

# The *Drosophila* Translational Control Element (TCE) Is Required for High-Level Transcription of Many Genes That Are Specifically Expressed in Testes

Rebecca J. Katzenberger<sup>1</sup>\*, Elizabeth A. Rach<sup>2</sup>\*, Ashley K. Anderson<sup>1</sup>, Uwe Ohler<sup>2\*</sup>, David A. Wassarman<sup>1\*</sup>

**1** University of Wisconsin School of Medicine and Public Health, Department of Cell and Regenerative Biology, Madison, Wisconsin, United States of America, **2** Institute for Genome Sciences and Policy, Departments of Biostatistics and Bioinformatics and Computer Science, Duke University, Durham, North Carolina, United States of America

## Abstract

To investigate the importance of core promoter elements for tissue-specific transcription of RNA polymerase II genes, we examined testis-specific transcription in *Drosophila melanogaster*. Bioinformatic analyses of core promoter sequences from 190 genes that are specifically expressed in testes identified a 10 bp A/T-rich motif that is identical to the translational control element (TCE). The TCE functions in the 5' untranslated region of *Mst(3)CGP* mRNAs to repress translation, and it also functions in a heterologous gene to regulate transcription. We found that among genes with focused initiation patterns, the TCE is significantly enriched in core promoters of genes that are specifically expressed in testes but not in core promoters of genes that are specifically expressed in other tissues. The TCE is variably located in core promoters and is conserved in *melanogaster* subgroup species, but conservation dramatically drops in more distant species. In transgenic flies, short (300–400 bp) genomic regions containing a TCE directed testis-specific transcription of a reporter gene. Mutation of the TCE significantly reduced but did not abolish reporter gene transcription indicating that the TCE is important but not essential for transcription activation. Finally, mutation of testis-specific TFIID (tTFIID) subunits significantly reduced the transcription of a subset of endogenous TCE-containing but not TCE-lacking genes, suggesting that tTFIID activity is limited to TCE-containing genes but that tTFIID is not an obligatory regulator of TCE-containing genes. Thus, the TCE is a core promoter element in a subset of genes that are specifically expressed in testes. Furthermore, the TCE regulates transcription in the context of short genomic regions, from variable locations in the core promoter, and both dependently and independently of tTFIID. These findings set the stage for determining the mechanism by which the TCE regulates testis-specific transcription and understanding the dual role of the TCE in translational and transcriptional regulation.

**Citation:** Katzenberger RJ, Rach EA, Anderson AK, Ohler U, Wassarman DA (2012) The *Drosophila* Translational Control Element (TCE) Is Required for High-Level Transcription of Many Genes That Are Specifically Expressed in Testes. PLoS ONE 7(9): e45009. doi:10.1371/journal.pone.0045009

**Editor:** Ferenc Mueller, University of Birmingham, United Kingdom

**Received:** May 15, 2012; **Accepted:** August 11, 2012; **Published:** September 11, 2012

**Copyright:** © 2012 Katzenberger et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by National Institutes of Health (NIH) grant R01-HG004065 (to U.O.) and National Science Foundation (NSF) grant MCB-1118456 (to D.A.W.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: dawassarman@wisc.edu (DAW); uwe.ohler@duke.edu (UO)

† These authors contributed equally to this work.

## Introduction

The core promoter is the region surrounding the transcription start site (TSS) of a gene that functions to recruit RNA polymerase II to initiate transcription [1]. Historically, core promoters were thought to function by a single generic mechanism. However, recent discoveries suggest that core promoters function by numerous different mechanisms and that the differences are important for tissue-specific transcription programs [2,3]. The first key discovery is the diversity of elements that contribute to core promoter function [4–7]. There are no universal core promoter elements. Familiar core promoter elements, such as the TATA box and initiator (Inr), are only present in a minority of promoters, and computational analyses, followed by experimental validation, have helped to identify an ever-increasing number of core promoter elements, such as the motif ten element (MTE) [4,8]. The observation that maternally expressed genes in *Drosophila* have different core promoter element compositions than zygotically

active genes highlights the importance of core promoter element diversity for transcription regulation [5,9].

The second key discovery is the existence of two classes of transcription initiation patterns that are associated with different types of core promoter elements and chromatin structures [1,9]. Genes with focused initiation patterns initiate transcription at a single nucleotide or within a region of several nucleotides, and genes with dispersed initiation patterns initiate transcription at numerous sites in a region of 100–200 bp. Focused initiation patterns tend to be associated with location-specific core promoter elements, such as a TATA box at –30 bp or an Inr at –2 bp relative to the TSS, and are generally associated with regulated genes. Dispersed initiation patterns tend to contain variably located core promoter elements, such as the DNA replication-related element (DRE), and are generally associated with constitutively expressed genes [9,10].

The third key discovery is the diversity of protein factors that recognize core promoter elements. The general transcription

factor TFIID is the major core promoter recognition factor [11]. TFIID is a multi-protein complex composed of TATA-binding protein (TBP) and ~15 TBP-associated factors (TAFs) [12,13]. Recognition of core promoters by TFIID is exemplified by TBP, which binds the TATA box, and TAF1 and TAF2, which together bind the Inr [1,3]. In metazoans, paralogs of TBP, such as TBP-related factor 2 (TRF2), and TAFs, such as TAF4b, play unique roles in regulating transcription [14,15]. Thus, combinatorial capacity, provided by diverse core promoter elements and TFIID subunits, is likely to play a major role in tissue-specific transcription.

*Drosophila* spermatogenesis is an ideal system to investigate roles for core promoter element and TFIID diversity in the regulation of tissue-specific transcription. Genome-wide analyses of mRNA expression in a variety of *Drosophila* tissues have determined that >1,000 genes are preferentially or uniquely expressed in testes [16,17]. Transcription of genes that are specifically expressed in testes is commonly controlled by 100–400 bp genomic regions that include the TSS [18,19]. Particular sequences have been identified that are important for the testis-specific transcription of genes; however, no common testis-specific core promoter elements have been identified. In regard to TFIID diversity, paralogs of ubiquitously expressed TAF4, TAF5, TAF6, TAF8, and TAF12 (No hitter (Nht), Cannonball (Can), Meiosis I arrest (Mia), Spermatocyte arrest (Sa), and Ryan express (Rye), respectively) are predominantly expressed in testes [19–21]. Flies mutant for any one of the testis-specific TAFs (tTAFs) are male sterile, they arrest spermatogenesis prior to meiosis, and they have reduced transcription of genes required for entry into meiosis. tTAFs may be components of a testis-specific TFIID (tTFIID) complex since they co-localize within spermatocyte nuclei [22,23].

Based on these findings, we hypothesized that *Drosophila* genes that are specifically expressed in testes contain novel core promoter elements that regulate testis-specific transcription. Furthermore, we hypothesized that testis-specific core promoter elements are recognized by a tTFIID complex. Here, we present experiments that test these hypotheses.

