Competing RNA pairings in complex alternative splicing of a 3' variable region

HUAWEI PAN,¹ YANG SHI,¹ SHUO CHEN, YUN YANG, YUAN YUE, LEILEI ZHAN, LANZHI DAI, HAIYANG DONG, WEILING HONG, FENG SHI, and YONGFENG JIN

Institute of Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang, ZJ310058, P.R. China

ABSTRACT

Alternative pre-mRNA splicing remarkably expands protein diversity in eukaryotes. *Drosophila* PGRP-LC can generate three major 3' splice isoforms that exhibit distinct innate immune recognition and defenses against various microbial infections. However, the regulatory mechanisms underlying the uniquely biased splicing pattern at the 3' variable region remain unclear. Here we show that competing RNA pairings control the unique splicing of the 3' variable region of *Drosophila* PGRP-LC pre-mRNA. We reveal three roles by which these RNA pairings jointly regulate the 3' variant selection through activating the proximal 3' splice site and concurrently masking the intron-proximal 5' splice site, in combination with physical competition of RNA pairing. We also reveal that competing RNA pairings regulate alternative splicing of the highly complex 3' variable regions of *Drosophila* CG42235 and *Pip*. Our findings will facilitate a better understanding of the regulatory mechanisms of highly complex alternative splicing as well as highly variable 3' processing.

Keywords: RNA secondary structure; alternative splicing; PGRP-LC; combinatorial mechanism

INTRODUCTION

Alternative pre-mRNA splicing greatly expands proteomic diversity and the complexity of gene expression regulation in eukaryotes (Black 2003; Matlin et al. 2005; Blencowe 2006; Chen and Manley 2009; Keren et al. 2010; Nilsen and Graveley 2010; Matera and Wang 2014; Lee and Rio 2015; Baralle and Giudice 2017). Alternative splicing includes different splicing processes, such as intron retention, exon skipping, alternative 5' or 3' splice sites, mutually exclusive splicing, and alternative 5' or 3' terminal exons (Black 2003; Matlin et al. 2005; Blencowe 2006; Keren et al. 2010; Nilsen and Graveley 2010). It is remarkable that mutually exclusive alternative splicing of the *Drosophila melanogaster* Down syndrome cell adhesion molecule (*Dscam1*) potentially encodes 38,016 different mRNA isoforms (Schmucker et al. 2000). In this study, 95 alternative exons were internally organized into clusters containing 12, 48, 33, and two variable tandem exons for exons 4, 6, 9, and 17, respectively. Analogous gene architecture is found in 3' tandem cassette clusters, which result from the duplication of the 3' terminal cassette (McManus et al. 2010; Hatje and Kollmar 2013). Recent studies indi-

cate that hundreds of fly and human genes contain tandem 3' organization (Hatje and Kollmar 2013; Hatje et al. 2017). These gene architectures are compatible with alternative splicing coupled with multiple polyadenylation (Elkon et al. 2013), resulting in mutually exclusive mRNA isoforms with various coding regions and a 3' untranslated region (3'-UTR) (Choe et al. 2002; Danielli et al. 2003; Goeke et al. 2003; Hatje et al. 2017). Protein isoforms with different carboxy-terminal domains could regulate a myriad of cellular and developmental processes, and their mutation may be associated with numerous diseases. In addition, alternative 3'-UTRs play diverse regulatory roles in RNA stability and translation as well as protein localization (Sandberg et al. 2008; Yatsenko et al. 2014; Berkovits and Mayr 2015; Rabani et al. 2017; Tian and Manley 2017).

Peptidoglycan recognition proteins (PGRPs) are ubiquitous proteins involved in innate immunity and are conserved from insects to mammals (Kang et al. 1998). PGRP-LC acts as the primary receptor in initiating an innate immune response via the immune deficiency (IMD) pathway (Royet and Dziarski 2007). The PGRP-LC gene encodes

¹These authors contributed equally to this work.

Corresponding author: jinyf@zju.edu.cn

Article is online at <http://www.majournal.org/cgi/doi/10.1261/rna.066225.118>.

© 2018 Pan et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

three major PGRP-LC isoforms (PGRP-LCx, PGRP-LCy, and PGRP-LCa) via alternative splicing (Werner et al. 2000, 2003; Choe et al. 2002, 2005). These isoforms share common cytoplasmic and transmembrane domains but have variable extracellular domains and display different recognition capabilities for various microbial patterns (Werner et al. 2000, 2003; Choe et al. 2002, 2005; Kaneko et al. 2004; Schmidt et al. 2008; Neyen et al. 2016). It is interesting that recent evidence indicates that PGRP-LC isoforms are also candidate receptors for retrograde trans-synaptic signaling in the *Drosophila* nervous system (Harris et al. 2015).

The PGRP-LCx transcript is entirely analogous to PGRP-LCy and uses the same 5' splice site in the second exon, 57 bp upstream of the one used for the PGRP-LCa transcript (Fig. 1A; Werner et al. 2003). As a result, the link between the PGRP and the transmembrane domain is 19 amino acid residues shorter in PGRP-LCx and LCy than in LCa. The 5' intron-distal splice site tends to be spliced with the 3'

splice site of the PGRP-LCx and LCy isoforms, whereas the 5' intron-proximal splice site is biased toward 3' distal splice site usage in isoform PGRP-LCa (Fig. 1A; Werner et al. 2003). However, the mechanism by which the uniquely biased splicing is regulated remains unknown. First, this splicing bias cannot be explained by splice site strength, because LCa is not biased toward the stronger 5' distal splice site like LCx and LCy. Second, proximity rules cannot account for this unique splice site selection (Roca et al. 2013). Finally, spliceosomal incompatibility involving the specific arrangement of splice sites in the U2- and U12-type spliceosome regulates alternative splicing of the *D. melanogaster prospero* twintron (Scamborova et al. 2004). However, this mechanism is not involved with PGRP-LC, because these splice sites conform to the major spliceosome types. Thus, the mechanism underlying 3' variable cassette choice in PGRP-LC may require the presence of additional splicing regulatory elements.

FIGURE 1. The inclusion of PGRP-LC isoforms at the 3' variable region is biased. (A) Exon–intron organization and multiple 3' alternative cassettes of *Drosophila* PGRP-LC. Constitutive exons (black boxes), alternative exons (colored boxes), and introns (lines) are shown. The strength of the splice site is presented with the scores shown above. Splice site scores were calculated using a splice site prediction program (Reese et al. 1997) (range of 0–1). Three PGRP domains, a, x, and y, are alternatively spliced to the common exon 2 of the PGRP-LC gene. Letters denote the PGRP domains included in the alternative isoforms. The PGRP-LCy transcript is analogous to PGRP-LCx and uses the same 5' splice site in exon 2, 57 bp upstream of the one used with the PGRP-LCa transcript (Werner et al. 2003). (B) 3D structural comparisons of PGRP splice isoforms. Green regions are encoded by alternative variants. (C) The alternative splicing pattern was analyzed by RT-PCR. The primers used in each experiment are depicted in A. (D) Heatmap of the expression of splice isoforms in different tissues. Based on the exon–exon junction reads from RNA-seq data, we calculated the splice isoforms in different tissues. These results indicate that PGRP-LC splice isoforms are biased toward x-S, y-S, and a-L in a tissue-specific manner.

