High Throughput Sequencing Identifies Misregulated Genes in the *Drosophila* Polypyrimidine Tract-Binding Protein (*hephaestus*) Mutant Defective in Spermatogenesis

Vinod Sridharan, Joseph Heimiller, Mark D. Robida, Ravinder Singh*

Department of Molecular, Cellular and Developmental Biology, University of Colorado at Boulder, Boulder, Colorado, United States of America

* rsingh@colorado.edu

Abstract

The *Drosophila* polypyrimidine tract-binding protein (dmPTB or *hephaestus*) plays an important role during spermatogenesis. The *heph²* mutation in this gene results in a specific defect in spermatogenesis, causing aberrant spermatid individualization and male sterility. However, the array of molecular defects in the mutant remains uncharacterized. Using an unbiased high throughput sequencing approach, we have identified transcripts that are misregulated in this mutant. Aberrant transcripts show altered expression levels, exon skipping, and alternative 5' ends. We independently verified these findings by reverse-transcription and polymerase chain reaction (RT-PCR) analysis. Our analysis shows misregulation of transcripts that have been connected to spermatogenesis, including components of the actomyosin cytoskeletal apparatus. We show, for example, that the *Myosin light chain 1* (*Mlc1*) transcript is aberrantly spliced. Furthermore, bioinformatics analysis reveals that *Mlc1* contains a high affinity binding site(s) for dmPTB and that the site is conserved in many *Drosophila* species. We discuss that *Mlc1* and other components of the actomyosin cytoskeletal apparatus offer important molecular links between the loss of dmPTB function and the observed developmental defect in spermatogenesis. This study provides the first comprehensive list of genes misregulated *in vivo* in the *heph²* mutant in *Drosophila* and offers insight into the role of dmPTB during spermatogenesis.

Introduction

Heterogeneous nuclear ribonucleoproteins (hnRNPs) are ubiquitously expressed and associate with primary transcripts. One of these RNA-binding proteins, the polypyrimidine-tract-binding protein (PTB), which is also referred to as hnRNP I, binds to polypyrimidine tracts. These binding sites contain UCUU and UUCU sequence motifs [1–5]. PTB has been linked to regulation of mRNA splicing, polyadenylation, translation, mRNA stability/degradation, and mRNA

localization (reviewed in [6–9]). Of the three human PTB genes (PTBP1, PTBP2, and PTBP3), PTBP1 is more widely expressed, and PTBP2 and PTBP3 show expression restricted to specific tissues [10].

There is one PTB homolog in *Drosophila melanogaster*, which is also known as *hephaestus* (*heph*). Varied phenotypes are associated with *Drosophila heph* mutants: embryonic lethality, sensory bristle and wing margin abnormalities [4, 11–13]. During wing development and embryogenesis, it has been implicated in modulating Notch signaling [11, 14–20]. During oogenesis, *heph* has been linked to *oskar* mRNA translational repression [21] and Grk signaling in the female germline [22].

We reported previously that a major *heph* or *dmPTB* transcript is expressed specifically in the male germline and that it correlates with male fertility in *Drosophila* [13]. Spermatogenesis, which is a complex process, shows remarkable cytological similarities between *Drosophila* and mammals: germ cell maintenance, mitotic divisions preceding meiosis, and morphological changes involving all cellular components (reviewed in [23]). Extensive genetic analysis of spermatogenesis has revealed multiple mutations that exhibit critical cytological blocks, for example, stem cell renewal, mitotic and meiotic amplification, and spermatid differentiation [23, 24]. In fact, *heph* was first identified in a genetic screen for male sterility [24]. Nonetheless, how relevant mutations influence specific stages during spermatogenesis had been a mystery. We showed that the dm*PTB* loss of function *heph²* mutant disrupts the cellular process of spermatid individualization, leading to male sterility [4]. Spermatid individualization represents the terminal step in spermatogenesis.

To understand the sex-specific biological role of dmPTB in male sterility, the question arises what are specific mRNAs that are misregulated in the *heph²* mutant? We undertook an unbiased high throughput sequencing approach to investigate the role of *heph* in *Drosophila* at a genome-wide level. We used RNA-Seq analysis of mRNAs from wild type and *heph²* mutant flies and show that loss of *heph* function results in misregulation of several specific transcripts, including mRNA splicing misregulation, that may provide a missing link(s) to its role in spermatogenesis.

Results

High throughput sequence analysis for transcriptome profiling

The molecular basis of the developmental defects in the *heph²* mutant is currently unknown. Therefore, we used high throughput sequencing to identify transcripts that are misregulated in *heph²* mutant flies. We performed the RNA-Seq protocol on polyadenylated RNA from wild-type control and *heph²* mutant adult flies. We obtained 57,738,593 sequence reads for the wild-type control and 44,791,181 sequence reads for the *heph²* mutant. Using TopHat [25], over 90% of the sequences could be mapped to the *Drosophila* genome and about 7% corresponded to splice junction reads ("Table 1").

The *heph²* mutant was generated via *P* element transposon mutagenesis [24]. Multiple sequence reads from the *heph²* mutant confirmed the insertion of the transposon *P* element in the *heph* locus. As shown in Fig 1, the *heph²* mutation disrupted synthesis of *dmPTB* transcripts in *heph²* homozygotes—a severe drop-off in mapped reads downstream of the *P* element insertion, which is located within a *heph* intron. These results show that the *heph²* mutation resulted in truncation of about 96% of the *dmPTB* transcripts. Considering full-

Table 1. Sequencing and mapping statistics for the *heph²* analysis.

Group	# of reads	# of spliced reads	% mapped to genome
heph2/TM3 adult males(control)	57,738,593	3,981,756 (7%)	91%
heph2/heph2 adult males	44,791,181	3,154,576 (7%)	92%

