Worm 3, e28459; March 2014; © 2014 Landes Bioscience

Comprehensive analysis of mutually exclusive alternative splicing in *C. elegans*

Hidehito Kuroyanagi[1,*], Satomi Takei[1], and Yutaka Suzuki[2]

[1]Laboratory of Gene Expression; Medical Research Institute; Tokyo Medical and Dental University; Bunkyo-ku, Tokyo, Japan;
[2]Department of Medical Genome Science; University of Tokyo; Kashiwa, Chiba, Japan

Keywords: alternative splicing, intron, mutually exclusive exons, NMD; pre-mRNA processing, proteome diversity

Abbreviations: GO, gene ontology; ME, mutually exclusive; NMD, nonsense-mediated mRNA decay; nt, nucleotide; pre-mRNA, precursor messenger RNA; PTC, premature termination codon; RT-PCR, reverse transcription-polymerase chain reaction

Mutually exclusive selection of one exon in a cluster of exons is a rare form of alternative pre-mRNA splicing, yet suggests strict regulation. However, the repertoires of regulation mechanisms for the mutually exclusive (ME) splicing in vivo are still unknown. Here, we experimentally explore putative ME exons in *C. elegans* to demonstrate that 29 ME exon clusters in 27 genes are actually selected in a mutually exclusive manner. Twenty-two of the clusters consist of homologous ME exons. Five clusters have too short intervening introns to be excised between the ME exons. Fidelity of ME splicing relies at least in part on nonsense-mediated mRNA decay for 14 clusters. These results thus characterize all the repertoires of ME splicing in this organism.

Introduction

Alternative processing of precursor mRNAs (pre-mRNAs) is a major source of protein diversity and plays crucial roles in development, differentiation, and diseases in higher eukaryotes.[1-4] One of the most elaborately regulated forms of alternative pre-mRNA processing is mutually exclusive (ME) alternative splicing, through which only one exon is mutually exclusively selected from a cluster of exons at a time to determine critical aspects of the target genes such as ligand-binding specificity of receptors and properties of enzymes and channels.[5-7] The ME exons occur only as pairs in vertebrates, but the number of ME exons in a cluster in invertebrates can be more than two in some genes. The extreme example is the *Dscam* gene of a fruit fly *Drosophila melanogaster*, which has four ME exon clusters containing 12, 48, 33, and two variants.[8]

Several mechanisms have been proposed for the mutually exclusive nature of the ME exon selection.[9-12] Steric interference between two splice sites of an intervening intron between two ME exons has been proposed to make the ME exons physically incapable of being spliced to each other in some mammalian genes.[9,12,13] This is due to a shorter distance between the 5' splice site and the branch point on the intron than minimal spacing required for a spliceosome to be productively assembled. Another mechanism proposed to prohibit double inclusion and double skipping of the ME exons is spliceosome incompatibility.[9,12] This is proposed for a tandem exon pair flanked by U2- and U12-type introns on either side of the exons.[14] Disposal of aberrantly spliced mRNAs by a surveillance system termed nonsense-mediated mRNA decay (NMD) is considered to play substantial roles for some genes.[15-17] Antagonism of repression by base-paring interaction of a docker with one of the selector sequences is proposed for the *Dscam* exon 6 cluster.[18,19]

In *C. elegans*, it is estimated by a recent genome-wide analysis that up to 25% of the protein-coding genes undergo alternative pre-mRNA processing and 55 events were assigned to ME alternative splicing.[20] In previous studies, we have elucidated tissue-specific and/or developmental selection patterns and regulation mechanisms for some of the ME exon clusters in *C. elegans* by generating fluorescence alternative splicing reporters and isolating splicing factor mutants. In the case of exons 5B/5A of the *egl-15* gene, encoding fibroblast growth factor receptors (FGFRs),[21-23] the RBFOX family and SUP-12 cooperatively bind to the upstream flanking intron of the upstream exon to repress the upstream exon in muscles. In the case of exons 9/10 of the *let-2* gene, encoding α2 subunit of collagen type IV,[24] ASD-2 binds to the downstream flanking intron of the downstream exon to promote inclusion of the downstream exon in muscle-specific and developmentally regulated manners.[25,26] In the case of exons 7a/7b of the *unc-32* gene, encoding subunit *a* of V_0 domain of vacuolar proton-translocating ATPase (V-ATPase),[27] UNC-75 binds to the intervening intron to repress the downstream exon and the RBFOX family binds to the downstream intron to promote inclusion of the upstream exon in the nervous system.[28] Thus,

*Correspondence to: Hidehito Kuroyanagi; Email: kuroyana.end@tmd.ac.jp
Submitted: 12/20/2013; Revised: 02/28/2014; Accepted: 03/06/2014
http://dx.doi.org/10.4161/worm.28459

Table 1. Summary of experimental validation of the putative ME exon clusters

		Name and ID in WS239				Position in WS235				Supporting Reads[*a]	Sequence Identity		Intervening Intron(s) [nt][*b]	NMD-Dependence[*c]
AS Type	Chr	Gene	WBGene ID	Str	Exon	Left end	Right end	Exon Length			Nucleotide	Amino Acid		
Homologous ME	I	*F30F8.9*	WBGene00009276	+	4a	7837844	7837926	83	3n +2	9	41.0%	17.9%	683	Yes
					4b	7838610	7838692	83	3n +2	12				
		lev-11	WBGene00002978	-	7b	14623010	14623148	139	3n +1	1615	71.2%	78.3%	422	No
					7a[*d]	14623571	14623709	139	3n +1	83				
					5b	14624030	14624147	118	3n +1	478	81.4%	74.4%	155, 123	No
					5c	14624303	14624420	118	3n +1	94	47.5%	35.9%		
					5a	14624544	14624661	118	3n +1	706	83.9%	71.8%		
	II	*snt-1*	WBGene00004921	+	6B[*d]	6815048	6815187	140	3n +2	0	64.3%	55.3%	116	No
					6A	6815304	6815437	134	3n +2	63				
		lat-1	WBGene00002251	+	3a	8903388	8903494	107	3n +2	24	55.8%	45.9%	112	Partially
					3b	8903607	8903719	113	3n +2	72				
	III	*let-805*	WBGene00002915	+	19a	2676547	2676664	118	3n +1	44	40.4%	33.3%	97	No
					19b	2676762	2676870	109	3n +1	34				
		gly-6	WBGene00001631	-	8b	4325419	4325523	105	3n	26	44.1%	28.9%	11	-
					8a	4325535	4325642	108	3n	49				
		pdfr-1	WBGene00015735	-	8b	6636402	6636556	155	3n +2	77	53.4%	30.8%	336	Partially
					8a	6636893	6637017	125	3n +2	24				
		coq-2	WBGene00000762	+	6a	6937600	6937733	134	3n +2	13	47.0%	40.0%	678	Yes
					6b	6938412	6938545	134[*e]	3n +2	12				
		unc-32	WBGene00006768	+	7a	8909233	8909355	123	3n	46	47.2%	35.4%	725	-
					7b	8910081	8910221	141	3n	56				
		gly-5	WBGene00001630	+	9b	13200771	13200881	111	3n	16	77.7%	54.1%	34, 1466	-
					9c	13200916	13201020	105	3n	13	66.1%	54.1%		
					9a	13202487	13202588	102	3n	6	71.4%	48.6%		
		W06F12.2	WBGene00012306	-	2b	13730072	13730243	172	3n +1	80	35.5%	17.5%	129	Yes
					2a	13730373	13730526	154	3n +1	28				
	IV	*gba-3*	WBGene00008706	+	6a	17478400	17478647	248	3n +2	78	75.4%	80.5%	207	Yes
					6b	17478855	17479102	248	3n +2	82				

