Developmental Cell, Volume *36*

Supplemental Information

Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues

Gabriela Ecco, Marco Cassano, Annamaria Kauzlaric, Julien Duc, Andrea Coluccio, Sandra Offner, Michaël Imbeault, Helen M. Rowe, Priscilla Turelli, and Didier Trono

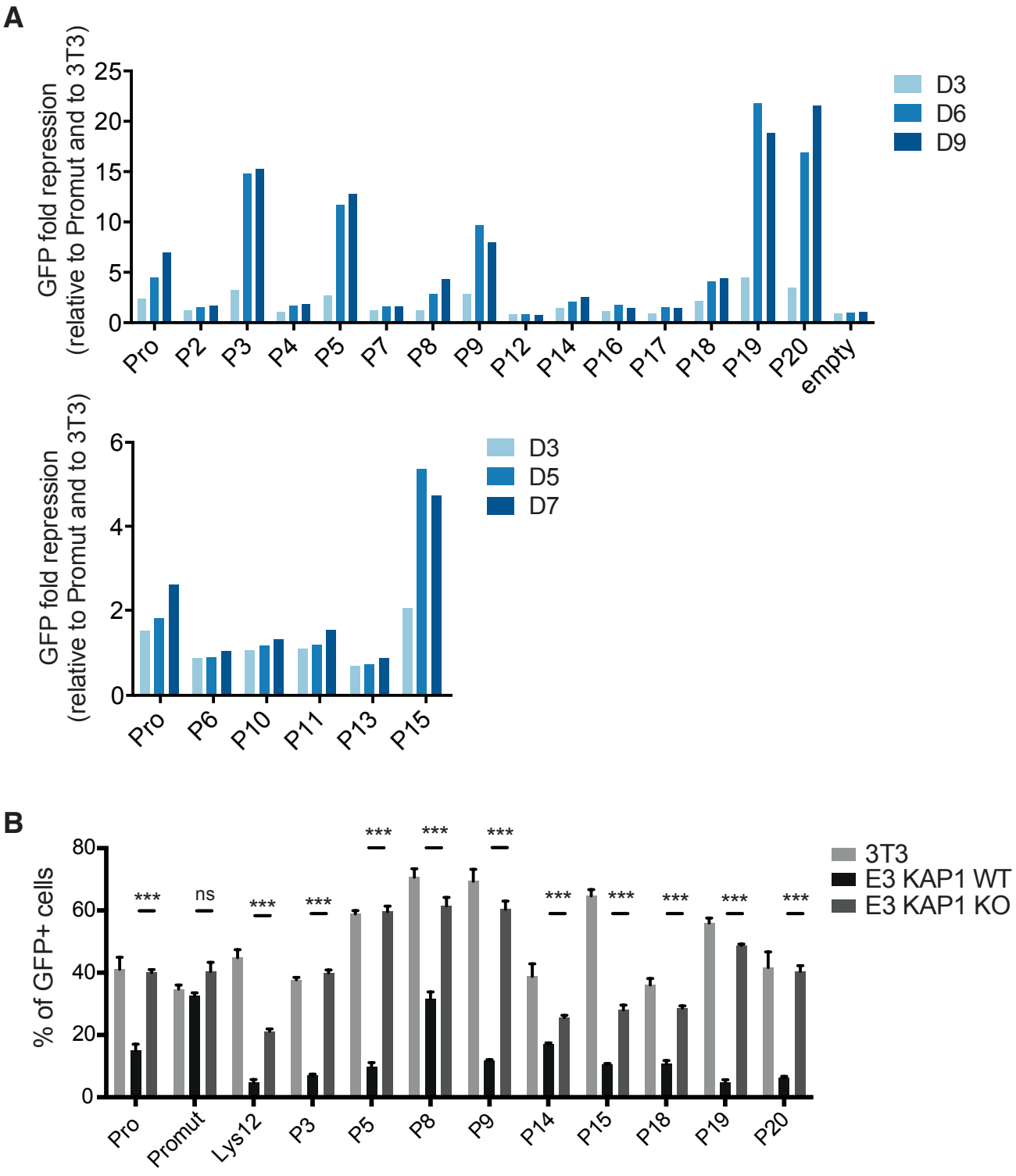

Figure S1, related to Figure 1. Ten of the DNA target sequences selected for the screen are repressed in murine ES in a KAP1-dependent manner.

(**A**) Repression assays in murine ES cells of selected DNA target sequences. ES or 3T3 cells were transduced with vectors containing the sequence of interest upstream a PGK-GFP cassette and percentage of GFP positive cells was measured by FACS at different time-points after transduction. Data are represented as GFP repression normalized to the negative control Promut and the corresponding 3T3 control. (**B**) GFP repression assay with strongly repressed target sequences in *Kap1* WT or KO mESC, using 3T3 cells as control (day 4 after transduction). Promut is a negative control, corresponding to a point mutant of PBS^{Pro} (also known as B2), which is not repressed in mESC. Error bars represent SD, *** $p \leq 0.001$, ns= not significant, Student's t test.

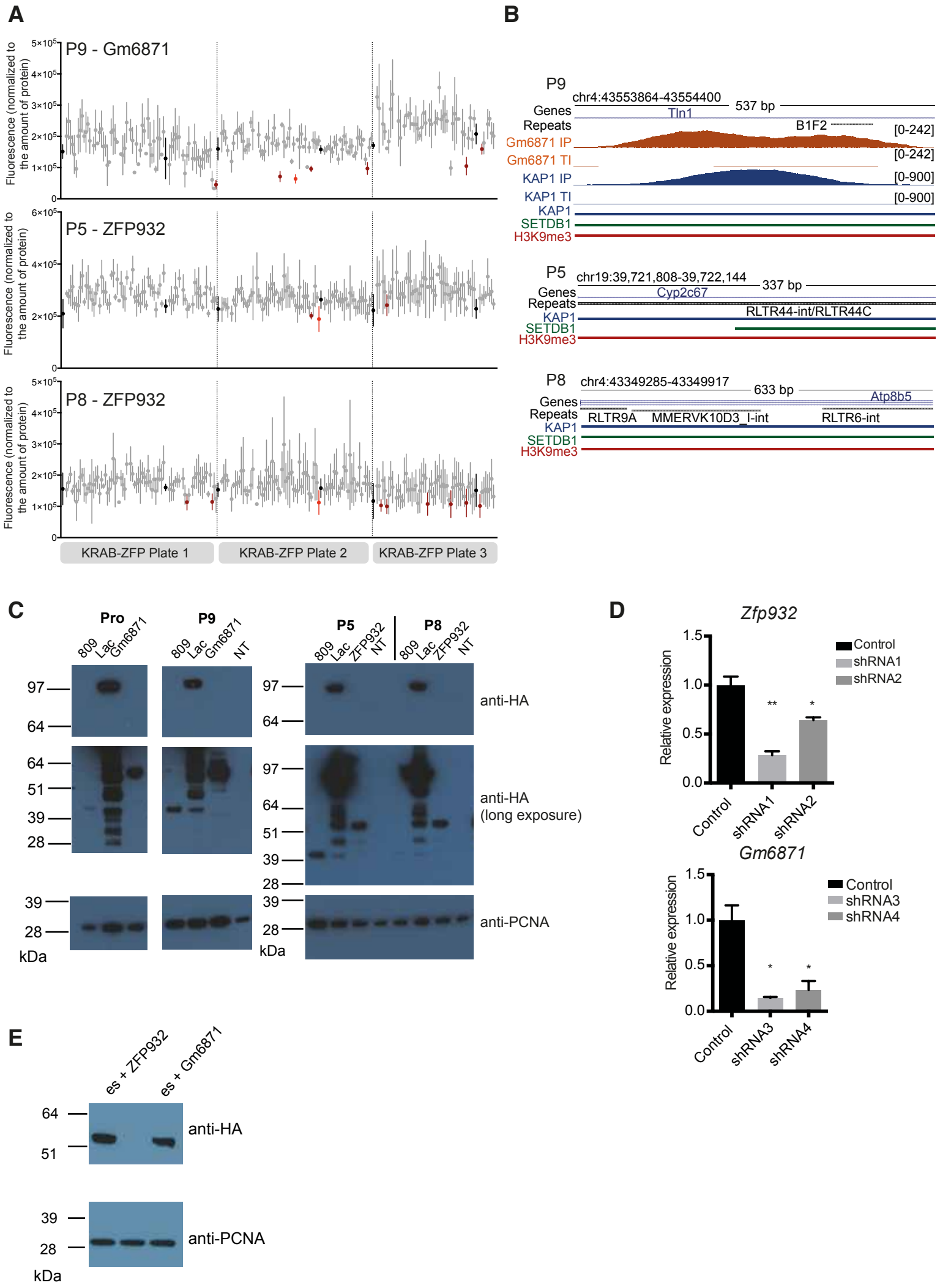
A
P9 - Gm6871
P5 - ZFP932
P8 - ZFP932
Fluorescence (normalized to the amount of protein)
KRAB-ZFP Plate 1
KRAB-ZFP Plate 2
KRAB-ZFP Plate 3
B
P9
chr4:43553864-43554400
537 bp
Genes
Repeats
Tln1
B1F2
Gm6871 IP
Gm6871 TI
KAP1 IP
KAP1 TI
[0-242]
[0-900]
KAP1
SETDB1
H3K9me3
P5
chr19:39,721,808-39,722,144
337 bp
Cyp2c67
RLTR44-int/RLTR44C
P8
chr4:43349285-43349917
633 bp
Atp8b5
RLTR9A
MMERVK10D3_I-int
RLTR6-int
C
Pro
P9
P5
P8
809
Lac
Gm6871
NT
ZFP932
anti-HA
anti-HA
(long exposure)
anti-PCNA
kDa
D
Zfp932
Relative expression
Control
shRNA1
shRNA2
Gm6871
shRNA3
shRNA4
E
es + ZFP932
es + Gm6871
anti-HA
anti-PCNA

