

RESEARCH

Open Access

The *Tug1* lncRNA locus is essential for male fertility



Jordan P. Lewandowski^{1†}, Gabrijela Dumbović^{2†}, Audrey R. Watson^{2,3†}, Taeyoung Hwang^{2†}, Emily Jacobs-Palmer⁴, Nydia Chang¹, Christian Much², Kyle M. Turner⁵, Christopher Kirby⁵, Nimrod D. Rubinstein⁵, Abigail F. Groff^{1,6}, Steve C. Liapis¹, Chiara Gerhardinger¹, Assaf Bester^{7,8}, Pier Paolo Pandolfi^{7,8}, John G. Clohessy^{7,8}, Hopi E. Hoekstra^{4,5,9*}, Martin Sauvageau^{10,11*} and John L. Rinn^{1,2,3*}

* Correspondence: hoekstra@oeb.harvard.edu; martin.sauvageau@ircm.qc.ca; john.rinn@colorado.edu

[†]Jordan P. Lewandowski, Gabrijela Dumbović, Audrey R. Watson and Taeyoung Hwang contributed equally to this work.

⁴Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, 4109 BioLabs, Cambridge, MA 02138, USA

¹⁰Montreal Clinical Research Institute, Montreal, QC H2W 1R7, Canada

¹Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA

Full list of author information is available at the end of the article

Abstract

Background: Several long noncoding RNAs (lncRNAs) have been shown to function as components of molecular machines that play fundamental roles in biology. While the number of annotated lncRNAs in mammalian genomes has greatly expanded, studying lncRNA function has been a challenge due to their diverse biological roles and because lncRNA loci can contain multiple molecular modes that may exert function.

Results: We previously generated and characterized a cohort of 20 lncRNA loci knockout mice. Here, we extend this 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 underlying defects including a low number of sperm and abnormal sperm morphology. Because lncRNA loci can contain multiple modes of action, we wanted to determine which, if any, potential elements contained in the *Tug1* genomic region have any activity. Using engineered mouse models and cell-based assays, we provide evidence that the *Tug1* locus harbors two distinct noncoding regulatory activities, as a *cis*-DNA repressor that regulates neighboring genes and as a lncRNA that can regulate genes by a *trans*-based function. We also show that *Tug1* contains an evolutionary conserved open reading frame that when overexpressed produces a stable protein which impacts mitochondrial membrane potential, suggesting a potential third coding function.

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

Keywords: *Tug1*, lncRNA, Fertility, DNA repressor, *Cis*-regulatory elements, RNA-seq, Allele-specific, Genetics, Genomics, Mouse



Background

Noncoding RNAs have been shown to play central roles in biology. Key cellular machines such as telomerase and the ribosome are comprised of proteins and non-coding RNAs and serve as classic examples of RNA-based functionalities [1, 2]. While thousands of lncRNA loci have been identified in mammalian genomes, characterizing their functions has been a challenge because of their diverse biological roles and due to complications in identifying their modes of action [3, 4]. Indeed, in addition to RNA function, lncRNA loci can harbor several potential regulatory modalities including DNA regulatory elements and the act of transcription [5–12]. Moreover, lncRNAs have been found to possess open reading frames (ORFs) [13, 14], and an increasing number have been found to encode proteins with biological roles [15–19]. With this in mind, it is likely that more regulatory DNA, RNA, and hidden protein activities will be uncovered at lncRNA loci. Thus, identifying the molecular activities present at lncRNA loci is important for further functional dissection of phenotypes attributed to lncRNA loci.

We previously reported the generation of 20 lncRNA loci knockout mouse strains, five of which displayed either viability, growth, or brain phenotypes [20, 21]. From the strains that did not initially display such phenotypes, we selected *Tug1* for further analysis because of its high conservation [22] as well as implications in diverse cellular functions and human malignancies [23].

Tug1 was first identified in a microarray screen for genes upregulated in response to taurine in a heterogenous culture of retinal cells [22]. In addition, there is some evidence that *Tug1* is transcriptionally regulated by p53 [24]. A number of studies have found evidence suggesting diverse RNA-based roles for *Tug1*, including but not limited to the development of photoreceptors [22] and in regulating gene expression in the nucleus by associating with the polycomb repressive complex 2 (PRC2) [25–27]. There is also some evidence for an RNA-based role for *Tug1* in cancer by acting as a tumor suppressor in human gliomas [28, 29] and by acting as a cytoplasmic miRNA sponge in prostate cancer cell lines [30]. Thus, the number of studies identifying broad roles for *Tug1* underscores the importance of this locus.

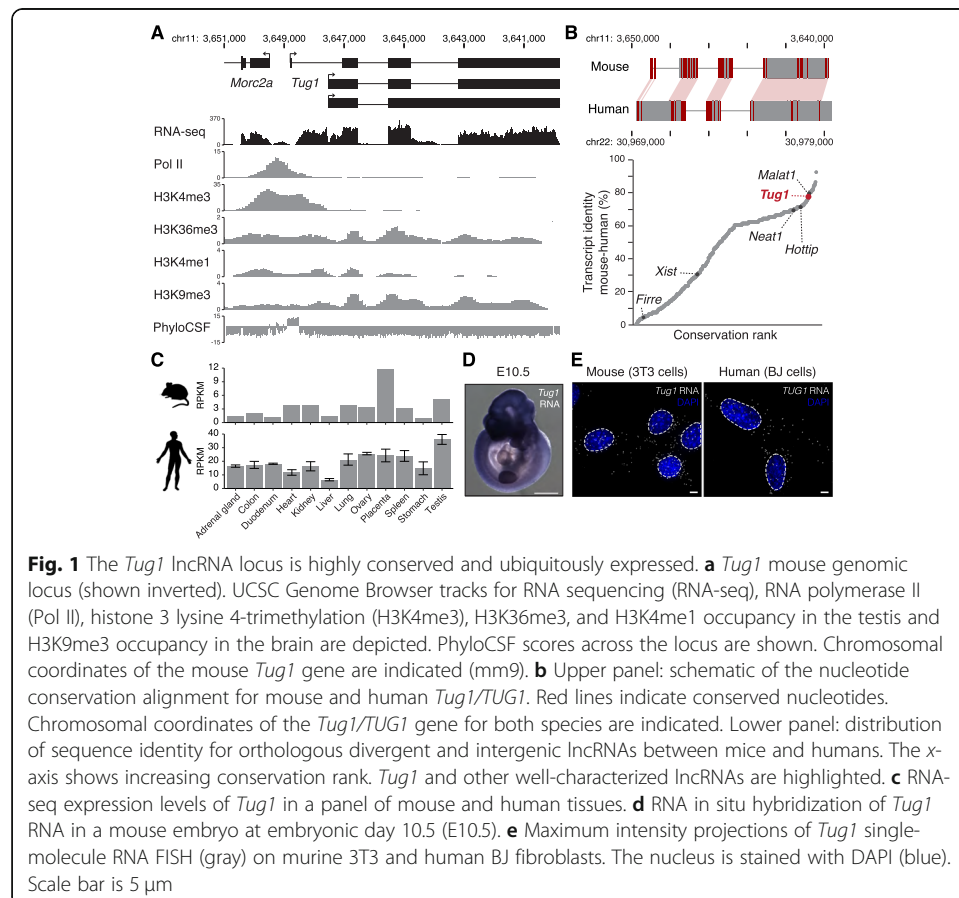
Here, we characterize the *Tug1* locus using multiple genetic approaches and describe a physiological function in spermatogenesis and male fertility. We show that deletion of the *Tug1* locus in mice leads to male sterility and also report an underappreciated molecular complexity at the *Tug1* locus. Using several complementary genetic approaches (gene body 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*. 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 conserved ORF, which can be translated into a stable protein that impacts mitochondrial membrane potential upon overexpression. Collectively, this study implicates the *Tug1* locus as essential in male fertility and provides evidence that the *Tug1* locus contains at least two noncoding regulatory activities and a putative coding function.

Results

The *Tug1* lncRNA locus is widely expressed and highly conserved

The murine *Tug1* lncRNA locus is located on chromosome 11 and has three annotated transcripts (Fig. 1a). *Tug1* shares a bidirectional promoter with its neighboring protein-coding gene *Morc2a*, whose transcription start site (TSS) is located approximately 680 bp upstream of the first *Tug1* TSS. The *Tug1* locus is enriched with hallmarks of active transcription, such as RNA polymerase II (Pol II) and histone H3 lysine 4-trimethylation (H3K4me3) at its promoter, H3K36me3 across its gene body, and abundant transcription as shown by RNA-seq (Fig. 1a). However, the *Tug1* locus is simultaneously enriched with the repressive histone mark H3K9me3 in several mouse cell types (Fig. 1a and Additional file 1: Fig. S1). This atypical combination of H3K9me3 and H3K36me3 histone marks at the *Tug1* locus is also conserved in human cells (Additional file 1: Fig. S1). Moreover, the binding of repressor proteins SIN3A and COREST has been detected at both the human and mouse promoters (Additional file 1: Fig. S1).

Tug1 is among the most conserved lncRNAs between human and mouse, with exonic nucleotide conservation levels reaching 77% (Fig. 1b). This level of sequence conservation is similar to the highly abundant lncRNA *Malat1* (79%) and higher than other well-characterized lncRNAs including *Hottip* (71%), *Neat1* (69%), *Xist* (30%), and *Firre* (4%) (Fig. 1b) [31]. Further conservation analyses lead us to identify a highly conserved open reading frame (ORF) in the *Tug1* locus, as indicated by phylogenetic codon



substitution frequencies (PhyloCSF) (Fig. 1a), a computational tool that can distinguish protein-coding and noncoding regions [32].

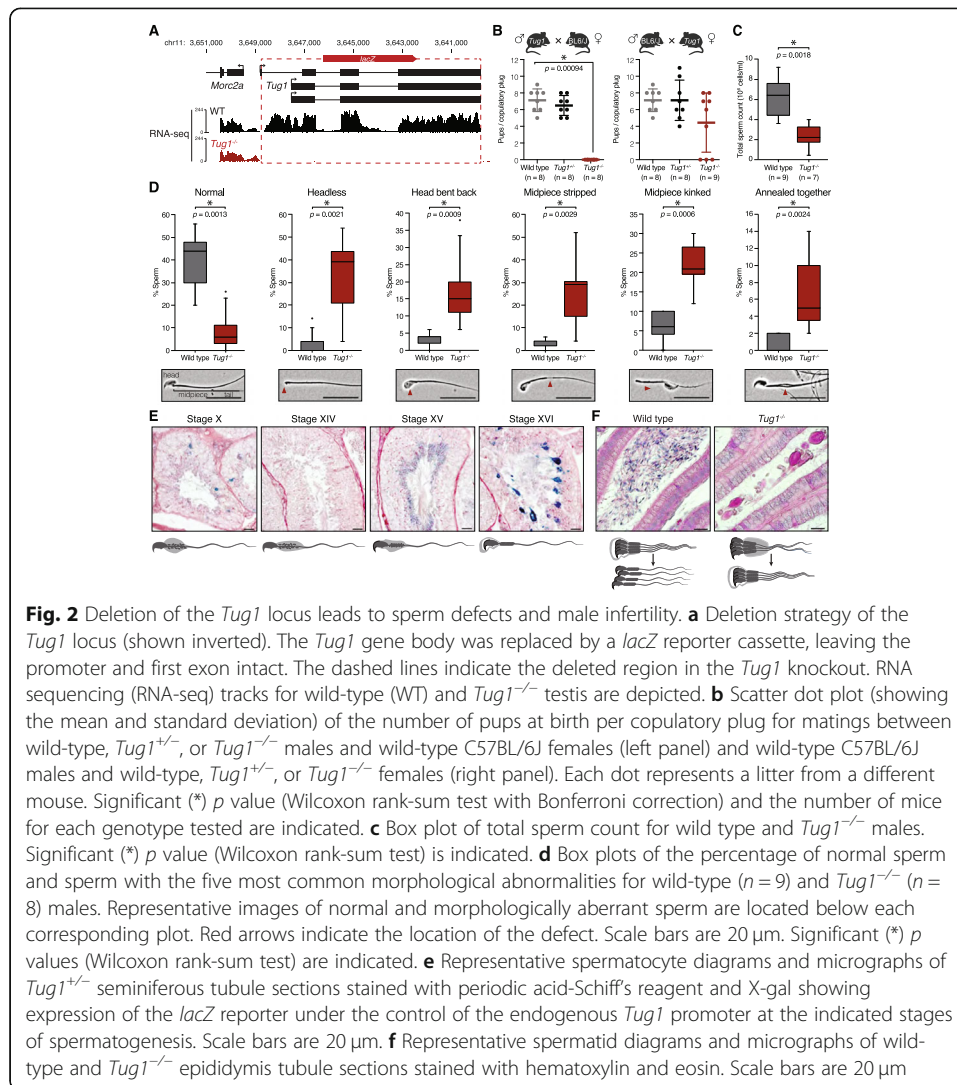
In addition to its high level of sequence conservation, *Tug1* RNA is expressed at moderate to high levels in several adult tissues in both mouse and human (Fig. 1c) [33, 34], and *Tug1* RNA is detected in a number of embryonic tissues at multiple embryonic stages (E8.0–E12.5) (Fig. 1d and Additional file 2: Fig. S2). Further, using single-molecule RNA fluorescence in situ hybridization (smFISH), we observed that *Tug1* RNA is detected in both the cytoplasm and the nucleus in human and mouse fibroblasts (Fig. 1e), which is consistent with previous reports [24, 35–37].

***Tug1*^{-/-} males are sterile due to impaired spermatogenesis**

To investigate the in vivo role of *Tug1*, we utilized a *Tug1*-knockout (*Tug1*^{-/-}) mouse model where the gene body of the *Tug1* locus was removed and replaced with a *lacZ* reporter cassette downstream of the endogenous promoter, thereby preserving the act of transcription (Fig. 2a) [20, 21]. Notably, this deletion strategy also removed 86 out of 143 amino acids in the predicted ORF (Additional file 3: Fig. S3). Thus, using this approach, any potential phenotype due to the lncRNA, potential DNA elements, or even a putative protein would be included. We confirmed the *Tug1* genetic model by genotyping and by RNA-seq analysis in wild-type and *Tug1*^{-/-} testes (Fig. 2a).

Tug1^{-/-} mice are viable and do not display any obvious physiological abnormalities up to 1 year of age, with the exception of a slight reduction in weight in male mice relative to wild-type littermates (Additional file 4: Fig. S4A). As previously reported, the progeny of *Tug1*^{+/-} intercrosses follow normal Mendelian ratios [20]. However, we noticed a complete absence of offspring from intercrosses between *Tug1*^{-/-} mice ($n = 4$ breeding pairs). Therefore, we sought to investigate the fertility of *Tug1*^{-/-} mutants in more detail. We separately mated *Tug1*^{-/-}, *Tug1*^{+/-}, and wild-type males or females to *C57BL/6J* mice. We did not observe a difference in the mounting behavior between wild-type and *Tug1*^{-/-} mice, as assessed by the presence of a vaginal plug. Strikingly, matings between *Tug1*^{-/-} males ($n = 8$) and *C57BL/6J* females did not produce any offspring, whereas matings involving either *Tug1*^{+/-} males ($n = 8$) or wild-type males ($n = 8$) with *C57BL/6J* females resulted in similar numbers of offspring (Fig. 2b). Moreover, 6 out of 9 *Tug1*^{-/-} females that mated with *C57BL/6J* males gave birth to pups (Fig. 2b), indicating that only *Tug1*^{-/-} males are sterile.

To further understand the underlying fertility defect in *Tug1*^{-/-} males, we examined the reproductive morphology of wild-type and *Tug1*^{-/-} male mice. Testicular descent appeared normal, and we did not observe any other gross morphological abnormalities in their reproductive system upon dissection (Additional file 4: Fig. S4B). We measured testis mass relative to total body weight and did not observe a significant decrease ($p = 0.0751$) in *Tug1*^{-/-} (mean = $0.25 \pm 0.020\%$, $n = 8$) compared to wild type (mean = $0.30 \pm 0.016\%$, $n = 9$) (Additional file 4: Fig. S4C). Next, we quantified sperm production and found a significant reduction in sperm number from *Tug1*^{-/-} males (mean = $2.35 \times 10^6 \pm 0.473 \times 10^6$ cells/mL, $n = 7$), which produced on average only 40% as many sperm as wild-type mice ($6.13 \times 10^6 \pm 0.636 \times 10^6$ cells/mL, $n = 9$, $p = 0.0018$) (Fig. 2c). Although *Tug1*^{-/-} males produce fewer sperm, none was found to completely lack sperm (a condition called azoospermia).