ME, mutually exclusive exons; CE, cassette exon. *[a]Total number of sequence reads mapped to each ME exon and its junctions with the flanking exons out of 10.3 million mapped reads in RNA-seq analysis of polyA+ RNA from synchronized smg-2 L1 larvae.[43] *[b]Intervening introns shorter than 40 nt are underlined. *[c]Yes, NMD isoform(s) were significantly increased in the smg-2 mutant ($P < 0.05$ in a modified χ^2 test of the RT-PCR products[43]). Partially, NMD isoform(s) were slightly increased in the smg-2 mutant (difference in the amount of the splice variant > 0.4 [% of sum] in molar concentration). No, no apparent difference in the RT-PCR patterns between N2 and the smg-2 mutant. *[d]These exons were expressed at very low levels at the L1 stage. At the young adult stage, lev-11 exon 7a and snt-1 exon 6B were still rare, while unc-62 exon 7a was readily detected. *[e]Out-of-frame tandem acceptor sites were frequently used for coq-2 exon 6b. *[f]mrp-1 exon 13D is aberrantly annotated in the RefSeq model used for mapping. *[g]These exons carry in-frame termination codons. *[h]Nucleotide and amino acid sequence identities of del-6 exons 5/6 appear high because of their short lengths. *[i]In-frame tandem acceptor sites were frequently used for unc-43 exon 14. *[j]Tandem donor sites were frequently used for K10D6.2 exon 5.

Table 1. Summary of experimental validation of the putative ME exon clusters (continued)

		Name and ID in WS239				Position in WS235				Supporting Reads*[a]	Sequence Identity		Intervening Intron(s) [nt]*[b]	NMD-Dependence*[c]
AS Type	Chr	Gene	WBGene ID	Str	Exon	Left end	Right end	Exon Length			Nucleotide	Amino Acid		
Homologous ME	V	*unc-62*	WBGene00006796	+	7a*[d]	4504975	4505120	146	3n +2	4	49.1%	30.2%	238	No
					7b	4505359	4505516	158	3n +2	162				
		gck-1	WBGene00001526	+	4a	8952725	8952879	155	3n +2	19	52.2%	48.1%	138	Partially
					4b	8953018	8953175	158	3n +2	31				
		akt-1	WBGene00000102	+	6a	10250841	10251032	192	3n	136	62.8%	63.8%	294	-
					6b	10251327	10251533	207	3n	41				
		atn-1	WBGene00000228	+	4a	12480728	12480888	161	3n +2	147	61.4%	57.1%	225	Partially
					4b	12481114	12481196	83	3n +2	26				
		slo-1	WBGene00004830	+	10a	18499582	18499694	113	3n +2	21	65.5%	64.1%	856	Partially
					10b	18500551	18500663	113	3n +2	11				
	X	*mrp-1*	WBGene00003407	-	13D	579498	579653	156	3n	31*[f]	68.6%	53.8%	334, 154, 435	-
					13C	579988	580143	156	3n	35	76.9%	59.6%		
					13B	580298	580453	156	3n	37	74.4%	53.8%		
					13A	580889	581044	156	3n	2	69.2%	50.0%		
		cca-1	WBGene00000367	-	18b	7854051	7854201	151	3n +1	8	53.0%	44.0%	97	Partially
					18a	7854299	7854428	130	3n +1	20				
		bet-2	WBGene00010199	+	3a	10571708	10571827	120	3n	35	45.8%	35.0%	37	-
					3b	10571865	10571984	120	3n	32				
		let-2	WBGene00002280	-	10	16386613	16386723	111	3n	164	65.8%	56.8%	30	-
					9	16386754	16386861	108	3n	104				
Non-Homologous ME	I	*pqn-72*	WBGene00004154	-	6	3274267	3274812	546	3n	53	37.0%	21.6%	117	-
					5	3274930	3275010	81	3n	3				
		tom-1	WBGene00006594	+	14a	5552880	5553369	490	3n +1	172	50.0%	16.4%	1371	Partially
					14b	5554741	5554864	124	3n +1	72				
		ttx-7	WBGene00008765	-	5b	7301663	7301774	112	3n +1	16	41.8%	17.1%	132	Partially
					5a	7301907	7302027	121	3n +1	54				
	III	*unc-32*	WBGene00006768	+	4a	8906524	8906678	155	3n +2	34	60.0%	17.3%	288, 237	Partially
					4b	8906967	8907073	107	3n +2	49	51.6%	18.0%		
					4c	8907311	8907432	122	3n +2	26	49.7%	23.1%		

