Mini-review and Review

Hidden Peptides Encoded by Putative Noncoding RNAs

Akinobu Matsumoto* and Keiichi I. Nakayama

Department of Molecular and Cellular Biology, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Fukuoka 812-8582, Japan

ABSTRACT. **Although the definition of a noncoding RNA (ncRNA) is an RNA molecule that does not encode a protein, recent evidence has revealed that some ncRNAs are indeed translated to give rise to small polypeptides (usually containing fewer than 100 amino acids). Despite their small size, however, these peptides are often biologically relevant in that they are required for a variety of cellular processes. In this review, we summarize the production and functions of peptides that have been recently identified as translation products of putative ncRNAs.**

Key words: long noncoding RNA (lncRNA), circular RNA (circRNA), primary miRNA (pri-miRNA), translation, peptide

Introduction

Noncoding RNAs (ncRNAs) have been defined as a class of RNA molecules that are transcribed from genomic DNA but do not encode proteins. They include traditional RNA molecules such as transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) as well as other small RNAs such as micro-RNAs (miRNAs), Piwi-interacting RNAs (piRNAs), small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs), many of which play a variety of important biological roles (Pasut *et al.*, 2016).

Among ncRNAs, a family of RNA molecules that contain >200 nucleotides are categorized as long ncRNAs (lncRNAs), with the identification of such molecules having been facilitated by the development of high-throughput sequencing of cDNA clones (Carninci *et al.*, 2005;

---

*To whom correspondence should be addressed: Akinobu Matsumoto, Department of Molecular and Cellular Biology, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Fukuoka 812-8582, Japan.

Tel: +81–92–642–6816, Fax: +81–92–642–6819

E-mail: akinobu@bioreg.kyushu-u.ac.jp

Abbreviations: ncRNA, noncoding RNA; miRNA, microRNA; lncRNA, long noncoding RNA; ORF, open reading frame; circRNA, circular RNA; IRES, internal ribosome entry site; pri-miRNA, primary miRNA; RNAPII, RNA polymerase II; DCL1, Dicer-like 1; MS, mass spectrometry; MWCO, molecular weight cutoff filter; SDS-PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; Pgc, Polar granule component; CTD, COOH-terminal domain; P-TEFb, positive transcription elongation factor b; Pri, Polished rice; Svb, Shavenbaby; RNAi, RNA interference; Aplnr, Apelin receptor; MLN, myoregulin; PLN, phospholamban; SLN, sarcolipin; SERCA, sarco/endoplasmic reticulum Ca^{2+}-ATPase; SR, sarcoplasmic reticulum; ELN, endoregulin; ALN, another-regulin; SPAR, small regulatory polypeptide of amino acid response; v-ATPase, vacuolar-type H^+-ATPase; mTORC1, mammalian target of rapamycin complex 1; DMD, Duchenne muscular dystrophy.

Consortium *et al.*, 2007). Although lncRNAs do not harbor apparent open reading frames (ORFs) for proteins of >100 amino acids, recent studies have identified a subset of lncRNAs that actually encode polypeptides of <100 amino acids (Fig. 1A) (Anderson *et al.*, 2015; Kondo *et al.*, 2010; Matsumoto *et al.*, 2017; Nelson *et al.*, 2016; Pauli *et al.*, 2014).

Another newly identified class of "translated" ncRNAs comprises circular RNAs (circRNAs) that are localized mainly in the cytosol (Hansen *et al.*, 2013; Memczak *et al.*, 2013). These RNA molecules consist of only exon sequences and are produced by a back-splicing reaction that generates a covalent bond between the 3' end of an exon and the 5' end of an upstream exon (Fig. 1B). This reaction is promoted by the presence of sequences of reverse complementarity—such as human Alu repeats—in introns (Zhang *et al.*, 2014). Given that circRNAs lack a 5'-cap modification, cap-dependent translation cannot occur in principle. However, a synthetic circRNA containing an internal ribosome entry site (IRES) was shown to be translated in vitro (Chen and Sarnow, 1995), suggesting the possibility that endogenous circRNAs with an IRES might be translated. Indeed, recent studies have shown that endogenous peptides are produced by circRNAs in a cap-independent manner (Fig. 1B) (Legnini *et al.*, 2017; Pamudurti *et al.*, 2017). Some circRNAs have even been found to generate proteins containing >100 amino acids (Legnini *et al.*, 2017).

