Supplementary Methods and Figure Legends

Alkbh1 and Tzfp Repress a Non-Repeat piRNA Cluster in Pachytene Spermatocytes

Line M. Nordstrand[1#], Kari Furu[1#], Jonas Paulsen[2], Torbjørn Rognes[1,3] and Arne Klungland[1,4*]

[1] Centre for Molecular Biology and Neuroscience, Department of Microbiology, Oslo University Hospital, Rikshospitalet, PO Box 4950 Nydalen, NO-0424 Oslo, Norway
[2] Institute for Medical Informatics, Oslo University Hospital, The Norwegian Radium Hospital, PO Box 4953 Nydalen, NO-0424 Oslo, Norway
[3] Department of Informatics, University of Oslo, PO Box 1080 Blindern, NO-0316 Oslo, Norway
[4] Institute of Basic Medical Sciences, University of Oslo, PO Box 1018 Blindern, NO-0315 Oslo, Norway

SUPPLEMENTARY METHODS

Yeast Two-Hybrid Screening (Y2H)

A LexA-based yeast two-hybrid (Y2H) screen of testis-specific transcripts with full-length human ALKBH1 as bait was conducted. Clones that grew under selection were pulled out from the primary screening plates, restreaked onto new selection plates and assayed for activity of the second reporter gene, lacZ, using a quantitative β-galactosidase assay (HTX assay, see http://www.dualsystems.com/ for details). Clones that survived the primary selection and lacZ tests were picked and the library plasmid isolated and amplified in E. coli. Since yeast cells have the potential to take up several independent library plasmids during the transformation procedure, two independent E. coli transformants per original yeast clone were picked for plasmid isolation.

Dot Blot Analysis

0.2 µg, 0.7 µg and 1.4 µg of purified ALKBH1 (a kind gift from Professor H. Krokan, NTNU in Trondheim) was dotted onto a nitrocellulose membrane and then blocked with skimmed milk. The membrane was incubated with full-length or truncated TZFP-Myc (approximately 10 µg in 5 ml PBST) and the ALKBH1-TZFP complex was detected using anti-Myc (Clontech). Full length TZFP-Myc was used as a positive control (12.5 ng, 50 ng and 100 ng) to check the efficiency of the antibody. BSA was used as a negative control (10 ng, 50 ng and 100 ng) to check for unspecific binding of TZFP to other proteins. In addition, one blot was not incubated

with TZFP and thus works as a negative control for unspecific binding of the antibody. TZFP-Myc was purified from transfected HEK-cells by immunoprecipitation using anti-Myc coated agarose beads (Pierce Biotechnology).

Immunofluorescent Staining of Pachytene Cells

The purity of pachytene spermatocytes isolated using the STAPUT method was also verified using pachytene specific antibodies. Isolated cells were fixed onto SuperFrost Plus slides (VWR) using Cell Adhesive Solution (Crystalgen, Lot no 425081). Before staining, the cells were washed in 1x PBS and blocked in 1% BSA for 1 hour at room temperature. The slides were then incubated with primary antibodies overnight at 4°C prior to detection with secondary antibodies. Primary antibodies used were mouse anti-H2A.X (1:500, Millipore) and rabbit anti-SCP3 (1:1000, Abcam). Secondary antibodies used were goat anti-mouse Alexa 488 (green dye) (1:200, Sigma-Aldrich) and goat anti-rabbit Alexa 594 (red dye) (1:500, Sigma-Aldrich), respectively. The slides were counterstained with DAPI (Invitrogen) and mounted with Mowiol (Merck Biosciences Ltd). Pictures were taken using an AxioCam MR Rev3 camera on an Axio Observer.Z1 microscope with Apotome (Carl Zeiss). The images were processed using AxioVision 4.8 software (Carl Zeiss) and Image J.

mRNA Sequencing and Computational Analysis

Large RNAs >200 nt were isolated from C57BL6/J and Alkbh1-/- pachytene cells using the mirVana miRNA Isolation Kit (Ambion) in line with the manufacturer's protocol. Any DNA remnants were removed from the RNA using TURBO DNase (Ambion) before mRNA isolation was performed. One µg RNA from wild-type and 1.6 µg RNA from Alkbh1-/- were diluted to 50 µL, and mRNA was isolated using the Sera-Mag® Magnetic Oligo(dT) Particles (Thermo Scientific part#1004815) according to the manufacturer's protocol. From here, the mRNA-Seq Sample Preparation Guide (Illumina part # 1004898, Rev D September 2009) was used to perform ligation of 5' and 3' adapters to the cDNA and purification of the ligation product. cDNA template was enriched by PCR amplification (15 cycles) using the mRNA Seq 8 Sample Prep Kit (RS-100-0801, Illumina). Finally, the mRNA library (230-233 bp) was sequenced (75 bp Paired-End Read) using the Illumina Genome Analyzer IIx (GAIIx).

