**Stem Cell Reports, Volume *14***

## Supplemental Information

# TET1 and 5-Hydroxymethylation Preserve the Stem Cell State of Mouse Trophoblast

**Claire E. Senner, Stephanie Chrysanthou, Sarah Burge, Hai-Yan Lin, Miguel R. Branco, and Myriam Hemberger**

**TET1 and 5-hydroxymethylation preserve the stem cell state of mouse trophoblast**

**Claire E. Senner[1,2,*], Stephanie Chrysanthou[1,3], Sarah Burge[1], Hai-Yan Lin[4], Miguel R. Branco[5] and Myriam Hemberger[1,2,6,7*]**

**Supplemental Figures and Methods**

Figure S1: Related to Figure 1

Figure S2: Related to Figure 2

Figure S3: Related to Figure 2

Figure S4: Related to Figure 3

Figure S5: Related to Figures 4 and 5

Figure S6: Related to Figure 6

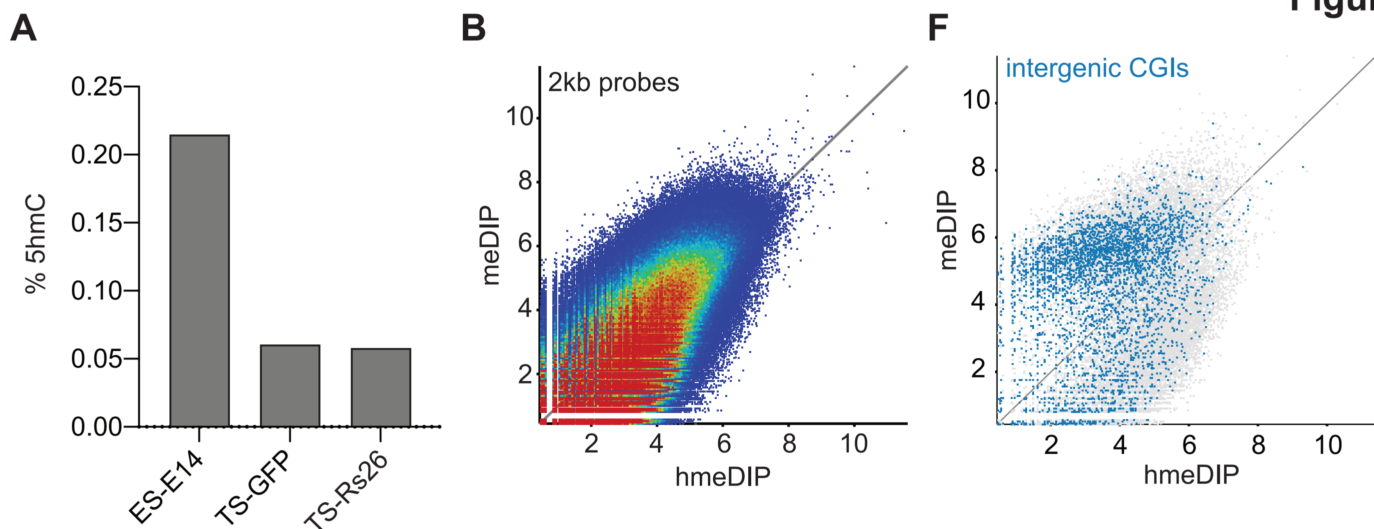
Table S1: Genes associated with CGIs that lose DNA methylation upon TS cell differentiation

Table S2: Intergenic TET1 peaks overlapping with TS cell promoter interaction sites (Hi-C)

Table S3: Detailed list of Data sets used in this study

Supplemental Methods giving details of immunofluorescence staining and RT-qPCR procedures.

# Figure S1




C

| meDIP-seq (data from Figure 1) | | | hmeDIP-seq (data from Figure 1) | | | hmeDIP-seq (data from Figure 2) | | |
|---|---|---|---|---|---|---|---|---|
| **total peaks** | **217656** | | **total peaks** | **222037** | | **total peaks** | **202687** | |
| | peaks | % of total | | peaks | % of total | | peaks | % of total |
| CGIs | 11581 | 5.32 | CGIs | 13178 | 5.94 | CGIs | 9103 | 4.49 |
| promoters | 8800 | 4.04 | promoters | 10944 | 4.93 | promoters | 7148 | 3.53 |
| enhancers | 8606 | 3.95 | enhancers | 8992 | 4.05 | enhancers | 7615 | 3.76 |
| genes | 129412 | 59.46 | genes | 102724 | 46.26 | genes | 91193 | 44.99 |
| exons | 56418 | 25.92 | exons | 30122 | 13.57 | exons | 21999 | 10.85 |
| introns | 124816 | 57.35 | introns | 98900 | 44.54 | introns | 87777 | 43.31 |
| intergenic | 94448 | 43.39 | intergenic | 124641 | 56.14 | intergenic | 113696 | 56.09 |

D

CGIs | Promoters | Enhancers | Genes | Exons | Introns | Intergenic

% of total

genome, meDIP (Fig.1), hmeDIP (Fig.1), hmeDIP (Fig.2)

E

Log2 RPKM

higher 5mC enrichment, higher 5hmC enrichment

**Figure S1. Additional detailed analysis of meDIP- and hmeDIP-seq data sets,** Related to Figure 1.

(A) Global 5hmC levels in one ES cell line (ES-E14) and two TS cell lines (TS-GFP and TS-R26) measured by mass spectrometry expressed as a percentage of total cytosines. The experiment was carried out in triplicate.