Primary miRNAs (pri-miRNAs), precursor ncRNAs that undergo enzymatic excision in the nucleus to yield mature miRNAs, have also been shown to encode peptides (Fig. 1C) (Ambros, 2004; Bartel, 2004; Lauressergues *et al.*, 2015). These RNA molecules are generated in both animals and plants by the same enzyme, RNA polymerase II (RNA-
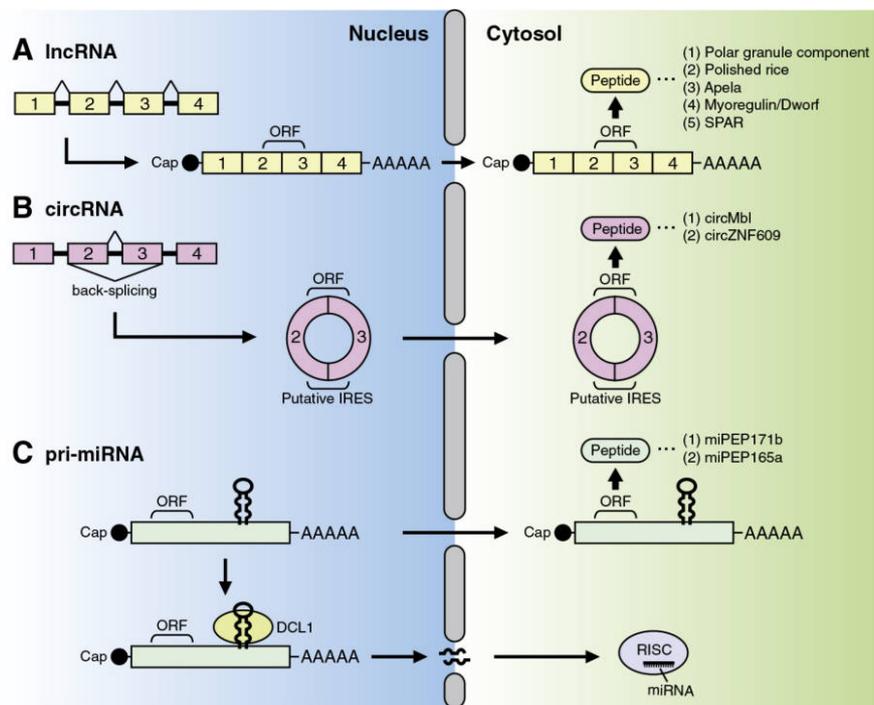
Fig. 1. Schematic models for the processing and translation of putative ncRNAs. (A) Peptides encoded by lncRNAs. Many mature lncRNAs are modified with a 5' cap and 3' poly(A) tail and are exported from the nucleus to the cytosol. Some of these molecules with small ORFs are taken up by ribosomes and translated. (B) Peptides encoded by circRNAs. A noncanonical form of splicing, known as back-splicing, is responsible for the generation of circRNAs, which do not contain a 5' cap or 3' poly(A) tail. These RNAs are exported to the cytosol and translated if they possess an IRES element. (C) Peptides encoded by pri-miRNAs. Plant pri-miRNAs have been shown to encode peptides. In both animals and plants, pri-miRNAs are synthesized by RNAPII and are modified with a 5' cap and 3' poly(A) tail. Most miRNA duplexes are processed by DCL1 and transported to the cytosol for formation of the RNA-induced silencing complex (RISC) that mediates suppression of target mRNAs. However, some pri-miRNAs containing small ORFs may be transported to the cytosol without processing and then translated.

PII), that mediates the synthesis of protein-coding mRNAs, and they are modified with a 5' cap and 3' polyadenylate tail. These properties suggest that pri-miRNAs may produce proteins if they are transported to the cytosol without processing (Waterhouse and Hellens, 2015). It is of note that pri-miR171b of *Medicago truncatula* and pri-miR165a of *Arabidopsis thaliana* were found to be translated to yield peptides that enhance the expression of the corresponding miRNAs (Fig. 1C). The mechanism by which these pri-miRNAs avoid processing by Dicer-like 1 (DCL1) remains unknown, however.

In this review, we focus on ncRNAs that do actually encode peptides—including lncRNAs, circRNAs, and pri-miRNAs—and we address the functions and biological relevance of these molecules.

Identification of hidden peptides

Most coding RNAs contain one long ORF and many short ORFs, with the long ORF typically encoding a functional protein (Fig. 2A). However, putative ncRNAs that encode hidden peptides contain only short ORFs that usually comprise <300 nucleotides, and the longest of these ORFs are frequently not responsible for encoding functional peptides, making it difficult to predict ORFs that do encode hidden peptides (Fig. 2A). Cross-species genomic comparisons are useful for the prediction of ORFs, given that the sequences of translated regions tend to be conserved under evolutionary pressure. PhyloCSF is a well-designed computational tool for ORF prediction that detects high frequencies of synonymous codon and conservative amino acid substitutions (Fig. 2B) (Lin *et al.*, 2011).

Experimental approaches, such as deep sequencing–based ribosome profiling and mass spectrometry (MS)–based peptidomics, have also been established for ORF identification. Ribosome profiling relies on deep sequencing to detect ribosome-protected RNA fragments, with such protection being indicative of active translation (Fig. 2C) (Ingolia, 2016; Ingolia *et al.*, 2009). In this approach, RNA molecules with bound ribosomes, known as polysomes, are purified by sucrose density gradient centrifugation or affinity pull-down of epitope-tagged ribosomes and are then treated with nucleases, resulting in the generation of
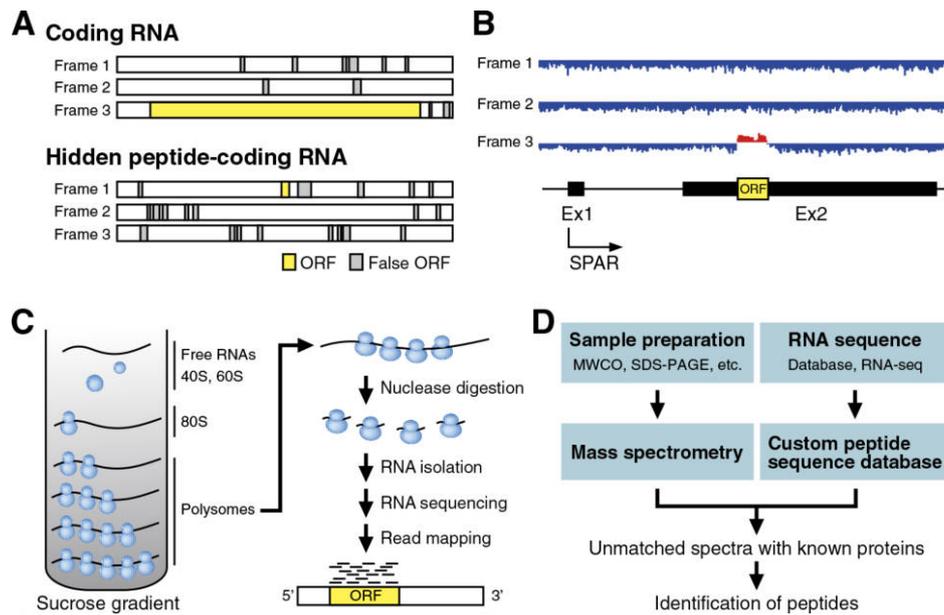
Fig. 2. Identification of hidden peptides. (A) ORF prediction. Whereas the longest ORFs encode functional proteins in typical coding RNAs, they do not necessarily encode functional peptides in hidden peptide–coding RNAs. (B) PhyloCSF computational analysis. The analysis for the SPAR polypeptide is shown, with the coding region yielding a positive score for frame 3. Ex, exon. (C) Ribosome profiling. The polysome fraction is purified by sucrose density gradient centrifugation and subjected to nuclease digestion, with the resultant ribosome-protected RNA fragments then being isolated, sequenced, and mapped to the genome. Polysomes can also be isolated by affinity pull-down of tagged ribosomes. (D) Peptidomics. Peptides are concentrated and then analyzed by MS, and the MS/MS spectra are compared with a custom peptide sequence database generated by prediction based on RNA sequences. Spectra that do not match known proteins are selected for identification of novel hidden peptides.