Mapping of sequence reads to the genome was done using TopHat (48). The output from TopHat was then converted to BED-format using a script from Ensembl (ftp://193.62.203.113/pub/ensembl-functgenomics/scripts/miscellaneous/sam2bed.pl). These reads were then intersected with all known mouse genes from Ensembl using intersectBed. Similar to the analysis of the piRNAs, reads were counted for each annotated gene. Reads that were annotated to multiple genes were discarded. The DESeq package was used for statistical analysis of differential expression, similar to the method used for the piRNA clusters.

REFERENCES:

50. Trapnell,C., Pachter,L. and Salzberg,S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics., 25, 1105-1111

SUPPLEMENTARY FIGURE AND TABLE LEGENDS

Figure S1. Tzfp is the interaction partner of Alkbh1 in testis. (A) Human TZFP sequences identified by a yeast 2-hybrid (Y2H) screen with testis-specific transcripts against full-length human ALKBH1 as bait. The screen yielded four positive clones, and sequencing analysis revealed that three of these clones represented the same gene (left panel). All three clones were assayed for activity by a second reporter gene, lacZ, using a quantitative β-galactosidase assay (right panel). (B) Plasmid control digest was performed on the three clones in two independent E. coli transformants. (C) Verification of interaction between human ALKBH1 and TZFP by Dot blot analysis. ALKBH1 protein was dotted onto a nitrocellulose membrane before the membrane was incubated with full-length or truncated TZFP-Myc. Positive control, full length TZFP-Myc; negative control, BSA.

Figure S2. Purity analysis of STAPUT isolated pachytene spermatocytes. (A) Image analysis of STAPUT purified pachytene cells reveals a homogenous population of cells. One isolation from six males yielded approximately 1.5×10^6 pachytene cells with an average size of 12.5 µm. Analysed on the Countess® Automated Cell Counter (Invitrogen). (B) Specific expression of *Lcn2* in pachytene spermatocytes confirms high purity of STAPUT isolated cell fractions. Analysed by SYBR Green qPCR of pachytene spermatocytes (PS) and round spermatids (RSd). (C) The meiotic pachytene cells are easily detected by double immunostaining. H2A.X marks the transcriptionally silenced XY chromatin domain (red), SYCP3 stains the chromosome axial elements of the synaptonemal complex (green), and DAPI labels DNA (blue). (Magnification: ×20).

Figure S3. Significantly up- and downregulated piRNA clusters in pachytene cells lacking Alkbh1 and Tzfp. Identical to scatterplot analysis in Fig. 4 except that all piRNA clusters significantly differentially expressed in wild-type versus mutants are presented with coloured, filled-circles. Expression of piRNA clusters in pachytene cells are plotted on a log2 scale in a pairwise comparison of wild-type (y-axis) and *Alkbh1*$^{-/-}$ (x-axis) (A) and of wild-type (y-axis) and *Tzfp*$^{GTi/GTi}$ (x-axis) (B). Cluster 1082B is the piRNA cluster deviating most dramatically from the center line in both mutant samples (black, filled-circle). In contrast, clusters with similar number of reads from the two samples being compared line up together on the center line (grey, open-circles).

Figure S4. Significantly upregulated mRNAs in pachytene cells from wild-type and *Alkbh1*$^{-/-}$ testes. (A) Scatterplot analysis of mRNA reads annotated by high-throughput sequencing. Expression of mRNAs in pachytene cells are plotted on a log2 scale in wild-type (y-axis) compared to *Alkbh1*$^{-/-}$ (x-axis). mRNAs with notably higher number of reads in the *Alkbh1*$^{-/-}$ sample than in the wild-type sample deviate from the center line (coloured, filled-circles). In contrast, mRNAs with similar number of reads from the two samples being compared line up

together on the center line (grey, open-circles). (B) mRNAs upregulated more than 4-fold in *Alkbh1*$^{-/-}$ versus wild-type pachytene cells are listed.

