PNAS
www.pnas.org

Supplementary Information for

The *Rhox* gene cluster suppresses germline *LINE1* transposition

Kun Tan, Matthew E. Kim, Hye-Won Song, David Skarbrevik, Eric Babajanian, Tracy A. Bedrosian, Fred H. Gage, Miles F. Wilkinson

Miles F. Wilkinson
Email: mfwilkinson@health.ucsd.edu

This PDF file includes:

Supplementary text
Figures S1 and S2
Table S1
SI References

Other supplementary materials for this manuscript include the following:

Dataset S1

Supplementary Information Text

SI Materials and Methods

Mammalian cell culture and plasmids

NIH/3T3, GC1, and HEK293T cells were grown in DMEM (Gibco), 10% fetal bovine serum (Sigma), and 1× penicillin/streptomycin (Gibco). TCam-2 cells were grown in advanced RPMI 1640 medium (Gibco), 10% fetal calf serum, 2 mM L-glutamine, and 1× penicillin/streptomycin (Gibco). GS cells were cultured in IMDM/SFM, as described previously (1). All cells were cultured in a humidified 5% CO_2, 37°C incubator.

For transient transfection experiments, the coding sequences of mouse *Rhox* and human *RHOX* genes were cloned into the pcDNA 5/FRT expression vector (Invitrogen). The *Piwil2* coding sequence was cloned into pcDNA 3.1 (-) expression vector. *RHOXF1* and *RHOXF2* shRNA plasmids were generated by inserting siRNA sequences (*RHOXF1*: GGGTTTGGTTTAAGAATAA; *RHOXF2*: GGGCATTAATGGCAAGAAA) into the pLLU2G backbone. Mouse (*ORFeus*) and human ($L1_{RP}$) LINE1 reporter constructs were generated as previously described (2). The mutant *RHOXF2* constructs were obtained from two previous studies (3, 4), and cloned into the pIRES-hrGFP expression vector. The *Piwil2* and *Tdrd1* promoters were separately cloned into the pGL3 vector (Promega).

Transfections and luciferase analysis

HEK-293T cells were transfected with Lipofectamine 2000 (Invitrogen). GC1, NIH/3T3, and TCam-2 cells were transfected with Lipofectamine 3000 (Invitrogen). GS cells were transfected with the Lipofectamine stem reagent (Invitrogen). For transfection experiments, most cells were trypsinized and seeded in 24-well plates at a density of ~70,000 cells per well and grown up to 60-70% confluence. TCam-2 cells were seeded at ~20,000 cells/well. All transfections were performed following the manufacturer's instructions. Cells were harvested 1-day post-transfection for luciferase activity analysis. Luminescence was measured using the Dual-Luciferase Reporter assay system (Promega) following the manufacturer's instruction. Statistical significance was determined using the paired Student's t-test.

Chromatin immunoprecipitation (ChIP) analysis

GS cells (5 million) were dissociated with 0.05% Trypsin-EDTA (Gibco), washed twice with PBS, and cross-linked with 1% formaldehyde (37°C for 10 min). Glycine (0.125 M final concentration) was added to quench the formaldehyde. After washing twice with ice-cold PBS plus 1x Protease Inhibitors (Roche) and 1 mM PMSF, the cells were resuspended in 500 µL of lysis buffer (1% sodium dodecyl sulfate [SDS], 50 mM Tris-HCl pH8.0, 10 mM EDTA, 1x Protease Inhibitors

(Roche), 1 mM PMSF) and incubated on ice for 10 min. Samples were sonicated for 20 min in a Diagenod Bioruptor (30 s interval, setting H). The sonicated lysate was centrifuged at 13,200 rpm for 10 min at 4°C and the supernatant was collected into a new tube. 50 µL of the supernatant was saved as input. 15 µL Protein A Dynabeads (Invitrogen) was washed with Dilution Buffer (10 mM Tris-HCl pH7.5, 0.5 mM EGTA, 140 mM NaCl, 0.1% SDS, 1% Triton X-100, 1 mM EDTA) and resuspended in 100 µL Dilution Buffer and added to the samples. Samples were split into two parts: one half for Rabbit-anti-HA (Abcam, ab9110), and the other half for Rabbit-IgG (Abcam, ab171870) as a control. Samples were incubated at 4°C overnight. On the second day, samples were placed on a magnetic rack to remove supernatant and the beads were washed twice with Dilution Buffer, and eluted with Elution Buffer (1% SDS, 20 mM Tris-HCl pH7.5, 5 mM EDTA, 50 mM NaCl, 200 µg/mL Proteinase K) in a thermomixer at 1300 rpm for 2 hr at 68°C. The samples were purified with MinElute PCR Purification Kit (Qiagen). qPCR was performed as described below. The primers for ChIP-qPCR are listed in *SI Appendix*, Table S1. The ChIP data represent IP values for each region relative to the input from three independent replicates.