ribosome-protected RNA fragments. These fragments are sequenced and mapped to the genome in order to identify the region of active translation (Fig. 2C).

Whereas ribosome profiling provides indirect evidence of translation, MS is able to detect translation directly. Peptidomics is a modified form of proteomics that combines MS and computational prediction (Fig. 2D) (Ma *et al.*, 2016; Matsumoto *et al.*, 2017; Slavoff *et al.*, 2013). Peptides are usually concentrated with the use of a molecular weight cutoff filter (MWCO) or sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) in order to reduce the background signal due to larger annotated proteins. MS/MS spectra are compared with a custom peptide database comprising a list of all possible short ORFs predicted from cDNA sequences in databases and RNA-sequencing data sets. Matched peptides corresponding to known annotated proteins or peptides are discarded, and the remaining unique peptides are annotated as novel hidden peptides (Fig. 2D).

Long ncRNAs

(1) Polar granule component (Pgc)

In *Drosophila* embryos, the specialized cytoplasm in the posterior region of the egg, referred to as pole plasm, is essential for the formation of germ cells. Pole plasm contains electron-dense structures known as polar granules that become incorporated into the germ cell precursors known as pole cells. An RNA molecule, designated *pgc*, was found to be localized to polar granules and was originally identified as an untranslatable RNA on the basis of the absence of conserved ORFs for products of >100 amino acids (Nakamura *et al.*, 1996). However, completion of the *Drosophila* genomic sequence revealed an error in the original *pgc* sequence and thereby led to the identification of an ORF encoding a 71–amino acid polypeptide that was highly conserved among 12 *Drosophila* species (Hanyu-Nakamura *et al.*, 2008). Immunostaining with antibodies to the Pgc polypeptide revealed its localization in pole cells.

The repression of RNAPII-dependent transcription in primordial germ cells is a common mechanism for the maintenance of these cells through prevention of somatic differentiation in many animal species. This transcriptional repression is associated with the loss of phosphorylation of the COOH-terminal domain (CTD) of RNAPII. The Pgc polypeptide was found to interact with positive transcription elongation factor b (P-TEFb), which phosphorylates the CTD of RNAPII, and Pgc-deficient pole cells failed to prevent this phosphorylation reaction (Hanyu-Nakamura *et al.*, 2008; Martinho *et al.*, 2004). Although Pgc had no
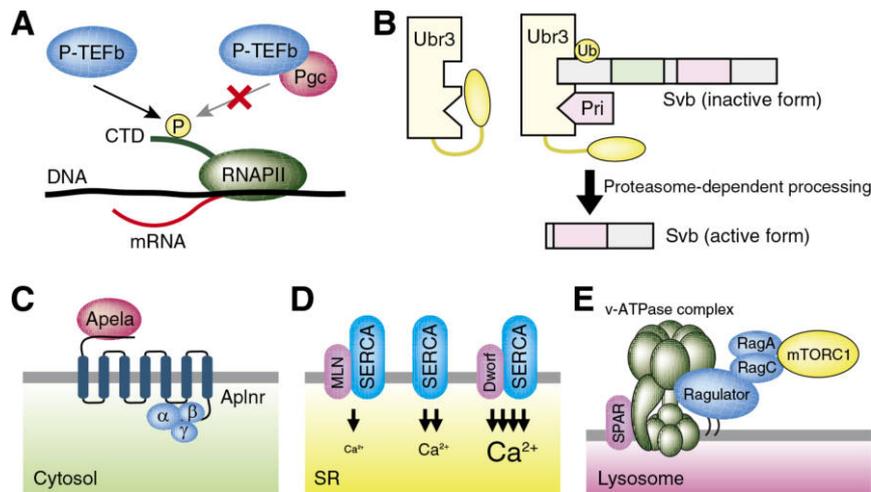
Fig. 3. Molecular mechanisms of action for peptides encoded by lncRNAs. (A) Pgc interacts with P-TEFb and thereby inhibits its recruitment to transcription sites, resulting in attenuation of CTD phosphorylation and RNAPII-dependent transcription. (B) The binding of Pri to Ubr3 promotes ubiquitylation of the NH_2-terminal region of Svb by Ubr3 and consequent proteasome-dependent processing of the long inactive form of Svb to the short active one. (C) Apela functions as an agonist for the G protein–coupled receptor Aplnr. (D) MLN and Dworf associate with and modulate the Ca^{2+} pump activity of SERCA, and they thereby regulate the Ca^{2+} concentration of the SR. (E) SPAR binds to the v-ATPase complex at the lysosomal membrane and inhibits amino acid–dependent mTORC1 activation through regulation of the v-ATPase–Ragulator supercomplex.

