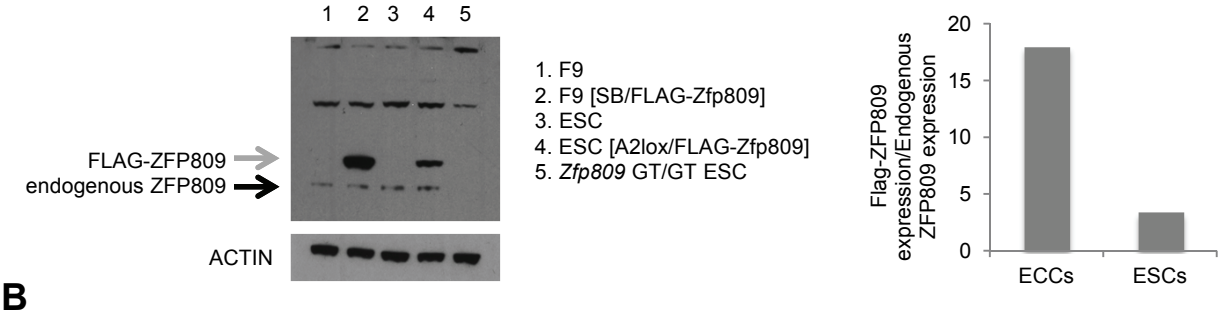
Supplemental Figure 1

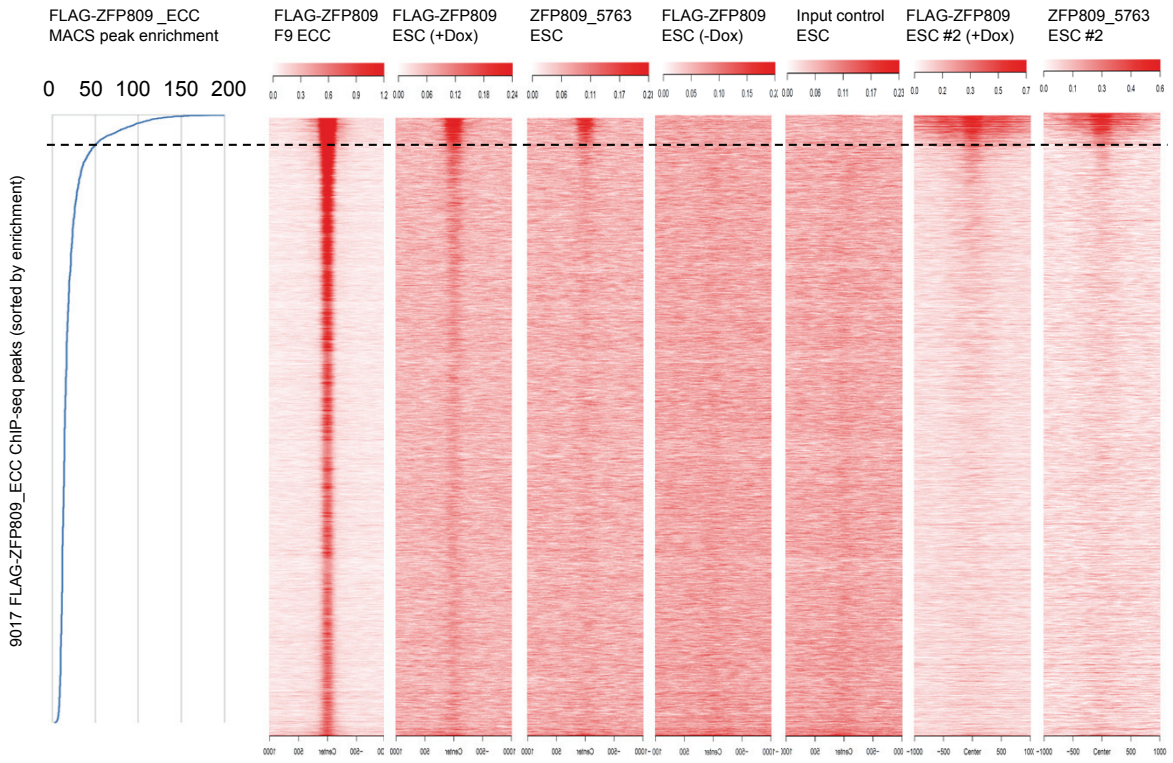
A



B

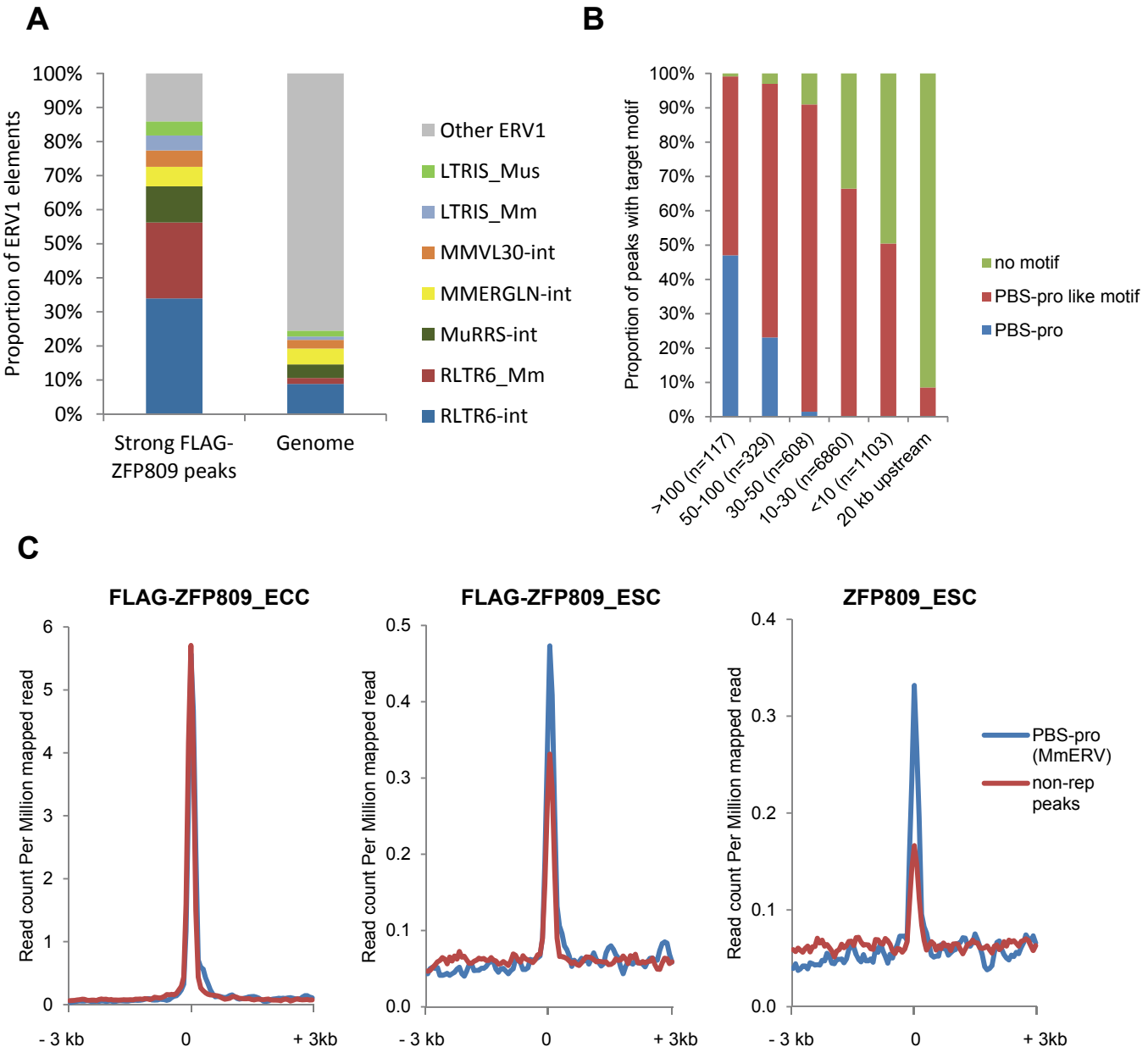
ChIP-seq	Cell type	Transgene	Antibody	ChIP protocol	Sequencing strategy	No. of called MACS peaks	Mean peak enrichment	Mean peak p-value	Mean peak FDR (%)
FLAG-ZFP809_ECC	F9 ECC	Multi copy SB transposon	FLAG	O'Geen et al. 2010	Illumina, single-end	9017	20.2	5.5E-14	0.6
FLAG-ZFP809_ESC	ESC	A2lox single copy insertion	FLAG	Barish et al. 2010	Illumina, single-end	1224	13.8	2.9E-07	85.3
ZFP809_ESC	ESC	-	ZFP809 (5763)	Barish et al. 2010	Illumina, single-end	1115	13.6	4.5E-07	69.3
FLAG-ZFP809_ESC #2	ESC	A2lox single copy insertion	FLAG	O'Geen et al. 2010	SOLiD, paired-end	n.d.	n.d.	n.d.	n.d.
ZFP809_ESC #2	ESC	-	ZFP809 (5763)	O'Geen et al. 2010	SOLiD, paired-end	n.d	n.d.	n.d	n.d