Figure S2, related to Figure 2. Hits identified by subjecting selected target sequences to functional screen.

(**A**) Normalized fluorescence high-throughput readout for P9 and Gm6871, and P5/P8 and ZFP932. Black: transfection control; red: hits identified by plate reader; light red: hits confirmed by FACS. (**B**) Genomic region of P9, P5 and P8. Different ChIP-seq peaks or reads densities are shown for each locus. (**C**) Western blot control of 293T cells overexpressing HA-tagged proteins, referent to experiment in Figure 2A. (**D**) Levels of *Zfp932* and *Gm15446* mRNA upon KRAB-ZFP depletion through shRNAs, corresponding to experiment depicted in Figure 2B. (**E**) Western blot of ES cell lines overexpressing HA-tagged ZFP932 or Gm6871, used for ChIP-PCR in Figure 2C. Error bars represent SD, * $p < 0.05$, ** $p \leq 0.01$, Student's t test.

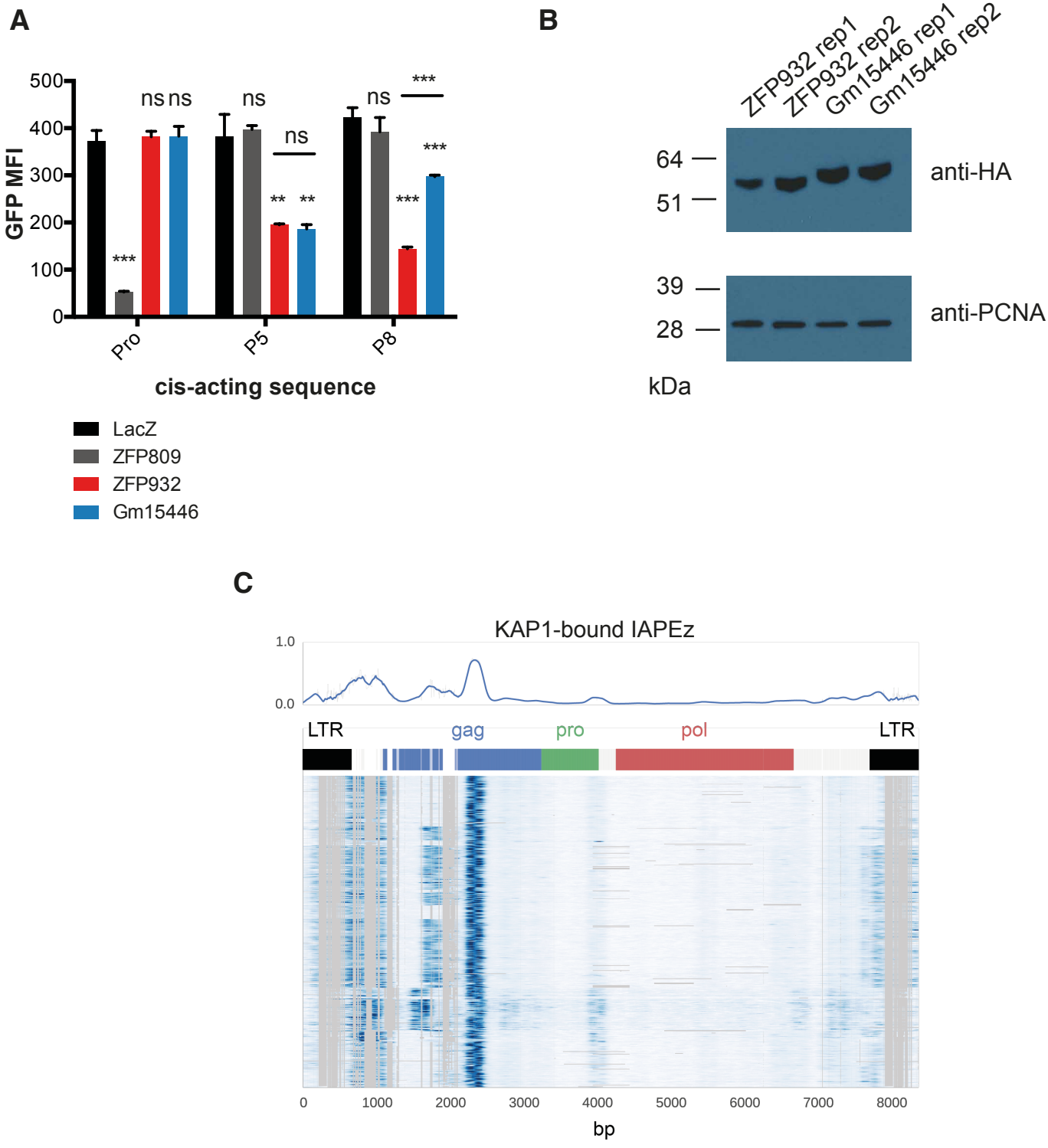
A
GFP MFI
500
400
300
200
100
0
ns ns
ns
ns
***
**
**
***
***
***
Pro
P5
P8
cis-acting sequence
LacZ
ZFP809
ZFP932
Gm15446
B
ZFP932 rep1
ZFP932 rep2
Gm15446 rep1
Gm15446 rep2
64
51
anti-HA
39
28
anti-PCNA
kDa
C
KAP1-bound IAPEz
1.0
0.0
LTR
gag
pro
pol
LTR
0
1000
2000
3000
4000
5000
6000
7000
8000
bp
D
chr8:123927320-123938011
10.7 kb
Repeats
IAPEz
LTR
-int
LTR
KAP1 IP
[0-900]
KAP1 TI
[0-900]
chr13:64978658-64987422
8.8 kb
Genes
Cntnap3
Repeats
IAPEz
LTR
-int
LTR
KAP1 IP
[0-900]
KAP1 TI
[0-900]

Figure S3, related to Figure 3. Differential regulation of TEs by the KRAB/KAP1 system.

(**A**) GFP repression assay of ZFP932/Gm15446 targets upon KRAB-ZFPs overexpression in 293T cell lines containing target *cis*-acting sequence upstream of PGK-GFP cassette. GFP mean fluorescence intensity is displayed and *P* value presented is relative to LacZ control, unless otherwise specified. (**B**) Western blot of ES cell lines used for ChIP-seq, overexpressing HA-tagged derivatives of ZFP932 or Gm15446. (**C**) Coverage plot of KAP1 ChIP-seq in ES cells on multiple alignment of "full length" (>6kb) IAPEz-int bound by KAP1. IAPEz repbase consensus is represented, and mean of binding coverage is depicted on top. Each row is individually normalized, with enrichment proportional to darkness of the blue color. (**D**) Two representative genomic regions containing full-length IAPEz. LTRs are highlighted and KAP1 ChIP-seq signal in murine ES is depicted. Error bars represent SD, ** $p \leq 0.01$, *** $p \leq 0.001$, ns= not significant, Student's t test.