qRT-PCR analysis

Total cellular RNA was isolated using TRIzol (Invitrogen), as previously described (5). Reverse transcription-PCR analysis was performed using 1 µg of total cellular RNA using iScript™ cDNA synthesis kit (Bio-Rad), followed by PCR amplification using SYBR Green (Bio-Rad) and the ΔΔCt method (with ribosomal L19 for normalization). The primers used are listed in *SI Appendix*, Table S1. Results were from at least three independent replicates. Statistical significance was determined using the paired Student's t-test.

TaqMan assay

Oct4-eGFP+ germ cells were purified by FACS. Total cellular RNA was isolated using TRIzol (Invitrogen) and DNA was removed using Ambion® TURBO DNA free kit (ThermoFisher Scientific). Reverse transcription was performed using TaqMan® microRNA reverse transcription kit (ThermoFisher Scientific). TaqMan-qPCR was performed (in triplicate for each sample) using TaqMan® fast advanced master mix (ThermoFisher Scientific). For each biological replicate, testes from three pups were pooled. Results were from three independent replicates. Statistical significance was determined using the paired Student's t-test.

Western blotting analysis

Testes tissue or cell lines were incubated in radioimmunoprecipitation assay (RIPA) buffer (Bio-Rad) supplemented with protease inhibitor cocktail (Sigma) on ice for 30 min, followed by centrifugation at 16,000 g for 15 min at 4°C. The lysates were then transferred to new tubes, and protein level was quantified using the DC Protein Assay kit (Bio-Rad). 10-20 µg of the protein

samples were separated on an 12% polyacrylamide gel, and Western blot analysis was performed as previously described (6). Quantification of the blots was performed using NIH ImageJ (1.8.0). Statistical significance was determined using the paired Student's t-test.

Testicular cell dissociation

Single testicular cells were isolated from fetal testes using a two-step enzymatic digestion protocol described previously (7), with minor modifications. In brief, testicular tissue was mechanically disrupted and enzymatically digested with 1 mg/ml collagenase type IV (Worthington Biochemical) in Hanks Balanced Salt Solution (HBSS; GIBCO) at 37 °C. The tubules were sedimented and washed with HBSS and digested in 0.25% Trypsin-EDTA (ThermoFisher) and Deoxyribonuclease I (Worthington Biochemical). The suspension was triturated vigorously ten times and incubated at 37°C for 5 min, followed by repeat trituration and incubation. The digestion was stopped by adding the same volume of αMEM + 10% FBS medium and the cells were size-filtered through 70 μm strainers (ThermoFisher) and pelleted by centrifugation at 300 g for 5 min.

RNAseq analysis

Single cells were isolated as described above from *Rhox10*-null;Oct4-eGFP$^{+/+}$ and littermate Oct4-eGFP$^{+/+}$ (control) mice, and the eGFP+ cells were purified using fluorescence-activated cell sorting (FACS). For each sample analyzed, testes from three to four fetuses were pooled. RNAseq was performed as described previously (8). Total RNA was extracted using the RNeasy Plus Micro Kit (Qiagen), following the manufacturer's protocol. 1 ng of total RNA was used to make the library using the SMARTer® stranded total RNAseq kit v2 - pico input system (Takara Bio USA), per manufacturer's instructions. Libraries were sequenced (single-end reads) with an Illumina HiSeq 4000 platform for 75 cycles at the UCSD IGM core. The average number of reads per sample ranged from approximately 38 to 46 million.

