The *Tug1* Locus is Essential for Male Fertility

Jordan P. Lewandowski[2,*], Gabrijela Dumbović[1,*], Audrey R. Watson[1,3,*], Taeyoung Hwang[1,*], Emily Jacobs-Palmer[4], Nydia Chang[2], Christian Much[1], Kyle Turner[4], Christopher Kirby[4], Jana Felicitas Schulz[5], Clara-Louisa Müller[5], Nimrod D. Rubinstein[6], Abigail F. Groff[7], Steve C. Liapis[2], Chiara Gerhardinger[2], Norbert Hubner[5], Sebastiaan van Heesch[5], Hopi E. Hoekstra[4,6,8,‡], Martin Sauvageau[2,9,10‡], John L. Rinn[1,2,3,8‡]

[1] BioFrontiers Institute, University of Colorado at Boulder, Boulder, CO, 80303, USA
[2] Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, 02138, USA
[3] Department of Biochemistry, University of Colorado at Boulder, Boulder, CO, 80303, USA
[4] Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, 02138, USA
[5] Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, Berlin, 13125, Germany
[6] Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, 02138, USA
[7] Department of Systems Biology, Harvard University, Cambridge, MA, 02138, USA
[8] Howard Hughes Medical Institute, Chevy Chase, MD, USA.
[9] Functional Genomics and Noncoding RNAs, Montreal Clinical Research Institute, Montreal QC, H2W 1R7, Canada
[10] Department of Biochemistry and Molecular Medicine, Université de Montréal, QC, H3C 3J7, Canada

* Co-first authors

‡ Corresponding Authors: John L. Rinn, PhD
BioFrontiers Institute
University of Colorado at Boulder
Boulder, CO 80303, USA
john.rinn@colorado.edu

Hopi E. Hoekstra, PhD
Harvard Organismic and Evolutionary Biology
16 Divinity Avenue, 4109 BioLabs
Cambridge, MA 02138, USA
hoekstra@oeb.harvard.edu

Martin Sauvageau, PhD
Functional Genomics and Noncoding RNAs
Montreal Clinical Research Institute
Montreal, QC, H2W 1R7, Canada
martin.sauvageau@ircm.qc.ca

ABSTRACT

Background: Several long noncoding RNAs (lncRNAs) have been shown to function as central components of molecular machines that play fundamental roles in biology. While the number of annotated lncRNAs in mammalian genomes has greatly expanded, their functions remain largely uncharacterized. This is compounded by the fact that identifying lncRNA loci that have robust and reproducible phenotypes when mutated has been a challenge.

Results: We previously generated a cohort of 20 lncRNA loci knockout mice. Here, we extend our initial study and provide a more detailed analysis of the highly conserved lncRNA locus, Taurine Upregulated Gene 1 (*Tug1*). We report that *Tug1* knockout male mice are sterile with complete penetrance due to a low sperm count and abnormal sperm morphology. Having identified a lncRNA loci with a robust phenotype, we wanted to determine which, if any, potential elements contained in the *Tug1* genomic region (DNA, RNA, protein, or the act of transcription) have activity. Using engineered mouse models and cell-based assays, we provide evidence that the *Tug1* locus harbors three distinct regulatory activities – two noncoding and one coding: (i) a *cis* DNA repressor that regulates many neighboring genes, (ii) a lncRNA that can regulate genes by a *trans*-based function, and finally (iii) *Tug1* encodes an evolutionary conserved peptide that when overexpressed impacts mitochondrial membrane potential.

Conclusions: Our results reveal an essential role for the *Tug1* locus in male fertility and uncover three distinct regulatory activities in the *Tug1* locus, thus highlighting the complexity present at lncRNA loci.

KEYWORDS

Tug1, lncRNA, fertility, DNA repressor, peptide, mouse, *in vivo*, RNA-seq

BACKGROUND

It has long been appreciated that noncoding RNAs play central roles in biology. Key cellular machines, such as telomerase and the ribosome, are comprised of both protein and noncoding RNAs and serve as classic examples of RNA-based functionalities (Feng et al., 1995; Sonenberg et al., 1975). LncRNAs have been shown to function in a variety of biological processes; however, different strategies to study lncRNA function have led to discrepancies in the observed phenotypes, thereby highlighting the challenges of finding robust and reproducible lncRNA phenotypes (Goudarzi et al., 2019). Moreover, another challenge presented when studying lncRNA loci, is that they can harbor several potential regulatory modalities including, DNA regulatory elements, misannotated protein-coding genes, and even the act of transcription. Therefore, it is important to determine what regulatory elements, if any, are active at lncRNA loci.

A number of studies have revealed that lncRNA loci can mediate their function through a variety of mechanisms (Kopp and Mendell, 2018). A few well-studied lncRNA examples include *Xist*, which is a key factor in the X inactivation pathway and acts locally (*cis*) (Lee and Jaenisch, 1997; Penny et al., 1996), *Malat1,* which modulates alternative splicing and acts distally (*trans*) (Tripathi et al., 2010), and other lncRNAs, such as *linc-Cox2*, which functions to activate and repress gene expression through local and distal mechanisms (Bester et al., 2018; Carpenter et al., 2013; Elling et al., 2018). While it is clear that a number of lncRNA loci have RNA-based roles, recent findings have shown that some lncRNA loci, such as *Lockd* (Paralkar et al., 2016), *lincRNA-p21* (Groff et al., 2016), and *Peril* (Groff et al., 2018), regulate gene expression in *cis* through DNA regulatory elements, independent of the noncoding transcript. Moreover, many lncRNAs possess small open reading frames (ORFs) (Housman and Ulitsky, 2016; Slavoff et al., 2013), and an increasing number encode small peptides that have biological roles (Anderson et al., 2015; Chng et al., 2013; Nelson et al., 2016; Stein et al., 2018). With this in mind, it is likely that more regulatory DNA, RNA and protein activities will be uncovered at lncRNA loci.

We previously reported the generation of 20 lncRNA loci knockout mouse strains, five of which displayed either viability, growth or brain phenotypes (Lai et al., 2015; Sauvageau et al., 2013). From the strains that did not initially display such phenotypes, we selected *Tug1* for further analysis because it is highly conserved between human and mouse and it has been reported to have a number of diverse cellular functions. *Tug1* was first identified to contain a lncRNA transcript that, upon RNAi-mediated knockdown, affects the development of photoreceptors in the mouse retina (Young et al., 2005). *Tug1* also has a human ortholog that has a number of unique molecular properties including being regulated by p53 (Guttman et al., 2009) and associating with polycomb repressive complex 2 (PRC2) (Khalil et al., 2009). In addition, *TUG1* RNA has also been proposed to play multiple cellular roles, such as acting as a tumor suppressor in human gliomas (Katsushima et al., 2016; Li et al., 2016), as a cytoplasmic miRNA sponge in prostate cancer cell lines (Du et al., 2016), and being involved in chromatin and gene regulation in the nucleus (He et al., 2018; Khalil et al., 2009; Long et al., 2016). Together, these studies highlight diverse cellular functions for the *Tug1* RNA.

Here, we characterize the *Tug1* locus using multiple genetic approaches and describe a physiological function in spermatogenesis and male fertility. We show that deletion of the *Tug1* locus in mice leads to male sterility due to reduced sperm production as well as a failure of spermatids to individualize during spermiation. Using several complementary genetic approaches (whole locus deletion with a *lacZ* reporter knock-in, an inducible *Tug1* transgene, and combinations thereof), we provide evidence of a DNA-based repressive element within the *Tug1* locus that regulates several genes in *cis*. Furthermore, we show that a gene-expression program dysregulated in *Tug1* knockout testes can be partially rescued by ectopic expression of *Tug1* RNA *in vivo*. Finally, we show that the *Tug1* locus contains an evolutionarily

132 conserved ORF, which is translated into a peptide and regulates mitochondrial function upon
133 overexpression. Collectively, our study implicates *Tug1* as an essential locus in male fertility and
134 demonstrates that the *Tug1* locus contains at least three regulatory activities – two noncoding and one
135 coding.
136

137 **RESULTS**
138 **The *Tug1* lncRNA locus is widely expressed and highly conserved**
139 The murine *Tug1* lncRNA locus is located on chromosome 11 and has three annotated transcripts (Figure
140 1A). *Tug1* shares a bidirectional promoter with its neighboring protein-coding gene *Morc2a*, whose
141 transcription start site (TSS) is located approximately 680 base pairs upstream of the first *Tug1* TSS. The
142 *Tug1* locus is enriched with hallmarks of active transcription, such as RNA polymerase II (Pol II) and histone
143 H3 lysine 4-trimethylation (H3K4me3) at its promoter, H3K36me3 across its gene body, and abundant
144 transcription as shown by RNA-seq (Figure 1A). However, the *Tug1* locus is simultaneously enriched with
145 the repressive histone mark H3K9me3 in several mouse cell types (Figure 1A and Figure S1). This atypical
146 combination of H3K9me3 and H3K36me3 histone marks at the *Tug1* locus is also conserved in human cells
147 (Figure S1). Moreover, the binding of repressor proteins SIN3A and COREST has been detected at both
148 the human and mouse promoters (Figure S1).
149   *Tug1* is among the most conserved lncRNAs between human and mouse, with exonic nucleotide
150 conservation levels reaching 77% (Figure 1B). This level of sequence conservation is similar to the highly
151 abundant lncRNA *Malat1* (79%), and higher than other well characterized lncRNAs including *Hottip* (71%),
152 *Neat1* (69%), *Xist* (30%) and *Firre* (4%) (Figure 1B) (Chen et al., 2016). Interestingly, further conservation
153 analyses lead us to identify a highly conserved putative open reading frame (ORF) in the *Tug1 locus,* as
154 indicated by phylogenic codon substitution frequencies (PhyloCSF) (Lin et al., 2011), a computational tool
155 for identifying protein-coding and non-coding regions (Figure 1A).
156   Apart from its high level of sequence conservation, *Tug1* RNA also has unique expression
157 properties. First, the *Tug1* lncRNA is expressed at moderate to high levels in several adult tissues in both
158 mouse and human (Figure 1C) (Fagerberg et al., 2014; The Mouse ENCODE Consortium, 2014). Second,
159 the *Tug1* lncRNA is abundantly detected in a number of embryonic tissues at different embryonic stages
160 (E8.0 – E12.5) (Figure 1D and Figure S2). Finally, using single molecule RNA fluorescence *in situ*
161 hybridization (smFISH) we observed that the *Tug1* lncRNA is detected in both the cytoplasm and the
162 nucleus in human and mouse fibroblasts (Figure 1E), which is consistent with previous reports (Cabili et al.,
163 2015; Khalil et al., 2009; van Heesch et al., 2014; Zhang et al., 2014).
164

165 ***Tug1-/-* males are sterile due to impaired spermatogenesis**
166 To investigate the *in vivo* role of *Tug1*, we utilized a previously generated full gene-ablation model (*Tug1-/-*
167), where after the promoter and first exon, the gene body of the *Tug1* locus was replaced with a *lacZ* reporter
168 cassette, thereby keeping the act of transcription intact (Figure 2A) (Lai et al., 2015; Sauvageau et al.,
169 2013). Notably, this deletion strategy also removed 86 out of 143 amino acids in the putative ORF (Figure
170 S3). Loss of *Tug1* was confirmed by genotyping and by RNA-seq analysis in wild type and *Tug1-/-* testes
171 (Figure 2A). Thus, through this approach any potential phenotype due to the lncRNA, potential DNA
172 elements or even the putative peptide would be included.
173   *Tug1-/-* mice are viable and do not display any obvious physiological abnormalities up to one year of
174 age, with the exception of a slight reduction in weight in male mice relative to wild type littermates (Figure
175 S4A). As previously reported, the progeny of *Tug1+/-* intercrosses follow normal Mendelian ratios

4

176   (Sauvageau et al., 2013). However, we noticed a complete absence of offspring from intercrosses between
177   *Tug1-/-* mice (n = 4 breeding pairs). Therefore, we sought to investigate the fertility of *Tug1-/-* mutants in more
178   detail. We separately mated *Tug1-/-*, *Tug1+/-* and wild type males or females to *C57BL/6J* mice. We did not
179   observe a difference in the mounting behavior between wild type and *Tug1-/-* mice, as assessed by the
180   presence of a vaginal plug. Strikingly, matings between *Tug1-/-* males (n = 8) and *C57BL/6J* females did not
181   produce any offspring, whereas matings involving either *Tug1+/-* males (n = 8) or wild type males (n = 8)
182   with *C57BL/6J* females resulted in similar numbers of offspring (Figure 2B). Moreover, six out of nine *Tug1-/-*
183   females that mated with *C57BL/6J* males gave birth to pups (Figure 2B), indicating that only *Tug1-/-* males
184   appear sterile. Thus, the *Tug1* locus is likely required for male fertility.

185   To further understand the underlying fertility defect in *Tug1-/-* males, we first examined the
186   reproductive morphology of wild type and *Tug1-/-* male mice. Testicular descent appeared normal and we
187   did not observe any other gross morphological abnormalities in their reproductive system upon dissection
188   (Figure S4B). We measured testes mass relative to total body weight and did not observe a significant
189   decrease ($p = 0.0751$) in *Tug1-/-* (mean = 0.25 ± 0.020 %, n = 8) compared to wild type (mean = 0.30 ±
190   0.016 %, n = 9) (Figure S4C). Next, we quantified sperm production and found a significant reduction in
191   sperm number from *Tug1-/-* males (mean = $2.35 \times 10^6 \pm 0.473 \times 10^6$ cells/mL, n = 7), which produced on
192   average only 40% as many sperm as wild type mice ($6.13 \times 10^6 \pm 0.636 \times 10^6$ cells/mL, n = 9, $p = 0.0018$)
193   (Figure 2C). Notably, although *Tug1-/-* males produce fewer sperm, none were found to completely lack
194   sperm (azoospermic).

195   Based on these results, we investigated whether perturbations in sperm morphology could explain
196   the complete infertility in *Tug1-/-* males. We examined the morphological features of sperm and quantified
197   the frequency of 15 different abnormalities (Table S1). Overall, the proportion of morphologically normal
198   sperm was significantly lower in *Tug1-/-* mice (mean = 8.3 ± 3.0 %, n = 8, $p = 0.0013$) compared to wild type
199   males (mean = 38.9 ± 4.3 %, n = 9) (Figure 2D). We observed significant morphological defects in *Tug1-/-*
200   sperm including: sperm with no head, misshapen head, head bent back, stripped midpiece, kinked
201   midpiece, curled midpiece, midpiece debris, broken tail, and the presence of multiple sperm attached along
202   the midpiece (Figure 2D, Figure S4D, and Table S1). Together, these results indicate that the sterility of
203   *Tug1-/-* males arises from a combination of low sperm count (oligozoospermia) and abnormal sperm
204   morphology (teratozoospermia).

205   To further investigate how the deletion of the *Tug1* locus leads to abnormal sperm morphology, we
206   examined the timing of *Tug1* expression at different stages of spermatogenesis. To this end, we took
207   advantage of the knock-in *lacZ* reporter driven by the endogenous *Tug1* promoter and assessed expression
208   by *lacZ* staining of histological sections of *Tug1+/-* testis and epididymis. From stages IX to XI of
209   spermatogenesis in the testis, *lacZ* staining was restricted to excess cytoplasm, known as residual bodies,
210   which are phagocytosed toward the basement membrane by Sertoli cells (Figure 2E) (Firlit and Davis,
211   1965). No expression was detected in the later stages XII to XIV (Figure 2E). However, we observed *lacZ*
212   staining in stage XV elongated spermatids and the *lacZ* staining became stronger at stage XVI, just before
213   spermiation (Figure 2E). The observed *lacZ* pattern indicates that *Tug1* expression is temporally controlled
214   during spermatogenesis.

215   In *Tug1-/-* testes, mature spermatids appeared to remain attached by their collective cytoplasm. This
216   was even more striking in the epididymis, where multiple sperm aggregates were observed in *Tug1-/-* mice,
217   while individual sperm appeared to migrate freely throughout the lumen in wild type mice (Figure 2F). These
218   aggregates were present in all regions of the epididymis (caput, corpus and cauda). Consistent with the

5

219   reduced sperm count, fewer individual sperm were observed in *Tug1*[-/-] epididymis tissue compared to wild
220   type. Together, our analyses of the *Tug1*[-/-] model provide evidence that the locus is required for male fertility.
221

222   *Tug1* DNA encodes a *cis* repressor regulatory element
223   Since we observed a robust phenotype in our *Tug1*[-/-] model, we next sought to investigate what, if any,
224   molecular activities (DNA, lncRNA, and protein) are present at the *Tug1* locus. We first focused on
225   determining if the DNA at the *Tug1* locus harbored any regulatory activity, because many lncRNA loci have
226   been reported to contain DNA regulatory elements that can regulate the expression of neighboring genes
227   (*cis*-acting) (Groff et al., 2016; Groff et al., 2018; Paralkar et al., 2016). Our *Tug1*[-/-] model enables us to test
228   for potential *cis* regulatory activity within the *Tug1* locus because our gene-ablation design removes
229   potential *cis*-acting elements, yet keeps the act of transcription intact (Figure 2A) (Lai et al., 2015;
230   Sauvageau et al., 2013). To determine if there is a local regulatory effect on gene expression, we performed
231   RNA-seq on testes from wild type and *Tug1*[-/-] mice and plotted significant changes in gene expression within
232   a 2-Mb region centered on the *Tug1* locus (FDR < 0.05, FC > 1.5). Of the 71 genes within this window, we
233   observed six differentially regulated genes: *Rnf185*, *Pla2g3*, *Selm*, *Smtn*, *Gm11946* and *8430429K09Rik*.
234   Notably, all of these genes were significantly upregulated in *Tug1*[-/-] compared to wild type and located
235   downstream of the *Tug1* TSS (Figure 3A). Because these six genes are all upregulated in *Tug1*[-/-] testes,
236   this local effect on neighboring gene expression provides evidence of a *cis* repressor function in the *Tug1*
237   locus.
238   To further investigate whether the *cis*-effect of the *Tug1* locus was more widespread, we performed
239   RNA-seq on six additional tissues (prostate, spleen, eyes, heart, liver and mouse embryonic fibroblasts
240   (MEFs)) as well as re-analyzed an existing brain dataset (Goff et al., 2015) from wild type and *Tug1*[-/-] mice
241   (Table S2). We examined whether genes within a 2-Mb window centered on the *Tug1* locus were similarly
242   dysregulated in the different tissues. Consistent with the testes, of the 71 genes within this window, nine
243   genes were dysregulated in one or more tissues (seven upregulated and two downregulated) (Figure 3A).
244   Notably, of the seven upregulated genes, the E3 ubiquitin ligase *Rnf185* was consistently upregulated in 8
245   of 8 *Tug1*[-/-] tissues, followed by the selenoprotein gene, *Selm* (7 of 8 samples), and *8430429K09Rik* (6 of
246   8 samples) (Figure 3B). This dysregulation is consistent with a previous study from our group in which we
247   observed a misregulation of genes located near the *Tug1* locus in the brain of our *Tug1*[-/-] model (Goff et al.,
248   2015). We also observed that *Morc2a*, the protein-coding gene that shares a promoter with *Tug1*, was
249   significantly downregulated in 4 of the 8 samples. Collectively, these data suggest that the *Tug1*-mediated
250   repressive *cis*-effect functions in a broad range of tissues.
251   Since the neighboring genes are upregulated upon deletion of the *Tug1* locus, we reasoned that the
252   repressive activity could be mediated either directly by the *Tug1* transcript or by regulatory DNA elements
253   within the locus. To determine if the repressive effect of *Tug1* on neighboring genes occurs on the same
254   allele (*cis*-acting), we performed allele-specific RNA-seq using a hybrid mouse strain. To generate this
255   strain, we crossed *Tug1*[+/-] C57BL/6J females with *Mus castaneus* (Cast/EiJ) males (Figure 3C). The
256   resulting polymorphisms in the F1 hybrid progeny (~1/150 bp between C57BL/6J and Cast/EiJ) allow
257   quantification of gene expression from each strain-specific allele (Keane et al., 2011). We thus harvested
258   testes from F1 hybrid males harboring a maternal C57BL/6J allele deletion and performed allele-specific
259   expression analysis (Figure 3B and Table S3). As a control for haplotype specific effects, we also analyzed
260   allele-specific expression differences in wild type F1 hybrid C57BL/6J::Cast/EiJ male littermates. We then
261   quantified the expression from each allele and found that *Rnf185*, *Selm*, and *Smtn* were significantly
262   upregulated and *Morc2a* slightly downregulated only on the C57BL/6J allele containing the *Tug1* deletion

263   (Figure 3D). Importantly, no change in expression was detected from any gene within 1 Mb of *Tug1* on the

264   Cast/EiJ allele, which contains an intact *Tug1* locus (Figure 3D). Moreover, it is notable that *Tug1* RNA from

265   the intact allele does not impact the dysregulated genes found on the *Tug1* knockout allele, thereby

266   suggesting a DNA-based repressor role at the *Tug1* locus. From the multiple mouse models, we conclude

267   that the *Tug1* DNA, rather than the lncRNA or the act of transcription, exerts a repressive effect in *cis* on

268   several genes up to 200 kb downstream of the *Tug1* transcription site.

269

270   **_Tug1_ lncRNA regulates gene expression in _trans_**

271   Previous studies have suggested a *trans* role for the *Tug1* lncRNA on chromatin regulation and gene

272   expression (Han et al., 2013; Li et al., 2016; Long et al., 2016; Xu et al., 2014; Young et al., 2005; Zhang et

273   al., 2014). Thus, we set out to determine if the lncRNA from the *Tug1* locus displays any trans regulatory

274   activity on gene expression *in vivo*. We analyzed the RNA-seq data for *Tug1*$^{-/-}$ tissues (testis, prostate,

275   spleen, eyes, liver, heart, brain and MEFs), and identified significant changes in gene expression relative

276   to wild type. Deletion of the *Tug1* locus was accompanied by 2139 significantly dysregulated genes across

277   all tissues examined. We observed that global changes in gene expression clustered by tissue-type,

278   indicating tissue-specific gene dysregulation (Figure 4A, Table S2, and Table S4). We found that while most

279   dysregulated genes (~89%) were perturbed in only a single tissue (Figure 4B), several genes were

280   commonly dysregulated across multiple tissues (Figure 4B, Table S2, Table S4). We then performed gene

281   set enrichment analysis (GSEA) using the differentially expressed genes for each tissue and observed an

282   enrichment of several pathways that were shared across the individual tissues. For example, oxidative

283   phosphorylation, Myc targets, and epithelial to mesenchymal transition were found enriched in 7 of the 8

284   *Tug1*$^{-/-}$ tissues (Figure 4C).

