

Genome-Wide Identification of Alternatively Spliced mRNA Targets of Specific RNA-Binding Proteins

Mark D. Robida, Andrew Rahn, Ravinder Singh*

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

Background. Alternative splicing plays an important role in generating molecular and functional diversity in multi-cellular organisms. RNA binding proteins play crucial roles in modulating splice site choice. The majority of known binding sites for regulatory proteins are short, degenerate consensus sequences that occur frequently throughout the genome. This poses an important challenge to distinguish between functionally relevant sequences and a vast array of those occurring by chance. **Methodology/Principal Findings.** Here we have used a computational approach that combines a series of biological constraints to identify uridine-rich sequence motifs that are present within relevant biological contexts and thus are potential targets of the *Drosophila* master sex-switch protein Sex-lethal (SXL). This strategy led to the identification of one novel target. Moreover, our systematic analysis provides a starting point for the molecular and functional characterization of an additional target, which is dependent on SXL activity, either directly or indirectly, for regulation in a germline-specific manner. **Conclusions/Significance.** This approach has successfully identified previously known, new, and potential SXL targets. Our analysis suggests that only a subset of potential SXL sites are regulated by SXL. Finally, this approach should be directly relevant to the large majority of splicing regulatory proteins for which bonafide targets are unknown.

Citation: Robida MD, Rahn A, Singh R (2007) Genome-Wide Identification of Alternatively Spliced mRNA Targets of Specific RNA-Binding Proteins. PLoS ONE 2(6): e520. doi:10.1371/journal.pone.0000520

INTRODUCTION

Intervening sequences called introns interrupt the majority of genes in multi-cellular organisms. Spliceosomal introns are rare (<4%) in the budding yeast but present in the majority (85–94%) of genes in metazoans. Introns are removed and the coding regions (exons) joined together via a process known as pre-mRNA splicing before an mRNA can be translated. The 5' and 3' splice sites, the branchpoint, and the polypyrimidine tract (Py-tract) are important splicing signals in metazoans. Five small nuclear ribonucleoproteins (U1, U2, U4, U5, and U6 snRNPs), along with several additional factors, recognize these signals and assemble onto the pre-mRNA to form a large complex called the spliceosome. Spliceosome assembly occurs in several distinct steps, involving RNA-RNA, protein-protein, and RNA-protein interactions, leading to two catalytic reactions [1,2].

Alternative splicing generates multiple mRNA and/or protein isoforms from a single gene through the use of alternative 5' splice sites, 3' splice sites, exons, and/or introns. Several genes are known to encode >1,000 alternatively spliced mRNA isoforms. For example, the *Drosophila* homolog of the human Down Syndrome Cell Adhesion Molecule (DSCAM) gene potentially encodes three times as many alternatively spliced transcripts (~38,000) as the total number of predicted genes (~13,600) in the fruitfly [3,4]. Thus, alternative splicing, among several processes [5], provides a mechanism to generate enormous molecular diversity from a single gene, and provides a rich source of functional diversity in multi-cellular eukaryotes [6,7]. Alternative splicing plays an important role in numerous cellular and developmental processes such as cell growth and differentiation, cell signaling, programmed cell death, and nonsense-mediated decay [8,9].

The best-studied example of a developmental process controlled by alternative splicing is the *Drosophila melanogaster* somatic sex-determination pathway. It involves a hierarchy of alternative splicing events in which the key sex determining genes (*Sex-lethal* (*Sxl*), *transformer* (*tra*), *double-sex* (*dsx*), and *male specific-lethal 2* (*msl2*) are spliced differently in male (XY) and female (XX) flies. The master sex-switch protein, SXL, is an RNA-binding protein that is

absent in male flies and present in females [10]. It affects the splicing of three known pre-mRNAs by binding to uridine-rich sequences or polypyrimidine-tracts (Py-tracts) that are present adjacent to splice sites, leading to exon skipping in *Sxl*, 3' splice site switching in *tra*, and intron retention in *msl2* [6,11]. In female somatic cells, SXL mediates sexual differentiation and courtship behavior by allowing synthesis of the TRA protein, and allows proper dosage compensation by preventing synthesis of the MSL2 protein [10,11]. In addition to its role in alternative splicing, SXL also represses translation by binding to uridine-rich sequences in the untranslated regions (UTRs) of the *Sxl* and *msl2* mRNAs [11]. Furthermore, SXL also controls female germline development [10]. Absence of SXL in the female germline causes mitotic and meiotic defects, resulting in ovarian tumors or multicellular cysts of small undifferentiated cells [12,13,14] and in defects in chromosome pairing and meiotic recombination [15].

Several independent studies have suggested that additional targets of SXL exist. First, SXL associates with numerous loci on polytene X-chromosomes, presumably binding to nascent transcripts [16]. Second, SXL regulates the *fit* (*female-specific independent of tra*) gene in a *tra*-independent manner in the soma, although it is unlikely to be a direct target of SXL because of the lack of sex-specific mRNA isoforms and lack of SXL-binding sites [17]. Third, SXL controls dosage compensation of some *msl2*-in-

.....
Academic Editor: Juan Valcarcel, Centre de Regulació Genòmica, Spain

Received: April 30, 2007; **Accepted:** May 17, 2007; **Published:** June 13, 2007

Copyright: © 2007 Robida 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: National Institutes of Health, the American Cancer Society, and the Butcher Foundation

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

* **To whom correspondence should be addressed.** E-mail: rsingh@colorado.edu

dependent gene(s) that remains to be identified [18]. Fourth, although SXL has several important functions in the female germline, previous attempts to develop a genetic handle on its germline-specific targets have been unsuccessful [19]. Thus, additional targets of SXL, particularly in the female germline, have gone unrecognized, most likely because of subtle phenotypes, redundant functions, or limitations of a particular genetic screen.

