Intronic Alternative Splicing Regulators Identified by Comparative Genomics in Nematodes

Jennifer L. Kabat[1], Sergio Barberan-Soler[1], Paul McKenna[2¤a], Hiram Clawson[2], Tracy Farrer[1¤b], Alan M. Zahler[1*]

1 Department of Molecular, Cell, and Developmental Biology and Center for Molecular Biology of RNA, University of California Santa Cruz, Santa Cruz, California, United States of America, 2 Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, California, United States of America

Many alternative splicing events are regulated by pentameric and hexameric intronic sequences that serve as binding sites for splicing regulatory factors. We hypothesized that intronic elements that regulate alternative splicing are under selective pressure for evolutionary conservation. Using a Wobble Aware Bulk Aligner genomic alignment of *Caenorhabditis elegans* and *Caenorhabditis briggsae,* we identified 147 alternatively spliced cassette exons that exhibit short regions of high nucleotide conservation in the introns flanking the alternative exon. In vivo experiments on the alternatively spliced *let-2* gene confirm that these conserved regions can be important for alternative splicing regulation. Conserved intronic element sequences were collected into a dataset and the occurrence of each pentamer and hexamer motif was counted. We compared the frequency of pentamers and hexamers in the conserved intronic elements to a dataset of all *C. elegans* intron sequences in order to identify short intronic motifs that are more likely to be associated with alternative splicing. High-scoring motifs were examined for upstream or downstream preferences in introns surrounding alternative exons. Many of the high-scoring nematode pentamer and hexamer motifs correspond to known mammalian splicing regulatory sequences, such as (T)GCATG, indicating that the mechanism of alternative splicing regulation is well conserved in metazoans. A comparison of the analysis of the conserved intronic elements, and analysis of the entire introns flanking these same exons, reveals that focusing on intronic conservation can increase the sensitivity of detecting putative splicing regulatory motifs. This approach also identified novel sequences whose role in splicing is under investigation and has allowed us to take a step forward in defining a catalog of splicing regulatory elements for an organism. In vivo experiments confirm that one novel high-scoring sequence from our analysis, (T)CTATC, is important for alternative splicing regulation of the *unc-52* gene.

Introduction

One of the interesting lessons learned from the analysis of the human genome is that we may possess fewer than 25,000 genes [1]. One mechanism to dramatically increase the complexity of the human proteome from this lower-than-expected number of genes is to allow some genes to encode multiple proteins. This process can be accomplished by alternative precursor messenger RNA (pre-mRNA) splicing. Studies that use expressed sequence tag (EST) alignments to identify alternatively spliced genes have led researchers to predict that up to 60% of human genes are alternatively spliced [2–5]. Alternative splicing events can be regulated in tissue-specific, developmental, and hormone-responsive manners, providing additional mechanisms for the regulation of gene expression [6,7]. Understanding alternative splicing and its regulation is a key component to understanding metazoan genomes.

The current models for alternative splicing regulation are based on the interactions of intronic or exonic RNA sequences, known as *cis* elements, with splicing regulatory proteins known as *trans*-acting splicing factors [8]. The binding of splicing factors to the pre-mRNA regulates the ability of the spliceosomal machinery to recognize and promote alternative splicing. The role of intronic elements in regulating splicing is well established and has been shown

to work in a combinatorial fashion based on the *trans*-acting factors that are present. For example, the inclusion of the 18-nucleotide-long, neural-specific N1 exon of the human *c-SRC* gene is regulated by the downstream control sequence found in the intron downstream of the N1 exon. This sequence serves as a recruitment site for both constitutive and neuronal cell-specific splicing factors such as nPTB, FOX-1,

* To whom correspondence should be addressed. E-mail: zahler@biology.ucsc.edu

¤a Current address: Department of Biology and Center for Computing for Life Sciences, San Francisco State University, San Francisco, California, United States of America

¤b Current address: University of Virginia School of Medicine, Charlottesville, Virginia, United States of America

Synopsis

Alternative splicing of precursor messenger RNA is a process by which multiple protein isoforms are generated from a single gene. As many as 60% of human genes are processed in this manner, creating tissue-specific isoforms of proteins that may be a key factor in regulating the complexity of our physiology. One of the major challenges to understanding this process is to identify the sequences on the precursor messenger RNA responsible for splicing regulation. Some of these regulatory sequences occur in regions that are spliced out (called introns). This study tested the hypothesis that there should be evolutionary pressure to maintain these intronic regulatory sequences, even though intron sequence is non-coding and rapidly diverges between species. The authors employed a genomic alignment of two roundworms, *Caenorhabditis elegans* and *Caenorhabditis briggsae,* to investigate the regulation of alternative splicing. By examining evolutionarily conserved stretches of introns flanking alternatively spliced exons, the authors identified and functionally confirmed splicing regulatory sequences. Many of the top scoring sequences match known mammalian regulators, suggesting the alternative splicing regulatory mechanism is conserved across all metazoans. Other sequences were not previously identified in mammals and may represent new alternative splicing regulatory elements in higher organisms or ones that may be specific to worms.

and FOX-2 [9–12]. The vertebrate RNA-binding protein FOX-1 can also regulate muscle-specific alternative splicing through interactions with the RNA sequence GCAUG [13], and repeats of this sequence have been shown to be important for alternative splicing regulation of the fibronectin exon EIIIB and the rat calcitonin/CGRP exon 4 [14,15]. Many other examples of complex and combinatorial regulation of alternative splicing through intronic *cis* elements have been demonstrated, and combinatorial interactions between proteins such as Nova-1, polypyrimidine tract binding protein (PTB), and ETR-3, with specific *cis* sequences, are important for alternative splicing regulation [16–20].

Intronic sequences are non-coding, and therefore they should have less evolutionary selective pressure to maintain their sequence. An exception to this should be intronic sequences that regulate alternative splicing. In an analysis of alternatively spliced human cassette exons, it was found that on average, approximately 100 nucleotides of intron sequence, flanking either side of the exon, tend to be highly conserved between the mouse and human genomes, with 88% identity in the upstream sequences and 80% identity in the downstream sequences [21]. Some clues to potential splicing regulatory motifs arise from these studies. For example, Sorek and Ast found that the sequence TGCATG was the second most common hexamer in the first 100 nucleotides downstream of alternatively spliced exons, appearing in 18% of these intronic regions [21]. Another study of aligned mouse/human alternative exons found that GCATG is the most overrepresented pentamer in the proximal conserved region of the intron downstream of alternative exons [22]. A third study found that TGCATG is frequently located in introns flanking brain-enriched alternative exons, and its presence and spacing are highly conserved in these genes from fish to man [23].

Using the nematode *Caenorhabditis elegans* as a model system, we have been working to take advantage of comparative genomics to identify *cis*-acting regulators of alternative splicing. The *C. elegans* gene structure, splicing machinery, and gene expression regulation is similar to that of other higher eukaryotes, with the exception that the average intron size is smaller. Our lab has previously developed methods for identification of alternatively spliced genes in *C. elegans* by aligning the genome sequence with ESTs and mRNA sequence [24]. We developed an algorithm, the Wobble Aware Bulk Aligner (WABA), for creating interspecies genome alignments between *C. elegans* and the related roundworm, *Caenorhabditis briggsae* [25]. WABA employs a hidden Markov model (HMM) to identify aligned regions as coding, high homology, low homology, and no homology. It also factors the AT richness of *C. elegans* introns into its calculations when it defines an intronic region as "high" homology [25]. *C. briggsae* and *C. elegans* diverged approximately 100 million years ago, yet are indistinguishable by eye [26]. Alignment of these two genomes revealed that exonic sequences are highly conserved between these species, but intronic and intergenic sequences are rarely conserved [25]. We found that these rare, high homology sequences in introns are more likely to occur in the introns flanking alternatively spliced exons than in total introns [25]. We hypothesized that these conserved intronic regions were *cis*-acting regulatory elements for alternative splicing. This nematode alignment, with relatively limited regions of high homology, provides the possibility for more specific pinpointing of intronic splicing regulatory elements than the much longer 100-nucleotide-long conserved regions flanking alternative exons in mouse/human alignments [21].

In this paper, we present the analysis of conserved regions of introns flanking alternatively spliced exons in *C. elegans* and correlate these conserved regions with alternative splicing regulation. We collected these conserved sequence regions into a database and searched for overrepresented pentamers and hexamers relative to a total intron database, similar to the method used by Burge's group to identify exonic splicing enhancers [27]. This allowed us to create a table of potential intronic alternative splicing *cis*-regulatory motifs. Since many RNA recognition motif–containing splicing factors recognize specific sequences on the order of 4–6 nucleotides in length [11,13,18,28–32], the high-scoring motifs in this catalog may represent specific binding sites for particular splicing factors. Several of our highest scoring motifs in this analysis correlate with known vertebrate splicing regulatory elements, for example, (T)GCATG [23], but several have not been previously identified. A number of candidates identified by this method were tested in an in vivo splicing reporter system in *C. elegans*. We have used this analysis to identify and confirm a new, highly conserved, alternative splicing regulatory element, (T)CTATC. We show that this sequence works in coordination with GCATG to regulate the alternative splicing of the *unc-52* gene.

Results

Identification of Highly Conserved Regions in Introns Flanking Alternatively Spliced Exons

In order to identify alternatively spliced cassette exons in the *C. elegans* genome, we used the Intronerator [24] to generate an initial set of 1,471 putative alternatively spliced genes. We did this by aligning over 200,000 ESTs and mRNAs to the *C. elegans* genome and identifying regions where the

alignments are consistent with more than one way to process a gene. The program could not distinguish between alternative splicing and alternative promoters, so we analyzed each of these alignments individually to verify alternative splicing. We found 449 examples of genes with strong cDNA evidence for alternative cassette exons, and 454 examples of genes with alternative promoters, which usually had unique first exons that are spliced to common second exons. Of the genes with alternative splicing, 162 also contained alternative promoter usage. The remaining genes in the initial set of putative alternatively spliced genes were mostly due to unusual ESTs in the database that did not fit a gene model. We also saw evidence of ESTs that showed internal deletions at short direct repeats. To the program, these indicated the potential for intron removal, but upon further inspection, these did not meet the criteria of introns (start with the dinucleotides GT or GC and end with AG) and were likely the result of cloning artifacts of the ESTs.

Intronic regions that are highly conserved between *C. elegans* and *C. briggsae* are rarely identified by WABA [25]. While analyzing our alternative cassette exon database, we were struck by the fact that we could identify many examples of WABA-defined high homology sequences in introns

flanking the alternatively spliced cassette exons, suggesting strong evolutionary conservation. For 142 of the 449 alternatively spliced genes, WABA identified a highly conserved sequence in flanking introns upstream or downstream of 147 alternative cassette exons. Figure 1 shows several examples of screen shots from the Intronerator browser of cDNA-confirmed *C. elegans* alternatively spliced isoforms (in blue) and a graphical representation of the WABA alignments between *C. elegans* and *C. briggsae* (in purple). Dark purple regions, indicating high homology, can be seen in these introns either upstream or downstream of the alternative exon. Sometimes a single conserved element is identified by WABA, while for other cases multiple regions in the introns are conserved.