285   To investigate the role of *Tug1* RNA, we sought to address whether ectopic expression of *Tug1* RNA

286   could restore the genes dysregulated in *Tug1*$^{-/-}$ testes. Given that *Tug1* harbors a putative peptide encoded

287   in the 5' region (discussed below), we focused on a *Tug1* isoform that lacks the 5' region, thus ensuring we

288   would address the role of *Tug1* RNA alone. To this end, we generated a doxycycline (dox)-inducible *Tug1*

289   transgenic mouse by cloning a *Tug1* isoform downstream of a tet-responsive element (henceforth called

290   tg(*Tug1*)) (Figure 4D and methods section). Next, we generated compound transgenic mice that contained

291   the *Tug1* transgene in the *Tug1*$^{-/-}$ background that also constitutively overexpressed the reverse tetracycline

292   transcriptional activator gene (CAG-rtTA3) (combined alleles henceforth called *Tug1*^{rescue}). This approach

293   enabled systemic induction of *Tug1* RNA in the presence of dox, allowing to distinguish DNA- and RNA-

294   based effects, and to test if *Tug1* RNA expression alone would be sufficient to rescue gene expression and

295   male fertility phenotypes arising in *Tug1*$^{-/-}$ mice.

296   Because *Tug1*^{rescue} mice lacked endogenous *Tug1*, we were able to assess the level of *Tug1* RNA

297   from the transgene. We performed RNA-seq on testes from *Tug1*^{rescue} mice (*n* = 3) and found that *Tug1*

298   RNA from the transgene was expressed at significantly lower levels than wild type in the testes (Figure 4E

299   and Figure S5A). Moreover, we sorted peripheral blood cell types (CD4, CD8, and NK) from *Tug1*^{rescue} mice

300   and also found lower levels of *Tug1* RNA induction relative to wild type (Figure S5A,B). Even though the

301   transgene expression was low, we reasoned that this would still be a valuable *in vivo* model to test RNA-

302   mediated effects on gene regulation. Thus, we tested whether genes found dysregulated in the testes from

303   *Tug1*$^{-/-}$ mice could be rescued by ectopic expression of the *Tug1* RNA in our *Tug1*^{rescue} model. Notably, 52

304   of the 1051 genes that were dysregulated in *Tug1*$^{-/-}$ testes were found significantly reciprocally regulated in

305   *Tug1*^{rescue} testes (Figure 4G, Table 1, and Table S4). For example, a mitochondrial related gene, *Mrarp*, and

306   an aquaporin gene, *Aqp2*, are significantly upregulated in *Tug1*$^{-/-}$ testes, but their expression was reduced

307 to wild type levels in *Tug1*[rescue] testes (Figure 4H). Conversely, the predicted lncRNA gene *Gm28181* that
308 is significantly reduced in *Tug1*[-/-] testes, is significantly upregulated to wild type levels in *Tug1*[rescue] testes
309 (Figure 4H). While we observed a *trans*-effect for *Tug1* RNA, we did not observe any changes in expression
310 for the genes neighboring the *Tug1* locus (Figure 4F and Table S4). Taken together, these data demonstrate
311 that *Tug1* lncRNA regulates a subset of genes by an RNA-based *trans* mechanism, evident even at low
312 levels of *Tug1* RNA.

313   We also tested if *Tug1*[rescue] male mice had normal fertility. We did not obtain any progeny from
314 matings between *Tug1*[rescue] male mice (n = 3) with C57BL6/J female mice (n = 12) (Figure S5C). Moreover,
315 we found that *Tug1*[rescue] males had a low sperm count (mean = $3.20 \times 10^5 \pm 8.0 \times 10^3$ cells/mL) which was
316 similar to the lower sperm count observed in *Tug1*[-/-] males (mean = $4.69 \times 10^5 \pm 1.6 \times 10^4$ cells/mL)
317 compared to wild type (mean = $9.32 \times 10^5 \pm 3.9 \times 10^3$ cells/mL). In addition, histological sections of *Tug1*[rescue]
318 testes and epididymis showed fewer sperm, thereby confirming the low sperm count (Figure S5E). In further
319 analysis, we observed that *Tug1*[rescue] mice had a low proportion of normal shaped sperm which was also
320 observed in *Tug1*[-/-] mice (Figure S5F). While this finding may indicate that the sterility phenotype is not due
321 to the lncRNA, the lack of a fertility rescue may also be due to the insufficient levels of *Tug1* expression
322 from the transgene in the testes.

323

324 **The *Tug1* locus encodes an evolutionary conserved peptide in human and mouse**
325 It has become increasingly clear that some lncRNA annotations also encode small peptides (Makarewich
326 and Olson, 2017). Since a PhyloCSF analysis revealed the presence of putative ORFs in the *Tug1* locus
327 (Figure 1A and Figure S6A), we further tested whether the *Tug1* locus could encode a peptide using
328 biochemical and cell-based assays. First, we systematically screened for ORFs that displayed strong
329 conservation across species, allowing for both canonical (AUG) and non-canonical (CUG and UUG)
330 translation start codons. We identified multiple short ORFs in human and mouse *TUG1*/*Tug1* (11 and 15,
331 respectively) (Figure 5A). Two ORFs (designated as ORF1 and ORF2) at the 5' region of *TUG1*/*Tug1* drew
332 our attention due to their conserved translational start and stop sites, as well as their high level of nucleotide
333 conservation between human and mouse (Figure 5A). ORF1 (154 amino acids in human) and ORF2 (153
334 amino acids in human) both start with a non-canonical start codon (CUG). On the amino acid level, ORF1
335 and ORF2 share 92% and 70% cross-species identity, respectively. Moreover, ORF1 has a high PhyloCSF
336 score (350) and shows conservation spanning its entire sequence, whereas ORF2 does not show patterns
337 of preserving synonymous mutations, indicating that ORF1 is more likely to be translated (Figure 5B and
338 Figure S6A).

339   To further hone in on translated regions of *Tug1*, we analyzed ribosome profiling data (Michel et al.,
340 2014), which identifies regions of RNA bound to ribosomes by high-throughput sequencing, thus indicating
341 actively translating portions of an RNA. We found pronounced ribosomal occupancy across the entire ORF1
342 sequence with a sharp decrease at its stop codon (Figure 5B) (Ingolia et al., 2009). A similar pattern
343 indicative of active translation of *Tug1* ORF1 is also observed from ribosome profiling in human, mouse,
344 and rat heart tissue (S. van Heesch, personal communication, September 2018). However, ORF2 does not
345 show ribosome occupancy above background level, particularly after the ORF1 stop codon (Figure 5B and
346 Figure S6A). Taken together, these results suggest that the most 5' region of *TUG1*/*Tug1* contains an ORF
347 that has evolutionary conservation characteristic of protein-coding genes. We designated the putative
348 peptide originating from ORF1 as TUG1-BOAT (*Tug1*-**B**ifunctional **O**RF **a**nd **T**ranscript).

349   To determine if ORF1 is translated, we first performed *in vitro* translation assays using [35S]-
350 methionine incorporation to detect newly synthesized proteins for three different constructs: (i) the

351   endogenous *TUG1* lncRNA (including the endogenous 5'UTR, ORF1 and a part of the 3'UTR), (ii) a codon
352   optimized ORF1-3xFLAG and (iii) a codon optimized ORF1-mEGFP (Figure S6B). For each construct, we
353   observed a protein product of the expected size, thereby supporting that ORF1 can produce a stable peptide
354   (Figure S6C). We next generated C-terminal epitope tagged human and mouse TUG1-BOAT expression
355   constructs with and without the 5' leader sequences (Figure 5C). As a negative control, we generated a
356   construct containing GFP in place of the TUG1-BOAT cDNA sequence. We then transfected 3T3 (mouse)
357   and HeLa (human) cells and tested for TUG1-BOAT translation by western blot analysis. We detected
358   peptides of approximately 19 kDa and 21 kDa in both cell lines (Figure 5D), which closely corresponds to
359   the predicted molecular weights of hTUG1-BOAT (18.7 kDa) and mTUG1-BOAT (19 kDa) fusion constructs,
360   respectively. Collectively, these results show that ORF1, with its 5' UTR and a native non-canonical
361   translational start site, can be translated into TUG1-BOAT in both human and mouse cells.

362   Having detected a peptide of expected size from human and mouse TUG1-BOAT constructs, we
363   next investigated the peptide's subcellular localization by immunofluorescence. We observed that human
364   and mouse TUG1-BOAT is distributed throughout the nucleus and cytoplasm in the majority of the cells
365   (>80 % cells, n = 50) (Figure 5E). However, in a subset of cells, TUG1-BOAT was predominantly
366   cytoplasmic (<20 % of cells, n = 50) (Figure 5E). Moreover, we found that TUG1-BOAT showed co-
367   localization with the mitochondria (Figure S6D).

368

369   **TUG1-BOAT overexpression compromises mitochondrial membrane potential**
370   We next sought to identify a potential cellular role for TUG1-BOAT. We used protein structure/domain
371   prediction tools to further characterize TUG1-BOAT. Based on predictions, TUG1-BOAT does not represent
372   any known homologs, and the predicted structures are conserved between human and mouse (template
373   modeling score of 0.658). Further investigation of putative functional domains revealed a conserved
374   mitochondrial localization domain (Figure 5F). Based on the predicted mitochondrial localization domain
375   encoded in TUG1-BOAT (Figure 5F), its co-localization with the mitochondria (Figure S6D), and given that
376   oxidative phosphorylation was one of the most affected pathways across multiple   *Tug1$^{-/-}$* tissues (Figure
377   4C), we hypothesized that TUG1-BOAT may have a role in the mitochondria.

378   To this end, we first examined mitochondrial membrane potential by using chloromethyl-X-rosamine
379   (CMXR), a lipophilic fluorescent cation that accumulates in the negatively charged interior of mitochondria
380   (Macho et al., 1996). We transfected human and mouse TUG1-BOAT expression constructs with and
381   without the 5' UTR, as well as a control GFP-containing plasmid and a *Tug1* construct that lacks ORF1
382   (*Tug1* cDNA ΔmORF1) into 3T3 cells (Figure 6A). Notably, cells with either human or mouse TUG1-BOAT
383   showed a reduction in mitochondrial staining by CMXR (22% and 44% CMXR stained cells, respectively),
384   compared to cells in the same culture not expressing TUG1-BOAT (Figure 6B). In contrast, cells expressing
385   GFP or *Tug1* cDNA ΔmORF1 were positive for CMXR staining in all cells examined, thus indicating that
386   CMXR staining deficiency is induced by the TUG1-BOAT peptide alone, rather than the *Tug*1 RNA.

387   Since CMXR is commonly used to measure mitochondrial membrane potential, we reasoned that
388   either impaired mitochondrial integrity or impaired redox potential at the mitochondrial membrane could
389   account for the accumulation defect of CMXR in mitochondria upon TUG1-BOAT overexpression. To
390   address these possibilities, we immunostained for TOM20, a redox independent translocase located on the
391   outer mitochondrial membrane (Likić et al., 2005) in cells overexpressing human or mouse TUG1-BOAT.
392   We observed staining for TOM20 in cells without CMXR staining, indicating that the mitochondria were
393   intact (Figure 6C). Collectively, these results provide evidence that human and mouse TUG1-BOAT have

394 conserved roles in mitochondrial function, and that *Tug1* RNA, DNA, and the TUG1-BOAT peptide have
395 distinct roles.

396
397

398 **DISCUSSION**
399 To date, there are a few well-established *in vivo* genetic models of lncRNAs with robust phenotypes and
400 lncRNAs remain understudied, as a class, in this context. This is further complicated by the fact that lncRNA
401 loci can contain multiple regulatory modalities including the DNA, RNA, protein, and the act of transcription.
402 Therefore, when a lncRNA locus presents a robust phenotype, understanding what molecular activities are
403 present at the locus is an important foundation in order to then address how it could potentially mediate an
404 effect. In this study, we characterized in more detail one of our previously published lncRNA knockout
405 mouse models, *Tug1*, and extended our understanding of the function of this locus *in vivo* by defining the
406 molecular properties present at the locus. By implementing multiple *in vivo* genetic strategies, we report a
407 number of key findings: (i) in our mouse model, deletion of the *Tug1* locus leads to completely penetrant
408 male sterility due to late stage spermatogenesis defects, (ii) we find evidence that *Tug1* harbors a *cis*-acting
409 DNA repressive element, and (iii) we find evidence that the *Tug1* RNA regulates a subset of genes in *trans*.
410 Moreover, using biochemical and cell-based assays, we find evidence that (iv) *Tug1* encodes a highly
411 conserved peptide in its 5' region that appears to have a mitochondrial role. Together, our results point to
412 an essential role for the *Tug1* locus in male fertility, where the locus harbors three distinct regulatory
413 activities.

414

415 *LncRNAs in spermatogenesis and male fertility*:
416 Infertility is estimated to affect approximately 15% of couples in developed countries, with male
417 infertility contributing up to 50% of cases. Much about the molecular regulation of male germ cell
418 development remains to be understood, especially the contribution of the noncoding genome. Indeed, a
419 number of studies have performed systematic gene expression profiling at defined stages during
420 spermatogenesis and have identified many developmentally regulated lncRNAs, suggesting that lncRNAs
421 may have a wide role during spermatogenesis (Wichman et al., 2017). In our gene ablation knockout for
422 *Tug1*, we observed a male sterility phenotype with complete penetrance. Although four other lncRNAs
423 (*Tslrn1*, *Tsx*, *Pldi*, and *Mrhl*) have been found to be important in spermatogenesis, none of them lead to
424 male sterility when disrupted (Anguera et al., 2011; Arun et al., 2012; Heinen et al., 2009; Wichman et al.,
425 2017).
426 Our stage-specific analysis of *Tug1*$^{-/-}$ sperm identified several key abnormalities. We observed
427 defects of cytoplasm removal during spermiation, causing mechanical strain on the midpiece region of
428 sperm that include the point of attachment with both head and tailpiece. Interestingly, similar aberrant sperm
429 phenotypes associated with defects of cytoplasm removal during spermiation have been described for the
430 protein-coding genes *Spem1* and *Ehd1* in loss-of-function models (Rainey et al., 2010; Zheng et al., 2007).
431 The timing and mechanism by which sperm shed their collective cytoplasm during individualization is highly
432 regulated (O'Donnell et al., 2011; Steinhauer, 2015); however, this process is poorly characterized in
433 mammals. Based on data from earlier studies, we speculate that the sterility of *Tug1*$^{-/-}$ males arises from a
434 combination of oligozoospermia (low sperm count) and teratozoospermia (abnormal morphology) resulting
435 from a failure of spermatids to individualize during spermatogenesis.

436
437

438 *Multiple molecular modalities of the Tug1 locus*:

439   Toward understanding the molecular activities present at the *Tug1* locus, we investigated the activity
440 of the DNA, lncRNA, and peptide and found that the *Tug1* locus harbors three distinct regulatory activities.
441 First, several lines of evidence indicate that the *Tug1* locus has a *cis*-acting repressive DNA function. We
442 observed that upon deletion of the *Tug1* locus, genes downstream of *Tug1* were consistently upregulated
443 across multiple tissues, and this effect was also observed in an allele-specific manner – thereby suggesting
444 that the local repressive *cis*-effect is mediated by DNA regulatory elements within the *Tug1* locus. Indeed,
445 DNA regulatory elements have been found overlapping gene bodies, whether protein coding or noncoding.
446 Contrary to enhancers, only a handful of repressor elements and silencers have been identified and
447 characterized to date in mammalian genomes (Li and Arnosti, 2011; Li et al., 2014; Qi et al., 2015; Tan et
448 al., 2010). In keeping with the idea of a repressive role for the DNA at the *Tug1* locus, it is notable that in a
449 recent study which systematically tested for enhancer activity across lncRNA loci, enhancer activity was
450 detected at all lncRNA loci examined except for *Tug1*, where only the promoter showed activity (Groff et al.,
451 2018). Further defining the precise DNA repressive elements as well as their mechanism will be of interest
452 to understand the regulatory abilities of the DNA elements within the *Tug1* locus.

453   Second, our study finds regulatory activity for the *Tug1* lncRNA and thus extends previous findings
454 for the role of *Tug1* RNA on gene expression (Long et al., 2016). Using compound transgenic mouse
455 models, we found that a subset of genes found dysregulated in *Tug1*$^{-/-}$ testes, could be reciprocally regulated
456 by ectopic expression of *Tug1* RNA, even at low levels. While our transgene was expressed at lower levels
457 than wild type *Tug1* RNA, other lncRNAs such as *Hottip* and *Xist* have been shown to exert a biological
458 activity at relatively low copy numbers (Sunwoo et al., 2015; Wang et al., 2011). In support of a *trans*-acting
459 role for *Tug1* RNA, a previous study found that *Tug1* RNA can regulate the levels of *Ppargc1a* mRNA in a
460 reciprocal manner in cultured podocytes (Long et al., 2016). In our RNA-seq dataset of eight different tissues
461 we did not find significant dysregulation for *Ppargc1a*, but it is important to note that our RNA-seq dataset
462 does not include the kidney. Expressing *Tug1* RNA at higher levels could uncover a more widespread *trans*
463 role for *Tug1* to regulate gene expression and is of interest for future work.

464   Finally, we demonstrate that *Tug1* encodes an evolutionarily conserved peptide in the 5' region that,
465 when overexpressed, impacts mitochondrial membrane potential. TUG1-BOAT has high positive charge
466 (net charge ~ +16.5), thus we speculate that the high accumulation of such a positively charged peptide at
467 the mitochondria in the overexpression experiments could lead to depolarization of the mitochondrial
468 membrane. A number of recent studies have identified peptides at candidate lncRNA loci, thereby
469 highlighting the complexity of these loci (Anderson et al., 2015; Chng et al., 2013; Nelson et al., 2016; Stein
470 et al., 2018). In one recent example, the candidate lncRNA *LINC00661* was shown to encode a conserved
471 peptide that has a role in mitochondrial function (Stein et al., 2018). In support of a mitochondrial role for
472 the *Tug1* locus, a previous report found that overexpression of a *Tug1* isoform impacted mitochondrial
473 bioenergetics in cultured podocytes from a murine diabetic nephropathy model (Long et al., 2016).

474

475 *The Tug1 locus has multiple regulatory modalities with potential function in spermatogenesis*:

476   Collectively, our study identifies that deletion of the *Tug1* locus results in a completely penetrant
477 male sterility phenotype, and that the *Tug1* locus contains three unique regulatory modalities: the DNA
478 repressive element, the *Tug1* lncRNA transcript, and the peptide (TUG1-BOAT). As such, our findings pose
479 an intriguing possibility that these features could individually or in combination mediate the observed fertility
480 defect in *Tug1* knockout mice. While our study does not resolve this outstanding question, there is evidence
481 to support a role for each modality for further investigation. First, there is some evidence that the cohort of

482　genes downstream of the *Tug1* locus that are transcriptionally upregulated when *Tug1* is deleted have a
483　role in male fertility. Two of the six dysregulated genes in *Tug1*$^{-/-}$ testes, *Smtn* and *Pla2g3*, have loss-of-
484　function mutations that have male fertility and sperm maturation defects (Niessen et al., 2005; Sato et al.,
485　2010). However, links between these genes and fertility in an overexpression context in an animal model
486　have not been reported. In addition, the effect on fertility for the other four genes (*Gm11946*, *Rnf185*, *Selm*,
487　and *8430429K09Rik*) that were found upregulated upon *Tug1* deletion have not been reported in either
488　loss-of-function or overexpression contexts. Second, in support of a potential role for the *Tug1* lncRNA in
489　male fertility, it is notable that *Selenop*, a gene found up-regulated in *Tug1*$^{-/-}$ testes and was reciprocally
490　regulated in the *Tug1*^{rescue} testes, has a known role in male fertility in the loss-of-function context (Hill et al.,
491　2003). Yet, the role of *Selenop* on male fertility in an overexpression context has yet to be reported.
492　Moreover, given that our *Tug1*^{rescue} mice did not restore fertility, this finding may indicate that it is not due to
493　the lncRNA; however, the lack of a rescue may also be explained by the low levels of *Tug1* expression from
494　the transgene. Finally, in our TUG1-BOAT experiments we observed altered mitochondrial membrane
495　potential, which has also been observed in male sterility (Wang et al., 2003). Thus, the TUG1-BOAT peptide
496　may also play a role in male fertility.

497
498

499　**CONCLUSIONS**
500　Our findings reveal an essential role for *Tug1* in male fertility, providing evidence that *Tug1* knockout male
501　mice are sterile with complete penetrance due to a low sperm count and abnormal sperm morphology.
502　Moreover, we show that the *Tug1* locus harbors three distinct regulatory activities that could account for the
503　fertility defect, including (i) a *cis* DNA repressor that regulates many neighboring genes, (ii) a lncRNA that
504　can regulate genes by a *trans*-based function, and (iii) an evolutionary conserved peptide that when
505　overexpressed impacts mitochondrial membrane potential. Thus, our study provides a roadmap for future
506　studies to investigate the individual and/or combined contributions of *Tug1* DNA, RNA, and/or peptide to
507　the male fertility defect, as well as in additional diseases in which *Tug1* is altered.

508
509

510　**METHODS**
511　**Mice and ethics statement**
512　Mice used in these studies were maintained in a pathogen-specific free facility under the care and
513　supervision of Harvard University's Institutional Animal Care Committee. *Tug1*$^{tm1.1Vlcg}$ knockout mice have
514　been described previously (Goff et al., 2015; Sauvageau et al., 2013). To remove the *loxP*-flanked neomycin
515　resistance gene included in the targeting construct, we crossed *Tug*$^{tm1.1Vlcg}$ mice to *C57BL6/J* mice and then
516　to a cre-recombinase strain (*B6.C-Tg*$^{(CMV-cre)1Cgn/J}$, The Jackson Laboratory, 006054). Mice free of both the
517　neomycin-resistance and cre-recombinase genes were selected for colony expansion and subsequently
518　backcrossed to *C57BL/6J* mice. The *Tug1* knockout allele was maintained by heterozygous breeding, and
519　mutant mice were identified by genotyping for loss of the *Tug1* allele and gain of the *lacZ* cassette
520　(Transnetyx, Inc.).
521　　　　For allele specific gene expression analyses, we generated *Tug1*$^{BL6-KO/Cast-WT}$ mice by crossing inbred
522　*Mus castaneus* (Cast/EiJ) males (The Jackson Laboratory, 000928) with inbred heterozygote *Tug1* females.
523　The F1 hybrid male progeny (three wild type *Tug1*$^{BL6-WT/Cast-WT}$ and four with a maternal *Tug1* knockout allele
524　*Tug1*$^{BL6-KO/Cast-WT}$) were used for allele-specific expression studies.