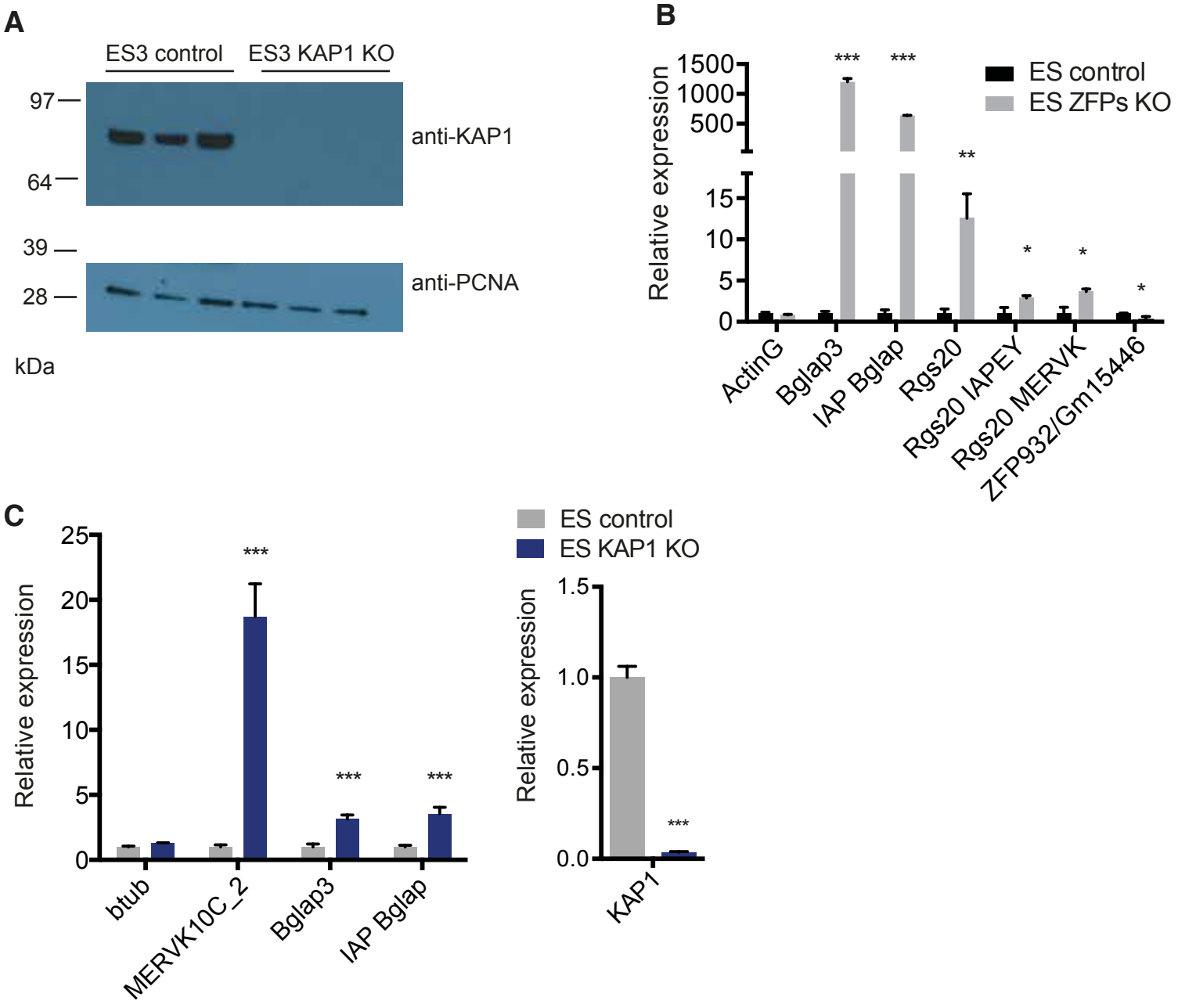

Figure S4, related to Figure 4. KAP1 and ZFP932/Gm15446 regulate the *Bglap3* locus in murine ESC.

(**A**) Western blot of *Kap1* KO ES cells used for RNA-seq. RT-qPCR analysis of mRNA expression of TE-gene pairs in *Zfp932/Gm15446* KO mES cells (**B**), *Kap1* KO in mES cells (**C**), and ZFP932/Gm15446 KD mES cells complemented with LacZ (control), ZFP932 or Gm15446 as indicated (**D**) (normalized to *Gapdh* and *ActinB*). For ZFP932/Gm15446 KD, ZFPs depletion was between 50-60%. Error bars represent SD, * $p < 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, Student's t test.

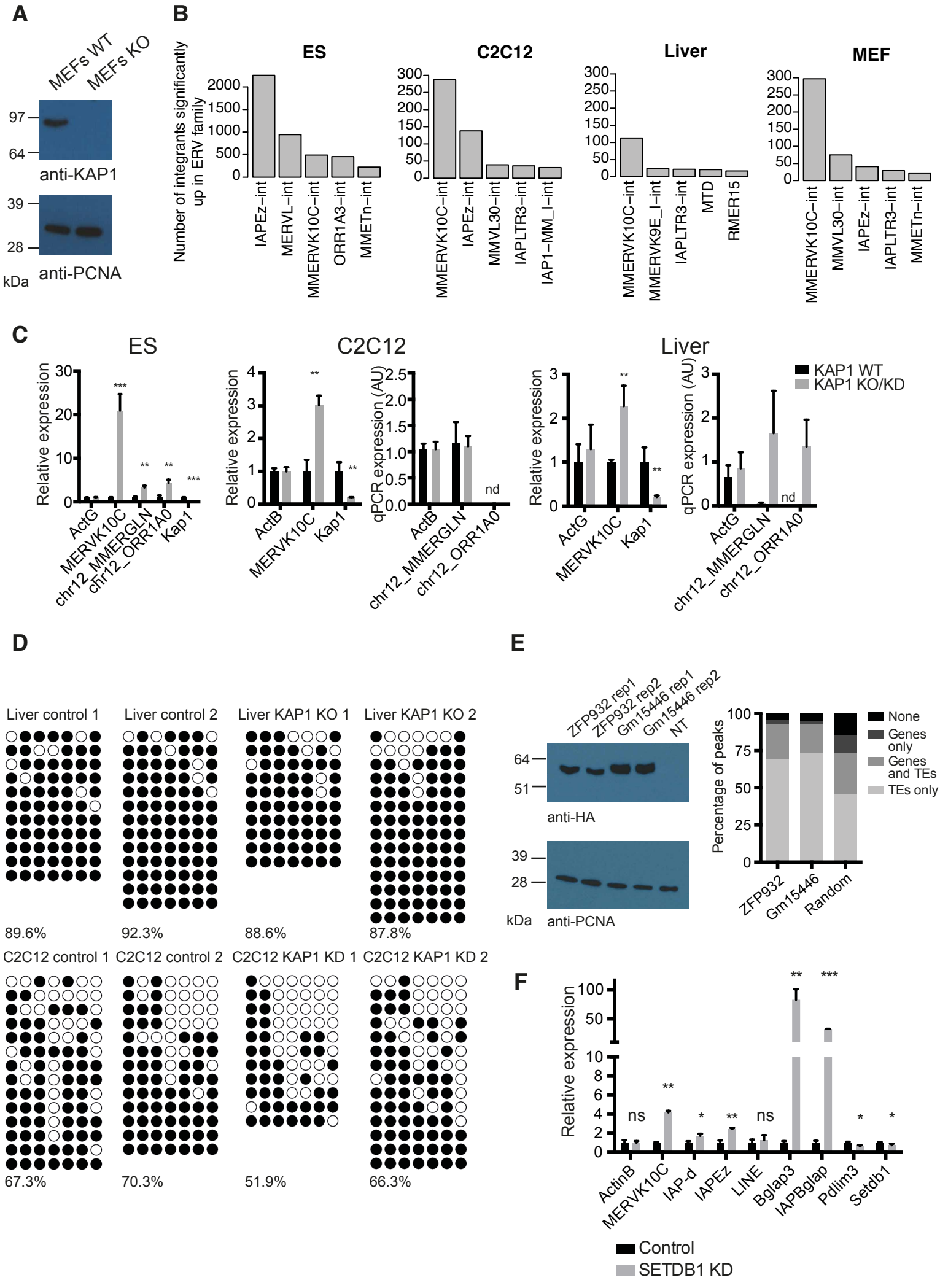
A
MEFs WT
MEFs KO
97
64
anti-KAP1
39
28
kDa
anti-PCNA
B
ES
C2C12
Liver
MEF
Number of integrants significantly up in ERV family
IAPEz−int
MERVL−int
MMERVK10C−int
ORR1A3−int
MMETn−int
MMERVK10C−int
IAPEz−int
MMVL30−int
IAPLTR3−int
IAP1−MM_I−int
MMERVK10C−int
MMERVK9E_I−int
IAPLTR3−int
MTD
RMER15
MMERVK10C−int
MMVL30−int
IAPEz−int
IAPLTR3−int
MMETn−int
C
ES
C2C12
Liver
Relative expression
qPCR expression (AU)
KAP1 WT
KAP1 KO/KD
ActG
MERVK10C
chr12_MMERGLN
chr12_ORR1A0
Kap1
ActB
nd
D
Liver control 1
Liver control 2
Liver KAP1 KO 1
Liver KAP1 KO 2
89.6%
92.3%
88.6%
87.8%
C2C12 control 1
C2C12 control 2
C2C12 KAP1 KD 1
C2C12 KAP1 KD 2
67.3%
70.3%
51.9%
66.3%
E
ZFP932 rep1
ZFP932 rep2
Gm15446 rep1
Gm15446 rep2
NT
64
51
anti-HA
39
28
kDa
anti-PCNA
Percentage of peaks
None
Genes only
Genes and TEs
TEs only
ZFP932
Gm15446
Random
F
Relative expression
ns
ActinB
MERVK10C
IAP-d
IAPEz
LINE
Bglap3
IAPBglap
Pdlim3
Setdb1
Control
SETDB1 KD

