

# Dynamic landscape of tandem 3' UTRs during zebrafish development

Yuxin Li,<sup>1</sup> Yu Sun,<sup>1</sup> Yonggui Fu,<sup>1</sup> Mengzhen Li, Guangrui Huang, Chenxu Zhang, Jiahui Liang, Shengfeng Huang, Gaoyang Shen, Shaochun Yuan, Liangfu Chen, Shangwu Chen, and Anlong Xu<sup>2</sup>

State Key Laboratory of Biocontrol, National Engineering Center of South China Sea for Marine Biotechnology, Guangdong Province Key Laboratory of Pharmaceutical Functional Genes, Department of Biochemistry, College of Life Sciences, Sun Yat-sen University, Higher Education Mega Center, Guangzhou, 510006, P.R. China

Tandem 3' untranslated regions (UTRs), produced by alternative polyadenylation (APA) in the terminal exon of a gene, could have critical roles in regulating gene networks. Here we profiled tandem poly(A) events on a genome-wide scale during the embryonic development of zebrafish (*Danio rerio*) using a recently developed SAPAS method. We showed that 43% of the expressed protein-coding genes have tandem 3' UTRs. The average 3' UTR length follows a V-shaped dynamic pattern during early embryogenesis, in which the 3' UTRs are first shortened at zygotic genome activation, and then quickly lengthened during gastrulation. Over 4000 genes are found to switch tandem APA sites, and the distinct functional roles of these genes are indicated by Gene Ontology analysis. Three families of *cis*-elements, including miR-430 seed, U-rich element, and canonical poly(A) signal, are enriched in 3' UTR-shortened/lengthened genes in a stage-specific manner, suggesting temporal regulation coordinated by APA and *trans*-acting factors. Our results highlight the regulatory role of tandem 3' UTR control in early embryogenesis and suggest that APA may represent a new epigenetic paradigm of physiological regulations.

[Supplemental material is available for this article.]

Embryonic development involves a series of complex but ordered cellular processes including cell proliferation, differentiation, and migration under robust and precise management by gene regulatory networks (Gilbert 2003). As a major regulatory region terminating a transcribed mRNA, the 3' UTR plays important roles in the transcriptional (Veraldi et al. 2001; Brown and Gilmartin 2003; Lutz 2008), post-transcriptional (Zhou and King 2004; Kloosterman and Plasterk 2006; Plasterk 2006; Zhao and Srivastava 2007; Stefani and Slack 2008), and translational (Kuersten and Goodwin 2003; de Moor et al. 2005) regulation of genes within the regulatory networks, often through the interactions of *cis*-elements and *trans*-acting factors. The process of alternative polyadenylation (APA) in the terminal exon enables the transcription of multiple mRNA isoforms with different 3' UTRs from a single gene (called tandem 3' UTRs) (Edwards-Gilbert et al. 1997; Lutz 2008; Tian and Graber 2011). Notably, shortening of tandem 3' UTRs has been found to associate with cell proliferation (Sandberg et al. 2008), while lengthening of tandem 3' UTRs has been observed during cell differentiation (Ji et al. 2009; Shepard et al. 2011). Based on integrative analysis of EST, SAGE, and microarray data, tandem 3' UTR lengthening has been observed during mouse embryogenesis (Ji et al. 2009). However, the specific genes directly regulated by APA during vertebrate development remain unknown.

Several high-throughput sequencing methods have been recently developed based on the second-generation sequencing platforms to deeply sample APA sites in a genome-wide fashion (Mangone et al. 2010; Fu et al. 2011; Jan et al. 2011; Shepard et al.

2011). Two of these methods have been applied to *Caenorhabditis elegans* along development (Mangone et al. 2010; Jan et al. 2011), in which widespread APA events were demonstrated, and a trend of 3' UTR shortening was observed from embryo to adult (Mangone et al. 2010). However, the functional consequences of 3' UTR dynamics on development have not been characterized. We previously proposed another high-throughput method, SAPAS (Fu et al. 2011). In this study, we used SAPAS to profile tandem poly(A) events throughout the embryonic development of zebrafish (*Danio rerio*), a well-studied model organism for developmental research, to address the functional role of tandem 3' UTR control during animal development.

## Results

### A comprehensive inventory of poly(A) sites and tandem 3' UTR genes

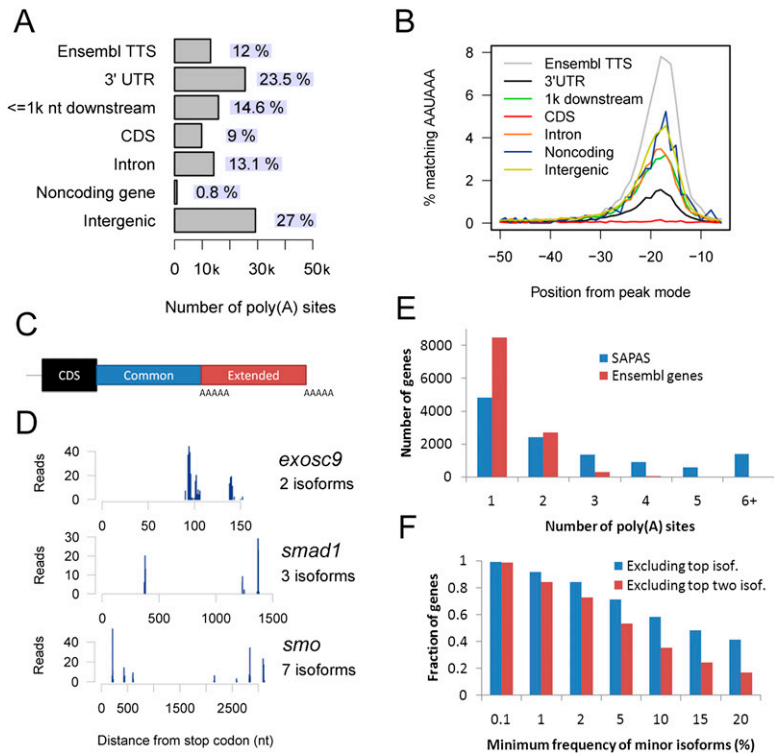
Eight time points across all major developmental stages were sampled with one lane of the Illumina Genome Analyzer 2 platform assigned for each time point (see Supplemental Fig. S1; Supplemental Table S1). Of the 59,294,885 reads obtained after mapping and filtering, nearly 90% were mapped to annotated 3' UTRs or 1-kb downstream regions (Supplemental Fig. S2). We pooled the data across all stages and focused on the 108,290 poly(A) sites with five or more normalized reads (Methods). Only 12% of these poly(A) sites were mapped to Ensembl transcription termination sites (TTS) (Fig. 1A) and another 1% to polyA\_DB2 (Lee et al. 2007), with the remaining 87% being unreported previously. To test the authenticity of the novel poly(A) sites, we performed 3'RACE on 30 novel poly(A) sites (five sites for each of the location categories shown in Fig. 1A, excluding Ensembl TTS) and 28 (93%) were validated (Supplemental Table S2). We also

<sup>1</sup>These authors contributed equally to this work.

