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

Hidehito Kuroyanagi, Satomi Takei, and Yutaka Suzuki

Laboratory of Gene Expression; Medical Research Institute; Tokyo Medical and Dental University; Bunkyo-ku, Tokyo, Japan

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. 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. 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 *Discam* gene of a fruit fly *Drosophila melanogaster*, which has four ME exon clusters containing 12, 48, 33, and two variants.

Several mechanisms have been proposed for the mutually exclusive nature of the ME exon selection. 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. 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. Disposal of aberrantly spliced mRNAs by a surveillance system termed nonsense-mediated mRNA decay (NMD) is considered to play substantial roles for some genes. Antagonism of repression by base-pairing interaction of a docker with one of the selector sequences is proposed for the *Discam* exon 6 cluster.

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. 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), 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, 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. In the case of exons 7a/7b of the *unc-32* gene, encoding subunit α of V$_0$ domain of vacuolar proton-translocating ATPase (V-ATPase), 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. Thus,
**Table 1. Summary of experimental validation of the putative mE exon clusters (continued)**

| Name and ID in WS239 | Position in WS235 | Supporting Reads** | Sequence Identity | Intervening Intron(s) [nt]** | NMD-Dependence** |
|----------------------|------------------|--------------------|-------------------|-----------------------------|------------------|
| **AS Type**<br>Homologous mE | **Chr** | **Gene** | **WBGene ID** | **Str**<br>**Exon**<br>**Left end**<br>**Right end** | **Exon Length** | **Nucleotide**<br>**Amino Acid** |
| I | F30F8.9 | WBGene00009276 | + 4a | 7837844<br>7837926 | 83<br>3n +2 | 9 | 41.0%<br>17.9% | 683 | Yes |
| | 4b | 7838610<br>7838692 | 83<br>3n +2 | 12 | |
| | 7b | 14623010<br>14623148 | 139<br>3n +1 | 1615 | |
| | 7a** | 14623571<br>14623709 | 139<br>3n +1 | 83 | |
| | 5b | 14624030<br>14624147 | 118<br>3n +1 | 478 | |
| | 5c | 14624303<br>14624420 | 118<br>3n +1 | 94 | 47.5%<br>35.9% |
| | 5a | 14624544<br>14624661 | 118<br>3n +1 | 706 | |
| II | snt-1 | WBGene00004921 | + 6b**<br>6a | 6815048<br>6815187 | 140<br>3n +2 | 0 | 64.3%<br>53.3% | 116 | No |
| | 6a | 6815304<br>6815437 | 134<br>3n +2 | 63 | |
| III | lev-11 | WBGene00002978 | - 7b<br>7a** | 14623571<br>14623709 | 139<br>3n +1 | 83 | |
| | 7a** | 14623571<br>14623709 | 139<br>3n +1 | 83 | |
| | 5b | 14624030<br>14624147 | 118<br>3n +1 | 478 | |
| | 5c | 14624303<br>14624420 | 118<br>3n +1 | 94 | 47.5%<br>35.9% |
| | 5a | 14624544<br>14624661 | 118<br>3n +1 | 706 | |
| gly-6 | WBGene00001631 | - 8b | 425419<br>4325523 | 105<br>3n | 26 | 44.1%<br>28.9% | 21 | - |
| pdfr-1 | WBGene00015735 | - 8b | 6636402<br>6636536 | 155<br>3n +2 | 77 | 53.4%<br>30.8% | 336 | Partially |
| coq-2 | WBGene00000762 | + 6a | 6937600<br>6937723 | 134<br>3n +2 | 13 | 47.0%<br>40.0% | 678 | Yes |
| gly-3 | WBGene00001630 | + 9b | 13200871<br>13200881 | 111<br>3n | 16 | 77.7%<br>54.1% | 26 | 1466 | - |
| unc-32 | WBGene00000768 | + 7a | 8909233<br>8909355 | 123<br>3n | 46 | 47.2%<br>35.4% | 725 | - |
| gly-5 | WBGene00001630 | + 9c | 13200916<br>13201020 | 105<br>3n | 13 | 66.1%<br>54.1% | 26 | 1466 | - |
| W06F12.2 | WBGene00012306 | - 2b | 13730072<br>13730243 | 172<br>3n +1 | 80 | 35.5%<br>17.5% | 129 | Yes |
| gdo-3 | WBGene00008706 | + 6a | 17478400<br>17478647 | 248<br>3n +2 | 78 | 75.4%<br>80.5% | 207 | Yes |