ME, mutually exclusive exons; CE, cassette exon. *[a]Total number of sequence reads mapped to each ME exon and its junctions with the flanking exons out of 10.3 million mapped reads in RNA-seq analysis of polyA+ RNA from synchronized smg-2 L1 larvae.[43] *[b]Intervening introns shorter than 40 nt are underlined. *[c]Yes, NMD isoform(s) were significantly increased in the smg-2 mutant ($P < 0.05$ in a modified χ^2 test of the RT-PCR products[43]). Partially, NMD isoform(s) were slightly increased in the smg-2 mutant (difference in the amount of the splice variant > 0.4 [% of sum] in molar concentration). No, no apparent difference in the RT-PCR patterns between N2 and the smg-2 mutant. *[d]These exons were expressed at very low levels at the L1 stage. At the young adult stage, lev-11 exon 7a and snt-1 exon 6B were still rare, while unc-62 exon 7a was readily detected. *[e]Out-of-frame tandem acceptor sites were frequently used for coq-2 exon 6b. *[f]mrp-1 exon 13D is aberrantly annotated in the RefSeq model used for mapping. *[g]These exons carry in-frame termination codons. *[h]Nucleotide and amino acid sequence identities of del-6 exons 5/6 appear high because of their short lengths. *[i]In-frame tandem acceptor sites were frequently used for unc-43 exon 14. *[j]Tandem donor sites were frequently used for K10D6.2 exon 5.

Table 1. Summary of experimental validation of the putative ME exon clusters (continued)

		Name and ID in WS239				Position in WS235				Supporting Reads*[a]	Sequence Identity		Intervening Intron(s) [nt]*[b]	NMD-Dependence*[c]
AS Type	Chr	Gene	WBGene ID	Str	Exon	Left end	Right end	Exon Length			Nucleotide	Amino Acid		
Non–Homologous ME	IV	*fbl-1*	WBGene00001403	+	5D	9542035	9542292	258	3n	57	35.0%	22.1%	435	-
					5C	9542728	9542868	141	3n	10				
	V	*del-6*	WBGene00011891	+	5	10574706	10574733	28	3n +1*[g]	4	53%*[h]	60%*[h]	200	Yes
					6	10574934	10575098	165	3n	230				
	X	*egl-15*	WBGene00001184	+	5B	11017580	11017930	351	3n	38	40.1%	10.3%	14	-
					5A	11017945	11018139	195	3n	5				
Tandem CE	I	*tom-1*	WBGene00006594	+	17	5557960	5557971	12	3n	-	-	-	-	-
Tandem CE					18	5558453	5558479	27	3n	-	-	-	-	-
Tandem CE	I	*lev-11*	WBGene00002978	-	9b	14622304	14622511	208	3n +1*[g]	-	-	-	-	-
Tandem CE					9a	14622658	14622743	86	3n +2*[g]	-	-	-	-	-
Single CE	II	*etr-1*	WBGene00001340	+	4	166071	166150	80	3n +2	-	-	-	-	-
Tandem CE	II	*zyg-12*	WBGene00006997	+	8	4952490	4952531	42	3n*[g]	-	-	-	-	-
Tandem CE					9	4952647	4952694	48	3n	-	-	-	-	-
Tandem CE	II	*C34F11.3*	WBGene00016415	+	10	5205040	5205075	36	3n	-	-	-	-	-
Tandem CE					11	5205444	5205632	189	3n	-	-	-	-	-
Tandem CE	III	*clp-1*	WBGene00000542	+	3	7981673	7981810	138	3n	-	-	-	-	-
Tandem CE					4	7982464	7982538	75	3n	-	-	-	-	-
Constitutive	IV	*unc-44*	WBGene00006780	+	13	5984796	5984980	185	3n +2	-	-	-	-	-
Single CE					14	5985093	5985170	78	3n	-	-	-	-	-
Undetected					15	5985577	5986355	779	3n +2	-	-	-	-	-
Alternative Acceptors	IV	*unc-43*	WBGene00006779	-	14S	10329097	10329195	99*[i]	3n	-	-	-	-	-
Tandem CE	V	*K10D6.2*	WBGene00010742	-	5L	11194203	11194342	140*[j]	3n +2*[g]	-	-	-	-	-
Tandem CE					4	11194574	11194627	54	3n	-	-	-	-	-
Tandem CE	V	*Y69H2.3*	WBGene00013481	-	7	18661622	18661711	90	3n	-	-	-	-	-
Tandem CE					6	18662119	18662313	195	3n	-	-	-	-	-
Constitutive	X	*F49E2.5*	WBGene00009888	+	5	9554299	9554547	249	3n	-	-	-	-	-
Tandem CE					6	9554692	9554823	132	3n	-	-	-	-	-
Constitutive	X	*T23E7.2*	WBGene00020732	+	16	17679922	17680035	114	3n	-	-	-	-	-

ME, mutually exclusive exons; CE, cassette exon. *[a]Total number of sequence reads mapped to each ME exon and its junctions with the flanking exons out of 10.3 million mapped reads in RNA-seq analysis of polyA+ RNA from synchronized smg-2 L1 larvae.[43] *[b]Intervening introns shorter than 40 nt are underlined. *[c]Yes, NMD isoform(s) were significantly increased in the smg-2 mutant ($P < 0.05$ in a modified χ^2 test of the RT-PCR products[43]). Partially, NMD isoform(s) were slightly increased in the smg-2 mutant (difference in the amount of the splice variant > 0.4 [% of sum] in molar concentration). No, no apparent difference in the RT-PCR patterns between N2 and the smg-2 mutant. *[d]These exons were expressed at very low levels at the L1 stage. At the young adult stage, lev-11 exon 7a and snt-1 exon 6B were still rare, while unc-62 exon 7a was readily detected. *[e]Out-of-frame tandem acceptor sites were frequently used for coq-2 exon 6b. *[f]mrp-1 exon 13D is aberrantly annotated in the RefSeq model used for mapping. *[g]These exons carry in-frame termination codons. *[h]Nucleotide and amino acid sequence identities of del-6 exons 5/6 appear high because of their short lengths. *[i]In-frame tandem acceptor sites were frequently used for unc-43 exon 14. *[j]Tandem donor sites were frequently used for K10D6.2 exon 5.

the tissue-specificity, *trans*-acting factors, positions of the *cis*-elements, and functions of the factors for the regulation of ME exon clusters, vary from gene to gene in *C. elegans*. These findings raise questions about to what extent the repertoires and regulation mechanisms for the ME exon clusters have evolved in this organism.