An important question is whether these 147 *C. elegans* alternative cassette exons are also alternatively spliced in *C. briggsae*. Due to a lack of *C. briggsae* transcript data, we have no direct evidence for alternative splicing of these exons. Therefore, we examined the *C. briggsae* homologs of these alternative exons for features of functional exons. A previous alignment of 8 MB of the *C. briggsae* genome with the complete *C. elegans* genome found that in coding exons there is 79.3% overall nucleotide identity [25]. Consistent with
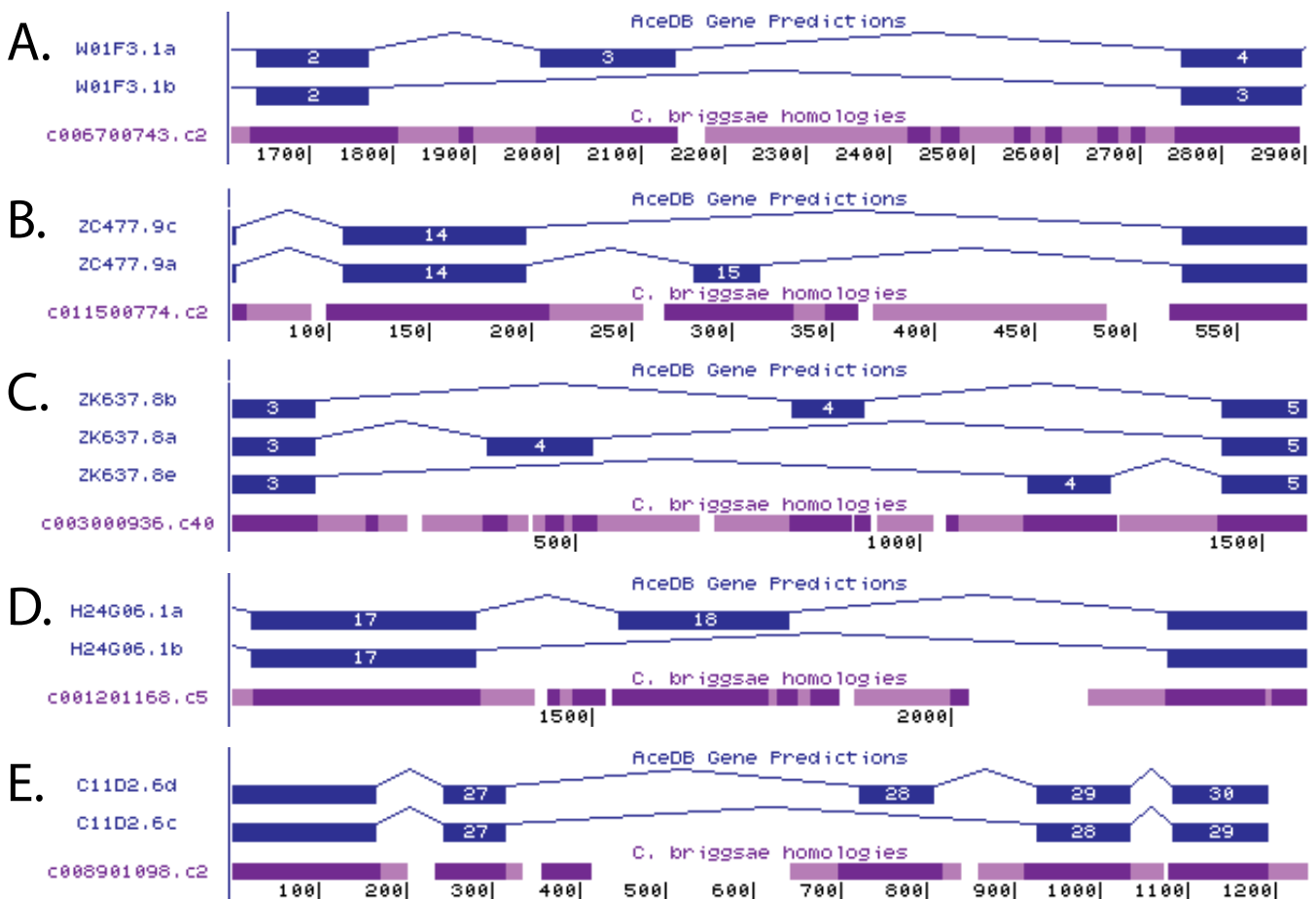
Figure 1. Images from the Intronerator Genome Browser Showing Alternatively Spliced Genes
Gene isoforms predicted by the Wormbase Consortium are shown in blue, and WABA homology alignments for *C. briggsae* to this region of the *C. elegans* genome are shown in purple. Dark purple indicates a region of WABA high homology, light purple corresponds to low homology, and white indicates no homology between species. Regions of alternatively spliced genes: (A) W01F3.1, (B) ZC477.9, (C) ZK637.8, (D) H24G06.1, and (E) C11D2.6 are shown.
DOI: 10.1371/journal.pcbi.0020086.g001

these findings, cross-species alignment of these 147 alternative cassette exons revealed an average nucleotide identity between *C. elegans* and *C. briggsae* of 81.8%. Amino acid identity of the open reading frame of these exons was 85.4%. Maintenance of the reading frame for these exons also appears to be well conserved: 57.0% of the 147 exons are the exact same size, 34.3% differ by a multiple of three nucleotides, and only 8.8% differ by a non-multiple of three nucleotides (unpublished data). The high conservation of nucleotide sequence, amino acid sequence, and maintenance of open reading frame is consistent with these being coding exons in *C. briggsae*.

One feature of alternative exons is that they are often flanked by weak splice sites which allow for splicing regulation (reviewed in [33]). If these exons are alternatively spliced in *C. briggsae* we would expect that they would be flanked by 3′ and 5′ splice sites of similar strength. We used the UCSC Genome Table Browser (http://www.genome.ucsc.edu) in an attempt to automate rapid alignment of the splice junctions for these 147 exons in *C. elegans* and *C. briggsae*. For the majority of these alternative exons, this automated method was successful in identifying *C. elegans* and *C. briggsae* 5′ and 3′ splice sites for each exon, and we used these data to compare the similarity of splice sites between the two species. For 5′ splice sites we examined the last three nucleotides of the exon and first six nucleotides of the intron, as these base-pair with U1 small nuclear RNA during initial 5′ splice-site recognition[31]. We found that these nine nucleotides were completely identical for 47.1% of the alternative exons and differed by only one nucleotide in 27.1% of the exons. For the 3′ splice sites, we examined the conservation of the last six nucleotides of the intron and the first nucleotide of the exon, as these are a binding site for the *C. elegans* homolog of the heterodimeric U2 auxiliary factor (U2AF) involved in initial 3′ splice-site recognition [34]. For 55.4% of the alternative exons examined, these seven nucleotides were completely identical, and for an additional 30.1% they differed by only one nucleotide. This strong conservation of splice-site sequence strength, along with regions of conservation in flanking introns, suggests that these exons may be substrates for alternative splicing in *C. briggsae*.

The Role of Conserved Elements in Alternative Splicing Regulation

There are many examples of regions of introns flanking alternatively spliced exons that function as *cis*-acting regulators of alternative splicing. We hypothesized that the reason that these high homology WABA-identified intronic elements have been conserved over 100 million years of evolution is that they may regulate alternative splicing. In order to test this hypothesis, we looked at the *C. elegans/C. briggsae* alignment of the alternatively spliced region of the *C. elegans* alpha(2) type IV collagen gene *let-2*. *Let-2* has two mutually exclusive alternative exons, exons 9 and 10, and their splicing pattern is evolutionarily conserved as far back as the distantly related parasitic nematode, *Ascaris suum* [35]. Messages incorporate either exon 9 or exon 10 in a developmentally regulated manner; embryos predominantly use exon 9, adults predominantly use exon 10, and there is a gradual shift in the usage of these two exons during the larval stages [36]. The WABA alignment of this 400-base intron identifies four conserved regions in the intron between exons 10 and 11

(Figure 2). To test the role of these conserved intronic elements in alternative splicing regulation, we employed a splicing reporter construct system containing the alternatively spliced region between exons 8 and 11 that mimics the developmental control of alternative splicing for this region when transformed into *C. elegans* [37]. We mutated the first conserved element in this intron and monitored the developmental regulation of this splicing. Deletion of this conserved element, in which 28 of 34 bases have been conserved between *C. elegans* and *C. briggsae,* results in a major reduction in the usage of exon 10 in L4 animals (unpublished data). The identical effect on splicing of an even smaller deletion within this element, del1.2, is shown in Figure 2. For the L4 time point, the developmental stage at which we should detect maximal exon 10 usage, only minimal splicing of this exon is detected. Since mutation of this element results in only minimal exon 10 inclusion, we hypothesize that either adults produce a splicing factor that interacts with this element to promote exon 10 splicing or that embryos produce a splicing repressor that inhibits exon 10 splicing. In the past, researchers have identified splicing regulatory elements by creating a series of mutations across an intron. In this computational approach, the interspecies genome alignments led us directly to a small element required for the developmentally regulated switch in this splicing.

Analysis of a Database of Conserved Intronic Elements Flanking Alternatively Spliced Exons

As demonstrated in the previous section, evolutionarily conserved elements in introns flanking alternatively spliced exons can be important for alternative splicing. We used a computational approach to identify sequences that are more likely to occur in these conserved elements than in total intron sequences. As described above, we have identified 142 alternatively spliced genes in which the introns flanking an alternatively spliced cassette exon contain high homology regions with *C. briggsae* as defined by WABA. We extracted the highly conserved *C. elegans* sequences from these introns and put them into a database. The first or last seven nucleotides of introns were excluded from this dataset as these contain conserved signals for the constitutive splicing machinery [38]. This dataset contained 537 conserved elements of average length, 38.5 bases (minimum length, seven bases, longest, 231 bases, and median length, 28 bases) for a total of 20,675 bases. See Table S1 for a list of these elements and the alternatively spliced genes from which they were derived. We also generated a control dataset of all introns annotated in Wormbase genome sequence release WS 120.