**ME**, mutually exclusive exons; **CE**, cassette exon. **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.** **Intervening introns shorter than 40 nt are underlined.** **Yes, NMD isoform(s) were significantly increased in the smg-2 mutant (P < 0.05 in a modified χ² test of the RT-PCR products).** **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.** **These exon 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.** **Out-of-frame tandem acceptor sites were frequently used for coq-2 exon 6b.** **mrd-1 exon 13D is aberrantly annotated in the RefSeq model used for mapping.** **These exons carry in-frame termination codons.** **Nucleotide and amino acid sequence identities of del-6 exons 5/6 appear high because of their short lengths.** **In-frame tandem acceptor sites were frequently used for unc-43 exon 14.** **Tandem donor sites were frequently used for K10D6.2 exon 5.**
| Name and ID in WS239 | Position in WS235 | Supporting Reads* | Sequence Identity | Intervening Intron(s) [nt]** | NMD-Dependence*** |
|----------------------|------------------|------------------|------------------|-----------------------------|------------------|
|                      | AS Type | Chr | Gene | WBGene ID | Str | Exon | Left end | Right end | Exon Length | Nucleotide | Amino Acid |                      |                      |                      |
| Homologous mE       | V       | unc-62 | WBGene00006796 | + | 7a*** | 4504975 | 4505120 | 146 | 3n +2 | 4 | 49.1% | 30.2% | 238 | No |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 138 | Partially |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 294 | - |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 225 | Partially |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 856 | - |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 334, 154, 435 | - |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 97 | Partially |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 17 | - |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 30 | - |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 117 | - |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 370 | 21.6% |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 1371 | Partially |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 164 | 50.0% |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 53.8% | 53.8% |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 7.9% | 59.0% |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 7.4% | 53.8% |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 69.2% | 50.0% |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 8.8% | 54.0% |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 35.0% | 35.0% |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 65.8% | 56.8% |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 172 | 172 |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 72 | 72 |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 54 | 54 |
|                      |         |       |       |           |     |       |          |          |             |              |                |               | 288, 237 | Partially |