Here we explore all the 55 putative ME splicing events in *C. elegans* listed in Ramani et al.[20] that utilized high-throughput sequencing and microarray profiling of polyA+ RNAs isolated from four and five different developmental stages, respectively. We experimentally test whether the putative ME exons are actually mutually exclusively selected by reverse transcription-polymerase chain reaction (RT-PCR). For the verified ME exons, we analyze the nucleotide and amino acid sequence identities of the ME exons in each cluster. To also elucidate to what extent the mutually exclusive nature of the exon selection rely on NMD, we compare the RT-PCR patterns between a wild-type strain N2 and an NMD-deficient mutant *smg-2*.[23,29]

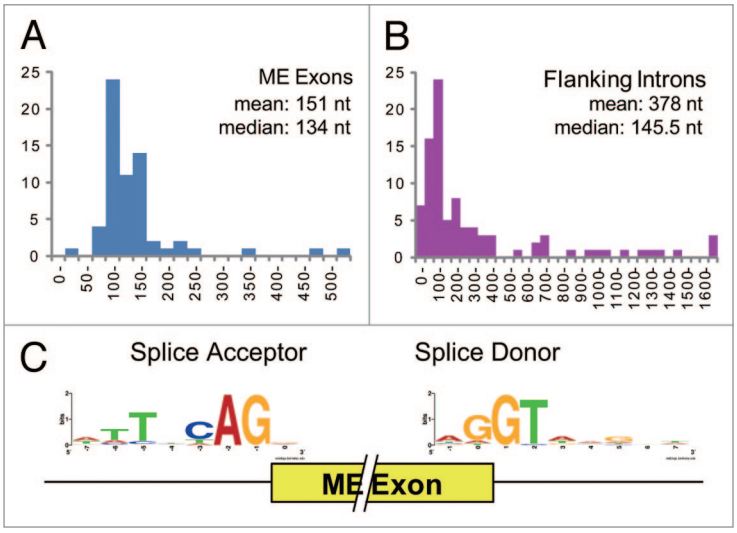

Figure 2. Statistics of the ME exons and their flanking introns. (**A and B**) Size distributions of the 63 verified ME exons (**A**) and their 92 flanking introns, including the intervening introns (**B**). The mean and median sizes are also indicated. (**C**) Sequence logos of the splice acceptor and donors sites of the 63 ME exons.

Results and Discussion

Table 1 summarizes the results of the comprehensive RT-PCR analyses at the L1 stage. The 55 events were assigned to 41 clusters in 37 genes. Eight of the clusters were considered to be tandem cassette exon pairs rather than ME exons because we detected in-frame double-inclusion and/or double-skipping isoforms (> 5% of the sum in molar concentration) in addition to the single-inclusion isoforms (**Table 1**; **Fig. 1A**; and data not shown). Notably, these cassette exons are multiple of three (3n) nucleotides (nt) in length except for those carrying natural termination codons (**Table 1**). Two exons in two genes were considered to be single cassette exons and two other exons in two genes appear to be constitutively included. The other 29 clusters in 27 genes were considered to be mutually exclusive (**Table 1**) since the single-inclusion isoforms were detected in our experiments and/or in the literature and other isoforms were almost undetectable or degraded by NMD in the wild-type background (see below). We confirmed that the single-inclusion isoforms were also almost exclusively expressed at the young adult stage (data not shown).

Features of the ME exon clusters

The 29 ME exon clusters can be divided into two groups according to sequence similarity of the ME exons. Homologous ME exon clusters include 19 pairs, two trios, and one quad of homologous ME exons, while non-homologous clusters include six pairs and one trio of non-homologous ME exons (**Table 1**). The homologous clusters may be originated from exon duplication.[14,30] The lengths of the homologous ME exons are close to or exactly the same as the counterpart(s) except for *atn-1* exons 4a/4b, while those of the non-homologous ME exons are often far different from the counterpart(s). Nevertheless, reading frames in the downstream common exons are preserved whichever exon in the clusters is selected in almost all cases. The only exception is exons 5/6 of the *del-6* gene, encoding a degenerin-like ion channel protein, where exon 5 consists of 3n+1 nt and carries a natural termination codon while exon 6 consists of 3n nt and has no termination codon (**Fig. 1B**).

Four of the ME exon clusters consist of more than two ME exons. Exons 13A/13B/13C/13D of the *mrp-1* gene, encoding an ATP-binding cassette (ABC) transporter,[31] is the only cluster with four ME exons. All the four exons are 156 (3n) nt in length and are homologous to each other (**Table 1**). Exons 9b/9c/9a of the *gly-5* gene, encoding a UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase,[32] is one of the three clusters with three ME exons. These exons are homologous to each other with almost the same size of 3n nt. The intervening intron between exons 9b and 9c is just 34 nt (discussed later). Exons 4a/4b/4c of the *unc-32* gene[27] are the only non-homologous trio of the ME exons. The *unc-32* gene has another cluster of ME exons 7a/7b (**Table 1**) and we have recently reported that these two clusters in the single gene are independently regulated in tissue-specific manners.[28] Exons 5a/5c/5b of the *lev-11* gene, encoding multiple tropomyosin isoforms,[33] are homologous and of exactly the same size (3n+1 nt). Pre-mRNA processing of the *lev-11* gene is complex due to the combination of tissue-specific promoters, clusters of ME exons 4a/4b, 5a/5b/5c, and 7a/7b, and tandem cassette exons 9a/9b (**Table 1**).[33] The complex structures and pre-mRNA processing patterns of the tropomyosin genes are evolutionarily conserved in metazoans,[34] suggesting functional significance of multiple tropomyosin isoforms.