Figure S5, related to Figure 5. ERVs are controlled by KAP1 in differentiated cells.

(**A**) Western blot of MEFs *Kap1* KO cells used for RNA-seq. (**B**) Top 5 upregulated ERV families upon KAP1 depletion in different cells and tissues. Criteria for inclusion were more then 2-fold upregulation when compared to control cells and adjusted *P* value ≤ 0.05. (**C**) mRNA expression of ERVs in liver, C2C12, and ES KAP1-depleted cells versus control (ES and liver samples are normalized to *Gapdh* and *ActinB*; C2C12 samples are normalized to *ActinG* and *Tbp*). MERVK10C and MERGLN/ORR1A0 expression was measured with general and locus-specific primers, respectively. nd, not detected. (**D**) Bisulfite sequencing of MMERVK10C-int sequences in *Kap1* KO vs. wild type liver and in KAP1 KD vs. control C2C12 cells. Empty and full circles correspond to unmethylated and methylated CpG dinucleotides, respectively. (**E**) Western blot of C2C12 cell lines used for ChIP-seq, overexpressing HA-tagged derivatives of ZFP932 or Gm15446; and percentage of ZFP932, Gm15446, or random control peaks on genic and repeated regions. Random control is based on the mean of 100 random shuffling of the peaks of Gm15446 (KRAB-ZFP ChIP with more peaks). (**F**) RT-qPCR analysis of mRNA expression of general ERVKs, *Bglap3* gene, and IAP Bglap in SETDB1 KD C2C12 cells (samples are normalized to *ActinG* and *Tbp*). *Plim3* is a control gene known to be downregulated. Error bars represent SD, * $p < 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, ns= not significant, Student's t test.

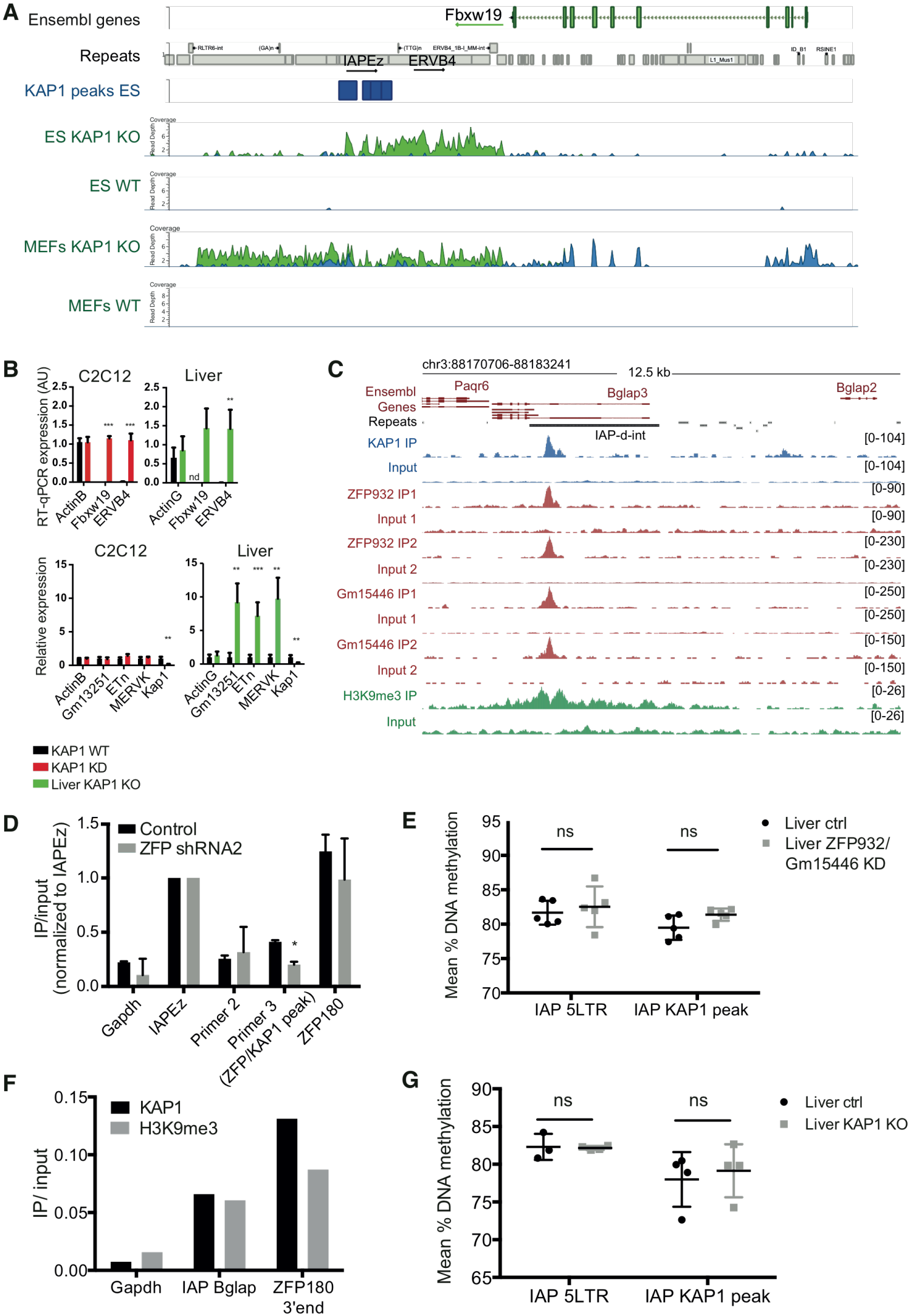
A
Ensembl genes
Fbxw19
Repeats
IAPEz
ERVB4
KAP1 peaks ES
ES KAP1 KO
ES WT
MEFs KAP1 KO
MEFs WT
B
C2C12
Liver
RT-qPCR expression (AU)
ActinB
Fbxw19
ERVB4
ActinG
nd
Relative expression
Gm13251
ETn
MERVK
Kap1
KAP1 WT
KAP1 KD
Liver KAP1 KO
C
chr3:88170706-88183241
12.5 kb
Paqr6
Bglap3
Bglap2
Ensembl Genes
Repeats
IAP-d-int
KAP1 IP
Input
ZFP932 IP1
Input 1
ZFP932 IP2
Input 2
Gm15446 IP1
Input 1
Gm15446 IP2
Input 2
H3K9me3 IP
Input
[0-104]
[0-90]
[0-230]
[0-250]
[0-150]
[0-26]
D
IP/input (normalized to IAPEz)
Control
ZFP shRNA2
Gapdh
IAPEz
Primer 2
Primer 3 (ZFP/KAP1 peak)
ZFP180
E
Mean % DNA methylation
ns
IAP 5LTR
IAP KAP1 peak
Liver ctrl
Liver ZFP932/ Gm15446 KD
F
IP/ input
KAP1
H3K9me3
Gapdh
IAP Bglap
ZFP180 3'end
G
Mean % DNA methylation
ns
IAP 5LTR
IAP KAP1 peak
Liver ctrl
Liver KAP1 KO