Based on these results, we investigated whether perturbations in sperm morphology could also contribute to the complete sterility observed in *Tug1*^{-/-} males. We examined the morphological features of sperm and quantified the frequency of 15 different abnormalities (Additional file 5: Table S1). Overall, the proportion of morphologically normal sperm was significantly lower in *Tug1*^{-/-} mice (mean = 8.3 ± 3.0%, *n* = 8, *p* = 0.0013) compared to wild-type males (mean = 38.9 ± 4.3%, *n* = 9) (Fig. 2d). We observed significant morphological defects in *Tug1*^{-/-} sperm including sperm with no head, misshapen head, head bent back, stripped midpiece, kinked midpiece, curled midpiece, midpiece debris, broken tail, and the presence of multiple sperm attached along the midpiece (Fig. 2d, Additional file 4: Fig. S4D, and Additional file 5: Table S1). Together, these results indicate that the sterility of *Tug1*^{-/-} males arises from a combination of low sperm count (oligozoospermia) and abnormal sperm morphology (teratozoospermia).

To further investigate how the deletion of the *Tug1* locus leads to abnormal sperm morphology, we examined the timing of *Tug1* expression at different stages of spermatogenesis. To this end, we took advantage of the knock-in *lacZ* reporter driven by the endogenous *Tug1* promoter and assessed the expression by *lacZ* staining of histological

sections of *Tug1*^{+/-} testis and epididymis. From stages IX to XI of spermatogenesis in the testis, *lacZ* staining was restricted to the excess cytoplasm, known as residual bodies, which are phagocytosed toward the basement membrane by Sertoli cells (Fig. 2e) [38]. No expression was detected in the later stages XII to XIV (Fig. 2e). However, we observed *lacZ* staining in stage XV elongated spermatids, and the *lacZ* staining appeared stronger at stage XVI, just before spermiation (Fig. 2e). The observed *lacZ* pattern indicates that *Tug1* expression is temporally controlled during spermatogenesis.

In *Tug1*^{-/-} testes, mature spermatids appeared to remain attached by their collective cytoplasm. This was even more striking in the epididymis, where multiple sperm aggregates were observed in *Tug1*^{-/-} mice, while individual sperm appeared to migrate freely throughout the lumen in wild-type mice (Fig. 2f). These aggregates were present in all regions of the epididymis (caput, corpus, and cauda). Consistent with the reduced sperm count, fewer individual sperm were observed in *Tug1*^{-/-} epididymis tissue compared to wild type. Together, our analyses of the *Tug1*^{-/-} mouse model provide evidence that the locus is required for male fertility.

***Tug1* DNA encodes a *cis* repressor regulatory element**

We next sought to investigate what, if any, molecular activities (DNA, lncRNA, or protein) are present at the *Tug1* locus. Many lncRNA loci have been reported to contain DNA regulatory elements that can regulate the expression of neighboring genes (*cis*-acting) [10–12]. The *Tug1*^{-/-} model enables us to test for potential *cis*-regulatory activity within the *Tug1* locus because the gene-ablation design removes potential *cis*-acting elements yet keeps the act of transcription intact (Fig. 2a) [20, 21]. To determine if there is a local regulatory effect on gene expression, we performed RNA-seq on the testes from wild-type and *Tug1*^{-/-} mice and plotted significant changes in gene expression within a 2-Mb region centered on the *Tug1* locus (FDR < 0.05, FC > 1.5) (Additional file 6: Table S2). Of the 71 genes within this window, we observed that 6 genes (*Rnf185*, *Pla2g3*, *Selm*, *Smtn*, *Gm11946*, and *8430429K09Rik*) were significantly upregulated in *Tug1*^{-/-} testes compared to wild type (Fig. 3a). Notably, these genes are downstream of the *Tug1* TSS and located within the same topological associated domain (TAD) in embryonic stem cells [39, 40].

To further investigate whether the *cis*-effect upon deletion of the *Tug1* locus is more widespread, we performed RNA-seq on 6 additional tissues (prostate, spleen, eyes, heart, liver, and mouse embryonic fibroblasts (MEFs)) as well as re-analyzed an existing brain dataset from wild-type and *Tug1*^{-/-} mice (Additional file 7: Table S3) [41]. Of the 71 genes within the 2-Mb region centered on the *Tug1* locus (FDR < 0.05, FC > 1.5), 9 genes were dysregulated in one or more tissues (7 upregulated and 2 downregulated) (Fig. 3a). Of the 7 upregulated genes, the E3 ubiquitin ligase, *Rnf185*, was consistently upregulated in 8 of 8 *Tug1*^{-/-} tissues, followed by the selenoprotein M gene, *Selm* (7 of 8 samples), and *8430429K09Rik* (6 of 8 samples) (Fig. 3b). We also note that *Morc2a*, the protein-coding gene that shares a promoter with *Tug1*, was significantly downregulated in 4 of the 8 samples, but this effect was not observed in the testes (Fig. 3a). Collectively, these data provide evidence of a *cis* repressor function at the *Tug1* locus in a broad range of tissues.

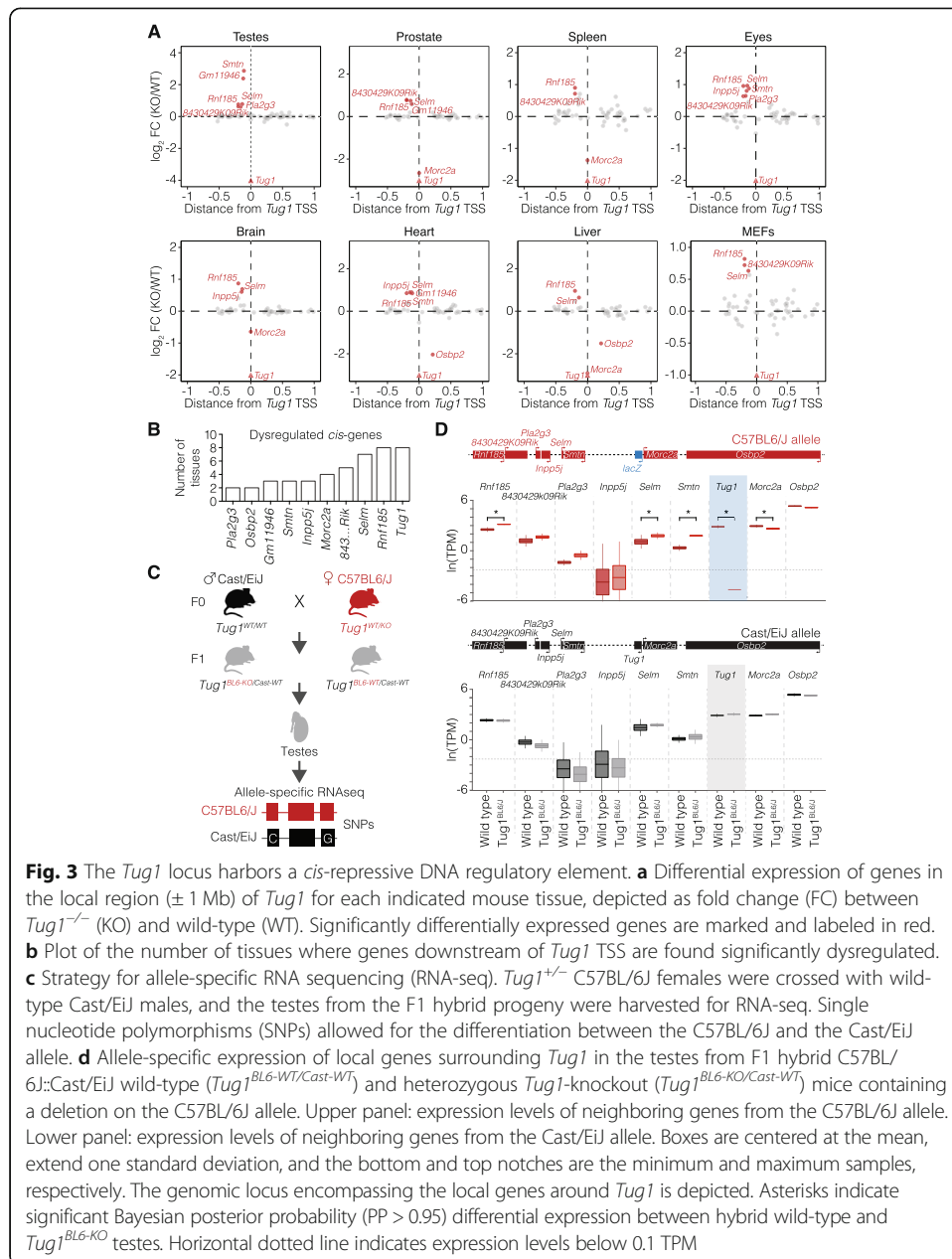


Fig. 3 The *Tug1* locus harbors a *cis*-repressive DNA regulatory element. **a** Differential expression of genes in the local region (± 1 Mb) of *Tug1* for each indicated mouse tissue, depicted as fold change (FC) between *Tug1*^{-/-} (KO) and wild-type (WT). Significantly differentially expressed genes are marked and labeled in red. **b** Plot of the number of tissues where 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 the 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 the 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, respectively. 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

Since the neighboring genes are upregulated upon deletion of the *Tug1* locus, we reasoned that the repressive activity could be mediated either directly by the *Tug1* transcript or by regulatory DNA elements within the locus. To determine if the repressive effect of *Tug1* on neighboring genes occurs on the same allele (*cis*-acting), we performed allele-specific RNA-seq using a hybrid mouse strain. To generate this strain, we crossed *Tug1*^{+/-} C57BL/6J females with *Mus castaneus* (Cast/EiJ) males (Fig. 3c). The resulting polymorphisms in the F1 hybrid progeny (~ 1/150 bp between C57BL/6J and Cast/EiJ) allow quantification of gene expression from each strain-specific allele [42]. We thus harvested the testes from F1 hybrid males harboring a maternal C57BL/6J allele deletion and performed allele-specific expression analysis (Fig. 3c; Additional file 8: Table S4; Additional file 9: Table S5). As a control for haplotype-specific effects, we

also analyzed allele-specific expression differences in wild-type F1 hybrid C57BL/6J::Cast/EiJ male littermates. We then quantified the expression from each allele and found that *Rnf185*, *Selm*, and *Smtn* were significantly upregulated and *Morc2a* slightly downregulated only on the C57BL/6J allele containing the *Tug1* deletion (Fig. 3d). Importantly, no change in the expression was detected from any gene within 1 Mb of *Tug1* on the Cast/EiJ allele, which contains an intact *Tug1* locus (Fig. 3d). Moreover, it is notable that *Tug1* RNA from the intact allele does not impact the dysregulated genes found on the *Tug1*-knockout allele, thereby suggesting a DNA-based repressor role at the *Tug1* locus. From these different mouse models, we conclude that the *Tug1* DNA, rather than the lncRNA or the act of transcription, exerts a repressive effect in *cis* on several genes up to 200 kb downstream of the *Tug1* transcription site.

Tug1* lncRNA regulates gene expression in *trans

To gain further insight into the molecular roles of *Tug1*, we took a gene expression approach. We analyzed RNA-seq data from WT and *Tug1*^{-/-} tissues (testes, prostate, spleen, eyes, liver, heart, brain, and MEFs) and identified significant changes in the gene expression relative to wild type. Deletion of the *Tug1* locus was accompanied by 2139 significantly dysregulated genes across all samples examined. We observed global changes in the gene expression clustered by tissue type, indicating tissue-specific gene dysregulation (Fig. 4a; Additional file 6: Table S2; Additional file 7: Table S3; Additional file 10: Fig. S5). We found that while most dysregulated genes (~89%) were perturbed in only a single tissue (Fig. 4b), several genes were commonly dysregulated across multiple tissues (Fig. 4b; Additional file 6: Table S2; Additional file 7: Table S3). We then performed a gene set enrichment analysis (GSEA) using the differentially expressed genes for each tissue and observed enrichment of several pathways that were shared across the individual tissues. For example, oxidative phosphorylation, Myc targets, and epithelial to mesenchymal transition were found enriched in 7 of the 8 *Tug1*^{-/-} tissues (Fig. 4c).

Previous studies have suggested a role for *Tug1* RNA acting in *trans* on chromatin regulation and gene expression [22, 26, 28, 36, 43, 44]. Thus, we wanted to further investigate a *trans* role of *Tug1* RNA on the gene expression in the testes. We reasoned that overexpressing *Tug1* RNA in the *Tug1*-knockout background would enable to test if *Tug1* RNA alone could rescue genes found dysregulated in *Tug1*^{-/-} testes (Fig. 4d). Given that *Tug1* harbors a conserved ORF in the 5' region (discussed in the next section), we identified an isoform of *Tug1* that lacks the 5' region, thus ensuring we would address the role of *Tug1* RNA alone. To this end, we generated a doxycycline (dox)-inducible *Tug1* transgenic mouse by cloning a *Tug1* isoform downstream of a tet-responsive element (henceforth called tg(*Tug1*)) (Fig. 4d). Next, we generated compound transgenic mice that contained tg(*Tug1*) in the *Tug1*^{-/-} background that also contained an allele that constitutively expresses the reverse tetracycline transcriptional activator gene (CAG-rtTA3) (combined alleles henceforth called *Tug1*^{rescue}) (Fig. 4d). This approach enabled systemic induction of *Tug1* RNA in the presence of dox, enabling to test if *Tug1* RNA expression alone would be sufficient to rescue gene expression in the testes and the male sterility phenotypes in *Tug1*^{-/-} mice.

Furthermore, we used fluorescence-activated cell sorting (FACS) and isolated peripheral blood cell types (CD4, CD8, and NK) from wild-type, *Tug1*^{+/-}, *Tug1*^{-/-}, and dox-fed *Tug1*^{rescue} mice. qRT-PCR also showed lower levels of *Tug1* RNA induction relative to wild type, but *Tug1* transgene expression in peripheral blood cell types was generally higher than in the testes (Additional file 11: Fig. S6A-B). Even though the transgene expression was low in the testes, we reasoned that this would still be a valuable in vivo model to test *Tug1* RNA-mediated effects on gene regulation. Thus, we tested whether genes found dysregulated in the testes from *Tug1*^{-/-} mice could be rescued by ectopic expression of the *Tug1* transgene RNA in the *Tug1*^{rescue} model (Additional file 12: Fig. S7). We found that there was significant overlap of dysregulated genes between *Tug1*^{-/-} and *Tug1*^{rescue} testes (p value < 2.23–16, Fisher exact test). Notably, 53 (including *Tug1*) of the 1051 genes that were significantly dysregulated in *Tug1*^{-/-} testes were found significantly and reciprocally regulated in the testes from dox-fed *Tug1*^{rescue} mice (Fig. 4g, Table 1, and Additional file 6: Table S2). For example, a mitochondrial-related gene, *Mrarp*, and an aquaporin gene, *Aqp2*, are significantly upregulated in *Tug1*^{-/-} testes, but their expression was reduced to wild-type levels in the testes from dox-fed *Tug1*^{rescue} mice (Fig. 4h). Conversely, the predicted lncRNA gene *Gm28181* that is significantly reduced in *Tug1*^{-/-} testes is significantly upregulated to wild-type levels in the testes from dox-fed *Tug1*^{rescue} mice (Fig. 4h). While we observed a *trans*-effect for *Tug1* RNA, we did not observe any changes in the expression for the neighboring genes near the *Tug1* locus (Fig. 4f; Additional file 6: Table S2). Taken together, these data demonstrate that the *Tug1* lncRNA regulates a subset of genes by a *trans*-RNA-based mechanism, evident even at low levels of *Tug1* RNA.

Next, we asked if dox-fed *Tug1*^{rescue} male mice had restored fertility and normal sperm production and morphology. Matings between dox-fed *Tug1*^{rescue} male mice ($n = 3$) and C57BL6/J female mice ($n = 12$) did not produce any progeny (Additional file 11: Fig. S6C). Moreover, we found that dox-fed *Tug1*^{rescue} males had a low sperm count (mean = $3.20 \times 10^5 \pm 8.0 \times 10^3$ cells/mL), similar to the levels observed in *Tug1*^{-/-} males (mean = $4.69 \times 10^5 \pm 1.6 \times 10^4$ cells/mL) compared to wild type (mean = $9.32 \times 10^5 \pm 3.9 \times 10^3$ cells/mL) (Additional file 11: Fig. S6D). These results were also confirmed with histological sectioning of the testes (Additional file 11: Fig. S6E). Further, we observed that dox-fed *Tug1*^{rescue} mice had a low proportion of normal shaped sperm, which was similar to the sperm observed in *Tug1*^{-/-} mice (Additional file 11: Fig. S6F). While this finding may suggest that the sterility phenotype is not due to the lncRNA transcript, we note that the lack of a fertility rescue may also be due to the low levels of *Tug1* expression from the transgene in the testes or because a different *Tug1* RNA isoform is required.

