Cell Reports
Supplemental Information

Tet3 reads 5-carboxylcytosine through its CXXC domain and is a potential guardian against neurodegeneration

Seung-Gi Jin, Zhi-Min Zhang, Thomas L. Dunwell, Matthew R. Harter, Xiwei Wu, Jennifer Johnson, Zheng Li, Jiancheng Liu, Piroska E. Szabó, Qiang Lu, Guo-liang Xu, Jikui Song, and Gerd P. Pfeifer

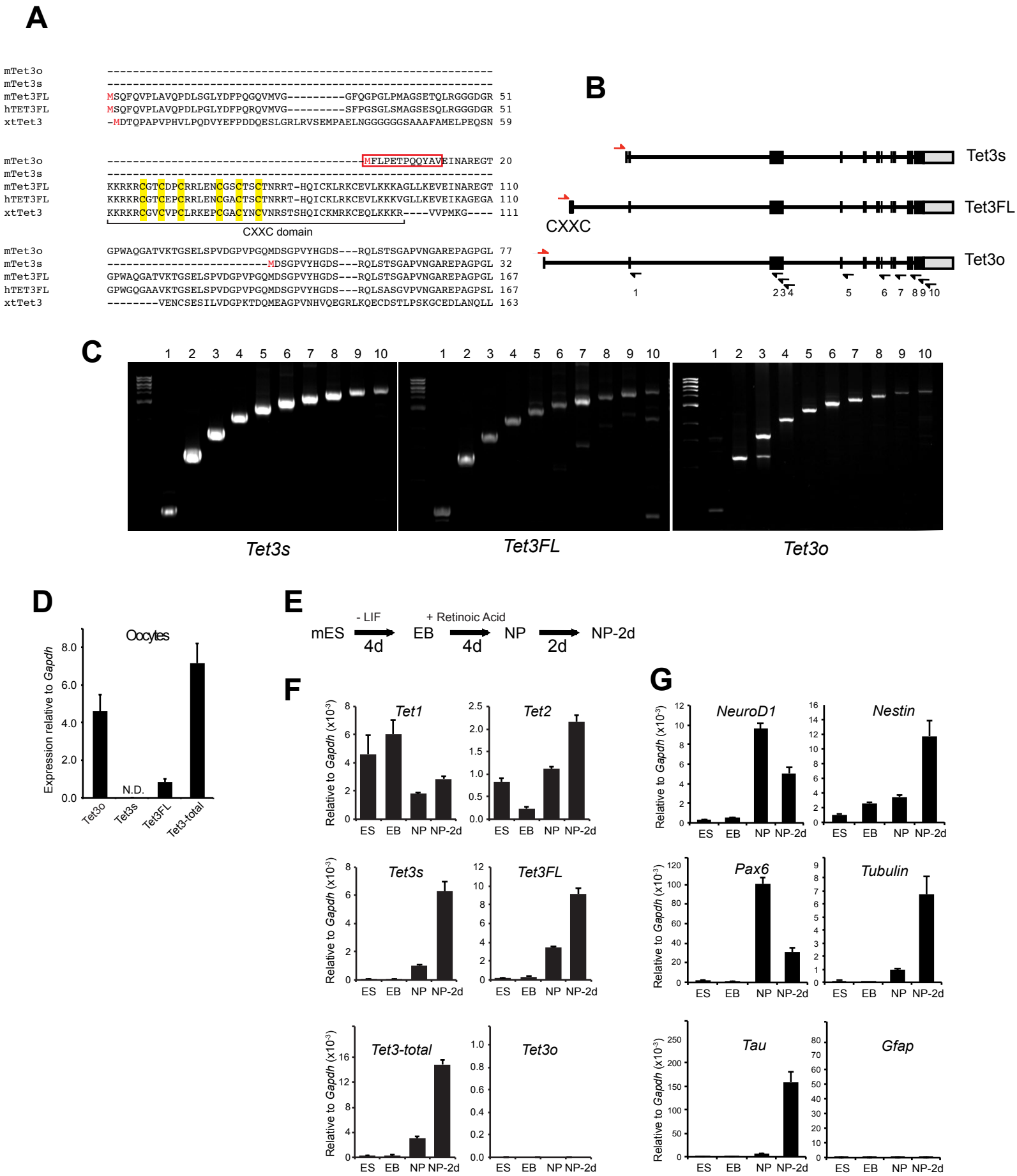
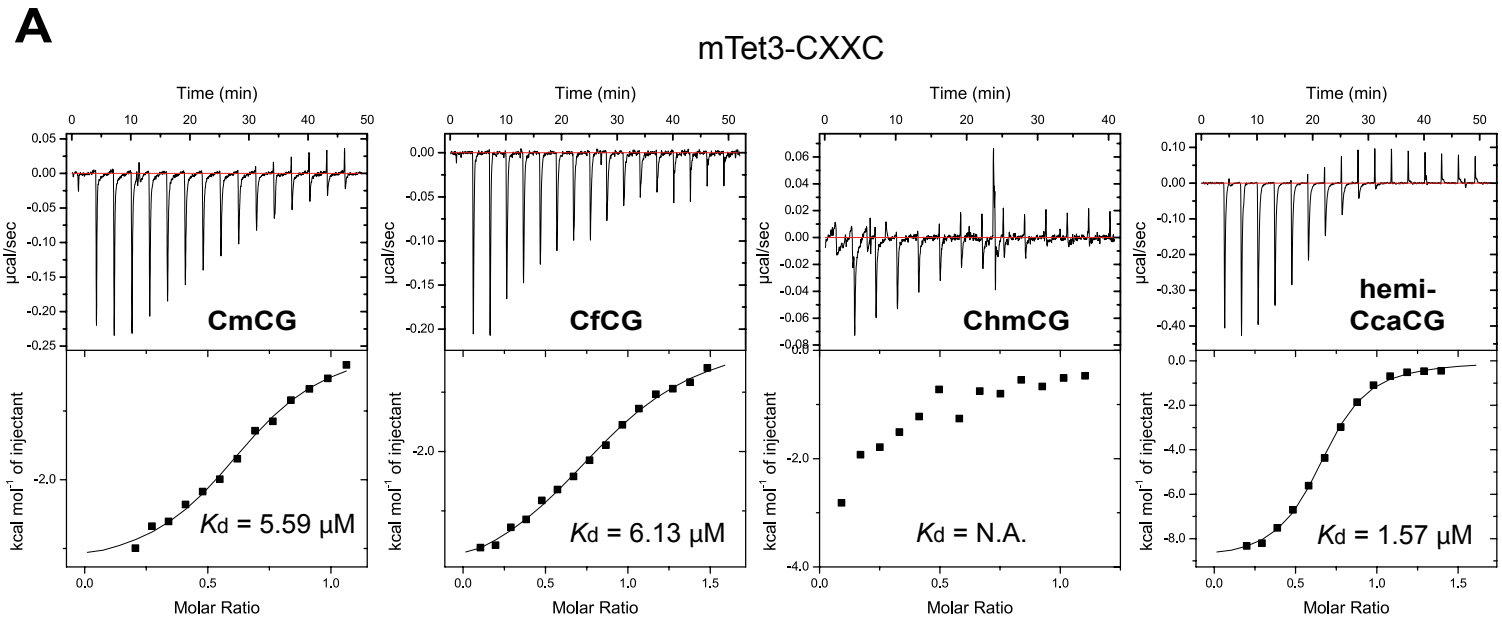
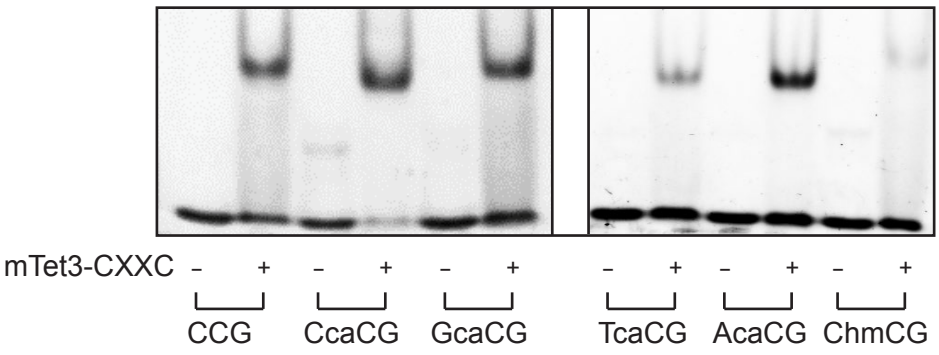