Steric interference to prohibit double inclusion of ME exons

Size distribution of the overall introns suggests that the minimal size of the introns is ~40 nt in *C. elegans*.[35] Among the ME exon clusters we have already reported, the intervening introns for exons 5B/5A of the *egl-15* gene[21-23] and exons 9/10 of the *let-2* gene[24,25] are 14 nt and 30 nt, respectively, and we

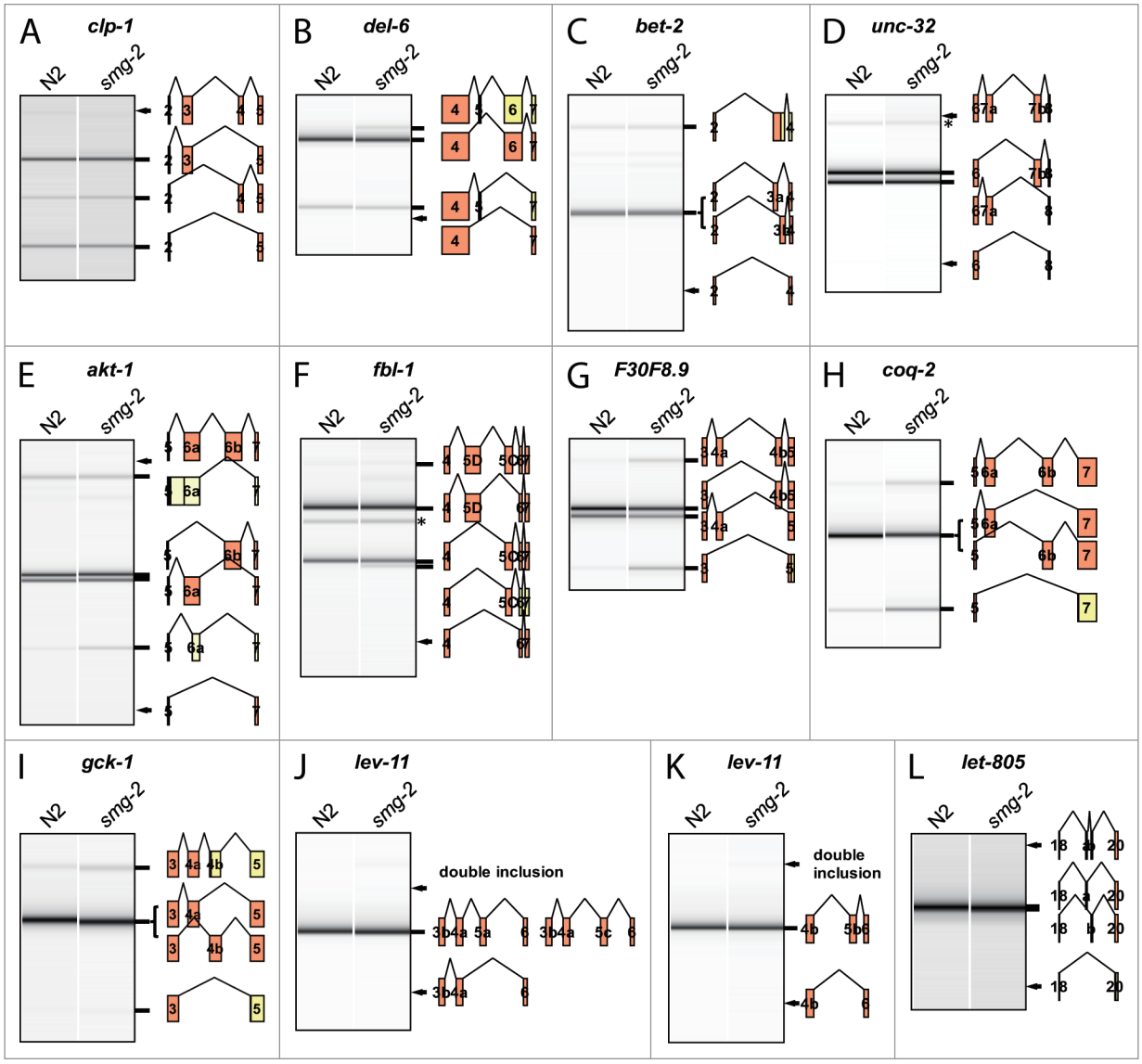

Figure 1. RT-PCR analyses of the putative ME exons in the wild-type (N2) and the *smg-2 (yb979)* mutant. (**A**) *clp-1* exons 3/4. (**B**) *del-6* exons 5/6. *del-6* exon 5 is unique in that it carries a natural termination codon. (**C**) *bet-2* exons 3a/3b. Note that the intervening intron is retained instead of double ME exon inclusion for this cluster. (**D**) *unc-32* exons 7a/7b.[28] (**E**) *akt-1* exons 6a/6b. A non-productive exon 6a isoform utilizing an aberrant acceptor site is detected in the *smg-2* mutant. (**F**) *fbl-1* exons 5D/5C.[50] (**G**) *F30F8.9* exons 4a/4b. (**H**) *coq-2* exons 6a/6b. (**I**) *gck-1* exons 4a/4b. (**J and K**) *lev-11* exons 5a/5c/5b.[37] (**L**) *let-805* exons 19a/19b. Splicing patterns are schematically indicated. Coding regions are in orange. Arrows indicate predicted positions of undetected isoforms indicated on the right. Asterisks indicate non-specific bands.

have never observed mRNA isoforms where these short introns are excised. These observations are consistent with the idea that the short introns of less than 40 nt cannot be excised because of the steric interference like in mammals, although no strong consensus are found for the branch point in *C. elegans*. According to this criterion, three more clusters are considered to be physically incapable of double exon inclusion: exons 3a/3b of *bet-2* encoding a BET (two bromodomains) family protein (**Fig. 1C**), *gly-5* exons 9b/9c, and exons 8a/8b of *gly-6*, a paralog of *gly-5* (**Table 1**).[32] Notably, the lengths of all the ME exons in these five clusters are 3n nt.

As U11 or U12 snRNA or an AT/AC splice-junction are not found in *C. elegans*,[36] we need not consider the spliceosome incompatibility in the regulation of the ME exons here.

NMD-dependence of the mutually exclusive selection

If the lengths of the ME exons are 3n nt, inclusion or skipping of the ME exons does not cause a frame-shift or a premature termination codon (PTC) in the mRNA isoforms. Consistent with this idea, there was no apparent difference in the amounts of multiple-inclusion and all-skipping isoforms between the wild-type and the *smg-2* mutant for such clusters (**Table 1**; **Fig. 1D–F**; and data not shown).