Here we present a computational strategy that allowed identification of both new and potential SXL targets. This approach may be used to identify potential targets of other RNA binding proteins.

RESULTS

Computational strategy for the identification of potential targets of SXL

Given that the *Drosophila* genome has been sequenced [3] and that the SXL-binding site has been well characterized [20,21,22,23], we searched the entire *Drosophila* genome using a weight matrix corresponding to the SXL-binding site. Unlike string matching, this approach provides a quantitative rather than a merely qualitative description of a binding site by assigning weights to the four nucleotides at each sequence position. We aligned the SXL-binding sites (UUUUGUU(G/U)U(G/U)UUU(G/U)UU) from sequences selected by SELEX from a random RNA library [20] and converted this alignment matrix into a weight matrix of log-likelihood scores (Supplementary Table S1), as described [24]. We searched each of the overlapping 16-nucleotide strings in the *Drosophila* genome and calculated the total score for each string based on the weight matrix. If the score was above a user defined cut-off value (5.1 was used here to obtain only high-affinity binding sites), the genomic location of the binding site was saved. However, if the score was below the cut-off, the search engine advanced to the next position (for additional details see materials and methods).

This score was carefully chosen to capture known high affinity, long SXL sites such as those adjacent to regulated splice sites of *tra*, *Sxl*, and *msl2* transcripts, but ignore the majority of the short Py-tracts, including those associated with 3' splice sites. We empirically determined how far apart the hits were in the genome sequence. For example, when the hits were on average 20,000 basepairs apart, we expected approximately 12,600 binding sites. The cut-off (5.1) used here ignored most of the uridine-tracts such as those present near 3' splice sites. We were aware that it eliminated multiple copies of clustered, short Py-tracts, which might be potentially regulated, because the number of hits became unmanageable. For the matrix for the SXL-binding site, a maximum possible score is 7.88 (Table S1).

Our search of both strands of the genomic DNA yielded 14,007 matches for putative high affinity SXL-binding sites (Figure 1A). Given that there are approximately 13,600 predicted genes in *Drosophila* [3], the initial number of matches was too large for experimental analysis. Therefore, our *in silico* analysis included the following filters in a step-wise manner to reduce the number of candidates to an experimentally manageable size (Figure 1B). First, we determined if there was an expressed sequence tag (EST) within 3 kb on either side of the potential SXL-binding sites. This was intended to eliminate the matches that were in the intergenic region, which is particularly AT-biased [3]. Second, since SXL controls the splicing of its known targets, the remaining candidates were filtered on the basis of the proximity of potential binding sites to known splice sites. For the initial screen, we selected only those candidates (807) in which SXL-binding sites were located within 100 nucleotides of known 5' or 3' splice sites. The splice site locations were assigned based on comparison of EST sequences to

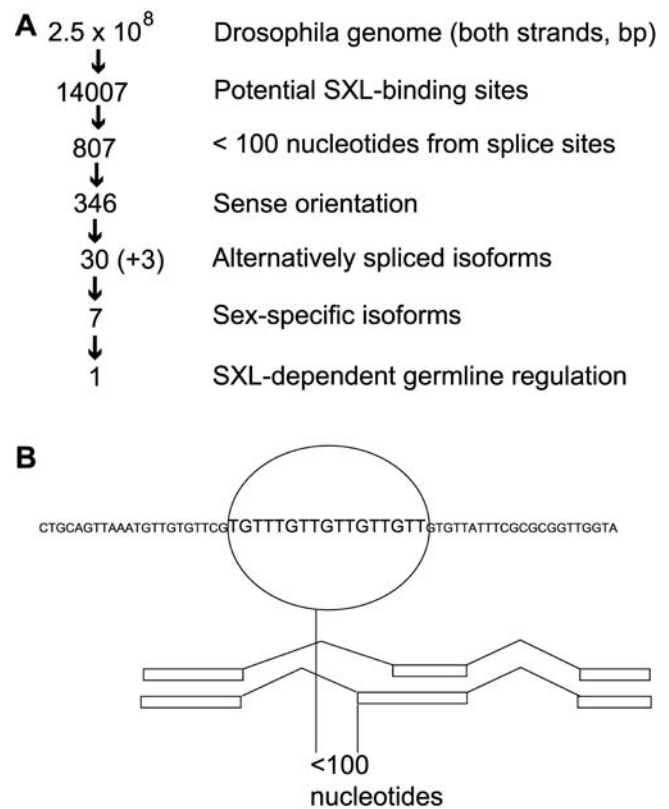


Figure 1. Computational search identifies potential targets of SXL. (A) Step-wise rationale for the identification of biologically relevant SXL targets. The number of potential candidates with SXL binding sites remaining after each step is indicated. (+3) represents that the three known targets of SXL (*Sxl*, *tra*, *msl2*) were also identified. (B) Schematics of how the search program works. Overlapping nucleotide windows are scanned for sequences that match the consensus-binding site. When a binding site is identified, the search program determines whether it is within 100 nucleotides of a splice site, whether it is in the sense orientation, and whether there is evidence of alternative splicing for that site.