(B) Scatter plot showing global meDIP-seq reads versus hmeDIP seq reads mapping to 2kb contiguous *in silico* probes. Experiments were carried out in triplicate. Values were normalized for total read count and converted to a $Log_2$ scale.

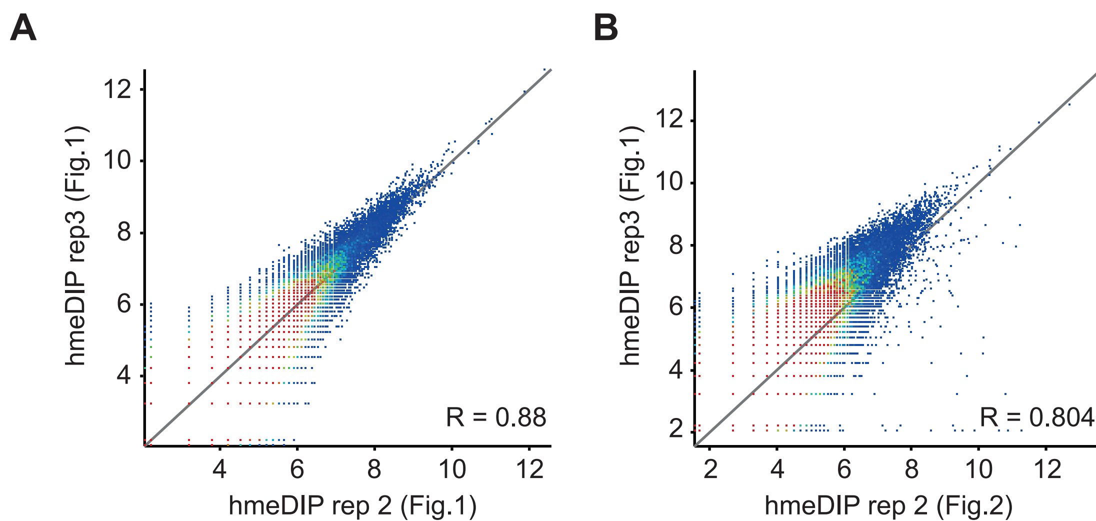
(C) Table showing absolute numbers and percentages of 5mC and 5hmC peaks, as determined by the MACS peak calling function within Seqmonk software, and their overlap with different genomic features. Experiments from Figure 1 were carried out in triplicate, those from Figure 2 in duplicate, for each modification and condition.

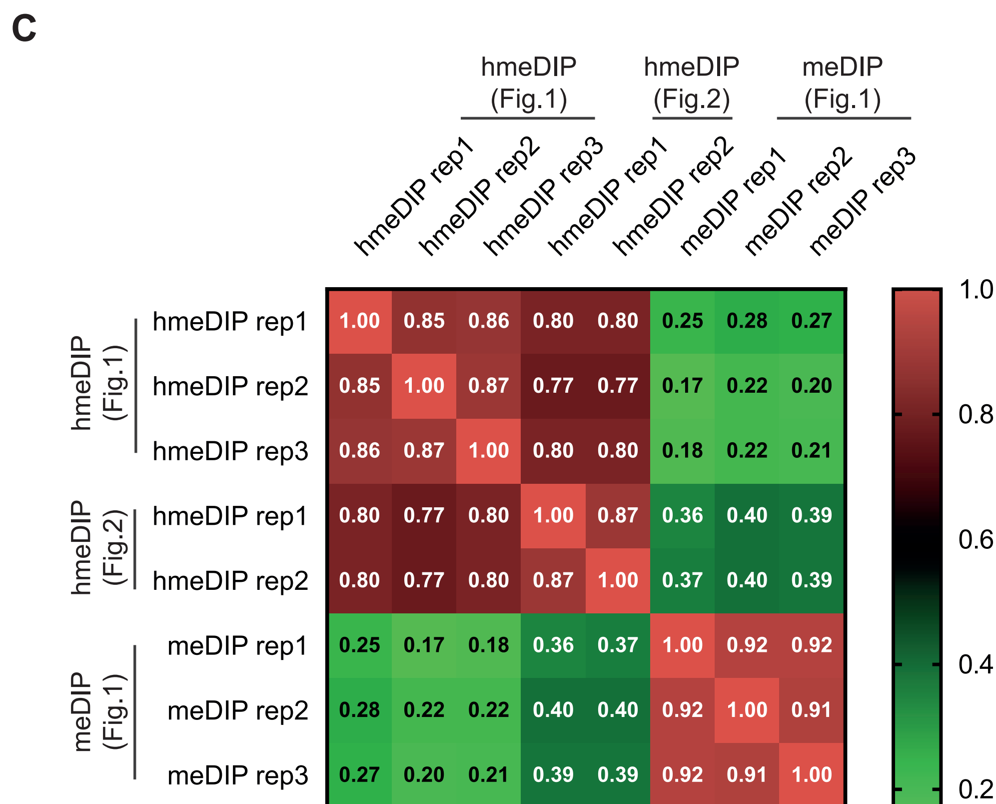
(D) Graphs showing percentages of meDIP (orange) and hmeDIP (blue) peaks aligning to various genomic features compared with the percentage of the genome that each feature represents (grey). Experiments from Figure 1 were carried out in triplicate, those from Figure 2 in duplicate, for each modification and condition.