Recent studies indicate that competing RNA pairings guarantee that only one variant exon is spliced from a cluster of duplicated exons (Graveley 2005; Anastassiou et al. 2006; May et al. 2011; Yang et al. 2011; Suyama 2013; Yue et al. 2016b, 2017; Hatje et al. 2017; Jin et al. 2018). This mechanism was initially proposed in the exon 6 cluster of *Drosophila Dscam*, which is involved in two types of intronic elements: the docking site and selector sequences (Graveley 2005). The former, which is located in the intron downstream from exon 5, may competitively pair with the selectors upstream of each exon 6 (48 total) (Graveley 2005; May et al. 2011). Other players are also involved in this competing RNA pairing system, such as a splicing regulator (the heterogeneous nuclear ribonucleoprotein, hrp36) (Olson et al. 2007) and a locus control region enhancer (Wang et al. 2012). Considering the similar origin of duplicated exons between the internal and 3' regions, such competing RNA pairings may provide an alternative explanation for alternative splicing of the highly complex 3' variable regions.

In this study, a combination of comparative genome analyses with mutation experiments revealed that competing RNA pairings control alternative splicing at the 3' variable region of *Drosophila PGRP-LC* pre-mRNA. We reveal the triple roles of these RNA pairings for combinatorial regulation of the selection of 3' variants through activating alternative exon 3 and concurrently masking the proximal 5' splice site of exon 2 in combination with physical competition for RNA pairing. The frequency of 3' alternative cassette selection is not only correlated with RNA pairing strength but is also closely related to its position. Moreover, we reveal that competing RNA pairings regulate alternative splicing at the highly complex 3' variable region of *CG42235* and *Pip*. Thus, our findings will facilitate a better understanding of the regulatory mechanisms of complex alternative splicing and highly variable 3' processing.

RESULTS

Competing RNA pairings at the 3' variable region of *Drosophila PGRP-LC*

Consistent with previous studies (Werner et al. 2003), RT-PCR and analysis of exon junctions from RNA-seq data from different developmental stages showed that the 5' intron-distal splice site was largely spliced with the 3' splice site of isoforms *PGRP-LCx* and *LCy*, whereas the 5' intron-proximal splice site was biased toward 3' distal splice site usage in the *PGRP-LCa* transcript (Fig. 1A,C,D). To identify the *cis*-elements involved in regulating this biased selection of the 3' alternative cassette, we used comparative sequence analysis to search for sequences that were conserved among *Drosophila* species. The organisms analyzed included 12 *Drosophila* species spanning approximately 40

million years. It is interesting that sequence alignment revealed two conserved intronic elements (CE1 and CE2) in an individual intron upstream of the tandem cluster (Fig. 2A) that showed high similarity to each other. We identified their reverse complementary sequences (RE) in the intron-exon boundary region of alternative exon 2 (Fig. 2A) by probing with the CE1 and CE2 sequences. The predicted RNA pairing between CE1 or CE2 and the RE sequence is highly conserved in *Drosophila* species (Supplemental Figs. S1, S2). This architecture is analogous to the competing RNA structures governing internal mutually exclusive splicing of *Drosophila Dscam1* and *14-3-3 ξ* pre-mRNA (Graveley 2005; Yang et al. 2011). Therefore, we propose that selection of the *PGRP-LC* 3' splice isoform is regulated by competing RNA pairings.

Competition between intra- and interintron RNA pairing directs *PGRP-LCx* and *LCy* selection

To explore how RNA elements and structures are involved in the selection of the *PGRP-LC* 3' alternative cassette, we first generated a minigene construct containing the 3' alternative cassette *LCx* and *LCy* under the inducible metallothionein promoter (Fig. 2C). Splice isoforms containing the 3' alternative cassette *LCx* and *LCy* were detected in transfection experiments using *Drosophila* S2 cells, which could help elucidate the splicing pattern of the endogenous gene (Fig. 2C–E). This system is therefore well suited for analyzing *cis*-elements involved in the selection of 3' alternative cassettes. Next we tested the effects on 3' alternative cassette selection of disruptive and compensatory mutations in the predicted stem structure. We demonstrated that mutation of CE1 (M2) markedly reduced *LCx* content, whereas *LCy* content increased (Fig. 2D,E), which indicates that *LCx* inclusion is dependent on the CE1 element. Conversely, mutation of CE2 (M3) eliminated *LCy* almost completely (Fig. 2D,E), which reveals that *LCy* inclusion is specifically dependent on the CE2 element. Mutated RE (M1), which is unable to pair with CE1 or CE2, markedly increased *LCx* levels, suggesting that disruption of competing RNA pairings may have a dissimilar effect on *LCx* and *LCy* inclusion (Fig. 2D,E). It is interesting that a structure-restoring double mutation (M12) increased the efficiency of *LCx* inclusion up to 95%, which was much higher than that of the wild type (WT), whereas *LCy* inclusion was fully inhibited (Fig. 2D,E). An explanation for this effect is that mutated RE and mutated CE1 have compensatory pairing, whereas mutated RE is clearly unable to pair with CE2. Conversely, RE and CE2 double compensatory mutation (M13) led to the predominant inclusion of *LCy*, whereas *LCx* was not included (Fig. 2D,E). Collectively, these observations confirm that the selection of *LCx* and *LCy* is dependent on competitive base pair interactions.

FIGURE 2. The inclusion of PGRP-LCx and LCy isoforms is directed by competing RNA pairings. (A) The arrangement of *cis* intronic elements at the 3' variable region. Symbols used are the same as those in Figure 1A. Shown above are the sequences of consensus intronic sequences for different species. The most identical nucleotides in the docking site (RE) and selector sequences (CE1 and CE2) are shown in green and red, respectively. Sequences of the same color are highly similar. For predicted intronic RNA pairings, see Supplemental Figure S1. Docking site (RE, marked ♥) was reverse-complementary to downstream selector sequences (CE1 and CE2, marked ♠) in a competitive mode. (B) Predicted mutually exclusive RNA pairings. Mutations introduced into dsRNA are indicated on the left or right mutated sequences (M1–M3). Green arrows depict activated inclusion of the alternative exon. (C) Overview of the minigene constructs of *D. melanogaster* used to assess the effects of RNA pairing on splicing. (D) Effects of competing RNA pairings were validated by disruptive single mutations (M1–M3) and compensatory double mutations (M12, M13). (E) Quantitation of the data in D. Data are expressed as the percentage mean $\pm$ standard deviation (SD).

Correlating RNA pairing strength with the frequency of 3' alternative cassette selection

To elucidate how the strength of RNA pairing modulates 3' alternative cassette selection, we performed a series of mutations of the CE1 element to alter the thermodynamic stability of RE-to-CE1 pairing (Fig. 3A). Mutations that strengthened RE-to-CE1 pairing prominently increased the 3' cassette LCx frequency (Fig. 3B,C). In contrast, weakening RE-to-CE1 pairing led to a dramatic reduction in 3' cassette LCx frequency (Fig. 3B,C). A similar trend was observed when we altered the strength of RE-to-CE2 RNA pairing to test LCy 3' cassette selection (Fig. 3D–F). Taken together, these mutation analyses indicate that the frequency of LCx or LCy 3' cassette selection is positively correlated with the strength of RNA pairing.

However, these results do not coincide with the correlation between the strength of RNA pairing and the targeted variant selection in endogenous PGRP-LC of WT flies (Fig. 1C). The LCx isoform was generally much more abundant than the LCy isoform in all tissues examined, although RE-to-CE1 and RE-to-CE2 RNA pairing exhibited similar strength (Fig. 2B). This result suggests that the inclusion frequency of the 3' cassette cannot be explained merely

by pairing strength between the docking site and selector sequence.