apparent effect on the kinase activity of P-TEFb, it was found to inhibit the recruitment of P-TEFb to transcription sites and thereby to prevent CTD phosphorylation (Fig. 3A).

(2) Polished rice (Pri)

An mRNA-like lncRNA, designated *pri*, was identified in *Drosophila melanogaster* as a poly(A)-containing RNA with no ORFs for products of >100 amino acids (Inagaki *et al.*, 2005). The *pri* RNA is expressed in seven stripes during early development and in metameric and bilateral clusters during mid-embryogenesis. Mutant flies deficient in *pri* RNA die during embryogenesis, with the embryos manifesting a smooth-cuticle phenotype that is characterized by loss of trichomes and aberrant tracheal architecture and was therefore named *polished rice (pri)*. Five small and highly conserved ORFs were detected in the *pri* RNA sequence, with ORFs 1 to 4 encoding products containing the amino acid sequence LDPTGQY or LDPTGTY. Translation of these four ORFs, but not that of ORF 5, was detected with the use of a series of constructs in which the sequence for green fluorescent protein was substituted for each ORF in the full-length transcript. ORFs 1 to 3 encode peptides of 11 amino acids, whereas ORF 4 encodes a peptide of 32 amino acids. The phenotype of *pri*-deficient mutant flies was completely rescued by introduction of ORFs 1 to 4 individually, suggesting that the encoded common amino acid sequence is responsible for *pri* function (Kondo *et al.*, 2007).

The peptides encoded by ORFs 1 to 4 of *pri* RNA were

shown to regulate the transcription factor encoded by *shavenbaby* (*svb*) (Kondo *et al.*, 2010), with expression of Svb target genes being greatly attenuated in *pri*-deficient mutants and mutation of *svb* resulting in an aberrant trichome pattern (Payre *et al.*, 1999; Sucena *et al.*, 2003; Zanet *et al.*, 2015). Indeed, Svb is translated as a long inactive form, and Pri peptides promote truncation of the NH_2-terminal region of Svb to generate the short active form of the protein (Kondo *et al.*, 2010).

A genome-wide RNA interference (RNAi) screen revealed that the ubiquitin-proteasome system likely contributes to the Pri-dependent processing of Svb. The NH_2-terminal region of Svb thus undergoes ubiquitylation mediated by Ubr3 and consequent proteasome-dependent processing that converts the long inactive form of Svb to the short active one. Pri binds to Ubr3, and this interaction promotes the association between Ubr3 and the NH_2-terminal region of Svb (Fig. 3B). Ubr3 is able to recognize other substrates in the absence of Pri, suggesting that Pri is required for the substrate selectivity of Ubr3 (Zanet *et al.*, 2015).

(3) Apela

Apela (also known as ELABELA, Ende, and Toddler) is a hormonal peptide that acts as an agonist for the G protein–coupled receptor Aplnr (Apelin receptor). The Apela transcript had been annotated as a lncRNA in zebrafish (ENSDARG00000094729), mouse (Gm10664), and human (LOC100506013), but was also found to be a peptide-coding mRNA in zebrafish (Chng *et al.*, 2013; Pauli *et al.*,

2014). Apela consists of 54 amino acids that include a signal peptide sequence and is highly conserved among vertebrates. Loss of Apela in zebrafish results in embryonic death, with the mutant embryos manifesting failure of heart formation, posterior accumulation of blood cells, malformation of pharyngeal endoderm, and abnormal left-right positioning and formation of the liver. The similarity of the phenotypes of Apela-deficient embryos to those of Aplnr mutants (Scott *et al.*, 2007; Zeng *et al.*, 2007) validates the function of Apela as the agonist for Aplnr (Fig. 3C).

About half of Apela-deficient mice also die during embryogenesis as a result of hypovascularity of the yolk sac and placenta as well as embryonic vascular malformations (Ho *et al.*, 2017), which are again similar to the phenotypes of Aplnr-deficient mice (Kang *et al.*, 2013). Although Aplnr is expressed in the yolk sac mesoderm that gives rise to endothelial cells, the Apela mRNA is not detectable in the endothelial precursors of the yolk sac. In contrast, Apela is expressed in the chorionic trophoblast of the developing placenta. The Apela peptide is secreted from the placenta and circulates in the blood of dams. Apela-deficient pregnant mice manifest preeclampsia-like signs including proteinuria and hypertension, indicating that Apela is a pregnancy hormone secreted by the developing placenta. Moreover, this preeclampsia-like condition is ameliorated by infusion of a recombinant Apela peptide, suggesting that Apela might have clinical applications in human pregnancy (Ho *et al.*, 2017).