Reads were filtered for quality and aligned with STAR (2.5.2b) (9) against *Mus musculus*, release-101, Ensembl genome (GRCm38). The exon counts were aggregated for each gene to build a read count table using SubRead function featureCounts (10). DEGs were defined using DESeq2 (11). The R package program "pheatmap" was used for clustering and to generate heatmap plots. The database for annotation, visualization and integrated discovery (DAVID) v6.8 was used for signaling pathway analysis (12). The search tool for retrieval of interacting genes (STRING) v11.0 was used for the analysis of gene molecular interactions (13, 14).

dd-PCR

dd-PCR reactions were performed on 40–60 ng of genomic DNA that were duplexed with EGFP and HPRT TaqMan probes (15). PCR cycling conditions: 95 °C for 10 min, 40 cycles of 94 °C for

30 s and 60 °C for 1 min, 98 °C for 10 min. Samples were analyzed in Bio-Rad QX200 Droplet Reader using QuantaSoft software v1.3.2.0.

Bisulfite sequencing analysis

Testes from six fetuses were pooled together in each group. EpCAM+ germ cells were purified using the MACS system by incubating dissociated fetal germ cells for 20 mins on ice in MACS buffer (PBS, 0.5% BSA, 2 mM EDTA) + Deoxyribonuclease I + EpCAM antibody (Developmental Studies Hybridoma Bank; 1:50). The cells were then washed with MACS buffer, incubated with anti-rat IgG microbeads for 20 mins on ice, washed in MACS buffer, re-suspended in 500 µl MACS buffer, run through a MACS MS column (Miltenyi Biotec, cat # 130-042-201), and eluted with MACS buffer.

Cells were processed for methylation analysis using the EpiTect Plus LyseAll Bisulfite Kit (Qiagen). Bisulfite treatment of fetal DNA was conducted with the EpiTect Plus DNA Bisulfite Kit (Qiagen). PCR amplification, gel extraction, TA cloning, and sequence analysis were performed as previously described (15). Statistical significance was determined using the Fisher's exact test from QUMA (http://quma.cdb.riken.jp/).

Immunofluorescence analysis

Testes were fixed in Bouin’s fixative for 0.5 h at room temperature, then cleared using multiple changes of 70% ethanol and stored at 4 °C. Tissues were dehydrated through an ethanol series (30%, 50%, 70%, 95%, and 100% twice) for 10 min at each step at room temperature, cleared with two changes of xylene, and embedded in paraffin. Tissue blocks were sectioned at 5 µm. As previously described (16, 17), sections were deparaffinized twice in xylene, followed by serial dilutions of ethanol. Unmasking was performed using a steamer (IHC World). Blocking was performed by incubating with 5% donkey serum (Sigma) at room temperature. The sections were then incubated overnight with the primary antibody (ORF1p rabbit monoclonal antibody [Abcam #ab216324; 1:200] or TRA98 rat monoclonal antibody [B-Bridge International; 1:200]) at 4°C and incubated with secondary antisera (Donkey anti-rabbit IgG (H+L) highly cross-adsorbed secondary antibody, or Alexa Fluor 488; Donkey anti-rat IgG (H+L) highly cross-adsorbed secondary antibody, Alexa Fluor 594) for 1 h at room temperature. The nuclei were counterstained with DAPI (Vector Laboratories), a coverslip was placed over the sections with mounting medium, and the images were viewed using a Leica DMI4000 B fluorescence microscope.

Statistical analysis

The details of the statistical method used for identifying the differential gene expression are provided in the detailed methods above. For luciferase, qPCR, ddPCR, and western blotting analyses, statistical significance was determined using the paired Student's t-test. For bisulfite

sequencing analysis, statistical significance was determined using the Fisher's exact test from QUMA (http://quma.cdb.riken.jp/).