<sup>2</sup>Corresponding author

E-mail lssxal@mail.sysu.edu.cn

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.128488.111>.



**Figure 1.** Summary of poly(A) sites and tandem 3' UTR genes. (A) Genomic locations of poly(A) sites. Only poly(A) sites with at least five normalized reads are considered. (B) Position-specific distributions of the canonical poly(A) signal hexamer AAUAAA for poly(A) sites from various genomic locations. (C) Schematics of tandem 3' UTRs. (D) Examples of tandem-UTR genes with two, three, and seven 3' UTR isoforms. (E) Number of genes with different numbers of tandem poly(A) sites and the comparison with Ensembl annotation. (F) Fraction of genes with the total frequency of minor 3' UTR isoforms exceeding given thresholds (0.1%, 1%, 2%, 5%, 10%, 15%, and 20%). Frequencies of minor isoforms are calculated as one minus the top isoform frequency for genes with 2+ poly(A) sites (blue bars), and as one minus the sum of the frequencies of the top two isoforms for genes with 3+ poly(A) sites (red bars).

compared our data with another RNA-seq data set (Aanes et al. 2011), and found that most (>70%) intergenic poly(A) sites were located within 5 kb downstream from RNA-seq reads (Supplemental Table S3), indicating the authenticity of these novel poly(A) sites.

The conventional mechanism of polyadenylation relies on two *cis*-elements in the neighborhood of a poly(A) site, i.e., the upstream poly(A) signal (PAS) hexamer AAUAAA/AUUAAA bound by CPSF and the downstream GU/U-rich element bound by CstF (Proudfoot 2011). These two signals are strongest for Ensembl-associated poly(A) sites, and are conspicuous for poly(A) sites from all of the other genomic regions except CDS (Fig. 1B; Supplemental Fig. S3). The CDS-associated poly(A) sites show very distinct features: almost no upstream canonical PAS AAUAAA/AUUAAA could be found (Supplemental Fig. S4) and there is relatively low U-content and higher preference for C/G-content surrounding the poly(A) site (Supplemental Fig. S3); a similar phenomenon is also observed in plants (Wu et al. 2011). These results suggest a potential novel mechanism of polyadenylation, which is independent of PAS and downstream GU/U-rich elements, in coding regions of genes of eukaryotes.

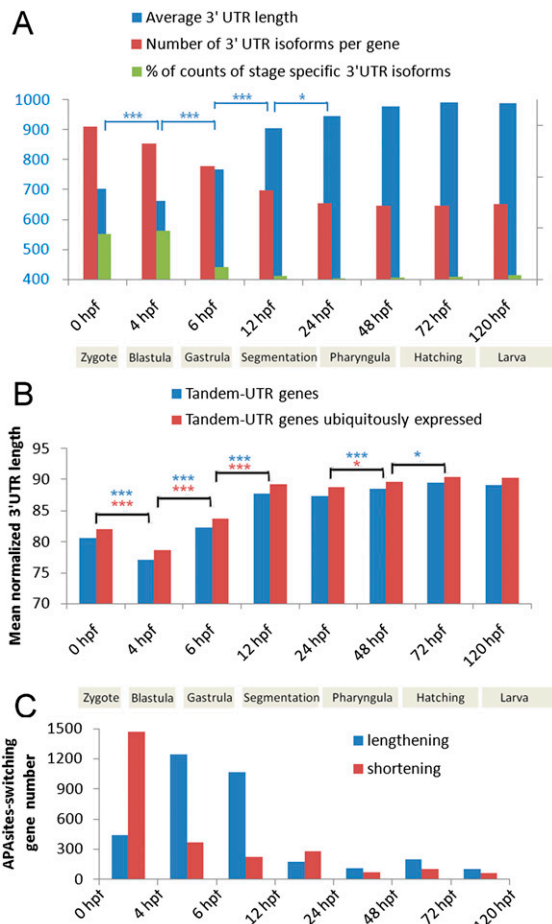
Based on this greatly expanded set of poly(A) sites, we defined tandem 3' UTRs as previously (Fu et al. 2011) and focused on the 11,505 genes with at least one SAPAS-sampled poly(A) site in the last exon. Up to 6660 (58% of the 11,505 genes, or 43% of all the expressed protein-coding genes) genes contained two or more

tandem poly(A) sites in our study (Fig. 1E; with three examples presented in Fig. 1D). Some genes with multiple 3' UTR isoforms display expression to a single isoform (the top isoform) and rarely express the other isoforms. However, this is not generally the case (Fig. 1F). By excluding the expression percentage of the top isoform, 58% of the genes with two or more poly(A) sites express the remaining isoforms as a minimum of 10% of the total expression. Even when excluding the top two isoforms [restricted to genes with three or more poly(A) sites], over half of the genes (53%) express the remaining isoforms at an appreciable level of 5% of total expression. These results not only reveal the complexity of the tandem UTR landscape during development, but also impose great challenges on data-processing methods, because systematic bias could be introduced if only two 3' UTR isoforms are considered per gene at a time (Fu et al. 2011).

### The dynamics of tandem 3' UTRs during development

We defined the 3' UTR length for a gene at a single time point as the average length of tandem 3' UTRs weighted by 3' UTR isoform expression levels (isoform-weighted 3' UTR length; Supplemental Data). Figure 2A shows that the average 3' UTR length varies dramatically during early development: first, the 3' UTR shortens from 0 hpf (zygote period) to 4 hpf (blastula period), a time point immediately following zygotic genome activation (ZGA; at ~2.75 hpf) (Kimmel et al. 1995), then significantly lengthens throughout the blastula and gastrula periods, and continues to lengthen for the duration of embryonic development. The dramatic early changes in 3' UTR length are accompanied by high 3' UTR diversity, which is the highest at the zygote period (0 hpf), with the presence of approximately three tandem poly(A) sites per gene and ~9% stage-specific 3' UTR isoforms (Fig. 2A). This diversity then rapidly decreases until the end of the segmentation period at 24 hpf, which displays ~1.5 3' UTR isoforms per gene and <1% stage-specific 3' UTR isoform.

