Casting CRISPR-Cas13d to fish for microprotein functions in animal development

Anthony James Treichel1 and Ariel Alejandro Bazzini1,2,*

SUMMARY
Protein coding genes were originally identified with sequence-based definitions that included a 100-codon cutoff to avoid annotating irrelevant open reading frames. However, many active proteins contain less than 100 amino acids. Indeed, functional genetics, ribosome profiling, and proteomic profiling have identified many short, translated open reading frames, including those with biologically active peptide products (microproteins). Yet, functions for most of these peptide products remain unknown. Because microproteins often act as key signals or fine-tune processes, animal development has already revealed functions for a handful of microproteins and provides an ideal context to uncover additional microprotein functions. However, many mRNAs during early development are maternally provided and hinder targeted mutagenesis approaches to characterize developmental microprotein functions. The recently established, RNA-targeting CRISPR-Cas13d system in zebrafish overcomes this barrier and produces potent knockdown of targeted mRNA, including maternally provided mRNA, and enables flexible, efficient interrogation of microprotein functions in animal development.

PIONEERING THE DISCOVERY OF EUKARYOTIC MICROPROTEINS
At the onset of the genomics era, stringent sequence-based gene definitions were adopted to avoid annotating biologically irrelevant open reading frames (ORFs)—especially short ORFs (≤ 100 codons long) that exhibit a high frequency based on probability alone. For example, in zebrafish (Danio rerio) there are over 425,000 ORFs with canonical (AUG) start codons between 10 and 100 codons within messenger RNAs (mRNAs) and long non-coding RNAs alone. Only approximately 1,000 (0.02%) of these short ORFs are currently described as protein coding, leaving countless short ORFs with the potential to be translated as “beautiful needles in the haystack”1 (Figure 1A).

Functional genetics in budding yeast first confirmed the existence of eukaryotic “needles” upstream of the canonical ORF encoding GCN4,2 a transcription factor that activates genes involved in amino acid biosynthesis.3 Translation of four short, upstream (within mRNA sequence once defined as the 5’ untranslated region (UTR)) ORFs (uORFs) in GCN4 is required for nutrient-dependent regulation of GCN4 translation but relies on the uORF’s ability to engage ribosomes rather than a biologically active peptide product. However, subsequent functional genetic studies revealed short ORFs that encode biologically active peptide products (from here on: microproteins) in diverse organisms from budding yeast3 to legumes4 and insects5,6–8 (Figure 1B). Notably, all initially discovered microproteins are encoded on polycistronic mRNAs, and parallel bioinformatic analyses revealed that many mRNAs contain uORFs9–15 including a population that shows amino acid conservation in mammals.9 This evidence collectively questioned the fundamental assumption that eukaryotic mRNAs are translated into a single peptide product.

Further supporting initial functional genetics and bioinformatics, the advent of ribosome profiling16–18 and targeted proteomic approaches19–25 over the last decade has revolutionized our ability to detect short ORF translation. Ribosome profiling, for example, leverages that ribosomes protect short fragments of bound mRNAs which are amenable to isolation and next-generation sequencing. Remarkably, some methods yield ribosome protected fragments that are consistent enough to define the reading frame18 (Figure 1B). Complementary detection of short ORF encoded peptide products with targeted proteomics has supported widespread translation of short ORFs on polycistronic mRNAs across many eukaryotic
taxa\textsuperscript{16,17,21,23,24,26–38} (Figure 1B). Omics-based approaches have consequently completed the paradigm shift in RNA biology—many eukaryotic mRNAs exhibit multiple, distinct translation events rather than producing a single peptide product.
Furthermore, omics techniques have defined diverse classes of translated short ORFs and polycistronic mRNA structures (Figure 1C, reviewed extensively in39–42). For instance (and like those in GCN4), translation of uORFs often acts to suppress translation of the canonical coding sequence.2,12,43–46 On the other hand, translation of downstream short ORFs (dORFs) (within mRNA sequence once called the 3’ UTR) enhances canonical ORF translation.38 Although many uORFs and dORFs regulate translation independently of microprotein activity, the catalog of uORFs that encode functional microproteins continues to grow.4,33,47,48 In fact, short ORFs nested within canonical coding sequences (overlapping ORFs) are translated into functional microproteins.27–29,51–62 However, although techniques continue to evolve and shine light on new short, translated ORFs, the potential functions encoded within the vast majority of identified short ORF peptide products remain unknown.

MICROPROTEINS ARE INTEGRAL IN DIVERSE EUKARYOTIC BIOLOGY, INCLUDING ANIMAL DEVELOPMENT

The arginine attenuator peptide (AAP) in budding yeast is one of the first functional microproteins identified in a eukaryote. AAP is a 25 amino acid microprotein produced from a uORF in the *Saccharomyces cerevisiae* CPA1 gene, which encodes an enzyme involved in arginine synthesis. AAP is required for free arginine to repress CPA1 translation, enabling arginine to negatively regulate its own biosynthesis.1 However, with a few notable exceptions5–8, most functional microproteins have been discovered in the last ten years with the aid of ribosome profiling16,17 and advances in proteomic and biochemical tools.21,24,25,63 Although microproteins encoded in animal genomes will be the main focus of this review, microproteins also play critical and emerging roles in fungi,4,12,64 bacteria65–67 (reviewed in68), and plants5,69,70 (reviewed in36,71).