Figure S6, related to Figure 6. **ZFP932/Gm15446 regulate the *Bglap3* locus in differentiated cells.**

(**A**) GenomeBrowse screenshot of *Fbxw19* locus with coverage of stranded RNA-seq data in mESC and MEFs WT and KO for *Kap1*. Coverage in the sense strand is represented in green and in the anti-sense strand in blue. The orientation of the genes and TEs is represented with arrows. (**B**) Liver and C2C12 mRNA expression upon KAP1 depletion of two genomic regions with gene and TEs controlled by KAP1 in differentiated cells (liver samples are normalized to *Gapdh* and *ActinB*, and C2C12 samples to *ActinG* and *Tbp*). nd, not detected. AU, arbitrary units. (**C**) ChIP-seq density of ZFP932/Gm15446 and associated effectors on the *Bglap3* locus in C2C12 are illustrated. (**D**) ChIP-PCR of KAP1 in ZFP932/Gm15446 KD cells, normalized to IAPEz positive control. KD levels are of 65%. Primer 2 and 3 correspond to numbering present in Fig 4C. (**E**) Pyrosequencing of IAP Bglap in ZFP932/Gm15446 KD mouse liver, compared to control. (**F**) ChIP-PCR of KAP1 and H3K9me3 in mouse liver. (**G**) Pyrosequencing of IAP Bglap in *Kap1* KO mouse liver, compared to control. For DNA methylation analysis, mean CpG DNA methylation per sequence and per replicate was calculated. Error bars represent SD, * $p < 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, ns= not significant, Student's t test.

SUPPLEMENTAL TABLES

Table S1, related to Figure1. Sequences selected from ChIP-seq analysis to be tested in the screen.

Sequence ID	chromosome	start	end	size (bp)	TE	gene
P2	chr15	96071841	96072638	797	IAPEY_LTR, MIRc	ENSMUSG00000085460,ENSMUSG00000086809
P3	chr16	16373546	16374572	1026	RLTR44-int	-
P4	chr17	17449713	17450247	534	Lx4A, LTRIS2	-
P5	chr19	39721807	39722144	337	RLTR44-int	ENSMUSG00000062624
P6	chr2	24366644	24367245	601	LTRIS2	-
P7	chr2	28285331	28285800	469	LTRIS2	-
P8	chr4	43349284	43349917	633	RLTR9A3, MMERVK10D3_I-int, RLTR6-int	ENSMUSG00000028457
P9	chr4	43553863	43554400	537	B1F2	ENSMUSG00000028465
P10	chr4	55012441	55013092	651	L1MA5	ENSMUSG00000060206
P11	chr4	152526608	152527445	837	MTC-int	-
P12	chr4	155549797	155550889	1092	B1_Mur3, MT2B	ENSMUSG00000041936
P13	chr5	114074078	114074817	739	L2b, MTEa	ENSMUSG00000042190
P14	chr5	130162484	130163546	1063	-	ENSMUSG00000051034
P15	chr6	4873614	4874324	710	RLTRETN_Mm	ENSMUSG00000032827
P16	chr6	48546386	48547039	653	-	ENSMUSG00000052751
P17	chr7	26652988	26653803	816	RLTR44-int, ORR1E	-
P18	chr7	86361075	86361707	632	B2_Mm2, LTRIS2	-
P19	chr8	88210944	88211630	686	ORR1A3-int	ENSMUSG00000036902
P20	chr9	87086351	87086989	638	RLTR44-int	ENSMUSG00000056919

Table S2, related to Figure1. Library of KRAB-ZFPs and controls used in the screen, with index number.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Plasmids and lentiviral vectors. For the screen, DNA targets were cloned into the pENTR/D/TOPO vector and transferred via gateway cloning into a pRRLSIN.cPPT.PGK-GFP.WPRE vector that was modified to contain R1-R2 gateway sites upstream the PGK-GFP cassette (pRRLSIN.cPPT.R1R2.PGK-GFP.WPRE). The KRAB-ZFP library was obtained partially from previous work (Gubelmann et al., 2013), and the remaining was codon optimized, synthetized into pENTR vectors, and further transferred via gateway cloning into a puromycin-selectable lentivector under a tetracyclin-inducible TRE promoter to obtain HA-tagged proteins (pSIN-TRE-R1R2-3xHA). pLKO.puro shRNA vectors or versions modified to contain the hygromycin resistance or tomato red cDNAs were used for KAP1 and ZFPs KD. The shRNA for *ZFP932/Gm15446* were obtained from the RNAi Consortium (http://www.broadinstitute.org/rnai/public/) or designed using the KRIBB siRNA AsiDesigner tool (http://sysbio.kribb.re.kr:8080/AsiDesigner/menuDesigner.jsf). For CRISPR-mediated KO, the lentiCRISPR vector (Shalem et al., 2014) was modified with an ubiquitin promoter in place of the EF1short promoter. sgRNA sequences were designed using the Zhang lab tool (http://crispr.mit.edu/). For KAP1 overexpression, pSicoR-KAP1-HA vector was used. Lentiviral vectors production protocols are detailed at http://tronolab.epfl.ch and backbones are available at Addgene (http://www.addgene.org).

Cell culture and cell-based assays. Mouse ES cells were cultured as previously described (Rowe et al., 2013a), using the *Kap1 loxP/loxP* ES3 line or its *Kap1* conditional KO derivative after transduction with a tamoxifen-inducible Cre vector (Rowe et al., 2010). For KO induction, cells were treated overnight with 1µM tamoxifen (Sigma) and collected at the latest 4-5 days later. *Kap1* KO and *loxP/loxP* control MEFs were generated and cultured as previously described (Rowe et al., 2013b). For KRAB-ZFP KO experiments, murine ES cells were transduced with an integrase defective lentiviral vector containing the pLentiCRISPR with an sgRNA for *Zfp932* and *Gm15446* or a control sgRNA against luciferase. Cells were selected for puromycin, and cloned from single cells. Two clones were selected and genotyped by PCR followed by TOPO cloning and Sanger sequencing. Clone 1 had a 909bp deletion in *Zfp932* disrupting exon 2 and a 4bp deletion in *Gm15446* (in exon 2); in clone 2 exons 2 and 3 of *Zfp932* and *Gm15446* were damaged. KRAB-ZFPs KD was induced with shRNA vectors for *Zfp932/Gm15446* on untransduced cells or on cells previously transduced with TRE vector containing LacZ control or ZFP for rescue experiments. For repression assays, cells were transduced at low multiplicity of infection (MOI), GFP fluorescence was assessed by flow cytometry, and ES cells pluripotency was monitored by anti-SSEA1 staining (BD Pharmingen, 560142). KAP1 KD was performed with shRNA vectors for *Kap1*. For KRAB-ZFPs ChIP experiments, C2C12 or ES cells were transduced with HA-tagged pSIN-TRE lentiviral vectors containing the codon-optimized KRAB-ZFP of interest, selected, and treated with doxycycline for 48h before harvesting. For KAP1 ChIP experiments in MEFs, MEF *Kap1* KO cells were complemented with HA-tagged KAP1 with levels similar to the endogenous protein. Cells were selected with 1µg/mL puromycin or 100 µg/mL hygromycin when necessary.