To determine whether the average pattern of 3' UTR dynamics is due to differential usage of alternative 3' UTR isoforms, we separated genes with two or more tandem poly(A) sites (tandem-UTR genes) from genes with a single site (single-UTR genes). We found that the 3' UTR dynamics of single-UTR genes contribute to the lengthening after ZGA, but not the initial shortening (Supplemental Fig. S5). By comparison, the 3' UTR dynamics of tandem-UTR genes recapture the early 3' UTR shortening before ZGA as well as the lengthening afterward (Supplemental Fig. S5). To better present the impact of differential usage of tandem 3' UTR isoforms, we proposed a normalized 3' UTR length (NUL), which is defined as the percentage of isoform-weighted 3' UTR length relative to the longest 3' UTR length sampled in our data. Thus, by averaging the NUL across all genes at a given developmental stage,



**Figure 2.** Tandem 3' UTRs during development. (A) Characteristics of tandem 3' UTRs during development. (\*\*\*)  $P < 0.001$ , and (\*)  $P < 0.05$ , two-tailed permutation test. (hpf) Hours post-fertilization. (B) Mean normalized 3' UTR length (NUL) across developmental stages. (\*\*\*)  $P < 0.001$ , and (\*)  $P < 0.05$ , two-tailed permutation test. (C) Number of APAsites-switching genes between consecutive developmental time points.

any change in the average NUL between stages should represent differential usage of tandem 3' UTR isoforms. In this way, a much sharper V-shaped pattern of 3' UTR dynamics was observed during early development (Fig. 2B). In addition, significant 3' UTR-lengthening was observed during the pharyngula and hatching periods (Fig. 2B).

To identify genes with their transcripts characterized by significantly shortened or lengthened 3' UTRs between stages, we previously developed a statistical method of testing tandem APA sites switching, in which multiple poly(A) sites for one gene could be considered simultaneously (Fu et al. 2011). By adopting this method, we identified a total of 4374 genes with tandem APA sites switched (called APAsites-switching genes hereafter) during development (Fig. 2C; Supplemental Table S4; Supplemental Data). Notably, 1471 genes shortened and 441 genes lengthened the 3' UTRs from 0 to 4 hpf, stressing the extreme divergence of 3' UTR usage profiles between the maternal state (0 hpf) and the developmental stage immediately following ZGA (4 hpf). Consistent with high 3' UTR diversity in early development, most (3581; 82%) of the APAsites-switching genes displayed significantly altered

3' UTR lengths within 12 hpf. Based on the APAsites-switching genes, we performed hierarchical clustering on the NUL values (Supplemental Fig. S6). The most frequent dynamic 3' UTR pattern is consistent with a "narrow" V-shape, with the 3' UTR shortened from 0 to 4 hpf, lengthened from 4 to 6 hpf, and remaining stable afterward (Supplemental Fig. S7). This pattern was further confirmed by quantitative RT-PCR on selected genes (Supplemental Fig. S8; Supplemental Notes 1). These results highlight the complex landscape of tandem 3' UTR dynamics during early vertebrate development and suggest that the tandem 3' UTR profiles may be tightly regulated during early embryogenesis.

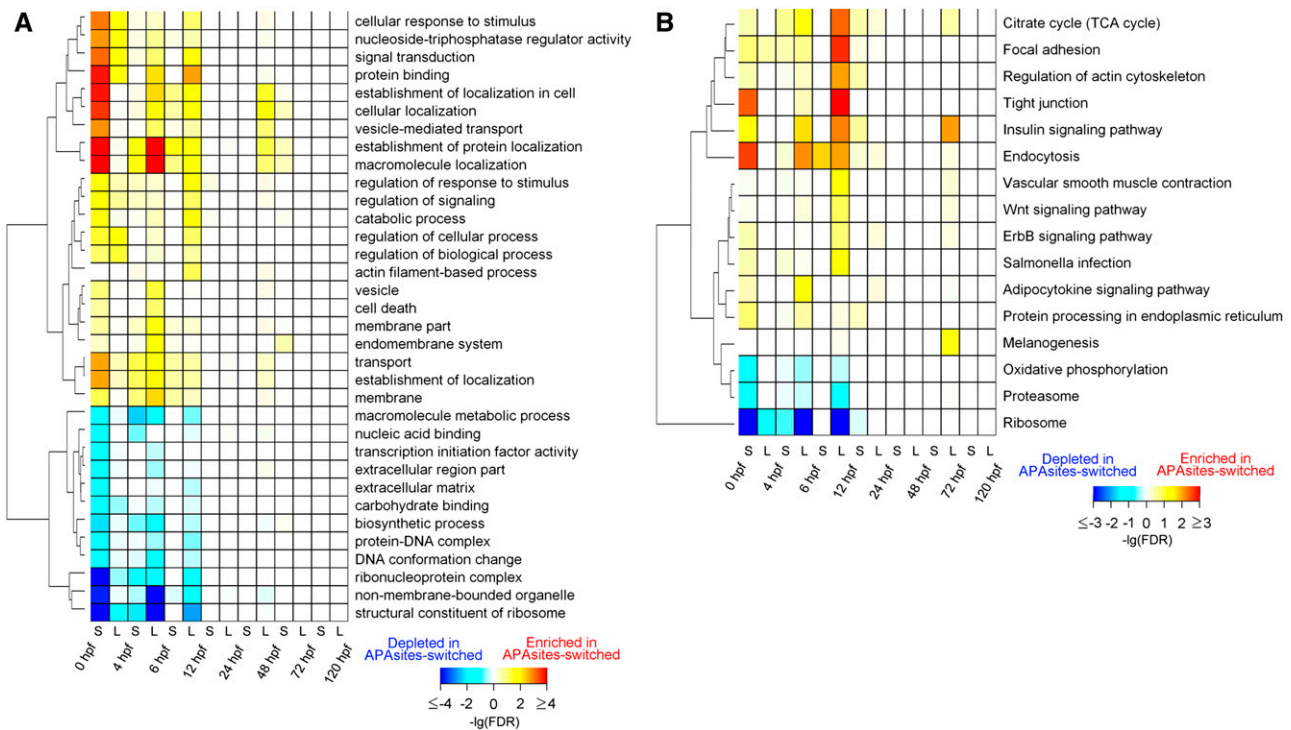
### Functional annotation analysis of APAsites-switching genes

We performed gene ontology (GO) enrichment/depletion analysis on APAsites-switching genes using single-UTR genes as the background (Fig. 3A; Supplemental Tables S5, S6) (False discovery rate: 10%, maximum  $P < 0.011$ ). GO terms related to protein localization, protein binding, the membrane/endomembrane system, signal transduction, and regulation of cellular and biological processes are significantly enriched across multiple developmental stages. However, these gene products are rarely found as the constituents of nonmembrane organelles (such as ribosomes) and seldom participate in biosynthetic processes. As for cellular localization, the products of these genes are enriched in intracellular processes and are rarely found in the extracellular region. These results imply that tandem 3' UTRs are tightly regulated for genes actively involved in intracellular processes, but not for genes involved in basic cell structure or routine biological processes.