A
mTet3o ----------
mTet3s ----------
mTet3FL MSQFQVPLAVQPDLSGLYDFPQGQVMVG---------GFQGPGLPMAGSETQLRGGGDGR 51
hTET3FL MSQFQVPLAVQPDLPGLYDFPQRQVMVG---------SFPGSGLSMAGSESQLRGGGDGR 51
xtTet3 -MDTQPAPVPHVLPQDVYEFPDDQESLGRLRVSEMPAELNGGGGGGSAAAFAMELPEQSN 59
mTet3o ----------MFLPETPQQYAVEINAREGT 20
mTet3s ----------
mTet3FL KKRKRCGTCDPCRRLENCGSCTSCTNRRT-HQICKLRKCEVLKKKAGLLKEVEINAREGT 110
hTET3FL KKRKRCGTCEPCRRLENCGACTSCTNRRT-HQICKLRKCEVLKKKVGLLKEVEIKAGEGA 110
xtTet3 KKRKRCGVCVPCLRKEPCGACYNCVNRSTSHQICKMRKCEQLKKKR----VVPMKG---- 111
CXXC domain
mTet3o GPWAQGATVKTGSELSPVDGPVPGQMDSGPVYHGDS---RQLSTSGAPVNGAREPAGPGL 77
mTet3s ----------MDSGPVYHGDS---RQLSTSGAPVNGAREPAGPGL 32
mTet3FL GPWAQGATVKTGSELSPVDGPVPGQMDSGPVYHGDS---RQLSTSGAPVNGAREPAGPGL 167
hTET3FL GPWGQGAAVKTGSELSPVDGPVPGQMDSGPVYHGDS---RQLSASGVPVNGAREPAGPSL 167
xtTet3 --------VENCSESILVDGPKTDQMEAGPVNHVQEGRLKQECDSTLPSKGCEDLANQLL 163
B
Tet3s
Tet3FL
CXXC
Tet3o
1
2 3 4
5
6 7 8 9 10
C
1 2 3 4 5 6 7 8 9 10
Tet3s
Tet3FL
Tet3o
D
Oocytes
Expression relative to Gapdh
8.0
6.0
4.0
2.0
0.0
N.D.
Tet3o
Tet3s
Tet3FL
Tet3-total
E
mES
- LIF
4d
EB
+ Retinoic Acid
4d
NP
2d
NP-2d
F
Tet1
Tet2
Tet3s
Tet3FL
Tet3-total
Tet3o
Relative to Gapdh (x10-3)
ES EB NP NP-2d
G
NeuroD1
Nestin
Pax6
Tubulin
Tau
Gfap
Relative to Gapdh (x10-3)
ES EB NP NP-2d

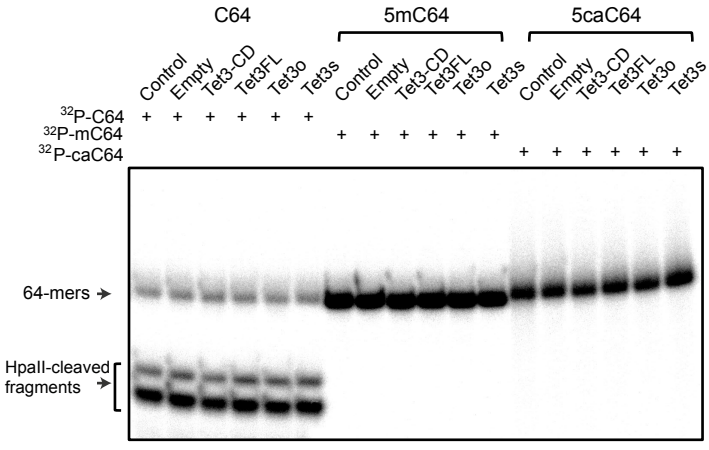
Fig. S1

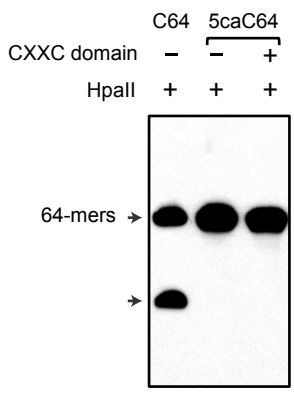


B

	mTet3-CXXC	
	K_d (µM)	N
CmCG	5.59±0.81	0.65±0.02
CfCG	6.13±0.57	0.89±0.02
ChmCG	N/A	N/A
Hemi-CcaCG	1.57±0.12	0.65±0.01







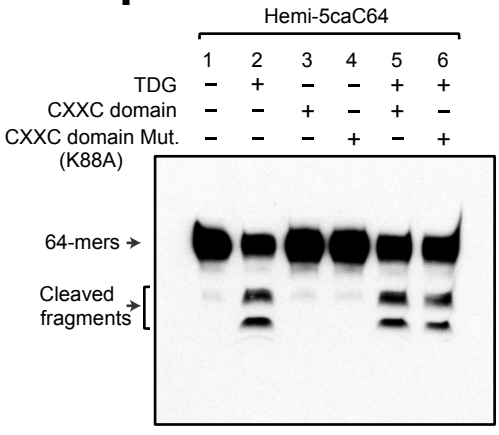
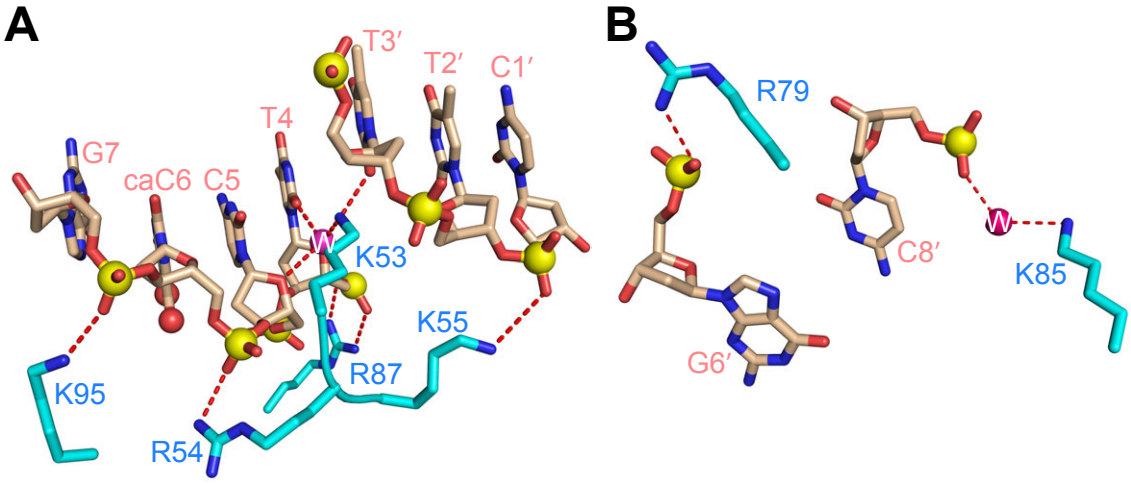
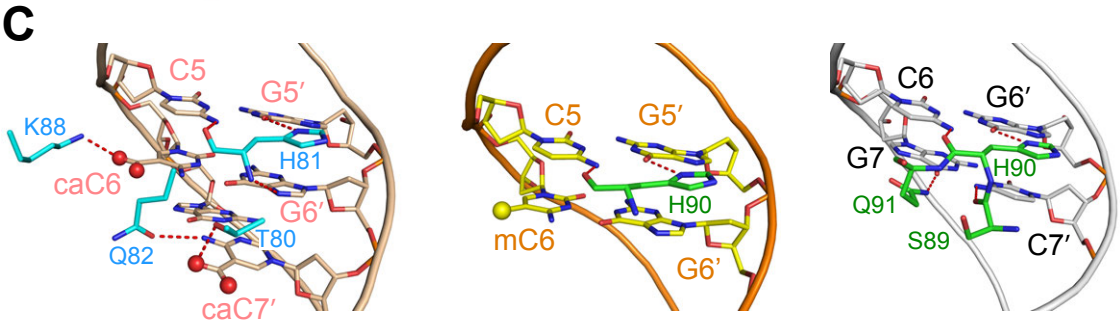
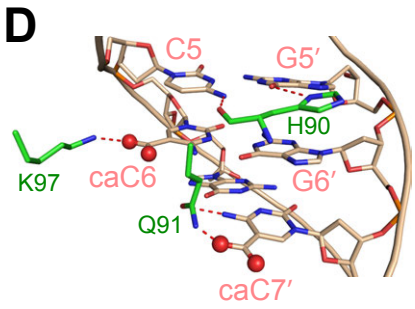