doi:10.1371/journal.pone.0150768.t001
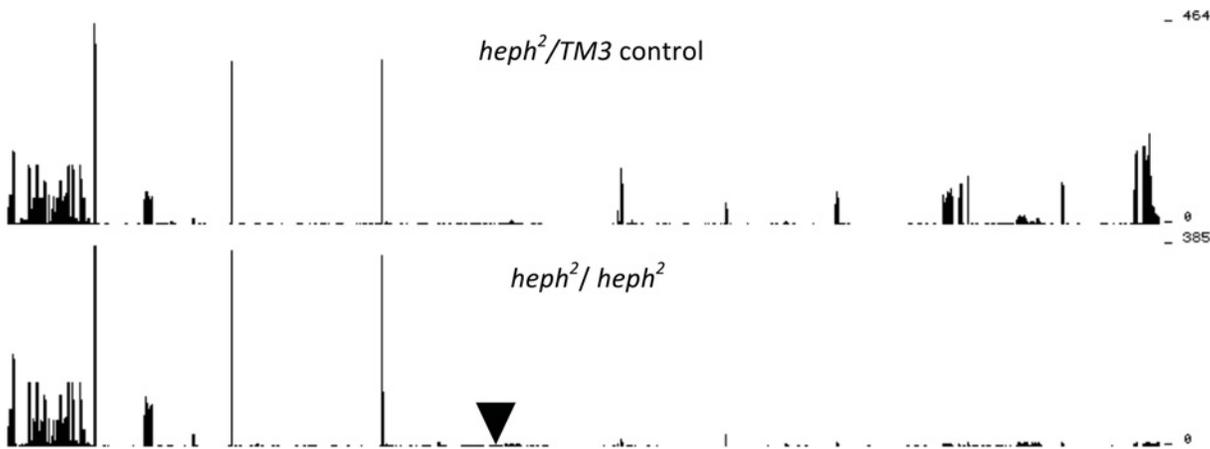
Fig 1. Loss of RNA-Seq coverage in the *heph²* homozygous mutant in the 3' portion of the *heph* gene, downstream of the *P* element insertion site. The control sample is shown on the top graph and the *heph²* homozygous mutant on the bottom graph. The inverted triangle represents the location of the *P* element insertion in an intron (intron has been condensed to save space). There are two transcription start sites for the *dmPTB or heph* locus, producing transcripts with proximal and distal 5'UTRs. The abundant male germline-specific transcript, which is disrupted in the *heph²* mutant, results from the use of the distal or upstream promoter. The *P* element in *heph²* is inserted into an intron corresponding to the transcript with distal 5' UTR. However, for the *heph^03429* mutant, showing embryonic lethality, the *P* element is inserted into an intron downstream of the proximal 5' transcript start site. This transcript start site is relevant for the non-sex-specific, including embryonic, function of *dmPTB*; no reads were found that spanned the other end of the *P* element insertion into the *heph* gene [26].

doi:10.1371/journal.pone.0150768.g001

length transcripts, the *heph²* homozygous mutants have roughly 4% of wild-type mRNA levels. Thus, transcriptome profiling establishes for the first time how *dmPTB* transcripts are disrupted in the *heph²* mutant.

Transcriptome profiling identifies mRNA expression level differences

We compared the two samples for expression levels of individual genes. In the *heph²* dataset, the Cufflinks software identified 493 genes as differentially expressed (Fig 2A, see Methods for details). There were numerous genes whose expression level changed over 2-fold (Fig 2A). Next, we asked if some functional gene ontology categories were over-represented in these differentially expressed genes. Using gene ontology analysis with DAVID (The Database for Annotation, Visualization and Integrated Discovery), we subjected to comparative analysis the above gene set that showed difference in expression levels between the wild-type control and the *heph²* mutant. Our analysis showed several over-expressed gene categories such as RNA metabolism, extracellular/secreted, hormones, and peptidases (Fig 2B and Table 2). Given that many genes play multiple cellular roles, further studies will be needed to establish whether and how the GO terms are related to the spermatogenesis phenotype. When we carefully inspected these genes, there were two differentially expressed genes that had been known to be involved in spermatogenesis: *Nc* (*Nedd2-like caspase* or *Dronc*) and *oxen* (*ox*) [24, 27, 28]. In the *heph²* mutant, *Dronc/Nc* levels were reduced to about 50% of wild-type levels, and *ox* levels were increased by 2.5 times compared to wild type (see Discussion for the functional relevance of these genes). Since *heph²* mutants cause male but not female sterility, we also compared genes differentially expressed in *heph²* to those expressed mostly in the male germline, but not in the female germline, using modEncode RNA Seq data [29]. The intersection of the male-specific genes (genes that were significantly higher in the adult male germline vs. the female germline) and the genes expressed in *heph²* yielded 185 genes (about 86 genes were expected by chance). None of these genes, however, was related to spermatid individualization based on prior annotation. Our results show significant changes in gene expression of specific genes in the *heph²* mutant.
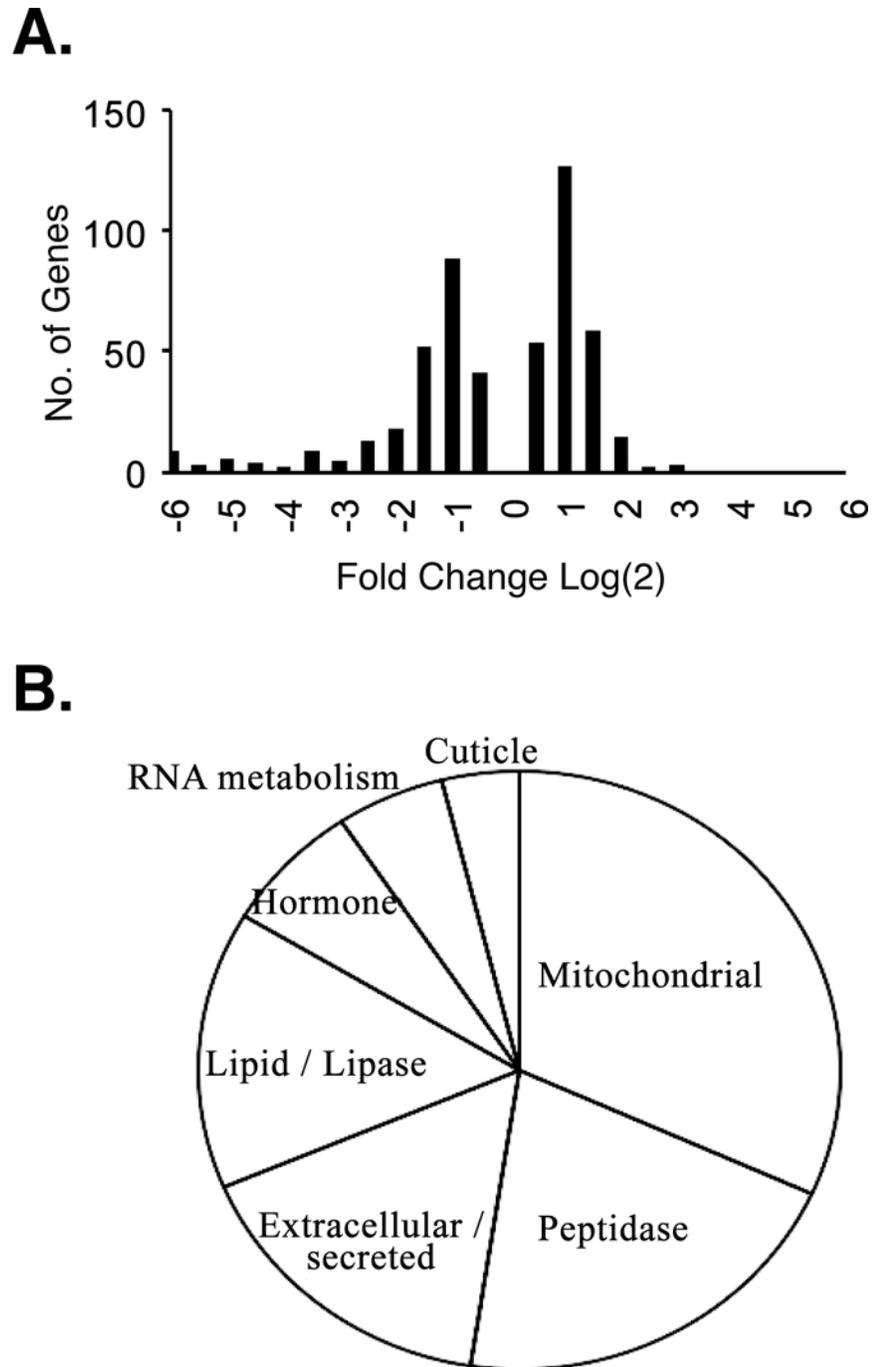
Fig 2. Effect of the *heph²* mutation on fold-changes in gene expression. Histogram of fold-changes in the *heph²* mutants of genes called significant by Cufflinks (note that Cufflinks did not call any genes significant below 0.78 log₂ change). The list is provided in S1 Table. (B). Pie chart representing gene ontology categories over-represented in the genes that are significantly affected in the *heph²* mutant.