To evaluate the roles of the regulations of tandem 3' UTRs in signaling cascades, we mapped the APAsites-switching genes to KEGG pathways (Fig. 3B; Supplemental Tables S7, S8) (FDR: 20%, maximum  $P < 0.013$ ). Four pathways, the citrate acid cycle (TCA cycle), endocytosis, insulin signaling, and tight junction, are significantly enriched in the APAsites-switching genes across early developmental stages. For stage-specific enrichment, two functionally related pathways, focal adhesion and regulation of actin cytoskeleton, are enriched in genes with 3' UTR lengthened during gastrulation (from 6 to 12 hpf). These two pathways are involved in the process of morphogenetic cell movement through interaction between the extracellular matrix and intracellular cytoskeleton during gastrulation (Bearer 1992). Interestingly, enrichment of genes with 3' UTR lengthened during gastrulation was also observed for Wnt and ErbB signaling, two pathways that are known to play pivotal roles in regulation of cell motility to achieve gastrulation (Nie and Chang 2007; Roszko et al. 2009). These results indicate a potential regulatory role of tandem 3' UTR regulation in cell localization during gastrulation.

### Developmental stage-specific *cis*-elements in tandem 3' UTRs

One potential mechanism of APA regulation of mRNAs involves the gain and loss of *cis*-elements in the 3' UTRs (Oh et al. 2000; Sandberg et al. 2008). In the case of studying genes with exactly two tandem poly(A) sites, the common 3' UTR is defined as the 3' UTR region between the stop codon and the proximal tandem poly(A) site, and the extended 3' UTR as the remainder of the 3' UTR, i.e., the 3' UTR region between the proximal poly(A) site and the distal site (see Fig. 1C; Sandberg et al. 2008). Comparisons of the common and extended 3' UTRs have demonstrated substantially different profiles of base composition (Ji et al. 2009). However, such comparisons have not been performed during specific

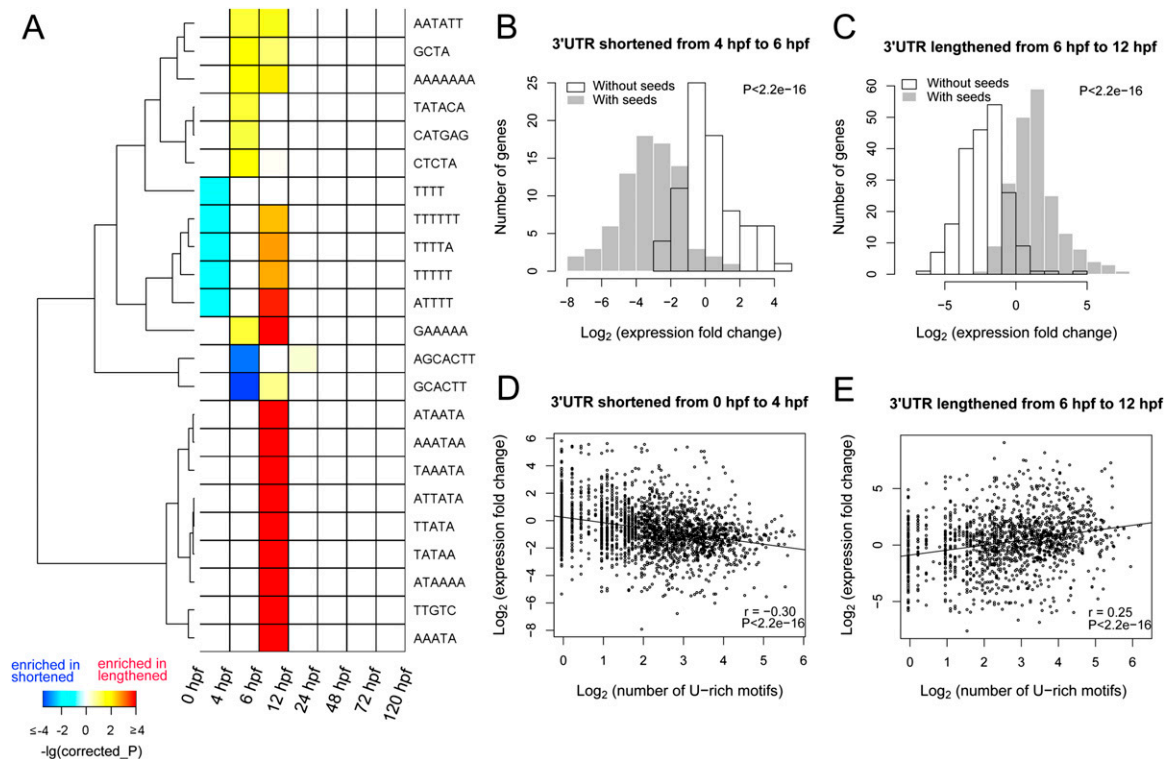


**Figure 3.** Functional annotations of APAsites-switching genes. (A) Gene ontology (GO) terms significantly ( $P < 0.011$ ; 10% false discovery rate) enriched and depleted in APAsites-switching genes by taking single-UTR genes as the background across developmental stages. GO terms enriched in APAsites-switching genes are assigned a positive  $P$ -value (hence, positive FDR in the BH sense), and depleted terms are assigned a negative  $P$ -value (hence, negative FDR). (S) Genes with 3' UTRs shortened; (L) genes with 3' UTRs lengthened. (B) KEGG terms significantly ( $P < 0.013$ ; 20% FDR) enriched and depleted in APAsites-switching gene. Data are presented as in A.

developmental stages or for the same 3' UTR region (either common or extended) between 3' UTR-lengthened and shortened genes. This study has provided an opportunity to perform such analysis in a developmental stage-specific manner. A major challenge, however, is the fact that most (3314 of 4374; 76%) APAsites-switching genes have three or more poly(A) sites, making it impossible to define common and extended 3' UTRs for these multipoly(A) genes using conventional definitions.