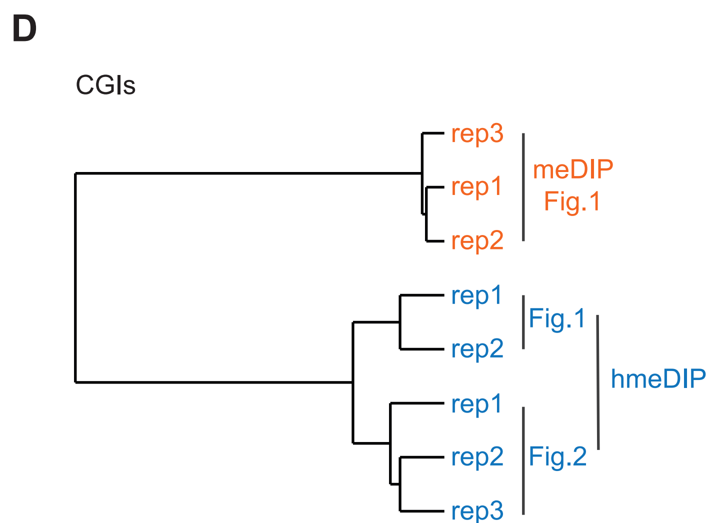
(E) Box-whisker plot showing expression ($Log_2$ RPKM) of genes where all exons have higher enrichment of 5mC than 5hmC (4234 genes in total), and genes where all exons have higher enrichment of 5hmC than 5mC (3897 genes in total). Data are of triplicate experiments.

(F) Scatter plot showing $Log_2$ normalised read counts mapping to CGIs from meDIP and hmeDIP-seq. Intergenic CGIs are highlighted in blue. Data are of triplicate experiments each.

**Figure S2**



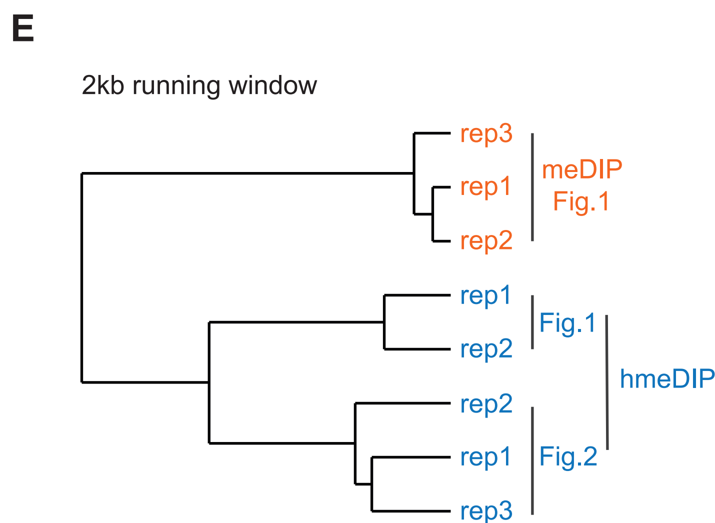



**Figure S2. Additional detailed validation of meDIP- and hmeDIP-seq replicate data sets,** Related to Figure 2.

(A) Scatter plot showing $Log_2$ normalised hmeDIP-seq read counts mapping to CGIs comparing two replicates of the data presented in Figure 1. R = Pearson's correlation.

(B) Scatter plot showing $Log_2$ normalised hmeDIP-seq read counts mapping to CGIs comparing one replicate of the data presented in Figure 1 with one replicate of the data presented in Figure 2. R = Pearson's correlation.