Table 2. Gene ontology analysis of 25 genes with the ME exons and GO terms

Ontology type	GO term ID	Fold enrichment	Count in 25 Genes with ME Exons and GO terms	Count in all genes with GO Terms (12,834)	P value (Fisher's Exact Test)	Term
Biological_process	GO:0046928	257	2	4	2.18E-05	regulation of neurotransmitter secretion
	GO:0030163	16	4	131	1.11E-04	protein catabolic process
	GO:0007166	79	2	13	2.80E-04	cell surface receptor linked signal transduction
	GO:0040011	3.5	9	1327	5.77E-04	locomotion
	GO:0034765	49	2	21	7.48E-04	regulation of ion transmembrane transport
	GO:0043050	45	2	23	8.99E-04	pharyngeal pumping
Cellular_component	GO:0005865	114	2	9	1.30E-04	striated muscle thin filament
	GO:0016021	2.9	12	2143	2.77E-04	integral to membrane
	GO:0016020	3.1	11	1847	3.37E-04	membrane
	GO:0005604	57	2	18	5.47E-04	basement membrane
	GO:0005578	47	2	22	8.22E-04	proteinaceous extracellular matrix
Molecular_function	GO:0005201	257	2	4	2.18E-05	extracellular matrix structural constituent
	GO:0005244	49	2	21	7.48E-04	voltage-gated ion channel activity

If the lengths of the ME exons are not 3n nt, multiple inclusion and all-skipping of the ME exons cause frame-shifts to create aberrant termination codons and such mRNA isoforms should be eliminated by NMD. We found that multiple-inclusion and/or all-skipping isoforms are evidently or slightly more abundant in the *smg-2* mutant than in the wild-type for 14 out of the 19 clusters where the ME exons are not 3n nt (**Table 1**; **Fig. 1G–I**; and data not shown), indicating that these non-productive mRNA isoforms are actually eliminated by NMD in the wild-type. For *lev-11* exons 7a/7b,[37] *snt-1* exons 6B/6A,[38] and *unc-62* exons 7a/7b,[39] we confirmed predominant use of only one of the two ME exons (**Table 1**) as in the literature and this can be a reason why aberrantly spliced isoforms are rare and undetectable for these clusters. For the other two clusters, *lev-11* exons 5a/5c/5b[37] and *let-805* exons 19a/19b, the RT-PCR patterns were indistinguishable between the wild-type and the *smg-2* mutant (**Table 1**; **Fig. 1J–L**). All these results indicate that the fidelity of the splicing regulation varies among the ME exon clusters and some of them rely on the mRNA surveillance system.

Statistics of the ME exons and flanking introns

Figure 2 summarizes the statistics of the 63 experimentally verified ME exons and their flanking and intervening introns.

The median size of the ME exons (134 nt) (**Fig. 2A**) is similar to those of the entire unique exons in confirmed genes (144 nt).[40] Most of the ME exons (60 of 63) are shorter than 260 nt and the average size (151 nt) is close to the median (**Fig. 2A**). In contrast, the size distribution of the entire unique exons has a fatter tail,[35] making the average of 201 nt.[40] The three ME exons longer than 350 nt belong to distinct non-homologous ME exon clusters (**Table 1**). The shortest ME exon (28 nt) exceptionally carries a natural termination codon (**Table 1**; **Fig. 1B**). Therefore, the size of the 48 homologous ME exons are in a relatively narrow range (83–248 nt) for *C. elegans*.

The mean size of the introns flanking the ME exons, including the five short intervening introns discussed above, is 378 nt (**Fig. 2B**), substantially longer than the overall average of the introns (267 nt).[35] The median size of the introns flanking the ME exons is 145.5 nt (**Fig. 2B**), whereas more than half of all the *C. elegans* introns are 100 nt or less and most of them are near the minimal length,[35] indicating that the introns flanking the ME exons tend to be longer than constitutive introns. This is consistent with a previous finding that many of *cis*-elements regulating alternative splicing in *C. elegans* are found in introns.[41] Eleven out of the 29 ME exon clusters have UGCAUG stretch(es) in the flanking introns and/or in the ME exons (data not shown), suggesting tissue-specific splicing regulation by the RBFOX family splicing factors ASD-1 and FOX-1.[42] Six out of the 29 clusters are affected in the *unc-75* mutant,[43] suggesting neuron-specific splicing regulation.

Figure 2C summarizes the sequences of the splice acceptor and donor sites for the verified ME exons. These are more diversified from the consensus sequences of the acceptor site (TTTTCAG/R)[44] and the donor site (AG/GTAAGTT)[45] in *C. elegans*, where R stands for A or G. Furthermore, two (2.2%) of the 92 flanking introns, *let-2* intron 10 and *del-6* intron 6, start with GC, a weaker donor than GT,[46,47] although GC-AG introns are rare (0.373%) in *C. elegans* like in other eukaryotes.[45] Therefore, the splice sites of the ME exons are considered to be weaker than those of constitutive exons, consistent with previous findings on alternative splice sites in higher organisms.[48]

Table 2 summarizes gene ontology (GO) analysis of 25 genes with GO terms out of the 27 genes with the verified ME exon clusters. It indicates enrichment of genes encoding membrane or extracellular matrix proteins ($P < 0.001$, Fisher's exact test).

Conclusion

We demonstrated that the 29 ME exon clusters in the 27 genes are actually regulated in a mutually exclusive manner in

C. elegans. Twenty-two of the 29 clusters consist of two to four homologous ME exons. Ten of the 29 clusters consist of ME exons with the lengths of 3n nt, five of which have too short intervening introns to be excised. Fourteen of the 19 clusters with the ME exons other than 3n nt in length rely at least in part on NMD. Nevertheless, many of the ME exon clusters appear to be strictly regulated. Further molecular and functional analyses of such clusters will elucidate novel mechanisms for mutually exclusive selection of the ME exons in vivo.

Materials and Methods

Total RNAs were extracted from synchronized L1 larvae of N2 and KH1668: *smg-2 (yb979) I* strains as described previously.[26] RT-PCR was performed essentially as described previously.[26] RT-PCR products were analyzed by using BioAnalyzer (Agilent) as described previously.[43] The sequences of the RT-PCR products were confirmed by direct sequencing or by cloning and sequencing. Sequences of the primers used in the RT-PCR assays are available upon request to Kuroyanagi H.

A list of the GO terms was retrieved from the Gene Ontology website (http://www.geneontology.org/). Fisher's exact test was performed by using Ekuseru-Toukei 2010 (Social Survey Research Information). Sequence logos were generated by using WebLogo3[49] at http://weblogo.threeplusone.com/create.cgi.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Arun K Ramani and Andrew G Fraser of University of Toronto for detailed information about alternative splicing events in *C. elegans*. We thank Hiroaki Iwasa of Tokyo Medical and Dental University (TMDU) for fruitful discussion. We thank *Caenorhabditis* Genetics Center for N2 and a bacterial strain OP50.