doi:10.1371/journal.pone.0150768.g002

Transcriptome profiling identifies misregulated mRNA isoforms

Differences in mRNA isoforms could result from use of alternative transcription start sites (TSS), alternative 5' splice sites, alternative 3' splice sites, exon skipping, and alternative 3'

Table 2. Selected gene ontology categories from the *heph²* analysis.

GO category	DAVID—*P* value(Benjamini-adjusted)	Up or down-regulation
extracellular / secreted	1×10^{-15}	up/down
mitochondrial	2×10^{-14}	up/down
peptidase	1×10^{-8}	up/down
stress response	1×10^{-8}	up
signal peptide	3×10^{-6}	down
lipase	3×10^{-5}	up/down

The p-values for ontology groups are shown from the DAVID analysis.

doi:10.1371/journal.pone.0150768.t002

ends. Alternative isoform candidates in the *heph²* dataset were first identified using MISO (Mixture of ISOforms) [30]. While MISO called 35 alternative isoform candidates as significant, we found, upon manual inspection, that most of these candidates appeared to be weak, and thus were not pursued. Our experience revealed the limitation of this algorithm. However, during our inspection we observed significant isoform differences in several categories. Table 3 shows a list of the filtered candidates and observed fold-changes. Transcripts such as *Paramyosin* (*Prm*) and *Tequila* were examples of alternative 5' untranslated regions resulting from alternative transcription start sites (5.3 fold). CG43293 was an example of alternative 5' splice site (significant fold change: 27% vs. 0%). *Myosin alkali light chain 1* (*Mlc1*), *CG1674*, and *Aldolase* (*Ald*) were examples of alternative exons (2–4 fold). The transcripts above showed the most significant isoform differences in the *heph²* mutant. We conclude that the *heph²* mutation results in misregulation of specific mRNA isoforms.

RT-PCR analysis confirms expression level differences

We randomly selected several candidates from specific gene ontology categories for analysis by RT-PCR for their expression level differences between the wild-type control and the *heph²* mutant. Fig 3 shows that RT-PCR analysis recapitulated the up and down regulation observed from the Illumina sequencing reads. For example, transcripts such as *Act88F*, *TpnC4*, *Hsp70Bc*, *CG11162*, and *CG42245* were significantly upregulated in the *heph²* mutant. On the other hand, transcripts such as *CG11598* and *Vm34Ca* exhibited downregulation (Table 4 and S1 Table). We conclude that RT-PCR and high throughput sequencing independently confirm misregulation of specific transcripts in the *heph²* mutant.

RNA analysis confirms misregulated isoforms

Next, we analyzed the *Prm* and *Tequila* transcripts for their isoform differences. Indeed, RT-PCR analysis showed the presence of alternative isoforms for the *Tequila* and *Prm* genes

Table 3. *heph²* differentially expressed isoforms.

Gene	Alternative Isoform	Fold-change	Gene ontology
Paramyosin (*Prm*)	TSS	5.3	Structural constituent of muscle
Tequila	TSS	5.3	Serine-type endopeptidase
Myosin alkali light chain 1 (*Mlc1*)	Exon	2.1	Actin motor
CG1674	Exon	4.1	Unknown
Aldolase (*Ald*)	Exon	1.8	Fructose-bisphosphate aldolase
CG43293	5' SS	Large	Unknown

The candidates were manually filtered. The fold change was calculated based on values from control (%) versus *heph²* mutant (%), calculated from exon-exon spanning 'spliced' reads supporting each isoform. The Bayes factor/odds ratio for these candidate as called by MISO was 10^9 or better.

doi:10.1371/journal.pone.0150768.t003
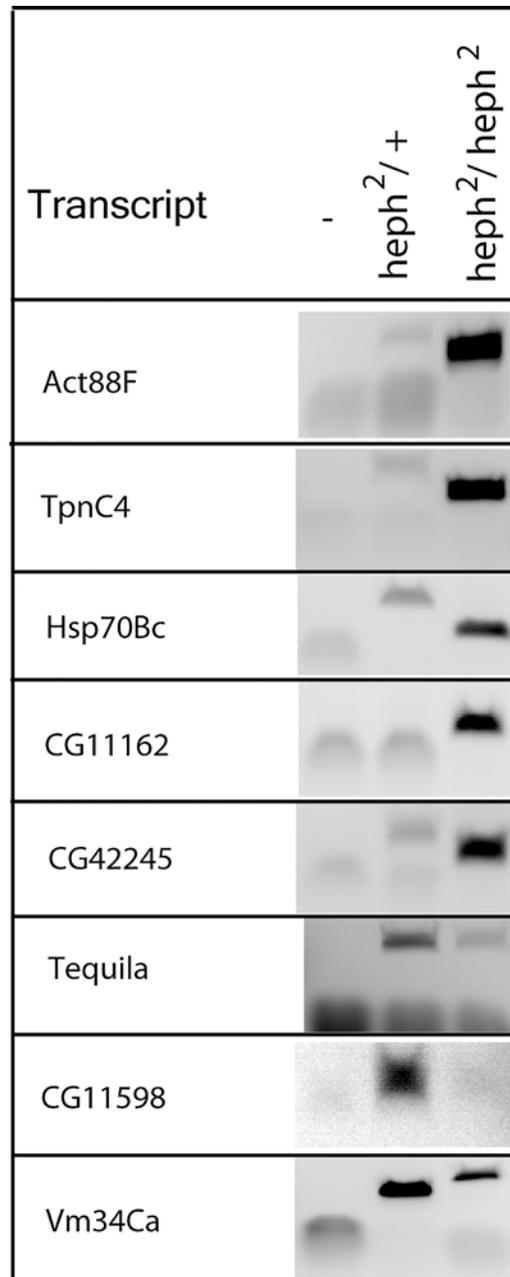
Fig 3. RT-PCR analysis reveals differentially expressed transcripts. RNA from the wild type control and the *heph²* mutant was analyzed by RT-PCR analysis. Gene-specific primer pairs were used for amplification of several randomly picked genes that showed significant differential expression. For *Tequila*, primer pairs were designed to test alternative 5' ends (transcription start site usage).