Table S1. Sequence of primers and oligonucleotides used in this study. Primers for genotyping of the *Alkbh1* gene (1-3) and the *Tzfp* gene (4-6), Taqman probes (7-10), miScript primers for the quantification of individual piRNAs (11-15), DIG-labeled oligonucleotide specific for piRNA used in FISH (16), and SYBR Green qPCR primers for expression analyses of transposons (17-24), the pachytene marker *Lcn2* (25-26), and the endogenous control *Gapdh* (27-28) are listed.

Table S2. Classification of small RNA sequences after Illumina sequencing of wild-type, *Alkbh1*$^{-/-}$ and *Tzfp*$^{GTi/GTi}$ pachytene cells. Number of sequence reads mapping to each category in wild-type, *Alkbh1*-/- and *Tzfp*$^{GTi/GTi}$ pachytene cells from 12-week old males. We did two separate runs with a 1-year interval between the mutants and corresponding wild-type samples. Small RNA reads were annotated as described in the Methods section.

Table S3. Significantly up- and downregulated piRNA clusters in *Alkbh1*$^{-/-}$ and *Tzfp*$^{GTi/GTi}$ samples versus wild-type. Number of piRNA cluster reads in wild-type, *Alkbh1*-/- and *Tzfp*$^{GTi/GTi}$ pachytene cells. Cluster name is the ID assigned by piRNABank (http://pirnabank.ibab.in) or by piRNAdb (http://kbrb.ioz.ac.cn/piRNA). The logfold change is calculated using log2 (exprKO+1) - log2 (exprWT+1), and the numbers are normalized against the total number of mapped reads for each sample. Cluster 1082B, the only piRNA cluster highly upregulated in both mutants, is presented with grey shading.

Table S4. Significantly differentially expressed genes found in *Alkbh1*-null versus wild-type pachytene cells. Number of mRNA sequence reads in wild-type and *Alkbh1*-/- pachytene cells from 12-week old males. Gene ID is the Ensembl accession number. The logfold change is calculated using log2 (exprKO+1) - log2 (exprWT+1), and the numbers are normalized against the total number of mapped reads for each sample.

Table S5. Potential target regions in the mouse genome of piRNAs from cluster 1082B identified by sequence similarity. The chromosomal location, direction and length of potential target regions for piRNAs derived from cluster 1082B.