Because many alternative splicing factors have binding preferences for relatively short sequence motifs [11,13,18,28–32], we decided to search the conserved intronic element dataset for short sequence motifs that appear more frequently here than in total introns. In order to identify sequences that are more prevalent in the conserved intronic regions bordering alternative exons, we counted the number of occurrences of every possible hexamer or pentamer motif in both our conserved intron element and total intron datasets. For each motif in the conserved element dataset, we determined the observed frequency, which was the count of each motif divided by the total number of possible motifs of that length in that dataset. We also determined an expected frequency based on the number of counts of each motif in the
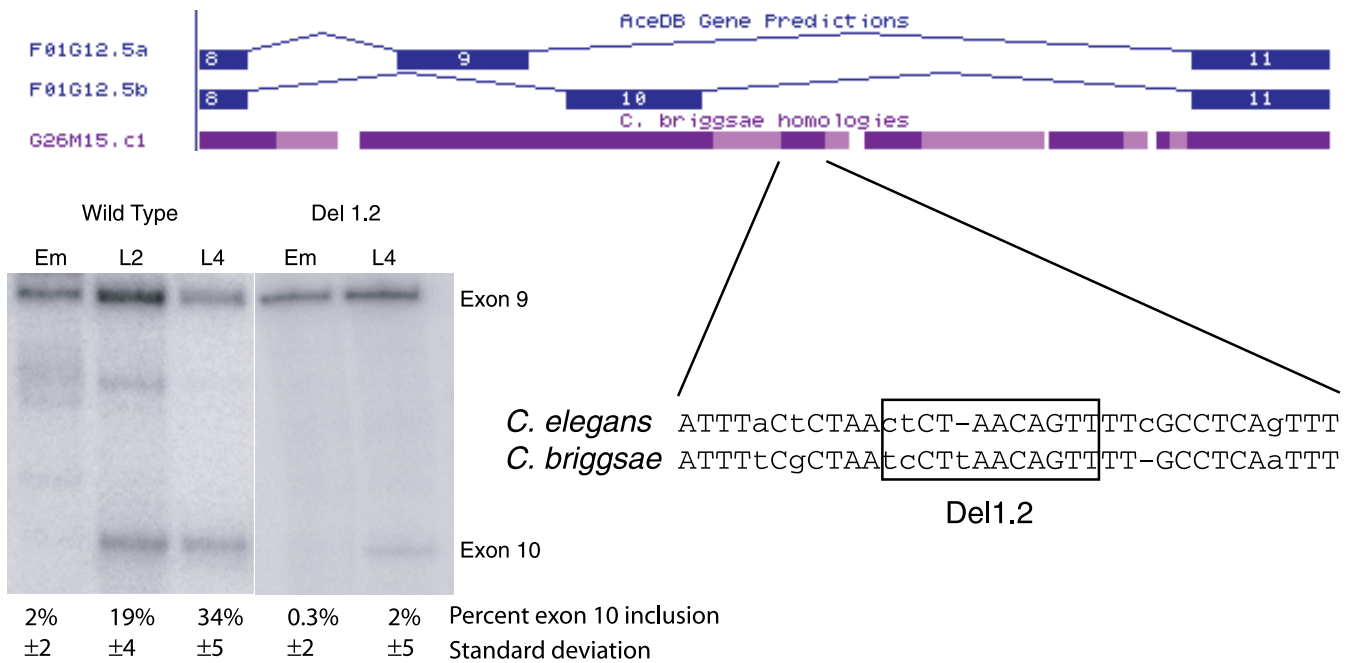
Figure 2. An Intronic Element Regulates *let-2* Alternative Splicing

The top of this figure shows the alignment of *C. elegans* and *C. briggsae* sequences for the alternatively spliced region of *let-2*. Blue tracks indicate the splicing for the embryonic (top) and adult (bottom) isoforms. The purple track indicates homology between the *C. briggsae* and *C. elegans* genomes as determined by WABA. Dark purple tracks indicate regions of strong homology (>70%). The sequence of the first conserved element of intron 10 is shown. The box indicates the part deleted and replaced with the sequence GAA in the del1.2 splicing reporter construct. The lower left part of the figure shows the results of ^{32}P RT-PCR reactions with primers specific for the splicing reporter. Products for exon 9- or exon 10-containing messages are indicated. In embryos, only usage of exon 9 is detected for either reporter. At the L4 stage, 34% of the wild-type reporter messages contain exon 10 while del1.2 mutant messages contain only a trace amount of this exon.
DOI: 10.1371/journal.pcbi.0020086.g002

total intron dataset divided by the total number of possible motifs in that dataset. We then calculated an observed over expected (obs/exp) ratio for each motif, which was the observed frequency of each motif in the conserved element dataset divided by the expected frequency of each motif based on the total intron dataset. Table 1 shows the 40 top scoring pentamer and hexamer motifs in the conserved element dataset as determined by the obs/exp ratio value.

Several previously identified mammalian splicing regulatory sequences are present in our lists of high-scoring hexamers and pentamers. This is encouraging because many of the known mammalian splicing factors have *C. elegans* homologs. For example, the sequence TGCATG and its subset pentamer GCATG have been shown to be important intronic regulators of splicing for the mammalian fibronectin and calcitonin/CGRP genes [14,15,39]. The mammalian FOX-1 protein selected the GCAUG sequence in a SELEX experiment, and it has been shown that the FOX-1 protein alters alternative splicing of several GCAUG-containing genes in transient transfection assays [13]. FOX-2, a homolog of FOX-1, has also been shown to regulate splicing of transcripts containing UGCAUG in neuronal cell culture [12]. FOX-1 was originally identified in *C. elegans* as a splicing factor of the transcript for the sex determination gene, *xol-1* [40]. Another well studied mammalian motif is that of intronic GU dinucleotide repeats, which have been shown to regulate splicing of the human *CFTR* gene [41] and serve as a binding site for the ETR-3 splicing regulatory factor [42]. CU dinucleotide repeats serve as a binding site for the PTB family members in the intronic regions that regulate

inclusion of the N1 exon of *c-src* [43]. While targets for *C. elegans* alternative splicing factors might be inferred from their homology to mammalian proteins, very little experimental evidence exists for nematode splicing-factor binding sites. The *C. elegans* muscle-specific splicing factor SUP-12 regulates *unc-60* alternative splicing. This protein interacts with a GU-rich region, similar to the GU-rich motifs identified in Table 1, of an *unc-60* intron [44]. (For a review of other mammalian splicing factors with *C. elegans* homologs see [33].) While many of the high-scoring conserved motifs that we identified match known mammalian splicing regulatory sequences, we have also identified new motifs that may also be involved in pre-mRNA splicing regulation in mammals as well as nematodes.

One important aspect of *C. elegans* introns is that they are shorter than introns in vertebrates. Half of *C. elegans* introns are below 60 nucleotides in length, too short to be spliced in vertebrates [38]. However, many *C. elegans* introns are long and resemble vertebrate introns. In our alternatively spliced intron dataset, the median intron size is 263 bases (shortest is 43, longest is 10,719, and average is 561), indicating that these regulated introns generally belong in the larger intron class. We asked whether a similar analysis of the pentamers and hexamers found in the full introns flanking alternatively spliced exons would yield similar high-scoring motifs as our analysis of the conserved sequence subset of these introns. We did a pentamer and hexamer motif count of the full introns flanking the 147 alternatively spliced exons from which we previously extracted evolutionarily conserved elements. We counted the occurrence of every pentamer and hexamer in

Table 1. Pentamer and Hexamer Sequence Motifs That Occur More Frequently in Conserved Intronic Elements Flanking Alternatively Spliced Exons than in the Total Intron Dataset

Pentamers		Obs/Exp	Hexamers		Obs/Exp
Word	Count		Word	Count	
GCATG	50	6.65	CCGTGC	9	9.29
GTGTG	58	4.63	GCATGC	13	8.88
CTATC	37	4.23	TGCATG	22	8.87
TGTGT	79	4.08	CGCATG	11	8.84
CTCTC	52	3.77	GCATGA	17	8.27
GCCCG	12	3.66	CGTTCC	10	7.45
CATGC	25	3.57	CTATCC	10	7.06
CTTCT	79	3.56	TCTATC	23	6.96
GTCGT	24	3.55	CTAACC	12	6.90
CGTCC	13	3.50	CGTGTG	11	6.93
CGTCG	17	3.46	GTGTTG	20	6.63
GTGTC	33	3.40	GTGTGC	17	6.58
CTCGC	17	3.40	TGCACG	8	6.54
TGTCC	26	3.35	GTCCGT	7	6.49
CATGT	43	3.32	CATGTC	12	6.28
TCCGT	28	3.30	CGGAGG	5	6.17
GTGCT	27	3.29	GCTTCT	24	6.17
TCCTC	36	3.27	CGCTTC	10	6.15
GCTTC	36	3.26	TGTGTC	19	6.13
TCTCT	72	3.25	CCGTCC	5	5.79
CGCTT	25	3.23	GTGTGT	27	5.77
CTAAC	32	3.22	GTCGTG	8	5.74
CATCC	21	3.19	GCAAGG	6	5.69
CGTGC	14	3.19	GCCCGG	3	5.69
GTTGT	52	3.12	TCCTCC	15	5.63
TGTGC	36	3.12	CCTTGT	12	5.61
TGTCT	43	3.11	GTTGTG	18	5.59
CCGTG	15	3.03	ACCGTG	6	5.55
CGGAC	10	3.01	GCCCGC	4	5.54
TCATG	37	2.90	GATGGT	9	5.39
TGCAC	29	2.90	CGAGTG	6	5.37
CTCCT	27	2.89	CCCGTC	5	5.32
GTCCA	19	2.87	CCAGGC	5	5.30
GTCCG	11	2.84	CCTCGC	4	5.27
CCTCC	20	2.79	GTCGCC	5	5.26
TGTTG	55	2.79	GACGTG	6	5.25
GTCGC	12	2.79	GTGCCC	5	5.25
TTGTG	63	2.77	TGTTGT	31	5.23
TTGTC	55	2.76	CTATCT	13	5.21
TGCTT	43	2.74	TCTCTC	34	5.20

The Count column indicates the number of occurrences of each motif in the conserved intron element dataset. Obs/Exp is the ratio of the observed frequency score of these motifs in the conserved element dataset divided by the expected occurrence of each motif based on the total intron dataset. The top 40 highest scoring motifs as determined by the obs/exp value for pentamers and hexamers are listed.
DOI: 10.1371/journal.pcbi.0020086.t001

Table 2. Pentamer and Hexamer Sequence Motifs That Occur More Frequently in the Introns Flanking Alternatively Spliced Exons than in Total Introns