## Results

### Identification of the TCE as a testis-specific core promoter element

To identify sequence elements that are enriched in core promoters of genes that are specifically expressed in testes, we examined core promoters surrounding TSSs with testis-specific activity. While recent sequencing protocols have made it possible to map TSSs at high resolution and throughput, such approaches are not straightforward to apply to small amounts of RNA from specific tissues. We therefore used a high quality set of 5,665 TSSs for 3,990 genes, which had been defined by hierarchical clustering of expressed sequence tags (ESTs) from 15 different *Drosophila melanogaster* libraries [9]. Shannon entropy was used as a metric to identify 1,997 condition-specific TSSs; those that occurred specifically in one of eight cDNA libraries with sufficient coverage (embryo, larva/pupa, head, ovary, testis, Schneider cell (S2), mbn2 hemocytic cell, and fat body). Among the 1,997 TSSs, 1,395 had focused initiation patterns. Of these, 190 were specifically expressed in testes and 1,205 were specifically expressed in another condition.

Computational sequence motif searches by MEME and Gibbs Sampling identified sequence elements that were overrepresented in the vicinity of the 190 testis-specific TSSs [24,25]. For these searches, core promoters were defined as the region from 60 bp upstream (–60) to 40 bp downstream (+40) of the TSS. Details of

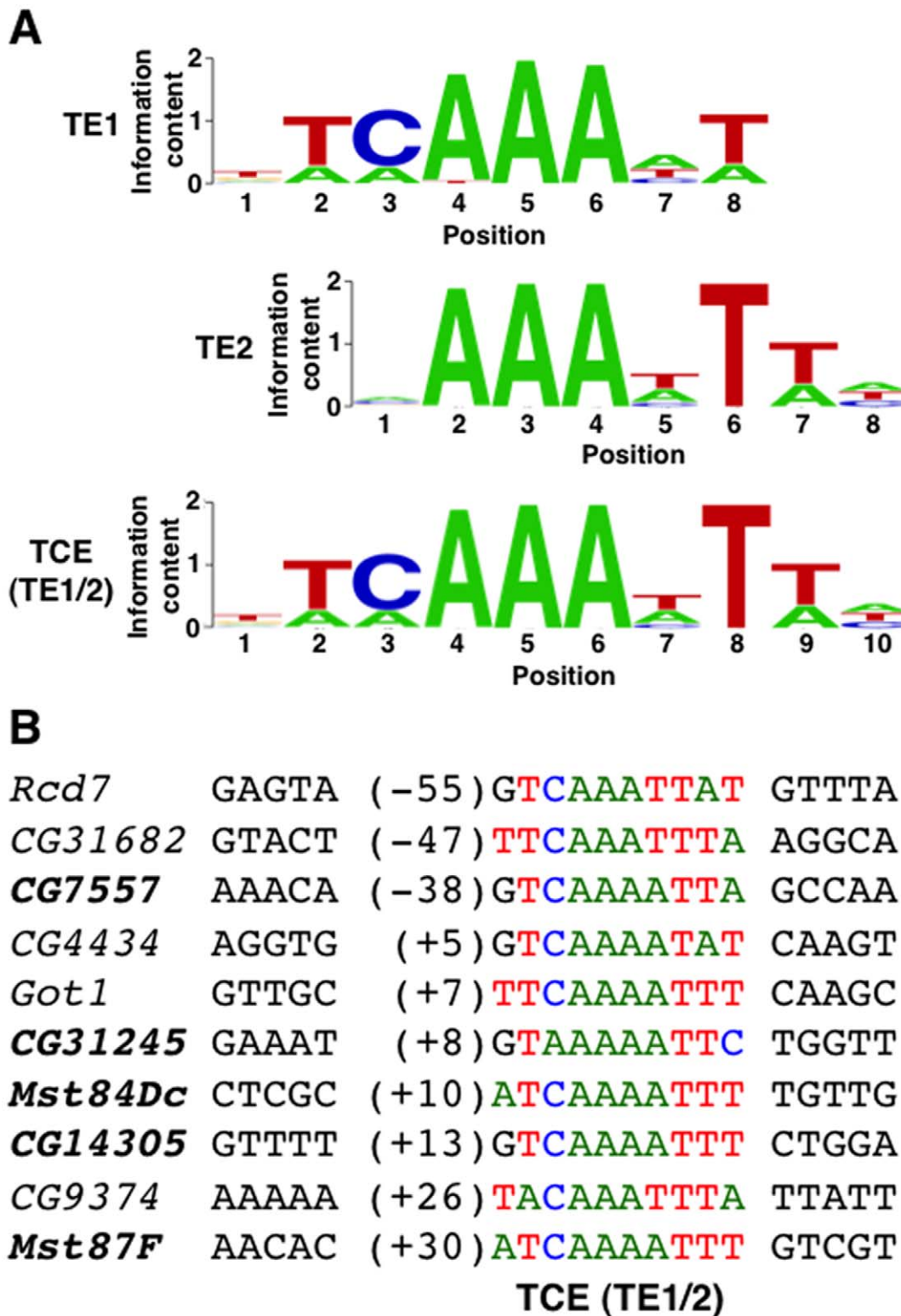
the search parameters are provided in the Materials and Methods. The independent searches returned highly similar results; two adenine/thymine (A/T)-rich elements were found to occur in ~60% of the testis-specific core promoters. DNA sequence logo representations of the elements identified by Gibbs Sampling are shown in Figure 1A and position frequency matrices for the elements are shown in Tables S1 and S2 [26]. The elements were named Testis Element 1 (TE1) and Testis Element 2 (TE2).

To determine the extent to which the elements are testis-specific, Patser was used to scan for matches of TE1 and TE2 in core promoters from each of the eight conditions [27]. This analysis revealed that TE1 and TE2 were present in 46% and 49% of the testis-specific core promoters, respectively (Table 1). At the same false positive stringency (i.e. at matrix scores corresponding to the same P-values), the frequency of TE1 or TE2 in testis-specific core promoters was higher than either the TATA box (33%) or Inr (40%) in embryo-specific core promoters [28]. In the sets of other condition-specific promoters, TE1 and TE2 occurred at a noticeably lower rate (9–11% in head and embryo promoters, for which the datasets were sufficiently large for analysis). These data indicate that TE1 and TE2 are overrepresented in testis-specific core promoters.

As illustrated in Figure 1A, both TE1 and TE2 had three consecutive invariant adenines, raising the possibility that TE1 and TE2 overlap to form a longer element, which was named Testis Element 1/2 (TE1/2). Indeed, 45% of testis-specific core promoters were found to contain TE1/2, a 10 nt position weight matrix (PWM) joined from TE1 and TE2 (Figure 1A and Table 1). Figure 1B provides examples of TE1/2 sequences in the core promoters of genes that are specifically expressed in testes. Thus, we consider TE1 and TE2 motifs non-canonical TE1/2 motifs rather than as separate or distinct functional motifs.