Mouse work. Lentiviral transgenesis was performed by perivitelline injection of pLKO.tomato vectors containing shRNA for *ZFP932/Gm15446* or empty control into fertilized oocytes (strain B6D2F1/J) that were transferred to foster mothers (strain NMRI). Lentiviral vectors for transgenesis were prepared as previously described (Barde et al., 2011), using Episerf medium (Invitrogen), the particle concentration obtained by p24 ELISA (PerkinElmer), and the infectious titer determined on HCT116 cells by tomato flow cytometry. Generated adults were genotyped by flow cytometry of peripheral blood cells and by quantitative PCR to assess the number of lentivector integration in the genome. *Kap1* KO and ZFP932/Gm15446 KD livers were harvested at 8 weeks of age.

Functional screen. DNA target sequences to be tested were chosen from the overlap of publicly available murine ESC ChIP-seq data for KAP1 (GSE41903), SETDB1 (GSE18371), H3K9me3 (GSE41903), and absence of ZFP57 (GSE31183) (Bilodeau et al., 2009; Quenneville et al., 2011; Rowe et al., 2013b). Nineteen sequences corresponding to KAP1 peaks were selected based on presence of TEs or of interesting KAP1 targets (such as 3'end of ZFP genes) (Table S2) and cloned into the pRRLSIN.cPPT.R1R2.PGK-GFP.WPRE vector. These sequences were tested for repression in murine ES and only repressed sequences were tested in the screen. $PBS^{Lys1,2}$ sequence was also cloned in the pRRLSIN.cPPT.R1R2.PGK-GFP.WPRE vector and tested in the screen. Lentiviral vectors with these sequences were used to transduce 293T cells at low MOI in order to favour integration of a single copy per cell. GFP positive cells were sorted and used in the screen. The ENSEMBL ID and amino

acid sequences of the KRAB-ZFPs used to establish the screen (Table S1) were obtained from our previously published curation (Corsinotti et al., 2013) and cloned as described above. Reverse transfections and cell culture were done in triplicates in 96-well plates, using a Sciclone ALH 3000 (Caliper Life Sciences) liquid handling robot and a Multidrop Combi dispenser (Thermoscientific). For each well, $5x10^3$ cells in culture media were added to a 10µL mix containing 100-150 µg of KRAB-ZFP plasmid and 0.6 µl of Fugene 6 Transfection Reagent (Promega) in Opti-MEM (Life Technologies). All screen plates contained two LacZ plasmids and two empty wells as controls. For induction and selection of the TRE vector, doxycycline (final concentration 5 µg/mL) was added 16-20h after transfection, and on the following day cultures were supplemented with puromycin (final concentration 1 µg/mL). Cells were harvested on day 6 after transfection, resuspended in RIPA buffer containing protease inhibitors, and GFP was measured using a Tecan Infinite F500 plate reader (Tecan) (excitation at 485 nm and emission at 520 nm). Total protein content was quantified using BCA (BCA Protein Assay Reagent, Thermoscientific), and used to calculate normalized GFP fluorescence. Candidate hits were identified by selecting the 10 KRAB-ZFPs with the lowest normalized fluorescence values per 96-well plate, and only the ones that were present in all 3 replicates were considered. Hits were identified by transfection of the 293T cell line of interest with candidate KRAB-ZFPs in 24-well plates ($1.5x10^4$ cells/well transfected with 150-200 µg of DNA and 0.38 µL of fugene mix in Opti-MEM); doxycycline and puromycin were added at the same concentrations, with FACS readout after 6 days.

ChIP-PCR and ChIP-seq. Cells were harvested, washed with Episerf, fixed in 2mL per $1x10^7$ cells (10 min in 1% formaldehyde), quenched with glycine in 10mL (at 125 mM final), washed three times with PBS, and pelleted. Each pellet containing $1x10^7$ cells was lysed, resuspended in 1 mL of sonication buffer on ice (for KRAB-ZFPs: 10 mM Tris at pH 8, 1 mM EDTA, 0.1% SDS, and protease inhibitors; for all others: 10 mM Tris at pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% NaDOC, 0.25% NLS, and protease inhibitors), transferred to TC 12x12 tubes (Covaris), and sonicated (Covaris settings: 20 min, 5% duty cycle, 140W, 200 cycles). Sonication was assessed by reverse cross-linking (65°C, RNAse A at 1µg/µL, overnight), followed by DNA extraction. Fragment size (between 200-400bp) was checked on a Bioanalyzer (Agilent 2100). Immunoprecipitations were performed with chromatin from $1x10^7$ cells (for KAP1, histone marks, and PolII), or $4x10^7$ (for KRAB-ZFPs), with Dynabeads (ThermoFisher) in IP buffer (for KRAB-ZFPS: 10 mM Tris at pH 8, 1 mM EDTA, 0.1% SDS, 150 mM NaCl, 10% Triton X-100, and protease inhibitors; for all others: 16.25 mM Tris at pH 8.1, 137.5 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1.25% Triton X-100, and protease inhibitors) overnight. Chromatin was reversed cross-linked (65°C, Proteinase K at 400ng/µL, overnight) and DNA was further extracted for analysis. Antibodies used were HA (Covance, MMS-101P), KAP1 (Tronolab, rabbit polyclonal S23470), H3K9me3 (Diagenode, C15410056), H3K4me1 (Diagenode, pAb-037-050), H3K27ac (Abcam, ab4729), RNA PolII CTD repeat (Abcam, ab817), H3 (Abcam, ab1791). ChIP samples were used for SYBER Green qPCR (Aplied Biosystems) or library preparation for sequencing. Primers (were designed using Primer 3 (Untergasser et al., 2007) or GETPrime (Gubelmann et al., 2011). Libraries of immunoprecipitated chromatin and total input control from ChIP were performed with single-end adaptors as previously described (Rowe et al., 2013b). Sequencing was performed on an Illumina HiSeq 2500 (Illumina), with each library sequenced in 100-bp reads run.

RT-qPCR and RNA-seq. Total RNA was extracted and DNAse-I treated using a spin column-based RNA purification kit (Macherey-Nagel). Reverse transcription was performed with 500ng of RNA using random primers and SuperScriptII (Invitrogen). Primers were designed using Primer 3 (Untergasser et al., 2007) or GETPrime (Gubelmann et al., 2011), and used for SYBER Green qPCR (Aplied Biosystems). For mRNA sequencing, 100-bp single-end RNA-seq libraries were prepared using 200ng of total RNA and the Illumina TruSeq Stranded mRNA reagents (Illumina). Cluster generation was performed with the resulting libraries using the Illumina TruSeq SR Cluster Kit v4 reagents and sequenced on an Illumina HiSeq 2500 (Illumina).

Primers, shRNAs and sgRNAs.

Primers used in this study

Primer	Sequence	Purpose
Gapdh F	GCCCTTCTACAATGCCAAAG	ChIP-qPCR
Gapdh R	TTGTGATGGGTGTGAACCAC	ChIP-qPCR