Figure S1

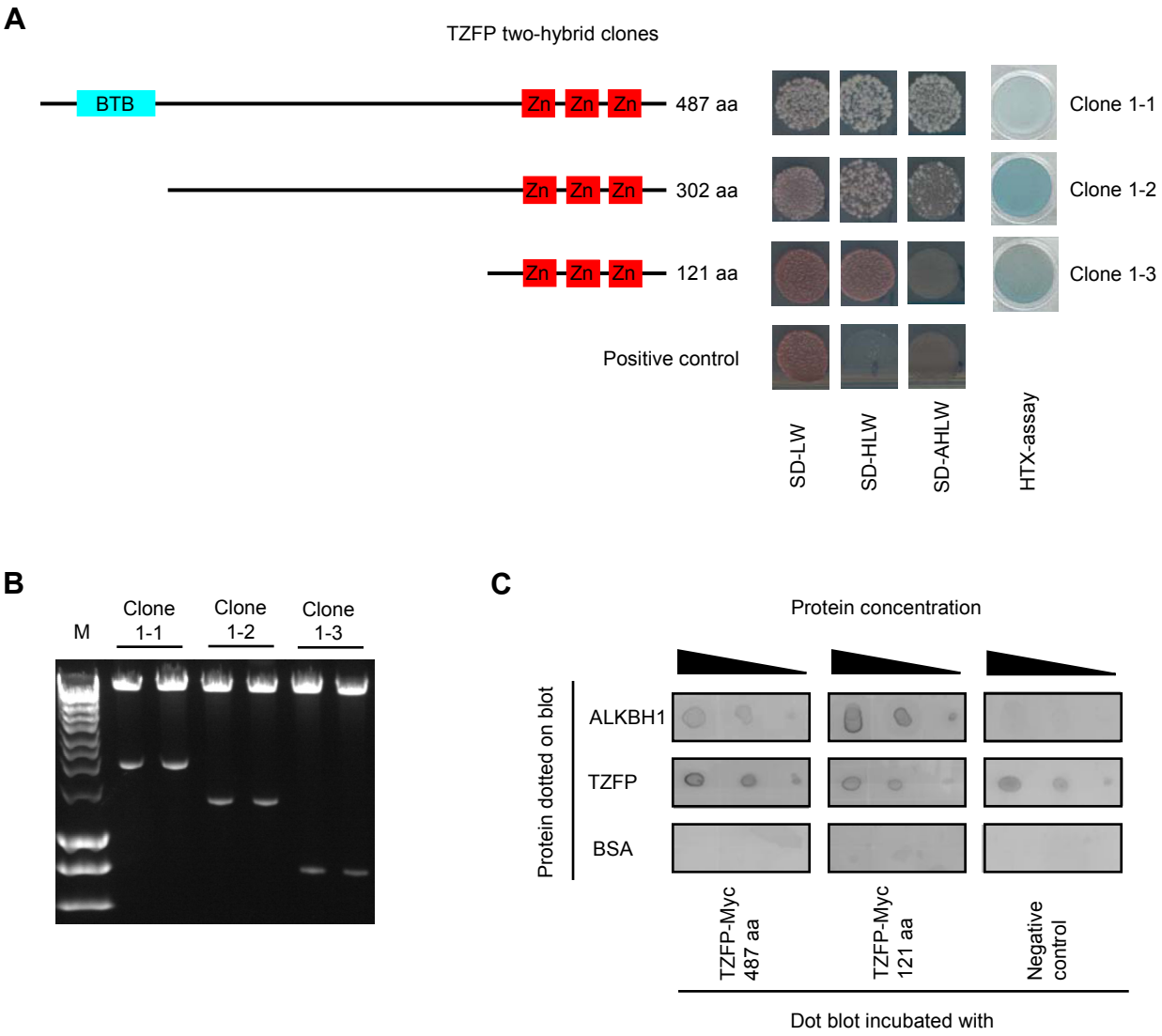
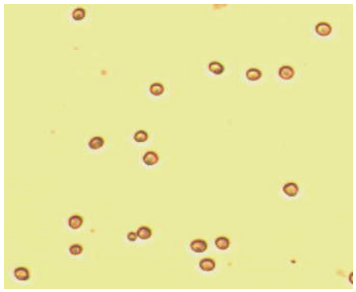

A
TZFP two-hybrid clones
BTB
Zn
Zn
Zn
487 aa
302 aa
121 aa
Positive control
Clone 1-1
Clone 1-2
Clone 1-3
SD-LW
SD-HLW
SD-AHLW
HTX-assay
B
M
Clone 1-1
Clone 1-2
Clone 1-3
C
Protein concentration
Protein dotted on blot
ALKBH1
TZFP
BSA
TZFP-Myc 487 aa
TZFP-Myc 121 aa
Negative control
Dot blot incubated with