doi:10.1371/journal.pone.0000520.g001

the genomic sequence and on their match to the splice site consensus [25]. Third, we discarded the candidates that were not present on the sense strand of the relevant genes. This left us with 346 candidates. Fourth, since all known targets of SXL are regulated by alternative splicing, we determined whether there was evidence of alternative splicing for the potential candidates based on the database of about 86,000 ESTs. We determined if splice sites adjacent to potential SXL-binding sites were alternatively spliced by aligning EST sets for each of the 346 candidates, using the ClustalW multiple sequence alignment program. The total number of potential candidates that met these multiple criteria was 33 (30 new) (Figure 1A); this number was experimentally amenable. It should be emphasized that this list included all 3 previously known targets of SXL (*Sxl*, *tra*, and *msl2*), and that several candidates contained multiple SXL-binding sites. Thus, this strategy successfully identified all previously known SXL targets as well as potential new targets.

Seven candidates show sex-specific mRNA isoforms in adult flies

SXL is present in females and absent in males. Furthermore, all known targets of SXL are alternatively spliced in a sex-specific

manner, generating sex-specific isoforms that differ in length and that can be identified using Northern analysis. Thus, it was anticipated that at least some of the additional candidates would also generate sex-specific isoforms. To pursue them, we obtained cDNA clones for each of the 30 new candidates and performed Northern analysis using poly(A)⁺ RNA from male or female adult flies. Twelve candidates showed expression but no sex-specific isoforms, and eight candidates showed no detectable signal on these RNA blots from adult flies (Figure 2). It remains possible that these candidates have similarly sized alternative exons, low abundance sex-specific transcripts, or sex-specific expression at other stages during development. These candidates were not pursued at this stage. Most importantly, seven candidates showed sex-specific mRNA isoforms (Figure 2, see asterisks), indicating that they might be potential SXL targets. CG3630 was found to have a longer non-sex-specific transcript and a shorter female-specific transcript. CG6422 had a shorter non-sex-specific transcript and a pair of longer female-specific transcripts. CG11737 had a longer non-sex-specific transcript and a shorter male-specific transcript. *Rm62* had a longer non-sex-specific transcript and two shorter sex-specific transcripts, one male-specific and the other female-specific. *Act5c* and *e(r)* both had a shorter non-sex-specific transcript and a longer female-specific transcript. Finally, *blow* had a longer non-sex-specific transcript and two shorter transcripts, one male-specific and another female-specific.

SXL-binding site and sex-specific isoforms

Examination of the EST database revealed potential sources for the sex-specific differences and suggested how SXL might regulate these targets (Figure 3A). Four candidates (CG3630, CG6422, CG11737, and *blow*) showed evidence of alternative 5' splice site choice adjacent to the SXL-binding site. The SXL-binding site in CG3630 was found within the exon upstream of the first alternative 5' splice site, the SXL-binding site in CG6422 was found downstream of the second alternative 5' splice site, and the SXL-binding sites in CG11737 and *blow* were found between alternative 5' splice sites. These scenarios are reminiscent of the way in which SXL regulates 3' splice site choice in its known target *tra* (Figure 3B) by binding to a site adjacent to the non-sex-specific 3' splice site [26,27]. Three candidates (*Rm62*, *Act5c*, and *e(r)*) showed evidence of alternative exon usage near SXL binding sites (Figure 3A). *Rm62* contained three identified SXL-binding

sites adjacent to alternative exons. *Act5c* and *e(r)* had SXL-binding sites located adjacent to alternative exons. Since the difference in the size of the alternative *Act5c* exons alone is insufficient to account for the sex-specific isoforms we believe that the female-specific isoform most likely reflects cross-hybridization to different members of the highly conserved actin family [28]. The use of alternative exons in *Rm62* and *e(r)* is reminiscent of the regulation of the known target *Sxl* (Figure 3B), in which exon skipping is caused by SXL binding to sites flanking an alternative exon [29,30]. As noted above, the new target *e(r)* was one of the candidates that contained multiple SXL-binding sites; one adjacent to an alternatively spliced exon and another downstream of an alternative polyadenylation site. Our molecular characterization of *e(r)* showed that both alternatively spliced and alternatively polyadenylated transcripts exist *in vivo* (data not shown). We found that the latter makes the primary contribution to sex-specific regulation, which occurs specifically in the female germline [31]. This candidate was pursued in significant detail because its regulation involved a novel mechanism. Our extensive molecular genetic analysis involving mutations in *Sxl* and the SXL-binding site and biochemical analysis using recombinant proteins showed that SXL-dependent regulation of *e(r)* provides a molecular mechanism for translational repression specifically within the female germline [31].

Somatic versus germline expression of the remaining sex-specific transcripts

As a first step towards determining any potential SXL-mediated regulation, we analyzed the tissue-specificity of the sex-specific transcripts for five of the remaining six candidates. The sixth candidate, *Act5c*, had several homologs that were highly conserved at the nucleotide level, raising the possibility that several individual genes likely contributed to the pattern of transcripts observed in Fig. 2. Thus, this candidate was not characterized further at this time. The expression patterns of the remaining five candidates were analyzed in the progeny of *tudor* (*tud*¹/*Df*) flies, which lack a germline, allowing the sex-specific transcripts to be sorted based on somatic or germline origin. Three of the five candidates showed consistent, reproducible results. The shorter male-specific transcripts found in two of the candidates, CG11737 and *blow-fuse* (*blow*), remained in the progeny of *tudor* (*tud*¹/*Df*) flies (Figure 4A,