TE1/2 is identical to a previously defined motif called the Translational Control Element (TCE) [29]. For this reason, we hereafter refer to TE1/2 as the TCE. The TCE is a 12 nt motif (ACATCNAATTT) that was defined based on sequence comparison among seven members of the *Mst(3)CGP* gene family, including *Mst(3)87F*, that are specifically expressed in testes. In *Mst(3)CGP* genes, the TCE is invariantly located at+28 relative to the TSS. A reporter gene containing a fragment of *Mst(3)87F* from –670 to+51 fused upstream of the *LacZ* gene was used to demonstrate that the TCE functions as a repressor of translation in diploid stage cells of third instar larval testes. Deletion or mutation of the TCE in the reporter gene results in premature translation, as observed by  $\beta$ -galactosidase ( $\beta$ -gal) activity in third instar larval testes. Additionally, moving the TCE 26 nt downstream to+54 results in premature translation, indicating that the TCE regulates translation in a location-specific manner [30]. However, it has been noted that other genes that are specifically expressed in testes and whose expression is under translational control contain TCE-like sequences at positions other than+28, bringing into question the location-specific requirement for activity [31]. Finally, there is limited evidence that the TCE regulates transcription in testes. Insertion of the TCE at+28 restores transcriptional activity to  $\beta$ -tubulin ( $\beta$ 2t) transgenes that lack a 14 bp fragment required for testis-specific transcription. In subsequent studies, we investigated the role of the TCE in testis-specific transcriptional regulation.

Since genes with focused initiation patterns typically have location-specific core promoter elements, we examined the location of the TCE in core promoters of genes that are specifically expressed in testes. TCE matches were located throughout the –60 to+40 core promoter region but were enriched in the –5 to+25 region (Figure 2A). In contrast, random intergenic sequences showed no such pattern, and enrichments



**Figure 1. The TCE (TE1/2) and related sequences, TE1 and TE2.** (A) Shown are DNA sequence logos of TE1, TE2, and TE1/2, which is identical to the previously defined TCE [29]. The logos have been aligned relative to the triplet adenines. (B) *Drosophila* genes that are specifically expressed in testes are aligned by the TCE. Genes indicated in bold font were subject to further analysis in Figure 4 and Table S3. Indicated in parentheses is the TCE location relative to the TSS. TCE nucleotides are colored to match panel A. doi:10.1371/journal.pone.0045009.g001

were lower in the set of core promoters from genes specifically expressed in embryos. Thus, among genes specifically expressed in

testes, the TCE is not strictly located within core promoters but is preferentially located at and immediately downstream of the TSS.

**Table 1.** The TCE (TE1/2) is enriched in testis-specific core promoters.

Condition	Specific TSSs	TE1	TE2	TCE (TE1/2)
Testis	190	87 (46%)	93 (49%)	85 (45%)
Embryo	765	71 (9%)	85 (11%)	79 (10%)
Head	370	38 (10%)	33 (9%)	34 (9%)
fat body	6	1 (17%)	1 (17%)	1 (17%)
larva/pupa	22	2 (9%)	0 (0%)	1 (5%)
Ovary	7	0 (0%)	1 (14%)	0 (0%)
S2 cell	25	3 (12%)	6 (24%)	4 (16%)
mbn2 cell	10	2 (20%)	0 (0%)	1 (10%)
Total	1395	204 (15%)	219 (16%)	205 (15%)

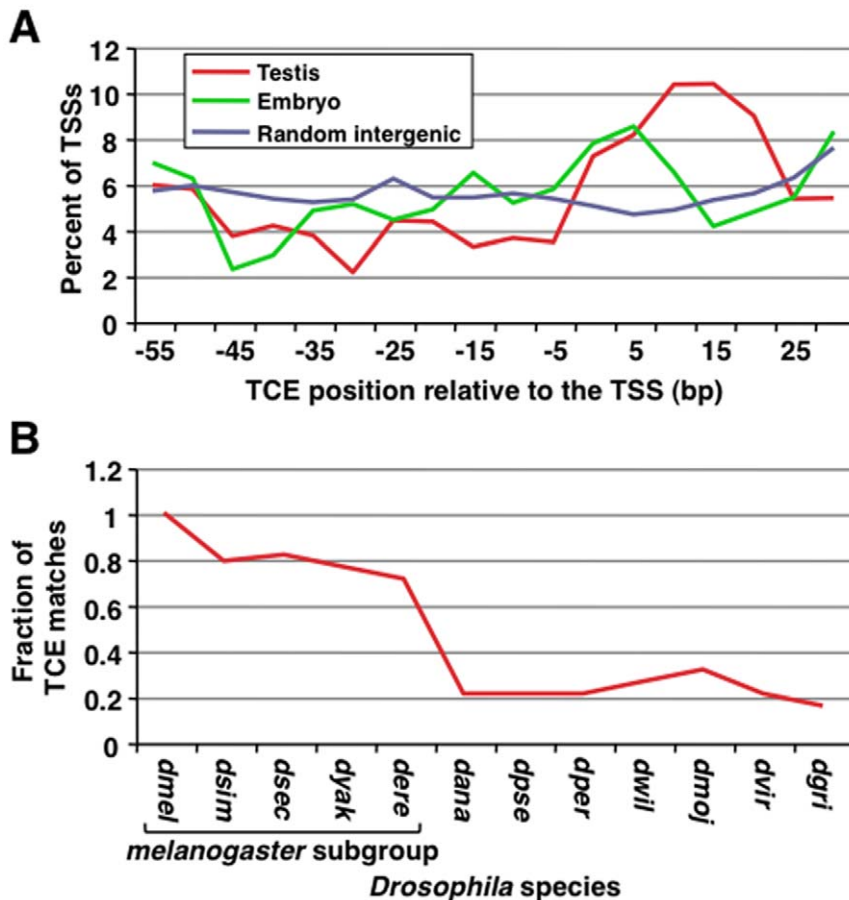
doi:10.1371/journal.pone.0045009.t001

Overall, the evolutionary conservation level of TCEs located in the  $-5/+25$  region is  $\sim 75\%$  in *melanogaster* subgroup species and

drops to  $\sim 25\%$  in more distant species (Figure 2B). The sharp decline in conservation outside of the *melanogaster* subgroup was not unexpected, as the level of preferential conservation is comparable to other variably located core promoter elements, such as the DRE [9,10]. These data suggest that testis-specific core promoters for species outside of the *melanogaster* subgroup either contain TCEs at different locations or contain different elements.