The *Tug1* locus contains an evolutionary conserved ORF in humans and mice

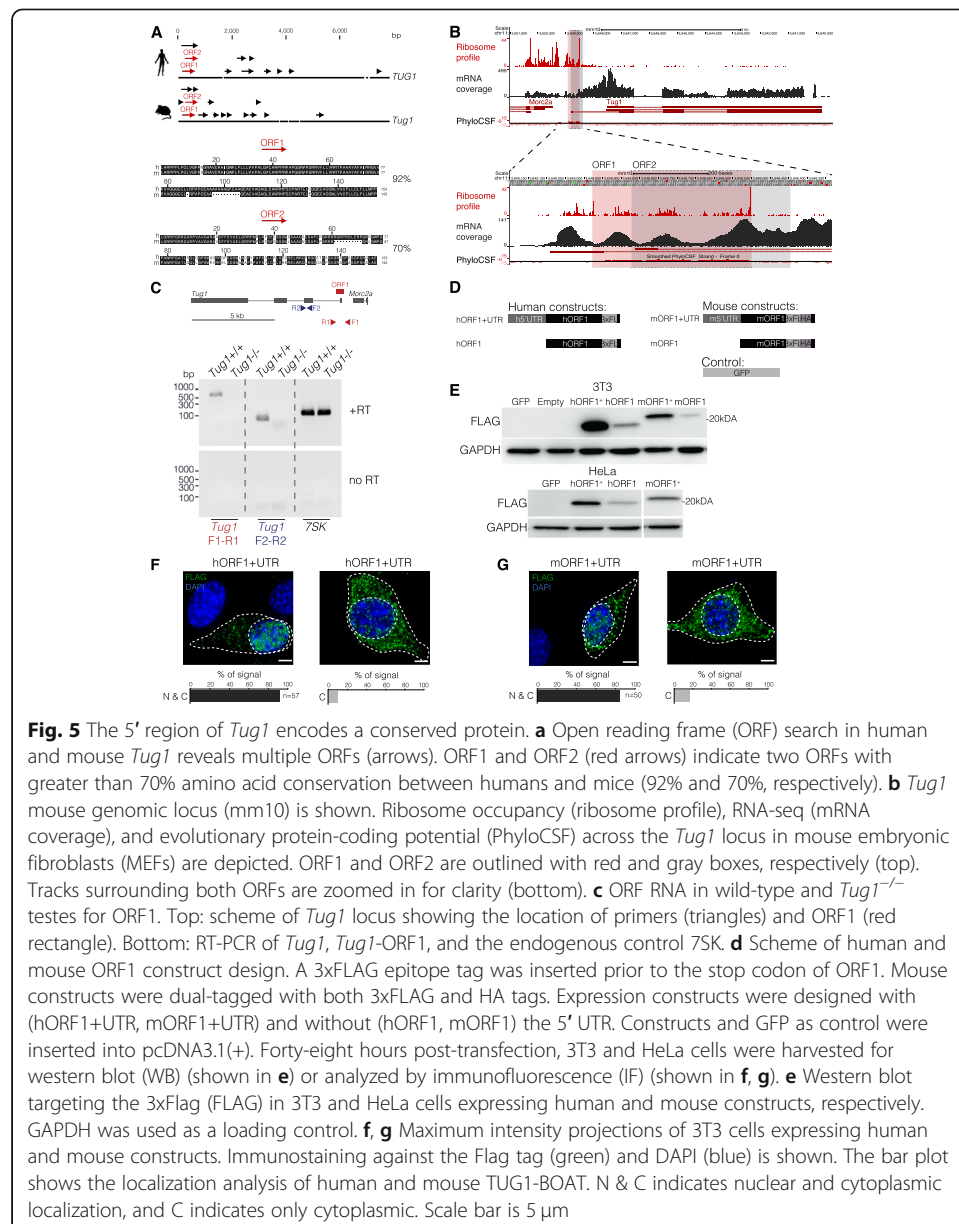
A number of studies have reported that some lncRNA loci can also harbor ORFs that can produce functional proteins [45]. Because PhyloCSF revealed a region with high coding potential in the human and mouse *TUG1/Tug1* locus (Fig. 1a; Additional file 13: Fig. S8A), we wanted to further explore this possibility using computational, biochemical, and cell-based assays. We systematically screened for ORFs that displayed strong conservation across species, allowing for both canonical

Table 1 Genes reciprocally regulated by *Tug1* lncRNA in the testes

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

List of genes with TPM ≥ 1 and significant changes in the 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 versus *Tug1*^{-/-} (WT-KO) and *Tug1*^{-/-} versus *Tug1*^{rescue} (KO-Rescue) is indicated by asterisks (* $p < 0.05$; ** $p < 0.01$, *** $p < 0.001$)

(AUG) and non-canonical (CUG and UUG) translation start codons, a feature that has been previously identified at lncRNA loci [13, 46]. We identified multiple short ORFs in the human and mouse *TUG1/Tug1* locus (11 and 15, respectively) (Fig. 5a), which is consistent with previous studies [22, 47, 48]. Two ORFs (designated as ORF1 and ORF2) at the 5' region of *TUG1/Tug1* drew our attention due to their conserved translational start and stop sites, as well as their high level of nucleotide conservation between humans and mice (Fig. 5a). ORF1 (154 amino acids in humans) and ORF2 (153 amino acids in humans) both start with a non-canonical start codon (CUG) and share 92% and 70% cross-species identity at the amino acid level, respectively. Notably, ORF1 has a high PhyloCSF score (350) and shows conservation spanning its entire sequence, whereas ORF2 does not show patterns



of preserving synonymous mutations past the ORF1 stop codon (Fig. 5b; Additional file 13: Fig. S8A).

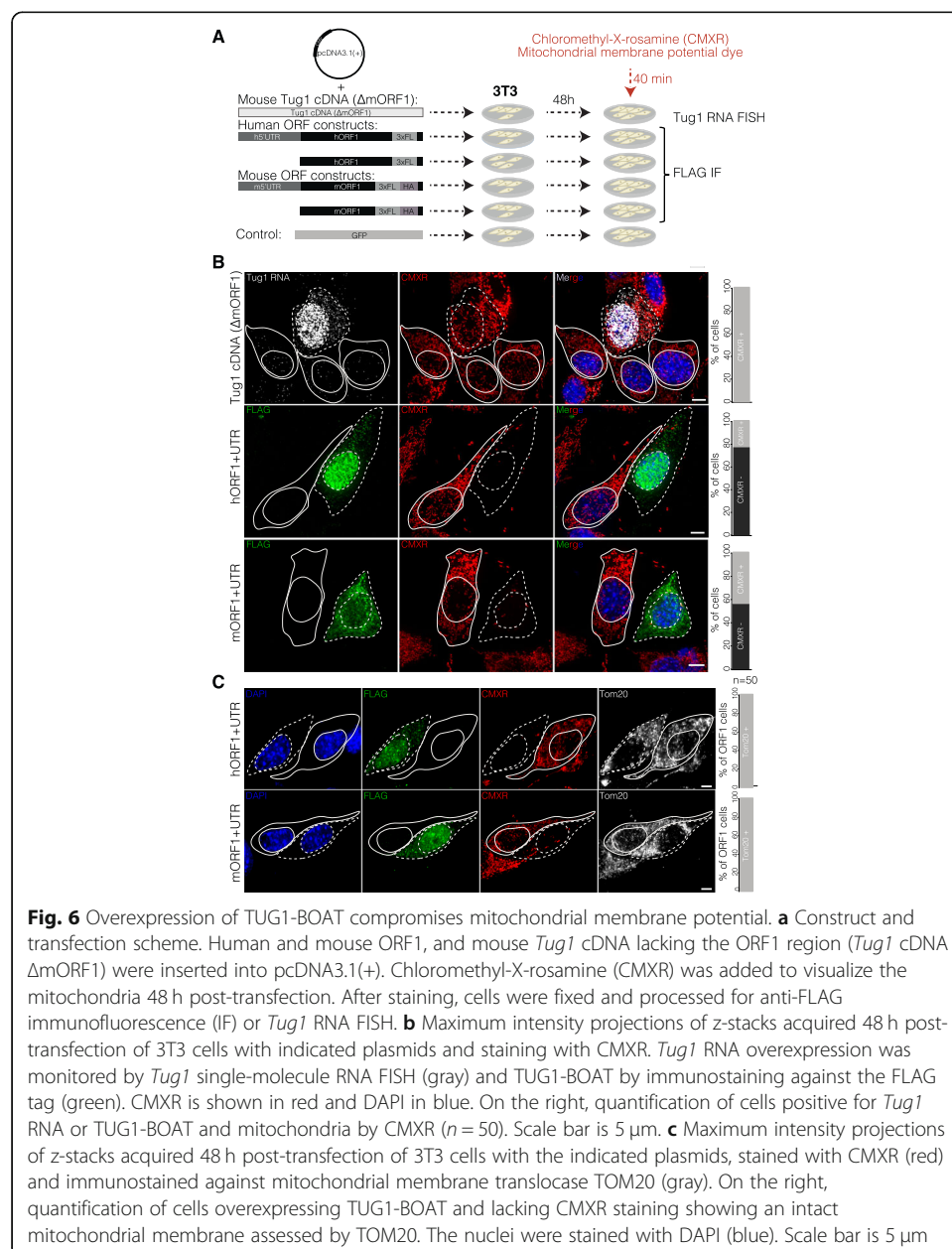
To further characterize the potential translated regions in *Tug1*, we analyzed publicly available ribosome profiling data [49], a technique used to identify regions where RNAs are bound to ribosomes by high-throughput sequencing [13, 50–52]. We found pronounced ribosomal occupancy across the entire *Tug1* ORF1 sequence with a sharp decrease at its stop codon (Fig. 5b). A similar pattern of *Tug1* ORF1 is also observed from ribosome profiling in human, mouse, and rat heart tissue (Additional file 13: Fig. S8) [47], whereas ORF2 does not show ribosome occupancy above the background level after the ORF1 stop codon (Fig. 5b, Additional file 13: Fig. S8A). To determine if a transcript containing ORF1 (located in the 5' region) is expressed in the testes, we performed reverse transcriptase PCR (RT-PCR) on wild-type and *Tug1*^{-/-} testes. Using a primer set that spans ORF1, we detected a band at approximately 700 bp in the wild-type sample in which the RT enzyme was added and did not detect a product in either the wild-type sample with no RT enzyme added or in the *Tug1*^{-/-} control samples (Fig. 5c). As a loading control, *7SK*, an abundant snRNA [53, 54], is detected in both wild-type and *Tug1*^{-/-} samples (Fig. 5c). Taken together, these results provide evidence that the 5' region containing ORF1 is expressed in the testes and presents the possibility that this ORF could generate the predicted protein. Thus, we designated the putative protein originating from ORF1 as TUG1-BOAT (*Tug1*-Bifunctional ORF and Transcript).

Next, to determine if TUG1-BOAT can produce a stable protein, we generated C-terminal epitope-tagged human and mouse TUG1-BOAT expression constructs with and without the 5' leader sequences (Fig. 5d). As a negative control, we generated a construct containing GFP in place of the TUG1-BOAT cDNA sequence. We then transfected 3T3 (mouse) and HeLa (human) cells and tested for TUG1-BOAT translation by western blot analysis. We detected bands at approximately 19 kDa and 21 kDa in both cell lines (Fig. 5e), which closely corresponds to the predicted molecular weights of hTUG1-BOAT (18.7 kDa) and mTUG1-BOAT (19 kDa) fusion constructs, respectively. Furthermore, the presence of the 5' UTR notably enhances the translation of human and mouse ORF1. Having detected a protein of the expected size from human and mouse TUG1-BOAT constructs, we next investigated the protein's subcellular localization by immunofluorescence. We observed that human and mouse TUG1-BOAT is distributed throughout the nucleus and cytoplasm in the majority of the cells (> 80% cells, *n* = 50) (Fig. 5f, g). However, in a subset of cells, TUG1-BOAT was predominantly cytoplasmic (< 20% of cells, *n* = 50) (Fig. 5f, g). We further found that TUG1-BOAT co-localizes with the mitochondria in an overexpression context (Additional file 13: Fig. S8C). Collectively, these results show that ORF1, with its 5' UTR and native non-canonical translational start site, can be translated into a stable protein in both human and mouse cells.

TUG1-BOAT overexpression compromises mitochondrial membrane potential

Given the localization of overexpressed TUG1-BOAT to the mitochondria and that oxidative phosphorylation was one of the most affected pathways across multiple

Tug1^{-/-} tissues (Fig. 4c), we hypothesized that TUG1-BOAT may have a role in the mitochondria. To this end, we first examined the mitochondrial membrane potential by using chloromethyl-X-rosamine (CMXR), a lipophilic fluorescent cation that accumulates in the negatively charged interior of mitochondria [55]. We overexpressed human and mouse TUG1-BOAT with and without the 5' UTR in 3T3 cells, along with the controls, a GFP-containing plasmid, and a *Tug1* construct lacking ORF1 (*Tug1* cDNA ΔmORF1) (Fig. 6a). Notably, cells with either human or mouse TUG1-BOAT showed a reduction in mitochondrial staining by CMXR (22% and 44% CMXR stained cells, respectively), compared to cells in the same culture not expressing TUG1-BOAT (Fig. 6b). In contrast, cells expressing GFP or *Tug1* cDNA ΔmORF1 were positive for CMXR staining in all cells examined, thus indicating that CMXR staining deficiency is



induced by overexpression of the TUG1-BOAT protein alone, rather than the *Tug1* RNA (Fig. 6b; Additional file 14: Fig. S9).

Since CMXR is commonly used to measure mitochondrial membrane potential, we reasoned that either impaired mitochondrial integrity or impaired redox potential at the mitochondrial membrane could account for the accumulation defect of CMXR in mitochondria upon TUG1-BOAT overexpression. To address these possibilities, we immunostained for TOM20, a redox independent translocase located on the outer mitochondrial membrane [56]. We observed staining for TOM20 in cells without CMXR staining, indicating that mitochondria were intact in cells overexpressing human or mouse TUG1-BOAT (Fig. 6c). Taken together, these results suggest that ectopic levels of TUG1-BOAT protein alter mitochondrial membrane potential.

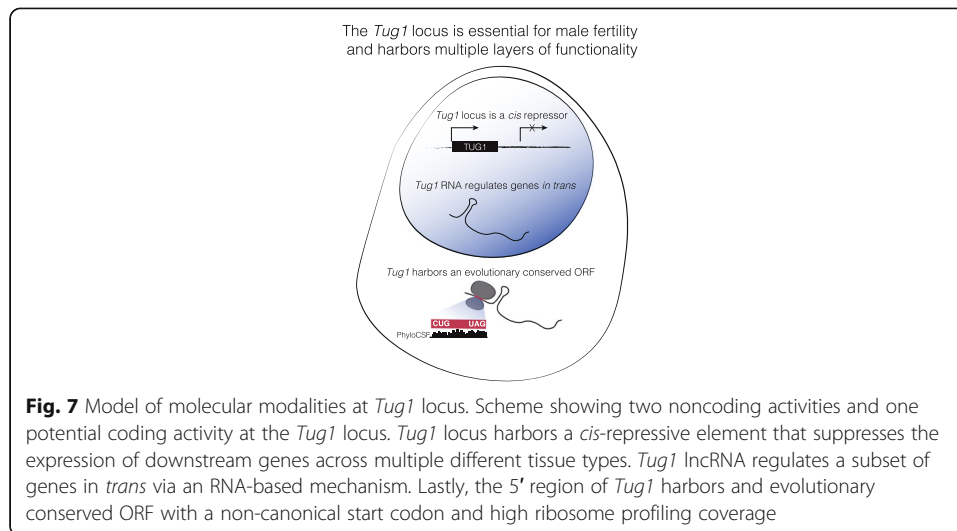
Discussion

Noncoding RNAs play roles in diverse biological functions, and their loci can possibly harbor multiple molecular modalities (DNA, RNA, the act of transcription, and misannotated proteins), each with the potential to exert function. In this study, we report an essential role in male fertility for the *Tug1* lncRNA locus in vivo and also report an underappreciated molecular complexity at the *Tug1* locus (Fig. 7). We find evidence that the *Tug1* locus harbors a (i) *cis*-acting DNA repressive element, (ii) a lncRNA that acts on gene expression in *trans*, and (iii) a conserved putative ORF which when overexpressed affects mitochondrial membrane potential (Fig. 7). Together, our results have several important biological implications.