The effect of splice site strength on RNA pairing-mediated 3' alternative cassette selection

Many studies have indicated that factors such as the splice site strength of the 3' cassette, the position of the 3' cassette, the rate of transcription, and RNA binding proteins may cooperate with secondary RNA structures to contribute to the outcomes of alternative exons (Jin et al. 2011, 2018; May et al. 2011; McManus and Graveley 2011). We first investigated how 3' splice site strength contributes to pairing-mediated selection of the 3' cassette. To this end, we performed a series of experiments by changing the 3' splice site strength (Fig. 4A). We hypothesized that inclusion of 3' cassette LCx and LCy could be regulated if splice sites were strengthened or weakened. Results showed that the inclusion efficiency of LCx decreased when LCx 3' splice sites were weakened (Fig. 4B,C). Conversely, the inclusion efficiency of LCx increased when the LCx 3' splice site strength improved (Fig. 4B, C). Similarly, strengthening the LCy 3' splice site led to a

FIGURE 3. RNA pairing strength is correlated with 3' alternative cassette selection of *PGRP-LC* pre-mRNA. (A) Schematic diagrams of the constructs used to assess the effects of RNA pairing on alternative cassette *LCx* inclusion. Symbols used are the same as those in Figure 1. Predicted RNA pairing for the WT and a series of mutants (point mutations are shown in red) with the estimated equilibrium free energies (given in kcal/mol). (B) The strength of RNA pairing modulated the alternative cassette *LCx* inclusion according to mutation analysis. (C) The strength of intronic RNA pairing was correlated with *LCx* selection. (D) Schematic diagrams of the constructs used to assess the effects of RNA pairing on *LCy* inclusion. (E) The strength of intronic RNA pairing modulated *LCy* inclusion according to mutation analysis. (F) The strength of intronic RNA pairing was correlated with *LCy* inclusion.

substantial increase in *LCy* inclusion. This result shows that the strength of the 3' alternative splice site is correlated with the frequency of alternative 3' cassettes. However, *LCy* was included at much lower levels than *LCx*, even when the 3' splice site of *LCy* was optimized. These results demonstrate that the difference in 3' splice site strength cannot account for the apparent discrepancy between *LCx* and *LCy* frequency. This also implies that the splice site strength of the alternative 3' cassette likely plays a partial, rather than a determining, role in pairing-mediated selection of the 3' cassette.

Proximity-mediated 3' alternative cassette selection

Next, to discern how the position (order) of the variable cassette within the cluster contributes to the inclusion of

the 3' alternative cassette, we examined the effects of exchanging the 3' cassette *LCx* and *LCy* positions on 3' alternative cassette selection (Fig. 4D). We were intrigued to find that *LCx* and *LCy* inclusion patterns were switched when the order of the 3' cassette *LCx* and *LCy* was changed (Fig. 4E,F). Given that the *LCx* cassette still contained a slightly stronger 3' alternative splice site than the *LCy* cassette, the difference in 3' alternative splice site strength could not account for this change in the splicing pattern. In addition, RE-to-CE1 and RE-to-CE2 RNA pairing showed similar strengths, which precludes the possibility that pairing strength is responsible for a notable difference. Moreover, switching the positions of the whole *LCx* and *LCy* cassettes avoided overlapping effects of their specific enhancers or silencers. We thus suggest that the frequency of alternative 3' cassette inclusion is associated

FIGURE 4. Splice site strength and order influence 3' alternative cassette selection. (A) An overview of the various mutant constructs generated to assess the effects of splice site strength on 3' alternative cassette selection. Red arrow depicts increasing or decreasing strength of the splice site. The WT and modified 3' splice site sequences are presented with the scores shown on the right. Uppercase letters indicate exon sequences; lowercase letters indicate intron sequences. Mutated nucleotides are marked in red. (B) Effects of splice site strength mutations on the inclusion of alternative cassette LCx or LCy. RT-PCR was performed to detect the RNA splicing pattern. (C) Effects of the mutations on the inclusion of alternative cassette LCx or LCy. (D) Schematic diagrams of the mutant constructs used to assess the effects of alternative cassette order on its selection. (E) Effects of alternative cassette order on its selection. LCx and LCy inclusion was switched when their order was exchanged. (F) The frequency of alternative 3' cassette is dependent on its position (order). Data are expressed as the mean $\pm$ SD from three independent experiments.

with the proximity of the selector sequence to the docking site. This idea is supported by the finding of a much higher frequency of the variant proximal to the docking site (RE) than that of the variant distal to RE, which is consistent with the much shorter distance of the docking site to the proximal selector (0.3 knt) compared with the distal selector (1.25 knt) (Fig. 4E,F). Therefore, we infer that RNA pairing acts in a distance-mediated manner in *PGRP-LC* pre-mRNA.

RNA pairing inhibits the *PGRP-LCx* and *LCy* long isoforms

Next, we sought to determine why *PGRP-LCx* and *LCy* transcripts are biased toward the short isoform. Because the intron-proximal 5' splice site is embedded in competing RNA structures, we speculated that these structures act concurrently to inhibit usage of the intron-proximal

5' splice site. To confirm this hypothesis, we tested disruptive and compensatory mutations in the RNA stem using transfection experiments (Fig. 5A). Mutations that lowered RE-to-CE1 pairing strikingly increased the frequency of long *LCx* (Fig. 5B,C) concomitant with unmasking the proximal 5' splice site from the RNA structure. A series of mutational analyses indicated that the strength of RNA pairing was positively correlated with the frequency of short *LCx* inclusion (Fig. 5B,C). Thus, the RNA structure inhibits the long variant of *LCx* by dynamically suppressing the proximal 5' splice site in combination with simultaneous activation of the 3' short *LCx* cassette. A similar trend was obtained for RE-to-CE2 double compensatory mutation, whereas CE2 mutation led to a shift from short to long cassette *PGRP-LCy* (Fig. 5D–F) in the splicing pattern. Similarly, the strength of RNA pairing was positively correlated with the frequency of short variant *LCy* selection (Fig. 5D–F). Together, these results indicate that

FIGURE 5. Long *PGRP-LCx* and *LCy* isoforms are inhibited by RNA pairing. (A) An overview of various mutants used to assess the effects of RNA pairing on inhibiting the long *PGRP-LCx* isoform. Predicted RNA pairing for the WT and a series of mutants (C1M1–C1M3, point mutations are shown in red) with the estimated equilibrium free energies (given in kcal/mol). Black arrow depicts an intron–exon boundary. Symbols used are the same as those in Figure 1. (B) The strength of RNA pairing modulated the *LCx* splice pattern according to mutation analysis. (C) Quantitation of the data in B. Data are expressed as the percentage mean $\pm$ SD from three independent experiments. *LCx* inclusion was negatively correlated with the strength of RNA pairing. (D) An overview of various mutants used to assess the effects on inhibiting long *PGRP-LCy*. Predicted RNA pairing for the WT and a series of mutants (C2M1–C2M3) is shown. (E) The strength of RNA pairing modulated the *LCy* splice pattern according to mutation analysis. (F) Quantitation of the data in E. Inclusion of long *LCy* was negatively correlated with the strength of RNA pairing.

RNA pairing controls the bias toward short isoforms in *LCx* and *LCy* transcripts by masking the intron-proximal 5' splice site.

The docking site acts as an enhancer of long *PGRP-LCa* 3' cassette selection

Next, we investigated how the 3' alternative cassette *PGRP-LCa* is regulated. One possibility is that splicing of *LCa* is directed by RNA pairing, as is splicing of *LCx* and *LCy*. Indeed, we observed short conserved sequences upstream of the 3' alternative cassette *LCa* that were predicted to pair with the docking site. However, mutation analysis indicated that this predicted RNA pairing could not activate the 3' alternative cassette *LCa* (data not shown). In contrast to the preferential intron-distal 5' splice site usage in isoforms *LCx* and *LCy*, *LCa* is biased toward proximal 5' splice site usage. Thus, the mechanism in-

involved in selecting the 3' alternative cassette *LCa* may be distinct from the interaction between the docking site and the selector sequence seen in *LCx* and *LCy*.