Fig. S2





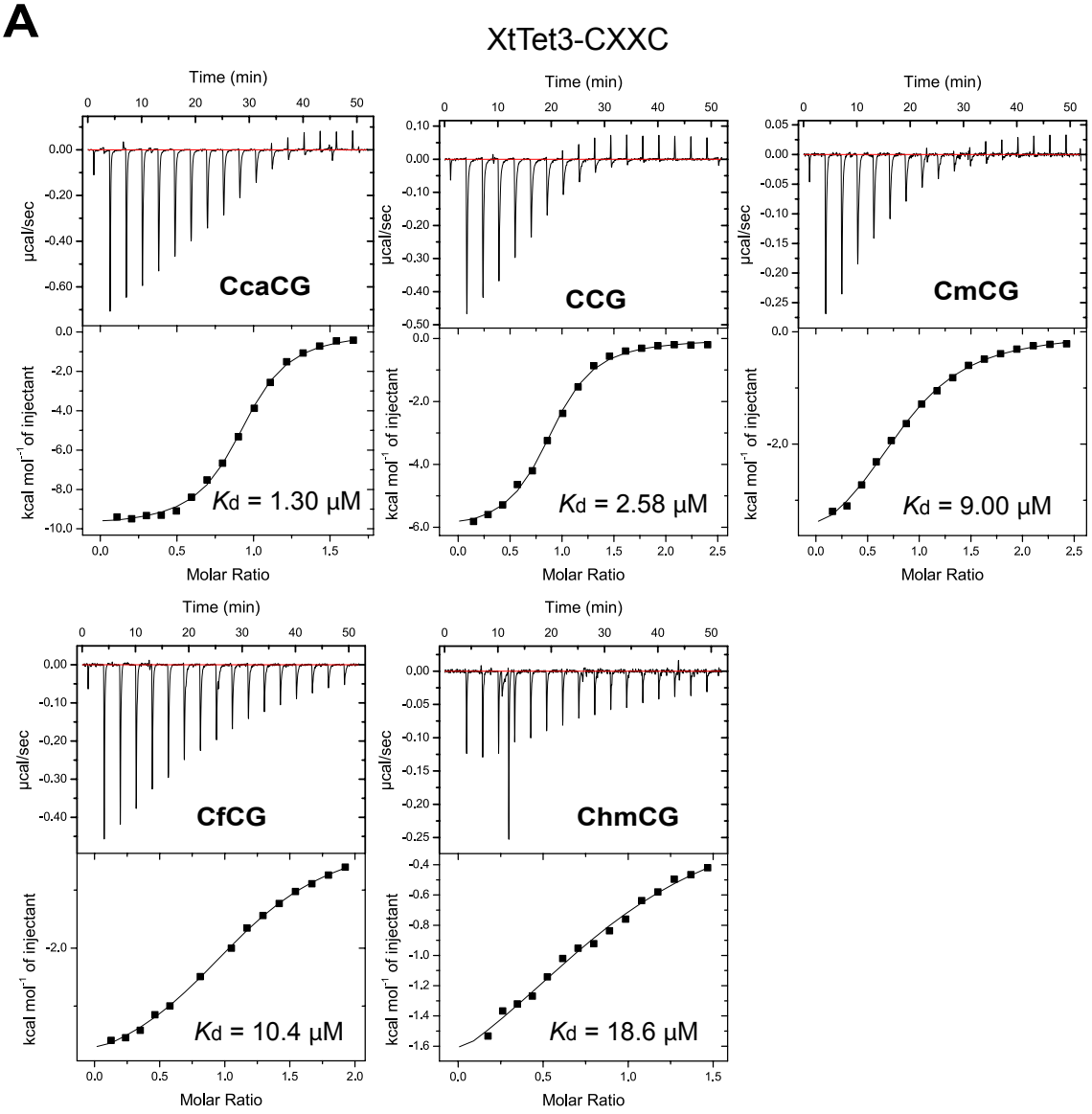
mTet3-CXXC - CcaCG XlTet3-CXXC - CmCG XlTet3-CXXC - ACG



XlTet3-CXXC - CaCG

Fig. S3

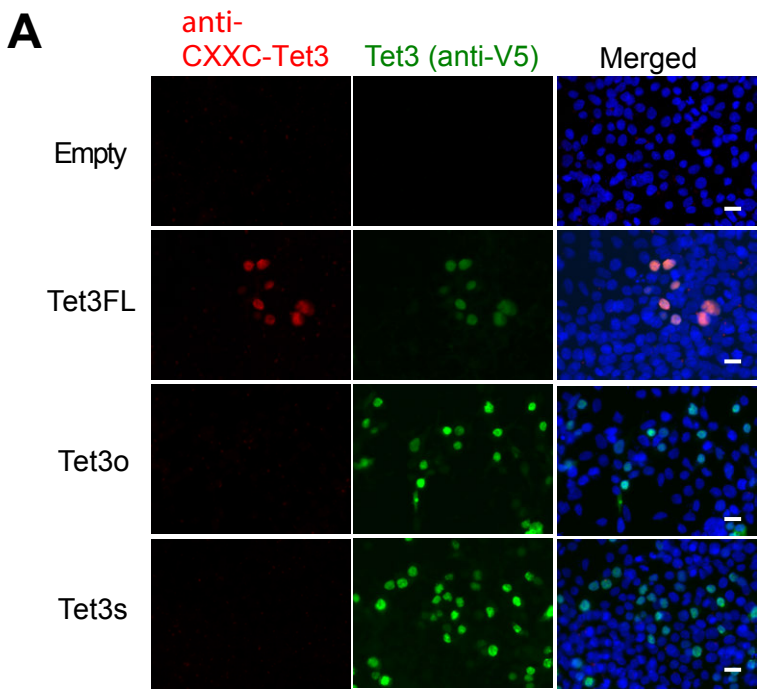
A

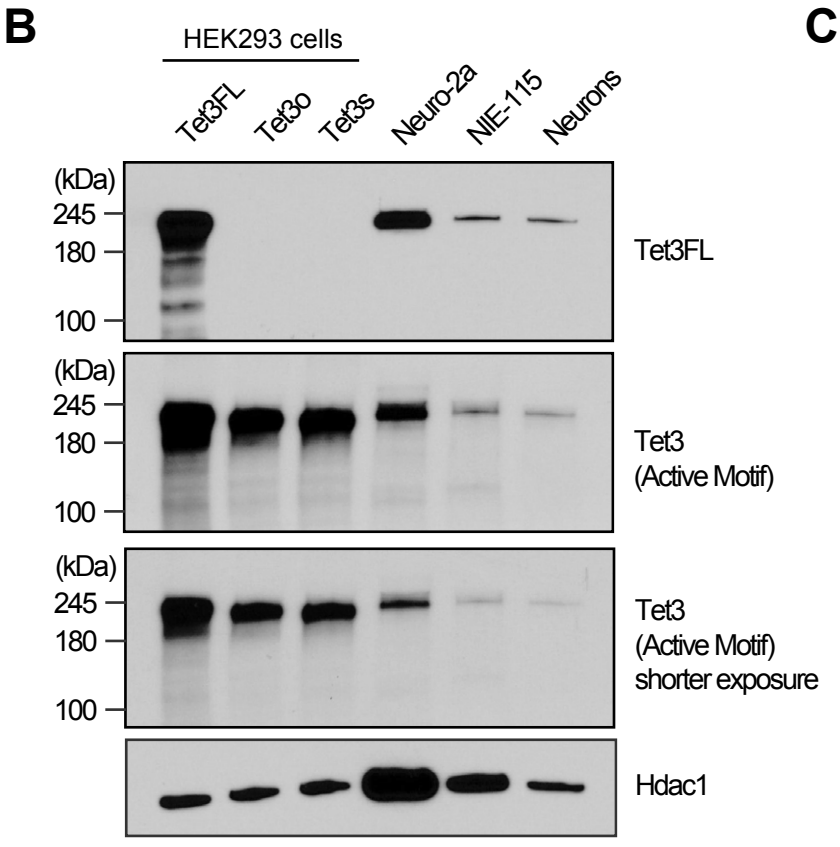


B

	XtTet3-CXXC	
	K_d (µM)	N
CcaCG	1.30±0.09	0.91±0.01
CCG	2.58±0.21	0.87±0.02
CmCG	9.00±0.60	0.83±0.02
CfCG	10.4±0.6	1.14±0.02
ChmCG	18.6±3.4	0.95±0.06