LncRNAs in male fertility

First, our study indicates that the *Tug1* locus has an essential role in male fertility in mice. Based on our results, we speculate that the sterility of *Tug1*^{-/-} males arises from a combination of oligozoospermia (low sperm count) and teratozoospermia (abnormal morphology). Notably, there is some evidence that *Tug1/TUG1* may have a conserved role in male fertility in humans. Microarray profiling of sperm from sterile men (a condition called teratozoospermia) has significantly less *TUG1* expression compared with sperm from fertile males (Additional file 15: Fig. S10) [57]. Thus, we speculate that the *Tug1*^{-/-} mouse model used in this study could potentially be of value to further investigate the functions of *Tug1/TUG1* with relevance to human male fertility.

Indeed, studies have performed systematic gene expression profiling at defined stages during spermatogenesis and have identified many developmentally regulated lncRNAs, suggesting that lncRNAs may have important roles during spermatogenesis [58]. Four other lncRNAs (*Tslrn1*, *Tsx*, *Pldi*, and *Mrhh*) have been found to be important in spermatogenesis, but in contrast to the *Tug1*-knockout phenotype reported in this study, none lead to male sterility when disrupted in mouse models [58–61]. On this note, identifying overt functions of lncRNA loci with respect to embryogenesis, viability, and fertility has been a challenge. A recent study that generated 32 deletion alleles for 25 lncRNA loci in zebrafish reported one mutant with abnormal development [62]. Thus, overt functions for individual lncRNA loci, such as the one observed for *Tug1*, may be less common.



Multiple molecular modalities of the *Tug1* locus

Second, we found that the *Tug1* locus harbors at least two distinct noncoding activities, and potentially a third coding one. Several lines of evidence indicate that the *Tug1* locus has a *cis*-acting repressive DNA function. We observed that upon deletion of the *Tug1* locus, several genes located downstream of *Tug1* were consistently upregulated across multiple tissues, and in an allele-specific manner in the testes. This dysregulation is consistent with a previous study from our group in which genes located near the *Tug1* locus were dysregulated in the brain of *Tug1*^{-/-} mice [41]. Collectively these data suggest that the local repressive *cis*-effect is mediated by DNA regulatory elements within the *Tug1* locus. Consistent with a repressive modality of the *Tug1* locus, a recent study which systematically tested for enhancer activity across lncRNA loci, reported a lack of enhancer activity across the *Tug1* gene body but found enhancer activity at all other lncRNA loci examined [11]. Contrary to enhancers, only a handful of repressor elements and silencers have been identified and characterized in mammalian genomes [63–66]. Further defining the precise DNA repressive elements as well as their mechanism will be of interest to understand the regulatory principles of DNA elements within the *Tug1* locus.

In addition to a DNA regulatory modality, this study finds evidence for an RNA-based role for *Tug1* in the testes by a *trans* mechanism. In support of this conclusion, we found that a subset of genes dysregulated in *Tug1*^{-/-} testes could be rescued by ectopic expression of *Tug1* RNA, even at low levels. While the *Tug1* transgene was expressed at lower levels than wild-type *Tug1* RNA, other lncRNAs such as *Hottip* and *Xist* have been shown to exert a biological activity at relatively low copy numbers [67, 68]; thus, *Tug1* RNA appears to have functional activity even at low levels. Consistent with our finding of a *trans*-acting role for *Tug1* RNA on gene expression, a previous study found that *Tug1* RNA can regulate the levels of *Ppargc1a* mRNA in cultured podocytes [26]. In our RNA-seq dataset of 8 different tissues, *Ppargc1a* was not significantly dysregulated, but it is important to note that our dataset does not include the kidney. We also note that low levels of *Tug1* transgene RNA were not sufficient to

rescue the *Tug1* male fertility defect. More studies will be needed to determine the direct or indirect mechanisms of *Tug1* RNA on gene expression in a male germ cell context in vivo.

A number of recent studies have identified ORFs at lncRNA loci and have also demonstrated that some ORFs can produce proteins [13, 15–18, 46, 50, 51]. Thus, we also explored the possibility of a protein modality embedded at the *Tug1* locus. We demonstrate that the 5' region of the *Tug1* locus contains an evolutionarily conserved ORF that when overexpressed, impacts mitochondrial membrane potential. The protein product which we named TUG1-BOAT is predicted to have a high positive charge (net charge $\sim +16.5$). Thus, we speculate that the accumulation of such a positively charged protein at the mitochondria in an overexpression context could lead to depolarization of the mitochondrial membrane. Consistent with this finding, a recent study also identified a conserved ORF in the 5' region of the human *TUG1* locus and showed that the ORF produces a stable protein that localizes to the mitochondria [47]. In addition, the evidence for a mitochondrial role for *Tug1* does not appear to be limited to the putative protein. One study found that overexpression of a *Tug1* RNA isoform lacking ORF1 affects mitochondrial bioenergetics in cultured podocytes derived from a diabetic nephropathy mouse model [26]. In our study, we did not observe any effect of *Tug1* RNA (lacking the ORF1 sequence) overexpression on mitochondrial membrane potential by CMXR staining in cell-based assays. Identifying endogenous and disease contexts in which *Tug1* RNA and TUG1-BOAT are detected and function endogenously will be important to further explore the potential link between *Tug1* and the mitochondria and determine whether the locus does contain a third coding function.

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

Finally, this study opens the possibility that *Tug1* DNA, RNA, or putative protein could individually or in combination mediate the observed male fertility defect in *Tug1*-knockout mice. There are few biological contexts in the literature describing noncoding loci implicated in human disease with potential DNA, RNA, and protein functionalities. However, one intriguing example is the human *D4Z4* locus, which is a repeat region that is reduced in copy number in humans with facioscapulohumeral dystrophy (FSHD) [69, 70]. Interestingly, in the disease context where the repeat number is reduced, local repressive chromatin is lost leading to transcriptional upregulation of proximally located genes [71, 72]. Mechanistic studies spanning decades have identified a number of candidate molecular modalities at the *D4Z4* locus, including a lncRNA, a protein that is detected in the disease context, and a DNA repressor element [71–76]. While our study does not resolve the molecular modes of action mediating the male fertility defect in *Tug1*-knockout mice, there is evidence to support a potential role in male fertility for each modality, thus warranting further investigation.

There is some evidence that the *cis* genes dysregulated upon deletion of the *Tug1* locus have a role in male fertility. Loss-of-function mutations in 2 of the 6 *cis* genes upregulated in *Tug1*^{-/-} testes, *Smtn* and *Pla2g3*, are characterized to have male fertility and sperm maturation defects [77, 78]. However, the impact of these genes on male fertility in an overexpression context has not been reported. A role in male fertility for the

other four upregulated *cis* genes upon deletion of *Tug1* (*Gm11946*, *Rnf185*, *Selm*, and *8430429K09Rik*) have not yet been reported. In addition, there is some evidence for a potential role of *Tug1* RNA in male fertility. Mice with a loss-of-function mutation for *Selenop*, a gene that was found significantly upregulated in *Tug1*^{-/-} testes and significantly and reciprocally regulated in *Tug1*^{rescue} testes, are reported to have reduced male fertility [79]. As for a role for the putative protein, TUG1-BOAT, there is currently limited evidence for its endogenous existence. Thus, future work clarifying an endogenous presence and function for TUG1-BOAT will be necessary to understand whether it contributes to the male fertility defect.

Conclusions

In summary, this study provides genetic evidence for an essential role of the *Tug1* locus in male fertility. This study also highlights the molecular complexity embedded at lncRNA loci as our findings provide evidence that the *Tug1* locus harbors two distinct noncoding activities: (i) a *cis* DNA repressor and (ii) a *trans*-acting lncRNA, and we find a potential third protein-coding activity. Therefore, going forward, it will be important to investigate the individual and/or combined roles of *Tug1* DNA, RNA, and/or putative protein in male fertility as well as in disease contexts in which *Tug1* is altered.

Methods

Mice and ethics statement

Tug1^{tm1.1Vlcg}-knockout mice have been described previously [20, 41]. To remove the *loxP*-flanked neomycin resistance gene included in the targeting construct, we crossed *Tug1*^{tm1.1Vlcg} mice to *C57BL6/J* mice and then to a cre-recombinase strain (*B6.C-Tg*^{(CMV-cre)1Cgn/J}, The Jackson Laboratory, 006054). Mice free of both the neomycin-resistance and cre-recombinase genes were selected for colony expansion and subsequently backcrossed to *C57BL/6J* mice. The *Tug1*-knockout allele was maintained by heterozygous breeding, and mutant mice were identified by genotyping for loss of the *Tug1* allele and gain of the *lacZ* cassette (Transnetyx, Inc.).

For allele-specific gene expression analyses, we generated *Tug1*^{BL6-KO/Cast-WT} mice by crossing inbred *Mus castaneus* (Cast/EiJ) males (The Jackson Laboratory, 000928) with inbred heterozygote *Tug1* females. The F1 hybrid male progeny (three wild-type *Tug1*^{BL6-WT/Cast-WT} and four with a maternal *Tug1*-knockout allele *Tug1*^{BL6-KO/Cast-WT}) were used for allele-specific expression studies.

To generate an inducible *Tug1*-overexpression mouse, tg(*Tug1*), we cloned *Tug1* cDNA (see the “Sequences and primers” section) into a Tet-On vector (pTRE2). Full-length *Tug1* (Ensembl ID: ENSMUST00000153313.2) was amplified from Riken cDNA clone E330021M17 (Source Bioscience) using specific primers containing MluI and EcoRV restriction sites (see the “Sequences and primers” section). After gel purification, we sub-cloned the amplicon using the MluI and EcoRV restriction sites into a modified Tet-On pTRE2pur vector (Clontech 631013) in which the bGlobin-intron was removed. We verified the absence of mutations from the cloned *Tug1* cDNA by sequencing (see the “Sequences and primers” section). We injected this cassette into the pronucleus of *C57BL/6J* zygotes (Beth Israel Deaconess Medical Center Transgenic Core). Two male founder mice containing random integration of the tg(*Tug1*) cassette were identified by

genotyping for the pTRE allele and individually mated to female C57BL/6J mice (The Jackson Laboratory, 000664) to expand the colonies. Next, we generated quadruple allele transgenic mice to test the functionality of the *Tug1* RNA by the following strategy. We mated tg(*Tug1*) males to *Tug1^{tm1.1Vlsg}* females and identified male progeny that were *Tug1^{+/-}*; tg(*Tug1*). These mice were then mated to female *rtTA* mice (B6N.FVB(Cg)-Tg(CAG-rtTA3)4288Slowe/J mice (The Jackson Laboratory, 016532)), and we identified male progeny that were *Tug1^{+/-}*; tg(*Tug1*), *rtTA*. Finally, we mated male *Tug1^{+/-}*; tg(*Tug1*), *rtTA* mice to *Tug1^{+/-}* females, and at the plug date, females were put on 625 mg/kg doxycycline-containing food (Envigo, TD.01306). We genotyped progeny from the above matings (Transnetyx, Inc.) and identified male progeny that were *Tug1^{-/-}*; tg(*Tug1*), *rtTA*, and maintained these mice on the doxycycline diet until the experimental end point.

Cell lines and cell culture

We derived primary wild-type and *Tug1^{-/-}* mouse embryonic fibroblasts (MEFs) from E14.5 littermates from timed *Tug1^{+/-}* intercrosses as described [80]. We maintained MEFs as primary cultures in DMEM, 15% FBS, pen/strep, L-glutamine, and non-essential amino acids. We genotyped MEFs derived from each embryo and used only male *Tug1^{-/-}* and wild-type littermate MEFs at passage 2 for all experiments. 3T3 (ATCC, CRL-1658), HeLa (ATCC, CRM-CCL-2), and BJ (ATCC, CRL-2522) cell lines were purchased from ATCC and cultured as recommended.

Whole-mount in situ hybridization

We generated an antisense riboprobe against *Tug1* (see the “Sequences and primers” section) from plasmids containing full-length *Tug1* cDNA (Ensembl id: ENSMUST00000153313.2) and performed in situ hybridization on a minimum of three C57BL6/J embryos per embryonic stage. For whole-mount staining, embryos were fixed in 4% paraformaldehyde for 18 h at 4 °C, followed by three washes in PBS for 10 min. We then dehydrated embryos through a graded series of 25%, 50%, and 75% methanol/0.85% NaCl incubations and then finally stored embryos in 100% methanol at -20 °C. Embryos were then rehydrated through a graded series of 75%, 50%, 25%, and methanol/0.85% NaCl incubations and washed twice with PBS with 0.1% Tween-20 (PBST). Embryos were treated with 10 mg/mL proteinase K in PBST for 10 min (E8.0, E9.5) or 30 min (E10.5, E11.5, and E12.5). Samples were fixed in 4% paraformaldehyde/0.2% glutaraldehyde in PBST for 20 min at room temperature and washed twice with PBST. We then incubated samples in pre-hybridization solution for 1 h at 68 °C and then incubated samples in 500 ng/mL of *Tug1* antisense or sense riboprobe at 68 °C for 16 h. Post-hybridization, samples were washed in stringency washes and incubated in 100 µg/mL RNaseA at 37 °C for 1 h. Samples were washed in 1× maleic acid buffer with 0.1% Tween-20 (MBST) and then incubated in Roche Blocking Reagent (Roche, #1096176) with 10% heat-inactivated sheep serum (Sigma, S2263) for 4 h at room temperature. An anti-digoxigenin antibody (Roche, 11093274910) was used at 1:5000 and incubated for 18 h at 4 °C. Samples were washed 8 times with MBST for 15 min, 5 times in MBST for 1 h, and then once in MBST for 16 h at 4 °C. Prior to developing, the samples were washed three times with NTMT (100 mM NaCl, 100 mM Tris-HCl (pH 9.5), 50 mM MgCl₂, 0.1% Tween-20, 2 mM

levamisole). The in situ hybridization signal was developed by adding BM Purple (Roche, 11442074001) for 4, 6, 8, and 12 h. After the colorimetric development, samples were fixed in 4% paraformaldehyde and cleared through a graded series of glycerol/1× PBS and stored in 80% glycerol. Imaging was performed on a Leica M216FA stereomicroscope (Leica Microsystems) equipped with a DFC300 FX digital imaging camera.

***Tug1* single-molecule RNA FISH**

We performed *Tug1* single-molecule RNA FISH as described previously [81]. Briefly, 48 oligonucleotides labeled with Quasar 570 and Quasar 670 tiled across human/mouse *Tug1* transcripts were designed with LGC Biosearch Technologies' Stellaris probe designer (Stellaris Probe Designer version 4.2) and manufactured by LGC Biosearch Technologies.

Human foreskin fibroblasts (ATCC[®] CRL-2522[™]) and mouse 3T3 fibroblasts (ATCC, CRL-1658[™]) were seeded on glass coverslips previously coated with poly-L-lysine (10 µg/mL) diluted in PBS. Prior to hybridization, coverslips were washed twice with PBS, fixed with 3.7% formaldehyde in PBS for 10 min at room temperature, and washed twice more with PBS. Coverslips were immersed in ice-cold 70% EtOH and incubated at 4 °C for a minimum of 1 h. We then washed the coverslips with 2 mL of Wash Buffer A (LGC Biosearch Technologies) at room temperature for 5 min. Next, we hybridized cells with 80 µL hybridization buffer (LGC Biosearch Technologies) containing *Tug1* probes (1:100) overnight at 37 °C in a humid chamber. The following day, we washed the cells with 1 mL of Wash Buffer A for 30 min at 37 °C, followed by another wash with Wash Buffer A containing Hoechst DNA stain (1:1000, Thermo Fisher Scientific) for 30 min at 37 °C. Coverslips were washed with 1 mL of Wash Buffer B (LGC Biosearch Technologies) for 5 min at room temperature, mounted with ProlongGold (Life Technologies) on glass slides, and left to curate overnight at 4 °C before proceeding to image acquisition (see below).