Next, we tested the effects of the docking site and RNA pairings on selection of the *LCa* isoform by disruptive and compensatory mutations. RE mutation (M3) or combined CE1 and CE2 mutation (M12), which disrupted RNA pairing, led to significant shift splicing to *LCa*, while CE1 or CE2 mutation (M1, M2) did not (Supplemental Fig. S3). However, compensatory mutations (M13, M23) did not revert this splicing shift (Supplemental Fig. S3). To further dissect how the docking site affects selection of the *LCa* isoform, we generated a series of constructs by changing the docking site (RE) context upstream of the 5' intron-proximal splice site, albeit with RNA secondary structures similar to WT (Fig. 6A). These mutations had little effect on the ratio of short to long form *LCx* and *LCy* (Fig. 6B). In contrast, these mutations significantly decreased and

FIGURE 6. The docking site acts as an enhancer for long *PGRP-LCa* 3' cassette selection. (A) Overview of the minigene constructs of *D. melanogaster* used to assess the effects of RNA pairing on splicing. Symbols used are the same as those in Figure 1. A series of constructs were generated by changing the docking site (CE1) context upstream of the 5' intron-proximal splice site, albeit with similar RNA secondary structures to those of the WT. Predicted RNA pairing for the WT and a series of mutants (Ma1–Ma6, point mutations are shown in red) with the estimated equilibrium free energies (given in kcal/mol). (B) Effects of RNA pairing on *LCx*, *LCy*, and *LCa* inclusion according to mutation analysis. These mutations had little effect on the ratio of long to short cassette *LCx* and *LCy*, but significantly decreased and even switched the ratio of long to short cassette *LCa*. (C) Mutations within the docking site modulated the ratio of long to short cassette *LCa*. Quantitation of the data from B. Data are expressed as the percentage mean $\pm$ SD from three independent experiments. (D) Schematic diagrams of the constructs used to screen splicing proteins by RNA affinity purification and mass spectrometry. The sequences of the WT and mutated (Mu) single-stranded RNAs used for affinity chromatography, respectively, are shown. (E) Effects of protein knockdown on the ratio of long to short cassette *LCa*. *PGRP-LC-M12*, in which the selector sequences (CE1 and CE2) have been deleted, was used for cotransfection. These experiments revealed that the depletion of B52 resulted in a significant decrease in the ratio of long to short cassette *LCa*. (F) Quantitation of the data from E. Data are expressed as the percentage mean $\pm$ SD from three independent experiments. (*) $P < 0.05$.

even reversed the ratio of long to short cassette *LCa* (Fig. 6B,C). For example, a structurally silent mutation (M6) caused a complete switch from the long to the short cassette *LCa* (Fig. 6B,C). It is interesting that one nucleotide change (M4), which had RNA secondary structures identical to WT, led to a significant shift from long to short cassette *LCa* (Fig. 6A–C). These observations further validated the idea that *LCa* selection is not depen-

dent on docking site-mediated base-pairing but rather on the sequence-specific context of the docking site. Based on these observations in mutation experiments, combined with the inverse correlation of the ratio of long to short cassette *LCa* with the strength of the proximal to distal 5' splice site (Fig. 1A), we speculate that an enhancer might be present in the middle region of the docking site.

To gain insights into how the docking site facilitates the inclusion of long cassette *L*Ca, we identified RNA binding proteins through affinity purification of in vitro-transcribed RNA with known sequences (Fig. 6D; Supplemental Fig. S4). Mass spectrometric analysis showed that RNA-associated candidate proteins were largely related to splicing, including the serine/arginine-rich (SR) proteins B52, U2af38, U2af50, and SmB (Fig. 6E; Supplemental Fig. S4). To elucidate the roles of candidate proteins in *PGRP-LC* splicing, we performed RNA interference, in which *D. melanogaster* S2 cells were treated with various double-stranded RNAs (dsRNAs). Depletion of B52 significantly decreased the ratio of long to short cassette *L*Ca (Fig. 6F), whereas depletion of most SR proteins did not have a significant effect. Gel shift assays showed that B52 slightly bound to WT but not to mutant RNA (Supplemental Fig. S4), suggesting that it promotes the inclusion of long *L*Ca in different ways (Olson et al. 2007). These results imply that the core motif in the middle of the docking site might recruit some activators (i.e., SR proteins) to enhance recognition of the relatively weak 5' splice site that is immediately downstream and thereby promote splicing. Overall, our data suggest that the docking site acts as an enhancer for long *L*Ca inclusion.

Similar RNA pairings harbored at the 3' variable region of other genes

Whether this competing RNA pairing system exists in the 3' variable region of other genes remains unknown. Given the similarity of the origin and evolutionary mechanisms, we focused on investigating RNA pairings in genes containing 3' tandem clusters, as previously described (Hatje and Kollmar 2013). As a result, we found that some genes have the potential to form RNA competing pairings. For example, similar arrangements of intronic elements and secondary structures have been discovered in the 3' variable region of *D. melanogaster* CG42235 (Fig. 7A,B; Supplemental Fig. S5). This gene contains five to seven tandemly arrayed cassettes with six to eight exons at the 3' variable region that can produce different 3' cassettes through highly complex alternative splicing in *Drosophila* species (Fig. 7A,B; Supplemental Fig. S6). In particular, the 3' splice site of exon 2 can only be spliced with the most proximal alternative 5' splice site and the three most distal alternative 5' splice sites, not with the two intermediate splice sites (Fig. 7A; Supplemental Fig. S7). In contrast, the 3' splice site of exon 3 can only be spliced with the two most proximal alternative 5' splice sites, not with the three most distal sites (Fig. 7A; Supplemental Fig. S7). Thus, a powerful mechanism must exist to ensure that the two alternative 5' splice sites are correctly spliced with the five alternative 3' splice sites. We reason that RNA pairing approximates the 3' splice site of exon 3 and the two most distal alternative 5' splice sites, thus precluding the

splicing of the 3' splice site of exon 2 with the intermediate alternative 5' splice site. To validate the predicted RNA pairings involved in 3' alternative cassette selection, we generated a minigene construct containing the first two 3' alternative cassettes and tested the effects of disrupting the RNA secondary structure on splicing (Fig. 7C). The reduced efficiency of cassette 2 inclusion caused by individual mutations (M1, M2) was restored by double compensatory mutations (M12, Fig. 7D,E). Combined, this evidence indicates that 3' cassette 2 inclusion is activated by intronic RNA pairing.

Furthermore, we discovered similar arrangements of intronic elements and secondary structures at the 3' variable region of *D. melanogaster* *Pip* (Supplemental Fig. S8A). These secondary structures were conserved across all *Drosophila* *Pips* investigated (Supplemental Figs. S9, S10). *Pip* genes contain 10 tandemly arrayed cassettes with two to four exons at the 3' variable region that can produce different 3' cassettes through alternative splicing in a tissue-differential manner (Supplemental Fig. S8B). Because of the strikingly large size (>12 kb in *D. melanogaster*) and complexity (10 copies) of the 3' cassette cluster, we generated a simplified construct to test the effects of RNA pairing on the selection of 3' alternative cassettes (Supplemental Fig. S8A). The reduced efficiency of 3' alternative cassette inclusion caused by individual mutations (M1, M2) was largely restored by compensatory double mutations (M12, Supplemental Fig. S8C–E). Overall, the data obtained from disruptive and compensatory mutations strongly suggest that RNA pairing is involved in the 3' alternative cassette selection of *Pip* pre-mRNA. Thus, we propose that competing RNA pairings play an important role in regulating the highly complex alternative splicing of the 3' variable region.