C



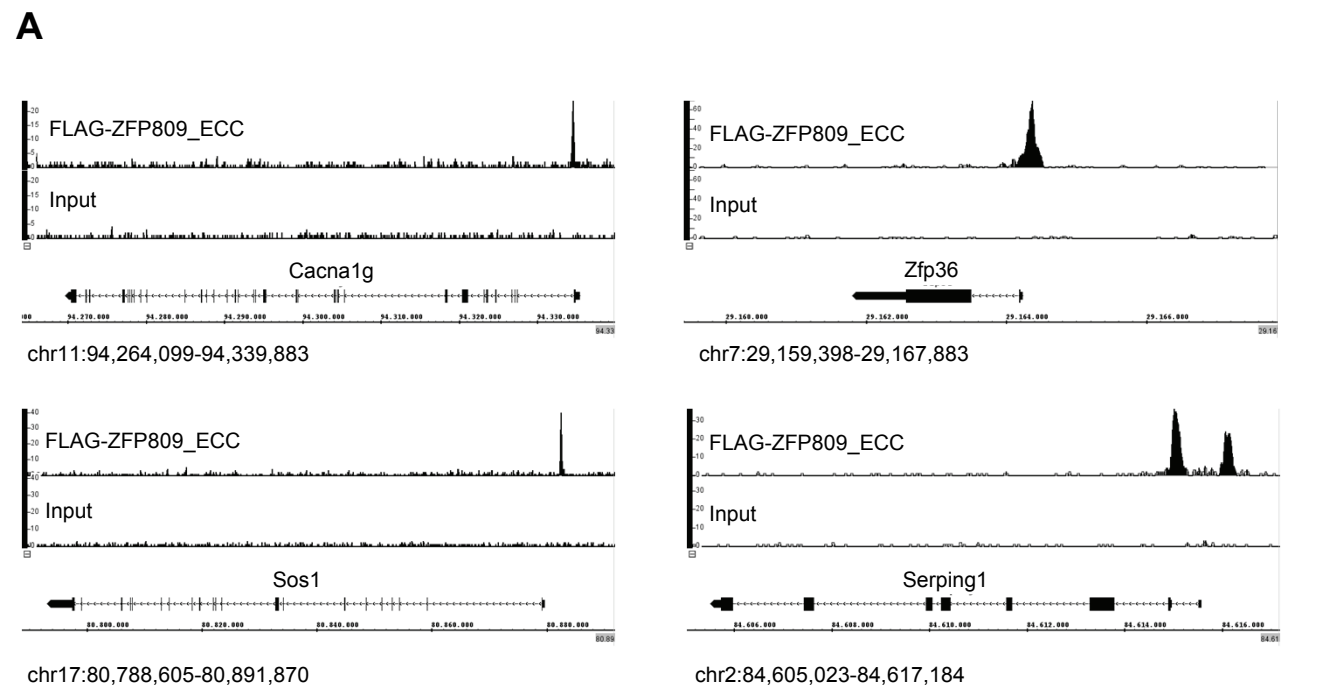
Supplemental Figure 1. (*A*) (*left panel*) Western blot was performed with whole cell lysates from FLAG-ZFP809 expressing F9 ECCs and ESCs that were used for ChIP-seq analysis. Endogenous ZFP809 and FLAG-ZFP809 were detected using a polyclonal ZFP809 antibody (kindly provided by Stephen Goff). (*right panel)* Quantification of FLAG-ZFP809 overexpression in F9 ECCs and ESCs using ImageJ. (*B*) Summary of ZFP809 ChIP-seq strategies. Peaks were called by Model-based Analysis of ChIP-Seq (MACS). Peak calling was not performed with FLAG-ZFP809_ESC #2 and ZFP809_ESC #2 data due to the lack of proper input control data. n.d.: not determined. (*C*) Heat maps with mapped ChIP-seq reads generated by different ChIP-seq strategies. Peaks were called using Model-based Analysis of ChIP-Seq (MACS) with FLAG-ZFP809_ECC data and used as reference sites (extending 1 kb up- and downstream of the peak summit). The dotted line indicates the 50-fold enrichment value that was set as threshold for strong FLAG-ZFP809_ECC peaks. ChIP-seq with FLAG antibody in uninduced A2lox [FLAG-Zfp809] ESCs (not expressing FLAG-ZFP809) was performed as negative control.