Sperm counts and morphology

Tug1^{-/-} ($n = 8$) and wild-type ($n = 9$) males between 8 and 41 weeks of age were sacrificed and weighed. We then dissected the entire male reproductive tract in phosphate-buffered saline (PBS). One testis was removed, weighed, and fixed in 4% paraformaldehyde (PFA) for histology (see below). Sperm were collected from one cauda epididymis by bisecting and suspending the tissue in a solution of Biggers-Whitten-Whittingham (BWW) sperm media at 37 °C. After a 15-min incubation, we used the collected sperm solutions to analyze sperm morphology and counts.

We characterized sperm morphology by fixing sperm in 2% PFA in PBS, mounting 20 µL of suspended sperm in Fluoromount-G media (Southern Biotech) on Superfrost glass slides (Thermo Fisher Scientific) and scanning each slide in a linear transect, recording the morphology as normal or abnormal for each sperm cell encountered (between 30 and 120 sperm). When abnormal, we also recorded the type of morphological defects: headless, head angle aberrant, head bent back to midpiece, debris on the head, debris on the hook, head misshapen, midpiece curled, midpiece kinked, midpiece stripped, debris on the midpiece, tailless, tail curled, tail kinked, broken tail, or multiple cells annealed together.

Sperm counts for each *Tug1*^{-/-} ($n = 7$) and wild-type ($n = 9$) mice were determined using a Countess Automated Cell Counter according to the manufacturer's protocol (Life Technologies, Carlsbad, CA). For the *Tug1*^{rescue} experiment, sperm counts for control (WT and

Tug1^{+/-} (*n* = 2), *Tug1*^{-/-} (*n* = 2), and *Tug1*^{-/-}; *tg(Tug1)*; *rtTA* mice (*n* = 3) were determined by manual counts using a hemocytometer. For all analyses, statistical comparisons between *Tug1*^{-/-} and wild type were performed using the two-tailed Wilcoxon rank-sum tests with an $\alpha = 0.05$. The results for testis, sperm count, and morphological parameters are presented in Additional file 5: Table S1. All statistical comparisons of *Tug1*^{-/-} versus wild type for relative testis size, sperm morphology, and sperm counts were performed using R (Wilcoxon rank-sum test and principal component analysis (PCA)).

***lacZ* and histological staining of male reproductive tissues**

Expression of the knock-in *lacZ* reporter and histological staining for morphological analysis of male reproductive tissues was conducted on the testes and epididymis from *Tug1*^{-/-} (*n* = 2) and wild-type (*n* = 2) mice. We fixed the testis and epididymis in 4% paraformaldehyde in PBS overnight at 4 °C and washed the tissues three times in PBS. For *lacZ* staining, we rinsed *Tug1*^{+/-} and wild-type tissues three times at room temperature in PBS with 2 mM MgCl₂, 0.01% deoxycholic acid, and 0.02% NP-40. We performed X-gal staining by incubating the tissues for up to 16 h at 37 °C in the same buffer supplemented with 5 mM potassium ferrocyanide and 1 mg/mL X-gal. The staining reaction was stopped by washing three times in PBS at room temperature, followed by 2 h post-fixation in 4% paraformaldehyde at 4 °C.

We then embedded the organs in paraffin, sectioned the organs at 6 μm thickness, and then mounted the sectioned samples onto glass microscope slides. The testis sections were additionally stained with Mayer's hematoxylin, periodic acid, and Schiff's reagent (VWR, 470302-348), and the epididymis sections were stained with eosin (VWR, 95057-848). Images were collected using a Zeiss AxioImager.A1 upright microscope or on an Axio Scan Z.1 (Zeiss).

RNA isolation and RNA-seq library preparation

We isolated total RNA from mouse tissues, mouse embryonic fibroblasts (MEFs), and blood cells using TRIzol (Invitrogen, 15596026) by chloroform extraction followed by spin-column purification (RNeasy mini or micro kit, Qiagen) according to the manufacturer's instructions. RNA concentration and purity were determined using a Nanodrop. We assessed RNA integrity on a Bioanalyzer (Agilent) using the RNA 6000 chip. High-quality RNA samples (RNA integrity number ≥ 8) were used for library preparation. We then constructed mRNA-seq libraries using the TruSeq RNA Sample Preparation Kit (Illumina) as previously described [82]. The libraries were prepared using 500 ng of total RNA as input and a 10-cycle PCR enrichment to minimize PCR artifacts. Prior to sequencing, we ran libraries on a Bioanalyzer DNA7500 chip to assess purity, fragment size, and concentration. Libraries free of adapter dimers and with a peak region area (220–500 bp) $\geq 80\%$ of the total area were sequenced. We then sequenced individually barcoded samples in pools of 6, each pool including *Tug1* mutant and wild-type samples, on the Illumina HiSeq platform using the rapid-full flow cell with the 101-bp paired-end reads sequencing protocol (Bauer Core, Harvard University FAS Center for System Biology).

RNA-seq and gene set enrichment analyses

We mapped sequencing reads to the reference mouse genome (GRCm38) by STAR [83] with the gene annotation obtained from GENCODE (vM16). We counted uniquely mapped reads for genes by featureCounts [84] and calculated transcripts per million (TPM) for genes to quantify the gene expression level after normalization of sequencing depth and gene length. Clustering of gene expression between tissues was done with Ward's method using Jensen-Shannon divergence between tissues as the distance metric. The R package, Philentropy, was used for calculation of the Jensen-Shannon divergence [85].

We identified differentially expressed genes by comparing the mean read counts of biological replicates between the groups (wild-type vs. *Tug1*^{-/-} and *Tug1*^{-/-} vs. *Tug1*^{rescue}) using the generalized linear model. Statistical significance was calculated with the assumption of the negative binomial distribution of the read counts and with the empirical estimation of variance by using the R packages DESeq2 [86] and fdrtool [87]. The genes were filtered if their read counts were less than three in every biological replicate. The genes were called significant if their adjusted *p* values by the false discovery rate (FDR) method were smaller than 0.05.

We performed gene set enrichment analysis (GSEA) to evaluate the enrichment of the gene sets available from MSigDB [88] after mapping genes to gene sets by gene symbols. The statistical significance of a gene set was calculated with the test statistics of individual genes computed by DESeq2. If the FDR-adjusted *p* value is less than 0.1, the term was called significant. We performed the GSEA analysis using the R package, CAMERA [89].

Allele-specific gene expression analysis

We performed allele-specific expression analysis as previously described [90]. For mouse testis samples, we created a *C57BL/6J*, *Cast/EiJ* diploid genome by incorporating single nucleotide polymorphisms and indels (obtained from the Mouse Genome Project: ftp://ftp-mouse.sanger.ac.uk/REL-1303-SNPs_Indels-GRCm38) from both strains into the *M. musculus* GRCm38 reference genome sequence. We created a transcriptome annotation set as follows. The gencode.vM2.annotation GTF file was downloaded, and Mt_rRNA, Mt_tRNA, miRNA, rRNA, snRNA, snoRNA, Mt_tRNA_pseudogene, tRNA_pseudogene, snoRNA_pseudogene, snRNA_pseudogene, scRNA_pseudogene, rRNA_pseudogene, and miRNA_pseudogene were removed (not enriched in our RNA-seq libraries). To create an extensive set of transcripts, we added to the gencode.vM2.annotation all transcripts from the UCSC knownGene mm10 annotation file, which are not represented in the gencode.vM2.annotation set. We also added all functional RNAs from the Functional RNA database (fRNAdb) [91], which did not intersect with any of the previously incorporated transcripts. From this, we then used the UCSC liftOver utility to generate a *C57BL/6J*, *Cast/EiJ* diploid transcriptome set.

Each RNA-seq library was first subjected to quality and adapter trimming using the Trim Galore utility (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore) with stringency level 3. We then mapped each of the *C57BL/6J*::*Cast/EiJ* hybrid RNA-seq libraries to the *C57BL/6J* and *Cast/EiJ* diploid genome and

transcriptome splice junctions using STAR RNA-seq aligner [83], allowing a maximum of 3 mismatches. The data were mapped twice, where after the first mapping step, we incorporated valid splice junctions that were reported by STAR to exist in the RNA-seq data. We then transformed the genomic alignments to transcriptomic alignments. Following that, we estimated the expression levels with their respective uncertainties for each transcript in our *C57BL/6J* and *Cast/EiJ* diploid transcriptome using MMSEQ [92]. The posterior FPKM samples were transformed into TPM units with a minimum expression TPM cutoff set to 0.01. In any RNA-seq sample, any transcript for which its MMSEQ posterior median TPM was lower than 0.01 was set to 0.01 (used as the minimal measurable expression level).

We adopted the approach of Turro et al. for combining lowly identifiable transcripts based on the posterior correlation of their expression-level estimates, tailored for a diploid transcriptome case [93]. In this approach, for any given RNA-seq sample, we compute the Pearson correlation coefficient of the posterior TPM samples of any pair of transcripts from the same locus and the same allele. Subsequently, if the mean Pearson correlation coefficient across all RNA-seq samples for a pair of transcripts in both alleles is lower than a defined cutoff (which we empirically set to -0.25), each of these pairs is combined into a single transcript. This process continues iteratively until no pair of transcripts (or pairs of already combined transcripts) can be further combined. This consistency between the alleles in the combining process ensures that the resulting combined transcripts are identical for the two alleles and can therefore be tested for allelically biased expression.

Amplification of full-length *Tug1*

We amplified the full-length *Tug1* isoform lacking the 5' region (Ensembl ID: ENSMUST00000153313.2) from Riken cDNA clone E330021M17 (Source Bioscience) using specific primers containing MluI and EcoRV restriction sites (see the “[Sequences and primers](#)” section). After gel purification, the amplicon was sub-cloned, using the MluI and EcoRV restriction sites, into a modified Tet-On pTRE2pur vector (Clontech, 631013) in which the bGlobin-intron was removed. We verified the absence of mutations from the cloned *Tug1* cDNA by sequencing using primers listed below. The plasmid was used also for sub-cloning *Tug1* into pcDNA3.1(+) (see below).

RT-PCR

The testes were collected from wild-type and *Tug1*^{-/-} mice (112 days old) and were homogenized in TRIzol (Invitrogen, 15596026) with a gentle MACS Dissociator (Miltenyi Biotec, 130-093-235). RNA was isolated by column purification (Qiagen, 74104) using a QiaCube (Qiagen). RNA was assessed and quantified on a Bioanalyzer (Agilent). cDNA was synthesized using 500 ng of total RNA as input with SuperScript IV VILO Master Mix with and without RT (Invitrogen, 11756050). PCR was performed on the synthesized cDNA (RT and no RT samples) using MyTaq Red Mix (Bioline, BIO-25043) with the cycling parameters: 95 °C for 1 min, followed by 35 cycles of 95 °C for 15 s, 60 °C for 15 s, and 72 °C for 10 s. PCR products were run a 1% agarose gel and verified by Sanger sequencing (GeneWiz). The following are the primers used for RT PCR: 7SK F1: GACATCTGTACCCCATTTGA and R1: TCCTCTATTCGGGGAAGGTC; *Tug1*

F1: CGGAGGAGCCATCTTGTCTTGTC and R1: GCTTCCAATTCATACACACACAC TG; Tug1 F2: CTCTGGAGGTGGACGTTTTGT and R2: GTGAGTCGTGTCTC TCTTTTCTC.

ORF search

We analyzed human and mouse *Tug1* cDNA sequences with CLC Genomics Workbench (Qiagen) for open reading frames (ORFs), allowing both canonical and non-canonical start codons (AUG, CUG, and UUG). After, sequences with annotated ORFs were aligned using MUSCLE alignment. All further sequence and amino acid alignments were performed with CLC Genomics Workbench.

Generation of human and mouse TUG1-BOAT overexpression constructs

We generated a synthesized construct for human *Tug1* ORF1 that contained an in-frame 3xFLAG epitope tag prior to the stop codon, with and without the 5' leader sequence (GeneWiz). We also synthesized a construct containing mouse ORF1 with an HA tag after the 3xFLAG before the stop codon, with and without the 5' leader sequence (GeneWiz).

We amplified the *Tug1* cDNA sequence with primers (see the “[Sequences and primers](#)” section) having KpnI and NotI restriction enzyme overhangs from the pTRE2-*Tug1* vector plasmid using Q5 polymerase (Roche) and under the following conditions: 96 °C for 2 min, 35 cycles of (96 °C for 30 s, 65 °C for 30 s, 72 °C for 4 min), 72 °C for 4 min, and gel purified the amplicon. We digested the inserts and pcDNA3.1(+) plasmid with proper restriction enzymes according to the manufacturer's instructions. After digestion, the plasmid was dephosphorylated using alkaline phosphatase. We then ligated the plasmid and inserts using T4 ligase (NEB) in a 1:3 ratio, respectively, followed by bacterial transformation, culture growth, and plasmid isolation (Qiagen Mini-Prep Kit).

Transfection of TUG1-BOAT constructs

We seeded 3T3 and HeLa cells in 10-cm plates containing poly-L-lysine-coated 18-mm glass coverslips. Next, we transfected the cells with 14 µg of plasmid (pcDNA3.1(+) containing each of the inserts) using Lipofectamine 3000 Transfection Reagent (Thermo Fisher Scientific) per manufacturer's recommendations. Forty-eight hours post-transfection, cell pellets were harvested for protein extraction (see below) and coverslips were processed for RNA FISH and/or immunofluorescence (see below).

Protein extraction and western blot

We resuspended 3T3 and HeLa cell pellets in RIPA Lysis and Extraction Buffer 48 h post-transfection (Thermo Fisher Scientific). Total protein was quantified with Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). We then separated a total of 20–25 µg of denatured protein on a 12.5% SDS polyacrylamide gel for 100 min at 120 V. We transferred proteins to an Immobilon-PSQ PVDF membrane (Sigma-Aldrich, ISEQ00010) at 400 mA for 75 min. After blocking in 5% dried milk in TBST, the membrane was incubated with properly diluted primary antibody (M2 Monoclonal ANTI-FLAG 1:1000, F1804, Sigma; Monoclonal GAPDH 1:5000, 2118S, CST) in 5% dried milk/TBST overnight at 4 °C. The next day, we washed

the membrane three times for 5 min each in TBST (0.5% Tween-20). We then incubated the membrane with horse radish peroxidase-conjugated secondary antibody (anti-mouse 1:15,000, A9044, Sigma; anti-rabbit 1:10,000, 711035152, Jackson ImmunoResearch), diluted in 5% dried milk/TBST for 1 h at room temperature. Following three 5-min washes in TBST, SuperSignal™ West Pico PLUS chemiluminescent substrate (Thermo Scientific, 34580) was added and chemiluminescence was detected using ImageQuant™ LAS 4000 imager. Original, uncropped images of western blots are provided in Additional file 16: Fig. S11.

TUG1-BOAT localization by immunofluorescence

We plated HeLa and 3T3 cells on poly-L-lysine-coated coverslips. Forty-eight hours post-transfection, we rinsed the coverslips twice with PBS and fixed the cells with 3.7% formaldehyde in PBS for 10 min at room temperature. After 2 washes with PBS, we permeabilized the cells with PBT (PBS, 0.1% Tween-20) for 15 min at room temperature. Next, we blocked the coverslips with 5% BSA in PBT for 1 h at room temperature and then incubated the coverslips with properly diluted primary antibody (mouse M2 monoclonal ANTI FLAG, 1:800, F1804, Sigma; rabbit polyclonal Tom20, 1:800, FL-145, Santa Cruz) in 5% BSA in PBT for 3 h at 37 °C in a humid chamber. Coverslips were washed three times for 5 min each with PBT and incubated with diluted secondary antibody (anti-mouse labeled with Alexa Fluor 488, 1:800, ab150113, Abcam; anti-rabbit labeled with Alexa Fluor 647, 1:800, 4414S, CST) in 5% BSA in PBT for 1 h at room temperature. Cells were then washed twice for 5 min with PBS, once for 20 min with PBS containing Hoechst DNA stain (1:1000, Thermo Fisher Scientific), rinsed in PBS, and then mounted on glass slides with ProLong Gold (Thermo Fisher Scientific).

Mitochondrial staining with MitoTracker® red chloromethyl-X-rosamine

We plated cells on poly-L-lysine-coated coverslips and transfected as described in the previous sections. Forty-eight hours post-transfection, cells were incubated with 200 nM MitoTracker red chloromethyl-X-rosamine (Thermo Fischer Scientific, M7512) in 1 mL FBS-free growth media for 40 min. We then washed the cells twice with PBS, fixed with 3.7% formaldehyde for 10 min at room temperature, and processed for immunofluorescence and/or RNA FISH (as described previously).