(C) Correlation matrix heat map showing Pearson's correlation values for all comparisons between meDIP-seq and hmeDIP-seq data generated from undifferentiated TS cells presented in Figures 1 and 2.

(D) Datastore tree diagram showing clustering of meDIP-seq and hmeDIP-seq data generated from undifferentiated TS cells presented in Figures 1 and 2 based on reads counted at CGIs.

(E) Datastore tree diagram based on a Pearson's correlation matrix showing clustering of meDIP-seq and hmeDIP-seq data generated from undifferentiated TS cells presented in Figures 1 and 2 based on reads counted over 2kb contiguous running windows covering the whole genome.

**Figure S3**

**A**

*Cdx2*
CGIs
TS rep1
TS rep2
TS Diff rep1
TS Diff rep2
2500
0
hmeDIP read count
147295000 147300000 147305000 147310000
Chromosome 5

*Nrip1*
CGIs
TS rep1
TS rep2
TS Diff rep1
TS Diff rep2
5000
0
76275000 76285000 76295000 76305000
Chromosome 16

*Lefty2*
CGIs
TS rep1
TS rep2
TS Diff rep1
TS Diff rep2
15000
0
hmeDIP read count
180885000 180890000 180895000 180900000
Chromosome 1

*Trpm6*
CGIs
TS rep1
TS rep2
TS Diff rep1
TS Diff rep2
4000
0
18740000 18750000 18760000
Chromosome 19

**B**

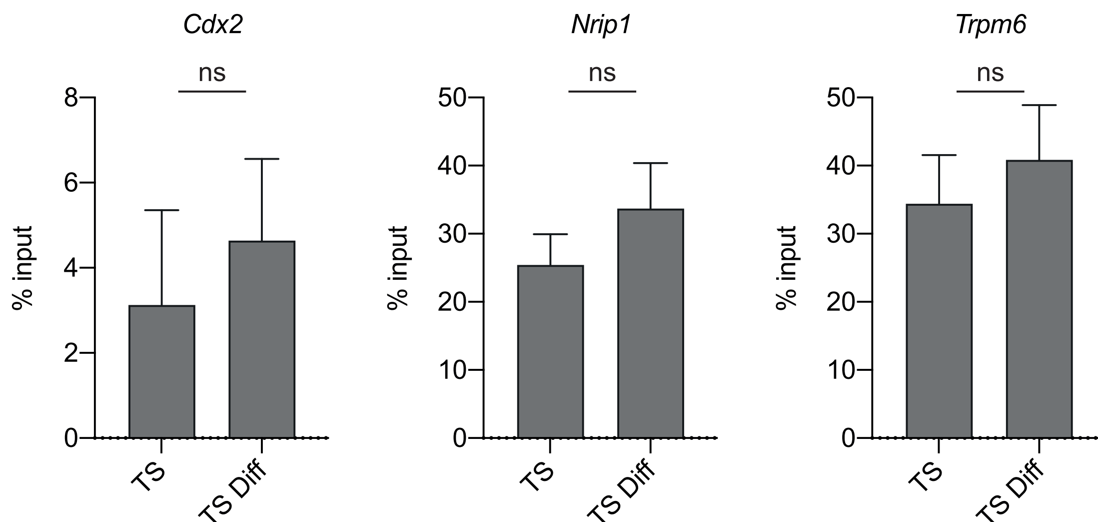


**Figure S3. DNA methylation and hydroxymethylation distribution detailed for each individual replicate data set, and additional meDIP-qPCR target gene analysis,** Related to Figure 2.

(A) Examples of hmeDIP-seq enrichment across genes that are down-regulated upon loss of 5hmC (*Cdx2, Lefty2*) or conversely that are upregulated with a gain of 5hmC (*Nrip1, Trpm6*) upon TS differentiation, showing individual replicates of the data shown in Figure 2E.

(B) meDIP-qPCR experiment showing methylation at *Cdx2*, *Nrip1* and *Trpm6* CGIs in TS cell differentiated in culture for 3 days. Data are mean +/- S.E.M. (n=3). Unpaired t-test revealed no significant difference.