References

1. Li Q, Lee JA, Black DL. Neuronal regulation of alternative pre-mRNA splicing. Nat Rev Neurosci 2007; 8:819-31; PMID:17895907; http://dx.doi.org/10.1038/nrn2237
2. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet 2007; 8:749-61; PMID:17726481; http://dx.doi.org/10.1038/nrg2164
3. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 2009; 10:741-54; PMID:19773805
4. Kalsotra A, Cooper TA. Functional consequences of developmentally regulated alternative splicing. Nat Rev Genet 2011; 12:715-29; PMID:21921927; http://dx.doi.org/10.1038/nrg3052
5. Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol 2005; 6:386-98; PMID:15956978; http://dx.doi.org/10.1038/nrm1645
6. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 2010; 11:345-55; PMID:20376054; http://dx.doi.org/10.1038/nrg2776
7. David CJ, Chen M, Assanah M, Canoll P, Manley JL. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. Nature 2010; 463:364-8; PMID:20010808; http://dx.doi.org/10.1038/nature08697
8. Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 2000; 101:671-84; PMID:10892653; http://dx.doi.org/10.1016/S0092-8674(00)80878-8
9. Smith CW. Alternative splicing--when two's a crowd. Cell 2005; 123:1-3; PMID:16213205; http://dx.doi.org/10.1016/j.cell.2005.09.010
10. Takeuchi A, Hosokawa M, Nojima T, Hagiwara M. Splicing reporter mice revealed the evolutionally conserved switching mechanism of tissue-specific alternative exon selection. PLoS One 2010; 5:e10946; PMID:20532173; http://dx.doi.org/10.1371/journal.pone.0010946
11. Yang Y, Zhan L, Zhang W, Sun F, Wang W, Tian N, Bi J, Wang H, Shi D, Jiang Y, et al. RNA secondary structure in mutually exclusive splicing. Nat Struct Mol Biol 2011; 18:159-68; PMID:21217700; http://dx.doi.org/10.1038/nsmb.1959
12. Pohl M, Bortfeldt RH, Grützmann K, Schuster S. Alternative splicing of mutually exclusive exons--a review. Biosystems 2013; 114:31-8; PMID:23850531; http://dx.doi.org/10.1016/j.biosystems.2013.07.003
13. Smith CW, Nadal-Ginard B. Mutually exclusive splicing of alpha-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. Cell 1989; 56:749-58; PMID:2924347; http://dx.doi.org/10.1016/0092-8674(89)90678-8
14. Letunic I, Copley RR, Bork P. Common exon duplication in animals and its role in alternative splicing. Hum Mol Genet 2002; 11:1561-7; PMID:12045209; http://dx.doi.org/10.1093/hmg/11.13.1561
15. Jones RB, Wang F, Luo Y, Yu C, Jin C, Suzuki T, Kan M, McKeehan WL. The nonsense-mediated decay pathway and mutually exclusive expression of alternatively spliced FGFR2IIIb and -IIIc mRNAs. J Biol Chem 2001; 276:4158-67; PMID:11042206; http://dx.doi.org/10.1074/jbc.M006151200
16. Spellman R, Rideau A, Matlin A, Gooding C, Robinson F, McGlincy N, Grellscheid SN, Southby J, Wollerton M, Smith CW. Regulation of alternative splicing by PTB and associated factors. Biochem Soc Trans 2005; 33:457-60; PMID:15916540; http://dx.doi.org/10.1042/BST0330457
17. Tang ZZ, Sharma S, Zheng S, Chawla G, Nikolic J, Black DL. Regulation of the mutually exclusive exons 8a and 8 in the $CaV_{1.2}$ calcium channel transcript by polypyrimidine tract-binding protein. J Biol Chem 2011; 286:10007-16; PMID:21282112; http://dx.doi.org/10.1074/jbc.M110.208116
18. Graveley BR. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell 2005; 123:65-73; PMID:16213213; http://dx.doi.org/10.1016/j.cell.2005.07.028
19. Olson S, Blanchette M, Park J, Savva Y, Yeo GW, Yeakley JM, Rio DC, Graveley BR. A regulator of Dscam mutually exclusive splicing fidelity. Nat Struct Mol Biol 2007; 14:1134-40; PMID:21188797; http://dx.doi.org/10.1038/nsmb1339
20. Ramani AK, Calarco JA, Pan Q, Mavandadi S, Wang Y, Nelson AC, Lee LJ, Morris Q, Blencowe BJ, Zhen M, et al. Genome-wide analysis of alternative splicing in Caenorhabditis elegans. Genome Res 2011; 21:342-8; PMID:21177968; http://dx.doi.org/10.1101/gr.114645.110
21. Goodman SJ, Branda CS, Robinson MK, Burdine RD, Stern MJ. Alternative splicing affecting a novel domain in the C. elegans EGL-15 FGF receptor confers functional specificity. Development 2003; 130:3757-66; PMID:12835392; http://dx.doi.org/10.1242/dev.00604
22. Kuroyanagi H, Kobayashi T, Mitani S, Hagiwara M. Transgenic alternative-splicing reporters reveal tissue-specific expression profiles and regulation mechanisms in vivo. Nat Methods 2006; 3:909-15; PMID:17060915; http://dx.doi.org/10.1038/nmeth944
23. Kuroyanagi H, Ohno G, Mitani S, Hagiwara M. The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo. Mol Cell Biol 2007; 27:8612-21; PMID:17923701; http://dx.doi.org/10.1128/MCB.01508-07
24. Sibley MH, Johnson JJ, Mello CC, Kramer JM. Genetic identification, sequence, and alternative splicing of the Caenorhabditis elegans alpha 2(IV) collagen gene. J Cell Biol 1993; 123:255-64; PMID:7691828; http://dx.doi.org/10.1083/jcb.123.1.255
25. Ohno G, Hagiwara M, Kuroyanagi H. STAR family RNA-binding protein ASD-2 regulates developmental switching of mutually exclusive alternative splicing in vivo. Genes Dev 2008; 22:360-74; PMID:18230701; http://dx.doi.org/10.1101/gad.1620608
26. Kuroyanagi H, Ohno G, Sakane H, Maruoka H, Hagiwara M. Visualization and genetic analysis of alternative splicing regulation in vivo using fluorescence reporters in transgenic Caenorhabditis elegans. Nat Protoc 2010; 5:1495-517; PMID:20725066; http://dx.doi.org/10.1038/nprot.2010.107
27. Pujol N, Bonnerot C, Ewbank JJ, Kohara Y, Thierry-Mieg D. The Caenorhabditis elegans unc-32 gene encodes alternative forms of a vacuolar ATPase a subunit. J Biol Chem 2001; 276:11913-21; PMID:11110798; http://dx.doi.org/10.1074/jbc.M009451200
28. Kuroyanagi H, Watanabe Y, Hagiwara M. CELF family RNA-binding protein UNC-75 regulates two sets of mutually exclusive exons of the unc-32 gene in neuron-specific manners in Caenorhabditis elegans. PLoS Genet 2013; 9:e1003337; PMID:23468662; http://dx.doi.org/10.1371/journal.pgen.1003337