Microscopy and image analysis

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

Fluorescence-activated cell sorting

Age- and sex-matched adult mice were used in all flow cytometry experiments. We obtained peripheral blood by cardiac puncture and collected blood into a 1.5-mL Eppendorf tube containing 4% citrate solution. Next, we added the blood-citrate mixture to 3 mL of 2% dextran/1× PBS solution and incubated for 30 min at 37 °C. The upper layer was transferred to a new 5-mL polystyrene FACS tube (Falcon, #352058) and centrifuged at 1200 rpm for 5 min at 4 °C. We then lysed red blood cells for 15 min at room temperature using BD Pharm Lyse (BD, 555899). Cells were washed twice with staining media (Hank's Balanced Salt Solution (HBSS) containing 2% FBS and 2 mM EDTA). The following antibodies were added (1:100) to each sample and incubated for 30 min at room temperature: Alexa Fluor 700 anti-mouse CD8a (Biolegend, 100730), PE/Dazzle-594 anti-mouse CD4 (Biolegend, 100456), APC anti-mouse CD19 (Biolegend, 115512), Alexa Fluor 488 anti-mouse NK-1.1 (Biolegend, 108718), and PE anti-mouse CD3 (Biolegend, 100205), and Zombie Aqua Fixable Viability Kit (Biolegend, 423101) was used as a live-dead stain. We washed samples twice with staining media and sorted directly into TRIzol LS using a BD Aria FACS.

qRT-PCR

We isolated and quantified RNA from sorted blood populations as described in the RNA Isolation and RNA-Seq Library Preparation. One hundred nanograms of total RNA was used as input to generate cDNA using SuperScript IV VILO Master Mix (Invitrogen, 11756050), according to the manufacturer's protocol. cDNA was diluted 1:3 with DNase- and RNase-free water, and 1 µL was used per each reaction. We performed qRT-PCR using FastStart Universal SYBR Green Master Mix with ROX (Sigma, 4913914001) on a ViiA 7 Real-Time PCR System (Thermo Fisher). Analysis was performed using the $\Delta\Delta C_t$ method [94]. Primers used in qRT-PCR experiments are listed in the “Sequences and primers” section.

GEO accession numbers

All primary RNA-seq data are available at the Gene Expression Omnibus (GSE124745 and GSE88819).

Sequences and primers

In situ hybridization riboprobe—mouse *Tug1* (492 bp)

GAGACACGACUCACCAAGCACUGCCACCAGCACUGUCACUGGGAACUUGAA
 GAUCCAAGUUUCUGUCCAGAACCUCAGUGCAAACUGACAACACUCCAUCCA
 AAGUGAACUACGUCCCGUGCCUCCUGAUUGCUGAAUGUUCACCUGGACCUG
 CCAAUGACCUUCCUUCUGCUACUCCAUCAGCCUACAGACCUGGUACUUGGA
 UUUUUGUCCAUGGUGAUUCCUACCUCUACUACUGAAGAAGACACCAUUC-
 CAGUGGACCACUGUGACCCAAGAAGCAUUCAGCCAUCAUGAUGUGGCCUUU
 ACCUCCACUCCUGUCCUACUCUGCCCAGAUUCAGCACAGCCCUUUAUAGUG-
 CAGUCAAGAGUCUUAAGCCAAAUAACUGAAGCUAUUUUAUCACAACAAAG
 GCCAGGUUUAUCCAUAUUAUGUACAGUUAUUCUGCAGUUUAUUCUUCAG
 AGACACAUAGUAAAUUUGGACCAGGGGAUUUUG

Genotyping primers***Tug1*^{tm1.1Vlg}-knockout mice and MEFs**

Tug1_5190-5166TD_Forward: TGACTGGCCCAGAAAGTTGTAAG

Tug1_5190-5166TD_Reverse: GCAAGCAGGTCTGTGAGACTATTC

lacZ_5_Forward: TTGAAAATGGTCTGCTGCTG

lacZ_5_Reverse: TATTGGCTTCATCCACCACA

Ychr_Forward: TCTTAAACTCTGAAGAAGAGAC

Ychr_Reverse: GTCTTGCCTGTATGTGATGG

Mouse *Tug1* (ENSMUST00000153313.2) cDNA clone

TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGGGGGTTTTTTGTTTTGTTTTTTT
 AAATTGAAGGCTAAAGTTTTTGAAAAAACTTTGTTGGACTCTGGCTGGGACA
 CAAAATCAGATATTTGGAATCATTTTGAAGCTTAACTTTTTCCTAACCCAGCC
 TTGTATTCTAATTGCTTGCAAATGTGAGACTGAATGGCCAAAATGC
 CGTTTGTGTTGTTGTTTATTGTCAGCTGCTTTTATCAAATTCAGGCCATT
 ATCCAGCAAACACTATTA AAAATGTTTGAACAGTTGGGTTTCAAACATTTTTG
 TTTTGTGGAGTGGTGTCTTATTAAGTGGTACAGCTCTCTAAGCAAGTGAACAC
 AAACATATTTAAGTGTATTTTGTATGATTAGATGTTACCAATTCTGATATTT
 TATTCAAATGTCTAAAAAAATAAGTTGACTTATCCCTTTACCAAAGGGCCA
 GAGACAAATGGTTTCCTTTCAAGAGAAATGACTGTTTTGAAGAAAAACTCTG
 TTGGTCTTAGCTCTTTTGTAATTA AATCTGGATGTACCTCAAAGACTCTTTA
 AGACTGTGGTGTTAAAAGGCTTTCCTCTGGAGAAGGAGAAAAAATAAAAATCA
 ACTGGAACCTAAAAGCTTGAAATTCCATGACAAAACACAGATGTCCAGGATT
 GGAGGTTCATAAAGTACATGCAGTAGTTGGAGTGGATTCCATTTTCAGTGTA
 GCTGCCACCATGGACTCCAGGCTCCAGATTTTCAAGA ACTGGACCTGTGAC
 CCAGAAGAGCTTGTCAAGATATGACAGGAACTCTGGAGGTGGACGTTTTGTA
 TTCAATTTTGGA ACTGTTGATCTTGCCGTGAGAAAAGAGAGACACGACTCAC
 CAAGCACTGCCACCAGCACTGTCACTGGGAACTTGAAGATCCAAGTTTCTGT
 CCAGAACCTCAGTGCAA ACTGACAACACTCCATCCAAAGTGA ACTACGTCCC
 GTGCCTCCTGATTGCTGAATGTTACCTGGACCTGCCAATGACCTTCCTTCTG
 CTACTCCATCAGCCTACAGACCTGGTACTTGGATTTTTGTCCATGGTGATTCC
 TTCCACCTTACTACTGAAGAAGACACCATTCCAGTGGACCACTGTGACCCAA
 GAAGCATT CAGCCATCATGATGTGGCCTTACCTCCACTCCTGTCTACTCTG
 CCCAGATT CAGCACAGCCCTTTATAGTGCAGTCAAGAGTCTTCAAGCCAAA
 TAACTGAAGCTATTTTATCACAACAAAGGCCAGGTTTATTCATAAATGTAC
 AGTTCATTTCTGCAGTTTATCTTTCAGAGACACATAGTAAATTTGGACCAGG
 GGATTTTGTGTTTGTATATTTGTCAACACTGTCTGAAGAAAGGCATCTCTGA
 GAACAGCATTGGACCCTACTCCACAATCTCAAATGATTGAAGTTTCATAAAC
 TGCCTAGGATCCTGTCAAGGCCACTGGACTCTTGTTCTTTTCCTACTTCAA
 ATCTGTAGCTGTCTACTAAATGACAAAGCAGATATTCTGACCCATTGGGATC
 AAAACCAAGGCATTTTGAATTCCTCATAGTATCATCTTCGGGTTACTCAGGA
 ACCAAA ACTTTTCACACCAATTTAAGAAATTCTACTGAGGAATCCCTTTACCT
 AACCATCTCACAAGGCTTCAACCAGATTCCTGAAAAGGCCTCTTGATATATC
 AAGATAGAACCTACATGCATTTTGTGAACAACTTATCACTGATTTTCCAAAG
 GCTTTGTGCTCTTGAAGTCTTTTGAAGGAAAGCTGTGTGGAAGTCCAGAGT

AAAGTGAAGCTGCTCTGGATGAAGTAGTGAAGTGGGAGTTGAGGTCTACAA
CCTGCCACAACCATCTTCCTTACCACCATGGTGATGCCAAAAGGGACTTCC
TTAAAGCTCTTCAGAAAATCCTGCTTGAAACCACTACCCTAGACAACATGTT
TGACCTGGATGGCATTCTCTTCAAAACAATTCATATTCAGTTGATGCTCAA
CATGTTTGGAGATGCTTTATTCAGAGAATGATGATAATTACAGCATTGTCTA
ATGAAGTTTTATTAATAGCATTCCATCCAAGGTGGACTTCCTGGAGTTGGAT
ATAACCAGAGAGCAATTCATATGTATCCTACACTGAAGAACACCATTAACTT
TCAGCAACCTATAGCTAGTGGTACTAGAAGTACGTGTCTTGGAAGTCTATGA
GAGCTGGTATTGAAGCTGATGCCTCCTTAAGGCCATCTTAGACCAAGTTGTT
TGTTTGACCTCTCCTCATTAACTATGGAGCAGAATTGAAATACAAATTTTTCC
TAAAGGGACTTGCAACCTGGTTATCATTATTATCTCAAGTTTCAAGTCATG
TTGATGCAACCAGTAGTTATTAAGTCTCCATGGTTTTTTGTTATTTAATAC
TTTTTCCAGGGCTTAAAAAACAATAAATTTCTCCAACACGTCTATACTT
GTCTGTTCAAAAGTAACTACTCACCCTATATGGAACAGATGATTCTGAAGA
CACTCTGAGCATCCTTTATGATATTTGTGACTTAAAATGTGGCTGGAAATTT
TCCTTCTACCCAGTGAAATATTTAATGATTAGTCTTCATGCCTGATACCATCA
ACTGTATATGCGTGATAGGCAAAGTTTGACATAGGCATTTGACTCTAGGCTA
TGATAGCTTGCTAGTAACTTCAAGTAGCATATTGTCAACCTGTTTGCTGGAA
AAGTAGAGTAACTTGGAATAAACTAAATGGCAGCTAAGGATTTTTTTCAG
TATTCCTGAGTTTCTGTCTTGGGATATTTCAATGAAATTTTCACCTGTCTCT
TCACTTAACAGAGTGACTGACTCCTTACTATGAAGTATTCTTAAGACATTA
GATTACTTTTTGTAGAAAGGATAAAATTCCTGACCATCCAAATCATCATAGTG
AACAAGACTTCAATTTGTGACCTGAGAAAATCTCATTTCTCTACTTCGTAGT
CAATGTAAGGGCCAATGCTATCAGCTACTCTGAGTGCCTGGGTAAACGTT
GGAAGTGCCTTCTTTATATCATTACTTTTTATCCTCTAAATTAATCATGGTTA
TGTAATTCCTCGCCACAAATCAGCAAAATCAGACTCAGATCTGGTTATTCTAGA
CTGCTCACAGTTAACAATCAAATCTGGATGACTTCTGCTTGTATATGCAA
CTACTATTTGTAAAGAAATGCAAAATCACTTTTCTATTACCTCTACATTGCT
AGCTCTTTCTTTTGTGTTTGTATTAATAAACAATAAAGCTACACTGCCAGCT
ATTCCTCCTGCCATACTCAGTTAAATGAAGAATCGGGAATCTAACAGTG
AATGGATAAGTAGAAAAACTAAACTTAAGGCAAAAAGCCTTAATCTAGGGC
CTTTTCTACTATCTTCATGTCTTGGATTCATCTAAAATCAACAGTGCCACCC
AACCAGTCTGAGGTCTTGACTTGCTTTTAAGATGATTCTTAGAGATGGGCT
GTATTACAGAAGGTGAAGACTTGATTACCAAAGAAAGTAGAGCCAACCTTG
ACAAACCTGGCTCTACAATCCTATTGCTTCCAGATGTAGCATAGACTCATAA
CTAGAACCTCAAGTCTGCATTGAGGATATAGCCTTCTAAGCTGACAGTTCTT
GCAACAGGTGAGCAAGAAAATGAAAGCTGTTATACCCAAGTGGCCCTTTAAG
ATCCAAAATAATGTCTGGACTAAACCCTATGGAGTACCCAGGACAAAAACT
AATTTACAGAGCTTCATTATTAATCTGCCTGTTCTTCTAGCTTAATTATTGGT
ATGGCTGGCCCTACTGAAGTAGTTTGTCTGTTTACCTGTCTTCAGCTCTTAAC
CTGGCTATTTTGACATGCTACTGCAATTAGACTAACTGGCTTTGAGAAGACT
ACAATCAGTTTTCAGCCTCTCCTTTGCCAATTTACCAAGGAATTTTGATAA
GAGGAACCCATACCTCACCCACCAGAACAGAAAGGACCATGCTGCATATT
CCTTGACCAGCAACTTTAAGTAGAGAACAACCCTGCTTGTTTTCAACATCTG
AAACACCATTTGATCTAATAGGAGTATAGAAGGTTGACAGCAGAGTACACTA
CTTACTTCTTTCATAACTCAGAAATGAATATGACTGGCCAGAAAGTTGTAA

GTTCACCTTGACAAGAAACAGCAACACCAGAAGTTTACTGCTGAACTTAACT
 TGCCACTTACTCGAATAGTCTCACAGACCTGCTTGCCAAGTAGGAGGCTAG
 TTTTCTGCTTCATATCACCATTGGAGTGGGGCTCAATGGGGTCAATGTTAA
 TACTGACTTGAATGGGGACCTTATGGTGAATCCTAGACTATGAGGCTAATGG
 AAATTATTGTCTATTCAAGTGGATTATAGATTTCTGAGGACAGAACAGAC
 ATCACTCCTGGTGATTTTTAGAACTTGATTACCAAGGAAGAAATACCAGCTG
 CTAACAGTCAACTTCATGGGCAAAGATTAAGCTCTCTATATCTGGTCGTATC
 CTGGATGCTAGTTTTTTATTGCCAGTGACCATTTCCATCTCACGCTTAACT
 TCCTGATGTTTTTTGGAACCATCTCTTCCAATTTTCAGTCCTGGTGATTTAGA
 CAGTCTTTTCATGCTGGACATTTTGTGCAACCTCATCAATCACAGCAAAGT
 CCATCTTGACTTTAGTGATTAGTTTCAAGGAATGGATGCATGATTCAAGTTTGT
 CCAATGATAATCAACCCTAGGTGTTTTCTCAGTTGTGGAGAAGTTCTCTTAG
 ATGCTTTAGCTTTGTAGGAGAAAACCTCAAACCAACAGGGCCTACCTACTATG
 TTGAATGATTGTAGGAGAAAACCTCAAACCAACAGGGCCTACCTACTATGTTG
 AATGAGCCAGGCAGAAAATGAAGCCAGTACAGAGGGAAAATGGAGCCAAA
 GAGGAAGAGACTTGAGTTCTGATGATCACATTTATGCCCTGTATCCAACCTG
 TGCTGAAGCTAATAGTACATCACCTGGACTTTTCAGTTATGTGAACCAATA
 AATTCCCCTTTTTGTTTAAGTTACTTTGAGTT

Full-length *Tug1* primers for cloning in pTRE2pur

Tug1_MluI_cDNA_Fwd gagaacgcgtTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGG
 GGGTTTTTTGTTTTGTTTTTTAAATTGAAGGCTAAAGTTTTTGA AAAA ACTTT
 GTTGGACTCTGGCTGG

Tug1_EcoRV_cDNA_RevgagatatcAACTCAAAGTAACTTAAACAAAAAGGG

Primers for full-length sequencing of pTRE2-*Tug1* expression vector

LNCXAGCTCGTTTAGTGAACCGTCAGATC
 TUG1_76TTTAAATTGAAGGCTAAAGTTTTTTGAA
 TUG1_266GGCCATTATCCAGCAAACAC
 TUG1_755ACTCCAGGCTCCCAGATTTT
 TUG1_1254TCTTCAAGCCAAATAACTGAAGC
 TUG1_1741AGAACCTACATGCATTTTGTGAA
 TUG1_2268ATGCCTCCTTAAGGCCATCT
 TUG1_2754TGTCAACCTGTTTGCTGGAA
 TUG1_3267TTGCAAATTCACTTTTCTATTACCTC
 TUG1_3746CCCAACTGGCCCTTTAAGAT
 TUG1_4241TGACAAGAAACAGCAACACCA
 TUG1_4740TCACAGCAAAGTCCATCTTGA