### The TCE is necessary for the transcription of a subset of genes that are specifically expressed in testes

To examine the importance of the TCE for testis-specific transcription, we focused our studies on five TCE-containing genes that are specifically expressed in testes, *Mst84Dc*, *Mst87F*, *CG7557*, *CG14305*, and *CG31245*, two of which (*Mst84Dc* and *Mst87F*) are members of the *Mst(3)CGP* gene family. These genes were chosen because they contain TCEs at different locations relative to the TSS and they differed in mRNA expression level in testes (Figure 1B and Table S3). To determine the extent to which genomic regions encompassing TCE-containing core promoters are sufficient to direct testis-specific transcription, transgenic reporter flies were generated. Short (300–400 bp) genomic regions

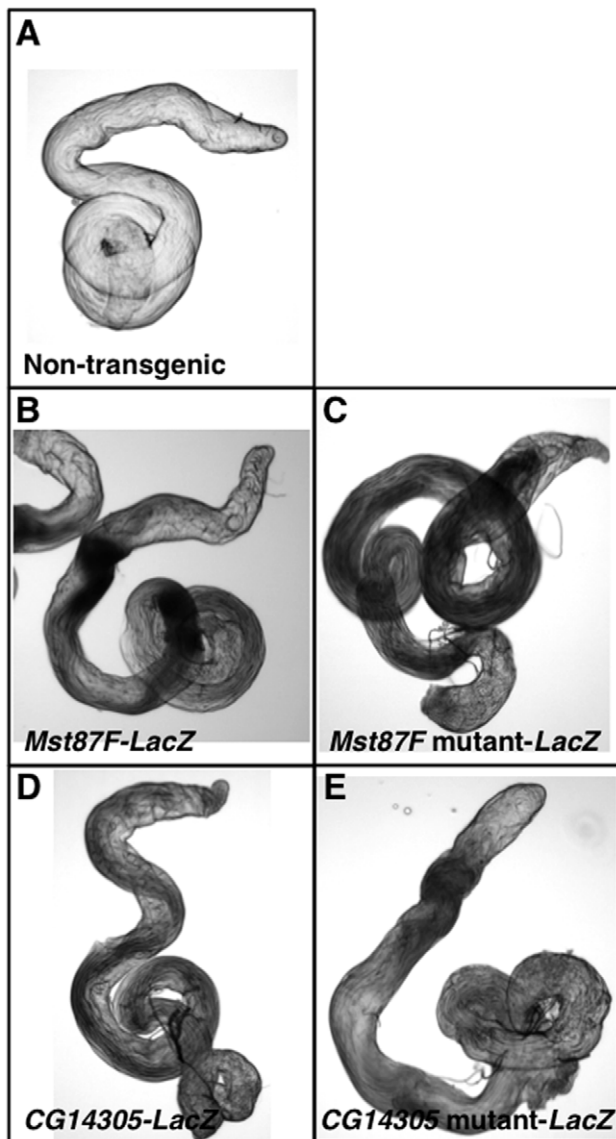


**Figure 2. TCEs are variably located in testis-specific core promoters and are conserved in *melanogaster* subgroup species.** (A) A smoothed plot of the location of the TCE in testis-specific core promoters, embryo-specific core promoters, and random intergenic regions, which serve as background. The calculated background is similar to the expected background of 5.6 (100/18 bins of 5-nt each). Plotted along the x-axis is the location of the TCE relative to the TSS. Plotted along the y-axis is the fraction of genes with a TCE located within a 5-nt interval. (B) Plotted is the TCE conservation across 12 *Drosophila* species, as determined by the fraction of matches in the  $-5/+25$  region in *D. melanogaster* as seen in panel A that are also found at corresponding locations in cross-species alignments. *Drosophila* species abbreviations are as follow: *dmel* (*melanogaster*), *dsim* (*simulans*), *dsec* (*sechellia*), *dyak* (*yakuba*), *dere* (*erecta*), *dana* (*ananassae*), *dpse* (*pseudoobscura*), *dper* (*persimilis*), *dwil* (*willistoni*), *dmoj* (*mojavensis*), *dvir* (*virilis*), and *dgri* (*grimshawi*).

doi:10.1371/journal.pone.0045009.g002

(−182 to+145 for *Mst84Dc*, −104 to+187 for *Mst87F*, −218 to+138 for *CG7557*, −199 to+209 for *CG14305*, and −224 to+177 for *CG31245*) were cloned directly upstream of the *LacZ* coding region in a transformation vector that can be targeted to specific genomic sites by the phiC31 integrase system [32]. To eliminate position effects, all of the transgenes were targeted to the identical site at 86F on the third chromosome. *LacZ* expression was examined in testes of transgenic flies by a whole mount assay for  $\beta$ -gal activity.  $\beta$ -gal activity was detected in testes of transgenic reporter lines but not non-transgenic lines (Figure 3A, B, and D). Thus, short genomic regions containing a TCE are sufficient to direct transcription in testes.

Quantitative real-time reverse transcription-PCR (qPCR) was used to determine whether *LacZ* expression was testis-specific.  $\beta$ -gal



**Figure 3. *Mst87F* and *CG14305* promoter regions direct expression in testes.** Shown is  $\beta$ -gal activity staining of testes from (A) a non-transgenic fly, (B) an *Mst87F-LacZ* transgenic fly, (C) an *Mst87F* mutant-*LacZ* transgenic fly, (D) a *CG14305-LacZ* transgenic fly, and (E) a *CG14305* mutant-*LacZ* transgenic fly. Staining was performed under identical reaction conditions.

doi:10.1371/journal.pone.0045009.g003

mRNA level relative to *actin* mRNA level was determined for RNA extracted from adult flies: whole males, whole females, testes, and male carcasses (the tissue remaining after testis dissection). This analysis revealed that reporters for all five genes were highly expressed in males relative to females (Table 2). Moreover, reporters for all of the genes were significantly enriched in testes relative to male carcasses. These results are consistent with those from a prior study of *Mst87F* transcription regulatory sequences [33]. Thus, short genomic regions encompassing the TCE and the TSS are sufficient to direct testis-specific transcription.