doi:10.1371/journal.pone.0150768.g003

(Figs 3 and 4A and 4B; data not shown), validating the analysis of high throughput sequencing. *Myosin light chain 1* (*Mlc1*) was an example of altered splicing pattern in the *heph²* mutant (Fig 5A). We analyzed *Mlc1* expression pattern by RT-PCR, and found that it recapitulated the quantitative difference in the *Mlc1* skipped exon in the mutant (Fig 5B). We also analyzed embryonic mRNA from another embryonic lethal allele *heph*^03429, which did not produce exon skipping. Thus, the altered splicing pattern observed here was specific to the *heph²* mutation

Table 4. Fold changes for selected candidates (ratio of RNA-Seq reads—mutant versus control).

Gene	Fold-change	Up or Down regulation
Act88F	+4.6	Up
TpnC4	+5.2	Up
Hsp70Bc	+10.5	Up
CG11162	+20.1	Up
CG42245	+79.8	Up
Tequila	-1.45	Down
CG11598	-32.1	Down
Vm34Ca	-88.1	Down

doi:10.1371/journal.pone.0150768.t004

that causes male sterility. We also excised the two bands from the gel and confirmed exon inclusion and exclusion in *Mlc1* by direct sequencing. These findings show that loss of *dmPTB* function in the *heph²* mutant results in significant differences in mRNA isoforms for specific transcripts.
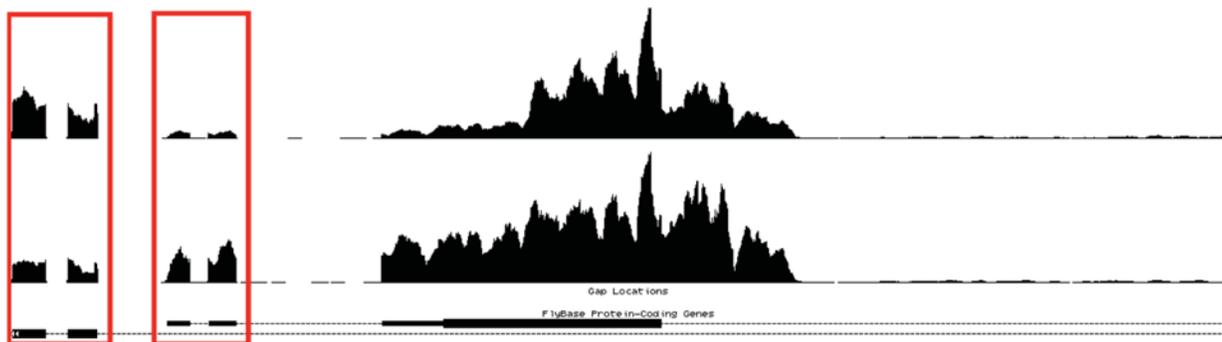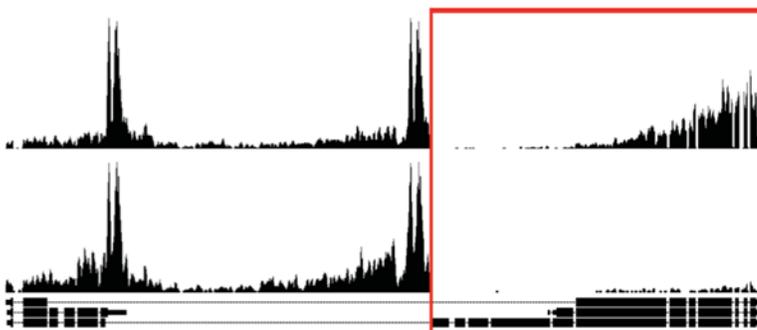
Fig 4. Use of alternative transcription start sites in the *heph²* mutant. (A) RNA Seq read pileup across a *Paramyosin* alternative transcription start site, where alternative transcription start sites highlighted by the box. **(B)** RNA Seq read pileup across a *Tequila* alternative transcription start site, where the 5' end showing the most difference is highlighted by the box. For RT-PCR validation of *Tequila* alternative transcription start site usage, see Fig 3.

doi:10.1371/journal.pone.0150768.g004

A. *Myosin light chain 1 (Mlc1)*
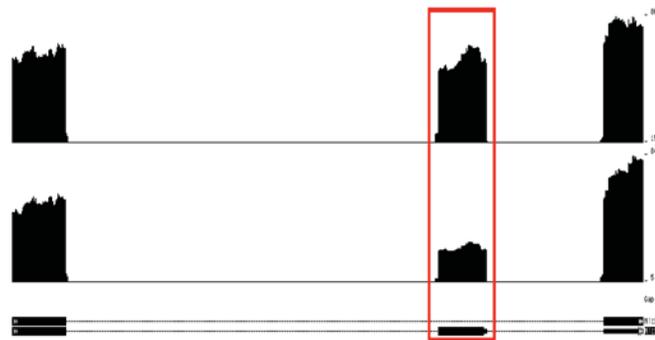
B.

Alternative Exon inclusion
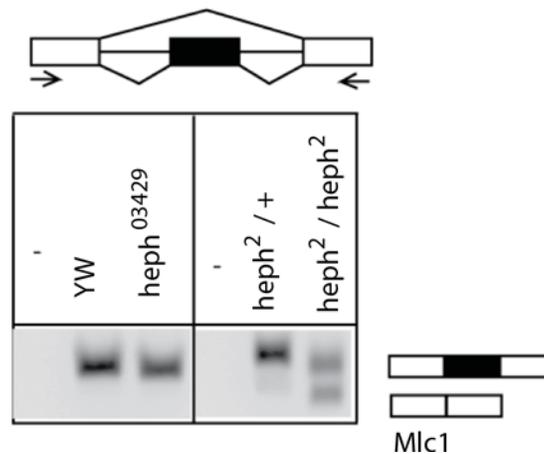
Mlc1

Fig 5. Alternative splicing of the *Mlc1* gene in the *heph²* mutant. RNA Seq pileups were generated using UCSC Genome Browser, y-axis is auto-scaled to show differences in isoform fractions, alternative exon is highlighted by the box. **(A)** RNA-Seq read pileup across the *Mlc1* skipped exon in control (top) and *heph²* mutant (bottom) flies, showing significantly increased exon skipped in the *heph²* mutant. **(B)** RT-PCR analysis of *Mlc1* for exon skipping, using primers in the flanking exons shown by the arrows.