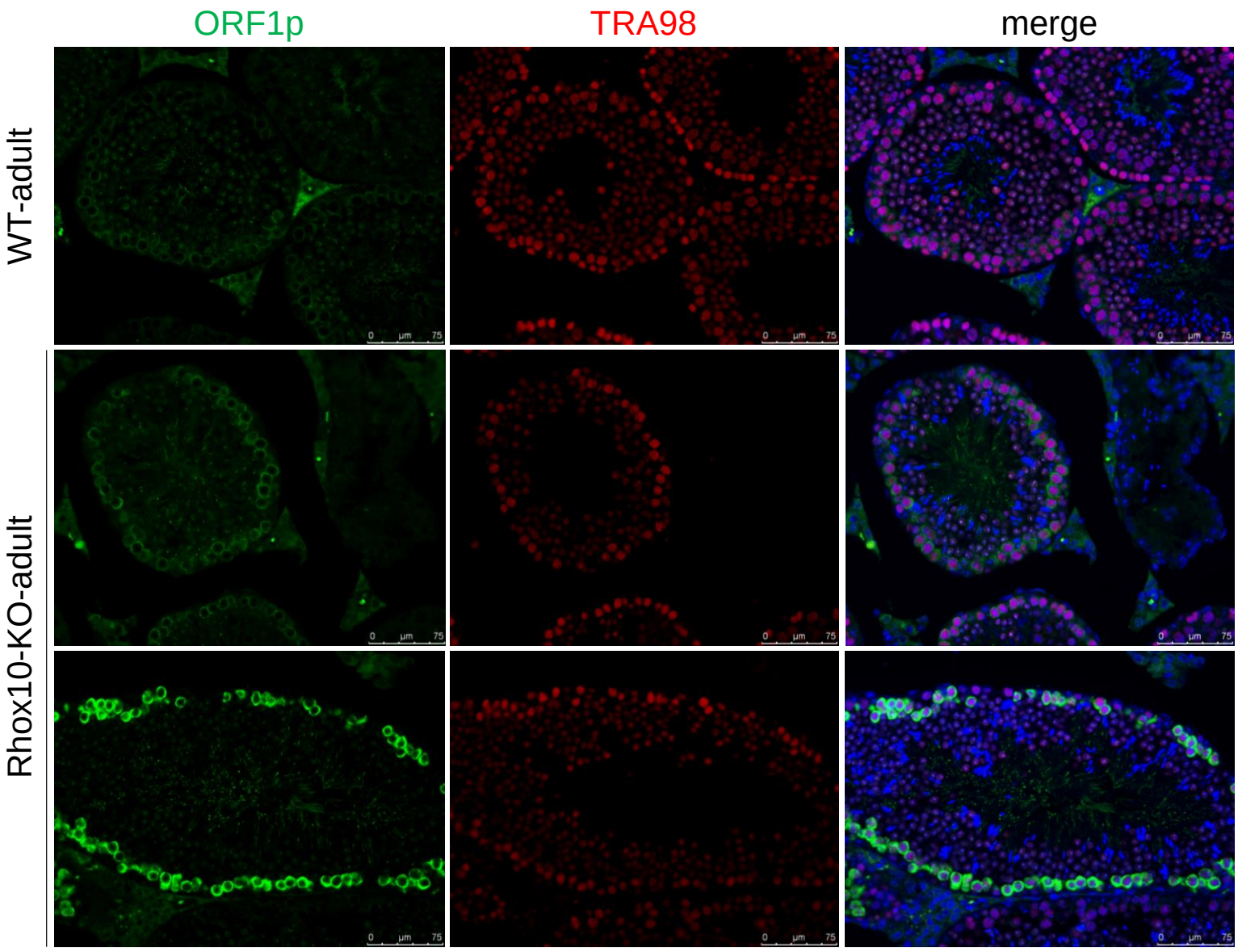
ORF1p
TRA98
merge
WT-adult
Rhox10-KO-adult

Fig. S1 Immunofluorescence analysis of ORF1p (green) in *Rhox10*-null (KO) and littermate control (WT) testes from adult mice. The sections were co-stained with an antibody against TRA98 to detect all germ cells (red) and DAPI to detect all cells (blue). (n=3).

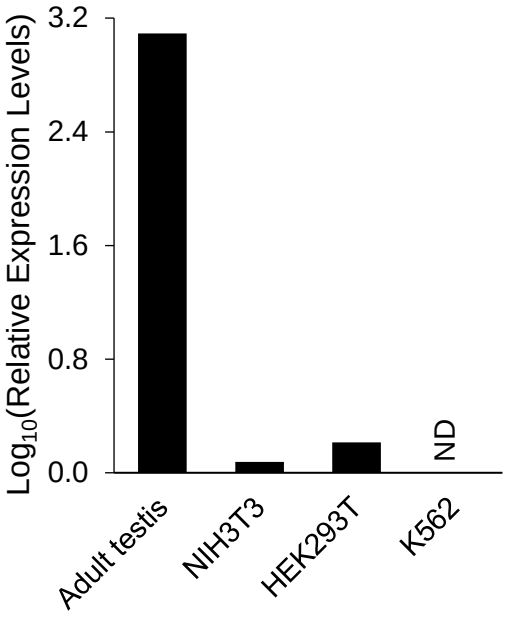
Log10(Relative Expression Levels)
3.2
2.4
1.6
0.8
0.0
ND
Adult testis
NIH3T3
HEK293T
K562

Fig. S2 qPCR analysis of *PIWIL2*/*Piwil2* mRNA expression in somatic cell lines. *RPL19*/*Rpl19* was used for normalization. Mouse adult testes is the positive control. The expression of *Piwil2* mRNA in NIH3T3 cells was set to 1; all values are Log10-transformed.