525 To generate an inducible *Tug1*-overexpression mouse, tg(*Tug1*), we cloned *Tug1* cDNA (see
526 Sequences and Primers below) into a Tet-On vector (pTRE2). Full length *Tug1* (Ensembl id:
527 ENSMUST00000153313.2) was amplified from Riken cDNA clone E330021M17 (Source Bioscience) using
528 specific primers containing MluI and EcoRV restriction sites (see Sequences and Primers below). After gel
529 purification, we subcloned the amplicon using the MluI and EcoRV restriction sites into a modified Tet-On
530 pTRE2pur vector (Clontech 631013) in which the bGlobin-intron was removed. We verified the absence of
531 mutations from the cloned *Tug1* cDNA by sequencing (see Sequences and Primers below). We injected
532 this cassette into the pronucleus of C57BL/6J zygotes (Beth Israel Deaconess Medical Center Transgenic
533 Core). Two male founder mice containing the tg(*Tug1*) cassette were identified by genotyping for the pTRE
534 allele and individually mated to female C57BL/6J mice (Jackson Laboratory, 000664) to expand the
535 colonies. Next, we generated quadruple allele transgenic mice to test the functionality of the *Tug1* RNA by
536 the following strategy. We mated tg(*Tug1*) males to *Tug1*$^{tm1.1Vlcg}$ females and identified male progeny that
537 were *Tug1*$^{+/-}$; tg(*Tug1*). These mice were then mated to female rtTA mice (B6N.FVB(Cg)-Tg(CAG-
538 rtTA3)4288Slowe/J mice (Jackson Laboratory, 016532)) and we identified male progeny that were *Tug1*$^{+/-}$;
539 tg(*Tug1*), *rtTA*. Finally, we mated male *Tug1*$^{+/-}$; tg(*Tug1*), *rtTA* mice to *Tug1*$^{+/-}$ females, and at the plug date,
540 females were put on 625 mg/kg doxycycline-containing food (Envigo, TD.01306). We genotyped progeny
541 from the above matings (Transnetyx, Inc) and identified male progeny that were   *Tug1*$^{-/-}$; tg(*Tug1*), *rtTA*,
542 and maintained these mice on the doxycycline diet until the experimental end point.

543

544 **Cell Lines and Cell Culture**
545 We derived primary wild type and *Tug1*$^{-/-}$ mouse embryonic fibroblasts (MEFs) from E14.5 littermates from
546 timed *Tug1*$^{+/-}$ intercrosses as described (Xu, 2005). We maintained MEFs as primary cultures in DMEM,
547 15% FBS, pen/strep, L-glutamine and non-essential amino acids. We genotyped MEFs derived from each
548 embryo and used only male *Tug1*$^{-/-}$ and wild type littermate MEFs at passage 2 for all experiments.
549 3T3 (ATCC, CRL-1658™), HeLa (ATCC, CRM-CCL-2), and BJ (ATCC, CRL-2522™) cell lines were
550 purchased from ATCC and cultured as recommended.

551

552 **Whole Mount *In Situ* Hybridization**
553 We generated an antisense riboprobe against *Tug1* (see Sequences and Primers below) from plasmids
554 containing full length *Tug1* cDNA (Ensembl id: ENSMUST00000153313.2) and performed *in situ*
555 hybridization on a minimum of three *C57BL6/J* embryos per embryonic stage. For whole-mount staining,
556 we fixed embryos in 4% paraformaldehyde for 18 hours at 4℃, followed by three washes for 10 minutes
557 each in PBS. We then dehydrated embryos through a graded series of 25%, 50%, 75% methanol / 0.85%
558 NaCl incubations and then finally stored embryos in 100% methanol at -20°C before *in situ* hybridization.
559 We then rehydrated embryos through a graded series of 75%, 50%, 25%, methanol/ 0.85% NaCl
560 incubations and washed in 2X PBS with 0.1% Tween-20 (PBST). Embryos were treated with 10mg/mL
561 proteinase K in PBST for 10 minutes (E8.0, E9.5) or 30 minutes (E10.5, E11.5 and E12.5). Samples were
562 fixed again in 4% paraformaldehyde/0.2% glutaraldehyde in PBST for 20 minutes at room temperature and
563 washed in 2X PBST. We then incubated samples in pre-hybridization solution for 1 hour at 68°C and then
564 incubated samples in 500 ng/mL of *Tug1* antisense or sense riboprobe at 68°C for 16 hours. Post-
565 hybridization, samples were washed in stringency washes and incubated in 100 µg/mL RNaseA at 37°C for
566 1 hour. Samples were washed in 1X maleic acid buffer with 0.1% Tween-20 (MBST) and then incubated in
567 Roche Blocking Reagent (Roche, #1096176) with 10% heat inactivated sheep serum (Sigma, S2263) for 4
568 hours at room temperature. We used an anti-digoxigenin antibody (Roche, 11093274910) at 1:5000 and

569   incubated the samples for 18 hours at 4°C. Samples were washed 8 times with MBST for 15 min, 5 times
570   in MBST for 1 hour, and then once in MBST for 16 hours at 4°C. To develop, samples were incubated in
571   3X NTMT (100 mM NaCl, 100 mM Tris-HCl (pH 9.5), 50 mM MgCl2, 0.1% Tween-20, 2 mM levamisole).
572   The *in situ* hybridization signal was developed by adding BM Purple (Roche, 11442074001) for 4, 6, 8, and
573   12 hours. After the colorimetric development, samples were fixed in 4% paraformaldehyde and cleared
574   through a graded series of glycerol/1X PBS and stored in 80% glycerol. Finally, we imaged embryos on a
575   Leica M216FA stereomicroscope (Leica Microsystems) equipped with a DFC300 FX digital imaging camera.
576

577   ***Tug1* Single Molecule RNA FISH**
578   We performed *Tug1* single molecule RNA FISH as described previously (Raj et al., 2008). Briefly, 48
579   oligonucleotides labeled with Quasar 570 and Quasar 670 tiled across human/mouse *Tug1* transcripts were
580   designed with LGC Biosearch Technologies' Stellaris probe designer (Stellaris® Probe Designer version
581   4.2) and manufactured by LGC Biosearch Technologies.
582   Human foreskin fibroblasts (ATCC® CRL-2522™) and mouse 3T3 fibroblasts (ATCC, CRL-1658™)
583   were seeded on glass coverslips previously coated with poly-L-lysine (10 µg/mL) diluted in PBS. Prior to
584   hybridization, coverslips were washed twice with PBS, fixed with 3.7% formaldehyde in PBS for 10 minutes
585   at room temperature, and washed twice more with PBS. Coverslips were immersed in ice-cold 70% EtOH
586   and incubated at 4°C for a minimum of 1 hour. We then washed the coverslips with 2 mL of Wash buffer A
587   (LGC Biosearch Technologies) at room temperature for 5 minutes. Next, we hybridized cells with 80 µL
588   hybridization buffer (LGC Biosearch Technologies) containing *Tug1* probes (1:100) overnight at 37°C in a
589   humid chamber. The following day, we washed cells with 1 mL of wash buffer A for 30 minutes at 37°C,
590   followed by another wash with wash buffer A containing Hoechst DNA stain (1:1000, Thermo Fisher
591   Scientific) for 30 minutes at 37°C. Coverslips were washed with 1 mL of wash buffer B (LGC Biosearch
592   Technologies) for 5 minutes at room temperature, mounted with ProlongGold (Life Technologies) on glass
593   slides and left to curate overnight at 4°C before proceeding to image acquisition (see below).
594

595   **Sperm Counts and Morphology**
596   *Tug1*$^{-/-}$ (n=8) and wild type (n=9) males between 8 and 41 weeks of age were sacrificed and weighed. We
597   then dissected the entire male reproductive tract in phosphate buffered saline (PBS). One testis was
598   removed, weighed and fixed in 4% paraformaldehyde (PFA) for histology (see below). Sperm were collected
599   from one cauda epididymis by bisecting and suspending the tissue in a solution of Biggers-Whitten-
600   Whittingham (BWW) sperm media at 37°C. After a 15-minute incubation, we used the collected sperm
601   solutions to analyze sperm morphology and counts.
602   We characterized sperm morphology by fixing sperm in 2% PFA in PBS, mounting 20 µL of
603   suspended sperm in Fluoromount-G media (Southern Biotech) on superfrost glass slides (Thermo Fisher
604   Scientific) and scanning each slide in a linear transect, recording the morphology as normal or abnormal for
605   each sperm cell encountered (between 30 to 120 sperm). When abnormal, we also recorded the type of
606   morphological defects: headless, head angle aberrant, head bent back to midpiece, debris on head, debris
607   on hook, head misshapen, midpiece curled, midpiece kinked, midpiece stripped, debris on midpiece,
608   tailless, tail curled, tail kinked, tail broken, or multiple cells annealed together.
609   Sperm counts for each *Tug1*$^{-/-}$ (*n* = 7) and wild type (*n* = 9) mice were determined using a Countess
610   Automated Cell Counter according to manufacturer's protocol (Life Technologies, Carlsbad, CA). For the
611   *Tug1*^{rescue} experiment, sperm counts for control (WT and *Tug1*$^{+/-}$) (*n* = 2), *Tug1*$^{-/-}$ (n = 2), and *Tug1*$^{-/-}$;
612   tg(*Tug1*); *rtTA* mice (*n* = 3) was determined by manual counts using a hemocytometer. For all analyses,

14

613 statistical comparisons between *Tug1^{-/-}* and wild type was performed using the two-tailed Wilcoxon rank
614 sum tests with an a = 0.05. Results for testes, sperm counts and morphological parameters are presented
615 in Extended Data Table 1. All statistical comparisons of *Tug1^{-/-}* versus wild type for relative testis size, sperm
616 morphology and sperm counts were performed using R (Wilcoxon rank-sum test, and principal component
617 analysis (PCA)).
618

619 ***lacZ* and Histological Staining of Male Reproductive Tissues**
620 Expression of the knock-in *lacZ* reporter and histological staining for morphological analysis of male
621 reproductive tissues was conducted on testes and epididymides from *Tug1^{-/-}* (n = 2) and wild type (n = 2)
622 mice. We fixed testis and epididymis in 4% paraformaldehyde in PBS overnight at 4°C and washed tissues
623 three times in PBS. For *lacZ* staining, we rinsed *Tug1^{+/-}* and wild type tissues three times at room
624 temperature in PBS with 2 mM MgCl2, 0.01% deoxycholic acid, and 0.02% NP-40. We performed X-gal
625 staining by incubating the tissues for up to 16 hours at 37°C in the same buffer supplemented with 5 mM
626 potassium ferrocyanide and 1 mg/mL X-gal. The staining reaction was stopped by washing three times in
627 PBS at room temperature, followed by 2 hours post-fixation in 4% paraformaldehyde at 4°C.
628   We then embedded organs in paraffin, sectioned the organs at 6 μm thickness, and then mounted
629 sectioned samples onto glass microscope slides. Testis sections were additionally stained with Mayer's
630 Hematoxylin, Periodic Acid and Schiff's Reagent (VWR, 470302-348), and epididymis sections were stained
631 with eosin (VWR, 95057-848). Images were collected using a Zeiss AxioImager.A1 upright microscope or
632 on an Axio Scan Z.1 (Zeiss).
633

634 **RNA Isolation and RNA-Seq Library Preparation**
635 We isolated total RNA from mouse tissues, mouse embryonic fibroblasts (MEFs), and blood cells using
636 TRIzol (Life Technology, Carlsbad, CA) by chloroform extraction followed by spin-column purification
637 (RNeasy mini or micro kit, Qiagen) according to the manufacturer's instructions. RNA concentration and
638 purity were determined using a Nanodrop. We assessed RNA integrity on a Bioanalyzer (Agilent) using the
639 RNA 6000 chip. High quality RNA samples (RNA Integrity Number ≥ 8) were used for library preparation.
640 We then constructed mRNA-seq libraries using the TruSeq RNA Sample Preparation Kit (Illumina) as
641 previously described (Sun et al., 2013). The libraries were prepared using 500 ng of total RNA as input and
642 a 10-cycle PCR enrichment to minimize PCR artifacts. Prior to sequencing, we ran libraries on a Bioanalyzer
643 DNA7500 chip to assess purity, fragment size, and concentration. Libraries free of adapter dimers and with
644 a peak region area (220-500 bp) ≥ 80% of the total area were sequenced. We then sequenced individually
645 barcoded samples in pools of 6, each pool including *Tug1* mutant and wild type samples, on the Illumina
646 HiSeq platform using the rapid-full flow cell with the 101 bp paired-end reads sequencing protocol (Bauer
647 Core, Harvard University FAS Center for System Biology).
648

649 **RNA-seq and Gene Set Enrichment Analyses**
650 We mapped sequencing reads to the reference mouse genome (GRCm38) by STAR (Dobin et al., 2013)
651 with the gene annotation obtained from GENCODE (vM16). We counted uniquely-mapped reads for genes
652 by featureCounts (Liao et al., 2014) and calculated TPM (Transcripts Per Million) for genes to quantify gene
653 expression level after normalization of sequencing depth and gene length. Clustering of gene expression
654 was done with Ward's method using Jensen-Shannon divergence between tissues as distance metric. The
655 R package, Philentropy was used for calculation (Drost, 2018),

656   We identified differentially-expressed genes by comparing the read counts of biological replicates
657   between the groups using the generalized linear model. Statistical significance was calculated with the
658   assumption of the negative binomial distribution of the read counts and the empirical estimation of variance
659   by using the R packages DESeq2 (Love et al., 2014) and fdrtool (Strimmer, 2008). The genes were filtered
660   if their read counts were less than three in every biological replicate. The genes were called significant if
661   their FDR-adjusted p-values were smaller than 0.05.

662   We performed Gene Set Enrichment Analysis (GSEA) to evaluate the enrichment of the gene sets
663   available from MSigDB (Subramanian et al., 2005) after mapping genes to gene sets by gene symbols. The
664   statistical significance of a gene set was calculated with the test statistics of individual genes computed by
665   DESeq2. If the FDR-adjusted p-value is less than 0.1, the term was called as significant. We did this
666   calculation using the R package, CAMERA (Wu and Smyth, 2012).

667

668   **Allele-Specific Gene Expression Analysis**
669   We performed allele-specific expression analysis as previously described (Perez et al., 2015). For mouse
670   testes samples, we created a *C57BL/6J*, *Cast/EiJ* diploid genome by incorporating single nucleotide
671   polymorphisms and indels (obtained from the Mouse Genome Project: ftp://ftp-mouse.sanger.ac.uk/REL-
672   1303-SNPs_Indels-GRCm38) from both strains into the *M. musculus* GRCm38 reference genome
673   sequence. We created a transcriptome annotation set as follows. The gencode.vM2.annotation GTF file
674   was downloaded and Mt_rRNA, Mt_tRNA, miRNA, rRNA, snRNA, snoRNA, Mt_tRNA_pseudogene,
675   tRNA_pseudogene,   snoRNA_pseudogene,   snRNA_pseudogene,   scRNA_pseudogene,
676   rRNA_pseudogene, miRNA_pseudogene were removed (not enriched in our RNA-seq libraries). To create
677   an extensive set of transcripts, we added to the gencode.vM2.annotation all transcripts from the UCSC
678   knownGene mm10 annotation file, which are not represented in the gencode.vM2.annotation set. We also
679   added all functional RNAs from the Functional RNA database (fRNAdb) (Mituyama et al., 2009), which did
680   not intersect with any of the previously incorporated transcripts. From this, we then used the UCSC liftOver
681   utility to generate a *C57BL/6J*, Cast/EiJ diploid transcriptome set.

682   Each RNA-seq library was first subjected to quality and adapter trimming using the Trim Galore
683   utility (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore) with stringency level 3. We then
684   mapped each of the *C57BL/6J::Cast/EiJ* hybrid RNA-seq libraries to the *C57BL/6J* and *Cast/EiJ*
685   diploid genome and transcriptome splice junctions using STAR RNA-seq aligner (Dobin et al., 2013),
686   allowing a maximum of 3 mismatches. The data were mapped twice, where after the first mapping step we
687   incorporated valid splice junctions that were reported by STAR to exist in the RNA-seq data. We then
688   transformed the genomic alignments to transcriptomic alignments. Following that, we estimated the
689   expression levels with their respective uncertainties for each transcript in our *C57BL/6J* and *Cast/EiJ* diploid
690   transcriptome using MMSEQ (Turro et al., 2011). The posterior FPKM samples were transformed to TPM
691   units with a minimum expression TPM cutoff set to 0.01. In any RNA-seq sample, any transcript for which
692   its MMSEQ posterior median TPM was lower than 0.01 was set to 0.01 (used as the minimal measurable
693   expression level).

694   We adopted the approach of Turro et al. for combining lowly identifiable transcripts based on the
695   posterior correlation of their expression level estimates, tailored for a diploid transcriptome case (Turro et
696   al., 2014). In this approach, for any given RNA-seq sample we compute the Pearson correlation coefficient
697   of the posterior TPM samples of any pair of transcripts from the same locus and the same allele.
698   Subsequently, if the mean Pearson correlation coefficient across all RNA-seq samples for a pair of
699   transcripts in both alleles is lower than a defined cutoff (which we empirically set to -0.25), each of these

16

700 pairs is combined into a single transcript. This process continues iteratively until no pair of transcripts (or
701 pairs of already combined transcripts) can be further combined. This consistency between the alleles in the
702 combining process ensures that the resulting combined transcripts are identical for the two alleles and can
703 therefore be tested for allelically biased expression.
704
705 **Amplification of Full Length *Tug1***
706 We amplified the full length *Tug1* isoform lacking the 5' region (Ensembl id: ENSMUST00000153313.2)
707 from Riken cDNA clone E330021M17 (Source Bioscience) using specific primers containing MluI and
708 EcoRV restriction sites (see Sequences and Primers below). After gel purification, the amplicon was sub-
709 cloned, using the MluI and EcoRV restriction sites, into a modified Tet-On pTRE2pur vector (Clontech,
710 631013) in which the bGlobin-intron was removed. We verified the absence of mutations from the cloned
711 *Tug1* cDNA by sequencing using primers listed below. The plasmid was used also for sub-cloning *Tug1* into
712 pcDNA3.1(+) (see below).
713
714 **ORF Search and TUG1-BOAT Structure and Subcellular Localization Prediction**
715 We analyzed human and mouse *Tug1* cDNA sequences with CLC Genomics Workbench (Qiagen) for open
716 reading frames (ORFs), allowing both canonical and non-canonical start codons (AUG, CUG and UUG).
717 After, sequences with annotated ORFs were aligned using MUSCLE alignment. All further sequence and
718 amino acid alignments were performed with CLC Genomics Workbench.
719   We predicted secondary and tertiary structure of TUG1-BOAT using RaptorX (Källberg et al., 2012;
720 Peng and Xu, 2011), based on the *Tug1* ORF1 amino acid sequence. RaptorX was chosen for structure
721 prediction due to its ability to predict structures of proteins without known homologs. The resulting PDB files
722 of the predicted structures were visualized using PyMOL. Subcellular localization of human and mouse
723 TUG1-BOAT was predicted with DeepLoc-1.0 (Armenteros et al., 2017).
724
725 **Generation of Human and Mouse TUG1-BOAT Overexpression Constructs**
726 We generated a synthesized construct for human *Tug1* ORF1 that contained an in-frame 3xFLAG epitope
727 tag prior to the stop codon, with and without the 5' leader sequence (GeneWiz). We also synthesized a
728 construct containing mouse ORF1 with an HA tag after the 3xFLAG before the stop codon, with and without
729 the 5' leader sequence (GeneWiz).
730   We amplified the *Tug1* cDNA sequence with primers (see Sequences and Primers below) having
731 KpnI and NotI restriction enzyme overhangs from the pTRE2-*Tug1* vector plasmid using Q5 polymerase
732 (Roche) and under following conditions: 96°C for 2 minutes, 35 cycles of (96°C for 30 seconds, 65°C for 30
733 seconds, 72°C for 4 minutes), 72°C for 4 minutes, and gel purified the amplicon. We digested the inserts
734 and pcDNA3.1(+) plasmid with proper restriction enzymes according to manufacturer's instructions. After
735 digestion, the plasmid was dephosphorylated using alkaline phosphatase. We then ligated the plasmid and
736 inserts using T4 ligase (NEB) in a 1:3 ratio respectively, followed by bacterial transformation, culture growth,
737 and plasmid isolation (Qiagen Mini-Prep Kit).
738
739 **Transfection of TUG1-BOAT Constructs**
740 We seeded 3T3 and HeLa cells in 10 cm plates containing poly-L-lysine coated 18 mm glass cover slips.
741 Next, we transfected the cells with 14 µg of plasmid (pcDNA3.1(+) containing each of the inserts) using
742 Lipofectamine™ 3000 Transfection Reagent (Thermo Fisher Scientific) per manufacturer's

17

743 recommendations. 48 hours post transfection, cell pellets were harvested for protein extraction (see below)
744 and coverslips were processed for RNA FISH and/or immunofluorescence (see below).
745

746 **Protein Extraction and Western Blot**
747 We resuspended 3T3 and HeLa cell pellets in RIPA Lysis and Extraction Buffer 48 hours post transfection
748 (Thermo Fisher Scientific). Total protein was quantified with Pierce™ BCA® Protein Assay Kit (Thermo
749 Fisher Scientific). We then separated a total of 20-25 µg of denatured protein on a 12.5% SDS
750 polyacrylamide gel for 100 minutes at 120V. We transferred proteins to an Immobilon-PSQ PVDF
751 membrane (Sigma-Aldrich, ISEQ00010) at 400 mA for 75 minutes. After blocking in 5% dried milk in TBST,
752 the membrane was incubated with properly diluted primary antibody (M2 Monoclonal ANTI-FLAG 1:1000,
753 F1804, Sigma; Monoclonal GAPDH 1:5000, 2118S, CST) in 5% dried milk/TBST overnight at 4°C. The next
754 day, we washed the membrane three times for 5 minutes each in TBST (0.5% Tween-20). We then
755 incubated the membrane with Horse Radish Peroxidase-conjugated secondary antibody (Anti-mouse
756 1:15,000, A9044, Sigma; Anti-rabbit 1:10,000, 711035152, Jackson Immunoresearch), diluted in 5% dried
757 milk/TBST for 1 hour at room temperature. Following three 5 minute washes in TBST, SuperSignal™ West
758 Pico PLUS chemiluminescent substrate (Thermo Scientific, 34580) was added and chemiluminescence
759 was detected using ImageQuant™ LAS 4000 imager.
760

761 **Mouse TUG1-BOAT Localization by Immunofluorescence**
762 We plated HeLa and 3T3 cells on poly-L-Lysine coated coverslips. 48 hours post transfection, we rinsed
763 coverslips twice with PBS and fixed cells with 3.7 % formaldehyde in PBS for 10 minutes at room
764 temperature. After 2 washes with PBS, we permeabilized cells with PBT (PBS, 0.1% Tween-20) for 15
765 minutes at room temperature. Next, we blocked coverslips with 5% BSA in PBT for 1 hour at room
766 temperature and then incubated coverslips with properly diluted primary antibody (mouse M2 monoclonal
767 ANTI FLAG, 1:800, F1804, Sigma; rabbit polyclonal Tom20, 1:800, FL-145, Santa Cruz) in 5 % BSA in PBT
768 for 3 hours at 37°C in a humid chamber. Coverslips were washed three times for 5 minutes each with PBT
769 and incubated with diluted secondary antibody (anti-mouse labelled with Alexa Fluor 488, 1:800, ab150113,
770 Abcam; anti-rabbit labelled with Alexa Fluor 647, 1:800, 4414S, CST) in 5% BSA in PBT for 1 hour at room
771 temperature. Cells were then washed twice for 5 minutes with PBS, once for 20 minutes with PBS containing
772 Hoechst DNA stain (1:1000, Thermo Fisher Scientific), rinsed in PBS, and then mounted on glass slides
773 with ProLong Gold (Thermo Fisher Scientific).
774