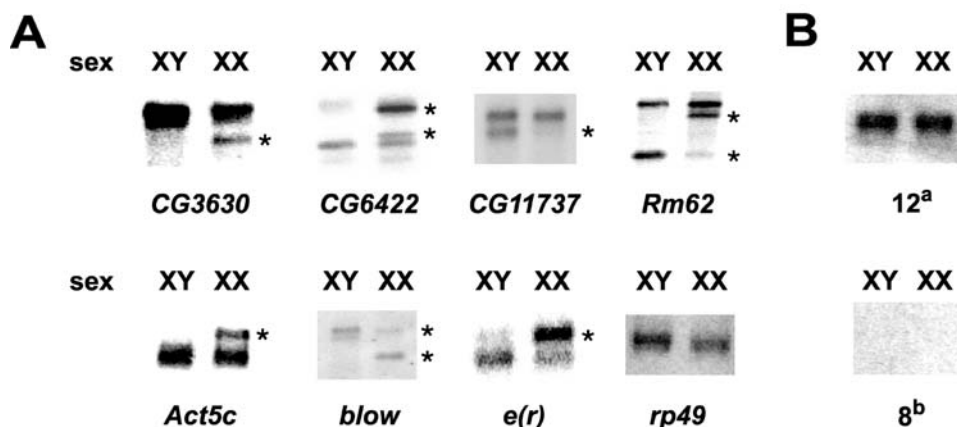


Figure 2. Screening for sex-specific isoforms. (A) Seven candidates show sex-specific mRNA isoforms by Northern analysis on poly(A)⁺ RNA from adult flies. Asterisks indicate sex-specific isoforms. XY and XX indicate chromosomal sex. (B) (Top) Several candidates (*bancal*, *Bap60*, *ImpL3*, *inx7*, *Moe*, *Rala*, *Ric*, *Top2*, *Frq2*, *Cyp28a5*, *RpL36*, and *Dhc16F*)^a were present at equal levels in both sexes. (Bottom) Several others (*fus*, *pUf68*, CG8370, *katanin-60*, *vap*, CG2967, CG5455, and *aralar1*)^b showed no hybridization in adult flies. doi:10.1371/journal.pone.0000520.g002

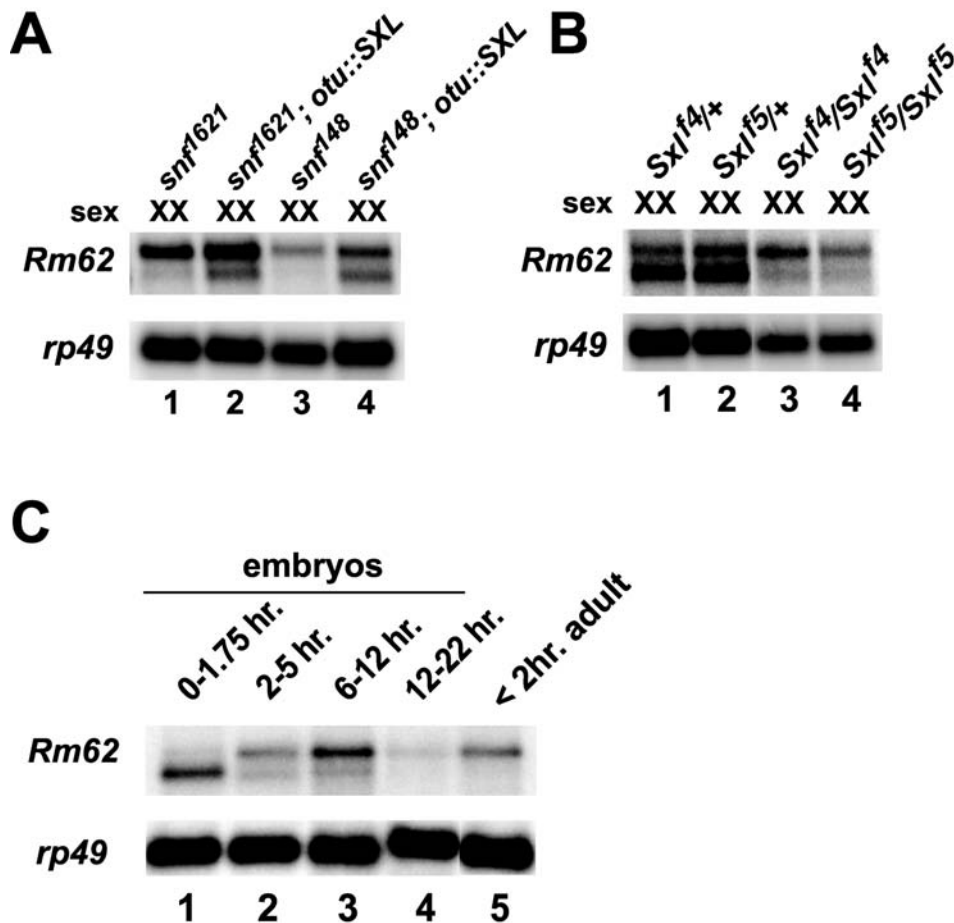


Figure 5. Germline SXL is necessary for expression of the female-specific *Rm62* transcript. A. XX flies homozygous for the *snf*¹⁶²¹ or *snf*¹⁴⁸ alleles, which disrupt SXL function specifically in the germline, do not express the female-specific *Rm62* transcript (lanes 1 and 3). Expression of a *Sxl* cDNA in *snf* mutant backgrounds under the control of the *otu* promoter restores the synthesis of the female-specific *Rm62* transcript (lanes 2 and 4). B. XX flies homozygous for the *Sxl*^{f4} or *Sxl*^{f5} alleles show a loss of the female-specific *Rm62* transcript (lanes 3 and 4). C. The female-specific *Rm62* transcript is maternally deposited, and is present only in mature ovaries.
doi:10.1371/journal.pone.0000520.g005