Fig. S4


A
anti-
CXXC-Tet3
Tet3 (anti-V5)
Merged
Empty
Tet3FL
Tet3o
Tet3s


B
HEK293 cells
Tet3FL
Tet3o
Tet3s
Neuro-2a
NIE-115
Neurons
(kDa)
245
180
100
Tet3FL
Tet3
(Active Motif)
Tet3
(Active Motif)
shorter exposure
Hdac1

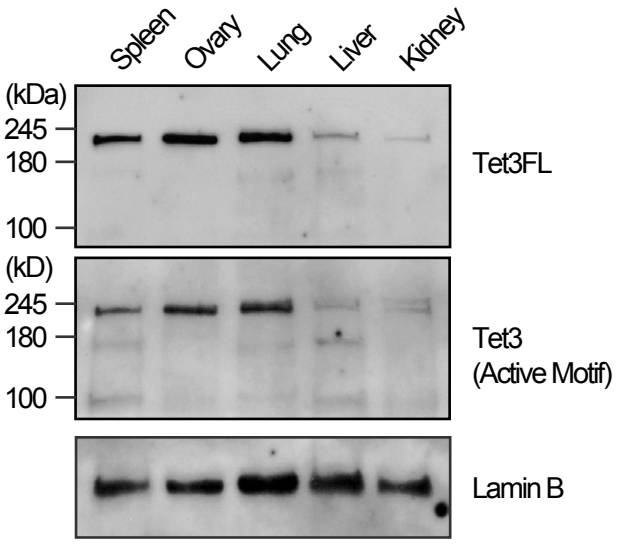
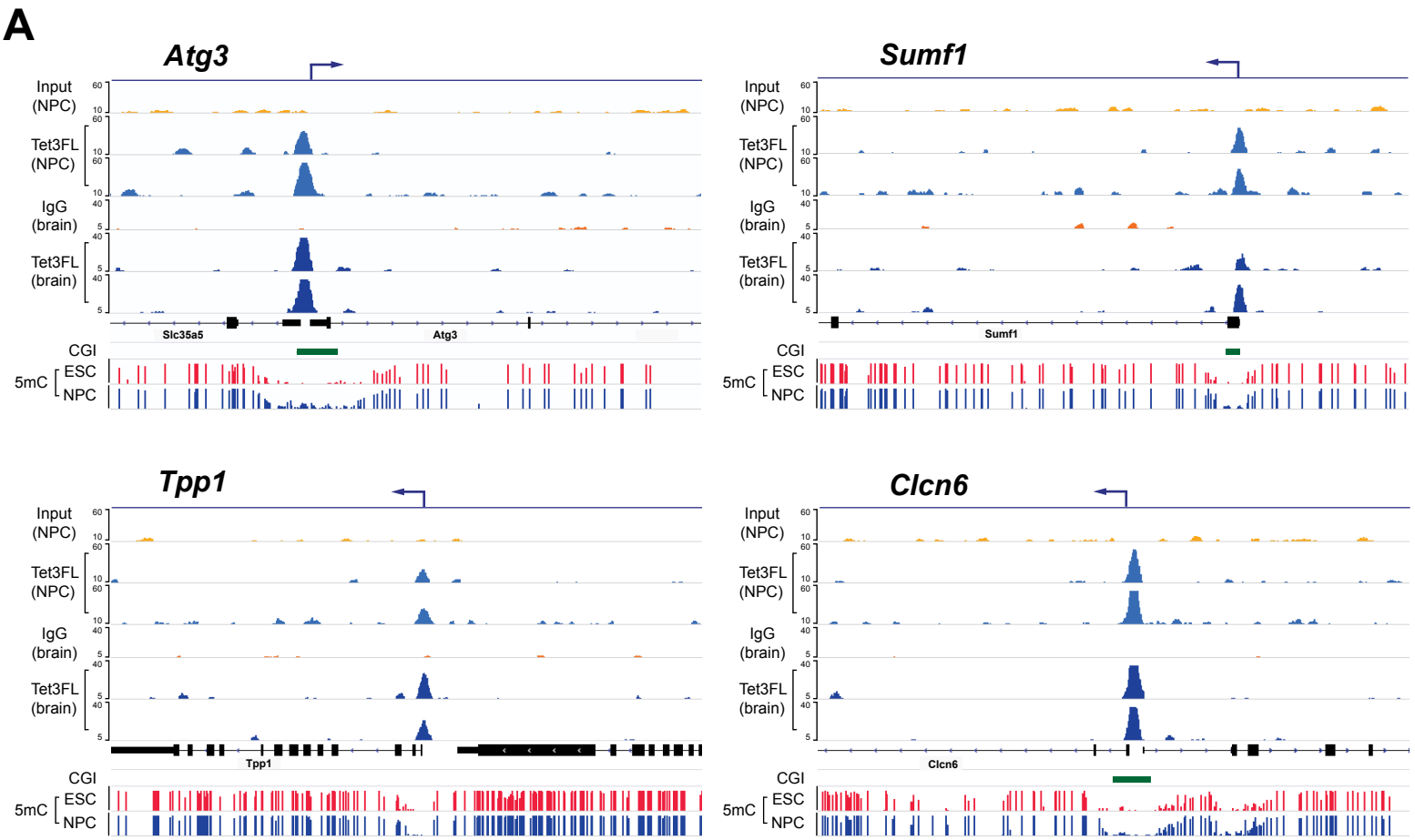

C
Spleen
Ovary
Lung
Liver
Kidney
(kDa)
245
180
100
Tet3FL
(kD)
Tet3
(Active Motif)
Lamin B

Fig. S5


A
Atg3
Input (NPC)
Tet3FL (NPC)
IgG (brain)
Tet3FL (brain)
Slc35a5
Atg3
CGI
5mC
ESC
NPC
Sumf1
Tpp1
Clcn6

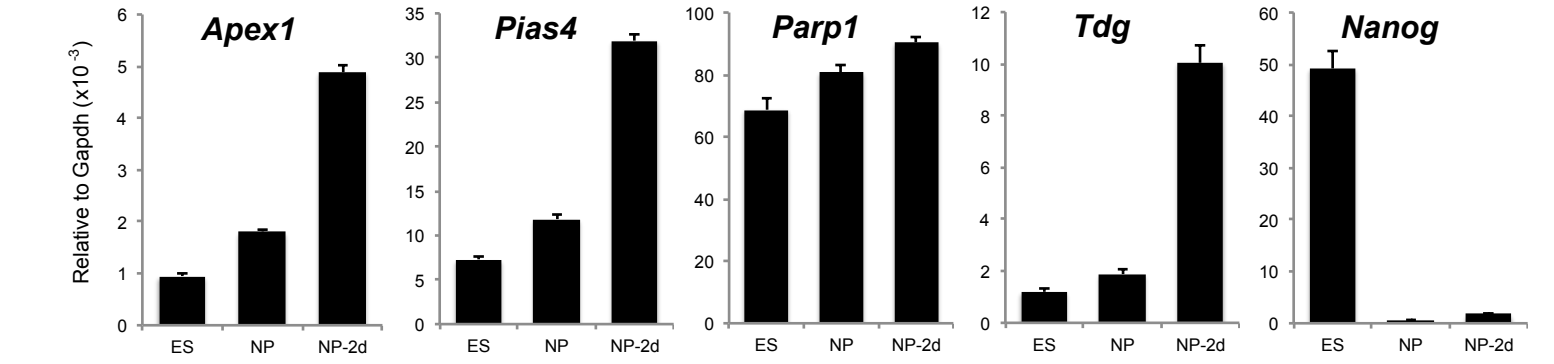

B
Relative to Gapdh (x10-3)
Apex1
Pias4
Parp1
Tdg
Nanog
ES
NP
NP-2d

Fig. S6

SUPPLEMENTAL FIGURE LEGENDS

Fig. S1: Expression of different *Tet3* isoforms. Related to Figure 1.
(A) The different isoforms of mouse and human Tet3 and of *Xenopus tropicalis* Tet3 are shown. The conserved cysteine residues of the CXXC domain are depicted in yellow. The unique N-terminus of Tet3o is marked by a red box.
(B) Isoform-specific primers (Table S1) were used to amplify specific *Tet3* isoforms.
(C) RT-PCR of different *Tet3* isoforms using primers as indicated in panel B and RNA from E15.5 brain (left and middle panel) or oocytes (right panel).
(**D)** Expression of *Tet3* isoforms in oocytes. N.D., not detected.
(**E)** Differentiation protocol used for production of neural progenitor (NP) cells from mouse ES cells.
(**F)** Expression of *Tet1*, *Tet2*, and the three *Tet3* isoforms in ES cells undergoing neuronal differentiation as outlined in panel E. Data are for three experiments +/- standard deviation. Expression of *Tet3o* was not detected.
(**G)** Expression of differentiation markers as analyzed by RT-PCR.