(4) SERCA regulators

The myoregulin (MLN) transcript was originally identified as a skeletal muscle–specific lncRNA conserved among vertebrates (LINC00948 in human and 2310015B20Rik in mouse), but it possesses a conserved ORF for a 46–amino acid peptide (Anderson *et al.*, 2015). The translation of MLN RNA was confirmed by both in vitro translation and in vivo FLAG-tag knock-in assays. The predicted ORF product is a type II single-pass transmembrane peptide that shows substantial structural similarity to phospholamban (PLN) and sarcolipin (SLN), both of which are also type II single-pass transmembrane peptides. PLN and SLN interact with and thereby regulate the Ca^{2+} pump activity of sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) in the sarcoplasmic reticulum (SR) membrane. Such regulation is important for muscle performance, as indicated by the fact that its impairment can give rise to cardiovascular disease (Kranias and Hajjar, 2012). The MLN peptide is also localized at the SR membrane and associates with SERCA to regulate the Ca^{2+} concentration of SR through inhibition of SERCA pump activity (Fig. 3D). Mice deficient in MLN manifest enhanced Ca^{2+} handling and skeletal muscle performance (Anderson *et al.*, 2015).

The Dworf peptide also regulates SERCA pump activity and was originally identified as the putative product of a lncRNA (LOC100507537 in human and NON-MMUG026737 in mouse) (Nelson *et al.*, 2016). Immunoblot analysis with antibodies to the encoded peptide sequence confirmed production of Dworf, which was found to be expressed specifically in heart and soleus muscle. Dworf is a transmembrane peptide localized at the SR membrane and binds to SERCA and thereby promotes its pump activity (Fig. 3D). The role of Dworf in vivo was examined by the generation of Dworf-deficient mice and mice overexpressing Dworf under control of the cardiomyocyte-specific promoter of the α–myosin heavy chain gene. SERCA activity was increased in the heart of the Dworf-overexpressing mice, whereas the affinity of SERCA for Ca^{2+} was slightly reduced in that of Dworf-deficient mice. Consistent with Dworf being most highly expressed in soleus, the decline in the affinity of SERCA for Ca^{2+} was more prominent in this muscle than in the heart of the Dworf-deficient mice. The lack of Dworf expression in quadriceps was also reflected by the lack of a change in SERCA affinity in this muscle of Dworf-deficient mice (Nelson *et al.*, 2016).

A bioinformatics approach identified two additional SERCA-inhibiting peptides, endoregulin (ELN) and another-regulin (ALN), both of which are translated from putative lncRNAs (1110017F19Rik/SMIM6 and 1810037I17Rik, respectively, in mouse) and contain a SERCA binding motif (Anderson *et al.*, 2016). Whereas MLN, PLN, and SLN are abundant in muscle, however, ELN and ALN are expressed in nonmuscle tissues, suggesting that a common mechanism may underlie the control of intracellular Ca^{2+} dynamics in both muscle and nonmuscle tissues.

(5) SPAR

To identify novel peptides encoded by putative lncRNAs, we have adopted a proteomics approach. We selected a polypeptide of 90 amino acids translated from human LINC00961, which we termed Small regulatory Polypeptide of Amino Acid Response (SPAR, also known as SPAAR), for further characterization given its high level of sequence conservation between human and mouse. Immunoblot analysis with antibodies to SPAR verified expression of the endogenous polypetide in both human cell line and mouse tissue. SPAR possesses a single putative transmembrane domain and is localized at the lysosomal membrane. Immunoprecipitation followed by MS identified proteins that interact with SPAR, including the vacuolar-type H^+-ATPase (v-ATPase) complex, the proton pump activity of which is required for lysosomal acidification. v-ATPase also regulates activation of mammalian target of rapamycin complex 1 (mTORC1) in response to amino acid stimulation through interaction with the Ragulator complex at the lysosome (Bar-Peled and Sabatini, 2014). Of note, SPAR had no obvious effect on lysosomal acidification but was

found to inhibit mTORC1 activation by amino acids through regulation of the v-ATPase–Ragulator supercomplex (Fig. 3E) (Matsumoto *et al.*, 2017).

To evaluate the biological functions of SPAR in vivo, we have generated SPAR-deficient mice. These animals manifest no apparent developmental abnormalities, but their skeletal muscle shows enhanced regeneration capacity after acute injury as well as enhanced mTORC1 activation. These observations are consistent with the abundance of *Spar* mRNA in skeletal muscle of wild-type mice as well as with the fact that amino acid–induced mTORC1 activation is required for regeneration after muscle injury, and they suggest that mTORC1 activation is precisely regulated by the SPAR polypeptide in a tissue-specific manner (Matsumoto *et al.*, 2017).

Circular RNAs

(1) circMbl

Analysis of published ribosome footprinting data sets resulted in the identification of 37 circRNAs that contain reads of the footprinting across the circRNA-specific junctions in *Drosophila*, suggesting that these circRNAs are actually translated (Pamudurti *et al.*, 2017). Translation of one of these 37 circRNAs, circMbl, was confirmed by forced expression of an intron-exon-intron minigene in both *Drosophila* S2 cells and transgenic flies. The minigene is able to produce a V5-tagged polypetide only if circularized, given that the V5 tag is located at the COOH-terminus of the predicted product upstream of the initiating methionine. By contrast, no translation was observed with corresponding constructs for circHaspin and circCamK1, neither of which was present in the ribosome footprinting data sets. As mentioned above, the absence of a 5' cap suggests that circRNAs are translated in a cap-independent manner (Fig. 1B). Indeed, the expression of the V5-tagged polypeptide encoded by circMbl was not affected by overexpression of the protein 4E-BP, which inhibits cap-dependent translation. Moreover, both in vivo and in vitro assays revealed that the untranslated region of circMbl supports cap-independent translation, suggesting the presence of a putative IRES element (Pamudurti *et al.*, 2017).

