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev Cell* **29**: 102-111.

Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R. 2013. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet* **9**: e1003602 doi: 10.1371/journal.pgen.1003602.

Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. 2010. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* **5**: e8888 doi: 10.1371/journal.pone.0008888.

Hussain S, Aleksic J, Blanco S, Dietmann S, Frye M. 2013. Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biol* **14**: 215.

Ignatiadis N, Klaus B, Zaugg JB, Huber W. 2016. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat Methods* **13**: 577-580.

Kellner S, Ochel A, Thüring K, Spenkuch F, Neumann J, Sharma S, Entian KD, Schneider D, Helm M. 2014. Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res* **42**: e142 doi: 10.1093/nar/gku733.

Laird PW. 2010. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* **11**: 191-203.

Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM et al. 2013. MODOMICS:

a database of RNA modification pathways--2013 update. *Nucleic Acids Res* **41**: D262-D267.

Okano M, Xie S, Li E. 1998. Dnmt2 is not required for *de novo* and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res* **26**: 2536-2540.

Rieder D, Amort T, Kugler E, Lusser A, Trajanoski Z. 2016. meRanTK: methylated RNA analysis ToolKit. *Bioinformatics* **32**: 782-785.

Schaefer M. 2015. RNA 5-Methylcytosine Analysis by Bisulfite Sequencing. *Methods enzymol* **560**: 297-329.

Schaefer M, Pollex T, Hanna K, Lyko F. 2009. RNA cytosine methylation analysis by bisulfite sequencing. *Nucleic Acids Res* **37**: e12 doi: 10.1093/nar/gkn954.

Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, Parker BJ, Suter CM, Preiss T. 2012. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res* **40**: 5023-5033.

Warnecke PM, Stirzaker C, Melki JR, Millar DS, Paul CL, Clark SJ. 1997. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res* **25**: 4422-4426.

Warnecke PM, Stirzaker C, Song J, Grunau C, Melki JR, Clark SJ. 2002. Identification and resolution of artifacts in bisulfite sequencing. *Methods* **27**: 101-107.

Zucker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**: 3406-3415.