775 **Mitochondrial Staining with MitoTracker® Red Chloromethyl-X-rosamine**
776 We plated cells on poly-L-lysisne coated coverslips and transfected as described in the previous sections.
777 48 hours post transfection, cells were incubated with 200 nM MitoTracker® Red Chloromethyl-X-rosamine
778 (Thermo Fischer Scientific, M7512) in 1 mL FBS-free growth media for 40 minutes. We then washed cells
779 twice with PBS, fixed with 3.7% formaldehyde for 10 minutes at room temperature, and processed for
780 immunofluorescence and/or RNA FISH (as described previously).
781

782 *In vitro* **Translation of Human *TUG1***
783 Synthetic gene constructs were produced by Genewiz (constructs available upon request) and designed to
784 capture (i) a selection of the endogenous human *TUG1* lncRNA, which includes the predicted ORF1 with a
785 CUG translation initiation site, the 5' UTR and 321 nucleotides of the 3' UTR (chr22:30,969261-
786 chr22:30,970,140), (ii) a codon-optimized sequence for the human *TUG1* translated ORF1 with a 3xFLAG

18

787 inserted before the termination codon, and (iii) a codon-optimized human *TUG1* translated ORF1 with an
788 mEGFP inserted before the termination codon. The sequence of the translated human *TUG1* lncRNA
789 transcript includes an alternative exon 1 transcriptional start site at chr22:30,969,261 (hg38) obtained from
790 a combination of *de novo* transcriptome assembly publicly available CAGE data. *TUG1* constructs were
791 transcribed and translated *in vitro* from 0.5 μg linearized plasmid DNA using the TnT® Coupled Wheat Germ
792 Extract system (Promega, Mannheim, Germany), in the presence of 10 mCi/mL [35S]-methionine
793 (Hartmann Analytic, Braunschweig, Germany), according to manufacturer's instructions. 5 μL of lysate was
794 denatured for 2 minutes at 85 °C in 9.6 μL Novex Tricine SDS Sample Buffer (2X) (Thermo Fisher Scientific)
795 and 1.4 μL DTT (500 mM). Proteins were separated on 16% Tricine gels (Invitrogen) for 1 hour at 50 V
796 followed by 3.5 hours at 100 V and blotted on PVDF-membranes (Immobilon-PSQ Membrane, Merck
797 Millipore). Incorporation of [35S]-methionine into newly synthesized proteins enabled the detection of
798 translation products by phosphor imaging (exposure time of 1 day).
799

800 **Human TUG1-BOAT Localization by Immunofluorescence**
801 HeLa cells were grown on glass slides for 24 hours and transfected with 3xFLAG-tagged codon optimized
802 human TUG1 ORF1 plasmid using Lipofectamine 2000 reagent for 24 hours. We fixed cells with 4%
803 paraformaldehyde for 10 minutes at room temperature and washed cells three times with ice-cold PBS. The
804 cells were permeabilized and blocked for 1 hour at room temperature using 2.5% bovine albumin serum,
805 10% anti-goat serum and 0.1% Triton X and washed again. Expressed TUG1 protein was stained for 1 hour
806 at room temperature using a monoclonal anti-FLAG mouse antibody (1:500, F1804, Sigma Aldrich) and co-
807 stained with organelle markers for mitochondria (1:1000, rabbit ATPIF1 #13268, Cell Signaling Technology;
808 Danvers, MA, USA). Afterwards, we washed the slide and incubated with fluorescently-labeled secondary
809 antibodies (1:500, Alexa Fluor 488 anti-rabbit & Alexa Fluor 594 anti-mouse (Invitrogen, Carlsbad, CA, USA)
810 for 30 minutes at room temperature. Cells were washed again, stained with 4-6-diamidino-2-phenylindole
811 (NucBlue™ Fixed Cell ReadyProbes™ Reagent, R37606, Thermo Fisher) for 5 minutes at room
812 temperature and mounted onto glass slides using ProLong™ Gold antifade reagent (Molecular Probes;
813 Invitrogen™). Images were visualized using a LEICA SP8 confocal microscope using a 63x objective. Image
814 analysis was performed using Leica confocal software Las X (v3.5.2) and ImageJ (v1.52a) (Schneider et
815 al., 2012).
816

817 **Microscopy and Image Analysis**
818 We acquired z-stacks (200 nm z-step) capturing entire cell volume for single molecule RNA FISH, single
819 molecule RNA FISH/CMXR staining, 3xFLAG tag immunofluorescence/CMXR staining and/or Tom20
820 immunofluorescence with a GE wide-field DeltaVision Elite microscope with an Olympus UPlanSApo
821 100x/1.40-NA Oil Objective lens and a PCO Edge sCMOS camera using corresponding filters. 3D stacks
822 were deconvolved using the built-in DeltaVision SoftWoRx Imaging software. Maximum intensity projections
823 of each image were subjected for quantification using Fiji.
824

825 **Fluorescence Activated Cell Sorting (FACS)**
826 Age- and sex-matched adult mice were used in all flow cytometry experiments. We obtained peripheral
827 blood by cardiac puncture and collected blood into a 1.5 mL Eppendorf tube containing 4% citrate solution.
828 Next, we added the blood-citrate mixture to 3 mL of 2% dextran/1X PBS solution and incubated for 30
829 minutes at 37°C. The upper layer was transferred to a new 5 mL polystyrene FACS tube (Falcon, #352058)
830 and centrifuged at 1200 rpm for 5 minutes at 4°C. We then lysed red blood cells for 15 minutes at room

831 temperature using BD Pharm Lyse (BD, 555899). Cells were washed twice with staining media (Hanks
832 balanced salt solution (HBSS) containing 2% FBS and 2 mM EDTA). The following antibodies were added
833 (1:100) to each sample and incubated for 30 minutes at room temperature: Alexa Fluor 700 anti-mouse
834 CD8a (Biolegend, 100730), PE/Dazzle-594 anti-mouse CD4 (Biolegend, 100456), APC anti-mouse CD19
835 (Biolegend, 115512), Alexa Fluor 488 anti-mouse NK-1.1 (Biolegend, 108718), PE anti-mouse CD3
836 (Biolegend, 100205) and Zombie Aqua Fixable Viability Kit (Biolegend, 423101) was used as a live-dead
837 stain. We washed samples twice with staining media and sorted directly into TRIzol LS using a BD Aria
838 FACS.

839

840 **qRT-PCR**
841 We isolated and quantified RNA from sorted blood populations as described in the RNA Isolation and RNA-
842 Seq Library Preparation. 100 ng of total RNA was used as input to generate cDNA using SuperScript IV
843 VILO Master Mix (Invitrogen, 11756050), according to the manufacture protocol. cDNA was diluted 1:3 with
844 DNase- and RNase-free water and 1 µL was used per each reaction. We performed qRT-PCR using
845 FastStart Universal SYBR Green Master Mix with ROX (Sigma, 4913914001) on a ViiA 7 Real-Time PCR
846 System (Thermo Fisher). Analysis was performed using the $\Delta\Delta$Ct method (Livak and Schmittgen, 2001).
847 Primers used in qRT-PCR experiments are listed in the Sequences and Primers section.

848

849 **GEO Accession Numbers**
850 All primary RNA-Seq data are available at the Gene Expression Omnibus (GSE124745 and GSE88819).

851

852

853 **SEQUENCES AND PRIMERS**
854 *In Situ* **Hybridization riboprobe - mouse** *Tug1* **(492 bp):**
855 GAGACACGACUCACCAAGCACUGCCACCAGCACUGUCACUGGGAACUUGAAGAUCCAAGUUUCUGUCCAGAACCUCAGUGC
856 AAACUGACAACACUCCAUCCAAAGUGAACUACGUCCCGGUGCCUCCUGAUUGCUGAAUGUUCACCUGGACCUGCCAAUGACC
857 UUCCUUCUGCUACUCCAUCAGCCUACAGACCUGGUACUUGGAUUUUUGUCCAUGGUGAUUCCUUCCACCUUACUACUGAAG
858 AAGACACCAUUCCAGUGGACCACUGUGACCCAAGAAGCAUUCAGCCAUCAUGAUGUGGCCUUUACCUCCACUCCUGUCCUA
859 CUCUGCCCAGAUUCAGCACAGCCCUUUAUAGUGCAGUCAAGAGUCUUCAAGCCAAAUAACUGAAGCUAUUUUAUCACAACA
860 AAGGCCAGGUUUAUUCCAUAAAUGUACAGUUCAUUUCUGCAGUUUAUUCUUCAGAGACACAUAGUAAAUUUGGACCAGGGG
861 AUUUUG

862

863 **Genotyping Primers:**
864 *Tug1*[tm1.1Vlcg] knockout mice and MEFs
865   Tug1_5190-5166TD_Forward: TGACTGGCCCAGAAGTTGTAAG
866   Tug1_5190-5166TD_Reverse: GCAAGCAGGTCTGTGAGACTATTC
867   lacZ_5_Forward: TTGAAAATGGTCTGCTGCTG
868   lacZ_5_Reverse: TATTGGCTTCATCCACCACA
869   Ychr_Forward: TCTTAAACTCTGAAGAAGAGAC
870   Ychr_Reverse: GTCTTGCCTGTATGTGATGG

871

872 **Mouse** *Tug1* **(ENSMUST00000153313.2) cDNA clone:**
873 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGGGGGTTTTTTGTTTTGTTTTTTAAATTGAAGGCTAAAGTTTTTGAAAAA
874 ACTTTGTTGGACTCTGGCTGGGACACAAAATCAGATATTTGGAATCATTTTGAAGCTTAACTTTTTCCTAACCAGCCTTGT

20

```
875   ATTCTAATTGCTTGCAAATGTGAGACTGAATGGCCAAAATGCCGTTTGTTTGTTTGTTTATTGTCAGCTGCTTTTATCAAA
876   TTCCAGGCCATTATCCAGCAAACACTATTAAAATGTTTGAACAGTTGGGTTTCAAACATTTTTGTTTTGTGGAGTGGTGCT
877   TATTAAGTGGTACAGCTCTCTAAGCAAGTGAACACAAACATATTTAAGTGTATTTTGTATGATTAGATGTTACCAATTCTG
878   ATATTTTATTCAAATGTCTAAAAAAATAAGTTGACTTATTCCCTTTACCAAAGGGCCAGAGACAAATGGTTTCCTTTCAAG
879   AGAAATGACTGTTTTGAAGAAAAACTCTGTTGGTCTTAGCTCTTTTGTAATTAAATCTGGATGTACCTCAAAAGACTCTTT
880   AAGACTGTGGTGTTAAAAGGCTTTCCTCTGGAGAAGGAGAAAAATAAAATCAACTGGAACTTAAAAGCTTGAAATTCCAT
881   GACAAAACACAGATGTCCAGGATTGGAGGTTCATAAAGTACATGCAGTAGTTGGAGTGGATTCCATTTTCAGTGTAGCTGC
882   CACCATGGACTCCAGGCTCCCAGATTTTCAAGAACTGGACCTGTGACCCAGAAGAGCTTGTCAAGATATGACAGGAACTCT
883   GGAGGTGGACGTTTTGTATTCAATTTTGGAACTGTTGATCTTGCCGTGAGAAAAGAGAGACACGACTCACCAAGCACTGCC
884   ACCAGCACTGTCACTGGGAACTTGAAGATCCAAGTTTCTGTCCAGAACCTCAGTGCAAACTGACAACACTCCATCCAAAGT
885   GAACTACGTCCCGTGCCTCCTGATTGCTGAATGTTCACCTGGACCTGCCAATGACCTTCCTTCTGCTACTCCATCAGCCTA
886   CAGACCTGGTACTTGGATTTTTGTCCATGGTGATTCCTTCCACCTTACTACTGAAGAAGACACCATTCCAGTGGACCACTG
887   TGACCCAAGAAGCATTCAGCCATCATGATGTGGCCTTTACCTCCACTCCTGTCCTACTCTGCCCAGATTCAGCACAGCCCT
888   TTATAGTGCAGTCAAGAGTCTTCAAGCCAAATAACTGAAGCTATTTTATCACAACAAAGGCCAGGTTTATTCCATAAATGT
889   ACAGTTCATTTCTGCAGTTTATTCTTCAGAGACACATAGTAAATTTGGACCAGGGGATTTTGTTTTGTTTATATTGTCAAC
890   ACTGTCTGAAGAAAGGCATCTCTGAGAACAGCATTGGACCCTACTCCACAATCTCAAATGATTGAAGTTTCATAAACTGCC
891   TAGGATCCTGTCAAGGCCACTGGACTCTTGTTCTTTTCCTACTTCAAAATCTGTAGCTGTCTACTAAATGACAAAGCAGAT
892   ATTCTGACCCATTGGGATCAAAACCAAGGCATTTTGAATTCCTCATAGTATCATCTTCGGGTTACTCAGGAACCAAAACTT
893   TTCACACCAATTTAAGAAATTCTACTGAGGAATCCCTTTACCTAACCATCTCACAAGGCTTCAACCAGATTCCTGAAAAGG
894   CCTCTTGATATATCAAGATAGAACCTACATGCATTTTGTGAACAACTTATCACTGATTTTCCAAAGGCTTTGTGCTCTTGA
895   AGTTCTTTGAAGGAAAGCTGTGTGGAAGTCCAGAGTAAAGTGAAGCTGCTCTGGATGAAGTAGTGAAGTGGGAGTTGAGGT
896   CTACAACCTGCCACAACCATCTTCCTTTACCACCATGGTGATGCCAAAAGGGACTTCCTTAAAGCTCTTCAGAAAATCCTG
897   CTTGAAACCACTACCCTAGACAACATGTTTGACCTGGATGGCATTCTCTTCAAAACAATTCATATTCAGTTGATGCTCAAC
898   ATGTTTGGAGATGCTTTATTCAGAGAATGATGATAATTACAGCATTGTCTAATGAAGTTTTATTAATAGCATTCCATCCAA
899   GGTGGACTTCCTGGAGTTGGATATAACCAGAGAGCAATTCATATGTATCCTACACTGAAGAACACCATTAACTTTCAGCAA
900   CCTATAGCTAGTGGTACTAGAAGTACGTGTCTTGGAAGTCTATGAGAGCTGGTATTGAAGCTGATGCCTCCTTAAGGCCAT
901   CTTAGACCAAGTTGTTTGTTTGACCTCTCCTCATTAACTATGGAGCAGAATTGAAATACAAATTTTTCCTAAAGGGACTTG
902   CAACCTGGTTATCATTCATTATCTCAAGTTTCAAGTCATGTTGATGCAACCAGTAGTTATTAAACTGCTCCATGGTTTTTT
903   GTTATTTAATACTTTTTCCAGGGCTTAAAAAAACAAAATTAAATTTCTCCAACACGTCTATACTTGTCTGTTCAAAAGTAA
904   CTACTCACCACTATATGGAACAGATGATTCTGAAGACACTCTGAGCATCCTTTATGATATTTGTGACTTAAAATGTGGCTG
905   GAAATTTTCCTTCTACCCAGTGAAATATTTAATGATTAGTCTTCATGCCTGATACCATCAACTGTATATGCGTGATAGGCA
906   AAGTTTGACATAGGCATTTGACTCTAGGCTATGATAGCTTGCTAGTAACTTCAAGTAGCATATTGTCAACCTGTTTGCTGG
907   AAAAGTAGAGTAACTTGGAAAAAAAACTAAATGGCAGCTAAGGATTTTTTTCAGTATTCCTGAGTTTCTGTCCTTGGGATA
908   TTTCAATGAAATTTTCACCTGTCTCTTCACTTAACAGAGTGACTGACTCCTTACTATGAAGTATTCTTAAGACATTAAGAT
909   TACTTTTGTAGAAAGGATAAAATTCCTGACCATCCAAATCATCATAGTGAACAAGACTTCAATTTGTGACCTGAGAAAATC
910   TCATTTCTCTACTTCGTAGTCAATGTAAGGGCCAATGCTATCAGCTACTCTGAGTGCACTGGGTAAACGTTGGAACTGCCT
911   TCTTTATATCATTACTTTTTATCCTCTAAATTAATCATGGTTATGTAATTCTCGCCACAAATCAGCAAATCAGACTCAGAT
912   CTGGTTATTCTAGACTGCTCACAGTTAACAAATCAAACTCTGGATGACTTCTGCTTGTATATGCAACTACTATTTGTAAAG
913   AAATTGCAAATTCACTTTTCTATTACCTCTACATTGCTAGCTCTTTCTTTTGTGTTTGTATTAAAAACAAAAATAAGCTAC
914   ACTGCCAGCTATTCCCTCCTGCCATACTCAGTTAAAATGAAGAATCGGGAATCTAACCAGTGAATGGATAAGTAGAAAAAA
915   CTAAAACTTAAGGCAAAAGCCTTAATCTAGGGCCTTTTCTACTATCTTCATGTCTTGGATTTCATCTAAAATCAACAGTGC
916   CACCCAACCAGTCTGAGGTCTTGACTTGCTTTTAAGATGATTCTTAGAGATGGGCTGTATTACAGAAGGTGAAGACTTGAT
917   TACCAAAGAAAGTAGAGCCAACTTTGACAAACCTGGCTCTACAATCCTATTGCTTCCAGATGTAGCATAGACTCATAACTA
918   GAACCTCAAGTCTGCATTGAGGATATAGCCTTCTAAGCTGACAGTTCTTGCAACAGGTGAGCAAGAAAATGAAAGCTGTTA
919   TACCCAACTGGCCCTTTAAGATCCAAAAATAATGTCTGGACTAAACCCTATGGAGTACCCAGGACAAAAACTAATTTACAG
920   AGCTTCATTATTAATCTGCCTGTTCTTCTAGCTTAATTATTGGTATGGCTGGCCCTACTGAAGTAGTTTGTCTGTTTACCT
921   GTCTTCAGCTCTTAACCTGGCTATTTTGACATGCTACTGCAATTAGACTAACTGGCTTTGAGAAGACTACAATCAGTTTCA
```

```
922  GCCTCTCCTTTGCCCAATTTCACCAAGGAATTTTGATAAGAGGAACCCATACCTCACCCCACCAGAACAGAAAGGACCATG
923  CTGCATATTCCTTGACCAGCAACTTTAAGTAGAGAACAACCCTGCTTGTTTTCAACATCTGAAACACCATTTGATCTAATA
924  GGAGTATAGAAGGTTGACAGCAGAGTACACTACTTACTTCTTTCATAACTCAGAAATGAATATGACTGGCCCAGAAGTTGT
925  AAGTTCACCTTGACAAGAAACAGCAACACCAGAAGTTTACTGCTGAACTTAACTTGCCACTTACTCGAATAGTCTCACAGA
926  CCTGCTTGCCAAGTAGGAGGCTAGTTTTCCTGCTTCATATCACCATTGGAGTGGGGCTCAATGGGGTCAATGTTAATACTG
927  ACTTGAATGGGGACCTTATGGTGAATCCTAGACTATGAGGCTAATGGAAATTATTGTCTATTCAAGTGGATTATAGATTTC
928  CTGAGGACAGAACAGACATCACTCCTGGTGATTTTTAGAACTTGATTACCAAGGAAGAAATACCAGCTGCTAACAGTCAAC
929  TTCATGGGCAAAGATTAAGCTCTCTATATCTGGTCGTATCCTGGATGCTAGTTTTTTATTGCCCAGTGACCATTTCCATCT
930  CACGCTTAACTTCCTGATGTTTTTTGGAACCATCTCTTCCAATTTTCAGTCCTGGTGATTTAGACAGTCTTTTCATGCTGG
931  ACATTTTGTTGCAACCTCATCAATCACAGCAAAGTCCATCTTGACTTTAGTGATTAGTTCAGGAATGGATGCATGATTCAA
932  GTTTGTCCAATGATAATCAACCCTAGGTGTTTTCTCAGTTGTGGAGAAGTTCTCTTAGATGCTTTAGCTTTGTAGGAGAAA
933  ACTCAAACCAACAGGGCCTACCTACTATGTTGAATGATTGTAGGAGAAAACTCAAACCAACCAGGCCTACCTACTATGTTG
934  AATGAGCCAGGCAGAAAATGAAGCCAGTACAGAGGGAAATGGAGCCAAAAGAGGAAGAGACTTGAGTTCTGATGATCACAT
935  TTATGCCCCTGTATCCAACTGTGCCTGAAGCTAATAGTACATCACCTGGACTTTTCAGTTATGTGAACCAATAAATTCCCC
936  TTTTTGTTTAAGTTACTTTGAGTT
937
938  Full Length Tug1 primers for cloning in pTRE2pur:
939       Tug1_MluI_cDNA_Fwd
940            gagaacgcgtTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGGGGGTTTTTTGTTTTGTTTTTTAAATTGA
941            AGGCTAAAGTTTTTGAAAAAACTTTGTTGGACTCTGGCTGG
942       Tug1_EcoRV_cDNA_Rev
943            gagagatatcAACTCAAAGTAACTTAAACAAAAAGGG
944
945  Primers for full length sequencing of pTRE2-Tug1 expression vector:
946       LNCX            AGCTCGTTTAGTGAACCGTCAGATC
947       TUG1_76         TTTAAATTGAAGGCTAAAGTTTTTGAA
948       TUG1_266        GGCCATTATCCAGCAAACAC
949       TUG1_755        ACTCCAGGCTCCCAGATTTT
950       TUG1_1254       TCTTCAAGCCAAATAACTGAAGC
951       TUG1_1741       AGAACCTACATGCATTTTGTGAA
952       TUG1_2268       ATGCCTCCTTAAGGCCATCT
953       TUG1_2754       TGTCAACCTGTTTGCTGGAA
954       TUG1_3267       TTGCAAATTCACTTTTCTATTACCTC
955       TUG1_3746       CCCAACTGGCCCTTTAAGAT
956       TUG1_4241       TGACAAGAAACAGCAACACCA
957       TUG1_4740       TCACAGCAAAGTCCATCTTGA
958
959  Primers for qRT-PCR:
960       Tug1_Fwd        CTCTGGAGGTGGACGTTTTGT
961       Tug1_Rev        GTGAGTCGTGTCTCTCTTTTCTC
962       Gapdh_Fwd       GGTGAAGGTCGGTGTGAACG
963       Gapdh_Rev       CTCGCTCCTGGAAGATGGTG
964
965
```