and *blow*, loss of *tra* (Figure 4C, lane 4) or loss of *dsx* (Figure 4C, lane 10) in females caused switching to a male expression pattern. Thus, the sex-specific expression patterns of CG11737 and *blow* are governed by genes in the somatic sex-determination pathway downstream of *Sxl*. Moreover, these findings emphasize that presence of an SXL-binding site is necessary but not sufficient for SXL-mediated regulation. We conclude that the two genes are indirectly regulated by SXL.

***Rm62* is downstream of SXL in the germline**

Since the female-specific transcript of *Rm62* is expressed in the germline, we tested whether SXL function in the germline was necessary for its production. First, we examined *Rm62* expression in the flies with mutations in the *sans fille* (*snf*) gene (*snf*¹⁶²¹ and *snf*¹⁴⁸) that disrupt *Sxl* function in the female germline. Female flies homozygous for either *snf* mutation did not express the female-specific *Rm62* transcript (Figure 5A, lanes 1 and 3), but introduction of an *Sxl* cDNA into these backgrounds restored expression of the transcript (lanes 2 and 4). Second, females homozygous for the *Sxl*^{f4} and *Sxl*^{f5} alleles also lacked the female-specific *Rm62* transcript (Figure 5B, lanes 3 and 4 versus 1 and 2). The combined results of these experiments demonstrate that SXL function in the

germline is necessary for expression of the female-specific, shorter *Rm62* isoform either directly or indirectly.

Given that the female-specific *Rm62* transcript was produced in the germline, it was possible that it was maternally deposited. Examination of the *Rm62* expression pattern in embryos showed that the female-specific *Rm62* transcript was specifically deposited into embryos (Figure 5C, lane 1) but was replaced by the non-sex-specific transcript after the maternal to zygotic transition (Figure 5C, lanes 2–4). We conclude that *Rm62* is downstream of SXL in the female germline and is maternally deposited.

Thus, our genome-wide search fulfilled its main purpose - to identify SXL targets near splice sites that showed evidence of alternative splicing, to identify all of the previously known targets, and to identify a novel target *e(r)* and a potential target *Rm62*. Only a subset of SXL-binding sites are regulated by SXL *in vivo*. The *e(r)* and *Rm62* transcripts provide important downstream handles to study the mysterious role of SXL in the female germline.

DISCUSSION

The genome-wide screen presented here, combining a computational search, biological constraints, and molecular genetic

analysis, identified both the previously known targets and a novel target of SXL. Identification of transcripts that appear downstream of SXL in the female germline is an important step toward understanding the role of SXL in the female germline.

Although previously known targets of SXL (*tra* and *msl2*) have exclusively sex-specific functions, it is not unreasonable to expect that certain targets could have both non-sex-specific and sex-specific functions at different times or in different tissues during development. These targets could have easily escaped previous genetic screens that identified the known components of the sex-determination pathway based on sex-specific phenotypes. In fact, the germline-specific *Sxl* target *e(r)*, which was identified in this screen, is essential in both sexes during embryogenesis and is regulated by *Sxl* in the female germline later during development [31]. Similarly, whereas certain mutations in the *Drosophila* PTB and the class VI unconventional myosin 95F (*jaguar*) are lethal, others specifically affect spermatogenesis, resulting in a male-sterile phenotype indicating that a gene can have both non-sex-specific and sex-specific function and/or regulation [36,37,38]. Therefore, we believe that additional *Sxl* targets that contribute to sexual dimorphism but do not solely have sex-specific functions remain to be identified. Among the potential new candidates that have SXL-binding sites in relevant biological contexts (Figure 1B) and that show sexually dimorphic expression patterns (Figure 2), *Rm62* is a potential target in the female germline, and the others (*blow*, CG3630, CG6422, and CG11737) are indirectly regulated by SXL via *dsx* in the soma. *blow* encodes a protein necessary for myoblast fusion and proper mesoderm development during embryogenesis, and the remaining candidates CG3630, CG6422, and CG11737 have no known function or previously recognized protein domain structure, although CG6422 is a putative member of the YT521-B-like family, which has been shown to modulate splice site selection *in vivo* [32,39,40]. *Rm62*, which is an ATP-dependent RNA helicase that contains a DEAD-box domain and an RRM-type RNA-binding motif [41,42], is inferred to be involved in the regulation of alternative splicing [35] and interacts with components of the RNAi machinery [43]. Given that the computational screen identified the three known targets of SXL as well as a novel target of SXL, characterization of *Rm62* using molecular genetics should provide important new insights into the function and regulation of both previously characterized and uncharacterized genes, the mechanisms of action of SXL, and the basis for sexual dimorphism. Known molecular differences in the *Rm62* transcripts represent both alternative splicing and polyadenylation variants. Future studies should address whether *Rm62* is a direct or indirect target of SXL, and what the molecular basis for the sex-specific *Rm62* transcript is.