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

Tug1_Tg F/KpnIataggtaccGCCCCGAATTCACGCGTT
 Tug1_Tg R/NotIatagcgccgcACCTGAGGAGTGAAGA

Primers for qRT-PCR

Tug1_Fwd CTCTGGAGGTGGACGTTTTGT
 Tug1_Rev GTGAGTTCGTGTCTCTCTTTTCTC
 Gapdh_Fwd GGTGAAGGTCGGTGTGAACG
 Gapdh_Rev CTCGCTCCTGGAAGATGGTG

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

CTGGCGGCCCTCCCCCTCCCGGTCTGGTAGGGCGAAGGAACGGGCGTGCGGTTCGATCGAGCGATCGGTTG
 GCGGCTCTTTCTCCTGCTCTGGCATCCAGCTCTTGGGGCGCAGGCCCGGCCCGCGCGCGCCCGGTGGCC
 GTTGGCGCTCGCGCCGCGTCTTTCTTCTCGTACGCAGAACTCGGGCGCGGCCATGCGTTTGCATTTCACGA
 GGAGTCGTCCGGTGGTTCGGCGCGCGGGCAGCTGCTCCGCCCGCTCCGGGGAGGCGCGCGCGCCAGCGGC
 CGCGGATTTGGAGCGCGCGGGGAGGCGGGGTGGCCGGGCGCGCTTGGAGGCTGGCGCCACCCTTCGGGGC
 CTGCAAGGACCCAGTTGGGGGGCAGGAGGGGGCGGAGGATGGTTGGTTGTGGGATTTCTACTTTGCCTTTTC
 CTCTTATGCGCCTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGA
 TGACAAGTAG

Human ORF1+UTR:

GGCCGAGCGACGCAGCCGGTAGCTGCGGTGCGGACCGGAGGAGCCATCTTGTCTCGTCGCGGGGAGTC
 AGGCCCTAAATCGAAGAAGCCCTGGCGGCCCTCCCCCTCCCGGTCTGGTAGGGCGAAGGAACGGGCGTG
 CGGTCGATCGAGCGATCGGTTGGCGCTCTTTCTCCTGCTCTGGCATCCAGCTCTTGGGGCGCAGGCCCGCCG
 CCGCGCGCGCGCCCGGTGGCCGTGGCGCTCGCGCCGCGTCTTCTTCTCGTACGCAGAACTCGGGCGCGGC
 CTATGCTTTGCGATTCGACGAGGAGTCCGCGGTGGTGGCGCGCGCGGGCAGCTGCTCCGCCCGCTCCGG
 GGGAGCGCGCGCGCGAGCGGCCCGGGATTTGGAGCGCGGGGAGGCGGGGTGGCCGGGGCGCGCTTGGAG
 GCCTGGCGCCACCCTTCGGGGCTGCAAGACCCAGTTGGGGGGCAGGAGGGGGCGCGGAGGATGGTTGGTTGT
 GGGATTTCTACTTTGCCTTTCTCTTATGCGCCTGACTACAAAGACCATGACGGTGATTATAAAGATCATG
 ACATCGACTACAAGGATGACGATGACAAGTAG

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

CTGGCGGCCCTCCCCCTCCCGGTCTGGTAGGGCGAAGGAGCGGGCGTGCGGTTCGATCGAGCGATCGGTTG
 GCGGCTCTTTCTCCTGCTCTGGCATCCAGCTCTTGGGGCGCAGGCCCGGCCCGCGCGCGCCCGGTGGCC
 GTTGGCGCTCGCGCCGCGTCTTTCTTCTCGTACGCAGAACTCGGGCGCGGCCATGCGTTTGCATTTCACGA
 GGGGTCTCCGGTGGTGGCGCGCGGGCAACTCCGCCCGCTCCCGGGAGGCGCGGGGAAGCTGGGGT
 GGCCGGGGCTGGCTGGAGGCTGGCGCCACCCTCGGGGCTGCTAGGACCCAGTTGGAGGGTCAAGAGGGAG
 CTGGAGGATGGTTGGTGGTGGGCTTCTCTTTGCTTTTCTACTTATGCCACCTGACTACAAAGACCATGAC
 GGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATGACAAGTACCCATACGATGTTCCAGATTACGC
 TTAG

Mouse ORF1+UTR:

GGCCGAGAGACGCAGCCGGTAGCTGACAGAGCAGCGGAGGAGCCATCTTGTCTTGTGCGGGGAGTC
 AGGCCCTAACTCGAAGAAGCCCTGGCGGCCCTCCCCCTCCCGGTCTGGTAGGGCGAAGGAGCGGGCGTG
 CGGTCGATCGAGCGATCGGTTGGCGCTCTTTCTCCTGCTCTGGCATCCAGCTCTTGGGGCGCAGGCCCGCCG
 CCGCGCGCGCGCCCGGTGGCCGTGGCGCTCGCGCCGCGTCTTCTTCTCGTACGCAGAACTCGGGCGCGGC
 CTATGCTTTTGCATCCGACGAGGGTTCGTCGGGTGGTGGCGCGCGGGCAACTCCGCCCGCTCCCGGGG
 AGGCGCGGGGAAGCTGGGGTGGCCGGGCTGGCCTGGAGGCTGGCGCCACCCTCGGGGCTGCTAGGACC
 CAGTTGGAGGTCAGAGGGAGCTGGAGGATGGTTGGTGGTGGGCTTCTCTTTGCTTTTCTACTTATGCC
 ACCTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATGACAAGTACC
 CATACGATGTTCCAGATTACGCTTAG

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13059-020-02081-5>.

Additional file 1: Fig. 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.

Additional file 2: Fig. 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.

Additional file 3: Fig. 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 promoter and first exon intact. The deleted region is indicated by red dashed lines. The open reading frame (ORF) encoding the TUG1-BOAT protein and PhyloCSF scores for the (-2) frame across the locus are depicted (grey). Chromosomal coordinates (mm10) are shown.

Additional file 4: Fig. 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.

Additional file 5: Table S1. *Tug1*^{-/-} and wild type sperm morphological defects.

Additional file 6: Table S2. Testes RNA-seq and *Tug1*^{rescue} RNA-seq in testes.

Additional file 7: Table S3. Prostate, spleen, eyes, brain, heart, liver, and MEF RNA-seq.

Additional file 8: Table S4. Allele-specific RNA-seq in testes, expression data.

Additional file 9: Table S5. Allele-specific RNA-seq in testes, differentially expressed genes.

Additional file 10: Fig. S5. Gene expression of multiple tissues in *Tug1* WT and KO mice. Gene expression of multiple samples (columns) were described with log scale of TPM: Log₂(TPM+1). Genes (rows) are clustered by hierarchical clustering with Ward's method based on Euclidean distance. Annotation of samples are provided in the top panel in terms of genotypes and tissues.

Additional file 11: Fig. S6. *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*^{rescue} (*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.

Additional file 12: Fig. S7. Changes of gene expression in two comparisons of *Tug1*^{-/-} (KO) vs. WT and *Tug1*^{rescue} (Rescue) vs. *Tug1*^{-/-} (KO). Each dot represents a gene whose x-axis value is the fold change of gene expression between KO and WT and y-axis value shows the fold change in Rescue vs. KO. Color describes statistical significance of fold change (adjusted *p*-value < 0.05): no statistical significance in either comparison (gray, *N* = 33688); significance in KO vs. WT (blue, *N* = 998); significance in Rescue vs. KO (orange, *N* = 126); significance in both comparisons, KO vs. WT and Rescue vs. KO (red, *N* = 52).

Additional file 13: Fig. S8. The 5' region of human *TUG1* contains a conserved ORF. **(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 showing human ORF1 construct design. hORF1 (labeled with a 3xFLAG epitope tag prior the stop codon) with the 5'UTR was inserted into pcDNA3.1(+) and transfected into HeLa cells. 48 hours post-transfection, TUG1-BOAT-3xFLAG localization was analyzed by immunofluorescence (IF) (shown in C). **(C)** Maximum intensity projection of HeLa cells expressing human 5'UTR-hORF1-3xFLAG. Localization of 3xFLAG tagged TUG1-BOAT was assessed by immunostaining against the 3xFLAG (green). Nucleus was monitored by DAPI (blue) and mitochondria was monitored by immunostaining against mitochondrial membrane translocase TOM20 (gray). Bar plot shows localization analysis of TUG1-BOAT. Scale bar is 5 μm.

Additional file 14: Fig. S9. GFP over-expression does not compromise mitochondrial membrane potential. Maximum intensity projections of z-stacks acquired 48 h post-transfection of 3T3 cells with GFP cloned into pcDNA3.1(+) under CMV promoter and staining with Chloromethyl-X-rosamine (CMXR). GFP (green) was used as control. CMXR is shown in red, DAPI in blue. On the right, quantification of cells expressing GFP and mitochondria membrane potential by CMXR (*n* = 50). Scale bar is 5 μm.

Additional file 15: Fig. S10. Loss of TUG1 expression in infertile human males. Heatmap of microarray data from three different probe sets showing decreased expression of TUG1 in sperm from infertile teratozoospermic men (T) compared to fertile (normospermic) individuals (ND). In all cases, *p* < 4.39 × 10⁻⁴.

Additional file 16: Fig. S11. Uncropped western blot images from Figure 5e. Original, uncropped images of western blots shown in Figure 5e. **(A)** Western blot of constructs overexpressed in 3T3 cells targeting the 3xFLAG

tag (top). GAPDH is used as a loading control (bottom). **(B)** Western blot of constructs overexpressed in HeLa cells targeting the 3xFLAG tag (top). GAPDH is used as a loading control (bottom).

Additional file 17. Review history.

Acknowledgements

We thank Dr. Diana Sanchez-Gomez for the assistance in the mouse facility; Catherine MacGillivray and Diane Faria in the HSCRB histology core facility; Joyce LaVecchio and Nema Kheradmand in the HSCRB flow cytometry core; the Broad Institute of MIT and Harvard Genomics Platform and the Bauer Core Facility at Harvard University for RNA sequencing support; and the University of Colorado BioFrontiers Institute Computing Core, which provided high-powered computing facilities and services. We also thank Dr. James C. Lee and Dr. Philipp Maass for the manuscript feedback.

Peer review information

Tim Sands was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 17.

Authors' contributions

Conceptualization: C.G., M.S., H.E.H., and J.L.R. Methodology: M.S., E.J.P., J.L., G.D., T.H., and J.L.R. Computational analyses: T.H., A.F.G., and 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., and C.M. Investigation: M.S., E.J.P., S.C.L., C.G., J.L., G.D., A.W., N.C., C.K., K.T., and C.M. Mouse husbandry: M.S., N.C., and J.L. Transgene funding acquisition and initial generation: M.S., A.B., J.G.C., P.P.P., and J.L.R. Data curation: M.S., E.J.P., T.H., J.L., G.D., and A.W. Writing of the draft: M.S., J.L.R., G.D., J.L., C.M., and T.H. Figures: M.S., E.J.P., C.M., J.L., G.D., A.W., and T.H. Supervision, administration, and Funding acquisition: J.L.R. and H.E.H. All authors have read and approved the final manuscript.

Authors' information

Not applicable

Funding

This work was supported by funds from Blavatnik Accelerator Program, P01GM099117 (J.L.R.); seed funding from the Harvard Initiative for RNA Medicine (J.L.R., P.P.P.); National Institutes of Health (NIH) General Medical Sciences postdoctoral fellowship award 1F32GM122335-01A1 (J.P.L.); M.S. is a Research Scholar Junior 1 of the Fonds de Recherche Santé du Québec (FRQS); H.E.H. is an Investigator of the Howard Hughes Medical Institute and J.L.R. is the Leslie Orgel Professor of RNA science and HHMI Faculty Scholar.

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.

Ethics approval and consent to participate

Mice used in these studies were maintained in a pathogen-specific free facility under the care and supervision of Harvard University's Institutional Animal Care and Use Committee.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA. ²BioFrontiers Institute, University of Colorado at Boulder, Boulder, CO 80303, USA. ³Department of Biochemistry, University of Colorado at Boulder, Boulder, CO 80303, USA. ⁴Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, 4109 BioLabs, Cambridge, MA 02138, USA. ⁵Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. ⁶Department of Systems Biology, Harvard University, Cambridge, MA 02138, USA. ⁷Cancer Research Institute, Beth Israel Deaconess Cancer Center, Department of Medicine and Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA. ⁸Harvard Initiative for RNA Medicine, Boston, MA 02215, USA. ⁹Howard Hughes Medical Institute, Chevy Chase, MD, USA. ¹⁰Montreal Clinical Research Institute, Montreal, QC H2W 1R7, Canada. ¹¹Department of Biochemistry and Molecular Medicine, Université de Montréal, Montreal, QC H3C 3J7, Canada.

Received: 6 February 2019 Accepted: 22 June 2020

Published online: 07 September 2020

References

1. Sonenberg N, Wilchek M, Zamir A. Identification of a region in 23S rRNA located at the peptidyl transferase center. *Proc Natl Acad Sci.* 1975;72:4332–6. <https://doi.org/10.1073/pnas.72.11.4332>.