Pentamers		Obs/Exp	Hexamers		Obs/Exp
Word	Count		Word	Count	
GCATG	160	2.43	TGCATG	75	3.35
CTCTC	273	2.26	CTCTCT	173	3.35
TCTCT	391	2.01	TCTCTC	186	3.15
CTATC	142	1.85	CCAGGG	20	3.06
CGCTT	125	1.84	CTAACC	46	2.93
TGTGT	306	1.80	CCCTGG	21	2.92
CTAAC	156	1.79	CGCATG	31	2.76
GTGTG	194	1.76	GGGTCG	16	2.74
GGGGG	93	1.71	CTACCC	28	2.67
TAACC	150	1.69	ACCGTG	25	2.56
ACCTC	110	1.66	GCATGT	58	2.55
GTCGT	98	1.65	TCTATC	76	2.55
CCCAC	108	1.64	GCATGA	47	2.53
CCTCT	131	1.62	GCATGC	33	2.50
GTTGT	237	1.62	TGTCCC	38	2.41
GTCCC	68	1.62	ACGGGC	18	2.41
TGTCC	110	1.62	CGCAGT	25	2.37
CTCCC	97	1.61	CTATCC	30	2.35
CCTCC	101	1.61	CCTCCC	33	2.34
CGCAC	69	1.60	CTATCT	51	2.27
CTTCT	310	1.59	GGGGGG	47	2.24
TGCAT	205	1.59	AGCCCG	17	2.22
CATGT	179	1.58	ACCTCT	46	2.19
CCGCA	86	1.56	CCAACC	32	2.16
CATGC	95	1.55	CTGCAC	27	2.16
CAACC	109	1.54	TGTGTG	105	2.16
CTCCT	126	1.54	CCGCAG	19	2.15
TCCTC	148	1.53	GTTCGC	22	2.15
ATCCT	137	1.52	CGAACC	24	2.14
TCCCT	117	1.50	GCAGTC	20	2.14
CGAAC	95	1.50	GTCCCA	33	2.13
TCCCC	108	1.50	ACTAAC	57	2.12
CGTCC	49	1.50	TCCTCC	51	2.12
TGTCT	182	1.50	CGCTTC	31	2.11
GCCCG	43	1.50	CTGCGC	18	2.09
CCGTG	65	1.49	TGCACG	23	2.08
GGCTT	152	1.49	AGCATG	36	2.08
CATCC	86	1.49	CCGCAC	23	2.07
CTACC	101	1.49	GCACGG	18	2.07
CCCTG	61	1.48	CCGTCT	26	2.06

We determined the number of occurrences of every pentamer and hexamer motif in the complete introns flanking 147 alternatively spliced exons that contain the evolutionarily conserved elements used for Table 1. The Count column indicates the number of occurrences of each motif in the dataset of introns flanking the 147 alternative exons. Obs/Exp is the ratio of the observed frequency of occurrence of these motifs in the introns flanking alternative exons dataset divided by the expected frequency score of each motif based on the control total intron dataset. The top 40 highest scoring motifs as determined by the obs/exp value for pentamers and hexamers are listed.
DOI: 10.1371/journal.pcbi.0020086.t002

these full-length alternative introns. We ranked these motifs by the obs/exp ratio when we compared the occurrence of these motifs to their prevalence in our total intron dataset. The top 40 pentamer and hexamer motifs ranked by their obs/exp ratios are shown in Table 2. Many of the top scoring pentamers and hexamers that are overrepresented in conserved elements in introns flanking alternatively spliced exons are also overrepresented in total introns flanking alternatively spliced exons (compare Table 1 with Table 2). However, the obs/exp scores for the motifs in the alternative splicing introns is 2- to 3-fold lower than in the conserved element dataset extracted from these introns. For example, TGCATG has an observed to expected ratio of 3.35 in the

whole introns that flank alternative exons but 8.87 for the conserved elements within those introns. TCTATC has a ratio of 2.55 in the total alternatively spliced intron analysis and 6.96 in the conserved element dataset analysis. The same holds true for high-scoring pentamers. CTATC goes from 1.85 for the observed to expected ratio in the introns flanking alternative exon dataset up to 4.23 in the conserved element dataset. From this analysis, the sequences of the introns flanking alternatively spliced exons can yield some data about potential pentamers and hexamers involved in alternative splicing regulation. However, limiting this analysis to WABA-

conserved regions between *C. elegans* and *C. briggsae* can substantially improve the signal to help us identify splicing regulatory elements.

We also used the database of introns flanking alternatively spliced exons to identify potential differences in the appearance of pentamer and hexamer motifs in the introns upstream or downstream of alternative exons. We grouped the introns upstream of the alternative exons and the introns downstream of the alternative exons separately into datasets and did motif counts and determined the obs/exp ratio for each when compared to the total intron dataset. These results are shown in Tables 3 and 4, with the top scoring pentamers and hexamers in the introns upstream and downstream of alternative exons ranked by their obs/exp ratios. Some of the high- scoring motifs show very little preference for the upstream or downstream intron. For example, high-scoring hexamer TGCATG has an obs/exp ratio of 3.90 in the downstream intron and 3.06 in the upstream intron. Others show a strong preference. CTCTCT and TCTCTC, likely

binding sites for PTB, have obs/exp ratios of 5.15 and 4.84, respectively, in upstream introns, but 1.36 and 1.30 in downstream introns. TCTATC, which will be analyzed below, also has a preference for upstream introns, with an obs/exp ratio of 3.25 in upstream introns and 1.79 in downstream introns. Conversely, CCAACC has a strong preference for downstream introns, with an obs/exp ratio of 3.17 in the downstream intron and 1.29 in the upstream intron. In addition to identifying potential binding sites for splicing regulatory factors, these differences in appearance in introns upstream or downstream of alternatively spliced exons may hold some keys to function.

To determine whether these pentamers and hexamers are found in introns flanking all alternative cassette exons or are specific for those introns with WABA-conserved elements, another dataset was constructed from the introns flanking the alternative cassette exons of 307 genes lacking WABA-defined evolutionarily conserved intronic elements. A motif search similar to that described in the previous paragraph was

Table 3. Comparison of the Occurrence of Pentamer Motifs in Introns Upstream versus Downstream of Alternative Cassette Exons

Pentamers

Upstream Intron Ranks								Downstream Intron Ranks							
US Rank	Motif	US Obs/Exp	DS Rank	DS Obs/Exp	NC Rank	NC Obs/Exp		DS Rank	Motif	DS Obs/Exp	US Rank	US Obs/Exp	NC Rank	NC Obs/Exp
1	CTCTC	3.18	162	1.25	130	1.31		1	GCATG	2.88	4	2.11	19	1.6
2	TCTCT	2.74	202	1.22	72	1.39		2	TGTGT	1.96	20	1.69	46	1.44
3	CTATC	2.28	81	1.39	166	1.27		3	TAACC	1.92	62	1.5	266	1.14
4	GCATG	2.11	1	2.88	115	1.33		4	GTGTG	1.9	23	1.66	114	1.29
5	CGTCC	1.92	457	1.05	8	1.82		5	ACCTC	1.9	82	1.46	15	1.64
6	CTCCC	1.87	116	1.31	296	1.15		6	CGCAC	1.88	147	1.37	511	1.01
7	GCACG	1.86	464	1.04	28	1.54		7	CGCTT	1.85	9	1.85	156	1.22
8	GGGGG	1.86	37	1.53	383	1.09		8	GTCCC	1.83	86	1.45	26	1.54
9	CGCTT	1.85	7	1.85	118	1.33		9	CCGTG	1.81	261	1.22	170	1.21
10	CTAAC	1.83	14	1.74	88	1.36		10	TGTCC	1.79	71	1.48	14	1.64
11	GTCGT	1.79	50	1.47	29	1.54		11	CATGC	1.77	121	1.39	9	1.69
12	CACTC	1.79	455	1.05	429	1.07		12	TGCAT	1.75	66	1.49	128	1.27
13	GCCCG	1.79	242	1.19	37	1.5		13	CCTCT	1.74	68	1.48	190	1.2
14	CCTCC	1.79	68	1.42	34	1.51		14	CTAAC	1.74	10	1.83	140	1.25
15	TCCGC	1.77	430	1.06	30	1.54		15	CAACC	1.71	144	1.37	543	1
16	CCGCT	1.76	263	1.17	181	1.25		16	GCACC	1.69	622	0.99	330	1.1
17	GGCTT	1.73	179	1.23	10	1.79		17	CATGT	1.69	65	1.49	298	1.11
18	CTTCT	1.72	41	1.49	102	1.35		18	GTTGT	1.65	26	1.64	226	1.17
19	CCCCG	1.7	174	1.24	178	1.25		19	CTACC	1.63	118	1.4	27	1.53
20	TGTGT	1.69	2	1.96	158	1.28		20	CCGCA	1.62	57	1.52	181	1.2
21	GCTTC	1.69	158	1.26	73	1.39		21	CGAAC	1.62	99	1.42	482	1.03
22	CCCAC	1.67	22	1.62	361	1.11		22	CCCAC	1.62	22	1.67	641	0.96
23	GTGTG	1.66	4	1.9	157	1.28		23	CTCCT	1.61	101	1.42	20	1.6
24	CGTCT	1.65	252	1.18	267	1.17		24	CCCCA	1.59	591	1.01	86	1.34
25	GGTGC	1.64	138	1.28	106	1.34		25	AACCT	1.58	617	0.99	834	0.86
26	GTTGT	1.64	18	1.65	112	1.33		26	CGGGG	1.57	935	0.79	89	1.34
27	AGAGT	1.63	365	1.09	303	1.14		27	CAGGG	1.57	111	1.4	987	0.71
28	TGCTG	1.63	350	1.11	200	1.22		28	CACCG	1.57	882	0.83	307	1.11
29	CCCTG	1.62	98	1.35	264	1.17		29	GAGCC	1.57	409	1.11	877	0.83
30	CGCAT	1.61	382	1.08	648	0.99		30	TACCC	1.57	329	1.17	148	1.24

Left side of table is analysis of upstream pentamer motifs. US Rank is the relative rank for the top 30 motifs based on Obs/Exp ratio (US Obs/Exp) for motifs in upstream introns that contain WABA-conserved elements (conserved introns). The relative rank of these motifs in the downstream conserved intron (DS Rank) and their obs/exp ratios for the downstream conserved intron dataset (DS Obs/Exp) are also indicated. NC Rank is based on Obs/Exp ratio for motifs found in upstream alternative introns lacking WABA-conserved elements (non-conserved introns). NC Obs/Exp indicates the ratios for each motif in the non-conserved intron analysis. Right side of table is analysis of downstream pentamers. DS Rank is the ranking of the motifs in conserved downstream introns based on the Obs/Exp ratio for motifs in downstream introns (DS Obs/Exp). The relative ranking of these motifs in the upstream introns (US Rank) and their Obs/Exp ratios in upstream introns (US Obs/Exp) are also indicated. NC Rank is based on Obs/Exp ratio from downstream non-conserved introns. NC Obs/Exp indicates the ratios for each motif in the non-conserved intron analysis. See Table S2 for a more complete analysis of motifs in non-conserved introns.
DOI: 10.1371/journal.pcbi.0020086.t003

Table 4. Comparison of the Occurrence of Hexamer Motifs in Introns Upstream versus Downstream of Alternative Cassette Exons