Fig. S2: Analysis of Tet3 CXXC domain properties. Related to Figure 2.
(A) Isothermal titration calorimetry analysis of the mTet3-CXXC domain binding to CmCG, CfCG, ChmCG and hemi-CcaCG DNA (cytosine on the opposite strand)**.**
(B) Summary of ITC parameters from (A).
(C) EMSA for the binding of mTet3-CXXC to CCG, CcaCG, GcaCG, TcaCG, AcaCG and ChmCG DNA.
(D,E) The Tet3 CXXC domain does not promote decarboxylation of 5-carboxylcytosine. (D) 64-mer double-stranded oligonucleotides containing cytosine (C64), 5-methylcytosine (5mC64) or 5-carboxylcytosine (5caC64) were incubated with lysates from HEK293 cells transfected with empty vector, Tet3-CD (catalytic domain), Tet3o, Tet3FL or Tet3s. After incubation, DNA was digested with HpaII, which cleaves unmethylated cytosine target sequences but does not cleave sequences containing 5mC or 5caC. (E) In vitro incubation of a 5caC-containing 64-mer with the CXXC domain of Tet3 does not lead to formation of unmethylated cytosines.
(F) Excision of 5caC by TDG is not affected by the CXXC domain of Tet3.

Fig. S3: Structural analysis of the Tet3 CXXC domain complex with 5-carboxylcytosine. Related to Figure 3.
(A,B) Hydrogen bonding interactions between the mTet3 CXXC domain and the backbone phosphate or the CcaCG-flanking DNA bases. The color scheme is the same as in Fig. 3.
(C) Molecular interactions of the recognition bases in the mTet3-CXXC - CcaCG DNA, XtTet3-CXXC - CmCG DNA (PDB 4HP1) and XtTet3-CXXC - ACG DNA (PDB 4HP3) complexes.
(D) Modeling analysis of the XtTet3-CXXC – CcaCG DNA complex.

Fig. S4: Isothermal titration calorimetry analysis of the XtTet3-CXXC domain binding to CCG with various modification states. Related to Figure 4.
(A) Interaction of XtTet3-CXXC with DNA containing CcaCG, CCG, CmCG, CfCG and ChmCG.
(B) Summary of ITC parameters from (A).

Fig. S5: Characterization of the anti-Tet3 CXXC domain antibody. Related to Figure 5.
(A) The different Tet3 isoforms were transfected into HEK293 cells. Cells were stained with anti-V5 tag antibody (green) and with the anti-Tet3FL (CXXC domain-specific) antibody (red). Scale bar, 20 μm.
(B) Western blot of transfected HEK293 cells and endogenous Tet3 in neuronal cells with anti-Tet3 antibodies. HEK293 cells were transfected with mouse Tet3 isoform expression vectors. Neuro-2a and NIE-115 are two mouse neuroblastoma cell lines. Neurons were derived by differentiation of ES cells. The top panel was created with the in-house CXXC domain-specific anti-Tet3 antibody revealing Tet3FL specifically. The two middle panels use a C-terminal specific anti-Tet3 antibody (Active Motif) that

recognizes also the shorter isoforms. Two different exposures are shown. An anti-Hdac1 antibody was used as a loading control.
(C) Western blot of Tet3 in different mouse tissues. The top panel was created with the in-house CXXC domain-specific Tet3 antibody revealing Tet3FL specifically. The middle panel uses a C-terminal specific anti-Tet3 antibody. An anti-Lamin B antibody was used as a loading control.

Fig. S6: Characterization of genes with Tet3FL peaks at the TSS. Related to Figure 6.
(A) Lysosome and autophagy pathway genes.
The genes shown are *Atg3*, *Sumf1*, *Tpp1*, and *Clcn6*. ChIP-sequencing shows Tet3FL enrichment at the transcription start sites (TSS) of genes in the lysosome/autophagy pathway in NPCs (light blue) and mouse brain (dark blue). The 5mC patterns (shown as vertical bars representing 0-100% methylation at individual CpG sites) were obtained from whole genome bisulfite sequencing data of mouse ESC (red) and NPC (blue) (Stadler et al., 2011).
(B) Increased expression of genes with Tet3FL peaks at the TSS in NPC. Expression of *Nanog* is shown as a control.

Table S1. RT-PCR primer sequences. Related to Figure 1.		
Gene	**Strand**	**Sequence**
PCR primers for the transcripts of Tet3 isoforms		
Tet3s	F	GCCGATGCAGTAGTGGAGG
Tet3FL	F	CTTCCTATGGCTGGGAGTGAG
Tet3o	F	CACATGTTCCTCCCAGAAAC
R-1	R	TTGGGTGGTTTCATGCCATGCCTTG
R-2	R	TGACTGGATGTAATCACTGGCGCTG
R-3	R	TTAGCCCTTCCAGAACAATCTGCCT
R-4	R	AGGTGTGTCTAGGTACTTTAGAGGTG
R-5	R	TAGGGTCTTTGCCTTGGCAAGCAC
R-6	R	GACCTCTCTGGGGAATGCTGTGAG
R-7	R	GAAGACAGGGTTGCTTGATGGAAAG
R-8	R	TGAGCTTGTCGGGAACAATGGTAGG
R-9	R	CTCATCCAGGAAGTTGTGTTCACTG
R-10	R	GATCCAGCGGCTGTAGGGGCCAGTGA
RT-PCR primers		
Oct4.	F	CGAGAAGGATGTGGTCCGAGT
	R	CAGTGAAGTGAGGGCTCCCATA
Nanog	F	ATGAAGTGCAAGCGGTGGCAGAAA
	R	CCTGGTGGAGTCACAGAGTAGTTC
Nestin	F	CGGGAGAGTCGCTTAGAGGTGCA
	R	ATCTTGAGGTGTGCCAGTTGCTGC
Pax6	F	CCGCCCTCACCAACACGTACAGT
	R	TTGCATGTGCGGAGGGGTGTAG
Tau	F	CTTTGAACCAGTATGGCTGACCCT
	R	CGAGGTGTGGCGATCTTCG
Neurod1	F	GTCCCAGCCCACTACCAATT
	R	CGGCACCGGAAGAGAAGATT
Tubulin	F	AGCGGATCAGCGTCTACTACA
	R	ATACTCCTCACGCACCTTGCT
Gfap	F	CCAAACTGGCTGATGTCTACC
	R	GCTTCATGTGCCTCCTGTCTA
Parp1	F	TGGCACTGCCCCTCTCTGTC
	R	GGCCTTGTTGGCAGATCCTGT
Apex1	F	AGTGGTGTGGGCCTACTTTCC
	R	CTGCATTGGGAACATAGGCTGT
Pias4	F	TGGACGAATTGCTGAAGCCCACT
	R	AGTAACAGATCCTCAGCACCACCT
Nanog	F	ATGAAGTGCAAGCGGTGGCAGAAA
	R	CCTGGTGGAGTCACAGAGTAGTTC
Gapdh	F	GCACAGTCAAGGCCGAGAAT
	R	GCCTTCTCCATGGTGGTGAA
Taqman primers and probes		

Tet3o	F	CACATGTTCCTCCCAGAAAC
	R	CTGCCTTGAATCTCCATGGTAC
	Probe	5'-6FAM-TCTTGACAGTCGCCCCTTGTGCCCA-Iowa Black FQ-3'
Tet3s	F	GCCGATGCAGTAGTGGAGG
	R	CTGCCTTGAATCTCCATGGTAC
	Probe	5'-6FAM-TCTTGACAGTCGCCCCTTGTGCCCA-Iowa Black FQ-3'
Tet3FL	F	TGGAAAACTGTGGGTCTTGTAC
	R	GAGCATTTATTTCCACCTCCTTA
	Probe	5'-6FAM-ACTTGCGGAGTTTGCAGATCTGGTGT-Iowa Black FQ-3'
TaqMan ID (Applied Biosystems)		
Tet1		Mm01169088_m1
Tet2		Mm01312907_m1
Tet3-total		Mm00805754_m1
Tdg		Mm02602088_g1

Table S2. Crystallographic statistics for the mTet3-CXXC-5caC DNA complex. Related to Figure 3.