We proposed a generalized framework to study motif enrichment and depletion in genes with any number of tandem poly(A) sites. Conceptually, we defined the extended 3' UTR as the "altered" portion of the 3' UTR that differs between two developmental stages (Methods; see details in Supplemental Methods), and the common 3' UTR as the shared portion of the 3' UTRs, i.e., the 3' UTR corresponding to the most proximal poly(A) site. To test motif enrichment and depletion, we considered all 4- to 7-mers and compared the distributions of all of these k-mers between two sets of 3' UTR regions. First, we compared the common 3' UTR with the extended 3' UTR defined by two consecutive developmental time points (Supplemental Figs. S9, S10). Consistent with previous studies (Beaudoing et al. 2000), the canonical PAS, AAUAAA, is always enriched in the extended 3' UTRs compared with the common 3' UTRs. Unsurprisingly, many U-rich elements are enriched in the extended 3' UTR because its AU-content is higher than that of the common 3' UTR (Ji et al. 2009). However, no developmental stage-specific motifs were identified, suggesting that the differences in base composition between the common and the "altered" regions of the 3' UTRs are intrinsic features and are independent of developmental timing.

Next we focused on the extended 3' UTRs and compared the 3' UTR-lengthened genes with the 3' UTR-shortened genes (Fig. 4A). A number of stage-specific motifs were identified, from which three families of motifs could be discerned. The first family contains two motifs, GCACTT and AGCACTT, the core "seed" region of miR-430. The miR-430 seed is first enriched in the extended 3' UTR of genes with 3' UTRs shortened from 4 to 6 hpf, leading to the loss of miR-430 target sites for these genes. This enrichment then switches to the extended 3' UTRs of genes with 3' UTRs lengthened from 6 to 12 hpf, resulting in the gain of miR-430 sites. miR-430 is expressed as early as 4 hpf and functions first as a maternal mRNA cleaner (Giraldez et al. 2006) and then regulates brain morphogenesis (Giraldez et al. 2005). We argue that the tandem 3' UTR dynamics from 4 to 12 hpf could be associated with the dual functions of miR-430. From 4 to 6 hpf, of the 3' UTR-shortened genes with miR-430 seed located in the extended 3' UTRs, the expression levels of the 3' UTR isoforms with miR-430 seed were mostly down-regulated (Fig. 4B), which is in agreement with the role of miR-430 in mRNA degradation during this period (Giraldez et al. 2006; Supplemental Notes 2). In contrast, the mRNA levels of the 3' UTR isoforms without miR-430 seed were largely stable. Some (15%) of these isoforms lacking miR-430 seed were even up-regulated (twofold,  $\text{FDR} < 0.05$ ), suggesting that these isoforms were preferentially selected, potentially by APA, to escape miR-430-mediated degradation. This situation is reversed for genes with 3' UTRs lengthened from 6 to 12 hpf; that is, the isoforms lacking miR-430 seeds were mostly down-regulated, while the miR-430 seed-containing isoforms were largely up-regulated (Fig. 4C). The genes that gained a miR-430 seed during gastrulation were enriched



**Figure 4.** Developmental stage-specific motifs in extended 3' UTRs. (A) Enrichment of motifs in extended 3' UTRs by comparing 3' UTR-lengthened genes with 3' UTR-shortened genes between consecutive developmental time points. Motifs enriched in 3' UTR-lengthened genes are assigned a positive  $P$ -value (hence, positive corrected  $P$  in the Bonferroni sense), whereas motifs enriched in 3' UTR-shortened genes are given a negative  $P$ -value (hence, negative corrected  $P$ ). Only motifs with corrected  $P$ -values less than 0.1 and ranked in the top 10 for a specific stage are shown. (B) Most 3' UTR isoforms with the miR-430 seed GCACTT were down-regulated for genes with a shortened 3' UTR from 4 to 6 hpf. The distributions of expression levels of 3' UTR isoforms with and without miR-430 seed were compared by Kolmogorov-Smirnov test and the  $P$ -values are shown for both B and C. (C) Most 3' UTR isoforms with the miR-430 seed GCACTT were up-regulated for genes with the 3' UTR lengthened from 6 to 12 hpf. (D) Expression levels of 3' UTR isoforms were negatively correlated with the number of U-rich motifs (UUUUU) for genes with the 3' UTR shortened from 0 to 4 hpf. Pearson correlation coefficients ( $r$ ) are shown for both D and E. (E) Expression levels of 3' UTR isoforms were positively correlated with the number of U-rich motifs (UUUUU) for genes with the 3' UTR lengthened from 6 to 12 hpf.

in the GO terms "protein binding" and "signal transduction" (Supplemental Table S9), and included a number of neurogenesis-associated genes such as *neo1* (Shen et al. 2002), *smo* (Philipp et al. 2008), and *paflah1b1b* (Supplemental Fig. S11; Sun et al. 2009). More interestingly, the set of genes that lost a miR-430 seed from 4 to 6 hpf significantly overlapped with the set that gained a seed from 6 to 12 hpf ( $P < 2.2 \times 10^{-16}$ , Fisher's exact test; Supplemental Fig. S12). This resulted in 47 genes with miR-430 seed lost in the blastula and immediately regained in the gastrula period (with four examples presented in Supplemental Fig. S13). The miR-430 targets in two of these genes (*mam2* and *uba1*) were validated by luciferase reporter assays (Supplemental Fig. S14). These results suggest a cooperative mechanism between APA and microRNAs in the regulation of development.