29. Pulak R, Anderson P. mRNA surveillance by the Caenorhabditis elegans smg genes. Genes Dev 1993; 7:1885-97; PMID:8104846; http://dx.doi.org/10.1101/gad.7.10.1885
30. Kondrashov FA, Koonin EV. Origin of alternative splicing by tandem exon duplication. Hum Mol Genet 2001; 10:2661-9; PMID:11726553; http://dx.doi.org/10.1093/hmg/10.23.2661
31. Yabe T, Suzuki N, Furukawa T, Ishihara T, Katsura I. Multidrug resistance-associated protein MRP-1 regulates dauer diapause by its export activity in Caenorhabditis elegans. Development 2005; 132:3197-207; PMID:15983401; http://dx.doi.org/10.1242/dev.01909
32. Hagen FK, Nehrke K. cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase sequence homologs from Caenorhabditis elegans. J Biol Chem 1998; 273:8268-77; PMID:9525933; http://dx.doi.org/10.1074/jbc.273.14.8268
33. Anyanful A, Sakube Y, Takuwa K, Kagawa H. The third and fourth tropomyosin isoforms of Caenorhabditis elegans are expressed in the pharynx and intestines and are essential for development and morphology. J Mol Biol 2001; 313:525-37; PMID:11676537; http://dx.doi.org/10.1006/jmbi.2001.5052
34. Irimia M, Maeso I, Gunning PW, Garcia-Fernàndez J, Roy SW. Internal and external paralogy in the evolution of tropomyosin genes in metazoans. Mol Biol Evol 2010; 27:1504-17; PMID:20147436; http://dx.doi.org/10.1093/molbev/msq018
35. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al.; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409:860-921; PMID:11237011; http://dx.doi.org/10.1038/35057062
36. Stricklin SL, Griffiths-Jones S, Eddy SR Sr. C. elegans noncoding RNA genes. WormBook 2005;1-7; PMID:18023116
37. Kagawa H, Sugimoto K, Matsumoto H, Inoue T, Imadzu H, Takuwa K, Sakube Y. Genome structure, mapping and expression of the tropomyosin gene tmy-1 of Caenorhabditis elegans. J Mol Biol 1995; 251:603-13; PMID:7666414; http://dx.doi.org/10.1006/jmbi.1995.0459
38. Mathews EA, Mullen GP, Crowell JA, Duerr JS, McManus JR, Duke A, Gaskin J, Rand JB. Differential expression and function of synaptotagmin 1 isoforms in Caenorhabditis elegans. Mol Cell Neurosci 2007; 34:642-52; PMID:17321753; http://dx.doi.org/10.1016/j.mcn.2007.01.009
39. Van Auken K, Weaver D, Robertson B, Sundaram M, Saldi T, Edgar L, Elling U, Lee M, Boese Q, Wood WB. Roles of the Homothorax/Meis/Prep homolog UNC-62 and the Exd/Pbx homologs CEH-20 and CEH-40 in C. elegans embryogenesis. Development 2002; 129:5255-68; PMID:12399316
40. Spieth J, Lawson D. Overview of gene structure. WormBook 2006:1-10; PMID:18023127
41. Zahler AM. Pre-mRNA splicing and its regulation in Caenorhabditis elegans. In: The C. elegans Research Community, ed. WormBook - Molecular biology -: http://www.wormbook.org, 2012:1-21.
42. Kuroyanagi H. Fox-1 family of RNA-binding proteins. Cell Mol Life Sci 2009; 66:3895-907; PMID:19688295; http://dx.doi.org/10.1007/s00018-009-0120-5
43. Kuroyanagi H, Watanabe Y, Suzuki Y, Hagiwara M. Position-dependent and neuron-specific splicing regulation by the CELF family RNA-binding protein UNC-75 in Caenorhabditis elegans. Nucleic Acids Res 2013; 41:4015-25; PMID:23416545; http://dx.doi.org/10.1093/nar/gkt097
44. Hollins C, Zorio DA, MacMorris M, Blumenthal T. U2AF binding selects for the high conservation of the C. elegans 3' splice site. RNA 2005; 11:248-53; PMID:15661845; http://dx.doi.org/10.1261/rna.7221605
45. Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R. Comprehensive splice-site analysis using comparative genomics. Nucleic Acids Res 2006; 34:3955-67; PMID:16914448; http://dx.doi.org/10.1093/nar/gkl556
46. Farrer T, Roller AB, Kent WJ, Zahler AM. Analysis of the role of *Caenorhabditis elegans* GC-AG introns in regulated splicing. Nucleic Acids Res 2002; 30:3360-7; PMID:12140320; http://dx.doi.org/10.1093/nar/gkf465
47. Kabat JL, Barberan-Soler S, McKenna P, Clawson H, Farrer T, Zahler AM. Intronic alternative splicing regulators identified by comparative genomics in nematodes. PLoS Comput Biol 2006; 2:e86; PMID:16839192; http://dx.doi.org/10.1371/journal.pcbi.0020086
48. Ast G. How did alternative splicing evolve? Nat Rev Genet 2004; 5:773-82; PMID:15510168; http://dx.doi.org/10.1038/nrg1451
49. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res 2004; 14:1188-90; PMID:15173120; http://dx.doi.org/10.1101/gr.849004
50. Hesselson D, Kimble J. Growth control by EGF repeats of the *C. elegans* Fibulin-1C isoform. J Cell Biol 2006; 175:217-23; PMID:17043142; http://dx.doi.org/10.1083/jcb.200608061