Data collection	
Wavelength (Å)	0.9774
Space group	*C2*
Cell dimensions	
a, *b*, *c* (Å)	75.0, 27.2, 47.1
α, β, γ (°)	90, 104.4, 90
Resolution (Å)*	50-1.3 (1.35-1.3)
R_{sym} or R_{merge}*	6.1 (25.6)
$I/\sigma I$	24.8 (5.2)
Completeness (%)*	99.1 (100.0)
Redundancy*	4.2 (3.9)
Unique reflections	23,079 (2,280)
Refinement	
Resolution (Å)	32.6-1.3
No. reflections	23,064
R_{work}/ R_{free}	16.8/18.7
No. atoms	
Protein	361
DNA	491
Zinc	2
Water	230
B-factors	
Protein	14.6
DNA	15.2
Zinc	11.2
Water	25.9
R.m.s deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.523

*Highest resolution shell is shown in parenthesis.

Table S3: List of Tet3FL peaks (TSS). Related to Figure 5.
Please see Excel file.

Table S4: Examples of genes with Tet3FL peaks at TSS (MSigDB Pathway). Related to Figure 6.

	Lysosome function	mRNA processing	Base excision repair
1	*Atp6v0b*	*Ccar1*	*Apex1*
2	*Atp6v0d1*	*Eftud2*	*Hmgb1*
3	*Atp6v1h*	*Eif4e*	*Parp1*
4	*Cln3*	*Gtf2f1*	*Parp3*
5	*Ctsa*	*Hnrnph2*	*Tdg*
6	*Ctsd*	*Nfx1*	*Stub1/ChIP*
7	*Gba*	*Nup188*	*Rad9*
8	*Gla*	*Ptbp1*	
9	*Hexa*	*Snrnp40*	
10	*Igf2r*	*Snrpa*	
11	*Lamp1*	*Srsf2*	
12	*Mcoln1/Trpml1*		
13	*Nagpa*		
14	*Sumf1*		
15	*Tpp1*		

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cells and tissue

HEK293 cells (ATCC; Manassas, VA) were maintained as a monolayer in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum at 37°C in a 5% CO_2 standard incubator. DNA transfections were carried out using BioT (Bioland; Paramount, CA) according to the manufacturer's instructions. Vitamin C (L-ascorbic acid, Sigma) was added on day 1 after transfection at a final concentration of 10 to 20 μM in culture medium.

J1 mouse embryonic stem cells (mESCs) (ATCC; Manassas, VA) were differentiated as previously described (Bibel et al., 2007). In brief, mESCs were grown on gelatin-coated plates in the presence of feeder cells at 37°C in knockout DMEM (Invitrogen; Carlsbad, CA) supplemented with 15% (v/v) Knockout Serum Replacement, 1x non-essential amino acids, 100 μM β-mercaptoethanol, 2 mM L-glutamine, and 1,000 U/ml LIF (ESGRO; Millipore, USA). To induce neuronal differentiation, mESCs were grown in the absence of feeder cells for three passages and then cultured on bacterial dishes (Greiner Bio-one) in medium lacking LIF and with 10% ES-qualified fetal bovine serum to induce embryoid body (EB) formation for 4 days. Next, EBs were further incubated in 5 μM retinoic acid in culture medium for 4 days and were then trypsinized. The dissociated EBs were pelleted by centrifugation at 700xg for 3 min and re-suspended in N2 medium (DMEM/F12 mixture, 1:1 with 1x N2 supplement). The cell suspension was filtered through a 40 μm nylon cell strainer (BD Biosciences) and plated on PORN/laminin-coated plates at a density of 1.5 x 10^7 cells per 100 mm plate. The cells were further incubated for 2 days with change of N2 medium 2 h and 24 h after plating.

Oocytes were collected from 6- to 8-week-old female C57Bl/6J mice. Females were super-ovulated by peritoneal injection with 5 IU pregnant mare serum gonadotropin (PMSG) followed by 5 IU of human chorionic gonadotropin (hCG) 46 hours later. Oocytes were teased from the ovaries 24 hours after hCG injection. Cumulus cells were completely removed from oocytes by 1% hyaluronidase treatment. Mouse embryo brain tissue was obtained from cortex, ganglionic eminence and hippocampus at E15.5. These tissues contain mostly neurons and few astrocytes. Gliogenesis in these areas starts around E17 to E18.

Quantitative reverse transcription PCR (RT-qPCR)

Poly(A) mRNAs were isolated from mouse oocytes (n=200) by using the Dynabeads mRNA DIRECT Micro Kit (Invitrogen). Total RNAs were isolated from mouse tissues and cultured cells by using PureLink RNA Mini Kit (Ambion). The RNAs were reverse transcribed with the SuperScript III reverse transcriptase (Invitrogen), according to the manufacturer's instructions. Real-time quantitative PCR reactions with target-specific primers (Table S1) were performed at 50°C for 2 min and 95°C for 10 min followed by 50 cycles at 95°C for 15 seconds and 60°C for 1 min using TaqMan Gene Expression master mix or Power SYBR Green master mix (Applied Biosystems; Foster City, CA) on a ViiA 7 real-time PCR cycler (Applied Biosystems). The cDNA levels of target genes were analyzed using comparative *Ct* methods and normalized to internal standard, *Gapdh*.

Plasmid constructs

For mammalian expression vectors, the open-reading frames (ORFs) of mouse Tet3 isoforms, *Tet3s*, *Tet3FL*, *Tet3o* and *Tet3-CD* (catalytic domain, residues 832-1803) were prepared by PCR using cDNA from mouse neurons or oocytes. The amplified fragments were cloned into the pEF-DEST51 Gateway vector (Invitrogen). The C60A, C72A or K88A *Tet3FL* constructs were obtained by site-directed mutagenesis.

Protein expression and purification

For the bacterial expression vectors, the mouse and human TET3 CXXC domains (residues 1-182) were cloned into the BamHI and SalI sites of the GST pGEX-5X-1 expression vector (GE Healthcare). The CXXC domains of Tet3 were expressed in BL21 competent cells and the cells were grown at 37°C until OD_{600} reached 0.4-0.5. Protein expression was then induced with addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to 0.25 mM and further incubation for 4 h at 37°C. The GST-tagged proteins were purified on Glutathione Sepharose 4B beads (GE Healthcare). The recombinant proteins were eluted from the beads using 7.5 ml elution buffer (50 mM Tris-HCl pH 8.5, 150 mM NaCl, 20 mM reduced

glutathione, 0.1% Triton X-100) and further purified using Amicon Ultra-15 centrifugal filter units by centrifugation at 3,500 rpm at 4°C for 30 min, followed by washing with cold PBS. The proteins were stored at -80°C in storage buffer containing 50 mM Hepes pH 7.4, 150 mM NaCl, 2 mM DTT, and 50% glycerol.

For the structural study, the DNA sequence encoding residues 51-96 of mouse Tet3 (mTet3-CXXC) was inserted into a modified pRSFDuet-1 vector (Novagen), in which mTet3 was separated from the preceding His_6-SUMO tag by a ubiquitin-like protease (ULP1) cleavage site. The fusion proteins were expressed in the BL21 (DE3) RIL cell strain (Stratagene). The cells were grown at 37°C until OD_{600} reached approx. 1.0. The temperature was then shifted to 20°C and the cells were induced by 0.4 mM IPTG. After induction, the cells continued to grow overnight. The fusion protein was purified through a Ni-NTA affinity column, and the His_6-SUMO tag was subsequently removed by ULP1 cleavage. The protein sample was further fractionated through a heparin column, followed by gel filtration on a 16/60 G200 Superdex column. The final sample for crystallization contained about 4-5 mM mTet3-CXXC, 20 mM Tris-HCl, pH 7.5, 50 mM NaCl. For ITC assay, the DNA sequence encoding the XtTet3-CXXC domain (residues 58-111), plus an extra C-terminal tryptophan, was inserted into the pRSFDuet-1 vector. The protein sample for the XtTet3-CXXC domain was produced in the same way as that described for mTet3-CXXC.

Electrophoretic mobility shift assay (EMSA)

We used double-stranded 64 bp oligonucleotides produced by the Keck oligo synthesis lab at Yale University. The sense strand was 5'-CCTCACCATCTCAACCAATATTAXGXGTATAT-CXGGTATTTTXGAATTGAGGGAGAAGTGGTGA-3' and the anti-sense strand was 5'-TCACCACTTCTCCCTCAATTXGAAAATACXGGATATAXGXGTAATATTGGTTGAGATGGTGAG G-3' where X is cytosine, 5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, or 5-carboxylcytosine. The oligonucleotides were labeled with biotin at the 3' end with DNA 3' End Biotinylation Kit (Thermo Scientific) following the manufacturer's instructions.