doi:10.1371/journal.pone.0150768.g005

A PTB binding site is phylogenetically conserved in Mlc1

We previously showed that the *heph²* mutation disrupted spermatid individualization, an acto-myosin cytoskeletal-related process [4, 31]. Thus, *Mlc1* was an attractive candidate for which misregulation could contribute to the observed male sterility phenotype of the *heph²* mutant. To determine if there was a dmPTB binding site in this transcript we analyzed sequences in the vicinity of the altered exon use in *Mlc1* from 12 *Drosophila* species. We found that the skipped *Mlc1* exon contained an extended C/U-rich site and that this site was conserved in 12 different Drosophila species ([Fig 6](#)). In fact, this C/U-rich sequence is very similar to the PTB consensus binding site obtained through the SELEX enrichment of RNAs that bound PTB with high affinity [4, 5]. The expression of other myosin light chains in *Drosophila* including *Androcam*, *Calmodulin*, *Mlc2*, and *Mlc-c* was not significantly affected in our analysis of the *heph²* mutant. We conclude that the *Mlc1* transcript contains a conserved high affinity PTB binding site.
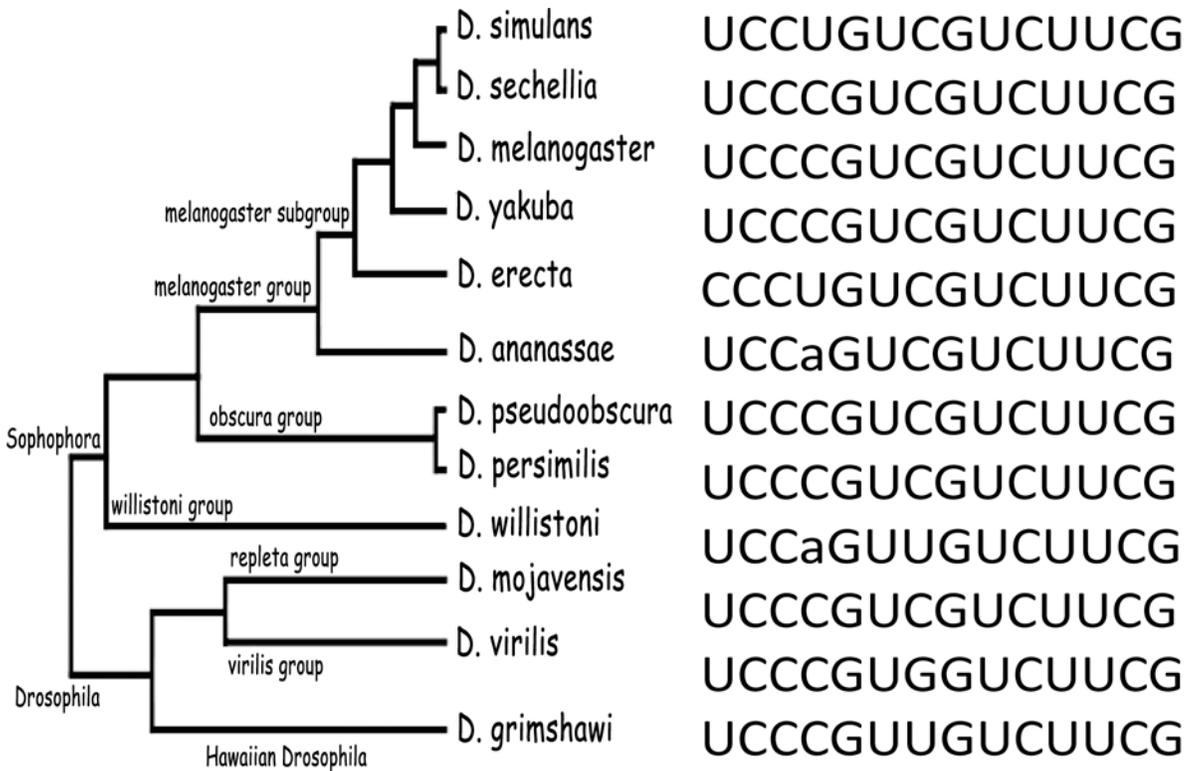
Fig 6. Identification of a conserved PTB binding site in *Mlc1*. Putative PTB binding site in *Mlc1* conserved in 12 Drosophila species is shown. The highest affinity SELEX sequence that binds PTB *in vitro* is shown for comparison.

doi:10.1371/journal.pone.0150768.g006

Discussion

We used a high throughput sequencing approach for transcriptome profiling and have identified transcripts that are misregulated in the *heph²* mutant. These include transcripts that show differences in expression levels (up or down regulation) and mRNA isoforms (with alternative 5' ends or exons) and some of them have been linked to spermatogenesis. The candidates revealed by the genome-wide analysis have been independently validated for their misregulation. Finally, phylogenetic comparison among *Drosophila* species reveals a conserved high affinity PTB binding site near a misregulated splice site in the *Mlc1* gene. These findings are consistent with the idea that the regulatory RNA binding protein dmPTB likely affects many molecular functions *in vivo*.

We chose to study the *heph²* allele because it shows a specific developmental defect during spermatogenesis, leading to male sterility. In contrast to the sex-specific expression of the major *dmPTB* transcript in *Drosophila* [13], the human PTB is ubiquitously expressed. Our high throughput analysis shows that in the *heph²* mutant *dmPTB* reads are absent past the site of transposon insertion. This lack of *dmPTB* expression manifests in significant misregulation of several specific downstream genes.

Among the misregulated transcripts revealed here, several are excellent candidates for their role in the spermatogenesis process, and thus provide connections to how loss of dmPTB function can lead to known spermatogenesis defects or some other male-specific defect that remains to be characterized. For example, *Dronc/Nc*, which is misregulated in the *heph²*

mutant, has been previously shown to be involved in one of the caspase activation pathways necessary for spermatid individualization [28]. This previous study elegantly showed that multiple caspases and caspase regulators, involving ARK- and HID- dependent activation of DRONC at sites of spermatid individualization, are required for this non-apoptotic process of spermatid individualization. Similarly, *oxen*, another misregulated transcript in the *heph²* mutant, was previously uncovered in a genetic screen that was designed to identify male-sterile mutants [24]. In fact, this same genetic screen also revealed the *heph²* mutant. Thus, our observation is consistent with a possible functional relationship between *heph* and *oxen*. We argue that a quantitative differences in *Dronc/Nc* and *ox* transcripts in the *heph²* mutant are biologically relevant because, based on our previous observation, quantitative reduction in dmPTB levels correlates well with the severity of the male-sterile phenotype [4]. These observations on *Dronc/Nc* and *ox* misregulation strongly suggest regulatory relationships, acting in the same developmental pathway–spermatogenesis.