The origin of RNA pairing is driven by complex splicing regulation

Finally, we explored what selection force drives the formation of RNA pairing at the 3' variable region. Because the genes share similar tandem multiexon clusters in the 3' variable region (Figs. 1A, 7A; Supplemental Fig. S8A), we initially speculated that RNA pairing evolved specifically to ensure correct splicing of multiple clustered exons. However, we did not observe RNA pairing in other genes with tandem multiexon clusters at the 3' variable region. In contrast, we found that these genes contained evolutionarily conserved small introns within the 3' tandem multiexon clusters (Supplemental Fig. S11) that are necessary for proper alternative splicing through mutational experiments (Supplemental Fig. S12). Therefore, the origin of competing RNA pairings might not be driven merely by duplication of the 3' tandem multiexons but by other selective pressures. Therefore, we investigated how such RNA pairings originated and have evolved in the 3'

FIGURE 7. Competing RNA pairings were validated in *Drosophila* CG42235 pre-mRNA. (A) Overview of the 3' variable region of *Drosophila* CG42235. Symbols used are the same as those in Figure 1. Introns are not drawn to scale. This gene contains five tandemly arrayed cassettes with five to eight exons at the 3' variable region, which are alternatively spliced to the common exon 3. Based on the exon-exon junction reads from RNA-seq data, we calculated the splice isoforms. Value represents the number of exon-exon junctions/million reads. The 3' splice site of exon 2 can be spliced only to the most proximal alternative 5' splice site and the three most distal alternative 5' splice sites (blue line), but not to two intermediate alternative 5' splice sites (red line). In contrast, the 3' splice site of exon 3 can be spliced only to the two most proximal alternative 5' splice sites (blue line), but not to the three most distal alternative 5' splice sites (red line). (B) Arrangement of cis intronic elements for RNA pairing interactions. For predicted intronic RNA pairing interactions, see Supplemental Figure S5. (C) Schematic diagrams of minigene constructs used to assess the effects of RNA secondary structure on 3' alternative cassette selection. Mutations introduced into dsRNA are indicated on the left or right mutated sequences (M1, M2). Green arrow depicts activated inclusion of the proximal exon. (D) Effects of intronic RNA pairing on alternative cassette inclusion were validated by disruptive single mutations (M1, M2) and compensatory double mutations (M12). (E) Quantitation of the data in D. Data are expressed as the percentage mean $\pm$ SD. (*) $P < 0.05$, (**) $P < 0.01$. (F) Evolutionary shifts from the 3' constitutive site to the alternative one are concomitant with the creation and expansion of RNA pairing. The origin and expansion of competing RNA pairings at the 3' variable region during *Drosophila* evolution are shown. Cassette duplications are indicated by red squares and green circles, respectively.

variable region of *PGRP-LC*. We found two or three copies in *PGRP-LCs* from mosquitoes, and a single copy exists in other insect species investigated (Supplemental Fig. S2).

However, we were unable to find similar RNA pairing in *Anopheles PGRP-LCs*, which contain similar tandem multi-exon clusters at the 3' variable region (Christophides et al.

2002). These results suggest that RNA pairings are specific to *Drosophila* PGRP-LCs. Note that 5' alternative sites were present in *Drosophila* PGRP-LCs, whereas similar 5' alternative sites were absent in the orthologous exon of *Anopheles* PGRP-LCs (Supplemental Fig. S2). Given the correlation between the emergence of alternative splice sites and the occurrence of RNA pairing in *Drosophila* PGRP-LCs (Fig. 1A), we speculate that the origin of RNA pairing is driven by the appearance of an alternative 5' splice site.

Similar results were obtained in *Drosophila* CG42235. By integrating the genetic and molecular data from *Drosophila* species, we suggest a credible phylogeny of 3' architecture directed by intronic RNA pairing in CG42235 (Fig. 7F). This 3' architecture with a recent origin might be of special interest, as it may bear marks from events that occurred during its genesis. Detailed genomic structure and phylogenetic analyses revealed that the ancestral CG42235 gene underwent tandem duplication of the 3' cassette before the divergence of the *obscura* and *melanogaster* groups, ~20 million years ago (Fig. 7F). Moreover, it is intriguing to note that another round of tandem duplication of the 3' cassette 2 occurred specifically in *D. bipectinata*, and the selector sequence required to control selection of the duplicated 3' cassettes was duplicated simultaneously (Fig. 7F; Supplemental Fig. S6). Therefore, the evolutionary shift from the 3' constitutive site to an alternative one that resulted from tandem cassette duplication was concomitant with the creation and expansion of RNA pairing. The emergence of a 3' alternative splice site might require RNA pairing to relieve the splicing regulatory conflict. Taken together, these findings suggest that the origin of RNA pairing is driven by complex splicing regulation.

DISCUSSION

Competing RNA pairings in alternative splicing at the 3' variable region

We revealed the multiple roles of competing RNA pairings in regulating highly complex alternative splicing at the 3' variable region. Based on the combined experimental evidence, a model can be proposed to explain how alternative splicing of the PGRP-LC 3' domain is regulated (Fig. 8). For an alternative cassette (LCx, LCy, and LCa) to be included in PGRP-LC mRNA, the selector sequence (CE1 and CE2) upstream of the target

exon must interact with the docking site (RE). When CE1 pairs with the docking site (RE), this RNA pairing interaction activates LCx splicing. In this case, because the long variant LCx is inhibited through suppression of the proximal 5' splice site by RNA structure, the short LCx cassette is dominant in the mRNA. Likewise, when CE2 pairs with the docking site (RE), the LCy cassette is activated. Conversely, long LCa will be included if the docking site assumes a linear conformation without specific RNA pairing interactions and acts as an enhancer (middle). In this case, the core motif of the docking site might recruit activators to promote splicing of the 5' splice site immediately downstream, thereby favoring long LCa inclusion. Our findings elucidate the molecular mechanism driving the biased splicing of the PGRP-LC alternative 3' domain and also provide a remarkable example of dynamic RNA pairing with multifaceted roles in alternative splicing.

Implications for complex 3'-end formation

Regulation of alternative splicing at the 3' variable region is intimately connected to 3'-end polyadenylation. It is worth considering how alternative splicing and polyadenylation can occur at the highly complex 3' variable region on transcription, particularly when complicated by long-range RNA pairing interactions. First, alternative transcription termination might favor the inclusion of the first 3' cassettes. For example, PGRP-LCx 3'-end formation could be

FIGURE 8. Model of the regulation of 3' alternative splicing of PGRP-LC pre-mRNA. For an alternative cassette (LCx, LCy, and LCa) to be included in the mRNA, the selector sequence (CE1 and CE2) upstream of the target exon interacts with the docking site (RE). When CE1 pairs with the docking site (left), such RNA pairing interaction functions to activate LCx splicing. In such cases, LCx is included in the mRNA because the remaining 3' cassette LCy and LCa might be cleaved prior to polyadenylation. In this case, long variant PGRP-LCx is inhibited by suppressing the proximal 5' splice site via the RNA structure, which also functions to sequester splicing activators away from the pre-mRNA targets. Likewise, when CE2 pairs with the docking site (RE), 3' cassette LCy is activated (middle). Conversely, long LCa would be included if the docking site assumes a linear conformation without specific RNA pairing interactions and acts as an enhancer (right). In this case, the RE core motif could bind to activators to promote splicing of its immediately downstream 5' splice site, thereby favoring long LCa inclusion.