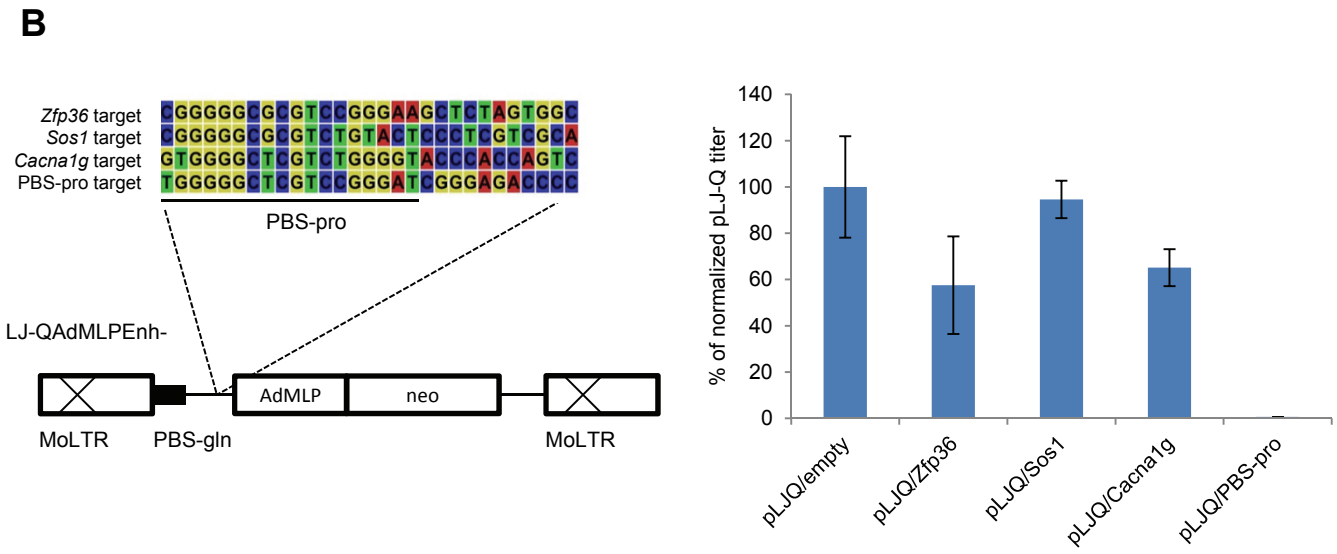
Supplemental Figure 2



Supplemental Figure 2. Analysis of FLAG-ZFP809 ChIP-seq peaks. (*A*) Categorization of ERV1 elements overlapping with strong (enrichment > 50) FLAG-ZFP809 ChIP-seq peaks in F9 ECCs compared to all genomic ERV1 elements in the mouse genome. ERV1 elements were categorized according to their UCSC Repeatmasker annotation. The suffix "-int" indicates that the repeat is an internal ERV1 sequence, all other repeats are LTR sequences. (*B*) Proportion of ChIP-seq peaks that contain a perfect PBS-pro site or a PBS-pro-like target motif (derived from the 100 ttop-scored non-repetitive FLAGZFP809_ECC peaks). The 200 bp core regions of all FLAG-ZFP809 peaks or control regions 20 kb upstream of the ChIP-seq peaks were screened for target motifs under stringency settings allowing five false hits per 10 kb. Peaks were categorized according to their enrichment over the input control. (*C*) NGS plots showing average FLAG-ZFP809_ECC, FLAG-ZFP809_ESC and and ZFP809_ESC ChIP-seq read densities at 6 kb regions around genomic MmERV PBS-pro (n=93) and strong non-repetitive FLAG-ZFP809_ECC ChIP-seq peaks (n=138).