Primers for sub-cloning full length *Tug1* from pTRE2-*Tug1* into pcDNA3.1(+) expression vector:

Tug1_Tg F/KpnI `ataggtaccGCCCCGAATTCACGCGTT`

Tug1_Tg R/NotI `atagcggccgcACCTGAGGAGTGAAGA`

Human TUG1-BOAT (ORF1) sequences with 3xFLAG (blue) for expression construct design:

Human ORF1:

```
CTGGCGCGCCCTCCCCCCCTCCCGGGTCTGGTAGGGCGAAGGAACGGGCGTGCGGTCGATCGAGCGATCGGTTGGC
GGCTCTTTCTCCTGCTCTGGCATCCAGCTCTTGGGGCGCAGGCCCGGCCGCCGCGGCGCGCGCCCGGTGGCCGTTG
GCGCTCGCGCCGCGTCTTTCTTCTCGTACGCAGAACTCGGGCGGCGGCCTATGCGTTTGCGATTCGACGAGGAGTC
GTCCGGGTGGTCGGCGGCGGCGGGCAGCTGCTCCGCCCCGCTCCGGGGGAGGCGGCGGCGGCAGCGGCCGCGGGAT
TTGGAGCGGCCGGGGAGGCGGGGGTGGCCGGGGCCGGCTTGGAGGCCTGGCGCCACCCTTCGGGGCCTGCAAGGAC
CCAGTTGGGGGGGCAGGAGGGGGCCGGAGGATGGTTGGTTGTGGGATTTCTACTTTGCCTTTTCCTCCTTATGCCG
CCTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATGACAAGTAG
```

Human ORF1+UTR:

```
GGCCGAGCGACGCAGCCGGGACGGTAGCTGCGGTGCGGACCGGAGGAGCCATCTTGTCTCGTCGCCGGGGAGTCAG
GCCCCTAAATCGAAGAAGCCCTGGCGCGCCCTCCCCCCCTCCCGGGTCTGGTAGGGCGAAGGAACGGGCGTGCGGT
CGATCGAGCGATCGGTTGGCGGCTCTTTCTCCTGCTCTGGCATCCAGCTCTTGGGGCGCAGGCCCGGCCGCCGCGG
CGCGCGCCCGGTGGCCGTTGGCGCTCGCGCCGCGTCTTTCTTCTCGTACGCAGAACTCGGGCGGCGGCCTATGCGT
TTGCGATTCGACGAGGAGTCGTCCGGGTGGTCGGCGGCGGCGGGCAGCTGCTCCGCCCCGCTCCGGGGGAGGCGGC
GGCGGCAGCGGCCGCGGGATTTGGAGCGGCCGGGGAGGCGGGGGTGGCCGGGGCCGGCTTGGAGGCCTGGCGCCAC
CCTTCGGGGCCTGCAAGGACCCAGTTGGGGGGGCAGGAGGGGGCCGGAGGATGGTTGGTTGTGGGATTTCTACTTT
GCCTTTTCCTCCTTATGCCGCCTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGA
TGACGATGACAAGTAG
```

Mouse TUG1-BOAT (ORF1) sequences with 3xFLAG (blue) and HA (red) tags for expression construct design:

Mouse ORF1:

```
CTGGCGCGCCCTCCCCCCCTCCCCGGTCTGGTAGGGCGAAGGAGCGGGCGTGCGGTCGATCGAGCGATCGGTTGGC
GGCTCTTTCTCCTGCTCTGGCATCCAGCTCTTGGGGCGCAGGCCCGGCCGCCGCGGCGCGCGCCCGGTGGCCGTTG
GCGCTCGCGCCGCGTCTTTCTTCTCGTACGCAGAACTCGGGCGGCGGCCTATGCTTTTGCGATCCGACGAGGGGTC
GTCCGGGTGGTTGGCGGCGGCGGGCAACTCCGCCCCGCTCCCGGGGAGGCGGCGGGGGAAGCTGGGGTGGCCGGGG
CTGGCCTGGAGGCCTGGCGCCACCCCTCGGGGCCTGCTAGGACCCAGTTGGAGGGTCAAGAGGGAGCTGGAGGATG
GTTGGTGGTGGGCTTCCTCCTTTGCCTTTTCCTACTTATGCCACCTGACTACAAAGACCATGACGGTGATTATAAA
GATCATGACATCGACTACAAGGATGACGATGACAAGTACCCATACGATGTTCCAGATTACGCTTAG
```

Mouse ORF1+UTR:

```
GGCCGAGAGACGCAGCCGGGACGGTAGCTGCAGAGCAGAGCGGAGGAGCCATCTTGTCTTGTCGCCGGGGAGTCAG
GCCCCTAACTCGAAGAAGCCCTGGCGCGCCCTCCCCCCCTCCCCGGTCTGGTAGGGCGAAGGAGCGGGCGTGCGGT
CGATCGAGCGATCGGTTGGCGGCTCTTTCTCCTGCTCTGGCATCCAGCTCTTGGGGCGCAGGCCCGGCCGCCGCGG
CGCGCGCCCGGTGGCCGTTGGCGCTCGCGCCGCGTCTTTCTTCTCGTACGCAGAACTCGGGCGGCGGCCTATGCTT
TTGCGATCCGACGAGGGGTCGTCCGGGTGGTTGGCGGCGGCGGGCAACTCCGCCCCGCTCCCGGGGAGGCGGCGGG
GGAAGCTGGGGTGGCCGGGGCTGGCCTGGAGGCCTGGCGCCACCCCTCGGGGCCTGCTAGGACCCAGTTGGAGGGT
CAAGAGGGAGCTGGAGGATGGTTGGTGGTGGGCTTCCTCCTTTGCCTTTTCCTACTTATGCCACCTGACTACAAAG
ACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATGACAAGTACCCATACGATGTTCCAGA
TTACGCTTAG
```

23

DECLARATIONS

Ethics Approval and Consent to Participate

Not applicable

Consent for Publication

Not applicable

Availability of Data and Materials

The datasets generated and/or analyzed during the current study are available in the Gene Expression Omnibus repository, GSE124745 and GSE88819.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

Conceptualization, M.S., J.L.R.; Methodology, M.S., E.J.P., J.L., G.D., T.H.; Computational Analyses, T.H., A.F.G., N.D.R.; Experimental Analyses, M.S., E.J.P., A.F.G., N.D.R., J.L., G.D., A.W., C.G., C.K., K.T.; J.S., C.M., S.H.; Investigation, M.S., E.J.P., S.C.L., C.G., J.L., G.D., A.W., N.C., C.K., K.T., J.S., C.M., S.H.; Data Curation, M.S., E.J.P. T.H., J.L., G.D., A.W.; Writing Draft, M.S., J.L.R., G.D., J.L., C.M. T.H.; Figures, M.S., E.J.P., C.M., J.L., G.D., A.W., T.H.; Supervision, Administration, and Funding Acquisition, J.L.R., H.E.H., N.H.; All authors have read and approved the final manuscript.

Authors' Information

Not applicable

REFERENCES

Anderson DM, Anderson KM, Chang C-L, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, Olson EN. 2015. A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance. *Cell* **160**:595–606. doi:http://dx.doi.org/10.1016/j.cell.2015.01.009

Anguera MC, Ma W, Clift D, Namekawa S, Kelleher RJ, Lee JT. 2011. Tsx produces a long noncoding RNA and has general functions in the germline, stem cells, and brain. *PLoS Genet* **7**:1–14. doi:10.1371/journal.pgen.1002248

Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. 2017. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**:3387–3395. doi:10.1093/bioinformatics/btx431

Arun G, Akhade VS, Donakonda S, Rao MRS. 2012. mrhl RNA, a Long Noncoding RNA, Negatively Regulates Wnt Signaling through Its Protein Partner Ddx5/p68 in Mouse Spermatogonial Cells. *Mol Cell Biol* **32**:3140–3152. doi:10.1128/MCB.00006-12

Bester AC, Lee JD, Chavez A, Lee YR, Nachmani D, Vora S, Victor J, Sauvageau M, Monteleone E, Rinn JL, Provero P, Church GM, Clohessy JG, Pandolfi PP. 2018. An Integrated Genome-wide CRISPRa Approach to Functionalize lncRNAs in Drug Resistance. *Cell* **173**:649–664.e20. doi:10.1016/j.cell.2018.03.052

Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, Rinn JL, Raj A. 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* **16**:20. doi:10.1186/s13059-015-0586-4

Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL, Byron M, Monks B, Henry-bezy M, Lawrence JB, Neill L a JO, Moore MJ, Caffrey DR, Fitzgerald KA. 2013. A Long Noncoding RNA Mediates Both Activation and Repression of Immune Response Genes. *Science (80-)* **341**:789–792. doi:10.1126/science.1240925

Chen J, Shishkin AA, Zhu X, Kadri S, Maza I, Guttman M, Hanna JH, Regev A, Garber M. 2016. Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol* **17**:1–17. doi:10.1186/s13059-016-0880-9

Chng SC, Ho L, Tian J, Reversade B. 2013. ELABELA : A Hormone Essential for Heart Development Signals via the Apelin Receptor. *Dev Cell* **27**:672–680. doi:10.1016/j.devcel.2013.11.002

Coutton C, Escoffier J, Martinez G, Arnoult C, Ray PF. 2014. Teratozoospermia: Spotlight on the main genetic actors in the human. *Hum Reprod Update* **21**:455–485. doi:10.1093/humupd/dmv020

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. doi:10.1093/bioinformatics/bts635

Drost H. 2018. Philentropy : Information Theory and Distance Quantification with R. *J Open Source Softw* **3**:1–4. doi:10.21105/joss.00765

Du Z, Sun T, Hacisuleyman E, Fei T, Wang X, Brown M, Rinn JL, Lee MGS, Chen Y, Kantoff PW, Liu XS. 2016. Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nat Commun* **7**:1–10. doi:10.1038/ncomms10982

Elling R, Robinson EK, Shapleigh B, Liapis SC, Covarrubias S, Katzman S, Groff AF, Jiang Z, Agarwal S, Motwani M, Chan J, Sharma S, Hennessy EJ, FitzGerald GA, McManus MT, Rinn JL, Fitzgerald KA, Carpenter S. 2018. Genetic Models Reveal cis and trans Immune-Regulatory Activities for lincRNA-Cox2. *Cell Rep* **25**:1511–1524.e6. doi:10.1016/j.celrep.2018.10.027

100   Fagerberg L, Hallstro M, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S,
101   Danielsson A, Edlund K, Asplund A, Sjo E, Lundberg E, Szigyarto CA, Skogs M, Ottosson J, Berling
102   H, Tegel H, Mulder J, Nilsson P, Schwenk JM, Lindskog C, Danielsson F, Mardinoglu A, Sivertsson
103   Å, Feilitzen K Von, Forsberg M, Zwahlen M, Olsson I, Navani S, Huss M, Nielsen J, Ponten F, Uhle
104   M. 2014. Analysis of the Human Tissue-specific Expression by Genome-wide Integration of
105   Transcriptomics and Antibody-based. *Mol Cell Proteomics* **13**:397–406.
106   doi:10.1074/mcp.M113.035600

107   Feng J, Funk W, Wang S-S, Weinrich S, Avilion A, Chiu C-P, Adams R, Chang E, Allsopp R, Yu J, Le S,
108   West M, Harley C, Andrews W, Greider C, Villeponteau B. 1995. The RNA component of human
109   telomerase. *Science (80-)* **269**:1236–1241.

110   Firlit BCF, Davis JR. 1965. Morphogenesis of the residual body of the mouse testis. *J Cell Sci* **106**:93–98.

111   Goff LA, Groff AF, Sauvageau M, Trayes-Gibson Z, Sanchez-Gomez DB, Morse M, Martin RD, Elcavage
112   LE, Liapis SC, Gonzalez-Celeiro M, Plana O, Li E, Gerhardinger C, Tomassy GS, Arlotta P, Rinn JL.
113   2015. Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the
114   mouse brain. *Proc Natl Acad Sci U S A* **112**:6855–6862. doi:10.1073/pnas.1411263112

115   Goudarzi M, Berg K, Pieper LM, Schier AF. 2019. Individual long non-coding RNAs have no overt
116   functions in zebrafish embryogenesis , viability and fertility. *Elife* **8**:1–17.

117   Groff AF, Barutcu AR, Lewandowski JP, Rinn JL. 2018. Enhancers in the Peril lincRNA locus regulate
118   distant but not local genes. *Genome Biol* **19**:1–14. doi:10.1186/s13059-018-1589-8

119   Groff AF, Sanchez-gomez DB, Gerhardinger C, Barutcu AR, Li E, Elcavage L, Plana O, Sanchez L V,
120   Sauvageau M, Rinn JL. 2016. In vivo characterization of Linc-p21 reveals functional cis- regulatory
121   DNA elements. *Cell Rep* **16**:2178–2186. doi:10.1016/j.celrep.2016.07.050.In

122   Groff AF, Sanchez-Gomez DB, Soruco MML, Gerhardinger C, Barutcu AR, Li E, Elcavage L, Plana O,
123   Sanchez L V., Lee JC, Sauvageau M, Rinn JL. 2016. In Vivo Characterization of Linc-p21 Reveals
124   Functional cis-Regulatory DNA Elements. *Cell Rep* **16**:2178–2186. doi:10.1016/j.celrep.2016.07.050

125   Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP,
126   Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn
127   JL, Lander ES. 2009. Chromatin signature reveals over a thousand highly conserved large non-
128   coding RNAs in mammals. *Nature* **457**:223–227. doi:10.1038/nature07672

129   Han Y, Liu Y, Gui Y, Cai Z. 2013. Long intergenic non-coding RNA TUG1 is overexpressed in urothelial
130   carcinoma of the bladder. *J Surg Oncol* **107**:555–559. doi:10.1002/jso.23264

131   He Q, Yang S, Gu X, Li M, Wang C, Wei F. 2018. Long noncoding RNA TUG1 facilitates osteogenic
132   differentiation of periodontal ligament stem cells via interacting with Lin28A. *Cell Death Dis* **9**:455.
133   doi:10.1038/s41419-018-0484-2

134   Heinen TJAJ, Staubach F, Häming D, Tautz D. 2009. Emergence of a New Gene from an Intergenic
135   Region. *Curr Biol* **19**:1527–1531. doi:10.1016/j.cub.2009.07.049

136   Hill KE, Zhou J, McMahan WJ, Motley AK, Atkins JF, Gesteland RF, Burk RF. 2003. Deletion of
137   selenoprotein P alters distribution of selenium in the mouse. *J Biol Chem* **278**:13640–13646.
138   doi:10.1074/jbc.M300755200

139   Housman G, Ulitsky I. 2016. Methods for distinguishing between protein-coding and long noncoding RNAs
140   and the elusive biological purpose of translation of long noncoding RNAs. *BBA - Gene Regul Mech*
141   **1859**:31–40. doi:10.1016/j.bbagrm.2015.07.017

142   Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-Wide Analysis in Vivo of
143   Resolution Using Ribosome Profiling. *Science (80-)* **324**:218–23. doi:10.1126/science.1168978

144 Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. 2012. Template-based protein structure
145   modeling using the RaptorX web server. *Nat Protoc* **7**:1511–1522. doi:10.1038/nprot.2012.085

146 Katsushima K, Natsume A, Ohka F, Shinjo K, Hatanaka A, Ichimura N, Sato S, Takahashi S, Kimura H,
147   Totoki Y, Shibata T, Naito M, Kim HJ, Miyata K, Kataoka K, Kondo Y. 2016. Targeting the Notch-
148   regulated non-coding RNA TUG1 for glioma treatment. *Nat Commun* **7**:1–14.
149   doi:10.1038/ncomms13616

150 Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson
151   M, Furlotte NA, Eskin E, Nellåker C, Whitley H, Cleak J, Janowitz D, Hernandez-Pliego P, Edwards
152   A, Belgard TG, Oliver PL, McIntyre RE, Bhomra A, Nicod J, Gan X, Yuan W, Van Der Weyden L,
153   Steward CA, Bala S, Stalker J, Mott R, Durbin R, Jackson IJ, Czechanski A, Guerra-Assunçao J,
154   Donahue LR, Reinholdt LG, Payseur BA, Ponting CP, Birney E, Flint J, Adams DJ. 2011. Mouse
155   genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**:289–294.
156   doi:10.1038/nature10413

157 Khalil A, Guttman M, Huarte M, Garber M, Raj A, Morales D, Thomas K, Presser A, Berstein B,
158   Oudenaarden A, Regev A, Lander E, Rinn J. 2009. Many human large intergenic noncoding RNAs
159   associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci*
160   **106**:11667–11672. doi:10.1073/pnas.0904715106

161 Kopp F, Mendell JT. 2018. Review Functional Classification and Experimental Dissection of Long
162   Noncoding RNAs. *Cell* **172**:393–407. doi:10.1016/j.cell.2018.01.011

163 Lai KMV, Gong G, Atanasio A, Rojas J, Quispe J, Posca J, White D, Huang M, Fedorova D, Grant C,
164   Miloscio L, Droguett G, Poueymirou WT, Auerbach W, Yancopoulos GD, Frendewey D, Rinn J,
165   Valenzuela DM. 2015. Diverse phenotypes and specific transcription patterns in twenty mouse lines
166   with ablated lincRNAs. *PLoS One* **10**:1–21. doi:10.1371/journal.pone.0125522

167 Lee JT, Jaenisch R. 1997. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome.
168   *Nature* **386**:275–279.

169 Li J, Zhang M, An G, Ma Q. 2016. LncRNA TUG1 acts as a tumor suppressor in human glioma by
170   promoting cell apoptosis. *Exp Biol Med* **241**:644–649. doi:10.1177/1535370215622708

171 Li LM, Arnosti DN. 2011. Long- and short-range transcriptional repressors induce distinct chromatin states
172   on repressed genes. *Curr Biol* **21**:406–412. doi:10.1016/j.neuroimage.2013.08.045.The

173 Li Q, Lewandowski JP, Powell MB, Norrie JL, Cho SH, Vokes S a. 2014. A Gli silencer is required for
174   robust repression of gremlin in the vertebrate limb bud. *Development* **141**:1906–14.
175   doi:10.1242/dev.104299

176 Liao Y, Smyth GK, Shi W. 2014. FeatureCounts: An efficient general purpose program for assigning
177   sequence reads to genomic features. *Bioinformatics* **30**:923–930. doi:10.1093/bioinformatics/btt656

178 Likić VA, Perry A, Hulett J, Derby M, Traven A, Waller RF, Keeling PJ, Koehler CM, Curran SP, Gooley
179   PR, Lithgow T. 2005. Patterns that define the four domains conserved in known and novel isoforms
180   of the protein import receptor Tom20. *J Mol Biol* **347**:81–93. doi:10.1016/j.jmb.2004.12.057

181 Lin M, Jungreis I, Kellis M. 2011. PhyloCSF: a comparative genomics method to distinguish protein coding
182   and non-coding regions. *Bioinformatics* **27**:275–282. doi:10.1093/bioinformatics/btr209

183 Livak K, Schmittgen T. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative
184   PCR and the delta delta Ct Method. *Methods* **25**:402–408. doi:10.1006/meth.2001.1262

185 Long J, Overbeek PA, Danesh FR, Long J, Badal SS, Ye Z, Wang Y, Ayanga BA, Galvan DL, Green NH,
186   Chang BH, Overbeek PA, Danesh FR. 2016. Long noncoding RNA Tug1 regulates mitochondrial
187   bioenergetics in diabetic nephropathy. *J Clin Invest* **126**:4205–4218. doi:10.1172/JCI87927.14

188   Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq
189   data with DESeq2. *Genome Biol* **15**:1–21. doi:10.1186/s13059-014-0550-8

190   Macho A, Decaudin D, Castedo M, Hirsch T, Susin SA, Zamzami N, Kroemer G. 1996. Chloromethyl-X-
191   rosamine is an aldehyde-fixable potential-sensitive fluorochrome for the detection of early apoptosis.
192   *Cytometry* **25**:333–340. doi:10.1002/(SICI)1097-0320(19961201)25:4<333::AID-CYTO4>3.0.CO;2-E

193   Makarewich CA, Olson EN. 2017. Mining for Micropeptides. *Trends Cell Biol* **27**:685–696.
194   doi:10.1016/j.tcb.2017.04.006

195   Michel AM, Fox G, M. Kiran A, De Bo C, O'Connor PBF, Heaphy SM, Mullan JPA, Donohue CA, Higgins
196   DG, Baranov P V. 2014. GWIPS-viz: Development of a ribo-seq genome browser. *Nucleic Acids Res*
197   **42**:859–864. doi:10.1093/nar/gkt1035

198   Mituyama T, Yamada K, Hattori E, Okida H, Ono Y, Terai G, Yoshizawa A, Komori T, Asai K. 2009. The
199   functional RNA database 3.0: Databases to support mining and annotation of functional RNAs.
200   *Nucleic Acids Res* **37**:89–92. doi:10.1093/nar/gkn805

201   Nelson B, Makarewich C, Anderson D, Winders B, Troupes C, Wu F, Reese A, McAnally J, Chen X,
202   Kavalai E, Cannon S, Houser S, Bassel-Duby R, Olson EN. 2016. A peptide encoded by a transcript
203   annotated as long noncoding RNA enhances SERCA activity in muscle. *Science (80-)* **351**:271–275.
204   doi:10.1126/science.aad4076

205   Niessen P, Rensen S, Van Deursen J, De Man J, De Laet A, Vanderwinden JM, Wedel T, Baker D,
206   Doevendans P, Hofker M, Gijbels M, Van Eys G. 2005. Smoothelin-A is essential for functional
207   intestinal smooth muscle contractility in mice. *Gastroenterology* **129**:1592–1601.
208   doi:10.1053/j.gastro.2005.08.018

209   O'Donnell L, Nicholls PK, O'Bryan MK, McLachlan RI, Stanton PG. 2011. Spermiation. *Spermatogenesis*
210   **1**:14–35. doi:10.4161/spmg.1.1.14525

211   Paralkar VR, Taborda CC, Huang P, Yao Y, Kossenkov A V., Prasad R, Luan J, Davies JOJ, Hughes JR,
212   Hardison RC, Blobel GA, Weiss MJ. 2016. Unlinking an lncRNA from Its Associated cis Element. *Mol*
213   *Cell* **62**:104–110. doi:10.1016/j.molcel.2016.02.029

214   Peng J, Xu J. 2011. RaptorX: exploiting structure information for protein alignment by statistical inference.
215   *Proteins* **79**:161–171. doi:10.1007/978-3-319-27282-5_30

216   Penny G, Kay G, Sheardown S, Rastan S, Brockdorff N. 1996. Requirement for Xist in X chromosome
217   inactivation. *Nature* **379**:131–137. doi:10.1038/379131a0

218   Perez JD, Rubinstein ND, Fernandez DE, Santoro SW, Needleman LA, Ho-Shing O, Choi JJ, Zirlinger M,
219   Chen SK, Liu JS, Dulac C. 2015. Quantitative and functional interrogation of parent-of-origin allelic
220   expression biases in the brain. *Elife* **4**:41. doi:10.7554/eLife.07860

221   Qi H, Liu M, Emery DW, Stamatoyannopoulos G. 2015. Functional validation of a constitutive autonomous
222   silencer element. *PLoS One* **10**:1–14. doi:10.1371/journal.pone.0124588

223   Rainey MA, George M, Ying G, Akakura R, Burgess DJ, Siefker E, Bargar T, Doglio L, Crawford SE, Todd
224   GL, Govindarajan V, Hess RA, Band V, Naramura M, Band H. 2010. The endocytic recycling
225   regulator EHD1 is essential for spermatogenesis and male fertility in mice. *BMC Dev Biol* **10**:37.
226   doi:10.1186/1471-213X-10-37

227   Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. 2008. Imaging individual mRNA
228   molecules using multiple singly labeled probes. *Nat Methods* **5**:877–879. doi:10.1038/nmeth.1253

229

230

231

232 Sato H, Taketomi Y, Isogai Y, Miki Y, Yamamoto K, Masuda S, Hosono T, Arata S, Ishikawa Y, Ishii T,
233   Kobayashi T, Nakanishi H, Ikeda K, Taguchi R, Hara S, Kudo I, Murakami M. 2010. Group III
234   secreted phospholipase A2 regulates epididymal sperm maturation and fertility in mice. *J Clin Invest*
235   **120**. doi:10.1172/JCI40493.1400

236 Sauvageau M, Goff L a., Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB,
237   Hacisuleyman E, Li E, Spence M, Liapis SC, Mallard W, Morse M, Swerdel MR, D'Ecclessis MF,
238   Moore JC, Lai V, Gong G, Yancopoulos GD, Frendewey D, Kellis M, Hart RP, Valenzuela DM,
239   Arlotta P, Rinn JL. 2013. Multiple knockout mouse models reveal lincRNAs are required for life and
240   brain development. *Elife* **2013**:1–24. doi:10.7554/eLife.01749

241 Schneider CA, Rasband WS, Eliceiri KW, Instrumentation C. 2012. NIH Image to ImageJ : 25 years of
242   Image Analysis. *Nat Methods* **9**:671–675.