completed even if transcription is terminated before the LCy 3' cassette. Thus, pre-mRNA terminated at the proximal site includes only the LCx 3' cassette. By analogy, LCx and LCy 3'-end formation could be completed even if transcription is alternatively terminated prior to LCa. Second, the pairing interaction between the docking site and selector sequence depends at least partially on the rate of transcription. Slower transcription favors sequential pairing between the docking site and the proximal selector sequence. Conversely, faster transcription permits more base-pairing between the docking site and the distal selector sequence. For example, based on the transcription rate of 1.1–1.4 kb/min in *Drosophila* (Thummel et al. 1990; Yao et al. 2007), the transcription interval between 3' cassette LCx and LCy is ~1 min. This interval means that the docking site may have sufficient time to pair with the initial selective sequences transcribed before transcription of later selective sequences. Thus, late-transcribed selective sequences have lower chances of pairing with the docking site; therefore, the associated 3' cassette should be selected less frequently. However, with a fast transcription rate of 50 kb/min typical of mammalian cells, the transcription interval between the 3' cassette LCx and LCy is only 0.02 min. Thus, pairing of the docking site with previously transcribed selective sequences before later selective sequences are transcribed is almost impossible. To what extent the transcription rate contributes to the 10-fold greater frequency of LCx over LCy remains unknown. Moreover, although the transcription interval between the first and last 3' cassette variants of *Pip* pre-mRNA is strikingly large (12,474 to 17,572 nucleotides [nt] in *Drosophila*), the 3' cassettes transcribed early are not chosen more frequently than those transcribed later (Supplemental Fig. S8B). Such a regulation pattern also argues against the classic cotranscriptional splicing mechanism (Maniatis and Reed 2002; Proudfoot et al. 2002; Kornblihtt 2007). Finally, factors such as RNA binding proteins may operate in conjunction with other players to modulate 3' cassette outcome (Olson et al. 2007; Fu and Ares 2014). Complex local splicing variation usually influences alternative 3' processing selection (Vaquero-Garcia et al. 2016). Collectively, these observations suggest that alternative transcription termination, transcription rate, and kinetic parameters influence RNA pairing-mediated selection of alternative 3' cassettes.

How common do competing RNA pairings act at the 3' variable region?

Many studies have revealed that competing RNA pairings between the docking site and selector sequences play a central role in mutually exclusive splicing of *Drosophila* *Dscam*, *14-3-3ξ*, *Myosin heavy chain*, *srp*, *RIC-3*, and human *dynamitin 1* and *CD55* (Graveley 2005; Anastassiou et al. 2006; May et al. 2011; Yang et al. 2011; Suyama

2013; Hatje et al. 2017; Jin et al. 2018). In contrast to internal mutually exclusive splicing, a classic scanning model can explain the selection of one 3' multiple tandem cluster at a time, where a common 5' splice site can be alternatively spliced to more than one 3' splice site (Black 2003; Matlin et al. 2005; Nilsen and Graveley 2010; Roca et al. 2013). In addition, proper selection of 3' variable exons can be achieved through the arrangement of intron size and/or splice sites (Supplemental Figs. S11, S12). Therefore, competing RNA pairings might be less common and conserved at the mutually exclusive 3' variable region than internal mutually exclusive exons.

However, some genes that contain complex 3' variable regions might require evolution of the RNA pairings to relieve splicing regulatory conflicts. For example, *Drosophila*-specific competing RNA pairings are concomitant with the emergence of an alternative 5' splice site (Fig. 1A; Supplemental Fig. S2). Therefore, the origin and biased selection of an alternative 5' splice site might drive the origin of competing RNA pairings in *Drosophila*. Similarly, RNA pairing acts specifically in CG42235 of the *D. melanogaster* subgroup, concomitant with the recent occurrence of subcassette duplication (Fig. 7F). These complex gene structures might require evolution of competing RNA pairings to ensure hierarchical splicing such that the internal 3' splice site within the duplicate cassette is not spliced to the external 5' splice site (Fig. 7A). Therefore, at least some genes may have evolved competing RNA pairings to relieve alternative splicing regulatory conflicts in highly complex 3' variable regions. Moreover, the arrangement and interaction of *cis*-elements would be advantageous to the evolutionary expansion of the cluster (Fig. 7F). Finally, because alternative splicing of the 3' variable region is coupled with cleavage, polyadenylation, and other events (Elkon et al. 2013; Bentley 2014), more protein regulators or complexes must be harbored. Thus, the involvement of the structural elements and dynamic interactions may simplify the regulatory proteins. This structural code increases the regulatory efficiency and flexibility of splicing-coupled multiple 3'-end formation.

MATERIALS AND METHODS

Materials

The insects and other species used in this study are presented in Supplemental Table S1. Fruit flies (*D. melanogaster*) were obtained as reported previously (Yang et al. 2011).

Annotation of *PGRP-LC*, *CG42235*, and *Pip* orthologs

Genomic DNA sequences and the corresponding protein sequences of homologs from other insects were obtained by

BLAST searches using the *D. melanogaster* homolog sequence as bait. Sequences from *Drosophila*, mosquito species (*A. gambiae*, *A. aegypti*, and *C. pipiens*), and other insect species were obtained from FlyBase (<http://flybase.org>) or NCBI (<https://www.ncbi.nlm.nih.gov/>). Potential alternative splice sites were verified either by expressed sequence tag and RNA-seq evidence or RT-PCR analysis.

Cloning and sequencing

Whole genomic DNA was isolated using the QIAamp DNA Kit (Qiagen). The primers used in this study are listed in Supplemental Table S2. The PCR parameters were denaturing at 95°C for 3 min, 35 cycles of denaturing at 95°C for 45 sec, annealing at 55°C for 50 sec, and extension at 72°C for 2 min and 10 sec, followed by extension at 72°C for 10 min. Amplification products were cloned into the pGEM-T Easy vector (Promega) for sequencing.

Sequence alignments and RNA pairing predictions

The conserved regions among species were aligned using ClustalW2 software (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) (Larkin et al. 2007). The genetic distances from pairwise mutually exclusive exons for each gene were estimated using MEGA 3.0 (Kumar et al. 2004). The phylogenetic tree was generated using the neighbor-joining method in the Clustal X 1.81 package. RNA pairings were predicted using Mfold software (Zuker 2003). The RNA folding temperature was fixed at 37°C, and the ionic conditions were 1 M NaCl without divalent ions. In some cases, the full sequences flanking the exon cluster were used as input for Mfold. In other cases, we only used the docking site, the selector, and their flanking sequences due to the limitations of the Mfold program. The 5' and 3' splice-site-motif scores were calculated using the splice site predictor (range of 0 to 1, with higher values predicting stronger splice sites) (Reese et al. 1997). Protein 3D structure modeling was acquired by Swiss-Model (automated mode) (www.swissmodel.expasy.org). The structures were displayed and processed using PyMOL (www.pymol.com). PGRP-LC-pa (FlyBase ID: FBpp0088491) modeling was based on template 2f2IX (PDB ID). The strength of the splice site is presented with the scores shown above.

RT-PCR

Total RNA was isolated from different developmental stages and tissue samples using the RNeasy Mini Kit (Qiagen), and isoform-specific primer pairs for each transcript were designed. Total RNA was reverse-transcribed using SuperScript III RTase (Invitrogen) with a gene-specific primer or an oligo(dT)₁₅ primer, and the resulting single-stranded cDNA product was treated with DNase I at 37°C for 30 min. Specific primers flanking the alternative exons were designed (Supplemental Table S2). PCR was performed as follows: denaturing at 95°C for 3 min, 35 cycles of denaturing at 95°C for 45 sec, annealing at 60°C for 50 sec, and extension at 72°C for 2 min and 10 sec, followed by extension at 72°C for 10 min. The PCR and RT-PCR products were purified

and cloned into the pGEM-T Easy vector (Promega) and then transformed into TG1 competent cells.