Hexamers

Upstream Intron Ranks							Downstream Intron Ranks						
US Rank	Motif	US Obs/ Exp	DS Rank	DS Obs/ Exp	NC Rank	NC Obs/ Exp	DS Rank	Motif	DS Obs/ Exp	US Rank	US Obs/ Exp	NC Rank	NC Obs/ Exp
1	CTCTCT	5.15	780	1.36	579	1.43	1	TGCATG	3.9	8	3.06	178	1.75
2	TCTCTC	4.84	914	1.3	688	1.38	2	ACCGTG	3.71	507	1.56	122	1.85
3	ACGGGC	3.56	1,464	1.14	1,405	1.15	3	CTACCC	3.65	239	1.81	1,325	1.15
4	CCAGGG	3.49	23	2.61	1,523	1.13	4	GGGTCG	3.65	146	1.95	3,866	0.52
5	TCTATC	3.25	208	1.79	721	1.37	5	CGAACC	3.24	1,453	1.19	2,333	0.94
6	CTAACC	3.15	14	2.85	197	1.73	6	GCATGT	3.19	114	2.01	1,150	1.19
7	CCTCCC	3.1	512	1.51	351	1.58	7	CCAACC	3.17	1,119	1.29	386	1.53
8	TGCATG	3.06	1	3.9	97	1.93	8	GCATGA	3.1	76	2.15	208	1.71
9	GCAGTC	3.06	1,456	1.14	2,636	0.93	9	GGGGTC	3.04	2,091	1.04	4,079	0.17
10	CTATCT	2.96	508	1.52	846	1.32	10	ACCCGC	3.04	2,355	0.99	3,762	0.59
11	TGCACG	2.93	1,395	1.16	545	1.45	11	TGTCCC	2.97	167	1.93	44	2.1
12	CCCTGG	2.91	12	2.96	2,045	1.03	12	CCCTGG	2.96	12	2.91	1,773	1.05
13	GGGGGG	2.9	480	1.53	3,439	0.77	13	GCATGC	2.9	48	2.3	24	2.4
14	CTATCC	2.83	175	1.83	389	1.54	14	CTAACC	2.85	6	3.15	193	1.73
15	CTGCAC	2.74	460	1.54	928	1.29	15	CGCATG	2.85	18	2.71	726	1.34
16	CGCTTC	2.72	594	1.45	581	1.43	16	GCACCG	2.81	3,362	0.75	611	1.39
17	CATCGC	2.72	2,373	0.94	3,219	0.81	17	AGCATG	2.71	544	1.54	93	1.92
18	CGCATG	2.71	15	2.85	3,923	0.55	18	CATGCT	2.69	1,203	1.26	500	1.45
19	TGCTGC	2.65	3,311	0.71	867	1.31	19	CGTGAC	2.67	1,440	1.19	190	1.73
20	GTTCGC	2.6	297	1.67	557	1.45	20	GGCACC	2.64	3,479	0.72	1,327	1.15
21	CGGGCT	2.59	1,389	1.16	3,713	0.67	21	TGTGTG	2.63	283	1.76	486	1.46
22	GCACGA	2.55	2,913	0.82	2,536	0.94	22	CGCAGT	2.63	72	2.16	877	1.29
23	CCTGGT	2.52	655	1.41	2,660	0.92	23	CCAGGG	2.61	4	3.49	4,061	0.23
24	GTGTTG	2.52	421	1.57	1,513	1.13	24	GTGGCT	2.58	3,884	0.54	2,189	0.97
25	ACGTCC	2.51	3,689	0.56	31	2.28	25	CGCACC	2.56	2,220	1.02	293	1.61
26	CCCGGT	2.51	3,772	0.51	290	1.63	26	TAACCT	2.56	1,196	1.26	1703	1.06
27	CACTGC	2.5	284	1.68	1,155	1.22	27	CGCTTG	2.55	596	1.51	1,606	1.08
28	ACGCTT	2.49	486	1.52	304	1.62	28	ACCTCT	2.54	181	1.9	330	1.58
29	TATCCT	2.48	611	1.44	313	1.61	29	GGTCGA	2.51	861	1.38	262	1.64
30	TCCGCT	2.47	3,099	0.77	442	1.51	30	CCGCAC	2.49	324	1.71	693	1.36

Analysis of hexamer motifs. The legend for Table 3 also applies, except that here hexamers are ranked. See Table S2 for a more complete analysis of motifs in non-conserved introns.
DOI: 10.1371/journal.pcbi.0020086.t004

conducted by dividing the new, non-conserved intron dataset into two sets: introns upstream of alternative exons and those downstream. Motifs were counted for each of the sets and compared to the total intron dataset to generate an obs/exp ratio and corresponding ranking for each element. It can be seen from Table 4 that the obs/exp scores of all of the top 30 hexamer motifs identified in conserved introns drop in the analysis of non-conserved introns, suggesting the motifs are indeed specific to introns containing conserved regions. For example, our top ranking hexamer from the conserved intron analysis, TGCATG, has an obs/exp score of 3.9 in downstream introns and 3.06 upstream. This same motif only scores 1.75 in downstream and 1.93 in upstream in non-conserved alternative introns (Table 4). For CTCTCT and TCTCTC, our two top hexamer upstream motifs in introns with WABA-conserved elements, the obs/exp ratio drops over 3-fold in this analysis, and they rank as 579 and 688, respectively, in the upstream non-conserved intron set (Table 4). Conversely, top scoring motifs from the non-conserved intron database are not highly represented in the conserved intron dataset. For example, GGCCAC, with an obs/exp ratio of 4.36 (Table S2) was the top scoring motif in the upstream non-conserved dataset. Strikingly, this same motif has an obs/exp score of 0.11 in total introns that contain WABA-conserved elements

upstream of an alternatively spliced exon (unpublished data). This dramatically different representation of top scoring pentamers and hexamers in introns containing WABA-conserved elements versus those that do not suggests that the presence or absence of intronic WABA-defined conservation may define two distinct classes of alternatively spliced exons with distinct splicing regulatory mechanisms.

To ensure that the motifs identified in the conserved element dataset were truly overrepresented in this region and not a consequence of WABA creating bias towards GC-rich regions in the AT-rich nematode genomes, we analyzed the base composition of our datasets. The GC content of our WABA-defined conserved element dataset was 39%, the total intron dataset that contained these conserved elements was 34%, and the total intron reference dataset was 32%. While this slight variation in GC content might suggest that the WABA HMM may have created a bias, two arguments can be made against this. The first is that the high-scoring motifs in the WABA-conserved region were also highly represented in the total introns containing these motifs (Table 2), consistent with these results not being derived from bias in GC content. The second is that GC enrichment bias is not observed in many of our highest scoring motifs. For example, TCTATC

has a GC content of 33%, yet is highly enriched in conserved elements.

The Candidate Motifs (T)GCATG and (T)CTATC Are Required for *unc-52* Alternative Splicing Regulation

The *C. elegans unc-52* gene is a homolog of the mammalian extracellular matrix protein perlecan and contains 37 exons. Exons 16, 17, and 18 are alternatively spliced in a complex and regulated manner. These three exons each encode an Ig protein motif and can be included or skipped in any number of combinations in the final *unc-52* mRNA transcript [45]. RT-PCR analysis shows *unc-52* is also alternatively spliced in *C. briggsae*, suggesting the splicing pattern may be controlled by similar methods (unpublished data). The use of a subset of these alternatively spliced forms is controlled by the splicing regulatory protein, MEC-8 [46,47]. The genetic regulation of *unc-52* alternative splicing makes this gene an attractive model for studying splicing regulation. Most importantly for this analysis, the introns flanking either side of alternative exon 16 contain regions of high nucleotide conservation as identified by WABA. A number of the top scoring pentamers and hexamers from our conserved element motif analysis are found in these conserved regions (Figure 3A).

To test the role of these motifs in alternative splicing of *unc-52*, we created a muscle-expressed alternative splicing reporter transgene derived from this region (Figure 3B). Expression studies using the native *unc-52* promoter to drive GFP expression have indicated that this extracellular matrix protein is expressed in the muscle and hypodermis [47,48]. However, studies of the *unc-52* splicing regulator *mec-8* have suggested that *mec-8* function on *unc-52* alternative splicing is mostly focused in hypodermal tissues [47]. Our intention is to mimic the native gene's splicing as closely as possible, so we tested whether the alternative splicing of our muscle-specific *unc-52* reporter could be regulated by *mec-8*. RT-PCR analysis of our wild-type *unc-52* alternative splicing reporter in *mec-8(+)* (wild-type) and *mec-8(e398)* (mutant) backgrounds indicates a dramatic difference in splicing (Table S3). When the reporter is spliced in a *mec-8(e398)* mutant that lacks its RNA-binding domain, an increase in exon 18-containing transcripts is observed, as well as a 75% drop in abundance of 15-16-19 and 15–19 isoforms. This closely mimics the effects of *mec-8(e398)* on native *unc-52* splicing [46]. These results demonstrate that the tissue-specific expression of our *unc-52* reporter is an adequate representation of the alternative splicing of the native gene, and that it is also under the control of the splicing factor MEC-8.

In order to test the role of evolutionarily conserved intronic elements and motifs in the regulation of *unc-52* alternative splicing, we created *unc-52* alternative splicing reporter constructs with different combinations of deletions in these sequences and monitored their in vivo splicing by RT-PCR. Due to the similar size of exons 17 and 18, a restriction digest with BamHI was performed on the RT-PCR
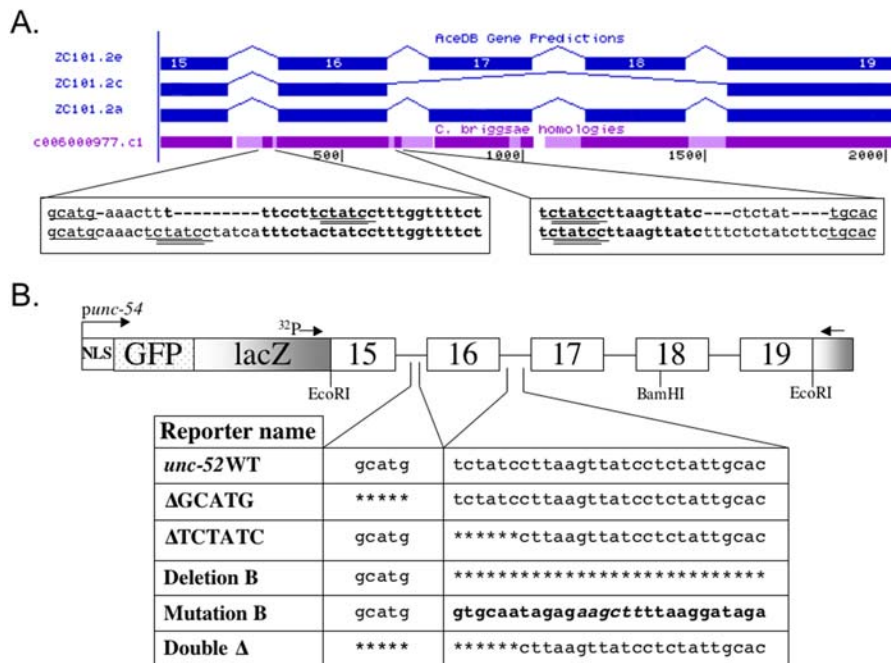

Figure 3. Conserved Intronic Elements of *unc-52* Contain Putative Regulators of Alternative Splicing

(A) *C. elegans* and *C. briggsae* sequence alignment is shown for the alternatively spliced portion of *unc-52*. Not all spliced isoforms are predicted by Wormbase software (blue); see Figure 4A for observed alternative splicing patterns. PhastCons sequence alignment is shown with WABA-designated conservation in bold. Upper line of sequence is *C. elegans*; *C. briggsae* is below. High-scoring conserved motifs identified in our pentamer/hexamer analysis of conserved intronic elements flanking alternatively spliced exons, GCATG, TCTATC, CTATCC, CTATC, and TGCAC are underlined.