243 Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL,
244   Saghatelian A. 2013. Peptidomic discovery of short open reading frame-encoded peptides in human
245   cells. *Nat Chem Biol* **9**:59–64. doi:10.1038/nchembio.1120

246 Sonenberg N, Wilchek M, Zamir A. 1975. Identification of a region in 23S rRNA located at the peptidyl
247   transferase center. *Proc Natl Acad Sci* **72**:4332–4336.

248 Stein CS, Jadiya P, Zhang X, Anderson EJ, Elrod JW, Boudreau RL, Stein CS, Jadiya P, Zhang X,
249   Mclendon JM, Abouassaly GM, Witmer NH. 2018. Mitoregulin : A lncRNA-Encoded Microprotein that
250   Supports Mitochondrial Supercomplexes and Report Mitoregulin : A lncRNA-Encoded Microprotein
251   that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. *CellReports* **23**:3710–
252   3720.e8. doi:10.1016/j.celrep.2018.06.002

253 Steinhauer J. 2015. Separating from the pack: Molecular mechanisms of drosophila spermatid
254   individualization. *Spermatogenesis* **5**:1–11. doi:10.1080/21565562.2015.1041345

255 Strimmer K. 2008. fdrtool : a versatile R package for estimating local and tail area-based false discovery
256   rates. *Bioinformatics* **24**:1461–1462. doi:10.1093/bioinformatics/btn209

257 Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert B, Gillette M, Paulovich A, Pomeroy S, Golub
258   T, Lander E, Mesirov J. 2005. Gene set enrichment analysis: A knowledge-based approach for
259   interpreting genome-wide expression profiles. *Proc Natl Acad Sci* **102**:15545–15550.
260   doi:10.3969/j.issn.0372-2112.2018.08.016

261 Sun L, Goff L a, Trapnell C, Alexander R, Lo KA, Hacisuleyman E, Sauvageau M, Tazon-Vega B, Kelley
262   DR, Hendrickson DG, Yuan B, Kellis M, Lodish HF, Rinn JL. 2013. Long noncoding RNAs regulate
263   adipogenesis. *Proc Natl Acad Sci U S A* **110**:3387–92. doi:10.1073/pnas.1222643110

264 Sunwoo H, Wu JY, Lee JT. 2015. The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist
265   stoichiometry and suggests a hit-and-run mechanism in mouse cells. *Proc Natl Acad Sci* **112**:E4216–
266   E4225. doi:10.1073/pnas.1503690112

267 Tan YP, Li S, Jiang XJ, Loh W, Foo YK, Loh CB, Xu Q, Yuen WH, Jones M, Fu J, Venkatesh B, Yu WP.
268   2010. Regulation of protocadherin gene expression by multiple neuron-restrictive silencer elements
269   scattered in the gene cluster. *Nucleic Acids Res* **38**:4985–4997. doi:10.1093/nar/gkq246

270 The Mouse ENCODE Consortium. 2014. A comparative encyclopedia of DNA elements in the mouse
271   genome. *Nature* **515**:355–364. doi:10.1038/nature13992

272 Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA,
273   Blencowe BJ, Prasanth SG. 2010. Article The Nuclear-Retained Noncoding RNA MALAT1 Regulates
274   Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Mol Cell* **39**:925–938.
275   doi:10.1016/j.molcel.2010.08.011

276 Turro E, Astle WJ, Tavaré S. 2014. Flexible analysis of RNA-seq data using mixed effects models.
277   *Bioinformatics* **30**:180–188. doi:10.1093/bioinformatics/btt624
278 Turro E, Su SY, Gonçalves Â, Coin LJM, Richardson S, Lewin A. 2011. Haplotype and isoform specific
279   expression estimation using multi-mapping RNA-seq reads. *Genome Biol* **12**:R13. doi:10.1186/gb-
280   2011-12-2-r13
281 van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers P, de Bruijn E, Hao W, MacInnes AW,
282   Cuppen E, Simonis M. 2014. Extensive localization of long noncoding RNAs to the cytosol and
283   mono- and polyribosomal complexes. *Genome Biol* **15**:1–12. doi:10.1186/gb-2014-15-1-r6
284 Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA,
285   Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY. 2011. A long noncoding RNA
286   maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**:120–4.
287   doi:10.1038/nature09819
288 Wichman L, Somasundaram S, Breindel C, Valerio DM, McCarrey JR, Hodges CA, Khalil AM. 2017.
289   Dynamic expression of long noncoding RNAs reveals their potential roles in spermatogenesis and
290   fertility. *Biol Reprod* **97**:313–323. doi:10.1093/biolre/iox084
291 Wu D, Smyth GK. 2012. Camera: A competitive gene set test accounting for inter-gene correlation.
292   *Nucleic Acids Res* **40**:1–12. doi:10.1093/nar/gks461
293 Xu J. 2005. Preparation, Culture, and Immortalization of Mouse Embryonic FibroblastsCurrent Protocols
294   in Molecular Biology. doi:10.1002/0471142727.mb2801s70
295 Xu Y, Wang J, Qiu M, Xu L, Li M, Jiang F, Yin R, Xu L. 2014. Upregulation of the long noncoding RNA
296   TUG1 promotes proliferation and migration of esophageal squamous cell carcinoma. *Tumor Biol*.
297   doi:10.1007/s13277-014-2763-6
298 Young TL, Matsuda T, Cepko CL. 2005. The noncoding RNA Taurine Upregulated Gene 1 is required for
299   differentiation of the murine retina. *Curr Biol* **15**:501–512. doi:10.1016/j.cub.2005.02.027
300 Zhang E, Yin D, Sun M, Kong R, Liu X, You L, Han L, Xia R, Wang K, Yang J, De W, Shu Y, Wang Z.
301   2014. P53-regulated long non-coding RNA TUG1 affects cell proliferation in human non-small cell
302   lung cancer, partly through epigenetically regulating HOXB7 expression. *Cell Death Dis* **5**:e1243.
303   doi:10.1038/cddis.2014.201
304 Zheng H, Stratton CJ, Morozumi K, Jin J, Yanagimachi R, Yan W. 2007. Lack of Spem1 causes aberrant
305   cytoplasm removal, sperm deformation, and male infertility. *Proc Natl Acad Sci* **104**:6852–6857.
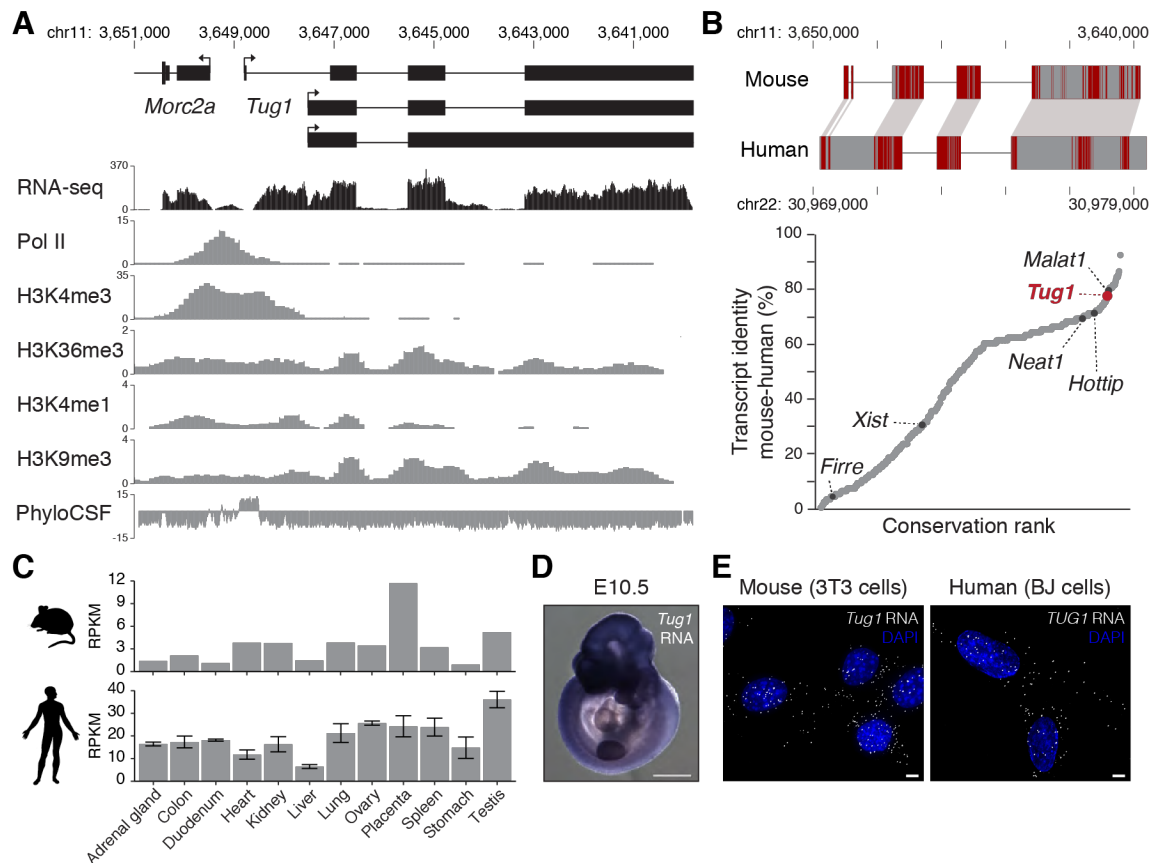306   doi:10.1073/pnas.0701669104
307

FIGURES

Figure 1. The *Tug1* lncRNA locus is highly conserved and ubiquitously expressed. (A) *Tug1* mouse genomic locus (shown inverted). UCSC Genome Browser tracks for RNA-sequencing (RNA-seq), RNA polymerase II (Pol II), histone 3 lysine 4-trimethylation (H3K4me3), H3K36me3, and H3K4me1 occupancy in testis and H3K9me3 occupancy in brain are depicted. PhyloCSF scores across the locus are shown. Chromosomal coordinates of the mouse *Tug1* gene are indicated (mm9). **(B)** Upper panel: Schematic of the nucleotide conservation alignment for mouse and human *Tug1*. Red lines indicate conserved nucleotides. Chromosomal coordinates of the *Tug1* gene for both species are indicated. Lower panel: Distribution of sequence identity for orthologous divergent and intergenic lncRNAs between mouse and human. X-axis shows increasing conservation rank. *Tug1* and other well characterized lncRNAs are highlighted. **(C)** RNA-seq expression levels of *Tug1* in a panel of mouse and human tissues. **(D)** RNA *in situ* hybridization of *Tug1* RNA in a mouse embryo at embryonic day 10.5 (E10.5). **(E)** Maximum intensity projections of *Tug1* single molecule RNA FISH (gray) on murine 3T3 and human BJ fibroblasts. Nucleus is stained with DAPI (blue). Scale bar is 5 μm.
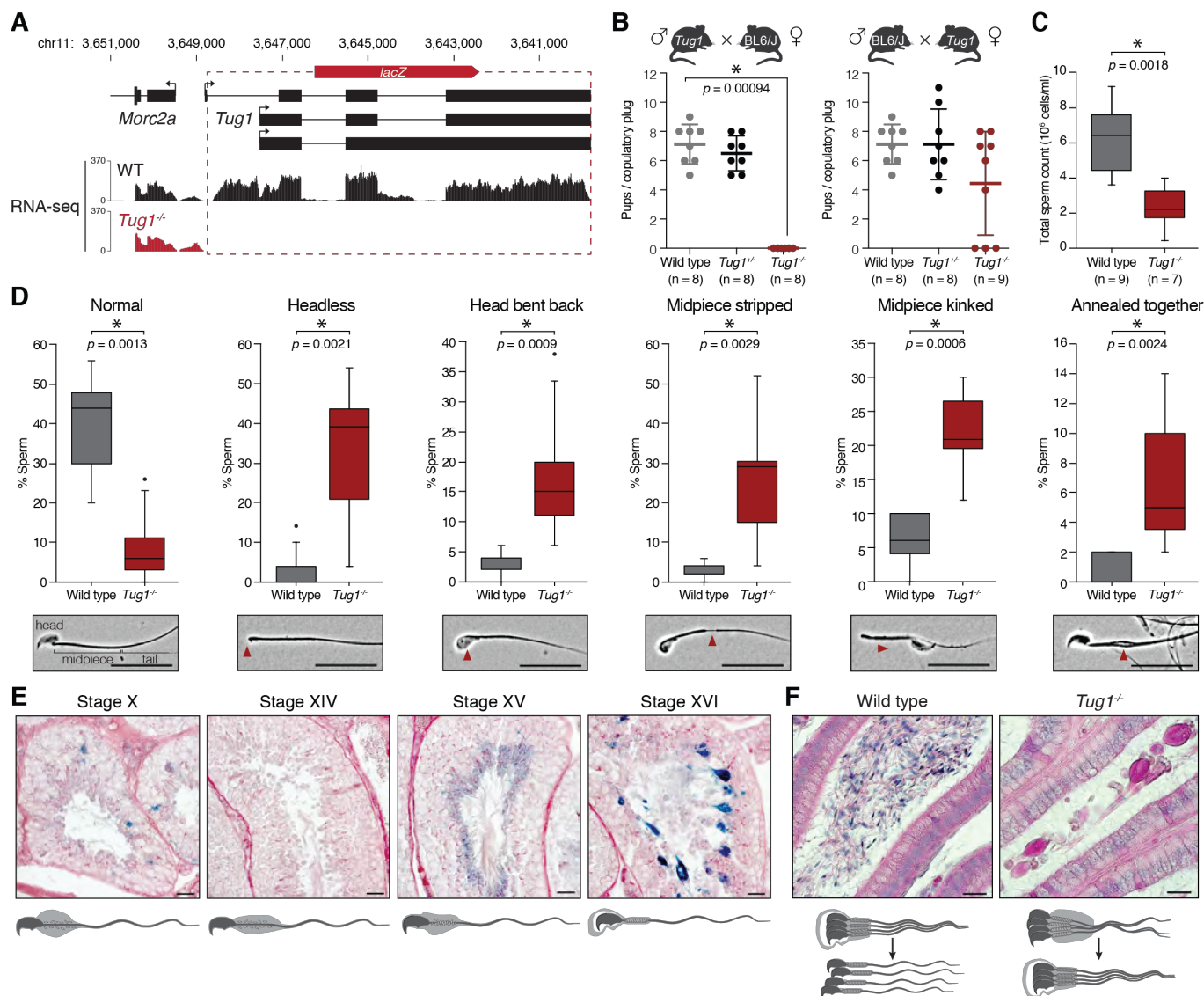
Figure 2. Deletion of the *Tug1* locus leads to sperm defects and male infertility. (A) Deletion strategy of the *Tug1* locus (shown inverted). The *Tug1* gene-body was replaced by a *lacZ* reporter cassette, leaving the promoter and first exon intact. The dashed lines indicate the deleted region in the *Tug1* knockout. RNA-sequencing (RNA-seq) tracks for wild type (WT) and *Tug1⁻/⁻* testis are depicted. **(B)** Scatter dot plot (showing the mean and standard deviation) of the number of pups at birth per copulatory plug for matings between wild type, *Tug1⁺/⁻* or *Tug1⁻/⁻* males and wild type C57BL/6J females (left panel) and wild type C57BL/6J males and wild type, *Tug1⁺/⁻* or *Tug1⁻/⁻* females (right panel). Each dot represents a litter from a different mouse. Significant (*) *p*-value (Wilcoxon rank sum test with Bonferroni correction) and number of mice for each genotype tested are indicated. **(C)** Box plot of total sperm count for wild type and *Tug1⁻/⁻* males. Significant (*) *p*-value (Wilcoxon rank sum test) is indicated. **(D)** Box plots of the percentage of normal sperm and sperm with the five most common morphological abnormalities for wild type (n = 9) and *Tug1⁻/⁻* (n = 8) males. Representative images of normal and morphologically aberrant sperm are located below each corresponding plot. Red arrows indicate the location of the defect. Scale bars are 20 μm. Significant (*) *p*-values (Wilcoxon rank sum test) are indicated. **(E)** Representative spermatocyte diagrams and micrographs of *Tug1⁺/⁻* seminiferous tubule sections stained with periodic acid-Schiff's reagent and X-gal showing expression of the *lacZ* reporter under the control of the endogenous *Tug1* promoter at the indicated stages of spermatogenesis. Scale bars are 20 μm. **(F)** Representative spermatid diagrams and micrographs of wild type and *Tug1⁻/⁻* epididymis tubule sections stained with hematoxylin and eosin. Scale bars are 20 μm.
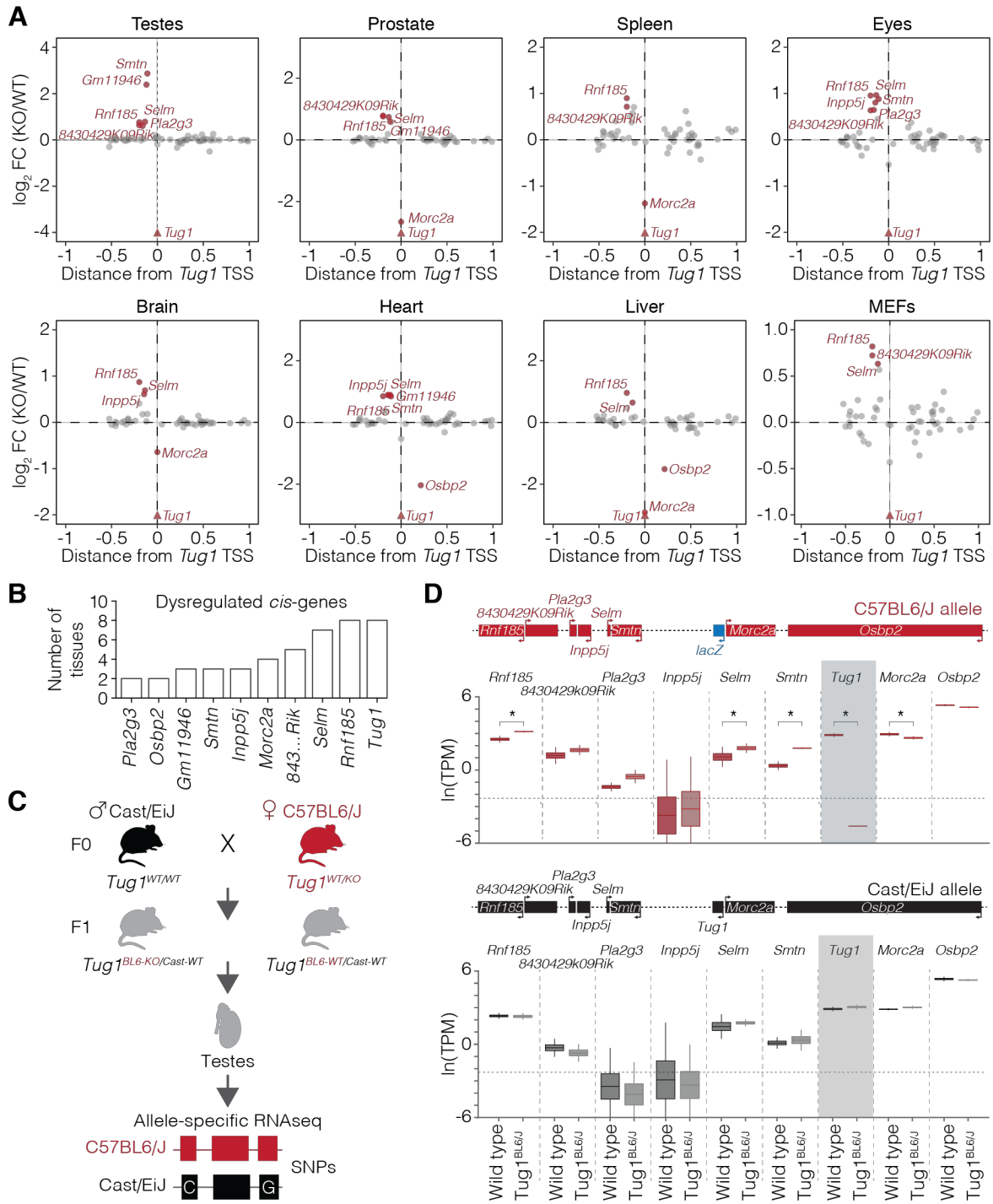
32