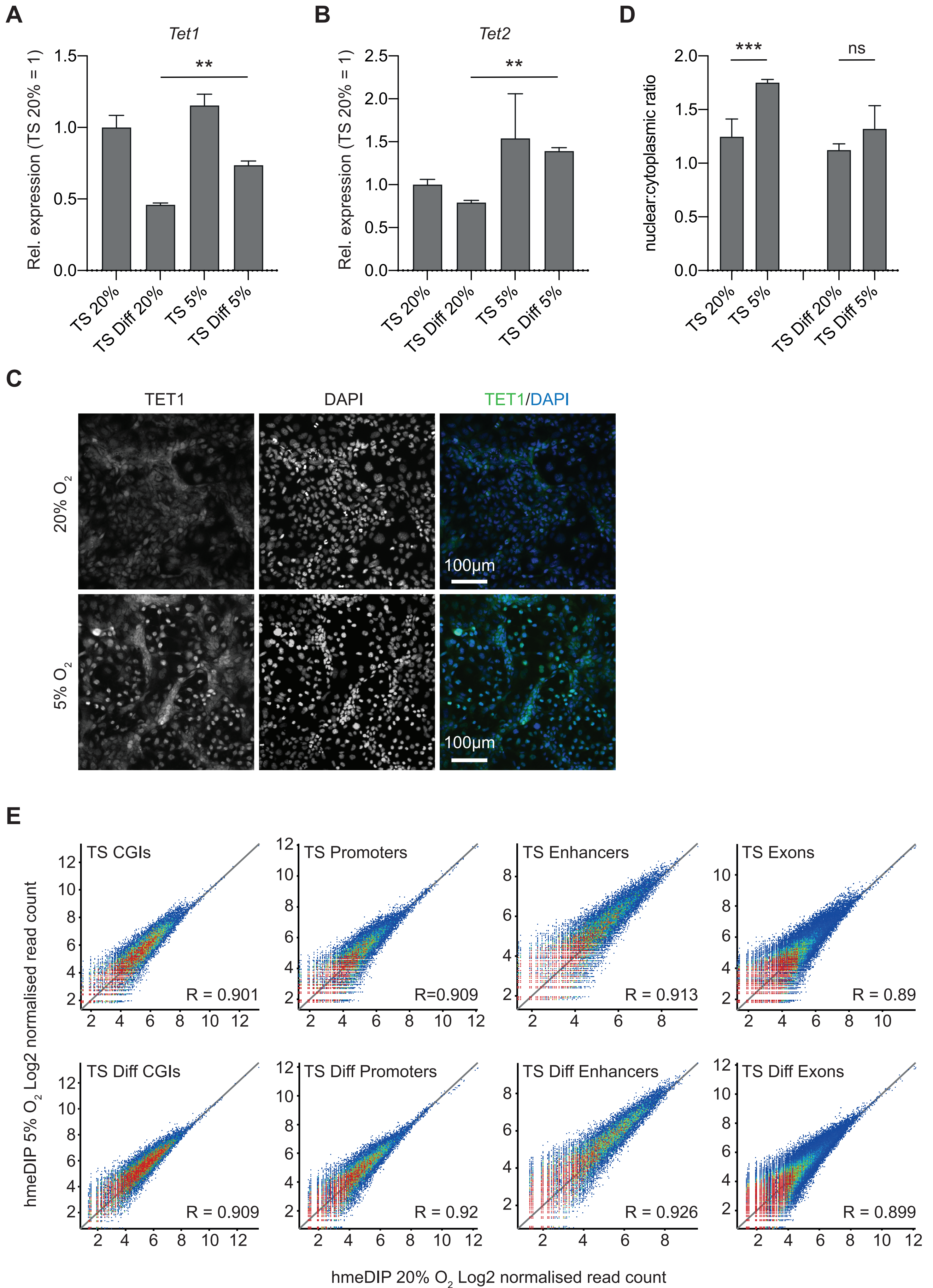


Figure S4
A
Tet1
Rel. expression (TS 20% = 1)
**
TS 20%
TS Diff 20%
TS 5%
TS Diff 5%
B
Tet2
Rel. expression (TS 20% = 1)
**
D
nuclear:cytoplasmic ratio
***
ns
C
TET1
DAPI
TET1/DAPI
20% O2
5% O2
100µm
E
TS CGIs
R = 0.901
TS Promoters
R=0.909
TS Enhancers
R = 0.913
TS Exons
R = 0.89
TS Diff CGIs
R = 0.909
TS Diff Promoters
R = 0.92
TS Diff Enhancers
R = 0.926
TS Diff Exons
R = 0.899
hmeDIP 5% O2 Log2 normalised read count
hmeDIP 20% O2 Log2 normalised read count

**Figure S4. Additional analysis of oxygen tension dependence of 5hmC dynamics with TS cell differentiation**, Related to Figure 3.

(A) RT-qPCR showing expression of *Tet1* in TS cells at 20% O2, TS cells differentiated for 3 days at 20% $O_2$, TS cells at 5% $O_2$ and TS cells differentiated for 3 days in 5% $O_2$. Expression is normalised to housekeeping gene *Dynein* and the TS cells at 20% $O_2$ samples set to 1. Data are mean +/- S.E.M. (n=2 separate experiments, ≥2 technical replicates each). **p<0.005.