2. Feng J, Funk WD, Wang S-S, Weinrich SL, Avilion AA, Chiu C-P, et al. The RNA component of human telomerase. *Science*. 1995;269:1236–41. <https://doi.org/10.1093/nar/gky705>.
3. Bassett AR, Akhtar A, Barlow DP, Bird AP, Brockdorff N, Duboule D, et al. Considerations when investigating lncRNA function in vivo. *Elife*. 2014;3:1–14. <https://doi.org/10.7554/eLife.03058>.
4. Kopp F, Mendell JT. Review functional classification and experimental dissection of long noncoding RNAs. *Cell*. 2018;172:393–407. <https://doi.org/10.1016/j.cell.2018.01.011>.
5. Penny G, Kay G, Sheardown S, Rastan S, Brockdorff N. Requirement for Xist in X chromosome inactivation. *Nature*. 1996;379:131–7. <https://doi.org/10.1038/379131a0>.
6. Lee JT, Jaenisch R. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature*. 1997;386:275–9. <https://doi.org/10.1038/386275a0>.
7. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*. 2010;39:925–38. <https://doi.org/10.1016/j.molcel.2010.08.011>.
8. Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL, et al. A long noncoding RNA mediates both activation and repression of immune response genes. *Science*. 2013;341:789–92. <https://doi.org/10.1126/science.1240925>.
9. Elling R, Robinson EK, Shapleigh B, Liapis SC, Covarrubias S, Katzman S, et al. Genetic models reveal cis and trans immune-regulatory activities for lincRNA-Cox2. *Cell Rep*. 2018;25:1511–1524.e6. <https://doi.org/10.1016/j.celrep.2018.10.027>.
10. Paralkar VR, Taborda CC, Huang P, Yao Y, Kossenkov AV, Prasad R, et al. Unlinking an lncRNA from its associated cis element. *Mol Cell*. 2016;62:104–10. <https://doi.org/10.1016/j.molcel.2016.02.029>.
11. Groff AF, Barutcu AR, Lewandowski JP, Rinn JL. Enhancers in the Peril lincRNA locus regulate distant but not local genes. *Genome Biol*. 2018;19:1–14. <https://doi.org/10.1186/s13059-018-1589-8>.
12. Groff AF, Sanchez-Gomez DB, Soruco MML, Gerhardinger C, Barutcu AR, Li E, et al. In vivo characterization of Linc-p21 reveals functional cis-regulatory DNA elements. *Cell Rep*. 2015;16:2178–86. <https://doi.org/10.1016/j.celrep.2016.07.050>.
13. Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, et al. Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol*. 2013;9:59–64. <https://doi.org/10.1038/nchembio.1120>.
14. Housman G, Ulitsky I. Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs. *BBA - Gene Regul Mech*. 1859;2016:31–40. <https://doi.org/10.1016/j.bbagr.2015.07.017>.
15. Anderson DM, Anderson KM, Chang C-L, Makarewich CA, Nelson BR, McAnally JR, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*. 2015;160:595–606. <https://doi.org/10.1016/j.cell.2015.01.009>.
16. Nelson B, Makarewich C, Anderson D, Winders B, Troupes C, Wu F, et al. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*. 2016;351. <https://doi.org/10.1126/science.aad4076>.
17. Stein CS, Jadiya P, Zhang X, Anderson EJ, Elrod JW, Boudreau RL, et al. Mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial supercomplexes and report mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial supercomplexes and respiratory efficiency. *Cell Rep*. 2018;23:3710–3720.e8. <https://doi.org/10.1016/j.celrep.2018.06.002>.
18. Chng SC, Ho L, Tian J, Reversade B. ELABELA: a hormone essential for heart development signals via the apelin receptor. *Dev Cell*. 2013;27:672–80. <https://doi.org/10.1016/j.devcel.2013.11.002>.
19. Matsumoto A, Pasut A, Matsumoto M, Yamashita R, Fung J, Monteleone E, et al. MTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature*. 2017;541:228–32. <https://doi.org/10.1038/nature21034>.
20. Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, et al. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife*. 2013;2:1–24. <https://doi.org/10.7554/eLife.01749>.
21. Lai KMV, Gong G, Atanasio A, Rojas J, Quispe J, Posca J, et al. Diverse phenotypes and specific transcription patterns in twenty mouse lines with ablated lincRNAs. *PLoS One*. 2015;10:1–21. <https://doi.org/10.1371/journal.pone.0125522>.
22. Young TL, Matsuda T, Cepko CL. The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. *Curr Biol*. 2005;15:501–12. <https://doi.org/10.1016/j.cub.2005.02.027>.
23. Balas MM, Johnson AM. Exploring the mechanisms behind long noncoding RNAs and cancer. *Non-Coding RNA Res*. 2018;3:108–17. <https://doi.org/10.1016/j.ncrna.2018.03.001>.
24. Khalil A, Guttman M, Huarte M, Garber M, Raj A, Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci*. 2009;106:11667–72. <https://doi.org/10.1073/pnas.0904715106>.
25. He Q, Yang S, Gu X, Li M, Wang C, Wei F. Long noncoding RNA TUG1 facilitates osteogenic differentiation of periodontal ligament stem cells via interacting with Lin28A. *Cell Death Dis*. 2018;9:455. <https://doi.org/10.1038/s41419-018-0484-2>.
26. Long J, Overbeek PA, Danesh FR, Long J, Badal SS, Ye Z, et al. Long noncoding RNA Tug1 regulates mitochondrial bioenergetics in diabetic nephropathy. *J Clin Invest*. 2016;126:4205–18. <https://doi.org/10.1172/JCI87927.14>.
27. Yang L, Lin C, Liu W, Zhang J, Ohgi KA, Jonathan D, et al. ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell*. 2011;147:773–88. <https://doi.org/10.1016/j.cell.2011.08.054.ncRNA>.
28. Li J, Zhang M, An G, Ma Q. LncRNA TUG1 acts as a tumor suppressor in human glioma by promoting cell apoptosis. *Exp Biol Med*. 2016;241:644–9. <https://doi.org/10.1177/1535370215622708>.
29. Katsushima K, Natsume A, Ohka F, Shinjo K, Hatanaka A, Ichimura N, et al. Targeting the Notch-regulated non-coding RNA TUG1 for glioma treatment. *Nat Commun*. 2016;7:1–14. <https://doi.org/10.1038/ncomms13616>.
30. Du Z, Sun T, Hacisuleyman E, Fei T, Wang X, Brown M, et al. Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nat Commun*. 2016;7:1–10. <https://doi.org/10.1038/ncomms10982>.
31. Chen J, Shishkin AA, Zhu X, Kadri S, Maza I, Guttman M, et al. Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol*. 2016;17:1–17. <https://doi.org/10.1186/s13059-016-0880-9>.
32. Lin M, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*. 2011;27:275–82. <https://doi.org/10.1093/bioinformatics/btr209>.
33. The Mouse ENCODE Consortium. A comparative encyclopedia of DNA elements in the mouse genome. *Nature*. 2014;515:355–64. <https://doi.org/10.1038/nature13992>.

34. Fagerberg L, Hallstro M, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based. *Mol Cell Proteomics*. 2014;13:397–406. <https://doi.org/10.1074/mcp.M113.035600>.
35. Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol*. 2015;16:20. <https://doi.org/10.1186/s13059-015-0586-4>.
36. Zhang E, Yin D, Sun M, Kong R, Liu X, You L, et al. P53-regulated long non-coding RNA TUG1 affects cell proliferation in human non-small cell lung cancer, partly through epigenetically regulating HOXB7 expression. *Cell Death Dis*. 2014;5. <https://doi.org/10.1038/cddis.2014.201>.
37. Van Heesch S, Van Irterson M, Jacobi J, Boymans S, Essers PB, De Bruijn E, et al. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol*. 2014;15:1–12. <https://doi.org/10.1186/gb-2014-15-1-r6>.
38. Firlit BCF, Davis JR. Morphogenesis of the residual body of the mouse testis. *J Cell Sci*. 1965;106:93–8.
39. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518:331–6. <https://doi.org/10.1038/nature14222>.
40. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376–80. <https://doi.org/10.1038/nature11082>.
41. Goff LA, Groff AF, Sauvageau M, Trayer-Gibson Z, Sanchez-Gomez DB, Morse M, et al. Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci U S A*. 2015;112:6855–62. <https://doi.org/10.1073/pnas.1411263112>.
42. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*. 2011;477:289–94. <https://doi.org/10.1038/nature10413>.
43. Xu Y, Wang J, Qiu M, Xu L, Li M, Jiang F, et al. Upregulation of the long noncoding RNA TUG1 promotes proliferation and migration of esophageal squamous cell carcinoma. *Tumor Biol*. 2014. <https://doi.org/10.1007/s13277-014-2763-6>.
44. Han Y, Liu Y, Gui Y, Cai Z. Long intergenic non-coding RNA TUG1 is overexpressed in urothelial carcinoma of the bladder. *J Surg Oncol*. 2013;107:555–9. <https://doi.org/10.1002/jso.23264>.
45. Makarewich CA, Olson EN. Mining for micropeptides. *Trends Cell Biol*. 2017;27:685–96. <https://doi.org/10.1016/j.tcb.2017.04.006>.
46. Ji Z, Song R, Regev A, Struhl K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife*. 2015;4:1–21. <https://doi.org/10.7554/eLife.08890>.
47. van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, et al. The translational landscape of the human heart. *Cell*. 2019;178:242–260.e29. <https://doi.org/10.1016/j.cell.2019.05.010>.
48. Gascoigne DK, Cheetham SW, Cattenoz PB, Clark MB, Amaral PP, Taft RJ, et al. PinStripe: a suite of programs for integrating transcriptomic and proteomic datasets identifies novel proteins and improves differentiation of protein-coding and non-coding genes. *Bioinformatics*. 2012;28:3042–50. <https://doi.org/10.1093/bioinformatics/bts582>.
49. Michel AM, Fox G, Kiran AM, De Bo C, O'Connor PBF, Heaphy SM, et al. GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res*. 2014;42:859–64. <https://doi.org/10.1093/nar/gkt1035>.
50. Bazzini AA, Johnstone TG, Christiano R, MacKowiak SD, Obermayer B, Fleming ES, et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J*. 2014;33:981–93. <https://doi.org/10.1002/emboj.201488411>.
51. Chew G-L, Pauli A, Rinn JL, Regev A, Schier AF, Valen E. Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development*. 2013;140:2828–34. <https://doi.org/10.1242/dev.098343>.
52. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of resolution using ribosome profiling. *Science*. 2009;324:218–23. <https://doi.org/10.1126/science.1168978>.
53. Wassarman DA, Steitz JA. Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. *Mol Cell Biol*. 1991;11:3432–45. <https://doi.org/10.1128/mcb.11.7.3432>.
54. Gurney T, Eliceiri GL. Intracellular distribution of low molecular weight RNA species in HeLa cells. *J Cell Biol*. 1980;87:398–403. <https://doi.org/10.1083/jcb.87.2.398>.
55. Macho A, Decaudin D, Castedo M, Hirsch T, Susin SA, Zamzami N, et al. Chloromethyl-X-rosamine is an aldehyde-fixable potential-sensitive fluorochrome for the detection of early apoptosis. *Cytometry*. 1996;25:333–40. [https://doi.org/10.1002/\(SICI\)1097-0320\(19961201\)25:4<333::AID-CYTO4>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1097-0320(19961201)25:4<333::AID-CYTO4>3.0.CO;2-E).
56. Likić VA, Perry A, Hulett J, Derby M, Traven A, Waller RF, et al. Patterns that define the four domains conserved in known and novel isoforms of the protein import receptor Tom20. *J Mol Biol*. 2005;347:81–93. <https://doi.org/10.1016/j.jmb.2004.12.057>.
57. Platts AE, Dix DJ, Chemes HE, Thompson KE, Goodrich R, Rockett JC, et al. Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. *Hum Mol Genet*. 2007;16:763–73. <https://doi.org/10.1093/hmg/ddm012>.
58. Wichman L, Somasundaram S, Breindel C, Valerio DM, McCarrey JR, Hodges CA, et al. Dynamic expression of long noncoding RNAs reveals their potential roles in spermatogenesis and fertility. *Biol Reprod*. 2017;97:313–23. <https://doi.org/10.1093/biolre/iox084>.
59. Heinen TJA, Staubach F, Häming D, Tautz D. Emergence of a new gene from an intergenic region. *Curr Biol*. 2009;19:1527–31. <https://doi.org/10.1016/j.cub.2009.07.049>.
60. Arun G, Akhade VS, Donakonda S, Rao MRS. mrhl RNA, a long noncoding RNA, negatively regulates Wnt signaling through its protein partner Ddx5/p68 in mouse spermatogonial cells. *Mol Cell Biol*. 2012;32:3140–52. <https://doi.org/10.1128/MCB.00006-12>.
61. Anguera MC, Ma W, Clift D, Namekawa S, Kelleher RJ, Lee JT. Tsx produces a long noncoding RNA and has general functions in the germline, stem cells, and brain. *PLoS Genet*. 2011;7. <https://doi.org/10.1371/journal.pgen.1002248>.
62. Goudarzi M, Berg K, Pieper LM, Schier AF. Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility. *Elife*. 2019;8:1–17.
63. Qi H, Liu M, Emery DW, Stamatoyannopoulos G. Functional validation of a constitutive autonomous silencer element. *PLoS One*. 2015;10:1–14. <https://doi.org/10.1371/journal.pone.0124588>.
64. Tan YP, Li S, Jiang XJ, Loh W, Foo YK, Loh CB, et al. Regulation of protocadherin gene expression by multiple neuron-restrictive silencer elements scattered in the gene cluster. *Nucleic Acids Res*. 2010;38:4985–97. <https://doi.org/10.1093/nar/gkq246>.

65. Li LM, Arnosti DN. Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes. *Curr Biol*. 2011;21:406–12. <https://doi.org/10.1016/j.cub.2011.01.054>.
66. Li Q, Lewandowski JP, Powell MB, Norrie JL, Cho SH, Vokes SA. A Gli silencer is required for robust repression of gremlin in the vertebrate limb bud. *Development*. 2014;141:1906–14. <https://doi.org/10.1242/dev.104299>.
67. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011;472:120–4. <https://doi.org/10.1038/nature09819>.
68. Sunwoo H, Wu JY, Lee JT. The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. *Proc Natl Acad Sci*. 2015;112:E4216–25. <https://doi.org/10.1073/pnas.1503690112>.
69. Lemmers RJLF, Wohlgemuth M, Van Der Gaag KJ, Van Der Vliet PJ, Van Teijlingen CMM, De Knijff P, et al. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *Am J Hum Genet*. 2007;81:884–94. <https://doi.org/10.1086/521986>.
70. Wijmenga C, Hewitt JE, Sandkuijper LA, Clark LN, Tracy J, Dauwerse WHG, et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet*. 1992;2:26–30.
71. Gabellini D, Green MR, Tupler R. Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell*. 2002;110:339–48. [https://doi.org/10.1016/S0092-8674\(02\)00826-7](https://doi.org/10.1016/S0092-8674(02)00826-7).
72. Rijkers T, Deidda G, Van Koningsbruggen S, Van Geel M, Lemmers RJLF, Van Deutekom JCT, et al. FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. *J Med Genet*. 2004;41:826–36. <https://doi.org/10.1136/jmg.2004.019364>.
73. Cabianca DS, Gabellini D. FSHD: copy number variations on the theme of muscular dystrophy. *J Cell Biol*. 2010;191:1049–60. <https://doi.org/10.1083/jcb.201007028>.
74. Snider L, Asawachaicharn A, Tyler AE, Geng LN, Petek LM, Maves L, et al. RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum Mol Genet*. 2009;18:2414–30. <https://doi.org/10.1093/hmg/ddp180>.
75. Zeng W, De Greef JC, Chen YY, Chien R, Kong X, Gregson HC, et al. Specific loss of histone H3 lysine 9 trimethylation and HP1 γ /cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). *PLoS Genet*. 2009;5. <https://doi.org/10.1371/journal.pgen.1000559>.
76. Dumbovic G, Forcales SV, Peruchio M. Emerging roles of macrosatellite repeats in genome organization and disease development. *Epigenetics*. 2017;12:515–26. <https://doi.org/10.1080/15592294.2017.1318235>.
77. Niessen P, Rensen S, Van Deursen J, De Man J, De Laet A, Vanderwinden JM, et al. Smoothelin-A is essential for functional intestinal smooth muscle contractility in mice. *Gastroenterology*. 2005;129:1592–601. <https://doi.org/10.1053/j.gastro.2005.08.018>.
78. Sato H, Taketomi Y, Isogai Y, Miki Y, Yamamoto K, Masuda S, et al. Group III secreted phospholipase A2 regulates epididymal sperm maturation and fertility in mice. *J Clin Invest*. 2010;120. <https://doi.org/10.1172/JCI40493.1400>.
79. Hill KE, Zhou J, McMahan WJ, Motley AK, Atkins JF, Gesteland RF, et al. Deletion of selenoprotein P alters distribution of selenium in the mouse. *J Biol Chem*. 2003;278:13640–6. <https://doi.org/10.1074/jbc.M300755200>.
80. Xu J. Preparation, culture, and immortalization of mouse embryonic fibroblasts. *Curr Protoc Mol Biol* 28, 2005. doi: <https://doi.org/10.1002/0471142727.mb2801s70>.
81. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods*. 2008;5:877–9. <https://doi.org/10.1038/nmeth.1253>.
82. Sun L, Goff LA, Trapnell C, Alexander R, Lo KA, Hacisuleyman E, et al. Long noncoding RNAs regulate adipogenesis. *PNAS*. 2013;110:3387–92. <https://doi.org/10.1073/pnas.1222643110>.
83. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
84. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30. <https://doi.org/10.1093/bioinformatics/btt656>.
85. Drost H. Philentropy: information theory and distance quantification with R. *J Open Source Softw*. 2018;3:1–4. <https://doi.org/10.21105/joss.00765>.
86. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:1–21. <https://doi.org/10.1186/s13059-014-0550-8>.
87. Strimmer K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics*. 2008;24:1461–2. <https://doi.org/10.1093/bioinformatics/btn209>.
88. Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert B, Gillette M, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102:15545–50. <https://doi.org/10.3969/jissn.0372-2112.2018.08.016>.
89. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res*. 2012;40:1–12. <https://doi.org/10.1093/nar/gks461>.
90. Perez JD, Rubinstein ND, Fernandez DE, Santoro SW, Needleman LA, Ho-Shing O, et al. Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *Elife*. 2015;4:41. <https://doi.org/10.7554/eLife.07860>.
91. Mituyama T, Yamada K, Hattori E, Okida H, Ono Y, Terai G, et al. The functional RNA database 3.0: databases to support mining and annotation of functional RNAs. *Nucleic Acids Res*. 2009;37:89–92. <https://doi.org/10.1093/nar/gkn805>.
92. Turro E, Su SY, Gonçalves A, Coin LJM, Richardson S, Lewin A. Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biol*. 2011;12:R13. <https://doi.org/10.1186/gb-2011-12-2-r13>.
93. Turro E, Astle WJ, Tavaré S. Flexible analysis of RNA-seq data using mixed effects models. *Bioinformatics*. 2014;30:180–8. <https://doi.org/10.1093/bioinformatics/btt624>.
94. Livak K, Schmittgen T. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods*. 2001;25:402–8. <https://doi.org/10.1006/meth.2001.1262>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.