The protein-DNA binding reaction was prepared in the following order: H_2O for a final volume of 20 µl, 2 µl 10x binding buffer (100 mM Tris-HCl pH 7.8, 500 mM NaCl, 100 mM $MgCl_2$, 0.5% NP-40, 50% glycerol, 250 ng/µl BSA), 1 µl fresh 20 mM DTT, 1 µl 1 mM $ZnCl_2$, 125 ng of sonicated JM110 *E. coli* genomic DNA, 1 µl (10 fmoles) double-stranded probe, and approximately 400 ng recombinant proteins. The reactions were incubated for 45 min at room temperature. The protein-DNA complexes were fractionated by electrophoresis on 6% non-denaturing polyacrylamide gels. Probes were then transferred to a GeneScreen Plus® hybridization transfer membrane (Perkin Elmer) using a semi-dry transfer cassette (Bio-Rad) for 45 min at 10 V. Biotin labeled probes were visualized using the Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Scientific) following the manufacturer's instructions.

To evaluate the binding of caCG DNA in different sequence contexts, we mixed 25 µM of mTet3-CXXC (residues 51-97; in addition, a Trp residue was introduced to the C-terminus for the purpose of protein concentration measurement using UV spectroscopy) with 7.5 µM of palindromic 12-mer DNA duplexes, containing a central 5-carboxylated CpG dinucleotide (strand sequence: GAATXcaCGGATTC, caC = 5-carboxyldeoxycytidine, X = C, G, T or A) individually in a buffer containing 20 mM Tris-HCl (pH 7.5), 5% glycerol, 150 mM NaCl. For comparison, 12-mer DNA duplexes containing a central unmodified or 5-hydroxymethylated CpG dinucleotide (strand sequence: GAATCXGGATTC, X = C or 5-hydroxymethyldeoxycytidine,) were also used. Binding reaction mixtures were electrophoresed in 8% DNA retardation gel in 1xTBE buffer (89 mM Tris-borate, 2 mM EDTA, pH 8.3) at 4°C at 100 V for 50 min. DNA was visualized by ethidium bromide staining.

Immunofluorescence staining

HEK293 cells were plated at a density of 2.5 x 10^5 cells per well directly on poly-L-lysine treated coverslips in a six-well tissue culture dish. The transfection was achieved using BioT (Bioland) according to the manufacturer's instructions using 2 µg of plasmids. At 48 h after transfection, cells were washed twice with PBS and fixed in 4% paraformaldehyde in PBS for 15 min at room temperature. The fixed cells were permeabilized in 0.4% Triton X-100 in PBS at room temperature for 30 min and then washed twice with PBS. Permeabilized cells were incubated with RNase A (100 µg/ml) for 30 min at 37°C and then denatured in 3 N HCl solution at room temperature for 30 min followed by neutralization in 100 mM Tris-HCl, pH 8.5, for 30 min and were washed twice with PBS. Cells were blocked overnight at 4°C in blocking solution (1% BSA, 0.05% Tween-20 in PBS). Cells were incubated with anti-5hmC (1:3,000, rabbit

polyclonal; Active Motif; Carlsbad, CA) and anti-V5 antibodies (1:500, mouse monoclonal; Invitrogen) in blocking solution for 1 hour at room temperature with gentle shaking. Then, cells were washed several times in 0.05% Tween-20 in PBS (PBST) with gentle shaking, and incubated in secondary antibody mixture of Alexa Fluor 488 goat anti-mouse (1:1,000) and Alexa Fluor 568 goat anti-rabbit (1:1,000, Invitrogen) in blocking solution at room temperature for 1 hour. The cells were washed several times with PBST prior to mounting on slides with ProLong Gold anti-fade reagent with DAPI (Invitrogen). Fluorescence images were acquired using an Olympus IX81 fluorescence microscope with Image Pro Plus version 7.0 software (Media Cybernetics Inc.).

Dot-blot analysis

Genomic DNAs from the cultured cells were prepared by using Quick-gDNA MiniPrep kit (Zymo Research; Irvine, CA) followed by RNase A treatment. The gDNAs were briefly sonicated and purified using QIAquick PCR purification kit (Qiagen). The gDNAs were serially diluted twofold and denatured in TE buffer for 10 min at 98°C, immediately chilled on ice for 10 min, and then spotted onto a wetted nylon membrane (GeneScreen plus hydridization membrane; Perkin Elmer) using a Bio-Dot apparatus (96-well; Bio-Rad). The blotted membrane was dried on 3M paper and then UV cross-linked. The membrane was blocked with 5% non-fat milk in 0.15% Tween-20/PBS (PBST) for 2 h at room temperature with gentle shaking, and the membrane was incubated with rabbit polyclonal anti-5hmC antibody (Active Motif, 1:8,000) for 1 hour at room temperature. After five times washing with PBST, the membrane was incubated with peroxidase-conjugated anti-rabbit IgG (Active Motif, 1:15,000) secondary antibody for 1 hour at room temperature. After the washing step, the signal was visualized by using ECL Prime detection reagent (GE Healthcare).

Western blot analysis

To detect the expression levels of transiently expressed Tet3 isoforms in HEK293 cells, the cells were harvested 48 hours after transfection. Whole-cell lysates from HEK293 cells were prepared by lysis in buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton X-100, 1 mM EDTA and proteinase inhibitor cocktail (Roche) on ice for 30 min and centrifugation at 10,000xg for 10 min at 4°C. Nuclear extracts from cells or mouse tissues were prepared by using NE-PER Nuclear and Cytoplasmic Extraction kit according to the manufacturer's instructions (Life Technologies). The lysates or extracts were separated by 4%-15% SDS-polyacrylamide gels and blotted onto PVDF membranes (Bio-Rad) by wet transfer overnight at 4°C. The membranes were blocked with blocking buffer (5% non-fat milk, 0.05% Tween-20 in PBS) for 2 hours at room temperature with gentle shaking, and the membranes were then incubated with rabbit anti-mouse Tet3 CXXC domain antibody (in house; 1:15,000), mouse monoclonal anti-V5 (Invitrogen, 1:5,000), anti-Tet3 (Active Motif, 1:5,000; No**:** 61395), anti-Gapdh (Genetex, 1:10,000), anti-Hdac1 (Abcam, 1:2,000), or anti-lamin B (Santa Cruz, C-5; 1:1,000) antibodies for 1 hour at room temperature. After washing with PBST, the membranes were incubated with peroxidase-conjugated anti-rabbit IgG or anti-mouse IgG (Active Motif, 1:15,000) secondary antibodies for 1 h at room temperature. Bands were detected using ECL Prime detection reagent (GE Healthcare).

Luciferase reporter assays

We used pM1-Luc (Roche) as the luciferase reporter vector, which contains the firefly luciferase gene driven by a CMV promoter. pM1-Luc was methylated *in vitro* with M.SssI DNA methylase (New England Biolabs) according to the manufacturer's instruction. The methylated pM1-Luc plasmid was then phenol/chloroform extracted, ethanol-precipitated and resuspended in TE buffer. The completeness of methylation was confirmed by digestion with the methylation-sensitive restriction enzyme HpaII. Unmethylated plasmid was used as a control.

HEK293 cells were seeded into poly-L-lysine (Sigma) coated 24-well plates at 2.5 x 10^5 cells per well one day before transfection. A total of 250 ng of plasmid DNAs including 200 ng of mouse Tet3 isoform mammalian expression vectors, 47.5 ng of in vitro methylated pM1-Luc vector and 2.5 ng of internal control *Renilla* luciferase reporter vector (pRL-CMV, Promega, Madison, WI) were transfected into cells. The cells were harvested 48 h after transfection. For vitamin C treatment, vitamin C was added to the culture medium at 24 h after transfection, and cells were harvested after an additional 24 h of incubation. All transfections were carried out at least in three independent experiments and in triplicate. Firefly and *Renilla* luciferase activities were assayed with the dual luciferase assay kit (Promega) according to the manufacturer's instructions. The firefly luciferase activities were normalized relative to *Renilla*

luciferase expression and then the ratios of firefly/*Renilla* activity were presented relative to control sample (empty vector).

Decarboxylation assay

To detect any 5caC decarboxylation activity of Tet3, nuclear extracts prepared from HEK293 cells overexpressing Tet3 isoforms or the recombinant Tet3 CXXC domain were incubated with ^{32}P-isotope or biotin-labeled 64 bp oligonucleotides, respectively. Potential decarboxylation was evaluated by cleavage with HpaII (NEB), a methylation- and carboxylation-sensitive restriction enzyme, and then visualized by autoradiography or by using the LightShift Chemiluminescent EMSA kit (Pierce).