Quantification of mRNA splice isoforms

We determined the RNA splice isoform ratio using RT-PCR followed by exon-specific restriction digestion. Total RNA was isolated from different developmental stages and tissue samples. RT-PCR was carried out under the following cycling conditions: initial denaturation at 94°C for 2 min followed by 30–35 cycles of denaturation at 94°C for 30 sec, annealing at 58°C–63°C for 30 sec, and extension at 72°C for 15 sec, with a final extension at 72°C for 10 min. PCR products were then digested by exon-specific restriction enzymes. Images were captured using a CCD camera, and the mRNA isoforms were quantified by comparison of the integrated optical densities of the detected bands measured using the GIS 1D Gel Image System, v. 3.73 (Tanon). Data are expressed as the percentage mean $\pm$ standard deviation. The error bars presented were calculated from the average of three independent experiments.

Analysis of RNA-seq data

To validate the exon–exon junctions, we selected publicly available RNA-seq data sets corresponding to various developmental stages, tissues, and organs of *D. melanogaster* (Graveley et al. 2011; Brown et al. 2014). For each sequencing data set, the count of each RNA-seq read was normalized to reads per million, thus enabling cross-sample comparison of the relative expression levels. An in-house computational program (Yue et al. 2016a) was used to search for sequencing reads supporting the exon–exon junctions. First, we generated exon sequences covering all possible exon junctions at the 5' variable region. We used 20 overlapping nucleotides from each exon in a pair to assign a given read to an exon junction. For example, the 160-nt exon sequences were composed of 80 nt upstream and 80 nt downstream from the exon junction for 100-nt RNA-seq reads. RNA-seq reads from an individual pool were then mapped onto the exon sequences, and the completely mapped reads covering exon–exon junctions are shown. These mappings enabled us to calculate the levels of exon–exon junctions per million mapped reads of the transcript.

Minigene construction, mutagenesis, and transfection

Minigene construction and transfection were carried out as reported previously (Yang et al. 2011). Genomic DNA isolated from *D. melanogaster* was used as the template to amplify the corresponding DNA segments by PCR. WT minigene DNA was cloned into the pGEM-T Easy vector (Promega). Disrupting and compensatory mutageneses were performed to restore RNA secondary structure based on the schematic diagrams of the minigene constructs (Figs. 2B, 7C; Supplemental Fig. S8C). All constructs were confirmed by sequencing. WT and mutated fragments were inserted into the pMT/V5-HisB vector (Invitrogen), and linearized DNAs were transfected into *Drosophila* S2 cells using Lipofectamine 2000 (Invitrogen) as described by the

manufacturer. The alternative splicing outcomes were detected by RT-PCR and exon-specific restriction enzyme analysis.

RNA affinity purification and mass spectrometry

Two RNAs (PGRP-LC-WT and PGRP-LC-Mu) were labeled with biotin at both ends for RNA pull-down analysis. *Drosophila* S2 cells were lysed with lysis buffer and centrifuged at 12,000 rpm. The supernatant was collected, and 200 μ L Pierce streptavidin agarose (Thermo Fisher Scientific) was added. The supernatant was then evenly divided into three tubes. PGRP-LC-WT and PGRP-LC-Mu were added separately to two tubes, and the other served as a control. Streptavidin agarose (100 μ L) was then added to each tube and incubated on ice for 2 h, followed by centrifugation at 12,000 rpm. The streptavidin agarose was collected and rinsed three times. After SDS-PAGE, the gel was stained with Coomassie Brilliant Blue, and specific proteins were analyzed by mass spectrometry (Supplemental Fig. S4).

RNA interference

Candidate proteins identified by mass spectrometry were evaluated by RNA interference according to a previous description (Park et al. 2004). Twenty micrograms of each dsRNA was added to 0.5×10^6 serum-free adapted Schneider S2 cells (Invitrogen). After 2 d, an additional 20 μ g dsRNA was added to each well, and the cells were incubated for 2 additional days. After a total of 4 d, total RNA was isolated using TRIzol (Invitrogen) and subjected to reverse transcription using Superscript II (Invitrogen) according to the manufacturer's instructions. Following treatment with dsRNA against each splicing protein, the efficiency of RNA interference was monitored by semiquantitative RT-PCR and compared with the results obtained from untreated cells. In vitro-transcribed RNAs of pasilla were added as an internal control. Specific primers used for RNA interference are shown in Supplemental Tables S3 and S4.

Gel shift assays

Candidate proteins were analyzed by electrophoretic mobility shift assay in vitro using the LightShift Chemiluminescent RNA EMSA Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. RNAs were incubated with the purified protein expressed in *Escherichia coli* in binding buffer at room temperature. After incubation for 15 min at room temperature, heparin was added to a final concentration of 0.5 mg/mL, and the reaction products were resolved on a 12% (w/v) (29:1) polyacrylamide gel. The reaction complex was electrophoresed on a 12% (w/v) (29:1) native polyacrylamide gel in 0.5 $\times$ TBE. The separated proteins and RNAs were transferred to a nylon membrane (Roche) in 20 $\times$ SSC. Following UV cross-linking, the membrane was visualized by chemiluminescence using the DIG Luminescent Detection Kit (Roche) according to the manufacturer's protocol.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was supported by research grants from the National Natural Science Foundation of China (31430050, 31630089, 91740104).

Received February 26, 2018; accepted July 24, 2018.

REFERENCES

- Anastassiou D, Liu H, Varadan V. 2006. Variable window binding for mutually exclusive alternative splicing. *Genome Biol* **7**: R2.
- Baralle FE, Giudice J. 2017. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* **18**: 437–451.
- Bentley DL. 2014. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15**: 163–175.
- Berkovits BD, Mayr C. 2015. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**: 363–367.
- Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**: 291–336.
- Blencowe BJ. 2006. Alternative splicing: new insights from global analyses. *Cell* **126**: 37–47.
- Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J, Park S, Suzuki AM, et al. 2014. Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512**: 393–399.
- Chen M, Manley JL. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**: 741–754.
- Choe KM, Werner T, Stöven S, Hultmark D, Anderson KV. 2002. Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in *Drosophila*. *Science* **296**: 359–362.
- Choe KM, Lee H, Anderson KV. 2005. *Drosophila* peptidoglycan recognition protein LC (PGRP-LC) acts as a signal-transducing innate immune receptor. *Proc Natl Acad Sci* **102**: 1122–1126.
- Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, Brey PT, Collins FH, Danielli A, Dimopoulos G, et al. 2002. Immunity-related genes and gene families in *Anopheles gambiae*. *Science* **298**: 159–165.
- Danielli A, Kafatos FC, Loukeris TG. 2003. Cloning and characterization of four *Anopheles gambiae* serpin isoforms, differentially induced in the midgut by *Plasmodium berghei* invasion. *J Biol Chem* **278**: 4184–4193.
- Elkon R, Ugalde AP, Agami R. 2013. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* **14**: 496–506.
- Fu XD, Ares M Jr. 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet* **15**: 689–701.
- Goeke S, Greene EA, Grant PK, Gates MA, Crouner D, Aigaki T, Giniger E. 2003. Alternative splicing of *lola* generates 19 transcription factors controlling axon guidance in *Drosophila*. *Nat Neurosci* **6**: 917–924.
- Graveley BR. 2005. Mutually exclusive splicing of the insect *Dscam* pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**: 65–73.
- Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, et al. 2011. The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**: 473–479.
- Harris N, Braiser DJ, Dickman DK, Fetter RD, Tong A, Davis GW. 2015. The innate immune receptor PGRP-LC controls presynaptic homeostatic plasticity. *Neuron* **88**: 1157–1164.
- Hatje K, Kollmar M. 2013. Expansion of the mutually exclusive spliced exome in *Drosophila*. *Nat Commun* **4**: 2460.