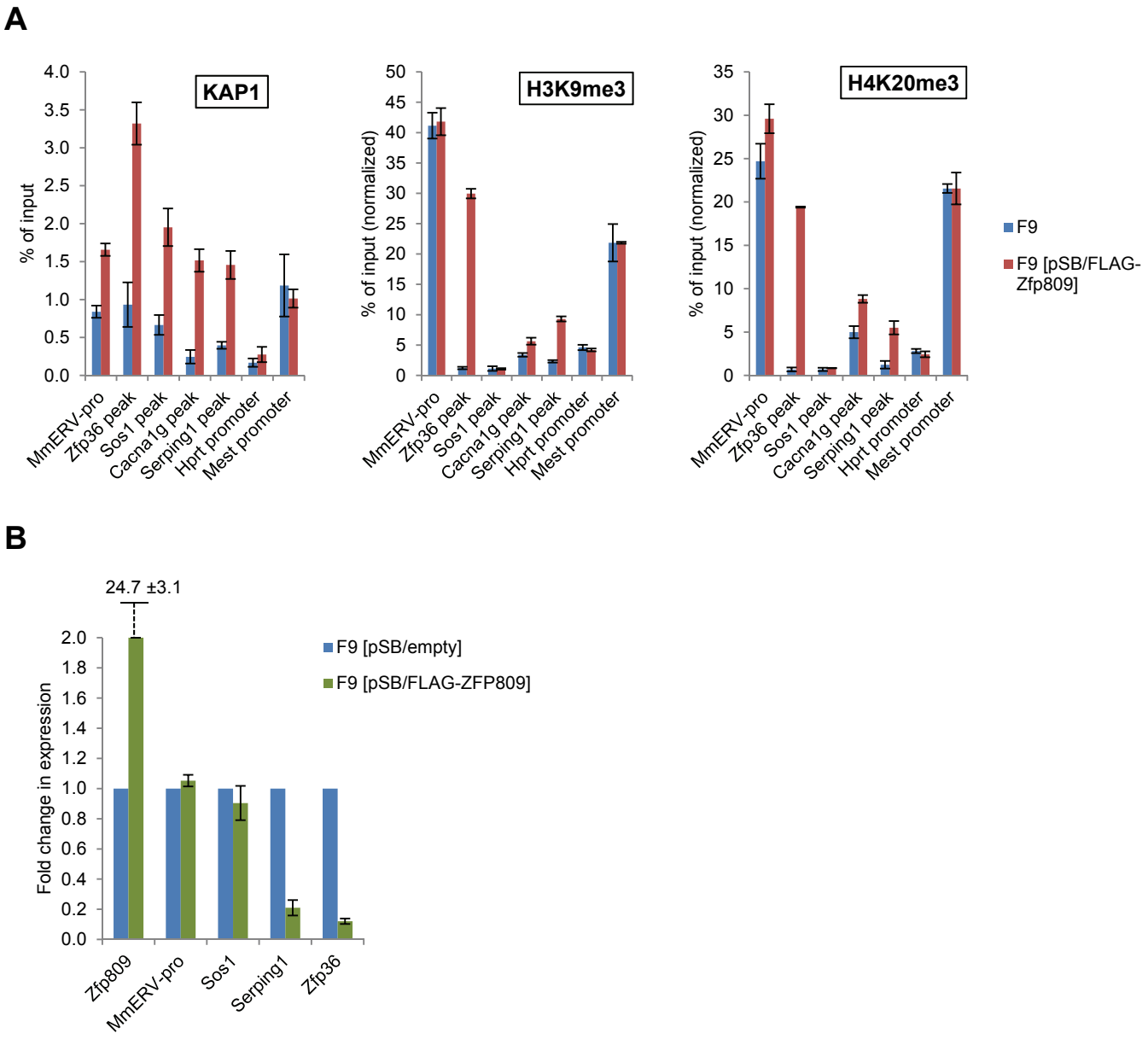
Supplemental Figure 3





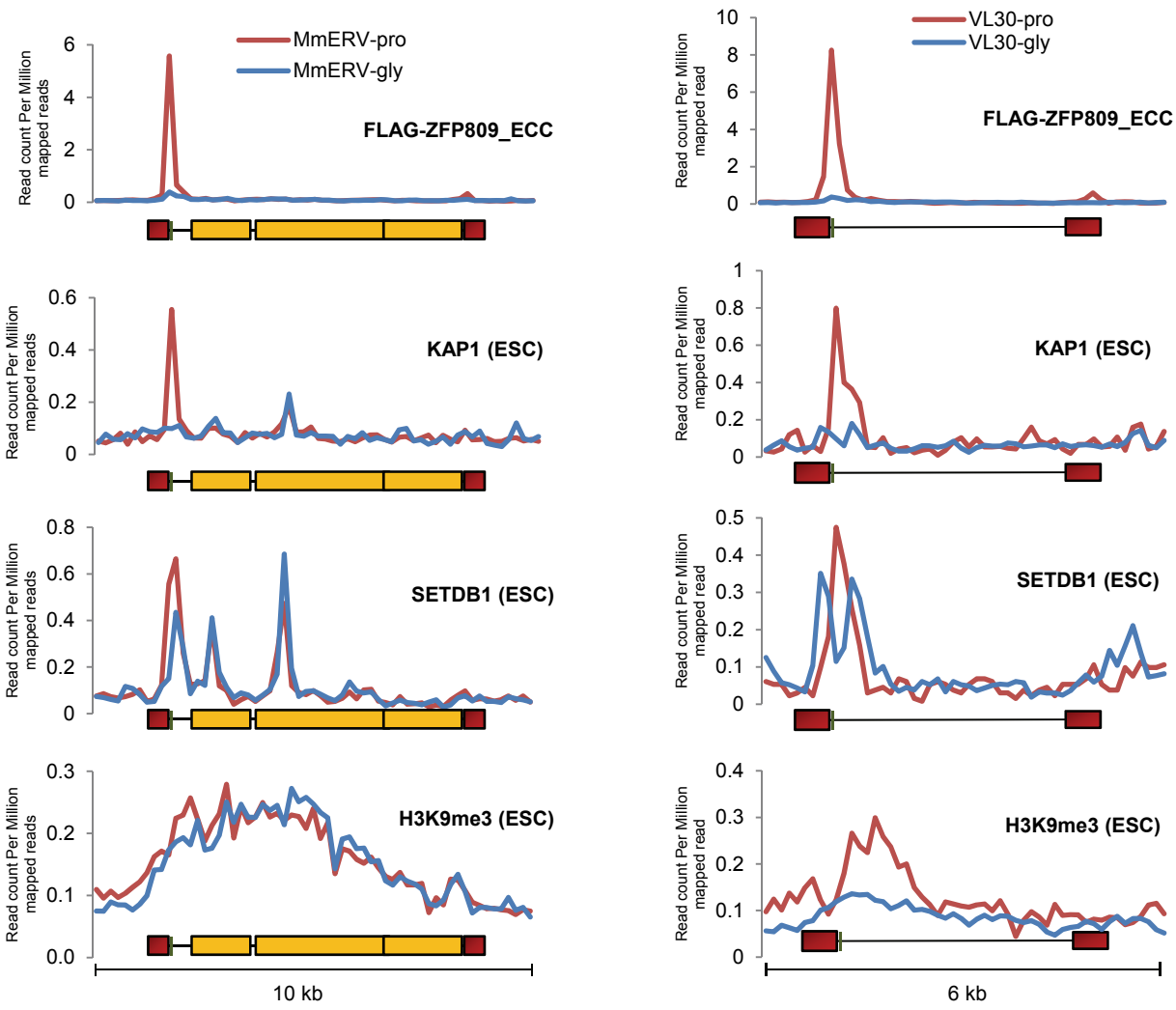
Supplemental Figure 3. Non-repetitive ZFP809 target sites do not induce efficient silencing of retroviral vectors. (*A*) Integrative Genome Browser (IGV) view of a selection of strong FLAG-ZFP809_ECC peaks near gene promoters containing PBS-pro-like target motifs with mapped FLAG-ZFP809_ECC ChIP-seq and Input control read densities. (*B*) (*left panel*) 29 bp oligos, containing PBS-pro like ZFP809 target motifs identified in ChIP-seq peaks close to gene promoters including 11 bp downstream flanking regions, or the PBS-pro sequence including the 11 bp downstream region of the Moloney Murine Leukemia Virus (MLV) PBS-pro were inserted upstream of the AdMLP promoter of the retroviral LJQAdMLPEnh- vector (Modin C, Pedersen FS, Duch M. 2000. Lack of shielding of primer binding site silencer-mediated repression of an internal promoter in a retrovirus vector by the putative insulators scs, BEAD-1, and HS4. *J Virol* **74**: 11697-11707). (*right panel*) F9 ECCs and NIH/3T3 fibroblasts were transduced in parallel with serial dilutions of retroviral particles containing the indicated ZFP809 target sequences. Viral titers (CFU/ml) for each vector were determined by counting resistant colonies after G418 selection. Viral titers in F9 cells were normalized to titers in NIH/3T3 cells and are shown relative to the normalized titer of a vector containing the empty LJQAdMLPEnh- vector. Graphs show mean normalized titers as percent relative to the empty control vector (n=2, technical replicates). Error bars represent s.d.

Supplemental Figure 4



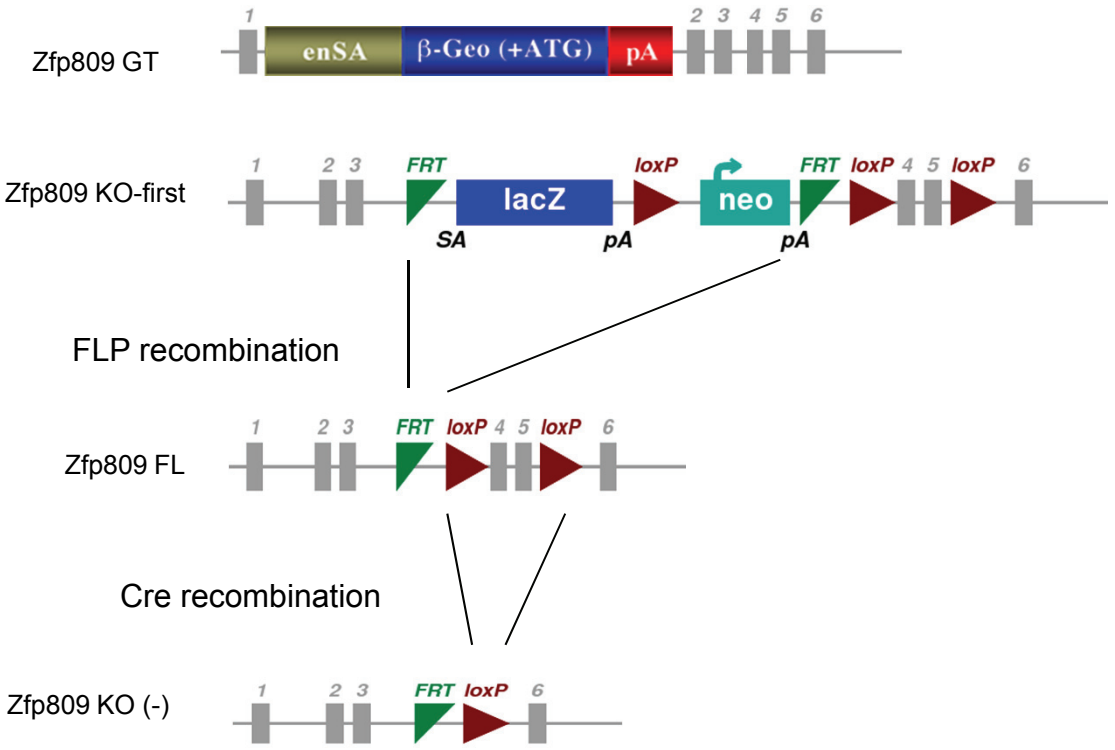
Supplemental Figure 4. Heterochromatin formation and transcriptional silencing of FLAG-ZFP809 target genes upon FLAG-ZFP809 overexpression. (*A*) ChIP-qPCR analysis of FLAG-ZFP809 binding sites in FLAG-ZFP809 overexpressing F9 cells. Cross-linked chromatin of F9 cells and F9 cells stably expressing FLAG-ZFP809 was immunoprecipitated with antibodies against KAP1, H3K9me3 (Abcam, ab8898) or H4K20me3 (Abcam, ab9053). Enrichment at target regions was determined by qPCR. Graphs show mean enrichment over input (n=3, technical replicates). Error bars represent s.d. H3K9me3 and H4K20me3 ChIP-qPCR data were normalized to equal enrichment at the *Mest* promoter, a known KAP1 binding site that is not enriched in FLAG-ZFP809. (*B*) RT-qPCR expression analysis of FLAG-ZFP809 target genes in FLAG-ZFP809 overexpressing F9 cells. Graphs show mean fold enrichment (n=3, technical replicates) over expression in F9 cells (normalized to ActinB) stably transfected with an empty control vector. Error bars represent s.d.