Base excision assay

The base excision assay was performed as described previously with slight modifications (Cortazar et al., 2011). Recombinant human TDG and/or the CXXC domain of human TET3 proteins were incubated with biotin-labeled hemi-modified 64 bp oligonucleotides (hemi-5caC64) in reaction buffer (50 mM Hepes pH 7.9, 1 mM DTT, 1 mM EDTA, and 1 mg/ml BSA) at 37°C for 2 hours. Resulting AP-sites were cleaved by the addition of 100 mM NaOH and 10 mM EDTA (final concentrations), and heating at 100°C for 5 min. Reaction products were separated by electrophoresis on 12% denaturing polyacrylamide gels in 1x TBE buffer and then visualized by using the LightShift Chemiluminescent EMSA kit (Pierce).

Structure determination

The sample for crystallizing the mTet3 CXXC domain - DNA complex was prepared by mixing mTet3-CXXC with a palindromic 12-mer DNA duplex, containing a central 5-carboxylated CpG dinucleotide (strand sequence: GAATCcaCGGATTC, caC = 5-carboxyldeoxycytidine), in a 1:1 molar ratio. Crystals of the mTet3-CXXC - DNA complex were generated from drops mixed from 1 µl of mTet3-CXXC - DNA solution and 1 µl of precipitant solution (0.1 M Hepes, pH 7.5, 50 mM calcium chloride, 41% PEG200). The crystals were flash frozen in liquid nitrogen before data collection.

The X-ray diffraction data set for mTet3-CXXC in complex with 5-carboxylated CpG DNA was collected on the BL 5.0.1 beamline at the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory. The diffraction data were indexed, integrated and scaled using the HKL 2000 program. The structure of the mTet3-CXXC - DNA complex was solved using the molecular replacement method in PHASER (McCoy et al., 2007) using the structure of the XtTet3-CXXC - DNA complex (PDB 4HP1) as a search model. There is one molecule of the mTet3-CXXC - DNA complex in the asymmetric unit (ASU). Further modeling of the mTet3-CXXC - DNA complex was carried out using COOT (Emsley and Cowtan, 2004) and was then subject to refinement using PHENIX (Adams et al., 2002). The same R-free test set was used throughout the refinement. The B-factors were refined with individual B values. The statistics for data collection and structural refinement for the mTet3-CXXC - DNA complex are summarized in Table S2.

ITC measurements

A Trp residue was introduced to the C-termini of mTet3-CXXC (residues 51-97) and XtTet3-CXXC (residues 58-111), respectively, for the purpose of protein concentration measurement using UV spectroscopy. The 12-mer CCGG (strand sequence: GAATCCGGATTC), CcaCGG (strand sequence: GAATCcaCGGATTC, caC = 5-carboxyldeoxycytidine), CmCGG (strand sequence: GAATCmCGGATTC, mC = 5-methyldeoxycytidine), CfCGG (strand sequence: GAATCfCGGATTC, fC = 5-formyldeoxycytidine) and ChmCGG DNA (strand sequence: GAATChmCGGATTC, hmC = 5-hydroxymethyldeoxycytidine) duplexes were synthesized by the Keck Oligonucleotide Synthesis facility at Yale University. Protein and DNA samples used for ITC measurements were subjected to overnight dialysis against buffer containing 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 50 µM $ZnCl_2$. Before the measurement, the protein and DNA concentrations were adjusted to about 0.04-0.1 mM and 0.3-0.5 mM, respectively. The ITC experiments were carried out using a MicroCal iTC200 instrument at 25°C. The titration curves were analyzed using software Origin7.0 (MicroCal, iTC200).

ChIP-seq

$1x10^7$ of mESC-derived neural progenitor cells (2 days; *see* Fig. S1E) or $1x10^7$ cells from E15.5 mouse embryo brain were used per ChIP (two biological replicates each). Samples were cross-linked with freshly prepared cross-linking solution (1% formaldehyde, 50 mM Hepes, pH 7.5, 1 mM EDTA, 0.5 mM EGTA,

and 100 mM NaCl) for 8 min at room temperature with shaking, followed by additional incubation for 5 min in 125 mM glycine. Cells were washed twice with PBS and pelleted by centrifugation at 3500xg for 5 min. After isolation of nuclei, the nuclear pellet was re-suspended in 1 ml of cold shearing buffer (10 mM Tris-HCl pH 8.1, 1 mM EDTA, and 0.1% SDS) and chromatin was then sheared to 100-300 bp fragments using a Covaris S220 sonicator (Covaris; Woburn, MA) with the following settings; 12.5 min, 5% duty cycle, 140 Watts peak incident power, 200 cycles per burst. After adding 1% Triton X-100 and 150 mM NaCl (final concentrations) to the sheared chromatin and mixing, the solution was cleared by centrifugation at 14,000xg for 10 min at 4°C. Chromatin fragments were incubated with 2.5 µg of affinity purified anti-Tet3 CXXC domain antibody (raised against the N-terminal region of Tet3FL, residues 1-102) or IgG control (Santa Cruz; SC-2027) overnight at 4°C with rotation, and then immunoprecipitated using protein A magnetic beads (Millipore; Billerica, MA). The libraries for Illumina sequencing of ChIP DNA were prepared using TruSeq Sample Prep kit (Illumina; San Diego, CA), and then samples were sequenced on an Illumina HiSeq instrument.

Data analysis

Illumina's Casava pipeline v1.8.2 was used to perform the base calling to extract sequences in Fastq format. Data were deposited into the GEO database (accession number GSE56821). Sequences were then aligned to the mm9 genome assembly using Novoalign (http://www.novocraft.com) with default settings. Only reads that were aligned to a unique genome location were kept for further analysis. We obtained between 56 and 64 million uniquely aligned reads for the brain samples and between 91 and 116 million uniquely aligned reads for the NPC samples. Peaks for Tet3FL were identified using IgG as control with in-house developed R scripts. Briefly, regions with coverage >= 8 were identified and merged into peaks if they were less than 250 bp apart. These peaks were further filtered to select the highly enriched ones with Tet3FL/IgG > 2.5-fold for the brain peaks and Tet3FL/input > 2.5-fold for the NPC peaks. The filtered peaks were then annotated to RefSeq databases into "TSS", "Extended promoter", "Intragenic", and "Intergenic" regions as described before (Rauch et al., 2009).

Sequence motifs that were enriched in Tet3FL TSS peaks were identified using the peak-motifs tools available on the RSAT Web server (http://www.rsat.eu/) (Thomas-Chollier et al., 2012).

ENCODE-derived ChIP-seq peaks for H3K4me3 and H3K27me3 located at the TSS in mouse E14.5 brain were downloaded from GEO data sets GSM1000095 and GSM1000143, respectively. Gene ontology analysis was performed using the Genomic Regions Enrichment Annotations tool (GREAT) (McLean et al., 2010).

The bisulfite DNA methylation data was obtained from Stadler et al (Stadler et al., 2011). This data is for NPC after 8 days EB formation. Tet3FL ChIP data was obtained after an additional 2 days on PORN/laminin coated dishes in N2 medium, which is designed to differentiate the NPCs into neurons. As described previously (Mohn et al., 2008), several hundred promoters become methylated when ES cells differentiate to NPC but the authors did not observe additional de novo methylation upon terminal differentiation.

RNA-seq data for E15.5 brain was used to determine differences in expression levels for genes with and without Tet3FL peaks at the TSS. For generation of histograms of Tet3FL distribution near the TSS, sequences near the TSS (+/- 1 kb on each side) were divided into 100 bp bins. If a Tet3FL peak overlapped more than one bin, it was counted only once at the peak center. Composite profiles were derived as described previously (Hahn et al., 2013).

SUPPLEMENTAL REFERENCES

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta crystallographica. Section D, Biological crystallography *58*, 1948-1954.

Bibel, M., Richter, J., Lacroix, E., and Barde, Y.A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. Nature protocols *2*, 1034-1043.

Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F.*, et al.* (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419-423.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta crystallographica. Section D, Biological crystallography *60*, 2126-2132.

Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.G., Jiang, Y., Pfeifer, G.P.*, et al.* (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. Cell Rep. *3*, 291-300.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. Journal of applied crystallography *40*, 658-674.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495-501.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol. Cell *30*, 755-766.

Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D., and Pfeifer, G.P. (2009). A human B cell methylome at 100-base pair resolution. Proc Natl Acad Sci U S A *106*, 671-678.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D.*, et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490-495.

Thomas-Chollier, M., Darbo, E., Herrmann, C., Defrance, M., Thieffry, D., and van Helden, J. (2012). A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs. Nat. Protoc. *7*, 1551-1568.