The second family of motifs contains a number of U-rich elements represented by UUUUA and stretches of U, which resemble cytoplasmic polyadenylation elements (CPE) (Mendez and Richter 2001) or AU-rich elements (ARE) (Barreau et al. 2005). From 0 to 4 hpf, these U-rich motifs were enriched in the extended 3' UTRs of genes with the 3' UTRs shortened, and a negative correlation between the expression levels of 3' UTR isoforms and the number of these U-rich motifs was observed (Fig. 4D). Interestingly, the investigation of single-UTR genes also identified a significant enrichment of U-rich motifs in genes that are down-regulated during

this period (Supplemental Fig. S15). Thus, these U-rich motifs are associated with decreased mRNA abundance during the cleavage and early blastula periods. Similar to the case of miR-430, the enrichment of this family of U-rich motifs switched to genes with the 3' UTRs lengthened from 6 to 12 hpf. However, a positive correlation between the expression levels of the 3' UTR isoforms and the number of U-rich motifs was observed during this interval (Fig. 4E), suggesting that the 3' UTR isoforms with these U-rich elements are likely stabilized. High mRNA levels of the class-III ARE-binding protein *elavl1* (also known as HuR) were detected (Supplemental Fig. S16), suggesting that *elavl1* bound to these U-rich motifs and increased the stability of these mRNAs. These data imply a complex network potentially coordinated by APA, *cis*-elements in the 3' UTRs, and RNA-binding proteins throughout the process of early development.

The third family of motifs contained a number of A-rich sequences similar to the canonical PAS, AAUAAA, and were exclusively enriched in the extended 3' UTRs of genes with 3' UTRs lengthened from 6 to 12 hpf. Indeed, the AAUAAA motif was significantly enriched in the extended 3' UTRs of 3' UTR-lengthened genes relative to the extended 3' UTRs of 3' UTR-shortened genes exclusively in this period (corrected  $P = 3.3 \times 10^{-7}$ ). Moreover, while the canonical PAS was usually enriched in the extended 3' UTR relative to the common 3' UTR of genes, regardless of whether

their 3' UTRs were shortened or lengthened, we found that this enrichment was not significant for genes with 3' UTRs shortened in this period (corrected  $P = 0.086$ , Supplemental Fig. S9). However, this enrichment was extremely strong for 3' UTR-lengthened genes (corrected  $P = 2.9 \times 10^{-138}$ , Supplemental Fig. S10). These data suggest that poly(A) sites with canonical PAS are preferentially selected during this particular developmental stage (gastrulation).

## Discussion

In this study we have presented a global picture of tandem 3' UTR dynamics during zebrafish development, which is characterized by a dramatic V-shaped pattern in early embryogenesis (Fig. 2). Tandem 3' UTR lengthening is also observed during mouse embryogenesis (Ji et al. 2009), a finding that is consistent with the 3' UTR dynamics observed after ZGA (at 4 hpf) in zebrafish development, indicating functional conservation of tandem 3' UTR regulation in the control of vertebrate embryogenesis. The initial 3' UTR shortening could be explained in two ways: (1) maternal mRNA isoforms with long 3' UTRs are cleared; and (2) zygotic mRNA isoforms with short 3' UTRs are selectively transcribed at ZGA. Our data suggest that both of these mechanisms make significant contributions (Supplemental Notes 3), implying that tandem 3' UTR dynamics are shaped by both transcriptional (through APA) and post-transcriptional (mainly through *trans*-acting factors) regulations.

Our investigation of sequence features of common and extended 3' UTRs suggests a coordinated network that is potentially modulated through APA and *trans*-acting factors by controlling the gain and loss of *cis*-elements within 3' UTRs. This coordinated network is exemplified in three critical phases of early development: the oocyte-to-embryo transition, the maternal-to-zygotic transition, and gastrulation (Fig. 5). In the oocyte-to-embryo transition, the zygotic genome is transcriptionally repressed and embryonic development is activated and controlled by maternal factors located in the oocyte cytoplasm (Stitzel and Seydoux 2007). Hundreds (871 in our data) of translation-dormant maternal mRNAs with the CPE UUUUUAU (a motif significantly enriched in the extended 3' UTRs of 3' UTR-shortened genes from 0 to 4 hpf) (Fig. 4A) in their relatively long 3' UTRs are pre-stored in the oocyte (de Moor et al. 2005). After fertilization, these dormant mRNAs are re-adenylated and become translationally active through a process mediated by CPE (de Moor et al. 2005; Richter 2007). Next, during the maternal-to-zygotic transition, the zygotic genome is activated and maternal mRNA is cleared (Schier 2007). While maternal mRNAs

are likely to be targeted for degradation through *cis*-elements (such as U-rich elements or miR-430 target sites based on the analysis of our data) located in their relatively long 3' UTRs (Walser and Lipshitz 2011), newly transcribed mRNAs preferentially display short 3' UTRs, possibly to escape improper degradation mediated by *trans*-acting targeting.

When gastrulation begins, dramatic cell movements take place, and more precise control must be imposed on mRNA transcription and translation through *cis*-elements in the 3' UTR regulated by *trans*-acting factors (Stefani and Slack 2008). Correspondingly, mRNAs with longer 3' UTRs are transcribed in a coordinated fashion. Additionally, the lengthening of 3' UTRs during gastrulation may act as the preparation for organogenesis, a delicate process that is tightly controlled. A typical example of this situation is the development of the nervous system. As miR-430 is known to be indispensable in brain morphogenesis (Giraldez et al. 2005), hundreds (229 in our data) of genes lengthen their 3' UTRs to gain miR-430 target sites during gastrulation. Interestingly, the coordination between APA and miRNAs has also been addressed for some critical genes during *Drosophila* embryogenesis (Thomsen et al. 2010), suggesting a general regulatory mechanism in control of animal development cooperated by APA and miRNAs. Genes with the 3' UTRs lengthened in this period are also enriched in the Wnt and ErbB signaling, two pathways that are known to regulate neural tube formation (Birchmeier 2009; Ulloa and Marti 2010). These genes include major receptors (such as *fzd3a*) and critical effectors (such as *shc1* and *rock2a*; Supplemental Fig. S17), suggesting a key role of APA in control of these signaling pathways.

Taken together, our data are consistent with a two-layer model in which (1) APA regulates the usage of the 3' UTR and, thus, the *cis*-elements of mRNAs; (2) mRNAs are or are not under the control of *trans*-factors, depending on the presence or absence of the corresponding *cis*-elements. Together, these two factors, APA and *trans*-factors, orchestrate animal development by modulating mRNAs in a spatiotemporal fashion.