Figure S2

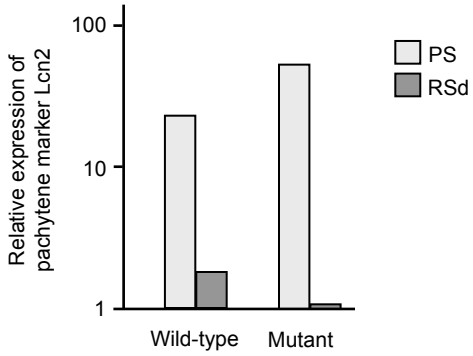
A



~1.5 x 10^6 pachytene cells

$\bar{X}_{size}$= 12.5 µm

B



C

STAPUT isolated pachytene cells

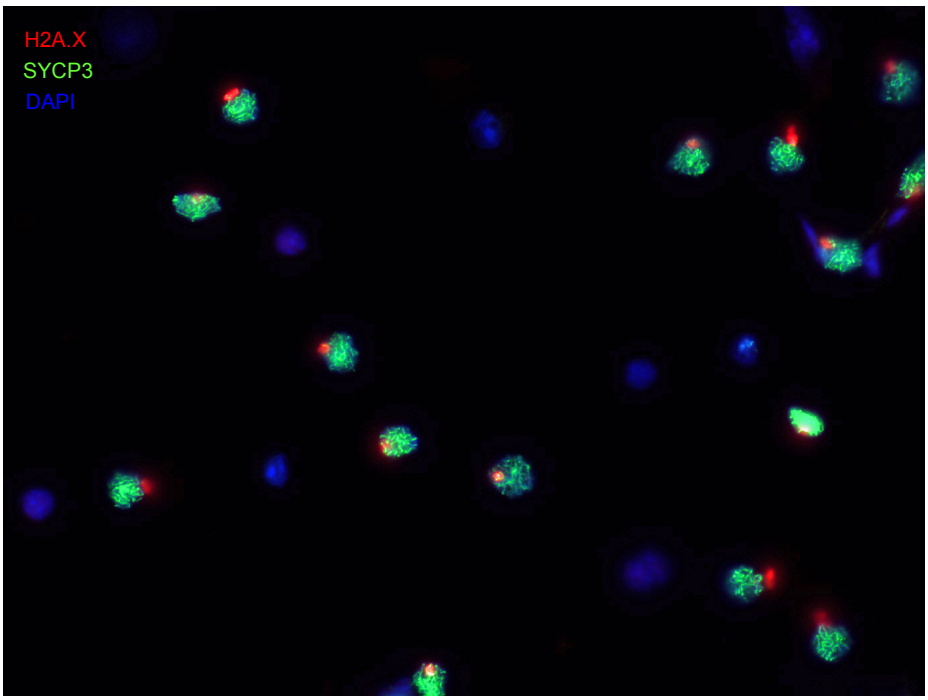


Figure S3

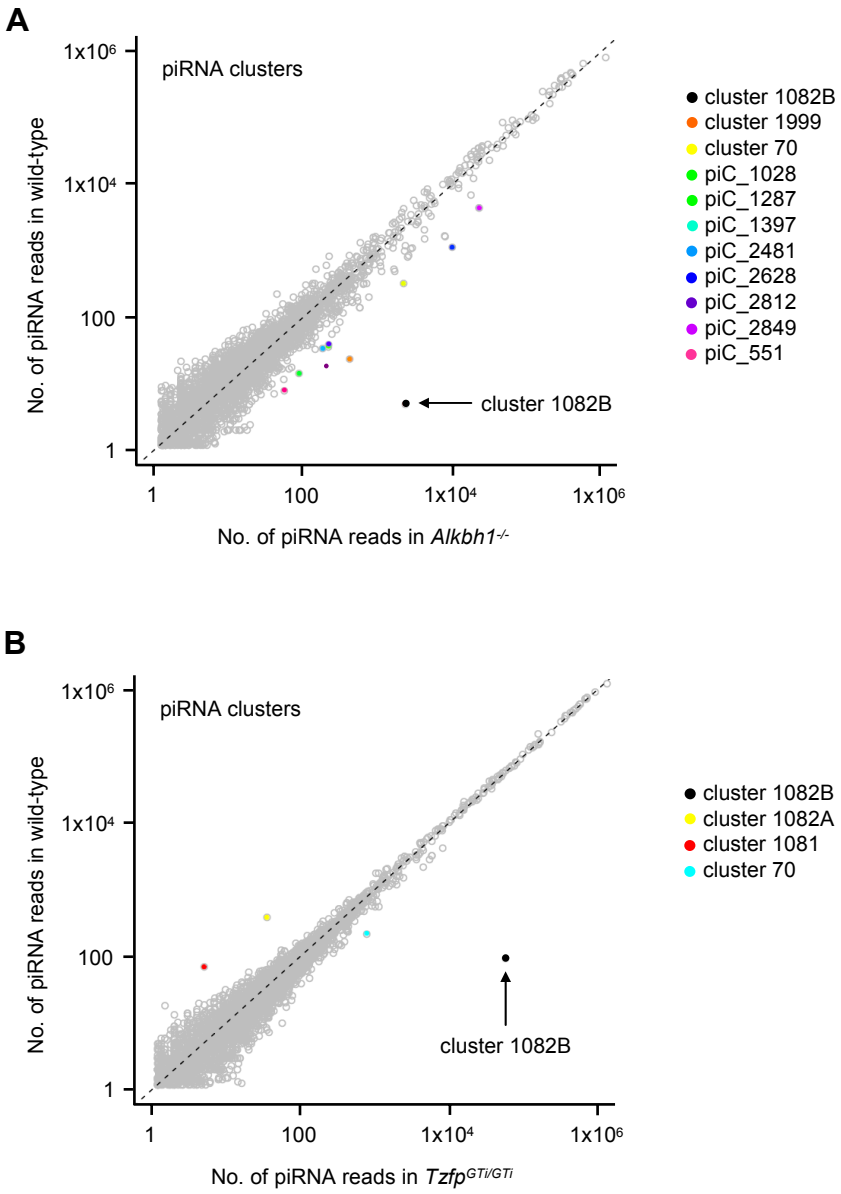
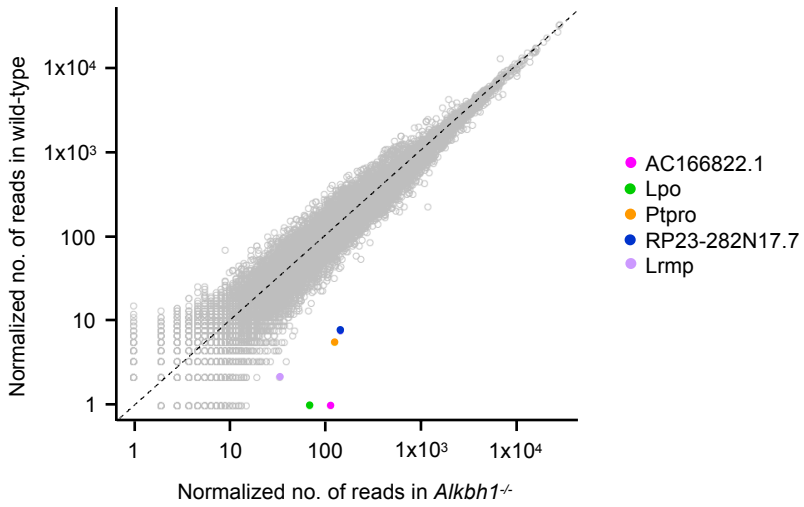