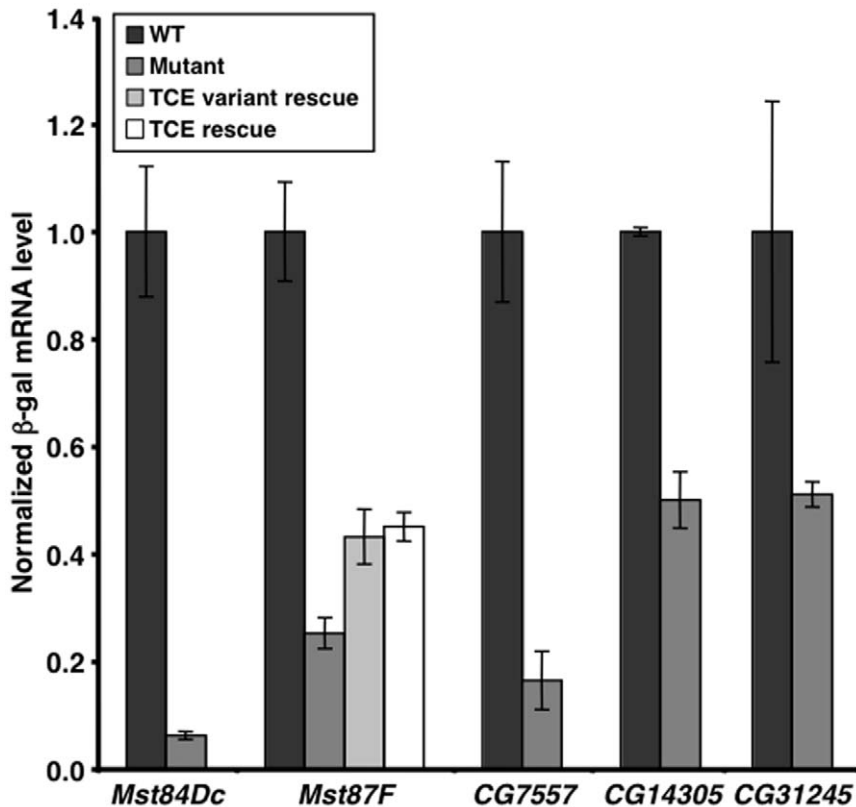
The qPCR transgenic reporter assay was used to determine the extent to which the TCE is necessary for testis-specific transcription of the genes analyzed in Table 2. Site-directed mutagenesis was used to introduce three point mutations into the TCE for each of the five *LacZ* reporter genes. For example, the *Mst87F* TCE sequence ATCAAAATTT was mutated to ATAACACTTT (Figure 4, bottom). Transgenic flies were generated with the reporter genes inserted at 86F on the third chromosome, so that they could be directly compared to the wildtype transgenes.  $\beta$ -gal activity in testes was not overtly different between wildtype and mutant transgenic flies; however, the assay is not sensitive to small changes in expression (Figure 3). To quantitatively assay transcription, qPCR was used to determine  $\beta$ -gal mRNA levels relative to *actin* mRNA levels for RNA extracted from testes of transgenic flies. This analysis revealed that mutation of the TCE significantly reduced expression of the *Mst84Dc*, *Mst87F*, *CG7557*, and *CG14305* transgenes ( $P < 0.05$ ) (Figure 4, top). The average steady-state mRNA level was reduced to 6.2%, 25.2%, 16.4%, and 50.0% of normal levels in *Mst84Dc*, *Mst87F*, *CG7557*, and *CG14305* mutants, respectively. Mutation of the *CG31245* transgene reduced expression to 51.2% of normal levels, but the effect did not meet the significance cutoff ( $P = 0.07$ ). In accord with the documented role of the TCE in translational repression, the lack of a change in  $\beta$ -gal protein expression in TCE mutant flies (Figure 3) may be due to compensatory decreases and increases in  $\beta$ -gal gene transcription and translation, respectively [29–31]. Reduced mRNA expression caused by the TCE mutation could be due to reduced gene transcription or increased mRNA decay. However, an effect on mRNA decay is unlikely because, for genes such as *CG7557*, the TCE is located upstream of the TSS and is not transcribed and, thus, cannot directly serve as a regulatory element for mRNA decay. Therefore, the TCE is necessary for activating the transcription of a subset of genes that are specifically expressed in testes.

To further examine the role of TCE in testis-specific transcription, we determined the extent to which introduction of a TCE motif at a new location within the core promoter could rescue the transcription defect caused by mutation of the normal TCE motif. Site-directed mutagenesis was used to introduce the

**Table 2. TCE-containing genomic regions direct testis-specific transcription.**

Gene	Testis	Male carcass	Whole male	female	male/female
<i>Mst84Dc</i>	2.435±0.268	0.005±0.001	0.699±0.037	0.005±0.001	140
<i>Mst87F</i>	1.681±0.070	0.005±0.001	0.114±0.040	0.003±0.001	38
<i>CG7557</i>	0.506±0.030	0.003±0.000	0.130±0.016	0.002±0.000	65
<i>CG14305</i>	0.348±0.007	0.005±0.000	0.032±0.005	0.002±0.001	16
<i>CG31245</i>	0.457±0.049	0.007±0.001	0.145±0.015	0.005±0.000	29

doi:10.1371/journal.pone.0045009.t002



*Mst84Dc* TGCTGCTCGCATCAAATTTTGTGTTTTG  
*Mst87F* GCCTGAACACATCAAATTTGTCGTCTAGT  
*CG7557* TGATTAACAGTCAAATTAGCCAACAAAT  
*CG14305* TGTAGGTTTTGTCAAAATTTCTGGATTCTGA  
*CG31245* TGTGAGAAATGTAAAATTTCTGGTTAACTA  
 Mutants **A C C**

*Mst87F* WT GCCTGAACACATCAAATTTGTCGTCTAGT  
*Mst87F* mutant GCCTGAACACAT**AACACTTT**GTCGTCTAGT  
 TCE variant rescue GCCTGAACACAT**AACACTTT**GCAAAATG  
 TCE rescue GCCTCAAATTT**AACACTTT**GTCGTCTAGT

**Figure 4. TCE mutations reduce the transcription of genes that are specifically expressed in testes.** Graphed is the mRNA expression of  $\beta$ -gal relative to *actin* in the testes of transgenic flies. Wildtype (WT) and TCE mutant transgenes were examined for each of the indicated genes. Additionally, for *Mst87F*, two rescue transgenes, containing ectopic TCE motifs in the context of a mutant TCE sequence, were examined. mRNA levels were normalized to those of the wildtype transgene. Statistically significant differences are indicated in the text. Below the graph are wildtype and mutant TCE sequences for the five genes analyzed in transgenic reporter gene assays, as well as the *Mst87F* wild type, mutant, and rescue sequences that were used in the transgenic reporter gene assays. Indicated in red are mutated nucleotides. Underlined are the ectopic TCE motifs. doi:10.1371/journal.pone.0045009.g004

TCE sequence CTCAAATTT immediately upstream of the mutant TCE sequence in the *LacZ* reporter gene for *Mst87F* (Figure 4, bottom). The reporter gene was then inserted at 86F on the third chromosome, so that it could be directly compared to the wildtype and the TCE mutant transgenes. qPCR was used to determine  $\beta$ -gal mRNA level relative to *actin* mRNA level for RNA extracted from testes of transgenic flies. This analysis revealed that introduction of the TCE motif caused a significant  $\sim$ 2-fold increase in transcription of the mutant reporter gene ( $P < 0.05$ ) (Figure 4, top). Similarly, introduction of a variant TCE sequence GTCAAAT immediately downstream of the mutant TCE caused a significant  $\sim$ 2-fold increase in the transcription of the

mutant reporter gene ( $P < 0.05$ ) (Figure 4, top and bottom). These data provide additional support for a role for the TCE in the activation of testis-specific transcription.

#### tTAFs are required for the transcription of a subset of TCE-containing genes that are specifically expressed in testes

Since TAF components of TFIID commonly bind core promoter elements, we hypothesized that tTAF components of tTFIID would be required for the transcription of TCE-containing genes that are specifically expressed in testes [3]. In fact, mutation

of the tTAF *can* was previously shown to reduce *Mst87F* and *Mst84Dc* transcription in testes [20]. qPCR of RNA extracted from testes was used to examine the requirement of the tTAF genes *can*, *rye*, and *sa* for transcription of the ten TCE-containing genes listed in Figure 1B. Mutation of any of the tTAF genes significantly reduced the mRNA level of five of the genes, *Rcd7*, *CG7557*, *CG31245*, *Mst84Dc*, and *Mst87F* ( $P < 0.05$ ) (Figure 5A). In contrast, tTAF mutations did not affect the transcription of ten genes that lack a TCE, all of which, except *twine* (*twe*), were chosen from among the 190 genes that were used to identify the TCE (Figure 5B). Taken together, these data suggest an intricate functional relationship between the TCE and a tTFIID complex containing Can, Rye, and Sa.