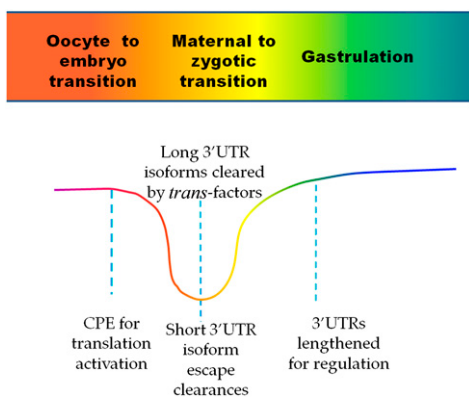
## Methods

### Collection of zebrafish embryos at different stages

Tubingen Zebrafish (*Danio rerio*) was provided by the North Center of National Zebrafish Resources of China. Embryos were obtained by natural mating and kept in Holt Buffer (NaCl 3.5 g, KCl 0.05 g, NaHCO<sub>3</sub> 0.025 g, CaCl<sub>2</sub> 0.1 g/1L) in 28°C water. Embryos were collected at eight time points (0, 4, 6, 12, 24, 48, 72, and 120 hpf), and washed twice in phosphate-buffered saline (PBS). A total of 1 mL of TRIzol was added per 100 embryos, swirled gently, and frozen at -80°C.

### SAPAS library construction

A sequencing library was constructed as described previously (Fu et al. 2011). Briefly, total RNA was extracted from zebrafish embryos by TRIzol, and ~10 µg of total RNA was randomly fragmented by heating. An anchored oligo d(T) primer and a 5' template switching linker tagged with Illumina adaptors were used in template switch reverse transcription (RT) by SuperScript II reverse transcriptase from Invitrogen. Two mutations in the poly(A) were introduced by PCR amplification with a determined number of cycles to ensure that the ds cDNA remain in the exponential phase of amplification. The PCR products were recovered after PAGE. The size-selection of 250–400 bp was performed by PAGE gel-excision. The recovery was quantified by a Qubit 2.0 Fluorometre, and the average



**Figure 5.** Schematic graph of potential biological functions of 3' UTR dynamics involved in early embryogenesis. See text for details.

size was determined by Agilent 2100 bioanalyzer. A quality control was performed by plasmid recombinant and Sanger sequencing. The recovery was ligated to pGEM-T Easy Vector and transformed into DH5a competent cells. Plasmid DNA was extracted and sequenced by ABI 3730 DNA Analyzer. Each end of the insertion sequence should be the Illumina sequence primer. The insertion sequence with long poly(A) stretch should be <5%, and most of the insertion sequence should be mapped to the zebrafish genome.

### 3'RACE and qRT-PCR

Embryos were recollected and the experiments were performed as described previously (Fu et al. 2011). Briefly, for 3'RACE, 30 novel poly(A) sites located in various genomic locations and with relative high-expression levels were chosen, and the PCR products were sequenced by ABI3730. For qRT-PCR, 17 genes with extreme 3' UTR length change at two developmental stages (0–4 and 4–6 hpf) were selected (details in Supplemental Notes 1). All of the primers were listed in Supplemental Table S2.

### Luciferase reporter assays

For plasmids construction, 3' UTRs with different lengths were cloned by PCR from zebrafish cDNA. To disrupt the miRNA complementary site, the nucleotides that paired to nucleotides 3 and 5 of the miRNA were substituted by mutant PCR. All wild-type and mutant 3' UTR isoforms were inserted into luciferase reporter vector psiCHECK-2 (Promega) between XhoI and NotI, downstream from the luciferase gene. All of the primers were listed in Supplemental Table S10. For luciferase reporter assays, HEK293T cells were seeded into each well of 6-well plates and incubated overnight, cotransfected with 3' UTR psiCHECK-2 luciferase reporter plasmid (100 ng/well) and miRNA (50 nmol/well) by Lipofectamine 2000 reagent (Invitrogen). After 24 h, cells were lysed and collected for luciferase reporter assay. Each experiment was performed at least in triplicate and repeated at least twice in all cases.

### Data analysis

A full description of the bioinformatics methods was given in the Supplemental Methods. Raw data were analyzed mainly as described previously (Fu et al. 2011). Briefly, raw reads were trimmed, filtered, and mapped to the zebrafish genome (Zv9) using Bowtie (version 0.12.5) (Langmead et al. 2009) and the 5' ends of uniquely mapped reads were clustered to define poly(A) sites. Total read counts from each stage were normalized to eliminate bias introduced by unequal read counts from different stages, and poly(A) sites with five or more normalized reads were used for further analysis. Poly(A) sites with reads that overlapped Ensembl-annotated 3' UTR(s) were defined as tandem poly(A) sites.

Tandem APA sites switching between stages was detected by a test of linear trend alternative to independence (Agresti 2002) as described previously (Fu et al. 2011). Normalized 3' UTR length (NUL) was defined as the percentage of isoform-weighted 3' UTR length relative to the longest 3' UTR length sampled in our data. Gene clustering was performed following the standard WGCNA procedure (Langfelder and Horvath 2008) for any pair of genes with evidence of APA sites switching.

For testing GO/KEGG item enrichment or depletion, a Fisher's exact test was performed to compare the proportion of APAsites-switching genes in one specific item with that proportion of single-UTR genes, and FDR was obtained in the Benjamini-Hochberg sense by using R. For testing motif enrichment or depletion, we considered all possible 4- to 7-mers and searched extended 3' UTR

regions of 3' UTR-lengthened genes against extended 3' UTR regions of 3' UTR-shortened genes. To count the occurrences of a particular k-mer word in the extended 3' UTR region for a gene with any number of tandem 3' UTR isoforms, we averaged the occurrences of this k-mer word for each 3' UTR isoform weighted by the isoform expression level change at that stage. *P*-values were obtained by Fisher's exact tests and Bonferroni corrected.

### Data access

The raw sequence data can be accessed from the NCBI Sequence Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>) under accession no. SRA036536.

### Acknowledgments

We thank Professor Bo Zhang from the North Center of National Zebrafish Resources of China for kindly providing the zebrafish. We also thank Professor Shuo Lin, Professor Xueqin Wang, and Dr. Yang Shen for fruitful discussions. This work was supported by the National Basic Research Program (no. 2007CB815800 to A.X.; no. 2011CB946101 to S.C.), the Key Project of the National Natural Science Foundation of China (no. 30730089 to A.X.), Fundamental Research Funds for the Central Universities (to Y.F.), and the China Postdoctoral Science Foundation (no. 2012M511618 to Y.S.).