We previously showed that adult *heph²* testes contain all spermatogenesis stages up to the elongated spermatids but no motile sperm [13] and accumulate cysts of elongated spermatids with numerous bulges along their length [4], which is characteristic of defects in the spermatid individualization process [32]. In fact, when we specifically examined, with DAPI or phalloidin staining, the integrity of the individualization complex (IC), which is characterized by a tightly cross-linked actin cytoskeletal cone, this complex was disrupted in the *heph²* mutant [4]. Given that this terminal step of spermatogenesis involves an actomyosin cytoskeletal apparatus or directional F-actin regulation, the question arises what components or regulators of this actomyosin cyskeletal apparatus are involved. To our knowledge, these candidates have not been experimentally shown to display male sterile defects. We speculate that *Act88F*, *Prm*, *TpnC4*, and/or *Myosin light chain 1*, which have emerged as candidates for regulation by dmPTB in our genome-wide screen, are likely important players in this process and their misregulation could in part account for male sterility in general and the disrupted IC in particular. In regards to other misregulated genes reported here, although UCUU and UUCU sequence motifs [3] can be found in most genes, we have been unable to find an obvious high affinity PTB binding similar to the one shown in Fig 6 [4, 5]. It is an open question whether these genes are further downstream or contain a functional PTB-binding site that remains to be identified. Future studies should distinguish between these possibilities. Our analysis has also revealed misregulated genes that have not been linked to spermatogenesis yet.

We note that all of the alternative isoform candidates were confirmed to be expressed in the testis using RNA-Seq data from the modEncode project [29]; the testis isoform expression pattern matched with that of the *heph²/TM3* control and not the *heph²* homozygotes. We also compared the differentially expressed genes in *heph²* and *heph⁰³⁴²⁹*. There was little overlap, beyond the number of genes that would be expected by chance, between the differentially expressed genes in the *heph²* and *heph⁰³⁴²⁹* analyses (S1 Fig). The *heph⁰³⁴²⁹* mutation was previously linked to misregulation of *Notch* in embryos [11, 17–19]. While Notch plays a role in spermatogenesis [33], whether it functions during spermatid individualization, which is defective in the *heph²* mutant [4], remains to be established.

A prominent role of sex-specific RNA expression and splicing has been extensively studied in *Drosophila melanogaster*, where a series of alternative splicing events (*Sxl → tra → dsx*) leads to somatic sexual differentiation [34]. In *Caenorhabditis elegans*, a cascade of regulatory events or repression at the level of transcription or translation mediates sexual differentiation [35]. In mammals, including mice and humans, the key regulator gene *SRY* on the Y chromosome leads to testes differentiation, which produces testosterone and masculinizes all somatic tissues. In fact, the *dsx* ortholog is conserved in *C. elegans* (*mab-3*) and humans (*DMRT*) [36, 37]. Recent, high throughput analysis clearly shows abundance of sex-specific RNA expression

differences in *Drosophila* and humans, including pre-gonadal tissues such as brain [38–43]. Recognition of sex-specific RNA differences in humans and vertebrates has important implications in normal development and clinical aspects of various diseases such as neurological disorders that show sex-biased disease development, pathology, and recovery (reviewed in [44]). While sex-specific alternative mRNA processing is well established in *Drosophila melanogaster*, environmental, genetic, and inherited epigenetic mechanisms (chromatin modifications and remodeling), non-coding RNA, and RNA and DNA editing are emerging as important additional regulatory players in sex-specific gene expression, leading to sexual dimorphism in mammals and vertebrates.

In addition to RNA-Seq [26] and the phylogenetic sequence analysis used here, several other approaches have been employed in other studies of RNA-protein interaction: yeast-three-hybrid assay [45]; CLIP [46]; Fast-FIND (*Fast-Fully Indexed Nucleotide Database*) [47]; PAR-CliP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) [48]; and RIP-Seq (or RIP-ChIP). As noted above, the short, degenerate (pyrimidine-rich sequence) PTB binding site appears frequently in the genome and makes identification of RNA-protein interactions that are functionally relevant even more difficult. Our analysis has provided a list of candidates for future functional analysis and a useful dataset that can be combined with other genome-wide studies to obtain a more complete list of misregulated transcripts. Follow-up molecular, genetic, and biochemical analysis should provide better understanding of the molecular basis for spermatogenesis.

Materials and Methods

Fly stocks and RNA preparation

The *heph²* stock was obtained from the Bloomington Stock Center. Flies were raised on standard cornmeal food at 17–25°C. Poly(A)⁺ RNA from 1 microgram total RNA was obtained from the *heph²* mutant and wild-type control (*heph²* / *TM3*) whole adult homozygous *heph²* males aged 1–3 days after eclosion, and mRNA-seq sequencing libraries (standard Illumina TruSeq) were prepared according to manufacturer's instructions (Illumina, Inc., San Diego CA, USA). The *heph²* mutant and wild-type control libraries were sequenced on an Illumina HiSeq 2000 sequencer (singleton 50 base-pair reads).

RNA Seq: alternative isoform analysis

TopHat 2.0.4 was used to map reads to the Flybase *Drosophila melanogaster* r5.44 genome annotation (GTF and Fasta files), with the—no-novel-juncs and—microexon-search parameters (mapping statistics are shown in Table 1) [25]. Alternative splicing analysis was first attempted using the MISO [30] software package using event-specific analysis (as opposed to whole-isoform analysis). Alternative splicing candidates were called using an isoform fraction difference threshold of 0.2 and a minimum Bayes Factor of 10 (at least 10-to-1 odds, or 0.0909 probability, that the alternative event call was not due to chance alone). Most of the MISO alternative isoform candidates, however, showed little observable difference upon manual inspection of the reads across alternative junctions and so were not pursued. Cufflinks did not call any alternative splicing or alternative transcription start sites as significant, and so was not used for isoform-specific analysis [49]. We also used an alternative isoform algorithm which called ratios of exon-exon spanning reads supporting each isoform, using the probability mass function of the binomial distribution to estimate a p-value (which was then converted into a log-odds value). De novo prediction of new splice sites using TopHat did not yield any significant differential isoforms, so only annotated splice junctions were used in this analysis. Traces

were generated using the UCSC genome browser/tools. For additional methods details on RNA-Seq expression and isoform analysis, see [26].