(B) Diagram of alternative splicing reporter constructs for testing putative *cis*-regulatory splicing motifs. Part of exon 15 through part of exon 19 of *unc-52* was cloned into a GFP/lacZ fusion vector with an *unc-54* promoter and nuclear localization sequence suitable for expression in *C. elegans*. Site-directed mutagenesis of the wild-type substrate was performed in order to test putative *cis*-splicing regulatory elements. A table of the splicing reporter constructs and their alterations is shown. Asterisks denote highly conserved intronic nucleotides deleted by site-directed mutagenesis. To maintain the intron length, yet remove motifs in question, a reporter was also made in which native sequence was replaced with the reverse complement sequence (shown in bold) and a HindIII site (italics) for diagnostic purposes.
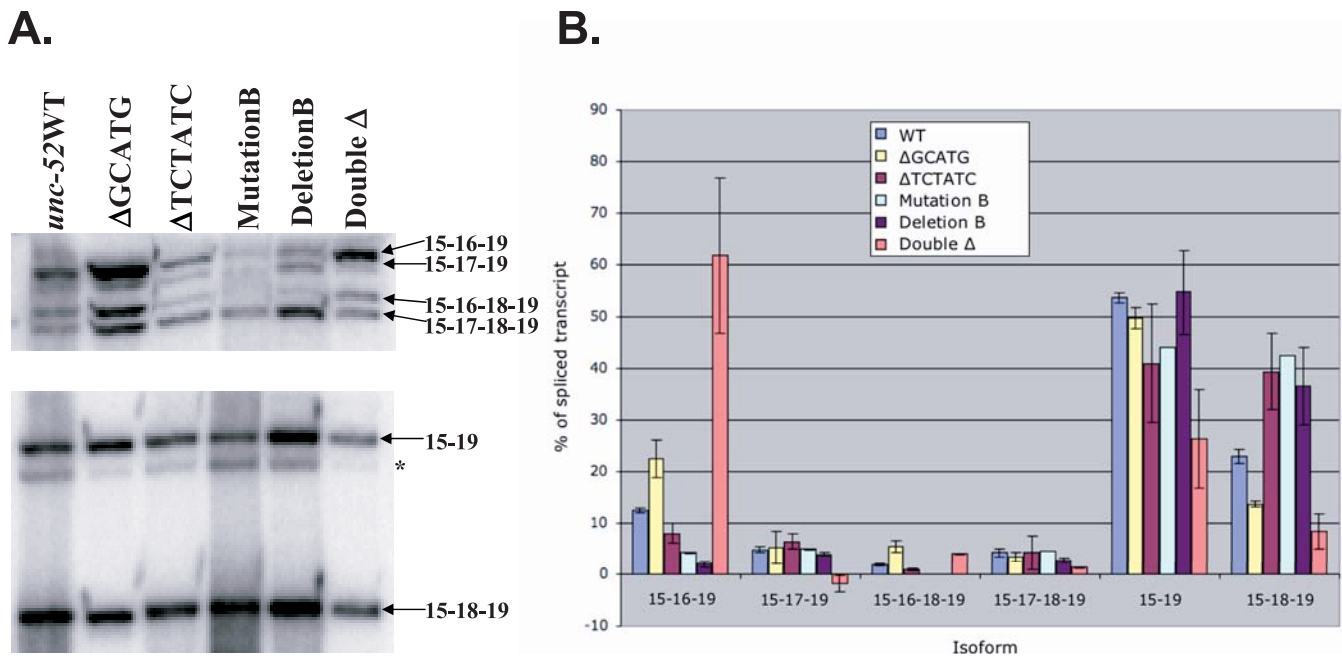
DOI: 10.1371/journal.pcbi.0020086.g003

Figure 4. Deletion or Mutation of Top Scoring Pentamers or Hexamers Alters *unc-52* Alternative Splicing

(A) Autoradiogram of *unc-52* reporter gene alternative splicing. Left side shows ^{32}P RT-PCR analysis (with BamHI digest) performed on strains carrying each of the reporter constructs described in Figure 3B. The reporter is indicated at top of gel. Table on right shows the relative percentage of each alternatively spliced isoform produced from the indicated in vivo splicing reporter constructs. Arrows point to the corresponding band on the gel. Asterisk denotes a non-specific band resulting from RT-PCR.

(B) Graphical summary of *unc-52* in vivo splicing reporter assays. Horizontal axis indicates spliced mRNA isoform. Vertical axis represents the relative percentage of each isoform. Color key on figure indicates which bars correspond to which strain. Standard deviation and mean values were calculated based on a minimum of three independent RNA extractions and subsequent RT-PCRs for each reporter.

DOI: 10.1371/journal.pcbi.0020086.g004

products. Spliced isoforms containing exon 18 would be cleaved by BamHI and thus run at smaller sizes on the gel (Figures 3B and 4A). A summary of these results is shown in Figure 4. All quantitation is based on the analysis of a minimum of three different RNA isolations from the indicated strains. In general, splicing reporters with deletions of the conserved intronic sequences flanking either side of exon 16 seemed to have the largest effect on three specific isoforms of mature *unc-52* mRNA, those containing exons 15-16-19, 15-16-18–19, and 15-18-19 (Figure 4B).

Our pentamer and hexamer motif analysis of conserved intronic elements flanking alternatively spliced exons identified GCATG and TGCATG as high-scoring motifs likely to be found in introns flanking either side of an alternative exon. Deletion of the GCATG upstream of alternative exon 16 from the *unc-52* splicing reporter resulted in nearly a doubling of the proportion of reporter transcripts containing exon 16 and a drop in the proportion using 15-18-19. This suggests that GCATG may play a role in repressing the inclusion of exon 16 (Figure 4B).

A conserved intronic element just downstream of exon 16 contains another of our top scoring motifs, (T)CTATC. A deletion of this TCTATC had the opposite effect of deleting the upstream intronic GCATG; the proportion of exon 16-containing transcripts decreased while the amount of 15-18-19 increased. A larger deletion of this downstream motif (Deletion B) or a replacement of the motif with the reverse complement sequence (Mutation B) amplified this effect as summarized in Figure 4B. Specifically, an approximately 75%

reduction in exon 16- containing transcripts relative to wild-type can be seen when this conserved region is altered in the Mutation B construct. A similar drop of almost 85% is seen when this conserved element is deleted in the Deletion B reporter construct. This indicates that TCTATC is part of a regulatory element that enhances the inclusion of exon 16, and its loss by mutation leads to a decrease in exon 16 inclusion.

Perhaps most interesting was the effect produced by the double deletion of two conserved motifs flanking exon 16: the upstream GCATG and the downstream TCTATC. Individual deletions of these two motifs suggested they work in opposition to include or exclude exon 16. However, the double mutant causes a shift in splice pattern similar to, but much more dramatic, than that seen in the ΔGCATG construct (Figure 4B). The double mutant had such a dramatic increase in 15-16-19 levels that it became the new predominant isoform, making up 61% of the mRNAs. This result suggests that the conserved motifs on either side of exon 16 are responsible for the inclusion of this exon and that they work in collaboration and as part of a larger regulatory network to produce a multilayered regulation of *unc-52* splicing. In general, changes to the evolutionarily conserved intronic elements flanking exon 16 did not alter the relative amounts of the 15-17-19, 15-17-18–19, and 15–19 isoforms dramatically. These elements flanking exon 16, for the most part, appear to regulate the decision to form either the 15-16-19 isoform or the 15-18-19 isoform (Figure 4B).

Discussion

Our comparative genomic analysis of *C. elegans* and *C. briggsae* allowed us to identify 142 alternatively spliced genes that exhibit regions of high nucleotide conservation in introns flanking alternatively spliced exons. These conserved regions were then analyzed for pentamer and hexamer motifs present at a statistically higher level than found in total introns. Many of the high scorers on these lists matched known mammalian splicing regulatory elements. This indicates that this approach can find alternative splicing regulatory sequences, and it is consistent with our observations that there are *C. elegans* homologs for the major mammalian splicing regulatory factors (unpublished data). In addition to finding known splicing regulatory sequences, this approach identified potentially novel splicing regulatory sequences. We have confirmed that the sequence (T)CTATC is important for alternative splicing regulation of the *unc-52* gene.

The limited sequence conservation between *C. elegans* and *C. briggsae* in introns flanking alternatively spliced exons contrasts with that seen in mammalian interspecies genome alignments. In mouse/human alignments of alternatively spliced human cassette exons, it was found that the 100 nucleotides of intron sequence flanking either side of alternative cassette exons tend to be highly conserved [21,22]. It is unclear from these mouse/human analyses which specific portion(s) of the introns immediately flanking these alternatively spliced exons are essential for alternative splicing regulation. Our analysis indicates that the WABA-identified regions of intronic homology flanking alternatively spliced exons are likely to be important for alternative splicing regulation. Therefore, the *C. elegans/C. briggsae* evolutionary distance, along with the sensitivity of WABA, provides for a more simplified method for pinpointing splicing regulatory sequences. One of the interesting questions resulting from our identification of intronic alternative splicing regulatory motifs in *C. elegans* is whether these will also function to regulate alternative splicing in mammals. It has been suggested that there is a difference in frequency of usage of constitutive intronic splicing enhancers between mammals and fish, which can explain why fish introns are not always spliced efficiently in mammalian cells [49]. No similar study for alternative splicing regulatory elements has yet been done. Our data on *C. elegans* alternative splicing regulatory elements suggest that there are many similarities in alternative splicing regulation across metazoans. This is seen in the fact that many of the motifs we identified in worms are known to regulate alternative splicing in mammals. Whether novel elements we have identified, such as TCTATC, can also function in mammals has yet to be determined.

The identification of (T)GCATG as a top scoring alternative splicing regulatory motif demonstrated the usefulness of this approach, as many examples of this motif as a splicing regulatory element are present in the literature [14,15,50]. (T)GCATG is statistically overrepresented in introns downstream of alternative exons conserved between human and mouse genomes [21,22] and may have a spatially conserved role in directing splice-site choice [23]. On the other hand, in vitro studies have demonstrated that the splicing factor FOX-1 affects splicing of transcripts with GCATG either upstream or downstream of an alternative exon [13]. One interesting finding of our *C. elegans* work is that GCATG in introns flanking alternatively spliced exons shows no preference for being upstream or downstream of alternative exons. This may be evidence that while the motifs that regulate alternative splicing between mammals and nematodes may be the same, they may be used in different ways to promote splicing.