P9 F	CTTGAGGCCAGCCAAGGA	ChIP-qPCR
P9 R	GCCTTAACAGCCTTACTTCTAGAATTG	ChIP-qPCR
P5 F	AGCCTTGGAAACAGGAACAG	ChIP-qPCR
P5 R	CACACTTTTGCCATCCTGTC	ChIP-qPCR
P8 F	ACACAATCTCCCCCCCTTTT	ChIP-qPCR
P8 F	GTGCCCCCTGTCCAGCTA	ChIP-qPCR
Bglap ChIP 1 F	CCCAGTGTCTGAAAGGGTAGG	ChIP-qPCR
Bglap ChIP 1 R	ATACTGGCCAACAGGAATGC	ChIP-qPCR
Bglap ChIP 2 F	TCATGGTGTCTGCTAGGTGTG	ChIP-qPCR
Bglap ChIP 2 R	TCAGAATCAGAGGCAACAGG	ChIP-qPCR
Bglap ChIP 4 F	TTGGTGCACTGTTTGACCTG	ChIP-qPCR
Bglap ChIP 4 R	AATAAGGTTCCCGGTCTTGG	ChIP-qPCR
Bglap ChIP 5 F	CAGCCCAACTGTGTGTTTTC	ChIP-qPCR
Bglap ChIP 5 R	ACATTTGGCCACGACCTATG	ChIP-qPCR
Bglap ChIP 6 F	TCTCTGATGTAAGCAGGAGGAG	ChIP-qPCR
Bglap ChIP 6 R	CAATCACCAACCACAGCATC	ChIP-qPCR
Bglap ChIP 7 F	CACACTGTACAAGAGGCTCCAG	ChIP-qPCR
Bglap ChIP 7 R	TTGTGCTGGAGTGGTCTCTATG	ChIP-qPCR
ZFP180 3'end F	CCGTACAGGTGCAATCTGTG	ChIP-qPCR
ZFP180 3'end R	GTTTGTAGCTCTGGCGGAAC	ChIP-qPCR
Gapdh F	TCCATGACAACTTTGGCATTG	RT-qPCR
Gapdh R	AGTCTTCTGGGTGGCAGTGA	RT-qPCR
Actin F (ActinB)	CTAAGGCCAACCGTGAAAAGAT	RT-qPCR
Actin R (ActinB)	CACAGCCTGGATGGCTACGT	RT-qPCR
ActinG F	TGGATCAGCAAGCAGGAGTATG	RT-qPCR
ActinG R	CCTGCTCAGTCCATCTAGAAGCA	RT-qPCR
TBP F	TTGACCTAAAGACCATTGCACTTC	RT-qPCR
TBP R	TTCTCATGATGACTGCAGCAAA	RT-qPCR
btub F	GCAGTGCGGCAACCAGAT	RT-qPCR
btub R	AGTGGGATCAATGCCATGCT	RT-qPCR
Bglap3 F	CTGACAAAGCCTTCATGTCC	RT-qPCR
Bglap3 R	TCAAGCTCACATAGCTCCC	RT-qPCR
IAP Bglap F/Bglap ChIP 3 F	AGGTGTTGCAGAGGTTTTGG	RT-qPCR and ChIP-qPCR
IAP Bglap R/Bglap ChIP 3 F	AATATCGGACACAGGGCAAG	RT-qPCR and ChIP-qPCR
Rgs20 F	CTACTTGTGGCCTCAATGG	RT-qPCR
Rgs20 R	GACAGTGAGGCAAGAACAG	RT-qPCR
Rgs20 IAPEY F	GTTCCTGCAAAACAGACTGC	RT-qPCR
Rgs20 IAPEY R	ATTGCTCTGGTCAGCCATTC	RT-qPCR
Rgs20 MERVK F	ACTGGAGGTCCTTGTCCTATG	RT-qPCR
Rgs20 MERVK R	AGGTTCGATGTGCTCTTTCC	RT-qPCR
ZFP932/Gm15446 F	TTGCACATCATTGTCATCTCC	RT-qPCR
ZFP932/Gm15446 R	CTGACCTACAAAGGCTTTACCAC	RT-qPCR
MERVK10C F	GCCCCTCAATTGGTAGAATG	RT-qPCR

MERVK10C R	TTTCCGGCAGTCTCTAATGC	RT-qPCR
MERVK10C_2 F	CAAATAGCCCTACCATATGTCAG	RT-qPCR
MERVK10C_2 R	GTATACTTTCTTCTTCAGGTCCAC	RT-qPCR
KAP1 F	CGGAAATGTGAGCGTGTTCTC	RT-qPCR
KAP1 R	CGGTAGCCAGCTGATGCAA	RT-qPCR
chr12_ORR1A0 F	GGTTGGAATGGGTGTTTCAC	RT-qPCR
chr12_ORR1A0 R	TCGTCCAACTTTCCAAGTCC	RT-qPCR
chr12_MMERGLN F	ACCCACAGTCTGCAAAATCC	RT-qPCR
chr12_MMERGLN R	AGTGATGCGGATTCCAACTC	RT-qPCR
Fbxw19 F	TGTGTACGTGTGGGAGGAGA	RT-qPCR
Fbxw19 R	AGAAAGCAGGGAATGGGACT	RT-qPCR
Fbxw19 ERVB4 F	TTAAAGCAGGGGAGGTGTTG	RT-qPCR
Fbxw19 ERVB4 R	GACCCCCTTTTCTTTTCTGG	RT-qPCR
Gm13251 F	GATGTGAAGTGTGCTTCGAGT	RT-qPCR
Gm13251 R	CACAACAGGACCAGACACCA	RT-qPCR
Gm13251 Etn F	TAATCTTTGGGCCAGGACTC	RT-qPCR
Gm13251 Etn R	CCAAAGAAATGCCACACCTG	RT-qPCR
Gm13251 MERVK F	AGCCCTTGGGATGATAACAG	RT-qPCR
Gm13251 MERVK R	GGATAACGCAATGCTGTGTG	RT-qPCR
Gm6871 F	ACCTACAGGAATCTCACCAC	RT-qPCR
Gm6871 R	GTTTGGTGCCTTCCATGTC	RT-qPCR
IAP-d F	CAGCTGAACACAATCACTCATC	RT-qPCR
IAP-d R	TCCAGTGCGGGAATCTATG	RT-qPCR
IAPEz F	CTTGCCCTTAAAGGTCTAAAAGCA	RT-qPCR and ChIP-qPCR
IAPEz R	GCGGTATAAGGTACAATTAAAAGATATGG	RT-qPCR and ChIP-qPCR
LINE F	TTTGGGACACAATGAAAGCA	RT-qPCR
LINE R	CTGCCGTCTACTCCTCTTGG	RT-qPCR
Pdlim3 F	AACCACAGGAATTCAAACCC	RT-qPCR
Pdlim3 R	TGTCATCAATGTTTGCTGCT	RT-qPCR
Setdb1 F	GATGTCCCCACTTCCTCTGA	RT-qPCR
Setdb1 R	GCATAGCTACGCCACACTGA	RT-qPCR
KVDMR F	GGGGTTTAAAGGGTTTTAAGATTAT	pyrosequencing
KVDMR R	TCATAACCTCCCCCTCCTC	pyrosequencing
KVDMR Seq	TGTAAGTTTGGGTTATAAAGA	pyrosequencing
IAP 5LTR F	GGAATAGTAGTTTATTGGTTAGAATTAT	pyrosequencing
IAP 5LTR R	CCCCACCACTCCTACTTACATCAAT	pyrosequencing
IAP 5LTR Seq	GGTTAGAATTATTATTTGTTATATG	pyrosequencing
KAP1 peak F	GGTTGGTGATAAGTTTAGGGAGTTT	pyrosequencing
KAP1 peak R	TACTAACAACTATCCCCCTCT	pyrosequencing
KAP1 peak Seq	GGTGATAAGTTTAGGGAGTTTTA	pyrosequencing
bisMERVK10C.1F	ATAGTTTAATTTAAGATATGGGGTT	bisulfite
bisMERVK10C.1R	ACAATAATCAATACCACTCTACAAC	bisulfite
GAG F	GGAGCTAGAACGATTCGCAGTTA	qPCR - mice genotyping

GAG R	GGTGTAGCTGTCCCAGTATTTGTC	qPCR - mice genotyping
GAG T (probe)	ACAGCCTTCTGATGTTTCTAACAGGCCAGG	qPCR - mice genotyping
Albumin F	GCTGTCATCTCTTGTGGGCTGT	qPCR - mice genotyping
Albumin R	ACTCATGGGAGCTGCTGGTTC	qPCR - mice genotyping
Albumin T (probe)	CCTGTCATGCCCACACAAATCTCTCC	qPCR - mice genotyping
ZFP932 F	TGACTTTTTAAAATAAGGGAACAACTG	CRISPR genotyping
ZFP932 R	TGGCTCACATTTGTAGCATCA	CRISPR genotyping
Gm15446 F	TTTCTCTCTCCCTTTCTCTCTCC	CRISPR genotyping
Gm15446 R	CATGGCTCACATTTGTACATCA	CRISPR genotyping
ZFP932/Gm15446 F	TGCAACATATATCCTATACAAAGAGCA	CRISPR genotyping
ZFP932/Gm15446 R	CATGCCTGCAAAGATAATTGG	CRISPR genotyping

shRNAs and sgRNAs used in this study

shRNA or sgRNA	**sequence (hairpin or sgRNA with PAM)**	**Target**
shRNA1	CCGGCCTCTCATGGTCAACTTCAAACTCGA GTTTGAAGTTGACCATGAGAGGTTTTTG	ZFP932/Gm15446
shRNA2	CCGGGCCTATTCACGACACAGCATTCTCGA GAATGCTGTGTCGTGAATAGGCTTTTTG	ZFP932/Gm15446
shRNA3	CCGGTATGATAACGTCTTCACATATCTCGA GATATGTGAAGACGTTATCATATTTTTG	Gm6871
shRNA4	CCGGAACGTCTTCACATATCACAGTCTCGA GACTGTGATATGTGAAGACGTTTTTTTG	Gm6871
shKAP1	CCGGCCGCATGTTCAAACAGTTCAACTCGA GTTGAACTGTTTGAACATGCGGTTTTTG	KAP1
shSETDB1	CCGGGCCTTGATCTTCCATGTCATTCTCGAG AATGACATGGAAGATCAAGGCTTTTTG	SETDB1
sg932Gm	TGTGAAAGCTCCAGAAGACATGG	ZFP932/Gm15446
sgLuc	ACGCTGGGCGTTAATCAAAGAGG	luciferase (control)