RNA Seq: differential gene expression analysis

Cufflinks version 2.0.0 was used to determine significant overall gene expression in the *heph²* mutant analyses. Cufflinks did not call any gene significant under an absolute fold-change value of 0.781 in the *heph²* analysis.

Search for PTB sites

Putative PTB sites were searched through the UCSC 15 species insect alignment, which includes 12 Drosophila species (ftp://hgdownload.cse.ucsc.edu/goldenPath/dm3/multiz15way/). A PTB binding site was defined as a stretch of at least nine C or U residues allowing at most one G, conserved within 25 nucleotides (not including indels) in at least 9 species in the 15-species genomic alignment. Conserved PTB sites were also searched manually around alternative splice sites in the UCSC genome browser.

RT-PCR validation of candidates

Alternative splicing candidates were validated using RT-PCR primers that spanned the alternative transcriptional area, using one PCR primer set to span both alternative isoforms to yield two product sizes. Alternative transcription site candidates were validated using an identical reverse primer and separate forward primers. RpS15 (a ribosomal protein gene) RT-PCR was used as a loading control with primers designed over a constitutive intron boundary to rule out the presence of genomic DNA. PCR primers used are shown in S1 Table.

Gene Ontology analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 was used for gene ontology analysis online. The pie chart (in Fig 2) represents the genes in only the most overrepresented ontology categories.

Supporting Information

S1 Fig. A comparison of sequence reads from the *heph²* and *heph^03429* mutants and primer sequences are provided.
(PDF)

S1 Table. Analysis of sequence reads between control and *heph²*.
(XLSX)

Author Contributions

Conceived and designed the experiments: VS JH RS. Performed the experiments: VS JH. Analyzed the data: VS JH RS. Contributed reagents/materials/analysis tools: MDR. Wrote the paper: VS JH MDR RS.

References

1. Clerte C, Hall KB. Characterization of multimeric complexes formed by the human PTB1 protein on RNA. RNA. 2006; 12(3):457–75. PMID: 16431980.

2. Lin CH, Patton JG. Regulation of alternative 3' splice site selection by constitutive splicing factors. RNA. 1995; 1(3):234–45. PMID: 7489496

3. Perez I, Lin CH, McAfee JG, Patton JG. Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. RNA. 1997; 3(7):764–78. PMID: 9214659

4. Robida M, Sridharan V, Morgan S, Rao T, Singh R. Drosophila polypyrimidine tract-binding protein is necessary for spermatid individualization. Proc Natl Acad Sci U S A. 2010; doi: 10.1073/pnas.1007935107 PMID: 20616016.

5. Singh R, Valcarcel J, Green MR. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. Science. 1995; 268(5214):1173–6. PMID: 7761834

6. Kafasla P, Mickleburgh I, Llorian M, Coelho M, Gooding C, Cherny D, et al. Defining the roles and interactions of PTB. Biochemical Society transactions. 2012; 40(4):815–20. PMID: 22817740. doi: 10.1042/BST20120044

7. Sawicka K, Bushell M, Spriggs KA, Willis AE. Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. Biochemical Society transactions. 2008; 36(Pt 4):641–7. PMID: 18631133. doi: 10.1042/BST0360641

8. Valcarcel J, Gebauer F. Post-transcriptional regulation: the dawn of PTB. Curr Biol. 1997; 7(11):R705–R8. PMID: 9382788

9. Wagner EJ, Garcia-Blanco MA. Polypyrimidine tract binding protein antagonizes exon definition. Mol Cell Biol. 2001; 21(10):3281–8. PMID: 11313454.

10. Coutinho-Mansfield GC, Xue Y, Zhang Y, Fu XD. PTB/nPTB switch: a post-transcriptional mechanism for programming neuronal differentiation. Genes & development. 2007; 21(13):1573–7. PMID: 17606635.

11. Dansereau DA, Lunke MD, Finkielsztein A, Russell MA, Brook WJ. Hephaestus encodes a polypyrimidine tract binding protein that regulates Notch signalling during wing development in Drosophila melanogaster. Development. 2002; 129(24):5553–66. PMID: 12421697.

12. Davis MB, Sun W, Standiford DM. Lineage-specific expression of polypyrimidine tract binding protein (PTB) in Drosophila embryos. Mech Dev. 2002; 111(1–2):143–7. PMID: 11804786.

13. Robida MD, Singh R. Drosophila polypyrimidine-tract binding protein (PTB) functions specifically in the male germline. EMBO J. 2003; 22(12):2924–33. PMID: 12805208.

14. Cruz C, Glavic A, Casado M, de Celis JF. A gain-of-function screen identifying genes required for growth and pattern formation of the Drosophila melanogaster wing. Genetics. 2009; 183(3):1005–26. PMID: 19737745. doi: 10.1534/genetics.109.107748

15. Kankel MW, Hurlbut GD, Upadhyay G, Yajnik V, Yedvobnick B, Artavanis-Tsakonas S. Investigating the genetic circuitry of mastermind in Drosophila, a notch signal effector. Genetics. 2007; 177(4):2493–505. PMID: 18073442.

16. Norga KK, Gurganus MC, Dilda CL, Yamamoto A, Lyman RF, Patel PH, et al. Quantitative analysis of bristle number in Drosophila mutants identifies genes involved in neural development. Curr Biol. 2003; 13(16):1388–96. PMID: 12932322.

17. Shepherd A, Wesley U, Wesley C. Notch and delta mRNAs in early-stage and mid-stage drosophila embryos exhibit complementary patterns of protein-producing potentials. Dev Dyn. 2010; 239(4):1220–33. PMID: 20201103. doi: 10.1002/dvdy.22262

18. Shepherd AK, Singh R, Wesley CS. Notch mRNA expression in Drosophila embryos is negatively regulated at the level of mRNA 3' processing. PLoS One. 2009; 4(11):e8063. PMID: 19956683. doi: 10.1371/journal.pone.0008063

19. Wesley CS, Guo H, Chaudhry KA, Thali MJ, Yin JC, Clason T, et al. Loss of PTB or negative regulation of Notch mRNA reveals distinct zones of Notch and actin protein accumulation in Drosophila embryo. PLoS One. 2011; 6(7):e21876. PMID: 21750738. doi: 10.1371/journal.pone.0021876

20. Tremmel DM, Resad S, Little CJ, Wesley CS. Notch and PKC are involved in formation of the lateral region of the dorso-ventral axis in Drosophila embryos. PLoS One. 2013; 8(7):e67789. doi: 10.1371/journal.pone.0067789 PMID: 23861806; PubMed Central PMCID: PMCPMC3701627.