## Discussion

Here, we have determined that the TCE is present in the core promoter of a substantial fraction of *Drosophila melanogaster* genes with focused transcription initiation patterns that are specifically expressed in testes. *In vivo* studies indicate that the TCE functions in a location-independent manner to activate the transcription of many genes that are specifically expressed in testes. Furthermore, *in vivo* studies indicate that a tTFIID complex containing tTAFs regulates testis-specific transcription of a subset of TCE-containing genes. To our knowledge, this is the first example in a metazoan organism of tissue-specific transcription regulation by a common, tissue-specific core promoter element.

Relatively little is known about *cis*-regulation of testis-specific transcription in *Drosophila*. A previous computational study applied a regression tree approach to explain sex-specific expression differences by the occurrence of co-occurring k-mers in 1 kb upstream regions but only identified female-specific significant sequence motifs [34]. Studies of several genes indicate that short genomic regions encompassing the TSS are sufficient to direct testis-specific transcription and that the combined activities of elements within these regions provide testis specificity and control transcription level [18,19]. For example, a genomic region from  $-53$  to  $+172$  of the  *$\beta 2t$*  gene is sufficient to direct transcription in testes [35,36]. A  $\beta 2UE1$  element at  $-45$  is necessary for testis-specific transcription, and two elements,  $\beta 2UE2$  at  $-29$  and  $\beta 2DE1$  at  $+60$ , are necessary for high-level transcription. Our data are consistent with these findings. Short genomic regions containing the TCE were sufficient to direct testis-specific transcription, and the TCE was necessary for high-level transcription in testes. What is unique about the TCE, relative to previously described testis elements, including those in the  *$\beta 2t$*  gene, is that the TCE is present in many genes that are specifically expressed in testes. Among the 190 genes with focused initiation patterns, 85 contained a TCE. Common core promoter elements may also coordinate testis-specific transcription programs in mammals. As in *Drosophila*, many mouse and human genes are specifically expressed in testes and short genomic regions surrounding the TSS are sufficient to direct testis-specific transcription [37–39].

An unexpected but potentially important feature of the testis-specific transcription mechanism is the variable location of the TCE in core promoters. The variable location was unexpected because genes with focused initiation patterns, which were used to identify the TCE, tend to be associated with location-specific core promoter elements, such as the TATA box, Inr, DPE, and MTE [9]. Furthermore, TFIID complexes have only been shown to bind location-specific core promoter elements, such as the TATA box, Inr, DPE, and MTE [3]. However, the TCE is located throughout core promoters. Moreover, TCEs located either upstream or downstream of the TSS were necessary for testis-specific

transcription, and insertion of a TCE at non-endogenous sites in a core promoter partially restored normal transcription levels to a TCE mutant core promoter. We cautiously note that for some genes that are specifically expressed in testes the location of the TSS and/or the designation of a focused initiation pattern may be inaccurate. In contrast to most mapped fly TSSs, testis-specific TSSs were based on 5' EST libraries prepared without cap trapping. While the same stringency was required as in previous work, in which promoters from cap-trapped libraries were examined, we expect the majority of promoters to be close to the indicated TSS but not necessarily exactly match in all cases [9]. In addition, the prevalence of focused initiation patterns appeared higher than was later determined by high-throughput sequencing [28].

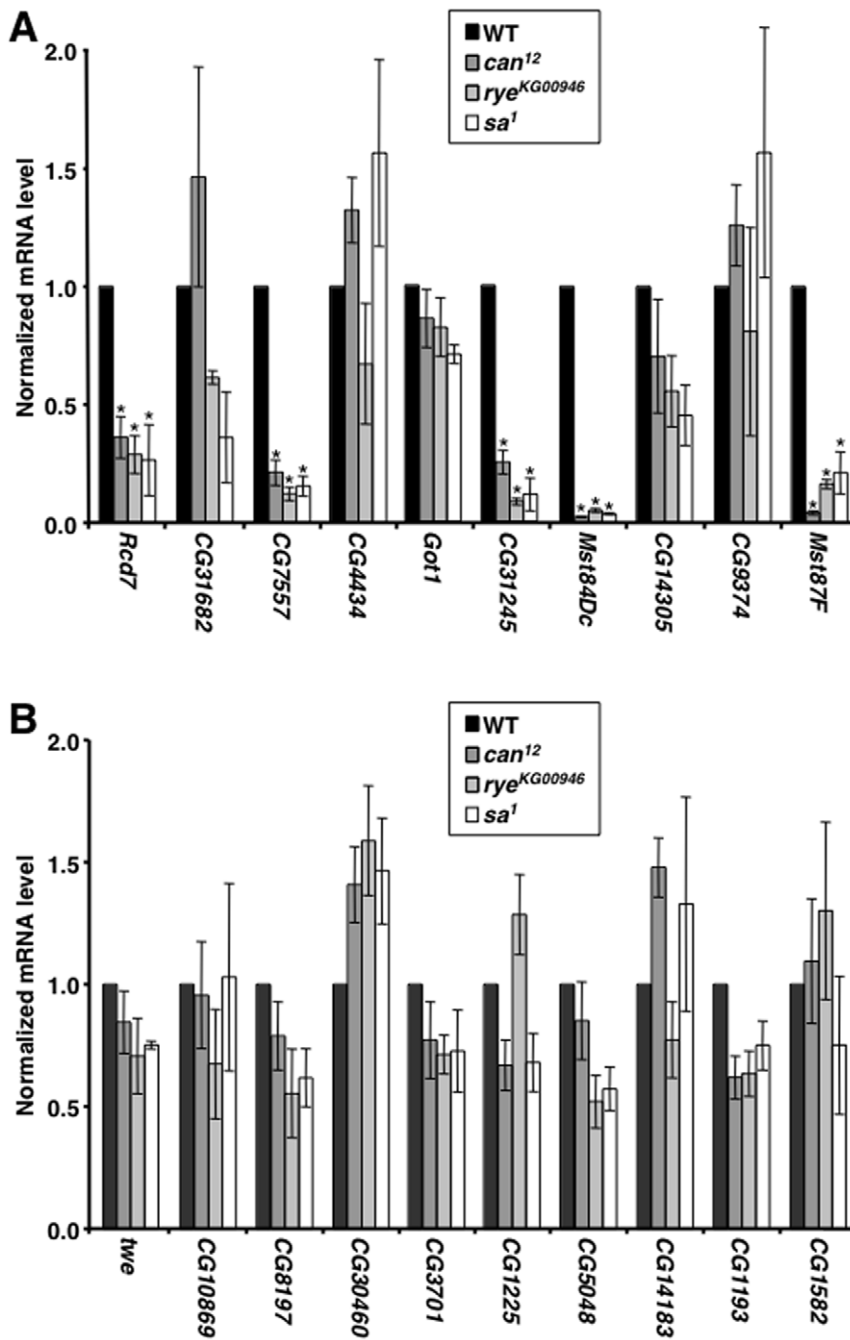
Different trans-acting factors presumably regulate the translational and transcriptional activities of the TCE. For genes specifically expressed in testes, the TCE functions as an RNA element in the 5' untranslated region of mRNAs to repress translation and as a DNA element in the core promoter of genes to activate transcription. Testis extracts are competent to form RNA-protein and DNA-protein complexes with the TCE *in vitro*, but specific nucleic acid-binding factors have not been identified [30]. Data presented here has shed some light on the DNA-protein complex. The variable location of the TCE raises the possibility that the TCE is a short-range enhancer that is bound by a transcription activator protein rather than a core promoter element that is bound by a component of the basal transcription machinery. Indeed, the TCE is unlikely to serve as a binding site for tTFIID since all TCE-containing genes did not require tTAFs for high-level transcription. As a short-range enhancer, the TCE may function analogously to OVO binding sites that regulate the transcription of genes in the ovary [40]. The lack of a strict requirement for tTFIID predicts the existence of a location-specific core promoter element that is recognized by tTFIID and functions in partnership with TCE at a subset of TCE-containing genes to determine testis-specific transcription. The predicted element is likely to only function in partnership with the TCE since tTAF mutations did not affect the transcription of genes that lacked a TCE. Additionally, the predicted element is likely to be of low sequence complexity since it was not identified in the bioinformatic analyses that minimally required a motif size of 5 nt. Finally, it is likely that transcription of TCE-containing genes is dependent on epigenetic features of the genome, such as nucleosome occupancy and histone modification [41,42]. Thus, the TCE is a component of a combinatorial control mechanism that determines testis-specific transcription.