Although expression of the endogenous peptide encoded by circMbl was barely detectable in total fly lysates, immunoblot analysis revealed a band of the expected size (slightly smaller than 10 kDa) in a synaptosomal fraction prepared from fly heads (Pamudurti *et al.*, 2017). The signal intensity for this band was increased by starvation, which promotes cap-independent translation, or by overexpression of FOXO, which inhibits cap-dependent translation. These results thus suggest that the endogenous peptide encoded by circMbl is indeed produced by translation in a cap-independent manner (Pamudurti *et al.*, 2017).

(2) circZNF609

The expression levels of genes for certain circRNAs that are conserved between human and mouse are high in myoblasts of both species and increase further during myoblast differentiation. A similar trend is observed for different circRNAs associated with neuronal differentiation, whereas the expression levels of a specific subset of circRNAs in myoblasts have been found to be altered in individuals with Duchenne muscular dystrophy (DMD) (Rybak-Wolf *et al.*, 2015). On the basis of criteria such as phylogenetic conservation and expression level, 29 molecules were chosen for further characterization from a list of circRNAs that are differentially expressed during human myoblast differentiation or in DMD. RNAi-based functional screening of these 29 circRNAs highlighted circZNF609, the depletion of which inhibited human myoblast proliferation. Whereas circZNF609 is down-regulated during myogenesis, it is up-regulated in myoblasts of DMD patients, suggesting that overexpression of circZNF609 might contribute to the delayed myoblast differentiation phenotype of such patients (Legnini *et al.*, 2017).

A predicted ORF of 753 nucleotides spanning the splicing junction has been identified in circZNF609. The protein-coding potential of circZNF609 was validated with an expression vector encoding the 3FLAG tag at the COOH-terminus of the predicted protein upstream of the initiating methionine, with production of the 3×FLAG-tagged protein being possible only from the circular form of the vector. Expression of the 3×FLAG-tagged protein was induced by heat shock stress, consistent with the activation of cap-independent translation by various external stimuli (Legnini *et al.*, 2017).

Primary miRNAs

(1) pri-miR171b and pri-miR165a

Peptide-encoding pri-miRNAs have recently been identified in plants (Lauressergues *et al.*, 2015). The observation that pri-miRNAs are transcribed by RNAPII is consistent with their coding potential. Indeed, pri-miR171b contains two putative ORFs for peptides of 20 or 5 amino acids, with miR171b regulating the formation of lateral roots in *M. truncatula*. Whether or not these ORFs of pri-miR171b are translated was examined with the use of a β-glucuronidase reporter gene fused to each ATG start codon. Translation was detected only when the reporter gene was fused to the start codon of the ORF encoding 20 amino acids. Endogenous expression of the encoded 20–amino acid peptide, designated miPEP171b, was detected by both immunoblot analysis and immunostaining with specific antibodies, with the expression being apparent at lateral root initiation sites. Another pri-miRNA–encoded peptide,

miPEP165a, was identified in *A. thaliana*. The amino acid sequence of miPEP165a is highly conserved in *Brassicales*. Immunoblot analysis with antibodies to miPEP165a confirmed its endogenous expression (Lauressergues *et al.*, 2015).

Both miPEP171b and miPEP165a have similar functions. Treatment of *M. truncatula* roots with synthetic miPEP171b was found to increase endogenous expression of miR171b and to reduce lateral root density, with the latter phenotype being mimicked by forced expression of miR171b. This effect of miPEP171b on the abundance of its corresponding miRNA was specific, with the peptide having no effect on the amount of other miRNAs. Treatment of *A. thaliana* seedlings with miPEP165a also resulted in the accumulation of miR165a. Additional treatment with cordycepin, an inhibitor of RNA synthesis, attenuated the induction of miR165a, indicating that miPEPs induce their pri-miRNAs at the transcriptional level (Lauressergues *et al.*, 2015).

Conclusion

We have here described examples of biologically relevant peptides encoded by lncRNAs, circRNAs, and pri-miRNAs. Although not covered by this review, many other peptides encoded by ncRNAs—or "hidden" peptides—have been identified. The existence of more such peptides is expected, but their detection is technically difficult given that they yield a limited number of fragments on trypsin digestion and are therefore not optimal for analysis by MS. Development of a technique geared to the identification of such short peptides is thus awaited. The biological activities of hidden peptides might provide the basis for the development of novel therapeutics for currently intractable and other diseases.

References

Ambros, V. 2004. The functions of animal microRNAs. *Nature*, **431**: 350–355.

Anderson, D.M., Anderson, K.M., Chang, C.L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R., and Olson, E.N. 2015. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*, **160**: 595–606.

Anderson, D.M., Makarewich, C.A., Anderson, K.M., Shelton, J.M., Bezprozvannaya, S., Bassel-Duby, R., and Olson, E.N. 2016. Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. *Sci. Signal.*, **9**: ra119.

Bar-Peled, L. and Sabatini, D.M. 2014. Regulation of mTORC1 by amino acids. *Trends Cell Biol.*, **24**: 400–406.

Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**: 281–297.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V.B., Brenner, S.E., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aidinis, V., Allen, J.E., Ambesi-Impiombato, A., Apweiler, R., Aturaliya, R.N., Bailey, T.L., Bansal, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N., Hill, D., Huminiecki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakauchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavesi, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S.L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Semple, C.A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugiura, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S.L., Tang, S., Taylor, M.S., Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Verardo, R., Wei, C.L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J., Hayashizaki, Y., Consortium, F., Group, R.G.E.R. and Genome Science, G. 2005. The transcriptional landscape of the mammalian genome. *Science*, **309**: 1559–1563.