Supplemental Figure 5



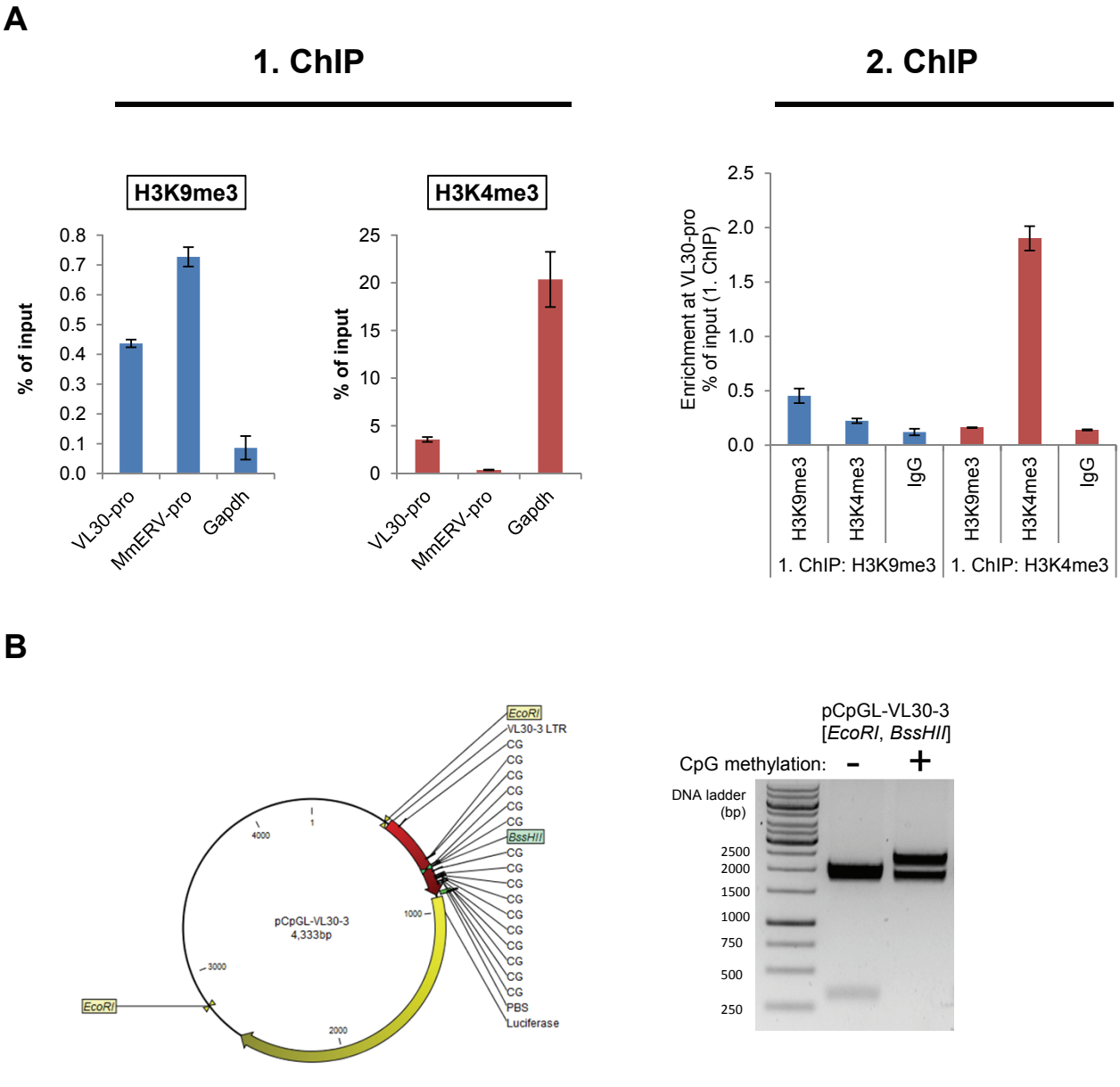
Supplemental Figure 5. Differential repressor binding patterns at MmERV and VL30 subgroups using different PBS sequences. NGS plots show average ChIP-seq read densities at 10 kb or 6 kb regions around genomic MmERV-pro (n=93) and MmERV-gly (n=69) (*left* panel), and VL30-pro (n=15) and VL30-gly (n=65) elements (*right* panel). Yellow rectangles indicate coding regions for retroviral proteins, red rectangles indicate LTRs. In addition to FLAG-ZFP809 ChIP-seq data from F9 ECCs, previously published KAP1 (SRA accession: SRR611529), SETDB1 (SRA accession: SRR031683) and H3K9me3 (SRA accession: SRR611521) ChIP-seq data (all generated in ESCs) were re-aligned to the mouse genome as described in the methods section.

Supplemental Figure 6



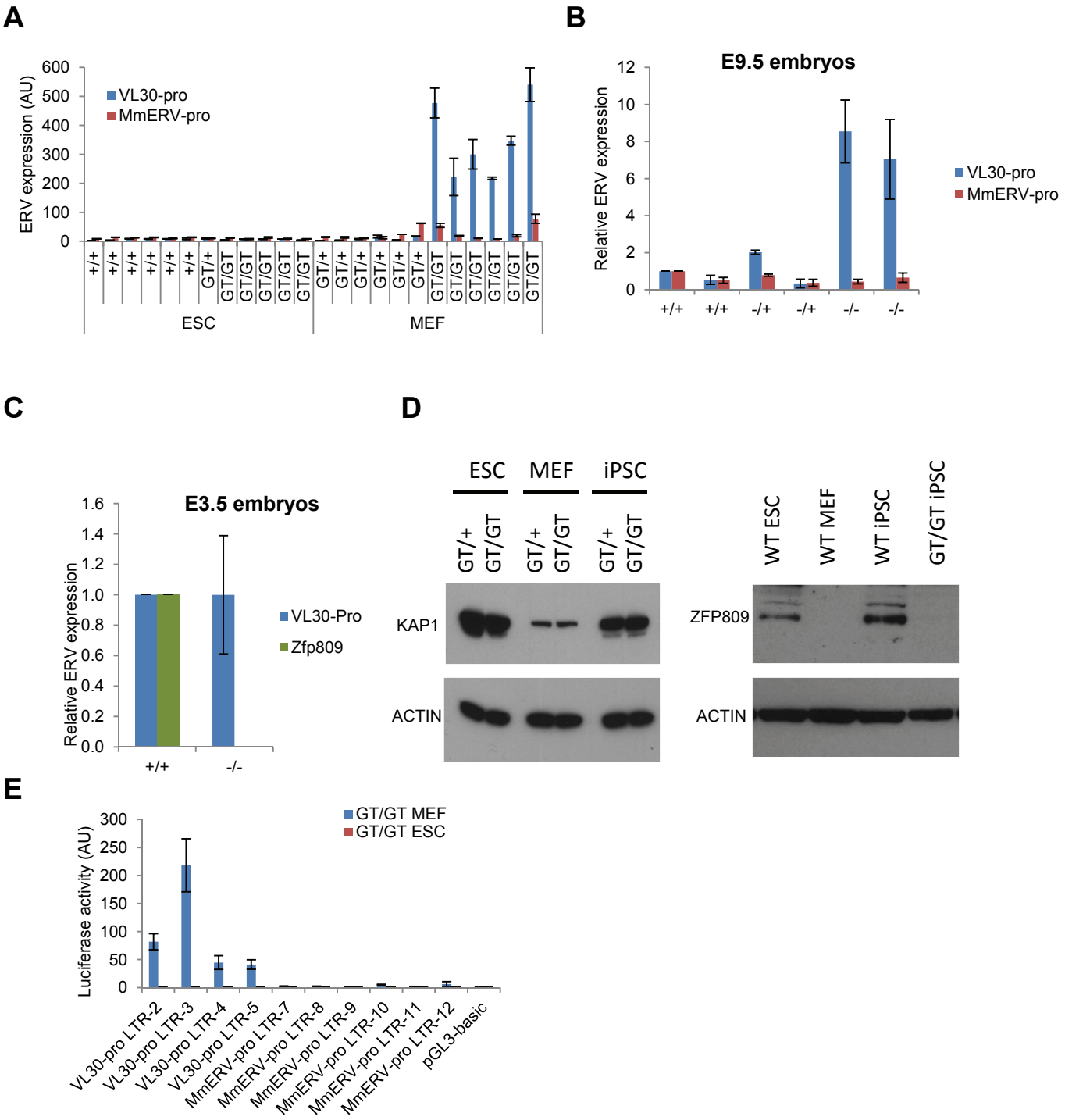
Supplemental Figure 6. Schematic representation of Zfp809 gene-trap (GT) and knock-out-first (KO-first) alleles used to generate mutant mice. Grey rectangles represent Zfp809 exons. The two recombination steps in which the KO-first allele turns into a KO (-) allele are shown.