A
piRNA clusters
No. of piRNA reads in wild-type
No. of piRNA reads in Alkbh1-/-
1x10⁶
1x10⁴
100
1
cluster 1082B
cluster 1999
cluster 70
piC_1028
piC_1287
piC_1397
piC_2481
piC_2628
piC_2812
piC_2849
piC_551
B
piRNA clusters
No. of piRNA reads in wild-type
No. of piRNA reads in TzfpGTi/GTi
cluster 1082B
cluster 1082A
cluster 1081
cluster 70

Figure S4

A



B

Gene Name	Gene Symbol	Relative Change
Ras-like guanine nucleotide exchange factor domain protein	AC166822.1	7.12
Lactoperoxidase	Lpo	6.38
Protein tyrosine phosphatase, receptor type, O	Ptpro	4.70
Krueppel-associated box (KRAB) containing protein	RP23-282N17.7	4.40
Lymphoid-restricted membrane protein	Lrmp	4.12

Table S1

Methods	Sequence (5'–3')	Details
PCR genotyping	1. AGTTATCAGGGCCATCCAGGGAGGT 2. AACTGAGAGGTACAGGGAAGCATAA 3. GCTTGCCGAATATCATGGTG 4. GCTCAACAAGTCAAGACTTT 5. ACTGTGGCAGACTAATACTT 6. CTTGCAAAATGGCGTTACTTAAGC	*Alkbh1* WT allele *Alkbh1* WT + KO allele *Alkbh1* KO allele *Tzfp* WT allele *Tzfp* WT + GTi allele *Tzfp* GTi allele
TaqMan qPCR	7. Mm01296827_m1 8. Mm00491292_g1 9. Mm01277432_m1 10. Mm99999915_g1	*Alkbh1* probe *Tzfp* probe *4933440M02Rik* probe *Gapdh* probe
miScript Primer Assay	11. UGGAACUCACUUUGUAGACCAGGCUGGCCU 12. UCUCAGUCCGUACCAGGUAGAGACCAGAUC 13. GAGGCUGUGGAGACAAUCACAAAUGACAGA 14. UGCCACAUGGGAGCUUCCUUCAAUAUGACU 15. UACAUUGACUGAUUCACUUGGUUUGUCCC	(1) piR-19852 (2) piR-12359 (3) piR-103121 (4) piR-17918 (5) piR-4749
piRNA FISH	16. GATCTGGTCTCTACCTGGTACGGACTGAGA-DIG	(2) piR-12359
SYBR Green qPCR Transposons	17. GGCGAAAGGCAAACGTAAGA 18. GGAGTGCTGCGTTCTGATGA 19. GGAGGGACATTTCATTCTCATCA 20. GCTGCTCTTGTATTTGGAGCATAGA 21. GCACATGCGCAGATTATTTGTT 22. CCACATTCGCCGTTACAAGAT 23. AACCAATGCTAATTTCACCTTGGT 24. GCCAATCAGCAGGCGTTAGT	*LINE1 5'UTR* forward primer *LINE1 5'UTR* reverse primer *LINE1 ORF2* forward primer *LINE1 ORF2* reverse primer *IAP 3'LTR* forward primer *IAP 3'LTR* reverse primer *IAP Gag* forward primer *IAP Gag* reverse primer
SYBR Green qPCR Pachytene marker	25. CCATCTTTCCTGTTGCCAGAG 26. GACGCCATTGGTGGTGTTAAG 27. TCGTCCCGTAGACAAAATGGT 28. CGCCCAATACGGCCAAA	*Lcn2* forward primer *Lcn2* reverse primer *Gapdh* forward primer *Gapdh* reverse primer

Table S2

	Wild-type	***Alkbh1*$^{-/-}$**	**Wild-type**	***Tzfp*$^{GTi/GTi}$**	
Total # mapped	10504223	8072816	15565882	15259047	
Non-coding RNA classes	reads	**reads**	**reads**	**reads**	**SUM**
piRNA	9761468	6985408	14653799	14125026	**45525701**
miRNA	157859	518065	195942	348381	**1220247**
rRNA	9513	7154	1016	1294	**18977**
tRNA	4525	6142	3631	4434	**18732**
snoRNA	10607	6642	3230	5172	**25651**
snRNA	688	932	492	579	**2691**
lincRNA	4749	5062	6724	6704	**23239**
miscRNA	152	142	108	122	**524**
Unknown	332716	318246	423160	468123	**1542245**
SUM (mapped <6)	**10282277**	**7847793**	**15288102**	**14959835**	**48378007**
Mapped > 5	221946	225023	277780	299212	**1023961**
% Mapped	79%	75%	78%	76%	49401968
Repeats	**reads**	**reads**	**reads**	**reads**	**SUM**
LTR	768396	498372	1144105	1094012	**3504885**
LINE	457412	274246	662868	618912	**2013438**
SINE	349614	248886	528862	511532	**1638894**
Other	221024	208632	372079	319968	**1121703**
SUM	**1796446**	**1230136**	**2707914**	**2544424**	**8278920**
SUM regulars	1575422	1021504	2335835	2224456	7157217
Non-repeats	8707777	6842680	12857968	12714623	**41123048**
Non-repeats (mapped <6)	8485831	6617657	12580188	12415411	40099087
Mappings to regions	**reads**	**reads**	**reads**	**reads**	**SUM**
Exonic	586034	822064	851772	999077	**3258947**
Intronic	313159	288309	419694	427714	**1448876**
5' UTR	124261	155553	154737	176812	**611363**
3' UTR	9489	15013	9688	14198	**48388**
Intergenic	9471280	6791877	14129991	13641246	**44034394**
SUM	**10504223**	**8072816**	**15565882**	**15259047**	**49401968**
SUM (mapped <6)	9249334	6566854	13852211	13342034	43010433