Independent biological information from one or more genome-wide analyses may also be integrated to further refine the list of candidates. For example, SXL may collaborate with cofactors for increased specificity, as documented for the repression of *msh2* translation in females by SXL and its cofactor UNR [44,45], and for sex-specific splicing regulation of the *dsx* and *fruitless* transcripts by TRA and its cofactor TRA2 [10]. Second, conservation of short degenerate binding sites for RNA-binding proteins may be revealed by cross-species sequence comparison as has been done for *C. elegans* [46]. Third, direct *in vivo* RNA binding can be revealed by immunoprecipitation coupled with microarray analysis as has been done in *Drosophila* [47]. Thus, incorporation of additional biological constraints should help overcome limitations unique to any given approach and reduce the number of candidates for detailed characterization. Incorporation of such constraints could also allow searches for clusters of short uridine-rich sequences, which occur frequently and were ignored in the present study.

We note that recent experiments employing tiling arrays argue for extensive transcription in the *Drosophila* genome [48], implying that a subset of the putative SXL binding sites in regions annotated as intergenic could have a role in post-transcriptional regulation of non-coding transcripts.

In the future, the computational search presented here could benefit from improvements in the following areas: availability of full-length ESTs; improved annotations of gene structure, especially at exon/intron junctions; identification of low-abundance alternative transcripts; incorporation of quantitative information with respect to the frequency of alternatively spliced isoforms; and improvement in the speed of the algorithm by using indexing techniques and relational databases [49].

The majority of RNA-binding proteins, including splicing regulators, tend to have short, degenerate binding sites that occur frequently throughout the genome. Therefore, identification of biologically relevant binding sites is one of the most important challenges in the area of gene regulation. The most important strength of the analysis presented here is that the SXL binding site, although frequent throughout the genome, has biological consequences only when present in specific contexts such as in the proximity of alternative splice or polyadenylation sites. Moreover, since not all SXL-binding sites are regulated by SXL, a co-factor may collaborate with SXL to specify those that are regulated. Our computational method is readily adaptable to any RNA-binding protein for which a consensus binding sequence is known but the targets are unknown [50], and should provide a powerful tool in the search for target pre-mRNAs.

In conclusion, this approach has identified a novel target of SXL, has provided an additional potential target, and can be extended to numerous other RNA-binding proteins for which binding sites have been identified.

MATERIALS AND METHODS

Databases and indexing

We downloaded the databases for the *Drosophila* genome (na_genome.dros.RELEASE2.Z) from the Genome Annotation Database of *Drosophila* Release 2 (GADFLY) (<http://www.fruitfly.org/sequence/dlMfasta.shtml>) and the expressed sequence tag (EST) database (na_EST.dros.Z) (<http://www.fruitfly.org/sequence/dlcDNA.shtml>) from the Berkeley *Drosophila* Genome Project (BDGP). Both the genomic and EST databases were converted, using the formatdb command, from a fasta format to a format usable by BLAST using the NCBI toolbox (<ftp://ftp.ncbi.nih.gov/toolbox/>). In addition, both databases were indexed to fetch sequences faster for intermediate steps during analysis.

Generation of the weight matrix for the SXL binding site

The SXL weight matrix for the search was created from twenty-six sequences selected by selection-amplification from a random RNA library [20]. First, the sequences were arranged into an alignment matrix, which defined the number of times each nucleotide was found at a specific position within the alignment. The alignment matrix was then converted into a weight matrix using the formula:

$$\log \frac{(n_{ij} + p_i)/(N + 1)}{p_i}$$

where N is the total number of sequences, p_i is the a priori probability of nucleotide i , and n_{ij} is the number of times nucleotide i appears at position j (for detailed description see [24]).

Sequence search

To identify SXL-binding sites, overlapping windows of 16 nucleotides were scored using the weight matrix (Table S1), and strings that scored higher than the cut-off value of 5.1 were labeled as potential high-affinity binding sites. The highest possible score, which can be obtained by adding the highest value in each row of Table S1, is 7.88 for the SXL matrix.

Biological filters

To determine if the identified binding site was near a gene, 6 kb of genomic DNA (3 kb on each side) was used to BLAST against the EST database using blastn. The blast results were used to align the ESTs against the genome using ClustalW. Each alignment was automatically screened in two ways. First, only alignments that had both putative exons and introns within 100 bases of the binding site were retained. Second, the putative exon and intron junctions were examined, and those that matched at least partially the splice-site consensus signals received high priority. Splice sites were identified by searching the genome with weight matrices created using the consensus 5' and 3' splice site signals [24,25]. To search for SXL-binding sites in the EST database, every EST containing a binding site was aligned with other ESTs from the same CLOT (a group of homologous ESTs as defined by BDGP); some CLOTs contained too many EST sequences to be aligned, and were skipped. For each binding site the alignments were converted into a post-script file and examined manually to ensure that they met the above criteria.