Bioinformatics analyses

Merged repeats track. All TE analyses were performed using a merged repeats track generated in-house. As the RepeatMasker annotation often contain TEs fragmented into smaller regions that not always correspond to their biological organization, we generated a new annotation that merges some of this fragments based on ERVs structure. The RepeatMasker 3.2.8 (Repeat Library 20090604, http://www.repeatmasker.org/species/mm.html) was used as basis, and we computed a frequency table of LTRs surrounding internal ERV ("ERV-int") parts for each ERV family. To infer the significance of the LTR-int associations we computed the frequencies distributions of the putative LTR for each "-int", and performed a Wald test to assign a *P* value to each LTR-int pair. The LTR was attributed to an "-int" family when the *P* value was smaller than 0.001, the frequency of occurrence next to an "-int" was higher than 2%, and it was annotated as an "LTR". ERV-int integrants that shared the same name, or attributed LTRs, were merged with their neighbouring elements when the distance was shorter than 400 bp and there were no SINEs or LINEs in between. We also kept the fragmented information in our new annotated list. For the final track, the other families of repeats were not modified and were added to the merged ERV/LTR elements.

ChIP-seq analyses. Reads were mapped to the mouse genome assembly mm9 using Bowtie 2 short read aligner (Langmead and Salzberg, 2012), using the --sensitive-local mode. The peaks were called using either the MACS program v1.4.2.1 (Zhang et al., 2008) or the SICER software v1.1 for histone modification marks (Zang et al., 2009), with the total input chromatin coverage as control. For MACS we used the default software parameters and selected MACS score above 100. For SICER we used the recommended parameters for histone marks (redundancy threshold: 1, window size: 200, fragment size: 150, effective genome fraction: 0.74, gap size: 400, FDR: 0.01). For KAP1 in ES cells, PeakSplitter was also used (Salmon-Divon et al., 2010) with the default software parameters. For KRAB-ZFPs, all ChIP-seq experiments were performed in duplicates and only the overlap of peaks of both duplicates was considered for the analyses. For KAP1 ChIP-seq in MEFs, experiments were

performed in duplicates and the union of both replicates was used for further analyses. Bigwig tracks were normalized to reads per 100 million mapped reads. Repeats enrichment of Fig 3C and Fig S5 were calculated from the overlap of repeats with ZFP932 or Gm15446 peaks containing KAP1 using the intersect tool (with default parameters) from the BEDtools 2.25.0 software (Quinlan and Hall, 2010) (minimum of 1 bp overlap). As control, we randomly assigned the peaks in the genome and used them in overlaps with repeats (this process was repeated 100 times). Motif search was performed using all ZFP932 and Gm15446 peaks with RSAT (Medina-Rivera et al., 2015), using unbound ERVK sequences as background control.

Coverage plots. ChIP-seq signal on each feature of interest were extracted from the bigwig file, and each sequence signal was scaled between 0 and 1. Features were aligned on their 3' end, and plotted on a heatmap using the matplotlib library of Python. For coverage plots with alignments, sequences were multiply aligned using the MAFFT software v. 7.245 (auto parameters and gap extension penalty of 0.123), and the coverage of each sequence was plotted on top of the alignment normalizing it as before. When gaps were created by only one sequence, it was removed from the alignment prior to plotting. Each consensus was then added to the corresponding alignment.

ERV annotation. For MMERVK10C-int, consensus sequence and coordinates were taken from Repbase (http://www.girinst.org/repbase/). For IAPEz, consensus sequence was taken from Repbase. Coordinates for IAPEz were estimated from an alignment with IAP1-MM_I (using IAP1-MM_I Repbase coordinates) and were checked using previous publications (Carmi et al., 2011). IAP Bglap LTR annotation was done by comparison with previously annotated IAP LTRs (Christy et al., 1985).

RNA-seq analysis. The RNA-seq reads were mapped to the mm9 genome using TopHat (Kim et al., 2013) with the following parameters: --b2-sensitive --no-novel-juncs --no-novel-indels. These settings allow multimapped reads to be randomly assigned once among the mapped loci. Gene counts were generated using the HTseq-count program with default parameters. TE counts were computed using the multicov script from the BEDtools software with the -split option. Only genes or TEs that had at least as many reads as samples present in the analysis were considered further. Sequencing depth normalization and differential expression analyses were performed using the voom function from the R package LIMMA (Law et al., 2014) from Bioconductor (Gentleman et al., 2004). The gene library sizes as given by voom were used to normalize the TEs counts. To be considered significantly upregulated, a gene or a TE had to have 2-fold increased expression (*P* value adjusted ≤ 0.05 for *Kap1* KO/KD; *P* value ≤ 0.01 for *Zfps* KO). *P* values were computed using a moderated t-test and corrected for multiple testing using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). MA plots were generated from normalized count values as given by VOOM and plotted using the matplotlib library of Python.

Public sequencing data. Raw reads were downloaded from publicly available ChIP-seq (GEO IDs: GSE41903, GSE18371, GSE31183, GSE48519, GSE62664, GSE58323) and RNA-seq data (GEO IDs: GSE62664). Mapping and data processing was done as described above. CAGE peaks based expression table for mouse of FANTOM 5 data was downloaded from the FANTOM consortium website (http://fantom.gsc.riken.jp), and their TSS peaks or max count were obtained from UCSC genome browser public tracks (Lizio et al., 2015). Representative samples were selected and expression in TPM (tags per million) was plotted using R software (http://www.R-project.org).

Heatmaps and boxplots. Heatmaps were generated with R, using the heatmap.2 function. Both rows and columns were clustered using hierarchical clustering. The agglomeration method was "complete", and the distance metric used was Pearson distance. For heatmaps of ERV expression, only integrants that had 10 or more reads on average per set of replicates were included. Boxplots were generated using R, and binding to transcription factors was determined using BEDtools with a minimum of 1 bp overlap.

SUPPLEMENTAL REFERENCES

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) *57*, 289-300.

Carmi, S., Church, G.M., and Levanon, E.Y. (2011). Large-scale DNA editing of retrotransposons accelerates mammalian genome evolution. Nature communications *2*, 519.

Christy, R.J., Brown, A.R., Gourlie, B.B., and Huang, R.C. (1985). Nucleotide sequences of murine intracisternal A-particle gene LTRs have extensive variability within the R region. Nucleic acids research *13*, 289-302.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J.*, et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome biology *5*, R80.

Gubelmann, C., Gattiker, A., Massouras, A., Hens, K., David, F., Decouttere, F., Rougemont, J., and Deplancke, B. (2011). GETPrime: a gene- or transcript-specific primer database for quantitative real-time PCR. Database : the journal of biological databases and curation *2011*, bar040.

Gubelmann, C., Waszak, S.M., Isakova, A., Holcombe, W., Hens, K., Iagovitina, A., Feuz, J.D., Raghav, S.K., Simicevic, J., and Deplancke, B. (2013). A yeast one-hybrid and microfluidics-based pipeline to map mammalian gene regulatory networks. Molecular systems biology *9*, 682.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A.*, et al.* (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Molecular cell *44*, 361-372.

Salmon-Divon, M., Dvinge, H., Tammoja, K., and Bertone, P. (2010). PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. BMC bioinformatics *11*, 415.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G.*, et al.* (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science (New York, N.Y.) *343*, 84-87.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J.A. (2007). Primer3Plus, an enhanced web interface to Primer3. Nucleic acids research *35*, W71-74.