### References

- Aanes H, Winata CL, Lin CH, Chen JP, Srinivasan KG, Lee SG, Lim AY, Hajan HS, Collas P, Bourque G, et al. 2011. Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res* **21**: 1328–1338.
- Agresti A. 2002. *Categorical data analysis*, 2nd ed. Wiley, Hoboken, NJ.
- Barreau C, Paillard L, Osborne HB. 2005. AU-rich elements and associated factors: Are there unifying principles? *Nucleic Acids Res* **33**: 7138–7150.
- Bearer E. 1992. *Cytoskeleton in development*. Academic Press, Waltham, MA.
- Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. 2000. Patterns of variant polyadenylation signal usage in human genes. *Genome Res* **10**: 1001–1010.
- Birchmeier C. 2009. ErbB receptors and the development of the nervous system. *Exp Cell Res* **315**: 611–618.
- Brown KM, Gilmartin GM. 2003. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor I<sub>m</sub>. *Mol Cell* **12**: 1467–1476.
- de Moor CH, Meijer H, Lissenden S. 2005. Mechanisms of translational control by the 3' UTR in development and differentiation. *Semin Cell Dev Biol* **16**: 49–58.
- Edwards-Gilbert G, Veraldi KL, Milcarek C. 1997. Alternative poly(A) site selection in complex transcription units: Means to an end? *Nucleic Acids Res* **25**: 2547–2561.
- Fu Y, Sun Y, Li Y, Li J, Rao X, Chen C, Xu A. 2011. Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res* **21**: 741–747.
- Gilbert SF. 2003. *Developmental Biology*. Sinauer Associates, Sunderland, MA.
- Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF. 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**: 833–838.
- Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**: 75–79.
- Jan CH, Friedman RC, Ruby JG, Bartel DP. 2011. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature* **469**: 97–101.
- Ji Z, Lee JY, Pan Z, Jiang B, Tian B. 2009. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci* **106**: 7028–7033.
- Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn* **203**: 253–310.
- Kloosterman WP, Plasterk RH. 2006. The diverse functions of microRNAs in animal development and disease. *Dev Cell* **11**: 441–450.
- Kuersten S, Goodwin EB. 2003. The power of the 3' UTR: Translational control and development. *Nat Rev Genet* **4**: 626–637.

- Langfelder P, Horvath S. 2008. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559. doi: 10.1186/1471-2105-9-559.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25. doi: 10.1186/gb-2009-20-3-r25.
- Lee JY, Yeh I, Park JY, Tian B. 2007. PolyA\_DB 2: mRNA polyadenylation sites in vertebrate genes. *Nucleic Acids Res* **35**: D165–D168.
- Lutz CS. 2008. Alternative polyadenylation: A twist on mRNA 3' end formation. *ACS Chem Biol* **3**: 609–617.
- Mangone M, Manoharan AP, Thierry-Mieg D, Thierry-Mieg J, Han T, Mackowiak SD, Mis E, Zegar C, Gutwein MR, Khivansara V, et al. 2010. The landscape of *C. elegans* 3'UTRs. *Science* **329**: 432–435.
- Mendez R, Richter JD. 2001. Translational control by CPEB: A means to the end. *Nat Rev Mol Cell Biol* **2**: 521–529.
- Nie S, Chang C. 2007. Regulation of *Xenopus* gastrulation by ErbB signaling. *Dev Biol* **303**: 93–107.
- Oh B, Hwang S, McLaughlin J, Solter D, Knowles BB. 2000. Timely translation during the mouse oocyte-to-embryo transition. *Development* **127**: 3795–3803.
- Philipp M, Fralish GB, Meloni AR, Chen W, MacInnes AW, Barak LS, Caron MG. 2008. Smoothed signaling in vertebrates is facilitated by a G protein-coupled receptor kinase. *Mol Biol Cell* **19**: 5478–5489.
- Plasterk RH. 2006. Micro RNAs in animal development. *Cell* **124**: 877–881.
- Proudfoot NJ. 2011. Ending the message: poly(A) signals then and now. *Genes Dev* **25**: 1770–1782.
- Richter JD. 2007. CPEB: A life in translation. *Trends Biochem Sci* **32**: 279–285.
- Roszko I, Sawada A, Solnica-Krezel L. 2009. Regulation of convergence and extension movements during vertebrate gastrulation by the Wnt/PCP pathway. *Semin Cell Dev Biol* **20**: 986–997.
- 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.
- Schier AF. 2007. The maternal-zygotic transition: Death and birth of RNAs. *Science* **316**: 406–407.
- Shen H, Illges H, Reuter A, Stuermer CA. 2002. Cloning, expression, and alternative splicing of neogenin1 in zebrafish. *Mech Dev* **118**: 219–223.
- Shepard PJ, Choi EA, Lu J, Flanagan LA, Hertel KJ, Shi Y. 2011. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* **17**: 761–772.
- Stefani G, Slack FJ. 2008. Small non-coding RNAs in animal development. *Nat Rev Mol Cell Biol* **9**: 219–230.
- Stitzel ML, Seydoux G. 2007. Regulation of the oocyte-to-zygote transition. *Science* **316**: 407–408.
- Sun C, Xu M, Xing Z, Wu Z, Li Y, Li T, Zhao M. 2009. Expression and function on embryonic development of lissencephaly-1 genes in zebrafish. *Acta Biochim Biophys Sin (Shanghai)* **41**: 677–688.
- Thomsen S, Azzam G, Kaschula R, Williams LS, Alonso CR. 2010. Developmental RNA processing of 3'UTRs in Hox mRNAs as a context-dependent mechanism modulating visibility to microRNAs. *Development* **137**: 2951–2960.
- Tian B, Graber JH. 2011. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA* **3**: 385–396.
- Ulloa F, Marti E. 2010. Wnt won the war: Antagonistic role of Wnt over Shh controls dorso-ventral patterning of the vertebrate neural tube. *Dev Dyn* **239**: 69–76.
- Veraldi KL, Arhin GK, Martincic K, Chung-Ganster LH, Wilusz J, Milcarek C. 2001. hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. *Mol Cell Biol* **21**: 1228–1238.
- Walser CB, Lipshitz HD. 2011. Transcript clearance during the maternal-to-zygotic transition. *Curr Opin Genet Dev* **21**: 431–443.
- Wu X, Liu M, Downie B, Liang C, Ji G, Li QQ, Hunt AG. 2011. Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation. *Proc Natl Acad Sci* **108**: 12533–12538.
- Zhao Y, Srivastava D. 2007. A developmental view of microRNA function. *Trends Biochem Sci* **32**: 189–197.
- Zhou Y, King ML. 2004. Sending RNAs into the future: RNA localization and germ cell fate. *IUBMB Life* **56**: 19–27.

Received July 1, 2011; accepted in revised form July 5, 2012.