All of the programs used here were written in Perl (<http://www.perl.org/>). The entire code and instructions for its use are available upon request.

cDNAs for analysis

ESTs, shown in parenthesis, for the following candidates were purchased from the Research Genetics Inc., CA: CG3630 (HL02887), *bancal* (LD15857), CG6422 (LD12853), *Bap60* (LD19076), *ImpL3* (LP10507), *inx7* (GH21056), *fus* (GH20047), *pUf68* (GH10982), CG11737 (LP01982), *Moe* (GH06344), *Rala* (SD01661), *Ric* (GH14071), CG8370 (LD46954), *Top2* (GH09845), *katanin-60* (SD02251), *vap* (LP02818), *Rm62* (LD17967), *Frq2* (LP01723), CG2967 (GH19107), CG5455 (GH11517), *Cyp28a5* (GH10483), *Act5c* (GH04613), *blow* (LP06243), *RpL36* (LP12131), *aralar1* (GH01348), *Dhc16F*

REFERENCES

- Jurica MS, Moore MJ (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 12: 5–14.
- Nilsen TW (2002) The spliceosome: no assembly required? *Mol Cell* 9: 8–9.
- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287: 2185–2195.
- Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, et al. (2000) *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101: 671–684.
- Singh R, Robida MD, Karimpour S (2006) Building Biological Complexity with Limited Genes. *Current Genomics* 7: 97–114.
- Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72: 291–336.
- Matlin AJ, Clark F, Smith CW (2005) Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* 6: 386–398.
- Valcarcel J, Singh R, Green MR (1995) Mechanisms of regulated pre-mRNA splicing. Lamond AI, ed. Austin, TX: The R. G. Landes Company, Biomedical Publisher.
- Lewis BP, Green RE, Brenner SE (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A* 100: 189–192.
- Schutt C, Nothiger R (2000) Structure, function and evolution of sex-determining systems in Dipteran insects. *Development* 127: 667–677.
- Forch P, Valcarcel J (2003) Splicing regulation in *Drosophila* sex determination. *Prog Mol Subcell Biol* 31: 127–151.
- Salz HK, Cline TW, Schedl P (1987) Functional changes associated with structural alterations induced by mobilization of a P element inserted in the Sex-lethal gene of *Drosophila*. *Genetics* 117: 221–231.
- Schupbach T (1985) Normal female germ cell differentiation requires the female X chromosome to autosome ratio and expression of sex-lethal in *Drosophila melanogaster*. *Genetics* 109: 529–548.
- Steinmann-Zwicky M, Schmid H, Nothiger R (1989) Cell-autonomous and inductive signals can determine the sex of the germ line of *Drosophila* by regulating the gene Sxl. *Cell* 57: 157–166.
- Bopp D, Schutt C, Puro J, Huang H, Nothiger R (1999) Recombination and disjunction in female germ cells of *Drosophila* depend on the germline activity of the gene sex-lethal. *Development* 126: 5785–5794.
- Samuels ME, Bopp D, Colvin RA, Roscigno RF, Garcia-Blanco MA, et al. (1994) RNA binding by Sxl proteins in vitro and in vivo. *Mol Cell Biol* 14: 4975–4990.
- Fujii S, Amrein H (2002) Genes expressed in the *Drosophila* head reveal a role for fat cells in sex-specific physiology. *EMBO J* 21: 5353–5363.
- Kelley RL, Kuroda MI (1995) Equality for X chromosomes. *Science* 270: 1607–1610.
- Hager J, Cline T (1997) Induction of female Sex-lethal RNA splicing in male germ cells: implications for *Drosophila* germline sex determination. *Development* 124: 5033–5048.

(LP05023), and *e(r)* (LD36385). ESTs HL02887, LD15857, LD12853, LD19076, and LD17967 had been cloned into the pBluescript SK+ vector, and ESTs LP10507, GH21056, GH20047, GH10982, LP01982, GH06344, SD01661, GH14071, LD46954, GH09845, SD02251, LP02818, LP01723, GH19107, GH11517, GH10483, GH04613, LP06243, LP12131, GH01348, LP05023, and LD36385 had been cloned into the pOT2 vector. Templates for Northern blot probes were generated by PCR from the pBluescript SK+ ESTs using the T7 primer (5' GTAATACGACTCACTATAGGG 3') and the T3 primer (5' AATTAACCCTCACTAAAGGG 3'), and from the pOT2 ESTs using the T7 primer and the pm001 primer (5' CGTTA-GAACGCGGTACAAT 3').

poly(A)⁺ RNA extraction

Total RNA was isolated using TRI reagent (Sigma-Aldrich, MO). Poly(A)⁺ RNA was isolated using the PolyATtract mRNA isolation system (Promega, WI).

Northern analysis

For each lane, approximately 0.5–1.0 µg of poly(A)⁺ RNA was separated by electrophoresis on a 1% agarose gel containing formaldehyde. RNA was transferred to a Duralose-UV membrane (Stratagene, CA), hybridized with ³²P labeled probe at 42°C overnight, washed extensively, and imaged on a Molecular Dynamic Phosphorimager. Additional details for various genotypes (*tra*, *Sxl*, *dsx*, and *tud*) can be found in [31,36]

SUPPORTING INFORMATION

Table S1 A weight matrix for the SXL binding site. Underlined positions reflect preferred residues.

Found at: doi:10.1371/journal.pone.0000520.s001 (0.44 MB TIF)

ACKNOWLEDGMENTS

We thank Dr. Bharat Gawande for his helpful comments on the manuscript, and Dr. Gerald Hertz for help with the weight matrices.

Author Contributions

Conceived and designed the experiments: RS. Performed the experiments: MR AR. Analyzed the data: RS MR AR. Wrote the paper: RS MR.