Figure 3. The *Tug1* locus harbors a *cis*-repressive DNA regulatory element. (A) Differential expression of genes in the local region (±1 Mb) of *Tug1* for each indicated mouse tissue, depicted as fold change (FC) between *Tug1^{-/-}* (KO) and wild type (WT). Significantly differentially expressed genes are marked and labeled in red. **(B)** Plot of the number of tissues that genes downstream of *Tug1* TSS are found significantly dysregulated. **(C)** Strategy for allele-specific RNA-sequencing (RNA-seq). *Tug1^{+/-}* C57BL/6J females were crossed with wild type Cast/EiJ males and testes from the F1 hybrid progeny were harvested for RNA-seq. Single nucleotide polymorphisms (SNPs) allowed for the differentiation between the C57BL/6J and the Cast/EiJ allele. **(D)** Allele-specific expression of local genes surrounding *Tug1* in testes from F1 hybrid C57BL/6J::Cast/EiJ wild type (*Tug1^{BL6-WT/Cast-WT}*) and heterozygous *Tug1* knockout (*Tug1^{BL6-KO/Cast-WT}*) mice containing a deletion on the C57BL/6J allele. Upper panel, expression levels of neighboring genes from the C57BL/6J allele; lower panel, expression levels of neighboring genes from the Cast/EiJ allele. Boxes are centered at the mean, extend one standard deviation, and the bottom and top notches are the minimum and maximum samples. The genomic locus encompassing the local genes around *Tug1* is depicted. Asterisks indicate significant Bayesian posterior probability (PP>0.95) differential expression between hybrid wild type and *Tug1^{BL6-KO}* testes. Horizontal dotted line indicates expression levels below 0.1 TPM.
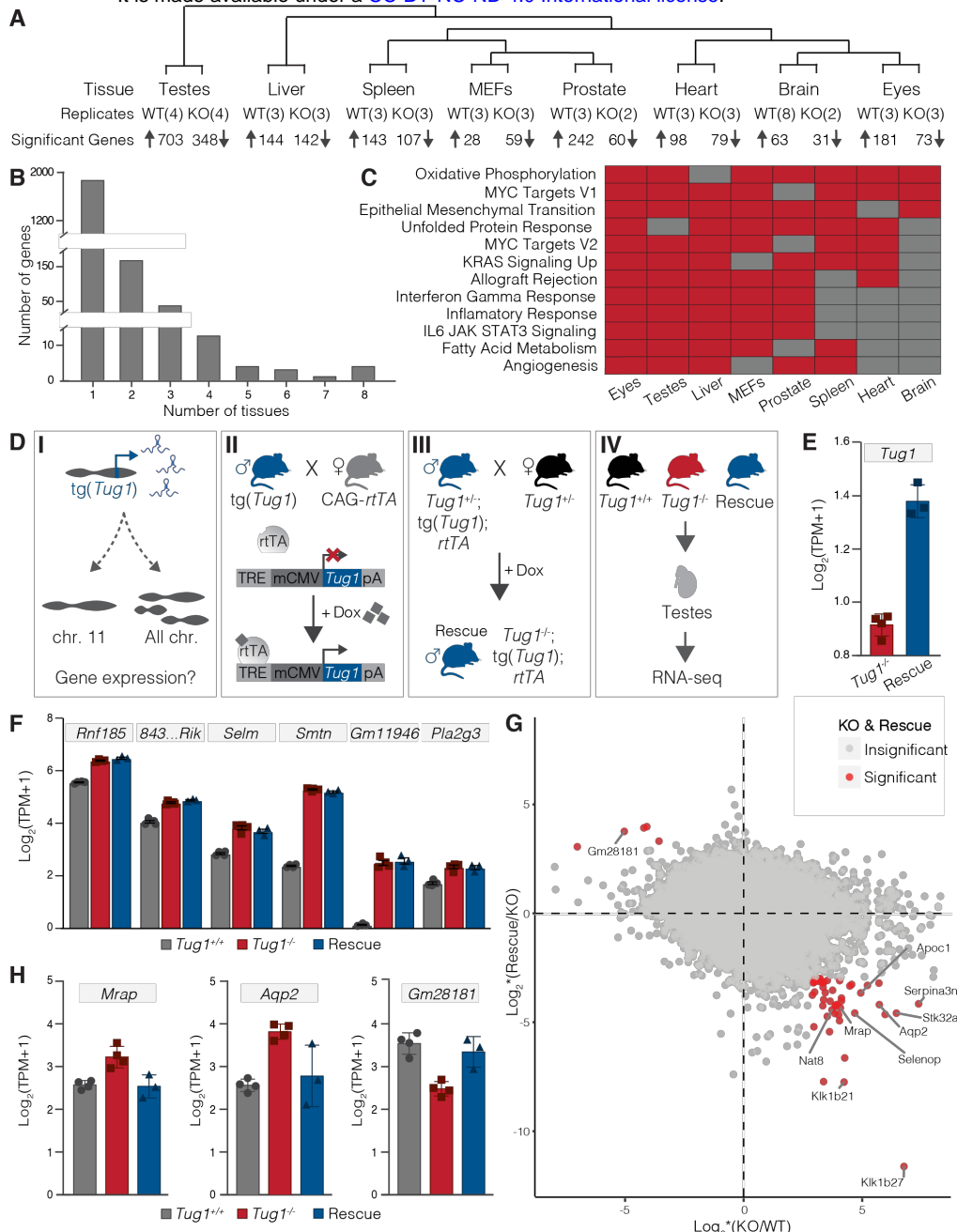
33

Figure 4. *Tug1* **lncRNA regulates gene expression in** *trans.* **(A)** Adult tissue types and mouse embryonic fibroblasts (MEFs) used for RNA sequencing of wild type (WT) and *Tug1-/-* (KO) mice. For each tissue, the number of biological replicates per genotype and the number of upregulated and downregulated genes (FDR<0.05) is shown from KO to WT comparisons. **(B)** The number of perturbed genes (y-axis) in KO animals according to the number of tissues in which the gene was found to be dysregulated (x-axis). **(C)** Gene Set Enrichment Analysis (GSEA) of differentially expressed genes found in wild type vs. *Tug1-/-* murine tissues and MEFs. Red shading indicates tissue in which gene set is perturbed, grey shading indicates tissue in which gene set is not different between WT and KO. **(D)** Schematic showing the experimental design to identify genes reciprocally regulated by *Tug1* RNA. **(DI)** Testing the impact of the *Tug1* transgene expression on gene expression *in vivo* **(DII)** Schematic of the *Tug1* transgene (tg(*Tug1*)) and systemic induction by mating to CAG-rtTA3 mice in the presence of doxycycline (dox). **(DIII)** Schematic of matings to generate *Tug1*rescue mice (*Tug1-/-*; tg(*Tug1*); *rtTA*), enabling dox-inducible *Tug1* expression in a *Tug1*-knockout background. **(DIV)** Collection of testes from WT (*Tug1+/+*) (n = 4), KO (*Tug1-/-*) (n = 4), and *Tug1*rescue (n = 3) mice for RNA sequencing. **(E)** *Tug1* gene expression level (log2TPM+1) in testes of KO (red) and doxycycline-induced *Tug1*rescue (blue) mice. Error bars represent standard error of the mean. **(F)** Expression levels (log2TPM+1) for *Tug1* neighboring genes in WT (grey), KO (red), and *Tug1*rescue (blue) mice. **(G)** Genome-wide profile of reciprocally regulated genes from *Tug1* RNA induction. The fold change score of KO-WT is plotted on the x-axis and the fold change score of Rescue-KO is plotted on the y-axis. The fold change score (*) is the fold change divided by standard deviation. Genes significantly differentially regulated in both comparisons of KO-WT and Rescue-KO are labeled in red, otherwise labeled in grey. Examples of reciprocally regulated genes are labeled with the gene name. **(H)** Examples of differentially expressed genes in testes showing significant reciprocal regulation in WT (grey), KO (red), and *Tug1*rescue (blue) mice.
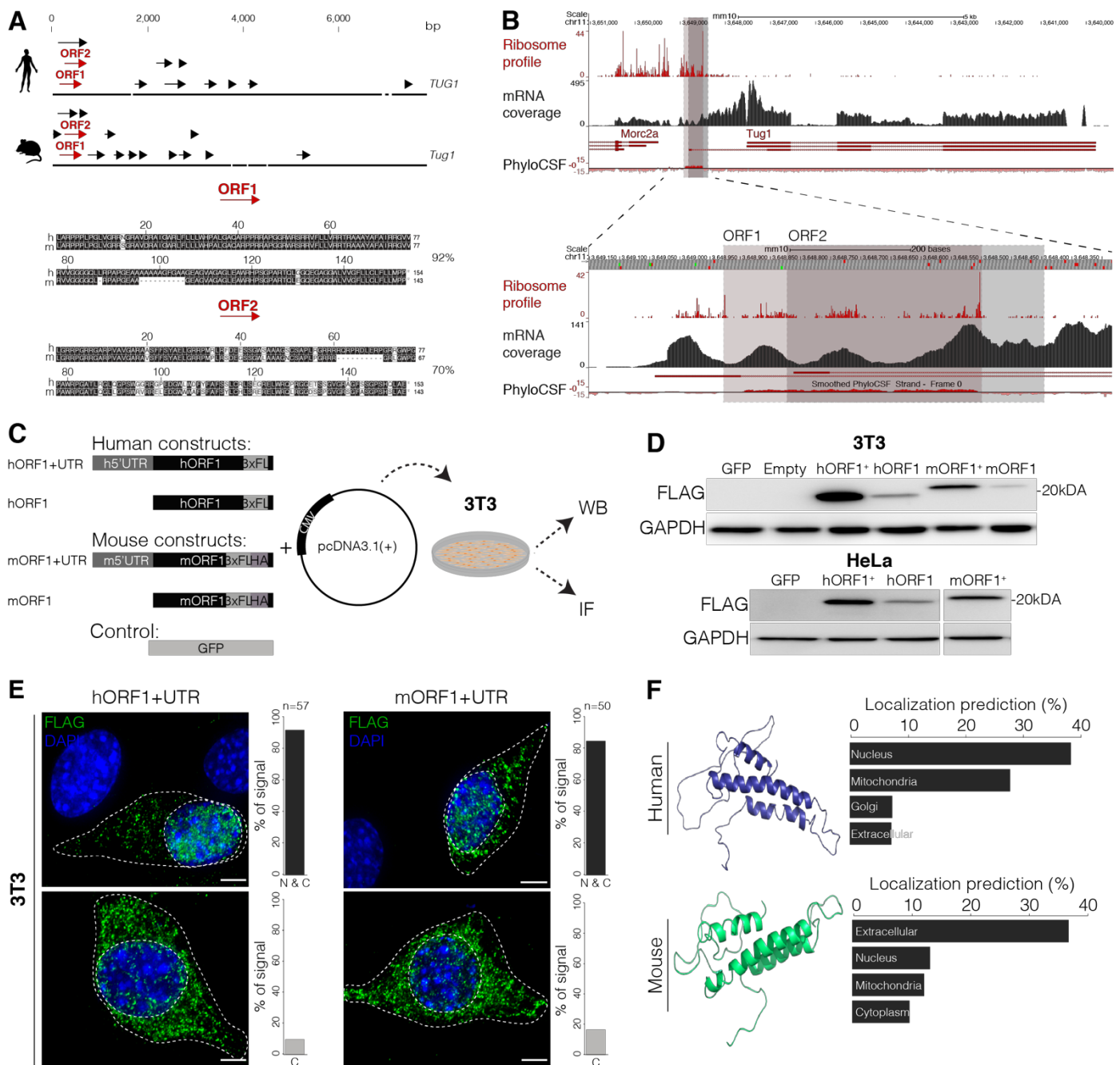
34

Figure 5. **The 5' region of *Tug1* encodes a conserved peptide. (A)** Open reading frame (ORF) search in human and mouse *Tug1* reveals multiple ORFs (arrows). ORF1 and ORF2 (depicted as red arrows) indicate two ORFs with greater than 70% amino acid conservation between human and mouse (92% and 70%, respectively). **(B)** *Tug1* mouse genomic locus (mm10) is shown. Ribosome occupancy (Ribosome profile), RNA-seq (mRNA coverage), and evolutionary protein-coding potential (PhyloCSF) across the *Tug1* locus in mouse embryonic fibroblasts (MEFs) is depicted. ORF1 and ORF2 are outlined with red and gray boxes, respectively (top). Tracks surrounding both ORFs are zoomed in for clarity (bottom). **(C)** Scheme of human and mouse ORF1 construct design. A 3xFLAG epitope tag was inserted prior the stop codon of ORF1. Mouse constructs were dual-tagged with both 3xFLAG and HA tags. Expression constructs were designed with (hORF1+UTR, mORF1+UTR) and without (hORF1, mORF1) the 5' UTR. Constructs and GFP as control were inserted into pcDNA3.1(+). 48 hours post-transfection, 3T3 and HeLa cells were harvested for western blot (WB) (shown in D) or analyzed by immunofluorescence (IF) (shown in E). **(D)** Western blot targeting the 3xFlag (FLAG) in 3T3 and HeLa cells expressing human and mouse constructs, respectively. GAPDH was used as a loading control. **(E)** Maximum intensity projections of 3T3 cells expressing human and mouse constructs. Immunostaining against the Flag tag (green) and DAPI (blue) are shown. Bar plot shows localization analysis of human and mouse TUG1-BOAT. N and C indicates nuclear and cytoplasmic localization, C indicates only cytoplasmic. Scale bar is 5 μm. **(F)** Human and mouse TUG1-BOAT structure (RaptorX) and localization (DeepLoc) prediction.
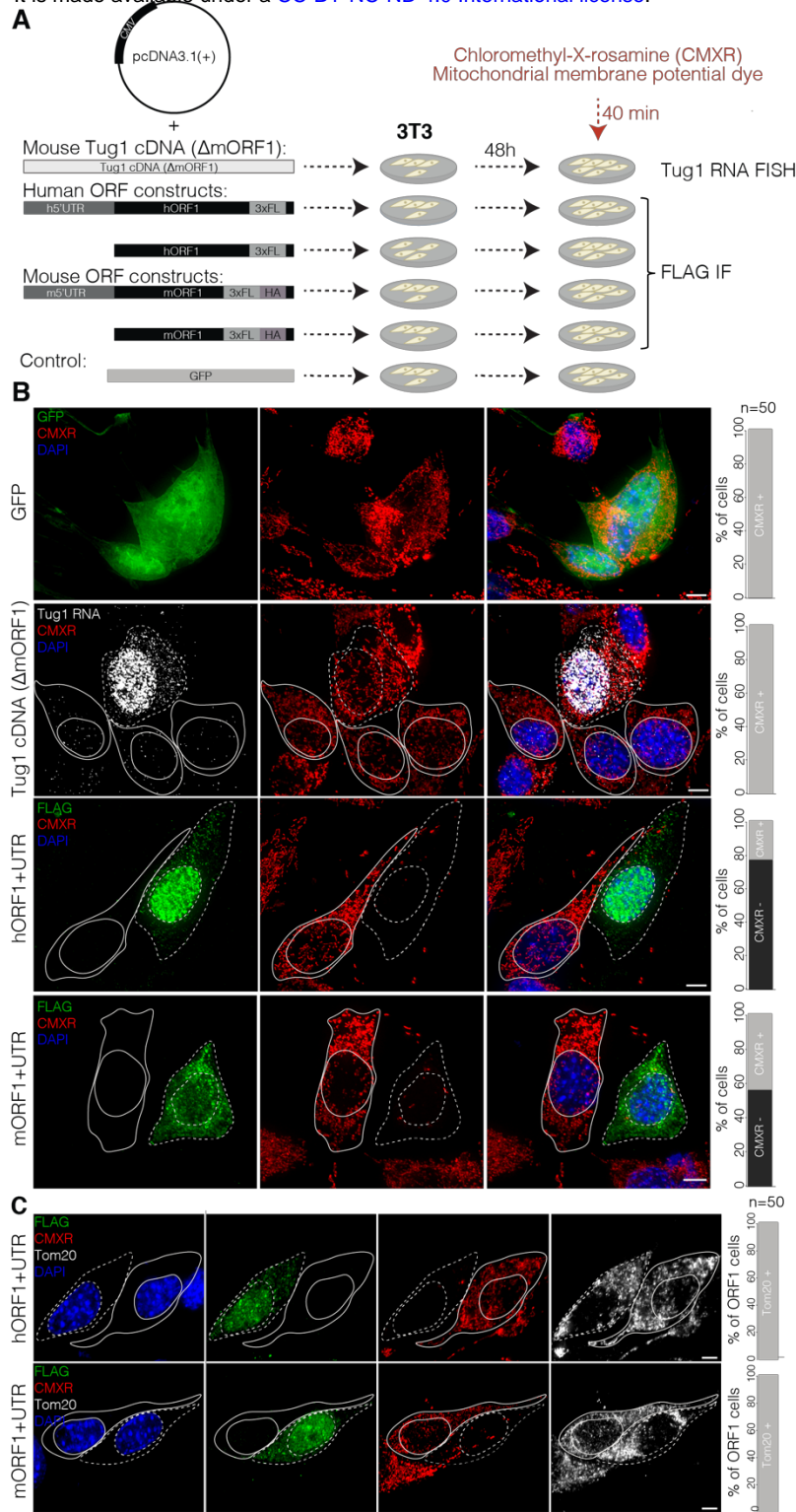
Figure 6. Overexpression of TUG1-BOAT compromises mitochondrial membrane potential. (A) Construct and transfection scheme. Human and mouse ORF1, and mouse *Tug1* cDNA lacking the ORF1 region (*Tug1* cDNA ΔmORF1) were inserted into pcDNA3.1(+). Chloromethyl-X-rosamine (CMXR) was added to visualize mitochondria 48 hours post transfection. After staining, cells were fixed and processed for anti-FLAG immunofluorescence (IF) or *Tug1* RNA FISH. **(B)** Maximum intensity projections of z-stacks acquired 48 hours post-transfection of 3T3 cells with indicated plasmids and staining with CMXR. *Tug1* RNA overexpression was monitored by *Tug1* single molecule RNA FISH (gray), TUG1-BOAT by immunostaining against the FLAG tag (green). GFP was used as a control. CMXR is shown in red, DAPI in blue. On the right, quantification of cells positive for GFP, *Tug1* RNA, or TUG1-BOAT and mitochondria by CMXR (n = 50). Scale bar is 5 μm. **(C)** Maximum intensity projections of z-stacks acquired 48 hours post-transfection of 3T3 cells with the indicated plasmids, stained with CMXR (red) and immunostained against mitochondrial membrane translocase TOM20 (gray). On the right, quantification of cells over-expressing TUG1-BOAT and lacking CMXR staining showing intact mitochondrial membrane assessed by TOM20. Nuclei were stained with DAPI (blue). Scale bar is 5 μm

36

TABLES

Gene name	Location	Mean TPM			Significance		Biological process
		WT	KO	Rescue	WT-KO	KO-Rescue	
Pard3b	chr1	2.66	2.13	2.74	***	***	microtubule cytoskeleton organization
Nav1	chr1	1.40	2.33	1.45	**	**	microtubule bundle formation
Gm28181	chr1	10.81	4.63	9.39	***	**	unknown
Col5a1	chr2	7.16	8.97	7.85	*	*	cell adhesion
Hoxd10	chr2	2.17	4.43	2.23	*	*	regulation of transcription
Sulf2	chr2	12.20	16.74	13.21	***	**	metabolic process
Amy1	chr3	16.43	23.06	14.58	*	***	metabolic process
Dhcr24	chr4	30.92	43.38	34.72	***	*	lipid metabolic process
Tmem176a	chr6	17.89	25.23	16.32	**	***	lipid metabolic process
Nat8f5	chr6	2.30	4.78	2.13	**	**	system development
Nat8	chr6	9.57	16.53	8.23	**	***	glutathione metabolic process
Apoc1	chr7	35.49	56.71	37.27	***	**	lipid metabolic process
Vmn1r181	chr7	2.65	4.26	1.88	**	***	unknown
Klk1b27	chr7	12.21	31.01	2.65	***	***	proteolysis
Klk1b21	chr7	23.81	56.40	5.45	***	***	proteolysis
Klk1b24	chr7	19.20	34.33	5.95	*	***	proteolysis
Cd209f	chr8	1.57	2.64	1.23	*	***	cell adhesion
Gpt2	chr8	25.46	33.70	24.06	**	***	biosynthetic process
Tug1	chr11	40.08	0.89	1.60	***	***	-
Spns2	chr11	6.30	7.92	6.47	*	*	sphingolipid metabolic process
Serpina3n	chr12	4.10	13.85	7.18	***	***	inflammatory response
Ankrd9	chr12	43.02	50.58	44.74	*	*	post-translational protein modification
Sv2c	chr13	5.59	7.53	5.15	***	***	transmembrane transport
3110070M22Rik	chr13	110.87	93.69	120.25	**	*	unknown
Tmem267	chr13	87.89	77.92	92.25	***	***	unknown
Il17rb	chr14	4.20	6.08	3.65	***	***	regulation of cell growth
Stab1	chr14	3.99	5.50	4.04	**	*	cell adhesion
Selenop	chr15	88.40	130.97	85.90	***	***	selenium compound metabolic process
C7	chr15	94.22	128.27	91.48	**	***	immune response
Aqp2	chr15	4.93	13.17	6.48	***	***	water transport
AU021092	chr16	30.38	44.08	33.37	***	*	unknown
Nrros	chr16	3.00	3.93	2.99	*	*	superoxide metabolic process
Mrap	chr16	4.83	8.21	4.75	***	***	protein localization to plasma membrane
Cyp21a1	chr17	1.99	3.12	1.22	*	***	steroid metabolic process
Ston1	chr17	28.51	34.85	29.11	***	***	regulation of endocytosis
Stk32a	chr18	1.22	2.78	1.50	***	***	protein phosphorylation
mt-Rnr1	chrM	211.26	253.11	219.24	*	*	ribosome biogenesis