ME, mutually exclusive exons; CE, cassette exon. *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. **Intervening introns shorter than 40 nt are underlined. ***Yes, NMD isoform(s) were significantly increased in the smg-2 mutant (P < 0.05 in a modified χ2 test of the RT-PCR products). 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. ****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. "Out-of-frame tandem acceptor sites were frequently used for coq-2 exon 6B. **mrp-1 exon 13D is aberrantly annotated in the RefSeq model used for mapping. ***These exons carry in-frame termination codons. ****Nucleotide and amino acid sequence identities of del-6 exons S/6 appear high because of their short lengths. *****In-frame tandem acceptor sites were frequently used for unc-43 exon 14. ******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** | Sequence Identity | Intervening Introns (nt)** | NMD-Dependence** |
|----------------------|-------------------|-------------------|-------------------|---------------------------|------------------|
| AS Type | Chr | Gene | WBGene ID | Str | E EXON | LEFT END | RIGHT END | EXON LENGTH | NUCLEOTIDE | AMINO ACID | NMD-DEPENDENCE |
| Non-Homologous mE | N | fbl-1 | WBGene0000403 | + | 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 | 1n | 4 | 53% | 60% | 200 | Yes |
| | | | | | 6 | 10574934 | 10575098 | 165 | 3n | 230 | - | - | - | - |
| | X | egl-15 | WBGene00001184 | + | SB | 11017580 | 11017930 | 351 | 3n | 38 | - | 40.1% | 10.3% | 14 | - |
| | | | | | 5A | 11017945 | 11018139 | 195 | 3n | 5 | - | - | - | - |
| | | | | | | | | | | | | | | | - |
| Tandem CE | I | som-7 | WBGene00006594 | + | 17 | 5557960 | 5557971 | 12 | 3n | - | - | - | - | - | - |
| | | | | | 18 | 5558453 | 5558479 | 27 | 3n | - | - | - | - | - | - |
| | Tandem CE | I | krv-11 | WBGene00002978 | - | 9b | 14622304 | 14622511 | 208 | 3n | 1n | - | - | - | - | - |
| | | | | | 9a | 14622658 | 14622743 | 86 | 3n | 2n | - | - | - | - | - | - |
| | Single CE | II | etr-7 | WBGene0000340 | + | 4 | 166071 | 166150 | 80 | 3n | 2 | - | - | - | - | - |
| | Tandem CE | II | zgo-72 | WBGene00006897 | + | 8 | 4952490 | 4952531 | 42 | 3n | 1n | - | - | - | - | - |
| | | | | | 9 | 4952647 | 4952694 | 48 | 3n | - | - | - | - | - | - |
| | Tandem CE | II | C34F11.3 | WBGene00016415 | + | 10 | 5205040 | 5205075 | 36 | 3n | - | - | - | - | - | - |
| | | | | | 11 | 5205444 | 5205632 | 189 | 3n | - | - | - | - | - | - |
| | Tandem CE | III | cdp-1 | WBGene0000542 | + | 3 | 7981673 | 7981810 | 138 | 3n | - | - | - | - | - | - |
| | | | | | 4 | 7982464 | 7982538 | 75 | 3n | - | - | - | - | - | - |
| | Constitutive | IV | unc-44 | WBGene00006780 | + | 13 | 5984796 | 5984980 | 185 | 3n | 2 | - | - | - | - | - |
| | | | | | 14 | 5985093 | 5985170 | 78 | 3n | - | - | - | - | - | - |
| | | | | | 15 | 5985577 | 5986355 | 779 | 3n | 2 | - | - | - | - | - |
| | Undetected | | | | | | | | | | | | | | - | - |
| | Alternative Acceptors | IV | unc-43 | WBGene00006779 | - | 14S | 10329097 | 10329195 | 99 | 3n | - | - | - | - | - | - |
| | | | | | 15 | 10329397 | 10329601 | 304 | 3n | - | - | - | - | - | - |
| | Tandem CE | V | K10D6.2 | WBGene000010742 | - | 5L | 11194203 | 11194342 | 140 | 3n | 2 | - | - | - | - | - |
| | | | | | 4 | 11194574 | 11194627 | 54 | 3n | - | - | - | - | - | - |
| | Tandem CE | V | ysh-23 | WBGene00013481 | - | 7 | 18661622 | 18661711 | 90 | 3n | - | - | - | - | - | - |
| | | | | | 6 | 18662119 | 18662313 | 195 | 3n | - | - | - | - | - | - |
| | Constitutive | X | F49E2.5 | WBGene00008888 | + | 5 | 9554299 | 9554547 | 249 | 3n | - | - | - | - | - | - |
| | | | | | 6 | 9554692 | 9554823 | 132 | 3n | - | - | - | - | - | - |
| | Constitutive | X | T22E7.2 | WBGene00020832 | + | 16 | 17679922 | 17680035 | 114 | 3n | - | - | - | - | - | - |

ME, mutually exclusive exons; CE, cassette exon. **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. **Intervening introns shorter than 40 nt are underlined. **Yes, NMD isoform(s) were significantly increased in the smg-2 mutant (P < 0.05 in a modified χ2 test of the RT-PCR products). 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 wild type and the smg-2 mutant. **These exons were expressed at very low levels at the L1 stage. At the young adult stage, lev-11 exon 7a and smt-1 exon 6b were still rare, while unc-62 exon 7a was readily detected. **Out-of-frame tandem acceptor sites were frequently used for coq-2 exon 6b. **mrp-1 exon 13D is aberrantly annotated in the RefSeq model used for mapping. **These exons carry in-frame termination codons. **Nucleotide and amino acid sequence identities of del-6 exons 5/6 appear high because of their short lengths. **In-frame tandem acceptor sites were frequently used for unc-43 exon 14. **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. 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*.

### 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 exon pairs 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.