Cluster Name	Wild-type reads	*Alkbh1*$^{-/-}$ reads	Logfold Change	Wild-type reads	*Tzfp*$^{GTi/GTi}$ reads	Logfold Change
cluster 1082B	4	1858	8.86	94	57188	9.25
cluster 1081	36	15	-1.26	69	4	-4.11
cluster 1999	23	330	3.87	17	63	1.92
cluster 1082A	258	199	-0.37	383	34	-3.5
piC_2628	1165	7923	2.77	4279	8156	0.93
piC_551	7	43	2.62	8	25	1.64
cluster 70	326	1755	2.43	218	768	1.82
piC_1287	14	68	2.33	24	21	-0.21
piC_1028	36	170	2.26	80	126	0.65
piC_2812	39	172	2.14	150	248	0.72
piC_2481	34	142	2.09	127	242	0.93
piC_1397	4586	18280	2	3264	5816	0.83
piC_2849	4586	18280	2	3265	5816	0.83

Gene ID	Wild-type reads	*Alkbh1*$^{-/-}$ reads	Logfold Change	Gene Symbol
ENSMUSG00000072616	0	138	7.12	AC166822.1
ENSMUSG00000009356	0	82	6.38	Lpo
ENSMUSG00000030223	4	154	4.7	Ptpro
ENSMUSG00000078503	6	178	4.4	RP23-282N17.7
ENSMUSG00000030263	1	38	4.12	Lrmp
ENSMUSG00000056220	65	9	-3.03	Pla2g4a
ENSMUSG00000009621	226	1611	2.51	Vav2
ENSMUSG00000029369	50	299	2.25	Afm
ENSMUSG00000041536	414	145	-1.82	Serpina3a
ENSMUSG00000000567	383	139	-1.77	Sox9
ENSMUSG00000037432	195	786	1.69	Fer1l5
ENSMUSG00000041245	531	209	-1.66	Wnk3-ps
ENSMUSG00000079173	228	889	1.65	Zan
ENSMUSG00000045330	548	220	-1.63	4933402E13Rik
ENSMUSG00000049775	626	271	-1.52	Tmsb4x
ENSMUSG00000038199	319	1097	1.47	4931409K22Rik
ENSMUSG00000040354	435	1483	1.45	Mars
ENSMUSG00000068522	1449	696	-1.37	Aard
ENSMUSG00000027199	1693	815	-1.37	Gatm
ENSMUSG00000034219	1155	559	-1.36	Sept14
ENSMUSG00000030605	1119	543	-1.36	Mfge8
ENSMUSG00000022390	1673	824	-1.34	Zc3h7b
ENSMUSG00000059248	1154	589	-1.28	Sept9
ENSMUSG00000039428	1023	525	-1.28	Tmem135
ENSMUSG00000027445	2150	1146	-1.22	Cst9
ENSMUSG00000021477	4746	2704	-1.13	Ctsl
ENSMUSG00000037625	1561	901	-1.11	Cldn11
ENSMUSG00000036687	2152	1274	-1.07	Tmem184a
ENSMUSG00000040260	2714	1612	-1.07	Daam2
ENSMUSG00000033161	2554	1537	-1.05	Atp1a1
ENSMUSG00000025151	2526	1532	-1.04	Maged1
ENSMUSG00000023047	3735	2304	-1.01	Amhr2
ENSMUSG00000022037	15030	9895	-0.92	Clu
ENSMUSG00000004207	4096	2930	-0.8	Psap

Cluster 1082B		Mouse genome (NCBI 37/mm9)			
Start	**End**	**Chr**	**Strand**	**Start**	**End**
1292	2078	3	-	20402645	20401882
1375	2031	6	+	121678236	121678932
1377	2031	6	+	121780705	121781412
2176	2645	6	+	121781344	121781828
2218	2739	6	+	121945714	121946245
2229	2641	6	+	122083487	122083901
2218	2739	6	+	122114964	122115495
2248	2538	11	-	21404638	21404353
260	1301	11	-	21407654	21406600
2	314	11	-	21407935	21407636
4098	4210	12	-	55274702	55274590
2792	3839	12	-	55276549	55275477
1354	2667	13	-	8929370	8928049
890	1217	13	-	8930278	8929945
2	1276	13	+	49859505	49860887
2797	3338	13	+	49940070	49940613
1244	2737	13	-	50160392	50158851
2	908	13	-	50510202	50509269
546	1150	13	+	52212995	52213592
1623	2452	13	+	52214343	52215179
1293	1937	13	-	52519668	52518988
1563	2775	13	-	52528902	52527636
1318	2086	16	-	16710254	16709448
260	1203	17	-	54051739	54050782
260	1246	17	-	54088485	54087479
1769	2060	18	+	6684413	6684697