Table S1. Primers used in this study

Symbol	Forward	Reverse
ChIP-qPCR primers		
Piwil2#1	CTAGCCGTCCACATCTGAG	TAGACTGCGTGAAAGGGTG
Piwil2#2	TGGAGCCTTGCCTAGATTTC	GCGAGACTCAGATGTGGAC
Piwil2#3	GCTGGATCTCATGGAGGC	GTAAGGGTTGAGGCTTAAAGT
qPCR primers		
Piwil2	CTGTCTCCAACTTCACTGCAC	CTCATGCTTTTGCATTCCACAC
Tdrd1	TCTACAGAGCTCAGGAAAGCAG	TGCTCTCAACTGTCACTCTTC
Rpl19	CCTGAAGGTCAAAGGGAATGTG	CTTTCGTGCTTCCTTGGTCTT
L1a	GGATTCCACACGTGATCCTAA	TCCTCTATGAGCAGACCTGGA
L1t	CAGCGGTCGCCATCTTG	CACCCTCTCACCTGTTCAGACTAA
L1gf	CTCCTTGGCTCCGGGACT	CAGGAAGGTGGCCGGTTGT
Primers for making reporters		
Piwil2_reporter	AAGGTACCGACTTGTCCGTATTGACACAAG	TTCTCGAGTGAGCCTCAACAGGAGTGTT
Tdrd1_reporter	AAGGTACCTGCACTCATTGTACGCTTCG	TTCTCGAGCACTGATGTCTGAGAAAAGAAATGT

Dataset S1 (separate file).

Differentially expressed genes identified from RNAseq analysis of testicular germ cell samples obtained from E16.5 *Rhox10*-null (KO) and control (WT) mice.

SI References

1. M. Kanatsu-Shinohara *et al.*, Improved serum- and feeder-free culture of mouse germline stem cells. *Biol Reprod* **91**, 88 (2014).
2. Y. Xie, J. M. Rosser, T. L. Thompson, J. D. Boeke, W. An, Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Res* **39**, e16 (2011).
3. C. Frainais *et al.*, RHOXF2 gene, a new candidate gene for spermatogenesis failure. *Basic Clin Androl* **24**, 3 (2014).
4. J. Borgmann *et al.*, The human RHOX gene cluster: target genes and functional analysis of gene variants in infertile men. *Hum Mol Genet* 10.1093/hmg/ddw313 (2016).
5. M. Ramaiah *et al.*, A microRNA cluster in the Fragile-X region expressed during spermatogenesis targets FMR1. *EMBO Rep* **20** (2019).
6. K. Tan *et al.*, Impaired imprinted X chromosome inactivation is responsible for the skewed sex ratio following in vitro fertilization. *Proc Natl Acad Sci U S A* **113**, 3197-3202 (2016).
7. A. Sohni *et al.*, The Neonatal and Adult Human Testis Defined at the Single-Cell Level. *Cell Rep* **26**, 1501-1517 e1504 (2019).
8. K. Tan *et al.*, Transcriptome profiling reveals signaling conditions dictating human spermatogonia fate in vitro. *Proc Natl Acad Sci U S A* **117**, 17832-17841 (2020).
9. A. Dobin, T. R. Gingeras, Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics* **51**, 11 14 11-19 (2015).
10. Y. Liao, G. K. Smyth, W. Shi, The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41**, e108 (2013).
11. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
12. W. Huang da, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
13. D. Szklarczyk *et al.*, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-D613 (2019).
14. K. Tan *et al.*, Dynamic integrated analysis of DNA methylation and gene expression profiles in in vivo and in vitro fertilized mouse post-implantation extraembryonic and placental tissues. *Mol Hum Reprod* **22**, 485-498 (2016).
15. S. J. Newkirk *et al.*, Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proc Natl Acad Sci U S A* **114**, E5635-E5644 (2017).
16. K. Tan, H. W. Song, M. F. Wilkinson, Single-cell RNAseq analysis of testicular germ and somatic cell development during the perinatal period. *Development* **147** (2020).
17. K. Tan *et al.*, The role of the NMD factor UPF3B in olfactory sensory neurons. *Elife* **9** (2020).