Table 1. Genes reciprocally regulated by *Tug1* lncRNA in the testes. List of genes with significant changes in expression between wild type (WT), *Tug1^-/-* (KO), and *Tug1^rescue^* testes. Chromosomal location of the genes, mean TPM for each condition, and the main biological processes associated with each gene are listed. Significance of the fold change between wild type verses *Tug1^-/-* (WT-KO) and *Tug1^-/-* verses *Tug1^rescue^* (KO-Rescue) is indicated by asterisks (*, p<0.05; **, p<0.01, ***, p<0.001).
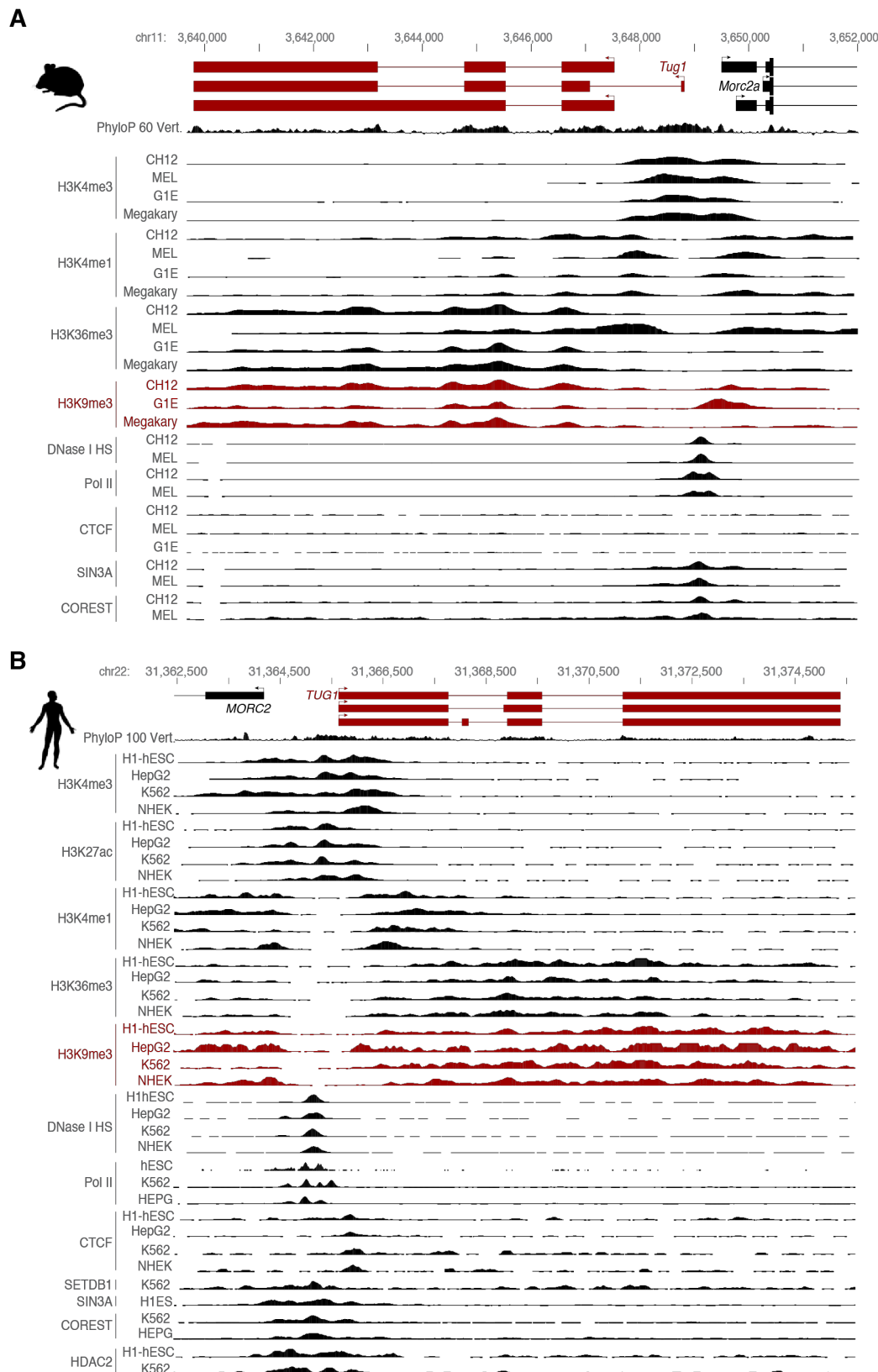
37

SUPPLEMENTARY FIGURES

Figure S1. Mouse and human *Tug1* locus and chromatin context in different cell types. (A) *Tug1* mouse and **(B)** human genomic loci. Evolutionary nucleotide conservation (PhyloP) of the locus are presented along with the chromatin context (DNase I hypersensitive regions, histone modifications) and protein binding ChIP-seq peaks (Pol2, CTCF, SIN3A, COREST, SETDB1, HDAC2) from ENCODE (UCSC Genome Browser, mm9) datasets in the indicated cell types.
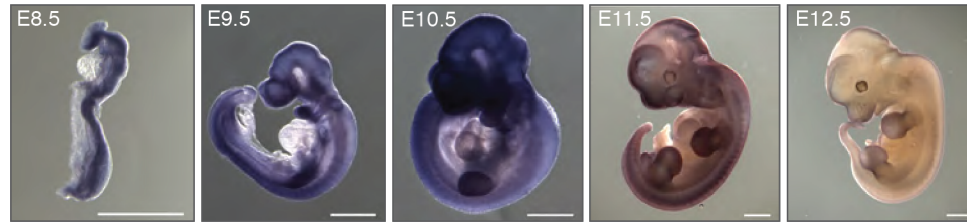
38

Figure S2. *In vivo* expression pattern of *Tug1* during murine embryogenesis. RNA *in situ* hybridization of *Tug1* RNA using a digoxigenin-labeled antisense RNA probe in mouse embryos at different developmental stages. Embryonic day (E)8.5, E9.5, E10.5, E11.5, and E12.5 are shown.
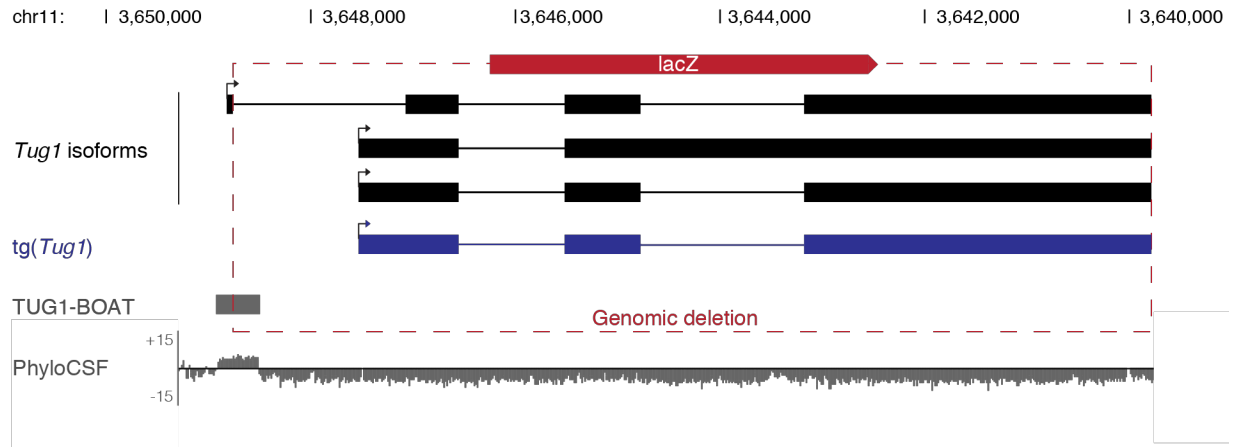
Figure S3. Overview of the *Tug1* locus in mouse. UCSC genome browser showing the murine*Tug1* locus. The three predominate *Tug1* isoforms are depicted (black) and the *Tug1* transgene (tg(*Tug1*)) is shown (blue). For *Tug1* knockout, the longest annotated *Tug1* isoform was replaced by a *lacZ* reporter cassette, leaving the promotor and first exon intact. The deleted region is indicated by red dashed lines. The open reading frame (ORF) encoding the TUG1-BOAT peptide and PhyloCSF scores for the (-2) frame across the locus are depicted (grey). Chromosomal coordinates (mm10) are shown.
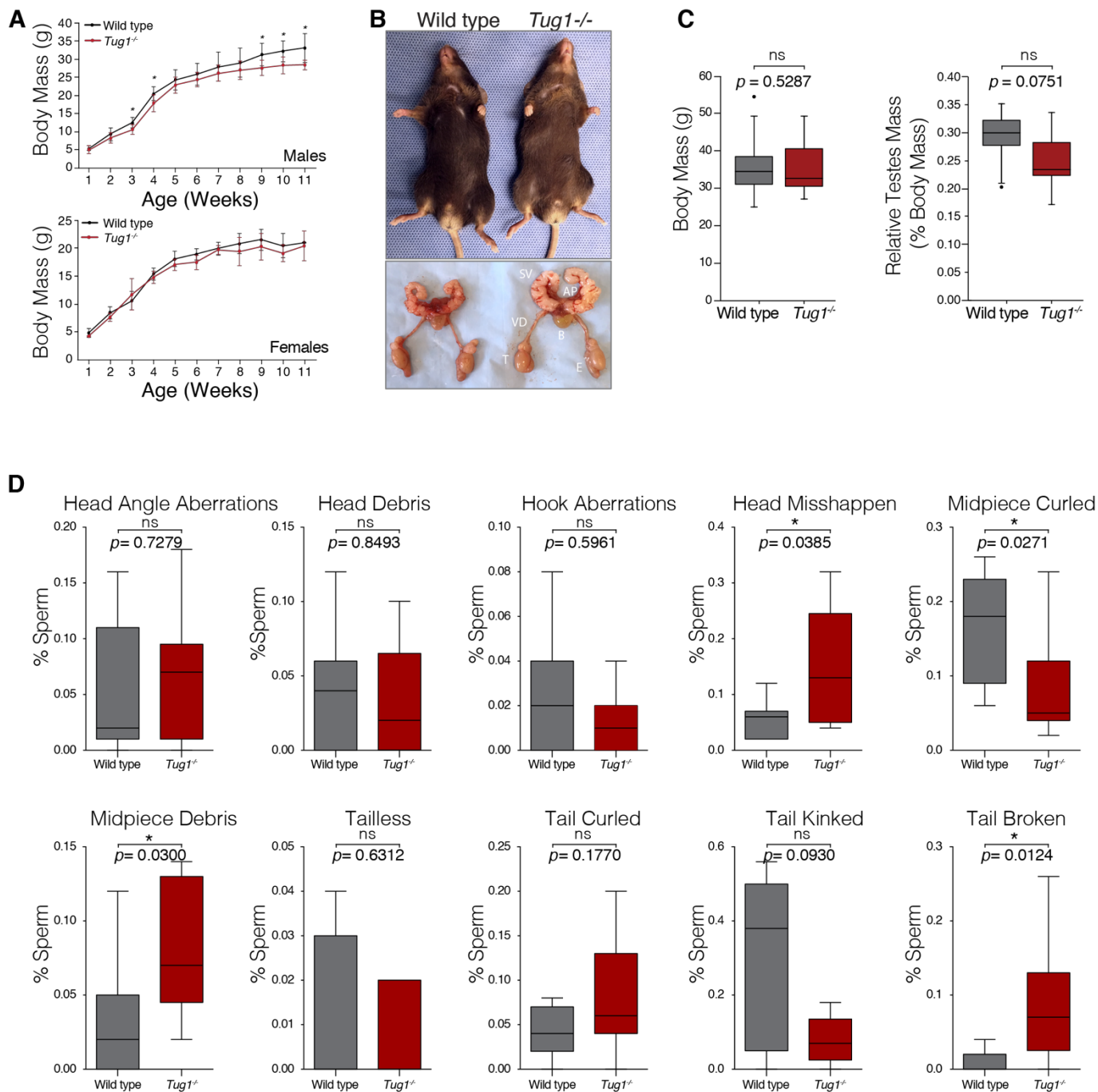
Figure S4. Morphology analysis of *Tug1-/-* mice and sperm (A) Body mass (g) measurements over 11 weeks of male and female *Tug1-/-* mice compared to wild type littermates. Males: *Tug1-/-* (n = 7); WT (n = 8). Females: *Tug1-/-* (n = 3), WT (n = 7)**.** Significant p values at specific time points are indicated (*). **(B)** Representative images from adult male mice (12 weeks old) show normal physiological appearance of external genitalia and reproductive tracks in *Tug1-/-* compared to WT. Seminal vesicles (SV), vas deferens (VD), bladder (B), testicle (T), epididymis (E), anterior prostate (AP). **(C)** Box plots of body mass (g) (left panel), relative testis mass (testis mass / body mass; middle panel) and total sperm count for wild type (n = 9) and *Tug1-/-* males. **(D)** Box plots of the percentage of different sperm morphological abnormalities for wild type (n = 9) and *Tug1-/-* (n = 8) males. Significant (*) p value (Wilcoxon rank sum test) is indicated**.**
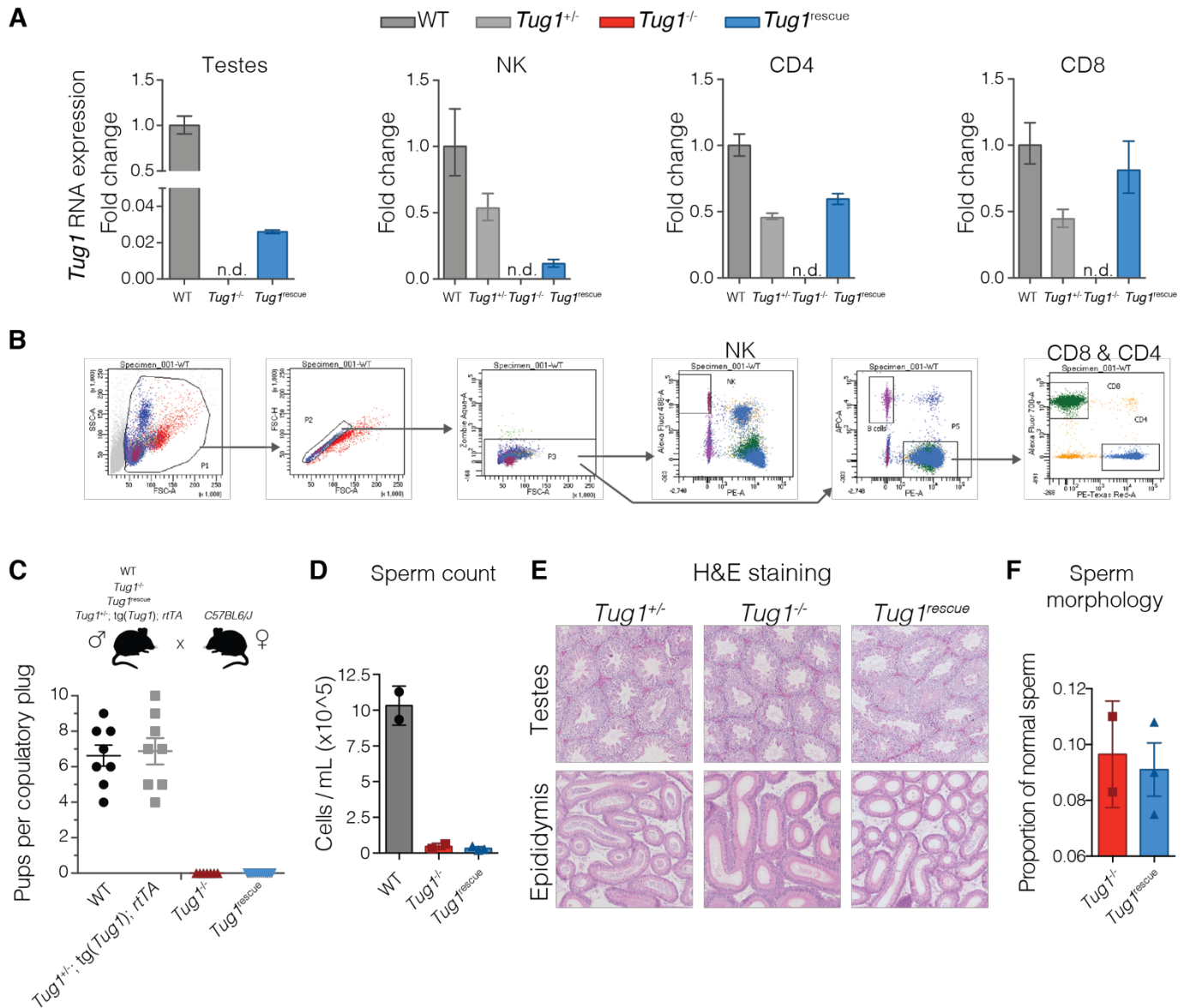
Figure S5. *Tug1* transgene expression and fertility assessment. (A) qRT-PCR for *Tug1* RNA expression in testes and sorted peripheral blood populations: WT (*n* = 1), *Tug1*[+/-] (*n* = 1), *Tug1*[-/-] (*n* = 1), and *Tug1*[rescue] (*n* = 1) and sorted peripheral blood populations. Error bars indicate the relative quantification minimum and maximum confidence interval at 98%. Not detected (n.d.). **(B)** Representative flow cytometry gating strategy for NK, CD4, and CD8 cells in peripheral blood from WT, *Tug1*[+/-], *Tug1*[-/-], and *Tug1*[rescue] mice (gating from WT peripheral blood shown). **(C)** Scatter dot plot (mean with standard error of the mean shown) of the number of pups at birth per copulatory plug for matings using male wild type, *Tug1*[+/-]; tg(*Tug1*); rtTA, *Tug1*[-/-], or *Tug1*[rescue] (on dox diet) with wild type C57BL/6J females. Each dot represents a litter from a different mouse. **(D)** Sperm count from control (WT and *Tug1*[+/-], n = 2), *Tug1*[-/-] (n = 2), and *Tug1*[resuce] (n = 3) mice. Each dot represents a different mouse and the error bars indicate the standard error of the mean. **(E)** Hematoxylin and eosin staining in *Tug1*[+/-], *Tug1*[-/-], and *Tug1*[rescue] testes and epididymis. **(F)** Morphological analysis of sperm from *Tug1*[-/-] (n = 2), and *Tug1*[rescue] (n = 3) mice.
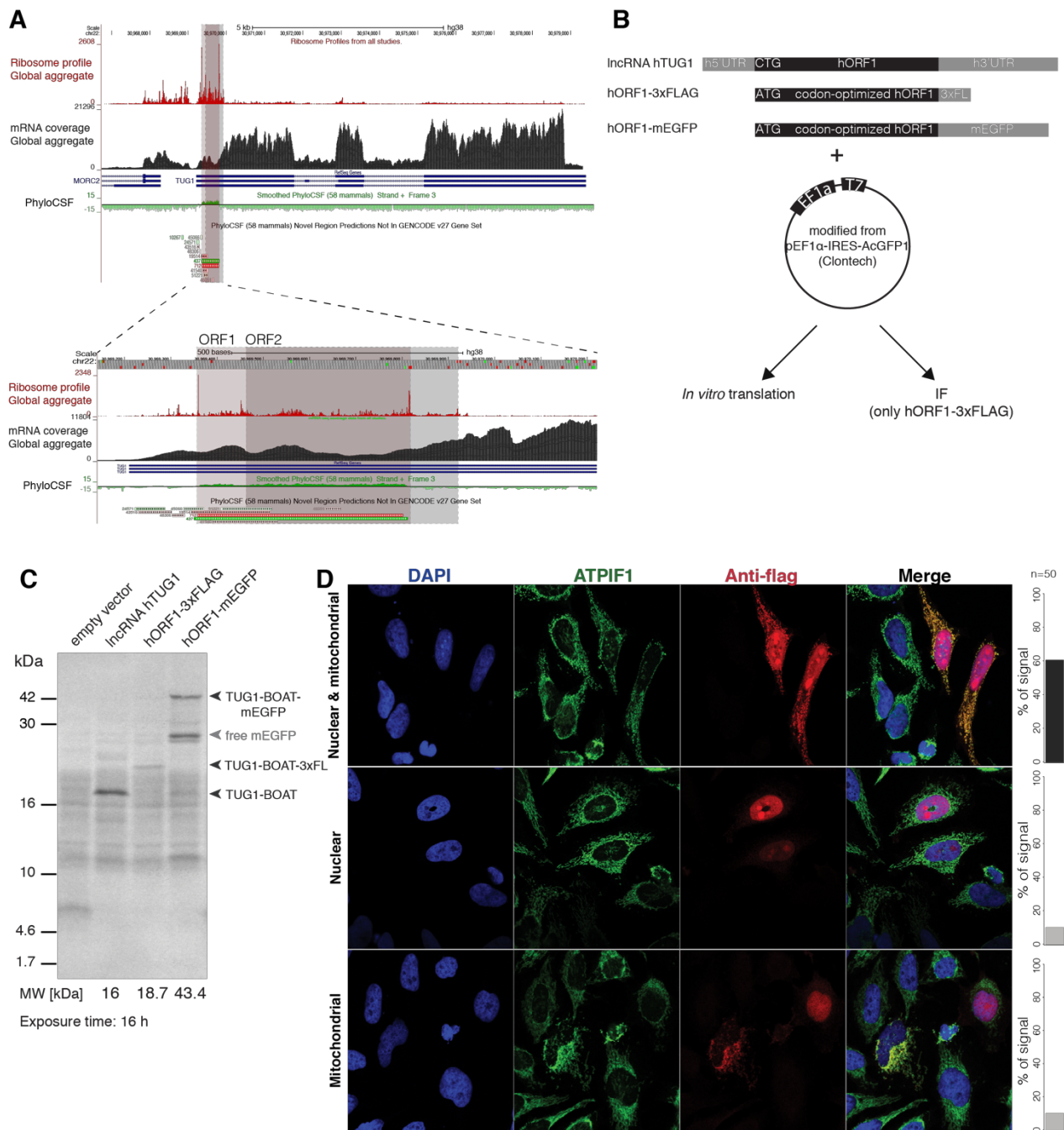
Figure S6. The 5' region of human *TUG1* encodes a conserved peptide. (A) GWIPS-viz tracks for human *TUG1* genomic locus (hg38) is shown. Global aggregate of ribosome occupancy (ribosome profile), RNA-seq (mRNA coverage), and evolutionary protein-coding potential (PhyloCSF) across the *TUG1* locus is shown. ORF1 and ORF2 are outlined with red and gray boxes, respectively. Tracks surrounding both ORFs are zoomed in for clarity (bottom). **(B)** Scheme of additional human ORF1 construct design. hORF1 was left unlabeled with its endogenous non-canonical start codon (CUG) and placed between its native 5' UTR and 321 nucleotides of its 3' UTR (lncRNA h*TUG1*). hORF1 was codon optimized to contain the canonical AUG start codon and labeled with either a 3xFLAG epitope tag (hORF1-3xFLAG) or mEGFP (hORF1-mEGFP), inserted prior to the stop codon. Constructs were inserted into a modified form of pEF1α-IRES-AcGFP1 and assessed for *in vitro* translation (shown in C). hORF1-3xFLAG was additionally analyzed by immunofluorescence (IF) (shown in D). **(C)** The synthesis of peptides from all three constructs and an empty vector control was assessed using a wheat germ extract *in vitro* translation assay. Newly synthesized peptides are labeled with arrows and correspond to their respective predicted molecular weights (16 kDa for hTUG1-BOAT, 18.7 kDa for TUG1-BOAT-3xFLAG, and 43.3 kDa for TUG1-BOAT-mEGFP). **(D)** Localization of codon-optimized 3xFLAG tagged ORF1 (hORF1-3xFLAG) was assessed by immunostaining against the 3xFLAG (red) in HeLa cells. Nuclear localization was monitored by DAPI (blue) and mitochondrial localization was monitored by the organelle marker ATPIF1 (green).

43

SUPPLEMENTARY TABLES

Table S1. *Tug1$^{-/-}$* and wild type sperm morphological defects.

Table S2. Prostate, spleen, eyes, brain, heart, liver, and MEF RNA-seq.

Table S3. Allele-specific RNA-seq in testes.

Table S4. Testes RNA-seq and *Tug1^{rescue}* RNA-seq in testes.

All supplementary tables are available at the Gene Expression Omnibus (GSEA124745 and GSE88819).