(B) RT-qPCR showing expression of *Tet2* in TS cells at 20% $O_2$, TS cells differentiated for 3 days at 20% $O_2$, TS cells at 5% $O_2$ and TS cells differentiated for 3 days in 5% $O_2$. Expression is normalised to housekeeping gene *Dynein* and the TS cells at 20% $O_2$ samples set to 1. Data are mean +/- S.E.M. (n=2 separate experiments, ≥2 technical replicates each). **p<0.005.

(C) Immunofluorescence staining showing TET1 (green) in TS cells differentiated for 3 days in 20% and 5% $O_2$. TET1 protein levels appeared higher in 5% $O_2$. Data are representative of three independent experiments. Cells were counterstained with DAPI (blue).

(D) Nuclear:Cytoplasmic ratios of TET1 protein in TS cells at 20% $O_2$, TS cells differentiated for 3 days at 20% $O_2$, TS cells at 5% $O_2$ and TS cells differentiated for 3 days in 5% $O_2$ (n=4, n=2, n=4, n=4, respectively, with total numbers of nuclei analysed ≥100 for each condition). Unpaired t-tests were carried out and revealed a statistically significant difference between TS cells grown in 20% and 5% $O_2$. ***p<0.0005.

(E) Scatter plots showing $Log_2$ normalised hmeDIP-seq read counts mapping to CGIs (left), promoters (middle left), enhancers (middle right) and exons (right) in TS cells (top panel) and TS cells differentiated for 3 days (bottom panel) in 20% and 5% $O_2$. R = Pearson's correlation. Data are of duplicate experiments each.