Supplemental Figure 7



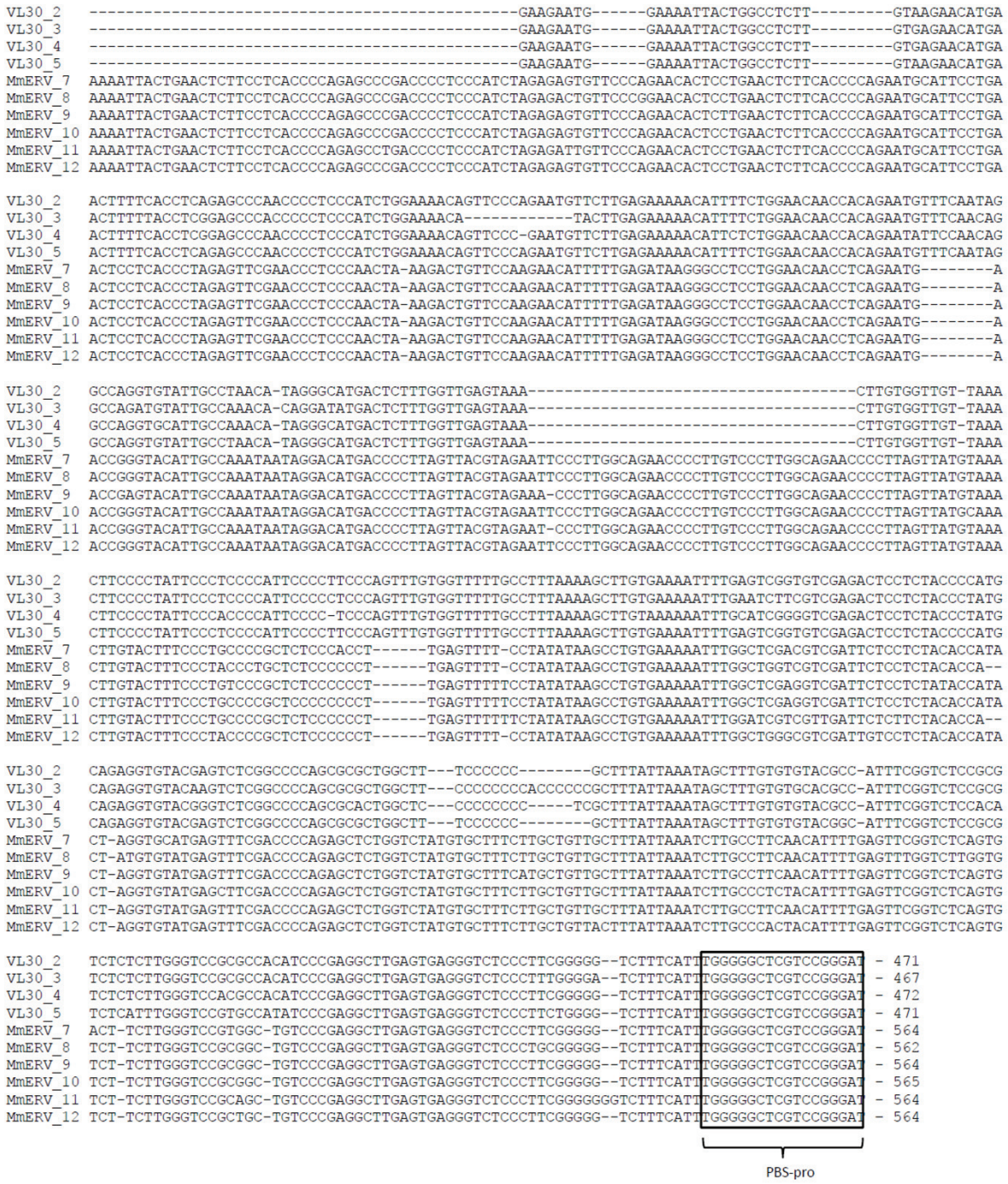
Supplemental Figure 7. (*A*) (*left* panel) H3K9me3 and H3K4me3 enrichment at ERVs and the Gapdh promoter in *Zfp809* GT/GT MEFs after one round of ChIP. (*right panel*) Re-ChIP analysis of H3K9me3 and H3K4me3 at VL30-pro. Chromatin immunoprecipitated with H3K9me3 or H3K4me3 specific antibodies was immunoprecipitated a second time with the indicated antibodies. Enrichment is shown as mean percentage of the corresponding input samples (n=3, technical replicates). Error bars represent s.d. (*B*) *In vitro* methylation of CpG-free luciferase vectors. (*left*) Vector map of the pCgGL plasmid with an inserted VL30-pro LTR element (including PBS). CpG sites and restriction sites for *EcoRI and BssHII* are indicated. The *BssHII* site overlaps with two CpG sites and is inhibited by CpG methylation. (*right*) Unmethylated and CpG methylated plasmids were digested with *EcoRI and BssHII* and run on an agarose gel to visualize fragment lengths. The two larger fragments of the digested unmethylated plasmid are both about 2 kb long and therefore appear as a single bond.