A recent analysis of the *C. elegans* genome used a support vector machine to identify predictive features of *C. elegans* alternative exons [51]. One of the predictive features of an alternatively spliced exon that they identified is the presence of several different hexamers in the introns surrounding alternative exons. In their supplemental materials (http://www2.fml.tuebingen.mpg.de/raetsch/projects/RASE) they list the top 12 predictive hexamers in the introns upstream and downstream of the alternative exon. Comparison of their high-scoring results with our conserved element analysis indicates that many of the top scorers in both approaches are identical (for example, TGCATG, CTAACC, GTGTGT, CTCTCT, and TCTATC). That both their support vector machine and our comparative genomics methods yielded similar results helps to confirm the validity of both methods.

Comparing the motif searches from different classes of intronic sequence (total introns, WABA-conserved elements, entire conserved introns flanking alternative exons, and non-conserved introns flanking alternative exons) brought about a number of insights concerning the presence of certain motifs in introns. First, it can be noted that many of these sequences are still found, although at lower frequency, in total introns as well as those that are alternatively spliced. They may therefore, in a method analogous to SR protein-binding sites in exons, play a role in constitutive as well as alternative splicing (SR proteins reviewed in [52]). Second, the comparison of motif analyses of introns lacking conserved elements and those containing WABA conservation reveals two dramatically different lists of overrepresented motifs. This suggests that there may in fact, be two separate classes of cassette exon splicing regulation, and WABA can help us distinguish between them.

Our splicing reporter system allowed us to directly test the role of the conserved elements GCATG and TCTATC flanking alternative exon 16 of the *unc-52* gene in vivo. Deletion of either of these elements individually led to measurable but opposing effects on the inclusion of this exon (Figure 4). These first results suggested the two *cis* elements might be counter-balancing each other to regulate splicing in this region. However, the creation of a double deletion of these two elements demonstrated that this was a naive assumption. The splicing program exhibited by the double deletion construct was similar to that seen from the ΔGCATG construct: an upregulation of the exon 16-containing transcripts. However, the proportional increase of these isoforms was far more dramatic in the double than the single deletion. The 15-16-19 isoform became the predominant spliced transcript, as opposed to 15–19 seen for all other splicing reporter constructs. This striking result is a good reminder that although these motifs can work independently and their deletion may produce measurable effects on splicing, they may work in a combinatorial way in vivo, and the effects of multiple deletions of important elements may be hard to predict. This type of combinatorial multifactor splicing regulatory mechanism has been described for several vertebrate genes including *c-src* [11] and cardiac troponin T [17].

While our comparative genomics approach gives us a method of identifying putative *cis*-regulatory elements of alternative splicing and our in vivo reporter assay provides a method for directly testing candidates, we still need more experimental information before we can predict the combinatorial effects of these elements on splicing.

Our comparative genomics analysis has not only provided a method for identifying intronic *cis*-regulatory elements, but also a starting point to investigate the mechanisms by which they control splicing. Consistent with current models of alternative splicing, *cis* elements are predicted to be binding sites for *trans* factors. Our identification of the novel splicing regulatory sequence (T)CTATC will lead us to subsequent experiments to identify its potential protein partner. The lists we generated of potential splicing regulatory elements include many sequences that have yet to be tested for a role in alternative splicing regulation. Experiments still need to be performed, similar to those done on *unc-52* and *let-2,* in which these elements are tested for a functional role in alternative splicing regulation. Due to the rapid release of new alignment programs and genomes, our list of conserved pentamers and hexamers will likely see some adjustments before all potential regulatory motifs have been tested. WABA has a limitation of needing to see relatively large sequence alignments in order to call a region as high homology (the minimum WABA high homology sequence run we detected was seven consecutive identical nucleotides). WABA high homology regions in introns flanking alternatively spliced exons were only detected for approximately 25% of alternative cassette exons in our dataset. The newer alignment algorithm PhastCons [53], which uses an HMM to align multiple genomes using a smaller window size than that of WABA, can allow us to more accurately pinpoint smaller regions of conserved nucleotides within introns flanking additional alternatively spliced exons (unpublished data). The release of additional nematode genomes such as *Caenorhabditis remanei* should provide us with the prospect of creating a three-way nematode genomic alignment, allowing us even more accuracy in our regulatory motif predictions. By systematically testing the pentamers and hexamers on our list of conserved motifs, we may be able to confirm more alternative splicing regulatory motifs. A compilation of these results will provide us with a better comprehension of the rules governing this complex and essential process.

Materials and Methods

Database construction. A control dataset of introns was obtained by downloading from the UCSC Genome Browser (http://www. genome.ucsc.edu) all of the introns in the *C. elegans* genome as annotated in Wormbase release WS120. From the sequences of these 118,492 introns, we removed the first and last seven nucleotides to avoid constitutive splicing signals. We then counted the number of occurrences of every possible hexamer or pentamer motif in the database by using the EMBOSS Compseq program [54]. For each motif in the database, we assigned a frequency score that was the number of occurrences of each motif divided by the total number of possible occurrences of motifs of that length in the dataset. This score for each possible motif in the control set was used in our analysis of both the introns that flank alternatively spliced exons and the conserved elements in these introns as our expected score. In order to obtain the obs/exp ratio for each motif in a test dataset, we divided the observed frequency score for each motif in that dataset by the expected frequency score for each motif based on the total intron control dataset.

Analysis of alternative exons. *C. elegans* and *C. briggsae* alignments for the 147 alternative spliced exons with intronic conservation were obtained from the UCSC Genome Table Browser conservation track [55]. The mean nucleotide percentage identity was computed from this alignment along with the number of insertions and deletions. In order to obtain the mean amino acid identity, all the 147 exons were compared to the complete *C. briggsae* genome (WormBase cb25.agp8) using TBLASTX [56], and the mean amino acid identity was then calculated from these alignments.

Generation of *let-2* splicing reporter constructs. We have previously described the development of an in vivo splicing reporter for the alternatively spliced region of the *let-2* gene fused to GFP and expressed in muscle cells [37]. Site-directed mutagenesis of the wild-type reporter for *let-2* alternative splicing from that paper was used to delete the core of the first conserved element, indicated in Figure 2, and replace it with GAA. Methods for detecting alternative splicing of this transcript with ^{32}P-labeled primers using RT-PCR, denaturing polyacrylamide gel electrophoresis, and autoradiography have been previously described [37].

Generation of *unc-52* splicing reporter constructs. The exon 15–19 region of *unc-52* was PCR-amplified from wild-type (N2) *C. elegans* genomic DNA using primers 5′-GGAATTCGATGAGTACATCTG-TATCGC and 5′-GGAATTCACATCTGAACTGATGTCGCTC for cloning into the vector pPD96.02, developed by Andrew Fire's lab. This vector contains the *unc-54* body wall muscle-specific promoter driving expression of a green fluorescent protein (GFP)/β galactosidase (lacZ) fusion protein with an N-terminal nuclear localization signal derived from SV40 T antigen. The final 127 nucleotides of exon 15 through the first 173 nucleotides of exon 19 of *unc-52* and all the genomic sequence in between were cloned into a unique EcoRI site, 16 codons before the end of the lacZ open reading frame in the final exon of the fusion protein. The Kunkel method of site-directed mutagenesis was used to create the larger deletion and mutations for the plasmid constructs Mutation B and Deletion B [57]. Pentamer and hexamer deletion constructs of the conserved *unc-52* intron sequence were created using the Quickchange site-directed mutagenesis kit (Stratagene, La Jolla, California, United States) according to the manufacturer's instructions. Animals carrying these constructs as extra-chromosomal arrays were generated by standard injection/transformation [58]. Transformed N2 animals were identified by GFP-expression in nuclei of body wall muscle cells.

Growth of *C. elegans* strains and isolation of RNA. *C. elegans* strains were grown on NGM agar plates using standard methods [59]. RNA was extracted from worms as previously described [60]. Strain CB938 carrying the *mec-8(e398)* mutant allele was obtained from the *C. elegans* Genetics Center, a National Institutes of Health funded Center for Research Resources at the University of Minnesota.

Production of cDNAs. cDNAs of the *unc-52* splicing reporters were made in 25-μl reaction mixtures. The annealing step contained 3 μg total RNA and 25-pmol oligodeoxynucleotide primer complementary to the lacZ construct (5′- GTTGAAGAGTAATTGGACTTA-3′ for *C. elegans* and 5′- AACTGGTGTCGCTCTCCT-3′ for *C. briggsae*) in 17-μl final volume. These were heated to 94 °C for 2 min and cooled to room temperature for 10 min. The rest of the reverse transcription reaction components were added to reach a final volume of 25 μl. The final reaction mixtures consisted of 1 mM each of dATP, dCTP, dGTP, and dTTP; 1 U RNA Guard (Promega, Madison, Wisconsin, United States); 1 × AMV RT buffer (Promega); and 10 U AMV reverse transcriptase (Promega). Reactions were incubated at 37 °C for 1.5 h and stored at −20 °C.

Polymerase chain reaction. 1.0 μl of the cDNA reaction mixture was used as the template in 25-μl PCR reaction mixtures. 1.0 pmoles of the same oligonucleotide used in the reverse transcription reactions was added to PCR reaction mixtures along with ^{32}P-labeled 5′ lacZ vector-specific primer (5′-CTGGAGCCCGTCAGTATCGGC-3′ for *C. elegans* and 5′-ATTCGTTGCTGGGTCCCAGG-3′ for *C. briggsae*). The reaction mixtures also contained 1 × PCR buffer, 0.25 mM of each of the four dNTPs, and *Taq* DNA polymerase. The 5′ oligo was labeled with (γ-^{32}P) ATP by T4 polynucleotide kinase. Reaction mixtures were incubated for 25 cycles at 94 °C for 1.0 min, 59 °C for 1.0 min, and 72 °C for 1.0 min. 2.0 μl of PCR product were digested with the restriction enzyme BamHI. There is a unique BamH1 site in *unc-52* exon 18, and this digestion step allows us to distinguish the different alternatively spliced isoforms more clearly on the gel. Digested PCR products were separated on 40-cm long, 0.4-mm thick, 5% or 6% polyacrylamide urea gels in TBE buffer. Gels were dried onto filter paper. These were then visualized using a Molecular Dynamics PhosphorImager (Sunnyvale, California, United States). Relative splice-site usage was quantified using ImageQuant software as previously described [37].

Supporting Information

Table S1. Complete Database of Highly Conserved Intronic Elements Used in Motif Analysis

Column (A) indicates *C. elegans* gene name. Columns (B–AI) list the WABA high homology sequences found in *C. elegans* introns flanking alternative exons of the corresponding gene.