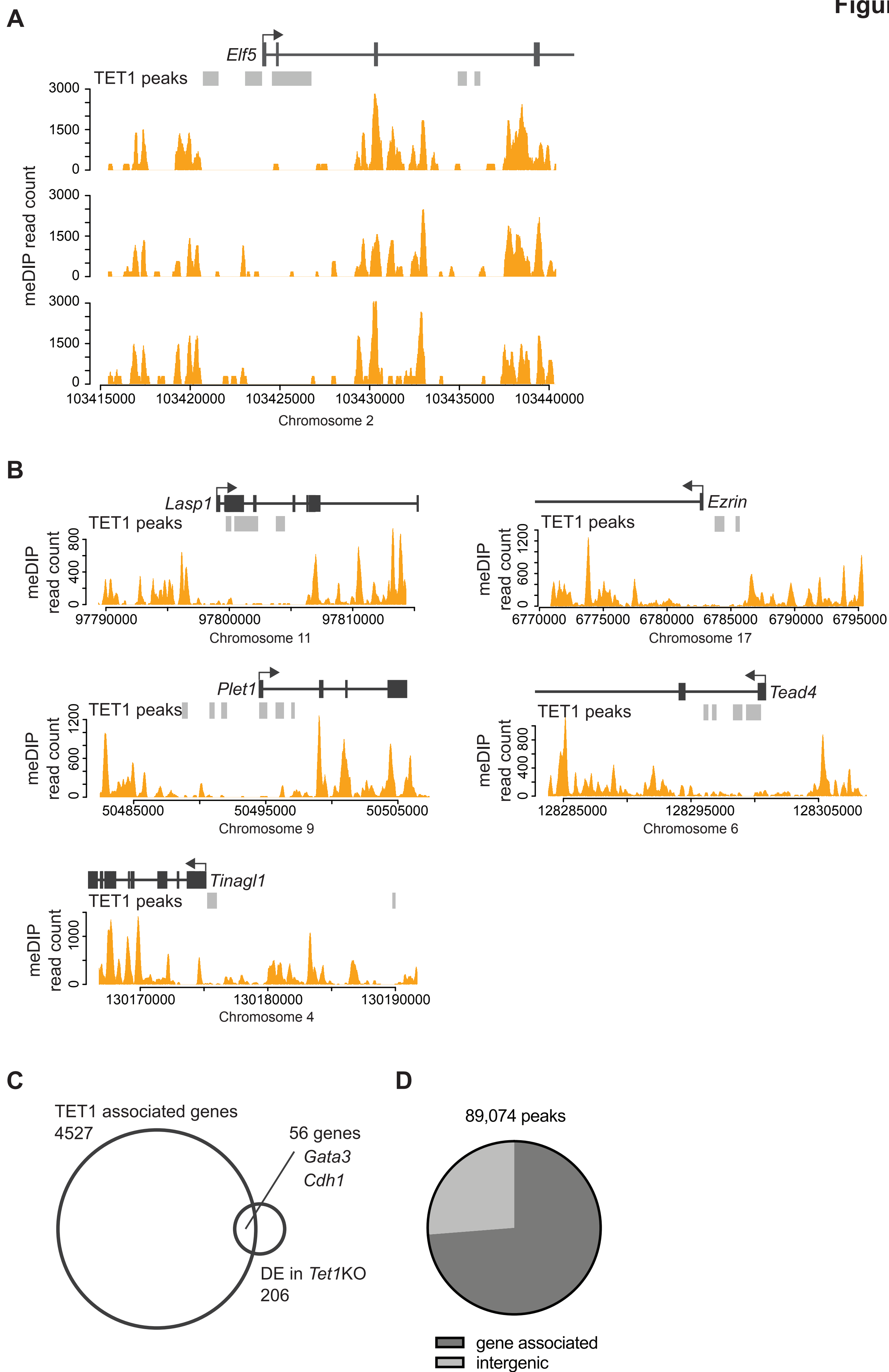


Figure S5
A
Elf5
TET1 peaks
meDIP read count
3000
1500
0
103415000
103420000
103425000
103430000
103435000
103440000
Chromosome 2
B
Lasp1
TET1 peaks
meDIP read count
97790000
97800000
97810000
Chromosome 11
Ezrin
6770000
6775000
6780000
6785000
6790000
6795000
Chromosome 17
Plet1
50485000
50495000
50505000
Chromosome 9
Tead4
128285000
128295000
128305000
Chromosome 6
Tinagl1
130170000
130180000
130190000
Chromosome 4
C
TET1 associated genes
4527
56 genes
Gata3
Cdh1
DE in Tet1KO
206
D
89,074 peaks
gene associated
intergenic

**Figure S5. meDIP-seq replicate data set visualization and additional TET1 ChIP-seq analysis,** Related to Figures 4 and 5.

(A) TET1 peaks at the hypomethylated region of the *Elf5* locus showing individual meDIP-seq replicates of the data shown in Figure 4G.

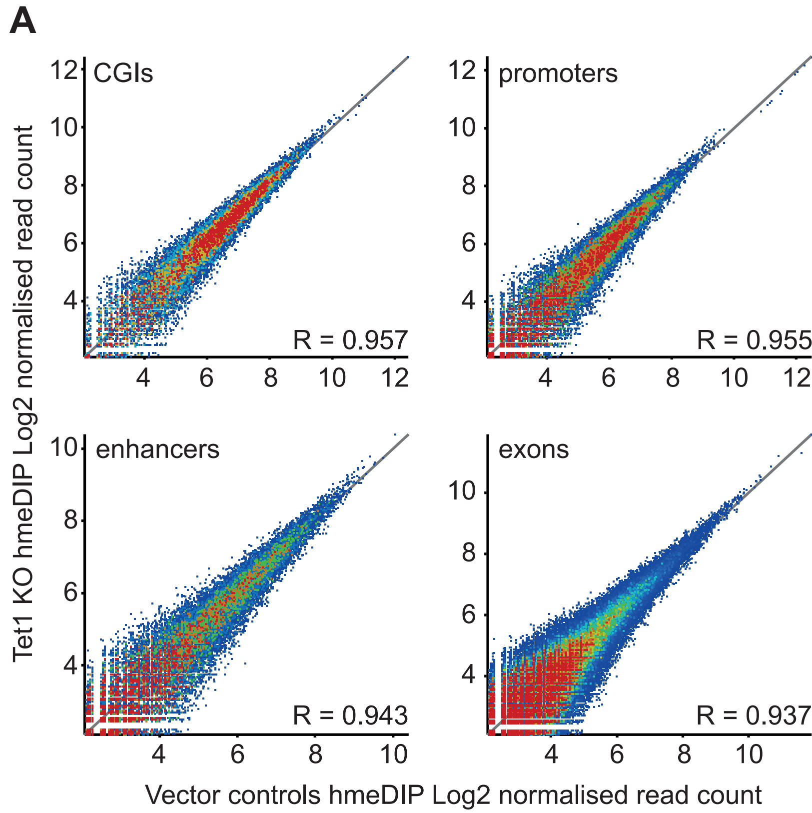
(B) TET1 peaks at hypomethylated regions at "gatekeeper" genes *Lasp1*, *Ezrin*, *Plet1*, *Tead4* and *Tinagl1*.

(C) Venn diagram showing overlap of TET1-associated genes and genes displaying differential expression in *Tet1* KO TS cells. ChIP-seq and RNA-seq experiments were performed in triplicate each.

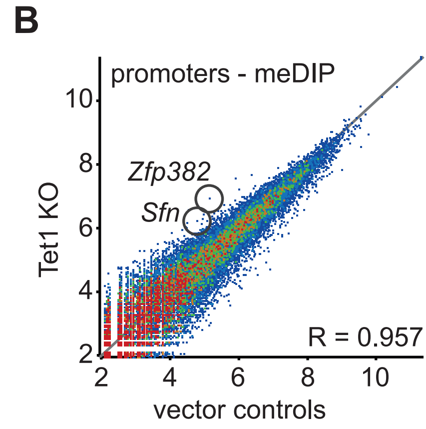
(D) Pie chart showing gene-associated and intergenic TET1 peaks detected in ES cells.