Chen, C.Y. and Sarnow, P. 1995. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science*, **268**: 415–417.

Chng, S.C., Ho, L., Tian, J., and Reversade, B. 2013. ELABELA: a hormone essential for heart development signals via the apelin receptor. *Dev. Cell*, **27**: 672–680.

Consortium, E.P., Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., Kuehn, M.S., Taylor, C.M., Neph, S., Koch, C.M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J.A., Andrews, R.M., Flicek, P., Boyle, P.J., Cao, H., Carter, N.P., Clelland, G.K., Davis, S., Day, N., Dhami, P., Dillon, S.C., Dorschner, M.O., Fiegler, H., Giresi, P.G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K.D., Johnson, B.E., Johnson, E.M., Frum, T.T., Rosenzweig, E.R., Karnani, N., Lee, K., Lefebvre, G.C., Navas, P.A., Neri, F., Parker, S.C., Sabo, P.J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F.S., Dekker, J., Lieb, J.D., Tullius, T.D., Crawford, G.E., Sunyaev, S., Noble, W.S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I.L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H.A., Sekinger, E.A., Lagarde, J., Abril, J.F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A.,

Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J.S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M.C., Thomas, D.J., Weirauch, M.T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K.G., Sung, W.K., Ooi, H.S., Chiu, K.P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M.L., Valencia, A., Choo, S.W., Choo, C.Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T.G., Brown, J.B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C.N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J.S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R.M., Rogers, J., Stadler, P.F., Lowe, T.M., Wei, C.L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S.E., Fu, Y., Green, E.D., Karaoz, U., Siepel, A., Taylor, J., Liefer, L.A., Wetterstrand, K.A., Good, P.J., Feingold, E.A., Guyer, M.S., Cooper, G.M., Asimenos, G., Dewey, C.N., Hou, M., Nikolaev, S., Montoya-Burgos, J.I., Loytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N.R., Holmes, I., Mullikin, J.C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W.J., Stone, E.A., Program, N.C.S., Baylor College of Medicine Human Genome Sequencing, C., Washington University Genome Sequencing, C., Broad, I., Children's Hospital Oakland Research, I., Batzoglou, S., Goldman, N., Hardison, R.C., Haussler, D., Miller, W., Sidow, A., Trinklein, N.D., Zhang, Z.D., Barrera, L., Stuart, R., King, D.C., Ameur, A., Enroth, S., Bieda, M.C., Kim, J., Bhinge, A.A., Jiang, N., Liu, J., Yao, F., Vega, V.B., Lee, C.W., Ng, P., Shahab, A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M.J., Inman, D., Singer, M.A., Richmond, T.A., Munn, K.J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J.C., Couttet, P., Bruce, A.W., Dovey, O.M., Ellis, P.D., Langford, C.F., Nix, D.A., Euskirchen, G., Hartman, S., Urban, A.E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T.H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C.K., Rosenfeld, M.G., Aldred, S.F., Cooper, S.J., Halees, A., Lin, J.M., Shulha, H.P., Zhang, X., Xu, M., Haidar, J.N., Yu, Y., Ruan, Y., Iyer, V.R., Green, R.D., Wadelius, C., Farnham, P.J., Ren, B., Harte, R.A., Hinrichs, A.S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A.S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R.M., Karolchik, D., Armengol, L., Bird, C.P., de Bakker, P.I., Kern, A.D., Lopez-Bigas, N., Martin, J.D., Stranger, B.E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrimsdottir, I.B., Huppert, J., Zody, M.C., Abecasis, G.R., Estivill, X., Bouffard, G.G., Guan, X., Hansen, N.F., Idol, J.R., Maduro, V.V., Maskeri, B., McDowell, J.C., Park, M., Thomas, P.J., Young, A.C., Blakesley, R.W., Muzny, D.M., Sodergren, E., Wheeler, D.A., Worley, K.C., Jiang, H., Weinstock, G.M., Gibbs, R.A., Graves, T., Fulton, R., Mardis, E.R., Wilson, R.K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D.B., Chang, J.L., Lindblad-Toh, K., Lander, E.S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B. and de Jong, P.J. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**: 799–816.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. 2013. Natural RNA circles function as efficient microRNA sponges. *Nature*, **495**: 384–388.

Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P., and Nakamura, A. 2008. Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. *Nature*, **451**: 730–U737.

Ho, L., van Dijk, M., Chye, S.T.J., Messerschmidt, D.M., Chng, S.C., Ong, S., Yi, L.K., Boussata, S., Goh, G.H., Afink, G.B., Lim, C.Y., Dunn, N.R., Solter, D., Knowles, B.B., and Reversade, B. 2017. ELABELA deficiency promotes preeclampsia and cardiovascular malformations in mice. *Science*, **357**: 707–713.

Inagaki, S., Numata, K., Kondo, T., Tomita, M., Yasuda, K., Kanai, A., and Kageyama, Y. 2005. Identification and expression analysis of putative mRNA-like non-coding RNA in Drosophila. *Genes Cells*, **10**: 1163–1173.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*, **324**: 218–33.

Ingolia, N.T. 2016. Ribosome Footprint Profiling of Translation throughout the Genome. *Cell*, **165**: 22–33.

Kang, Y., Kim, J., Anderson, J.P., Wu, J., Gleim, S.R., Kundu, R.K., McLean, D.L., Kim, J.D., Park, H., Jin, S.W., Hwa, J., Quertermous, T., and Chun, H.J. 2013. Apelin-APJ signaling is a critical regulator of endothelial MEF2 activation in cardiovascular development. *Circ. Res.*, **113**: 22–31.

Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S., and Kageyama, Y. 2007. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat. Cell Biol.*, **9**: 660–665.

Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., Kobayashi, S., Payre, F., and Kageyama, Y. 2010. Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. *Science*, **329**: 336–339.

Kranias, E.G. and Hajjar, R.J. 2012. Modulation of cardiac contractility by the phospholamban/SERCA2a regulatome. *Circ. Res.*, **110**: 1646–1660.

Lauressergues, D., Couzigou, J.M., Clemente, H.S., Martinez, Y., Dunand, C., Becard, G., and Combier, J.P. 2015. Primary transcripts of microRNAs encode regulatory peptides. *Nature*, **520**: 90–93.

Legnini, I., Di Timoteo, G., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., Fatica, A., Santini, T., Andronache, A., Wade, M., Laneve, P., Rajewsky, N., and Bozzoni, I. 2017. Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cell*, **66**: 22–37 e29.

Lin, M.F., Jungreis, I., and Kellis, M. 2011. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*, **27**: i275–282.

Ma, J., Diedrich, J.K., Jungreis, I., Donaldson, C., Vaughan, J., Kellis, M., Yates, J.R., and Saghatelian, A. 2016. Improved Identification and Analysis of Small Open Reading Frame Encoded Polypeptides. *Anal. Chem.*, **88**: 3967–3975.

Martinho, R.G., Kunwar, P.S., Casanova, J., and Lehmann, R. 2004. A noncoding RNA is required for the repression of RNApolII-dependent transcription in primordial germ cells. *Curr. Biol.*, **14**: 159–165.

Matsumoto, A., Pasut, A., Matsumoto, M., Yamashita, R., Fung, J., Monteleone, E., Saghatelian, A., Nakayama, K.I., Clohessy, J.G., and Pandolfi, P.P. 2017. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature*, **541**: 228–232.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., le Noble, F., and Rajewsky, N. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495**: 333–338.

Nakamura, A., Amikura, R., Mukai, M., Kobayashi, S., and Lasko, P.F. 1996. Requirement for a noncoding RNA in Drosophila polar granules for germ cell establishment. *Science*, **274**: 2075–2079.

Nelson, B.R., Makarewich, C.A., Anderson, D.M., Winders, B.R., Troupes, C.D., Wu, F., Reese, A.L., McAnally, J.R., Chen, X., Kavalali, E.T., Cannon, S.C., Houser, S.R., Bassel-Duby, R., and Olson, E.N. 2016. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*, **351**: 271–275.

Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., Shenzis, S., Samson, M., Dittmar, G., Landthaler, M., Chekulaeva, M., Rajewsky, N., and Kadener, S. 2017. Translation of CircRNAs. *Mol. Cell*, **66**: 9–21 e27.

Pasut, A., Matsumoto, A., Clohessy, J.G., and Pandolfi, P.P. 2016. The pleiotropic role of non-coding genes in development and cancer. *Curr.*

Opin. Cell Biol., **43**: 104–113.

Pauli, A., Norris, M.L., Valen, E., Chew, G.L., Gagnon, J.A., Zimmerman, S., Mitchell, A., Ma, J., Dubrulle, J., Reyon, D., Tsai, S.Q., Joung, J.K., Saghatelian, A., and Schier, A.F. 2014. Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science*, **343**: 1248636.

Payre, F., Vincent, A., and Carreno, S. 1999. ovo/svb integrates Wingless and DER pathways to control epidermis differentiation. *Nature*, **400**: 271–275.

Rybak-Wolf, A., Stottmeister, C., Glazar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., Ashwal-Fluss, R., Herzog, M., Schreyer, L., Papavasileiou, P., Ivanov, A., Ohman, M., Refojo, D., Kadener, S., and Rajewsky, N. 2015. Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol. Cell*, **58**: 870–885.

Scott, I.C., Masri, B., D'Amico, L.A., Jin, S.W., Jungblut, B., Wehman, A.M., Baier, H., Audigier, Y., and Stainier, D.Y. 2007. The g protein-coupled receptor agtrl1b regulates early development of myocardial progenitors. *Dev. Cell*, **12**: 403–413.

Slavoff, S.A., Mitchell, A.J., Schwaid, A.G., Cabili, M.N., Ma, J., Levin, J.Z., Karger, A.D., Budnik, B.A., Rinn, J.L., and Saghatelian, A. 2013. Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.*, **9**: 59–+.

Sucena, E., Delon, I., Jones, I., Payre, F., and Stern, D.L. 2003. Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature*, **424**: 935–938.

Waterhouse, P.M. and Hellens, R.P. 2015. Plant biology: Coding in noncoding RNAs. *Nature*, **520**: 41–42.

Zanet, J., Benrabah, E., Li, T., Pelissier-Monier, A., Chanut-Delalande, H., Ronsin, B., Bellen, H.J., Payre, F., and Plaza, S. 2015. Pri sORF peptides induce selective proteasome-mediated protein processing. *Science*, **349**: 1356–1358.

Zeng, X.X., Wilm, T.P., Sepich, D.S., and Solnica-Krezel, L. 2007. Apelin and its receptor control heart field formation during zebrafish gastrulation. *Dev. Cell*, **12**: 391–402.

Zhang, X.O., Wang, H.B., Zhang, Y., Lu, X., Chen, L.L., and Yang, L. 2014. Complementary sequence-mediated exon circularization. *Cell*, **159**: 134–147.