Found at DOI: 10.1371/journal.pcbi.0020086.st001 (51 KB XLS).

Table S2. Complete Table Detailing Ranks of Pentamer and Hexamer Motifs in Introns Lacking Evolutionary Conservation Surrounding 307 Alternatively Spliced Exons

On both worksheets: Column (A) represents rank of motif from search based on obs/exp frequency in introns lacking evolutionary conservation downstream of alternative introns. Column (B) denotes motif and (C) contains the number of times the motif was observed in the dataset. (D) The observed frequency of the motif in the dataset. (E) Contains the expected frequency of motif based on total intron dataset. Column (F) is the ratio of observed frequency to expected frequency. Columns (H–M) are data from the same analysis in the same order for upstream introns.

Found at DOI: 10.1371/journal.pcbi.0020086.st002 (1.2 MB XLS).

Table S3. Each Row in the Table Represents a Different Spliced Isoform of the *unc-52* Reporter Constructs, and Each Column Denotes the Reporter Being Monitored

The first six columns show the relative percentage of alternative splice-site usage of the specified reporter in a *mec-8(+)* background as quantified from ^{32}P RT-PCR. The seventh column shows results of the wild-type sequence reporter that has been crossed into *mec-8(e398)* mutant animals.

Found at DOI: 10.1371/journal.pcbi.0020086.st003 (185 KB TIF).

References

1. International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature 431: 931–945.
2. Hide WA, Babenko VN, van Heusden PA, Seoighe C, Kelso JF (2001) The contribution of exon-skipping events on Chromosome 22 to protein coding diversity. Genome Res 11: 1848–1853.
3. Kan Z, Rouchka EC, Gish WR, States DJ (2001) Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. Genome Res 11: 889–900.
4. Modrek B, Resch A, Grasso C, Lee C (2001) Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Res 29: 2850–2859.
5. Croft L, Schandorff S, Clark F, Burrage K, Arctander P, et al. (2000) ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. Nat Genet 24: 340–341.
6. Singh R (2002) RNA-protein interactions that regulate pre-mRNA splicing. Gene Expr 10: 79–92.
7. Caceres JF, Kornblihtt AR (2002) Alternative splicing: Multiple control mechanisms and involvement in human disease. Trends Genet 18: 186–193.
8. Matlin AJ, Clark F, Smith CW (2005) Understanding alternative splicing: Towards a cellular code. Nat Rev Mol Cell Biol 6: 386–398.
9. Black DL (1992) Activation of c-src neuron specific splicing by an unusual RNA element in vivo and in vitro. Cell 69: 795–807.
10. Modafferi EF, Black DL (1997) A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon. Mol Cell Biol 17: 6537–6545.
11. Markovtsov V, Nikolic JM, Goldman JA, Turck CW, Chou MY, et al. (2000) Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. Mol Cell Biol 20: 7463–7479.
12. Underwood JG, Boutz PL, Dougherty JD, Stoilov P, Black DL (2005) Homologs of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. Mol Cell Biol 25: 10005–10016.
13. Jin Y, Suzuki H, Maegawa S, Endo H, Sugano S, et al. (2003) A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. Embo J 22: 905–912.
14. Huh GS, Hynes RO (1994) Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element. Gene Dev 8: 1561–1574.
15. Hedjran F, Yeakley JM, Huh GS, Hynes RO, Rosenfeld MG (1997) Control of alternative pre-mRNA splicing by distributed pentameric repeats. Proc Natl Acad Sci U S A 94: 12343–12347.
16. Ryan KJ, Cooper TA (1996) Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle. Mol Cell Biol 16: 4014–4023.
17. Charlet BN, Logan P, Singh G, Cooper TA (2002) Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. Molecular Cell 9: 649–658.
18. Jensen KB, Musunuru K, Lewis HA, Burley SK, Darnell RB (2000) The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. Proc Natl Acad Sci U S A 97: 5740–5745.
19. Jensen KB, Dredge BK, Stefani G, Zhong R, Buckanovich RJ, et al. (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. Neuron 25: 359–371.
20. Polydorides AD, Okano HJ, Yang YY, Stefani G, Darnell RB (2000) A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. Proc Natl Acad Sci U S A 97: 6350–6355.
21. Sorek R, Ast G (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res 13: 1631–1637.
22. Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB (2005) Identification and analysis of alternative splicing events conserved in human and mouse. Proc Natl Acad Sci U S A 102: 2850–2855.
23. Minovitsky S, Gee SL, Schokrpur S, Dubchak I, Conboy JG (2005) The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. Nucleic Acids Res 33: 714–724.
24. Kent WJ, Zahler AM (2000) The Intronerator: Exploring introns and alternative splicing in *Caenorhabditis elegans*. Nucleic Acids Res 28: 91–93.
25. Kent WJ, Zahler AM (2000) Conservation, regulation, synteny, and introns in a large-scale *C. briggsae/C. elegans* genomic alignment. Genome Res 10: 1115–1125.
26. Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, et al. (2003) The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. PLoS Biol 1: e45. DOI: 10.1371/journal.pbio.0000045
27. Fairbrother WG, Yeh RF, Sharp PA, Burge CB (2002) Predictive identification of exonic splicing enhancers in human genes. Science 297: 1007–1013.
28. Burd CG, Dreyfuss G (1994) RNA-binding specificity of hnRNP A1: Significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. EMBO J 13: 1197–1204.
29. Tacke R, Manley JL (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA-binding specificities. EMBO J 14: 3540–3551.
30. Caputi M, Zahler AM (2001) Determination of the RNA-binding specificity of the heterogenous nuclear ribonucleoprotein (hnRNP) H/H′/F/2H9 family. J Biol Chem 276: 43850–43859.
31. Moore MJ, Query CC, Sharp PA (1993) Splicing of precursors to mRNA by the spliceosome. In: Gesteland RF, Atkins JF, editors. The RNA world. New York: Cold Spring Harbor Press. pp. 303–357.
32. Ladd AN, Charlet N, Cooper TA (2001) The CELF family of RNA-binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol Cell Biol 21: 1285–1296.
33. Zahler AM (2005) Alternative splicing in *C. elegans*. The *C. elegans* research community. Wormbook. DOI: 10.1895/wormbook.1.31.1. Available: http://www.wormbook.org. Accessed 25 September 2005.
34. Zorio DA, Blumenthal T (1999) Both subunits of U2AF recognize the 3′ splice site in *C. elegans*. Nature 402: 835–838.
35. Pettitt J, Kingston IB (1994) Developmentally regulated alternative splicing of a nematode type IV collagen gene. Dev Biol 161: 22–29.
36. Sibley MH, Johnson JJ, Mello CC, Kramer JM (1993) Genetic identification, sequence, and alternative splicing of the *Caenorhabditis elegans* alpha 2 (IV) collagen gene. J Cell Biol 123: 255–264.
37. Farrer T, Roller AB, Kent WJ, Zahler AM (2002) Analysis of the role of *Caenorhabditis elegans* GC-AG introns in regulated splicing. Nucleic Acids Res 30: 3360–3367.

38. Blumenthal T, Steward K (1997) RNA processing and gene structure. In: Riddle DL, Blumenthal T, Meyer BJ, Priess JR, editors. C. elegans II. Plainview (New York): Cold Spring Harbor Laboratory Press. pp. 117–145.

39. Lim LP, Sharp PA (1998) Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. Mol Cell Biol 18: 3900–3906.

40. Nicoll M, Akerib CC, Meyer BJ (1997) X-chromosome-counting mechanisms that determine nematode sex. Nature 388: 200–204.

41. Buratti E, Brindisi A, Pagani F, Baralle FE (2004) Nuclear factor TDP-43 binds to the polymorphic TG repeats in CFTR intron 8 and causes skipping of exon 9: A functional link with disease penetrance. Am J Hum Genet 74: 1322–1325.

42. Faustino NA, Cooper TA (2005) Identification of putative new splicing targets for ETR-3 using sequences identified by systematic evolution of ligands by exponential enrichment. Mol Cell Biol 25: 879–887.

43. Chou MY, Underwood JG, Nikolic J, Luu MH, Black DL (2000) Multi-site RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. Mol Cell 5: 949–957.

44. Anyanful A, Ono K, Johnsen RC, Ly H, Jensen V, et al. (2004) The RNA-binding protein SUP-12 controls muscle-specific splicing of the ADF/cofilin pre-mRNA in C. elegans. J Cell Biol 167: 639–647.

45. Rogalski TM, Gilchrist EJ, Mullen GP, Moerman DG (1995) Mutations in the unc-52 gene responsible for body wall muscle defects in adult Caenorhabditis elegans are located in alternatively spliced exons. Genetics 139: 159–169.

46. Lundquist EA, Herman RK, Rogalski TM, Mullen GP, Moerman DG, et al. (1996) The mec-8 gene of C. elegans encodes a protein with two RNA recognition motifs and regulates alternative splicing of unc-52 transcripts. Development 122: 1601–1610.

47. Spike CA, Davies AG, Shaw JE, Herman RK (2002) MEC-8 regulates alternative splicing of unc-52 transcripts in C. elegans hypodermal cells. Development 129: 4999–5008.

48. Mullen GP, Rogalski TM, Bush JA, Gorji PR, Moerman DG (1999) Complex patterns of alternative splicing mediate the spatial and temporal distribution of perlecan/UNC-52 in Caenorhabditis elegans. Mol Biol Cell 10: 3205–3221.

49. Yeo G, Hoon S, Venkatesh B, Burge CB (2004) Variation in sequence and organization of splicing regulatory elements in vertebrate genes. Proc Natl Acad Sci U S A 101: 15700–15705.

50. Min H, Turck CW, Nikolic JM, Black DL (1997) A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. Genes Dev 11: 1023–1036.

51. Ratsch G, Sonnenburg S, Scholkopf B (2005) RASE: Recognition of alternatively spliced exons in C. elegans. Bioinformatics 21 Suppl 1: i369–377.

52. Sanford JR, Ellis J, Caceres JF (2005) Multiple roles of arginine/serine-rich splicing factors in RNA processing. Biochem Soc Trans 33: 443–446.

53. Siepel A, Haussler D (2004) Combining phylogenetic and hidden Markov models in biosequence analysis. J Comput Biol 11: 413–428.

54. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 16: 276–277.

55. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, et al. (2004) The UCSC Table Browser data retrieval tool. Nucl Acids Res 32: D493–D496.

56. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucl Acids Res 25: 3389–3402.

57. Kunkel TA, Roberts JD, Zakour RA (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol 154: 367–382.

58. Mello C, Fire A (1995) DNA transformation. Methods Cell Biol 48: 451–482.

59. Lewis JA, Fleming JT (1995) Basic culture methods. Methods Cell Biol 48: 3–29.

60. Roller AB, Hoffman DC, Zahler AM (2000) The allele-specific suppressor sup-39 alters use of cryptic splice sites in C. elegans. Genetics 154: 1169–1179.