Supplemental Figure 8



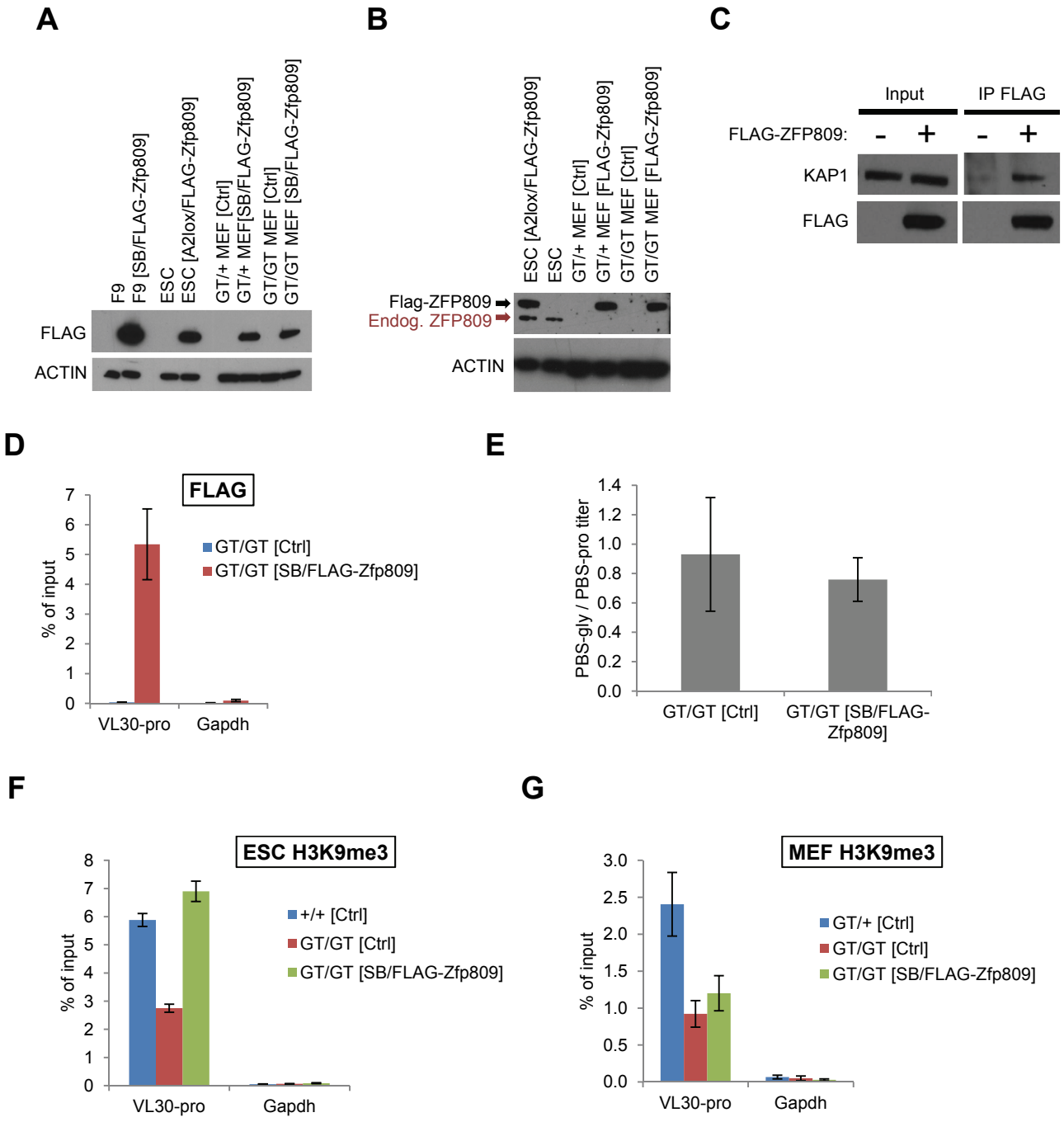
Supplemental Figure 8. (*A*) RT-qPCR analysis of MmERV-pro and VL30-pro expression in *Zfp809* GT ESCs and MEFs. ERV expression was normalized to Hprt and is shown in arbitrary units (AU). Graphs show mean expression (n=3, technical replicates). Error bars represent s.d. (*B*) VL30-pro expression in *Zfp809* knock-out embryos. Two E9.5 littermate embryos of each genotype were analyzed. Values represent mean expression (normalized to Hprt) relative to one WT embryo (n=3, technical replicates). Error bars represent s.d. (*C*) Five to six preimplantation embryos per genotype were harvested at E3.5 and pooled for RNA purification and RT-qPCR analysis. Values represent mean expression (normalized to Gapdh) relative to *Zfp809* +/+ embryos (n=3, technical replicates). Error bars represent s.d. (*D*) (*left panel*) Western blot with whole cell lysates of *Zfp809* GT/+ and GT/GT ESCs, MEFs and iPSCs using antibodies against KAP1 or ACTIN. (*right panel*) Western blot with whole cell lysates of WT ESCs, MEFs and iPSCs using antibodies against ZFP809 (5763) or ACTIN. Cell lysates of *Zfp809* GT/GT iPSCs were blotted to confirm the identity of the ZFP809 band. (*E*) Promoter activity of subcloned VL30-pro and MmERV-pro LTR elements (including PBS-pro) in indicated cell types using dual luciferase assay in transiently transfected cells. Luciferase activity was normalized to Renilla Luciferase activity of a cotransfected plasmid. Assay was performed in triplicates. Error bars represent s.d. LTR sequences are shown in Supplemental Fig. 9.

Supplemental Figure 9



Supplemental Figure 9. Alignment of VL30-pro and MmERV-pro LTR sequences that have been tested for transcriptional activity (Supplemental Fig. 8E).

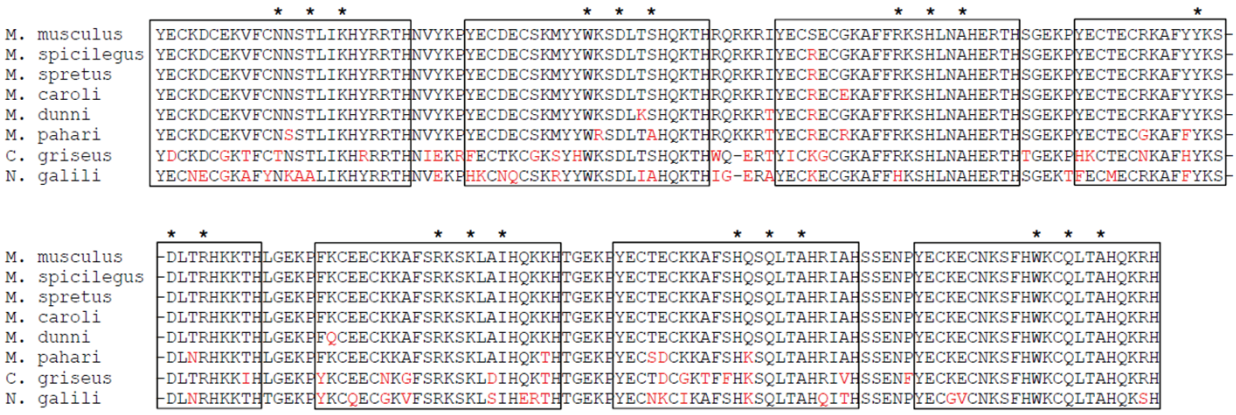
Supplemental Figure 10



Supplemental Figure 10. (*A*) Western blot with whole cell lysates of FLAG-ZFP809 overexpressing F9 and ESCs and FLAG-ZFP809 complemented MEFs using antibodies against the FLAG epitope or ACTIN. (*B*) Western blot with whole cell lysates of FLAG-ZFP809 overexpressing ESCs and FLAG-ZFP809 complemented MEFs using antibodies against endogenous ZFP809 (kindly provided by Stephen Goff) or ACTIN. (*C*) Whole cell lysates from *Zfp809* GT/GT MEFs stably transfected with SB/FLAG-Zfp809 (+) or an empty control vector (-) were immunoprecipitated with FLAG antibody. 10% input controls and FLAG immunoprecipitates were analyzed by western blot using anti-FLAG and anti-KAP1 antibodies. (*D*) ChIP-qPCR analysis of FLAG-ZFP809 recruitment to VL30-pro ERVs in FLAG-ZFP809 complemented *Zfp809* GT/GT MEFs. Enrichment is shown as mean percentage of input DNA (n=3, technical replicates). Error bars represent s.d. (*E*) Control and FLAG-ZFP809 complemented *Zfp809* GT/GT MEFs were transduced with retroviral vectors containing PBS-pro or the repression escape mutant PBS-gly upstream of a hygromycin resistance gene and treated with 200 µg/ml hygromycin B for ten days to determine viral titers (CFU/ml). Graphs show the mean ratios of PBS-gly/PBS-pro viral titers as determined in three replicates. Error bars represent s.d. (*F*) ChIP-qPCR analysis of H3K9me3 at VL30-pro elements in FLAG-ZFP809 complemented *Zfp809* GT/GT MEFs. Enrichment is shown as mean percentage of input DNA (n=3, technical replicates). Error bars represent s.d. (*G*) NChIP-qPCR analysis of H3K9me3 at VL30-pro elements in FLAG-ZFP809 complemented *Zfp809* GT/GT ESCs. Enrichment is shown as mean percentage of input DNA (n=3, technical replicates). Error bars represent s.d.