**Figure S6**

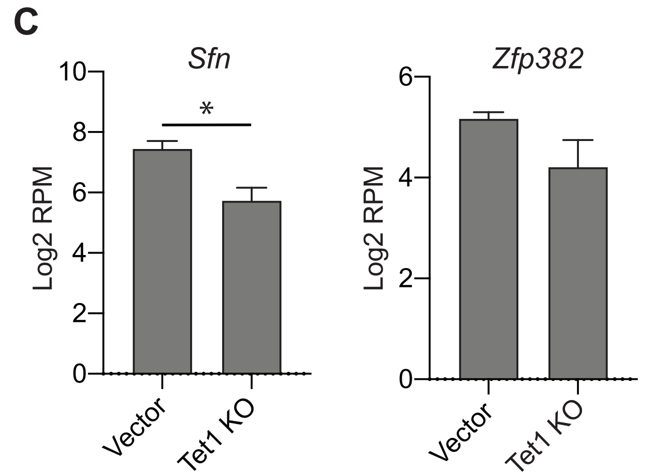
A




B




C




D

*Gpn2* *Sfn* *Zdhhc18*

CGIs

Vector controls rep1

Vector controls rep2

Vector controls rep3

*Tet1* KO rep1

*Tet1* KO rep2

*Tet1* KO rep3

meDIP read count

133590000 133600000 133610000

Chromosome 4

*Zfp382*

CGIs

Vector controls rep1

Vector controls rep2

Vector controls rep3

*Tet1* KO rep1

*Tet1* KO rep2

*Tet1* KO rep3

30120000 30130000 30140000

Chromosome 7

**Figure S6. Additional (h)meDIP-seq and candidate gene expression analysis in *Tet1* KO TS cells,** Related to Figure6.

(A) Scatter plots showing $Log_2$ normalised hmeDIP-seq read counts mapping to CGIs (top left), promoters (top right), enhancers (bottom left) and exons (bottom right) in vector controls and *Tet1* KO TS cells. R = Pearson's correlation. Data are of triplicate experiments.

(B) Scatter plot showing $Log_2$ normalised meDIP-seq read counts mapping to promoters in vector controls and *Tet1* KO TS cells. The *Sfn* and *Zfp382* promoters are circled. Data are of triplicate experiments.

(C) Expression (RPKM) of *Sfn* and *Zfp382* in vector controls and *Tet1* KO TS cells. Values are extracted from previously published RNA-seq data (Chrysanthou et al, 2018). Unpaired t-tests were carried out. *$p<0.05$ (n=3 each).

(D) meDIP-seq reads mapping to the *Sfn* and *Zfp382* loci in vector controls and *Tet1* KO TS cells. Individual replicates of data shown in Figure 6C.

**Supplemental Methods**

For immunofluorescence detection of 5hmC, DNA was denatured for 30 min with 2N HCl followed by neutralization with 100mM Tris, pH8.0 for 5 min. Cells were blocked with PBS, 0.1% Tween 20, 0.5% BSA (PBT/BSA), followed by 5hmC antibody (Active Motif, 39769 diluted 1:2000) incubation for 60 min. Primary antibody was detected with anti-rabbit Alexa Fluor 568 (Thermo Fisher Scientific) diluted 1:500. Nuclei were counter-stained with DAPI. Photographs were taken with an Olympus BX61 epifluorescence microscope or a Zeiss LSM 780 confocal microscope.

TET1 immunofluorescence was carried out using an anti-TET1 antibody (Genetex, GTX125888) diluted 1:750. Primary antibody was detected with anti-rabbit Alexa Fluor 488 (Thermo Fisher Scientific) diluted 1:500. Images to quantify TET1 nuclear and cytoplasmic signal were taken with Zeiss 780 confocal microscope and analysed with Fiji software using a macro code generated by the Babraham Imaging facility. The Fiji analysis of the confocal images was done as follows: a maximum intensity projection image was created, then a binary mask generated using the DAPI channel and 'Huang dark' auto-threshold. Average pixel intensity inside the masked regions was measured in the green channel-Tet1 (=nuclear signal), then the average pixel intensity above background was measured outside of the masked regions (=cytoplasm). Values were copied into Excel and ratios calculated.

RT-qPCR primers for *Tet1*, *Tet2*, *Sdha* and *Dynein* were as previously published (Chrysnathou et al, 2018). *Sfn* primers were forward 5'-GGAGGGGTCAGAAGAGAAGG and reverse 5'-CTTTGATGAGGTGCGAGTCC; *Gapdh* forward 5'-ACATCTCACTCAAGATTGTCAGC and reverse 5'-ATGGCATGGACTGTGGTCAT; and *Zfp382* forward 5'-TCAGACAAGGAGGCTCGT and reverse 5'-CTGTAGAGGGCTTTCTGG.