20. Singh R, Valcarcel J, Green MR (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* 268: 1173–1176.
21. Singh R, Banerjee H, Green MR (2000) Differential recognition of the polypyrimidine-tract by the general splicing factor U2AF65 and the splicing repressor sex-lethal. *RNA* 6: 901–911.
22. Sakashita E, Sakamoto H (1994) Characterization of RNA binding specificity of the *Drosophila* sex-lethal protein by in vitro ligand selection. *Nucleic Acids Res* 22: 4082–4086.
23. Kanaar R, Lee AL, Rudner DZ, Wemmer DE, Rio DC (1995) Interaction of the sex-lethal RNA binding domains with RNA. *EMBO J* 14: 4530–4539.
24. Hertz GZ, Hartzell GW, Stormo GD (1990) Identification of consensus patterns in unaligned DNA sequences known to be functionally related. *Comput Appl Biosci* 6: 81–92.
25. Burge CB, Tuschl T, Sharp PA (1999) Splicing of precursors to mRNAs by the spliceosomes. In: Gesteland RF, Cech TR, Atkins JF, eds. *The RNA World*. New York: Cold Spring Harbor Laboratory Press. pp 525–560.
26. Sosnowski BA, Belote JM, McKeown M (1989) Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage. *Cell* 58: 449–459.
27. Valcarcel J, Singh R, Zamore PD, Green MR (1993) The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* 362: 171–175.
28. Fyrberg EA, Kindle KL, Davidson N (1980) The actin genes of *Drosophila*: a dispersed multigene family. *Cell* 19: 365–378.
29. Lallena MJ, Chalmers KJ, Llamazares S, Lamond AI, Valcarcel J (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* 109: 285–296.
30. Nagengast AA, Stitzinger SM, Tseng CH, Mount SM, Salz HK (2003) Sex-lethal splicing autoregulation *in vivo*: interactions between SEX-LETHAL, the U1 snRNP and U2AF underlie male exon skipping. *Development* 130: 463–471.
31. Gawande B, Robida MD, Rahn A, Singh R (2006) *Drosophila* Sex-lethal protein mediates polyadenylation switching in the female germline. *EMBO J* 25: 1263–1272.
32. Doberstein SK, Fetter RD, Mehta AY, Goodman CS (1997) Genetic analysis of myoblast fusion: *blown fuse* is required for progression beyond the prefusion complex. *J Cell Biol* 136: 1249–1261.
33. Taylor BJ (1992) Differentiation of a male-specific muscle in *Drosophila melanogaster* does not require the sex-determining genes *doublesex* or *intersex*. *Genetics* 132: 179–191.
34. Eisen A, Sattah M, Gazitt T, Neal K, Szauter P, et al. (1998) A novel DEAD-box RNA helicase exhibits high sequence conservation from yeast to humans. *Biochim Biophys Acta* 1397: 131–136.
35. Park JW, Parisky K, Celotto AM, Reenan RA, Graveley BR (2004) Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc Natl Acad Sci U S A* 101: 15974–15979.
36. Robida MD, Singh R (2003) *Drosophila* polypyrimidine-tract binding protein (PTB) functions specifically in the male germline. *EMBO J* 22: 2924–2933.
37. Deng W, Leaper K, Bownes M (1999) A targeted gene silencing technique shows that *Drosophila myosin VI* is required for egg chamber and imaginal disc morphogenesis. *J Cell Sci* 112(Pt 21): 3677–3690.
38. Hicks JL, Deng WM, Rogat AD, Miller KG, Bownes M (1999) Class VI unconventional myosin is required for spermatogenesis in *Drosophila*. *Mol Biol Cell* 10: 4341–4353.
39. Schroter RH, Lier S, Holz A, Bogdan S, Klambt C, et al. (2004) *kette* and *blown fuse* interact genetically during the second fusion step of myogenesis in *Drosophila*. *Development* 131: 4501–4509.
40. Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, et al. (2005) CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* 33: D192–196.
41. Dorer DR, Christensen AC, Johnson DH (1990) A novel RNA helicase gene tightly linked to the Triplo-lethal locus of *Drosophila*. *Nucleic Acids Res* 18: 5489–5494.
42. Lasko P (2000) The *Drosophila melanogaster* genome: translation factors and RNA binding proteins. *J Cell Biol* 150: F51–56.
43. Ishizuka A, Siomi MC, Siomi H (2002) A *Drosophila* fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev* 16: 2497–2508.
44. Duncan K, Grskovic M, Strein C, Beckmann K, Niggeweg R, et al. (2006) Sex-lethal imparts a sex-specific function to UNR by recruiting it to the msl-2 mRNA 3' UTR: translational repression for dosage compensation. *Genes Dev* 20: 368–379.
45. Abaza I, Coll O, Patalano S, Gebauer F (2006) *Drosophila* UNR is required for translational repression of male-specific lethal 2 mRNA during regulation of X-chromosome dosage compensation. *Genes Dev* 20: 380–389.
46. Kabat JL, Barberan-Soler S, McKenna P, Clawson H, Farrer T, et al. (2006) Intronic alternative splicing regulators identified by comparative genomics in nematodes. *PLoS Comput Biol* 2: e86.
47. Blanchette M, Green RE, Brenner SE, Rio DC (2005) Global analysis of positive and negative pre-mRNA splicing regulators in *Drosophila*. *Genes Dev* 19: 1306–1314.
48. Manak JR, Dike S, Sementchenko V, Kapranov P, Biemar F, et al. (2006) Biological function of unannotated transcription during the early development of *Drosophila melanogaster*. *Nat Genet* 38: 1151–1158.
49. Hamady M, Peden E, Knight R, Singh R (2006) Fast-Find: a novel computational approach to analyzing combinatorial motifs. *BMC Bioinformatics* 7: 1.
50. Singh R, Valcarcel J (2005) Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol* 12: 645–653.