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, 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, 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 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. Exons 5a/5c/5b of the *lev-11* gene, encoding multiple tropomyosin isoforms, 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). The complex structures and pre-mRNA processing patterns of the tropomyosin genes are evolutionarily conserved in metazoans, 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*. Among the ME exon clusters we have already reported, the intervening introns for exons 5B/5A of the *egl-15* gene and exons 9/10 of the *let-2* gene are 14 nt and 30 nt, respectively, and we...
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 for this cluster. (D) *unc-32* exons 7a/7b. (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. (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. (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.

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. (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. (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. (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.

As U11 or U12 snRNA or an AT/AC splice-junction are not found in *C. elegans*, 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).
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 let-11 exons 7a/7b, snt-1 exons 6B/6A, and unc-62 exons 7a/7b, 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, let-11 exons 5a/5c/5b and let-805 exons 19a/19b, the RT-PCR patterns were indistinguishable between the wild-type and the smg-2 mutant (Table 1; Fig. 1G–I). 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).43 Most of the ME exons (60 of 63) are shorter than 260 nt and the average size (151 nt) is close to the median (201 nt).43 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, 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.44 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-1.45 Six out of the 29 clusters are affected in the unc-75 mutant, 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)43 and the donor site (AG/GTAAGTT)43 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.48 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. RT-PCR was performed essentially as described previously. RT-PCR products were analyzed by using BioAnalyzer (Agilent) as described previously. 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 Eksueru-Toukei 2010 (Social Survey Research Information). Sequence logos were generated by using WebLogo39 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:21922127; http://dx.doi.org/10.1038/nrg3152
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:20367054; 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 decay pathway and mutually exclusive expression of alternatively spliced FGFR2IIIb and -IIIc mRNAs. J Biol Chem 2001; 276:4158-67; PMID:11044206; http://dx.doi.org/10.1074/jbc.M00651200
8. Spellman R, Rideau A, Marlin A, Gooding C, Robinson F, McGlincy N, Grellischeid SN, Southby J, Wölkert M, Smith CW. Regulation of alternative splicing by PTB and associated factors. Biochimie Soc Trans 2005; 33:457-60; PMID:15996540; http://dx.doi.org/10.1014/jbt.03034045
9. 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 an unusual lariat branch point location: implications for constitutive splicing. J Cell Biol 2011; 184:10007-16; PMID:21282112; http://dx.doi.org/10.1083/jcb.2011081161
10. Grønberg-Borg J, Branda CS, Robinson MK, Burdine RD, Stem 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/10.1042/dve06604
11. 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; 3:999-1005; PMID:17060915; http://dx.doi.org/10.1038/nmeth894
12. Kuroyanagi H, Ohno G, Mitani S, Hagiwara M. The Fox1 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
13. Silby 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:7601828; http://dx.doi.org/10.1083/jcb.123.1.255
14. 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.1620680
15. 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
16. Pujol N, Bonnerot C, Ebwbank J, Kohara Y, Thierry-Mieg D. The Caenorhabditis elegans unc-32 gene encodes alternative forms of a vacuolar ATPase subunit. J Biol Chem 2001; 276:19131-21; PMID:11100798; http://dx.doi.org/10.1074/jbc.M009451200
17. Kuroyanagi H, Watanebe 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:e1003357; PMID:23468662; http://dx.doi.org/10.1371/journal.pgen.1003357
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.7.1885
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, Imadu H, Takawa 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/10.1101/gad.7.10.1885
38. Mathews EA, Mullen GP, Crowell JA, Duerr JS, McMannis 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 Aulen K, Weaver D, Robertson B, Sundaram M, Saldi T, Edgar L, Lee M, Boese Q, Wood WB. Roles of the Homorox/Meis/Prep homolog UNC-62 and the Exd/Pbx homologs 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 C. elegans RNA-binding protein UNC-75 in Caenorhabditis elegans. Nucleic Acids Res 2013; 41:4015-25; PMID:23416453; 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:15660145; http://dx.doi.org/10.10121/rna.222609
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/gk0156
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. Kahalu 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. pone.0020086
48. Ast G. How did alternative splicing evolve? Nat Rev Genet 2004; 5:773-82; PMID:15510168; http://dx.doi.org/10.1038/nrg1541
49. Crooks GE, Hon G, Chand proxies BM, Brenner SE, WebLogo: a sequence logo generator. Genome Res 2004; 14:1188-90; PMID:15373120; 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