Supplemental Figure 11

	Zfp809 homology (%)		dS/dN	dS/dN (specificity aa)
	DNA	Protein		
M. spicilegus	98.3	99.5	34.9	-
M. spretus	98.8	99.5	23.2	-
M. caroli	97.9	99.0	19.4	-
M. dunni	97.2	97.9	11.6	-
M. pahari	94.6	92.7	4.4	3.6
C. griseus	81.9	80.2	4.3	21.5
N. galili	80.7	76.6	4.0	8.7



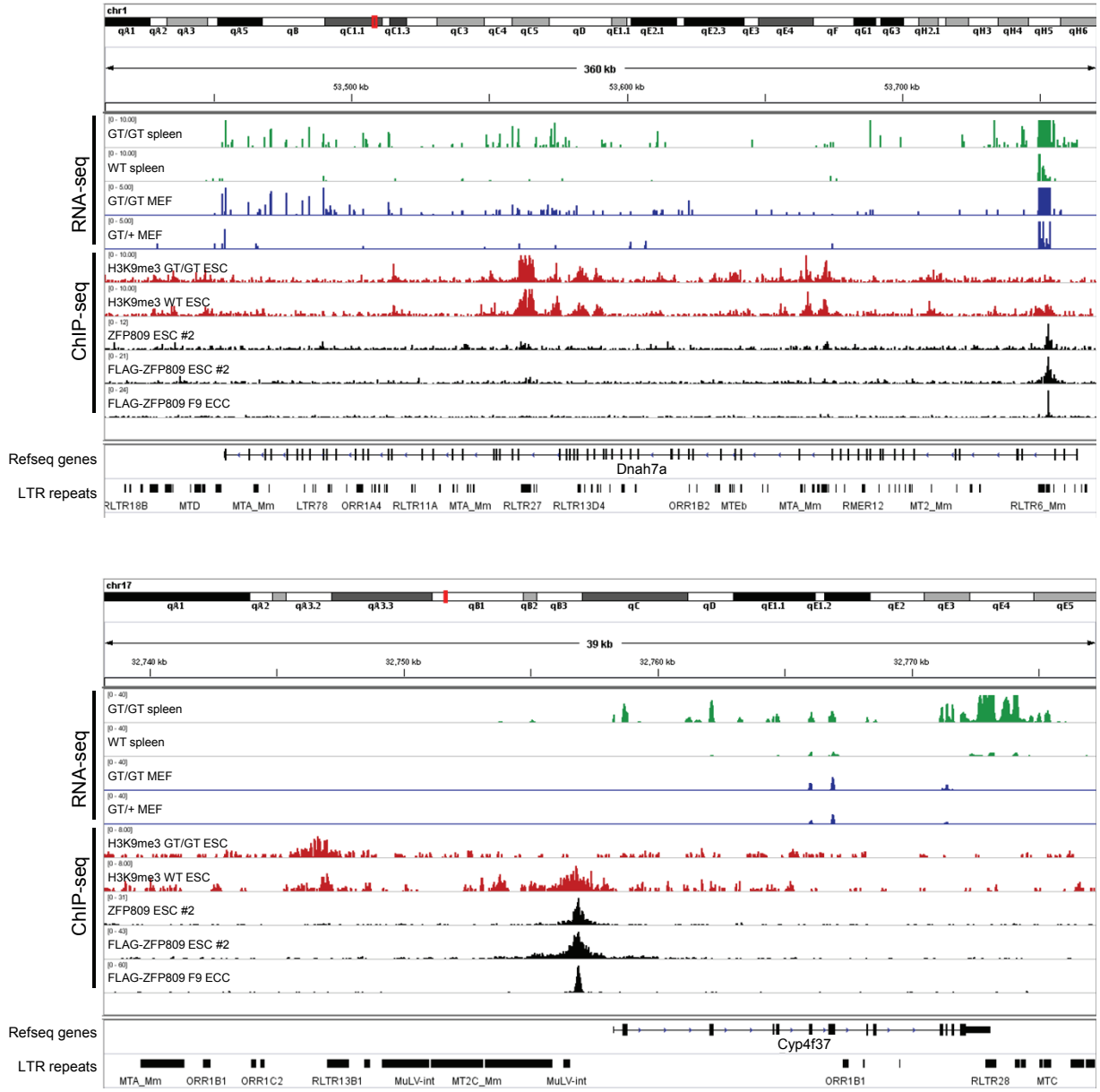
Supplemental Figure 11. Conservation of the DNA binding domain of putative ZFP809 orthologues. (*top*) Homology and synonymous to non-synonymous mutation ratios (dS/dN) of the 579 bp DNA binding region of *Zfp809* in wild *Mus* species relative to *M. musculus*. (*bottom*) Alignment of translated ZFP809 DNA binding domains identified in *Muroidea* species. The positions of the seven C2H2 zinc finger domains are marked by black rectangles. The amino acids known to be directly responsible for binding specificity (specificity aa) are marked by asterisks on top of the alignment. ZFP809 DNA binding domains of *Mus* species were sequenced from genomic DNA samples. The putative ZFP809 DNA binding sequences of *C. griseus* (NCBI accession: XP_003516040) and *N. galili* (GenBank: JO002639) were identified by BLAST searches. The JO002639 sequence contains a single nucleotide deletion in the fourth zinc finger domain, causing a frame shift mutation. The deletion, located after six consecutive adenine bases and possibly the result of a sequencing error was manually corrected to reconstruct the ORF.

Supplemental Figure 12

A

	Spleen						MEF					
Gene	base Mean	base GT/GT	log2Fold change	lfcSE	pvalue	padj	base mean	base GT/GT	log2Fold change	lfcSE	pvalue	padj
Zfp809	196.0	17.8	-3.46	0.24	6.0E-48	2.3E-44	125.4	20.3	-2.63	0.25	1.4E-26	7.0E-23
Tdrd9	173.5	6248.1	5.17	0.32	1.7E-57	9.8E-54	34.3	105.2	1.62	0.28	1.0E-08	2.6E-05
Gm9705	135.6	1041.7	2.94	0.27	1.4E-28	2.8E-25	25.8	31.0	0.26	0.28	3.5E-01	1.0E+00
Dnahc7a	59.1	584.5	3.30	0.37	1.8E-19	7.4E-17	30.5	85.4	1.48	0.28	1.4E-07	3.0E-04

B



Supplemental Figure 12. (*A*) Statistical analysis of differential gene expression in Zfp809 GT mice. RNA-seq data from *Zfp809* GT/+ and GT/GT MEFs or +/+ and GT/GT spleen samples (two independent MEF lines/mice per genotype) were analysed using DESeq (Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.). (*B*) RNA-seq and ChIP-seq read density at *Dnah7a* (*top*) and *Cyp4f37 (bottom)*.

Supplementary methods:

Re-ChIP

Chromatin was harvested from cross-linked Zfp809 GT/GT MEFs as described previously (Barish et al. 2010). Briefly, chromatin from 1.4×10^6 cells was immunoprecipitated with 1 µg H3K9me3 (abcam ab8898), 1 µg H3K4me3 (abcam ab8580). After washing, immunoprecipitated chromatin was eluted with Re-ChIP-IT elution buffer (Re-ChIP-IT kit, Active Motif), desalted, and immunoprecipitated for the second time with 1 µg H3K9me3 (abcam ab8898), 1 µg H3K4me3 (abcam ab8580) or 1 µg unspecific IgG according to the manufacturers protocol. Target enrichment was determined by qPCR.

Immunoprecipitation (IP)

Cells were lysed in RIPA buffer and whole cell extracts were diluted 1:10 in IP buffer (10 mM Tris-HCl [pH 7.5]; 1% NP40; 150 mM NaCl) and incubated with Flag Agarose beads (Sigma) over night at 4°C. Beads were washed 5 times with IP buffer before western blot analysis.