21. Besse F, Lopez de Quinto S, Marchand V, Trucco A, Ephrussi A. Drosophila PTB promotes formation of high-order RNP particles and represses oskar translation. Genes & development. 2009; 23(2):195–207. PMID: 19131435.

22. McDermott SM, Davis I. Drosophila hephaestus/polypyrimidine tract binding protein is required for dorso-ventral patterning and regulation of signalling between the germline and soma. PLoS One. 2013; 8(7):e69978. PMID: 23894566. doi: 10.1371/journal.pone.0069978

23. Fuller MT. Genetic control of cell proliferation and differentiation in Drosophila spermatogenesis. Semin Cell Dev Biol. 1998; 9(4):433–44. PMID: 9813190.

24. Castrillon DH, Gonczy P, Alexander S, Rawson R, Eberhart CG, Viswanathan S, et al. Toward a molecular genetic analysis of spermatogenesis in Drosophila melanogaster: characterization of male-sterile mutants generated by single P element mutagenesis. Genetics. 1993; 135(2):489–505. PMID: 8244010.

25. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology. 2013; 14(4): R36. PMID: 23618408. doi: 10.1186/gb-2013-14-4-r36

26. Heimiller J, Sridharan V, Huntley J, Wesley CS, Singh R. Drosophila polypyrimidine tract-binding protein (DmPTB) regulates dorso-ventral patterning genes in embryos. PLoS One. 2014; 9(7):e98585. doi: 10.1371/journal.pone.0098585 PMID: 25014769; PubMed Central PMCID: PMCPMC4094481.

27. Arama E, Agapite J, Steller H. Caspase activity and a specific cytochrome C are required for sperm differentiation in Drosophila. Dev Cell. 2003; 4(5):687–97. PMID: 12737804.

28. Huh JR, Vernooy SY, Yu H, Yan N, Shi Y, Guo M, et al. Multiple apoptotic caspase cascades are required in nonapoptotic roles for Drosophila spermatid individualization. PLoS Biol. 2004; 2(1):E15. doi: 10.1371/journal.pbio.0020015 PMID: 14737191; PubMed Central PMCID: PMCPMC300883.

29. The modENCODE Consortium, Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, et al. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science. 2010; 330 (6012):1787–97. doi: 10.1126/science.1198374 PMID: 21177974; PubMed Central PMCID: PMCPMC3192495.

30. Katz Y, Wang ET, Airoldi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nature methods. 2010; 7(12):1009–15. PMID: 21057496. doi: 10.1038/nmeth.1528

31. Noguchi T, Miller KG. A role for actin dynamics in individualization during spermatogenesis in Drosophila melanogaster. Development. 2003; 130(9):1805–16. PMID: 12642486.

32. Fabrizio JJ, Hime G, Lemmon SK, Bazinet C. Genetic dissection of sperm individualization in Drosophila melanogaster. Development. 1998; 125(10):1833–43. PMID: 9550716.

33. Xu T, Caron LA, Fehon RG, Artavanis-Tsakonas S. The involvement of the Notch locus in Drosophila oogenesis. Development. 1992; 115(4):913–22. PMID: 1451667.

34. Schutt C, Nothiger R. Structure, function and evolution of sex-determining systems in Dipteran insects. Development. 2000; 127(4):667–77. PMID: 10648226.

35. Cline TW, Meyer BJ. Vive la difference: males vs females in flies vs worms. Annu Rev Genet. 1996; 30:637–702. PMID: 8982468

36. Raymond CS, Murphy MW, O'Sullivan MG, Bardwell VJ, Zarkower D. Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. Genes Dev. 2000; 14 (20):2587–95. PMID: 11040213.

37. Raymond CS, Shamu CE, Shen MM, Seifert KJ, Hirsch B, Hodgkin J, et al. Evidence for evolutionary conservation of sex-determining genes. Nature. 1998; 391(6668):691–5. PMID: 9490411

38. Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. Sex-specific and lineage-specific alternative splicing in primates. Genome Res. 2010; 20(2):180–9. doi: 10.1101/gr.099226.109 PMID: 20009012; PubMed Central PMCID: PMCPMC2813474.

39. Catalan A, Hutter S, Parsch J. Population and sex differences in Drosophila melanogaster brain gene expression. BMC Genomics. 2012; 13:654. doi: 10.1186/1471-2164-13-654 PMID: 23170910; PubMed Central PMCID: PMCPMC3527002.

40. Chang PL, Dunham JP, Nuzhdin SV, Arbeitman MN. Somatic sex-specific transcriptome differences in Drosophila revealed by whole transcriptome sequencing. BMC Genomics. 2011; 12:364. doi: 10.1186/1471-2164-12-364 PMID: 21756339; PubMed Central PMCID: PMCPMC3152543.

41. Lott SE, Villalta JE, Zhou Q, Bachtrog D, Eisen MB. Sex-specific embryonic gene expression in species with newly evolved sex chromosomes. PLoS Genet. 2014; 10(2):e1004159. doi: 10.1371/journal.pgen.1004159 PMID: 24550743; PubMed Central PMCID: PMCPMC3923672.

42. Perry JC, Harrison PW, Mank JE. The ontogeny and evolution of sex-biased gene expression in Drosophila melanogaster. Mol Biol Evol. 2014; 31(5):1206–19. doi: 10.1093/molbev/msu072 PMID: 24526011; PubMed Central PMCID: PMCPMC3995337.

43.   Trabzuni D, Ramasamy A, Imran S, Walker R, Smith C, Weale ME, et al. Widespread sex differences in gene expression and splicing in the adult human brain. Nat Commun. 2013; 4:2771. doi: 10.1038/ncomms3771 PMID: 24264146; PubMed Central PMCID: PMCPMC3868224.

44.   Jazin E, Cahill L. Sex differences in molecular neuroscience: from fruit flies to humans. Nat Rev Neurosci. 2010; 11(1):9–17. doi: 10.1038/nrn2754 PMID: 20019686.

45.   Zhang B, Kraemer B, SenGupta D, Fields S, Wickens M. Yeast three-hybrid system to detect and analyze interactions between RNA and protein. Methods in enzymology. 1999; 306:93–113. PMID: 10432449.

46.   Ule J, Jensen K, Mele A, Darnell RB. CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods. 2005; 37(4):376–86. PMID: 16314267.

47.   Hamady M, Peden E, Knight R, Singh R. Fast-Find: a novel computational approach to analyzing combinatorial motifs. BMC Bioinformatics. 2006; 7:1. PMID: 16393334.

48.   Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, et al. PAR-CliP—a method to identify transcriptome-wide the binding sites of RNA binding proteins. J Vis Exp. 2010;In press(41). PMID: 20644507.

49.   Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome biology. 2011; 12(3):R22. PMID: 21410973. doi: 10.1186/gb-2011-12-3-r22