- Hatje K, Rahman RU, Vidal RO, Simm D, Hammesfahr B, Bansal V, Rajput A, Mickael ME, Sun T, Bonn S, et al. 2017. The landscape of human mutually exclusive splicing. *Mol Syst Biol* **13**: 959.
- Jin YF, Yang Y, Zhang P. 2011. New insights into RNA secondary structure in the alternative splicing of pre-mRNAs. *RNA Biol* **8**: 450–457.
- Jin Y, Dong H, Shi Y, Bian L. 2018. Mutually exclusive alternative splicing of pre-mRNAs. *Wiley Interdiscip Rev RNA* **9**: e1468.
- Kaneko T, Goldman WE, Mellroth P, Steiner H, Fukase K, Kusumoto S, Harley W, Fox A, Golenbock D, Silverman N. 2004. Monomeric and polymeric gram-negative peptidoglycan but not purified LPS stimulate the *Drosophila* IMD pathway. *Immunity* **20**: 637–649.
- Kang D, Liu G, Lundström A, Gelius E, Steiner H. 1998. A peptidoglycan recognition protein in innate immunity conserved from insects to humans. *Proc Natl Acad Sci* **95**: 10078–10082.
- Keren H, Lev-Maor G, Ast G. 2010. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* **11**: 345–355.
- Komblitt AR. 2007. Coupling transcription and alternative splicing. *Adv Exp Med Biol* **623**: 175–189.
- Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**: 150–163.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.
- Lee Y, Rio DC. 2015. Mechanisms and regulation of alternative pre-mRNA splicing. *Annu Rev Biochem* **84**: 291–323.
- Maniatis T, Reed R. 2002. An extensive network of coupling among gene expression machines. *Nature* **416**: 499–506.
- Matera AG, Wang Z. 2014. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* **15**: 108–121.
- Matlin AJ, Clark F, Smith CW. 2005. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* **6**: 386–398.
- May GE, Olson S, McManus CJ, Graveley BR. 2011. Competing RNA secondary structures are required for mutually exclusive splicing of the *Dscam* exon 6 cluster. *RNA* **17**: 222–229.
- McManus CJ, Graveley BR. 2011. RNA structure and the mechanisms of alternative splicing. *Curr Opin Genet Dev* **21**: 373–379.
- McManus CJ, Duff MO, Eipper-Mains J, Graveley BR. 2010. Global analysis of trans-splicing in *Drosophila*. *Proc Natl Acad Sci* **107**: 12975–12979.
- Neyen C, Runchel C, Schüpfer F, Meier P, Lemaître B. 2016. The regulatory isoform rGRP-LC induces immune resolution via endosomal degradation of receptors. *Nat Immunol* **17**: 1150–1158.
- Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**: 457–463.
- Olson S, Blanchette M, Park J, Savva Y, Yeo GW, Yeakley JM, Rio DC, Graveley BR. 2007. A regulator of *Dscam* mutually exclusive splicing fidelity. *Nat Struct Mol Biol* **14**: 1134–1140.
- Park JW, Parisky K, Celotto AM, Reenan RA, Graveley BR. 2004. Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc Natl Acad Sci* **101**: 15974–15979.
- Proudfoot NJ, Furger A, Dye MJ. 2002. Integrating mRNA processing with transcription. *Cell* **108**: 501–512.
- Rabani M, Pieper L, Chew GL, Schier AF. 2017. A massively parallel reporter assay of 3' UTR sequences identifies in vivo rules for mRNA degradation. *Mol Cell* **68**: 1083–1094.e5.
- Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *J Comput Biol* **4**: 311–323.
- Roca X, Krainer AR, Eperon IC. 2013. Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes Dev* **27**: 129–144.
- Royet J, Dziarski R. 2007. Peptidoglycan recognition proteins: pleiotropic sensors and effectors of antimicrobial defences. *Nat Rev Microbiol* **5**: 264–277.
- Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**: 1643–1647.
- Scamborova P, Wong A, Steitz JA. 2004. An intronic enhancer regulates splicing of the twintron of *Drosophila melanogaster prospero* pre-mRNA by two different spliceosomes. *Mol Cell Biol* **24**: 1855–1869.
- Schmidt RL, Trejo TR, Plummer TB, Platt JL, Tang AH. 2008. Infection-induced proteolysis of PGRP-LC controls the IMD activation and melanization cascades in *Drosophila*. *FASEB J* **22**: 918–929.
- Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. 2000. *Drosophila Dscam* is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**: 671–684.
- Suyama M. 2013. Mechanistic insights into mutually exclusive splicing in dynam 1. *Bioinformatics* **29**: 2084–2087.
- Thummel CS, Burtis KC, Hogness DS. 1990. Spatial and temporal patterns of E74 transcription during *Drosophila* development. *Cell* **61**: 101–111.
- Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **18**: 18–30.
- Vaquero-Garcia J, Barrera A, Gazzara MR, González-Vallinas J, Lahens NF, Hogenesch JB, Lynch KW, Barash Y. 2016. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife* **5**: e11752.
- Wang X, Li G, Yang Y, Wang W, Zhang W, Pan H, Zhang P, Yue Y, Lin H, Liu B, et al. 2012. An RNA architectural locus control region involved in *Dscam* mutually exclusive splicing. *Nat Commun* **3**: 1255.
- Werner T, Liu G, Kang D, Ekengren S, Steiner H, Hultmark D. 2000. A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster*. *Proc Natl Acad Sci* **97**: 13772–13777.
- Werner T, Borge-Renberg K, Mellroth P, Steiner H, Hultmark D. 2003. Functional diversity of the *Drosophila* PGRP-LC gene cluster in the response to lipopolysaccharide and peptidoglycan. *J Biol Chem* **278**: 26319–26322.
- Yang Y, Zhan L, Zhang W, Sun F, Wang W, Tian N, Bi J, Wang H, Shi D, Jiang Y, et al. 2011. RNA secondary structure in mutually exclusive splicing. *Nat Struct Mol Biol* **18**: 159–168.
- Yao J, Ardehali MB, Fecko CJ, Webb WW, Lis JT. 2007. Intranuclear distribution and local dynamics of RNA polymerase II during transcription activation. *Mol Cell* **28**: 978–990.
- Yatsenko AS, Marrone AK, Shcherbata HR. 2014. miRNA-based buffering of the cobblestone-lissencephaly-associated extracellular matrix receptor dystroglycan via its alternative 3'-UTR. *Nat Commun* **5**: 4906.
- Yue Y, Meng Y, Ma H, Hou S, Cao G, Hong W, Shi Y, Guo P, Liu B, Shi F, et al. 2016a. A large family of *Dscam* genes with tandemly arrayed 5' cassettes in *Chelicerata*. *Nat Commun* **7**: 11252.
- Yue Y, Yang Y, Dai L, Cao G, Chen R, Hong W, Liu B, Shi Y, Meng Y, Shi F, et al. 2016b. Long-range RNA pairings contribute to mutually exclusive splicing. *RNA* **22**: 96–110.
- Yue Y, Hou S, Wang X, Zhan L, Cao G, Li G, Shi Y, Zhang P, Hong W, Lin H, et al. 2017. Role and convergent evolution of competing RNA secondary structures in mutually exclusive splicing. *RNA Biol* **14**: 1399–1410.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

RNA

A PUBLICATION OF THE RNA SOCIETY

Competing RNA pairings in complex alternative splicing of a 3' variable region

Huawei Pan, Yang Shi, Shuo Chen, et al.

RNA 2018 24: 1466-1480 originally published online July 31, 2018

Access the most recent version at doi:[10.1261/rna.066225.118](https://doi.org/10.1261/rna.066225.118)

Supplemental Material

<http://rnajournal.cshlp.org/content/suppl/2018/07/31/rna.066225.118.DC1>

References

This article cites 62 articles, 15 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/24/11/1466.full.html#ref-list-1>

Creative Commons License

This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:

<http://rnajournal.cshlp.org/subscriptions>