## Materials and Methods

### Bioinformatic analyses

Methods for Expressed Sequence Tag (EST) filtering and clustering and TSS identification are described in Rach *et al.* (2009) [9]. Briefly, TSSs were called based on multiple 5' ESTs that were consistent with the gene annotation, upstream of annotated start codon(s), and whose 5' ends clusters in short genomic windows. Condition-specific assignments were made on a per promoter basis by evaluating the Shannon entropy for tag frequencies from the individual libraries.

Testis elements were identified by MEME v3.0.14 run on 100 nt long core promoter sequences, using the “zero or one occurrence” parameter and a motif size of 5–15 nt [24]. Additionally, BioProspector, a Gibbs sampling based motif finder, was used, as provided in the BEST suite [25]. The motif width was set to 8 nt, and sampling was repeated 10 times and allowed to



**Figure 5. tFAs regulate the transcription of a subset of TCE-containing genes.** Graphed is the mRNA expression level for the indicated genes in testes of wildtype (WT) and mutant flies. Expression levels were normalized to *actin* and to the expression level in wildtype flies. Asterisks indicate significant ( $P < 0.05$ ) changes in expression level in mutant relative to wildtype flies as determined by one-way ANOVA in conjunction with Bonferroni post-tests. (A) Graphed is the relative transcription level of genes with focused initiation patterns that contain a TCE in the core promoter. (B) Graphed is the relative transcription level of genes with focused initiation patterns that do not contain a TCE in the core promoter. *twe* is specifically transcribed in testes and lacks a TCE but was not included in the 190 genes that are specifically expressed in testes. doi:10.1371/journal.pone.0045009.g005

return up to 25 motifs. Both motif finders were provided with non-testis promoters to estimate background nucleotide frequencies, and reported motifs were required to occur in at least 20 sequences. MEME searches resulted in two significant hits, TE1 ( $E < 10^{-42}$ ) and TE2 ( $E < 10^{-7}$ ). Gibbs sampling reported both elements in all 10 runs, with TE1 consistently found as the top element. TE1/2 was formed by concatenating positions 1–3 from TE1 and positions 2–8 from TE2.

To identify motif matches in different sets of promoters, Patser was run with a P-value threshold of  $10^{-3}$ , using background nucleotide frequencies obtained from promoters not specific for any particular condition [27]. Position frequency matrices are presented in Tables S1 and S2. *Mst87F* was identified in an earlier iteration of the analysis that used a larger set of promoters and is not included in the set of 190 genes that are specifically expressed in testes.



The extent of motif conservation was evaluated using the protocol in Rach *et al.* (2009) [9]. Orthologous promoters were taken from the study describing the sequencing and analysis of 12 related *Drosophila* species [43]. Only promoters with identifiable alignments across all 12 species were analyzed further, comprising a set of 137 testis promoters. Motif conservation was evaluated in a pairwise manner by assessing the presence of co-occurring matches in *D. melanogaster* and one of the other 11 species. Matches to motifs were first determined independently in the promoters of each species and then counted as conserved if they occurred within  $\pm 5$  nt of the aligned position corresponding to the match in *D. melanogaster*.

### Transgenic flies

To generate TCE-containing transgenes, short genomic regions containing a TCE, as specified in the text, were amplified from *Drosophila melanogaster* genomic DNA by PCR and cloned into the *pattB-LacZ* plasmid [32]. TCE mutant and rescue transgenes were generated by site-directed mutagenesis using the QuickChange kit (Stratagene). Transgenic flies were generated by Rainbow Transgenics. Flies were maintained on standard cornmeal/molasses medium at 25°C. Molecular analyses were carried out on flies that were homozygous for the transgenes.  $\beta$ -gal staining of transgenic flies was performed as described by Glaser *et al.* (1986) [44].

### qPCR

For mRNA quantitation, total RNA was collected from 0–4 d whole flies or fly tissues or S2 cells, reversed transcribed, and analyzed by real-time PCR as described by Katzenberger *et al.*

### References

- Juven-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcription machinery. *Dev Biol* 339: 225–229.
- Müller F, Zaucker A, Tora L (2010) Developmental regulation of transcription initiation: more than just changing the actors. *Curr Opin Genet Dev* 20: 533–540.
- Ohler U, Wassarman DA (2010) Promoting developmental transcription. *Development* 137: 15–26.
- Ohler U, Liao G, Niemann H, Rubin GM (2002) Computational analysis of core promoters in the *Drosophila* genome. *Genome Biology* 3: research0087.1–0087.12.
- FitzGerald PC, Sturgill D, Shyakhthenko A, Oliver B, Vinson C (2006) Comparative genomics of *Drosophila* and human core promoters. *Genome Biology* 7: R53.
- Gershenson NI, Trifonov EN, Ioshikher IP (2006) The features of *Drosophila* core promoters revealed by statistical analyses. *BMC Genomics* 7: 161.
- Down TA, Bergman CM, Su J, Hubbard TJP (2007) Large-scale discovery of promoter motifs in *Drosophila melanogaster*. *PLoS Computational Biol* 3: e7.
- Lim CY, Santoso B, Boulay T, Dong E, Ohler U, *et al.* (2004) The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev* 18: 1606–1617.
- Rach EA, Yuan HY, Majoros WH, Tomancak P, Ohler U (2009) Motif composition, conservation and condition-specificity of single and alternative transcription start sites in the *Drosophila* genome. *Genome Biol* 10: R73.
- Hochheimer A, Zhou S, Zheng S, Holmes MC, Tjian R (2002) TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. *Nature* 420: 429–445.
- Goodrich JA, Tjian R (2010) Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet* 8: 549–558.
- Tora L (2002) A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev* 16: 673–675.
- Müller F, Tora L (2004) The multicoloured world of promoter recognition complexes. *EMBO J* 23: 2–8.
- Jacobi UG, Akkers RC, Pierson ES, Weeks DL, Dagle JM, *et al.* (2007) TBP paralogs accommodate metazoan- and vertebrate-specific developmental gene regulation. *EMBO J* 26: 3900–3909.
- Freiman RN (2009) Specific variants of general transcription factors regulate germ cell development in diverse organisms. *Biochim Biophys Acta* 1789: 161–166.

(2006) [45]. Sequences of qPCR primers are provided in Table S4. Experiments were performed for four independent samples. Real-time data was analyzed using the  $2^{-\Delta C_t}$  method [46]. Unpaired *t* tests were performed using Prism 4.0c (Graphpad Software).

### Supporting Information

**Table S1 Position frequency matrix for TE1.**  
(DOC)

**Table S2 Position frequency matrix for TE2.**  
(DOC)

**Table S3 qPCR of testis-specific genes relative to *actin*.**  
(DOC)

**Table S4 Primer sequences (5'-to-3').**  
(DOC)

### Acknowledgments

We thank Stein Aerts and Bassem Hassan for conducting preliminary analyses that showed that *Drosophila* testis-specific core promoters are A/T-rich, Johannes Bischof and Konrad Basler for providing phiC31 plasmids, Margaret Fuller for providing tTAF flies, and Nicole Bertram for comments on the manuscript. We also thank an anonymous reviewer for pointing out that TE1/2 is identical to the TCE.

### Author Contributions

Conceived and designed the experiments: EAR AKA UO DAW. Performed the experiments: RJK EAR AKA UO. Analyzed the data: RJK EAR AKA UO DAW. Contributed reagents/materials/analysis tools: RJK EAR AKA UO. Wrote the paper: EAR AKA UO DAW.

- Parisi M, Nuttall R, Edwards P, Minor J, Naiman D, *et al.* (2004) A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. *Genome Biol* 5: R40.
- Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, *et al.* (2011) The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471: 473–479.
- Renkawitz-Pohl R, Hempel L, Hollmann M, Schäfer MA (2005) Spermatogenesis. In: *Comprehensive Molecular Insect Science*, eds. L. I. Gilbert, K. Iatrou and S. S. Gill, New York: Elsevier BV, 157–178.
- White-Cooper H (2010) Molecular mechanism of gene regulation during *Drosophila* spermatogenesis. *Reproduction* 139: 11–21.
- Hiller MA, Lin T-Y, Wood C, Fuller MT (2001) Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes Dev* 15: 1021–1030.
- Hiller M, Chen X, Pringle MJ, Suchorski M, Sancak Y, *et al.* (2004) Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development* 131: 5297–5308.
- Chen X, Hiller M, Sancak Y, Fuller MT (2005) Tissue-specific TAFs counteract polycomb to turn on terminal differentiation. *Science* 310: 869–872.
- Metcalf CE, Wassarman DA (2007) Nucleolar colocalization of TAF1 and testis-specific TAFs during *Drosophila* spermatogenesis. *Dev Dyn* 236: 2836–2843.
- Bailey TL, Elkan C (1995) The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol* 3: 21–29.
- Che D, Jensen S, Cai L, Liu JS (2005) BEST: binding-site estimation suite of tools. *Bioinformatics* 21: 2909–2911.
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14: 1188–1190.
- Hertz GZ, Stormo GD (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* 15: 563–577.
- Ni T, Corcoran DL, Rach EA, Song S, Spana EP, *et al.* (2010) A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat Methods* 7: 521–527.
- Schäfer M, Kuhn R, Bosse R, Schäfer U (1990) A conserved element in the leader mediates post-meiotic translation as well as cytoplasmic polyadenylation of a *Drosophila* spermatocyte mRNA. *EMBO J* 9: 4519–4525.
- Kempe E, Muhs B, Schäfer M (1993) Gene regulation in *Drosophila* spermatogenesis: Analysis of protein binding at the translational control element TCE. *Dev Genet* 14: 449–459.
- Schäfer M, Nayernia K, Engel W, Schäfer U (1995) Translational control in spermatogenesis. *Dev Biol* 172: 344–352.

32. Bischof J, Maeda RK, Hediger M, Karch F, Basler K (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci* 104: 3312–3317.
33. Kuhn R, Schafer U, Schafer M (1988) Cis-acting regions sufficient for spermatocyte-specific transcriptional and spermatid-specific translational control of the *Drosophila melanogaster* gene *mst(3)gl-9*. *EMBO J* 7: 447–454.
34. Honeycutt E, Gibson G (2004) Use of regression methods to identify motifs that modulate germline transcription in *Drosophila melanogaster*. *Genet Res* 83: 177–188.
35. Michiels F, Gasch A, Kaltschmidt B, Renkawitz-Pohl R (1989) A 14 bp promoter element directs the testis specificity of the *Drosophila*  $\beta 2$  tubulin gene. *EMBO J* 8: 1559–1565.
36. Santel A, Kaufmann J, Hyland R, Renkawitz-Pohl R (2000) The initiator element of the *Drosophila*  $\beta 2$  tubulin gene core promoter contributes to gene expression in vivo but is not required for male germ-cell specific expression. *Nucleic Acids Res* 28: 1439–1446.
37. Reddi PP, Flickinger CJ, Herr JC (1999) Round spermatid-specific transcription of the mouse SP-10 gene is mediated by a 294-base pair proximal promoter. *Biol Reprod* 61: 1256–1266.
38. Skawran B, Schubert S, Dechend F, Vervoorts J, Nayernia K, et al. (2005) Characterization of the human TSPY promoter. *Mol Cell Biochem* 276: 159–167.
39. Yamashita A, Goto N, Nishiguchi S, Shimada K, Yamanishi H, et al. (2008) Computational search for over-represented 8-mers within the 5'-regulatory regions of 634 mouse testis-specific genes. *Gene* 427: 93–98.
40. Lü J, Oliver B (2001) *Drosophila* OVO regulates ovarian tumor transcription by binding near the transcription start site. *Development* 128: 1671–1686.
41. Sassone-Corsi P (2002) Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* 296: 2176–2178.
42. DeJong J (2006) Basic mechanisms for the control of germ cell gene expression. *Gene* 366: 39–50.
43. *Drosophila* 12 Genomes Consortium, et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203–218.
44. Glaser RL, Wolfner MF, Lis JT (1986) Spatial and temporal pattern of hsp26 expression during normal development. *EMBO J* 5: 747–754.
45. Katzenberger RJ, Marengo MS, Wassarman DA (2006) ATM and ATR pathways signal alternative splicing of *Drosophila* TAF1 pre-mRNA in response to DNA damage. *Mol Cell Biol* 26: 9256–9267.
46. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta Ct}$  method. *Methods* 25: 402–408.