In animals, microproteins play integral roles in diverse biological processes including transcription54,72,73 (Figure 2A), ribosome biogenesis,49 cell growth and viability,33 cancer cell survival,35 embryonic stem cell pluripotency74 (Figure 2B) and differentiation,75 immune responses48,60,62 and both cardiac34,53,54,76 (Figure 2B) and skeletal27–29 muscle function and development (Figure 2C). In fact, developmental contexts have provided significant insights into microprotein functions4–8,27–29,52,53,74–77 (Figures 2A–2C). For instance, parallel developmental studies in *Tribolium castaneum* and *Drosophila melanogaster* identified the first animal microprotein gene called *mille-pattes* or *tarsal-less/polished rice*, respectively, that produces a polycistronic mRNA that encodes functionally redundant 11–32 amino acid microproteins.5–8 Tarsal-less microproteins interact with Ubr3, an E3 ubiquitin ligase, and enable it target the transcription factor shavenbaby.27,78 Consequent proteasome-mediated cleavage of the shavenbaby N-terminus switches it from a transcriptional repressor to a transcriptional activator, inducing the expression of genes critical for distal limb patterning in *Drosophila*81 and abdominal segment identity in many other insects73,77 (Figure 2A). Similarly, characterized microproteins often function as essential regulators of protein activity or key signaling molecules, supporting that uncharacterized short ORF peptide products likely have additional functions in orchestrating animal development.

SHORT ORF TRANSLATION IS PREVALENT DURING VERTEBRATE DEVELOPMENT

Early stages of vertebrate development and vertebrate pluripotent stem cells show widespread short ORF translation.17,32,33,38,54,79,80 Among many vertebrate model systems, zebrafish (*D. rerio*) has emerged as a premier model system because of its high offspring number, ex-tero development, and genetic tractability. Not to mention, zebrafish exhibit many conserved cellular, molecular, and genetic foundations of development,81 and more than 70% of human genes have at least one ortholog in zebrafish, making zebrafish studies highly informative for human biology.82 Critically, zebrafish exhibit hundreds of translated short ORFs across their first five days of development.32,54 Although first identified in 2014, only twelve zebrafish developmentally translated short ORFs are currently characterized.53,54,77,83 Targeted mutagenesis has revealed that four of these short ORFs produce functional microproteins within the first seven days of development. The pioneer vertebrate functional microprotein called apela (elabela, toddler) is a conserved signaling molecule that is critical for zebrafish mesodermal migration, heart formation in zebrafish and mouse,33,54,76 pluripotency in human embryonic stem cells,74 and placental development in mouse76 (Figure 2B). Specifically in zebrafish, the microprotein bouncer localizes to the oocyte membrane and is necessary and sufficient for fertilization.63 Two additional zebrafish microproteins encoded in *linc-mipep* and *linc-wrb* have been recently reported to coordinate to regulate locomotor activity in larvae (4–7 days...
MATERINALLY PROVIDED RNAs IMPEDE MICROPROTEIN CHARACTERIZATION IN EARLY DEVELOPMENT

Animal development presents unique challenges to elucidating microprotein functions. During maturation, animal oocytes are packed with maternal RNAs and proteins. Immediately after fertilization, maternally provided RNAs and proteins control the developmental program while the zygote remains transcriptionally silent. As the zygote awakens its genome and ramps up production of its own transcripts, it is weaned from maternal control in a process known as the maternal-to-zygotic transition (reviewed extensively in84). Still, the majority of RNAs during the maternal-to-zygotic transition are maternally provided post-fertilization), directly or indirectly changing chromatin accessibility in CNS cells called oligodendrocytes.77 Still, hundreds of developmental short ORF encoded peptide products in zebrafish alone remain without insights into their functions.

Figure 2. Established microproteins play critical roles in animal development

(A) Tarsal-less/polished rice/mille-pattes is conserved throughout insects and produces a polycistronic mRNA that encodes functionally redundant microproteins critical for regulating the activity of the transcription factor shavenbaby for its role in specifying distal limb identity in Drosophila melanogaster and abdominal segment identity in the larvae of many insects, including Tribolium castaneum. 

(B) apela/elabela/toddler encodes a signaling molecule that is conserved from zebrafish to humans that is critical for mesodermal migration and heart formation in both zebrafish and mouse, resulting in a lethal phenotype. Loss of Apela/Elabela/Toddler function in mouse also affects placental development and maternal health. Furthermore, APELA/ELABELA/TODDLER is required for maintaining pluripotency in human embryonic stem cells.

(C) Myomixer/Minion encodes a conserved, membrane-tethered ligand that is both necessary and sufficient for the cellular fusion required for muscle cells to develop into multinucleated muscle fibers in mouse. Because Myomixer/Minion is necessary for the development of the diaphragm muscle, knockout mice fail to inflate their lungs, resulting in perinatal lethality.
a few developmental microproteins (Figure 3B). Alternatively, conventional RNAi is generally ineffective in zebrafish and other teleosts (e.g., medaka, killifish, Mexican tetra), which has led to widespread use of morpholinos as the preeminent knockdown technology. Although morpholinos have correctly elucidated gene function in select studies, the general utility of morpholinos has been questioned over off-target effects and general toxicity. In some cases, loss-of-function mutant and morpholino-treated animals exhibit discordant phenotypes. More strikingly, additional phenotypes have been observed on morpholino treatment in loss-of-function mutants, indicating cellular effects unrelated to the targeted RNAs. In fact, some morpholinos can trigger cellular immune responses and off-target mis-splicing. On top of these concerns, morpholinos are expensive and their validation can be cumbersome.

CRISPR-CAS13D KNOCKS DOWN BARRIERS TO STUDYING MATERNALLY PROVIDED MRNAS

Following the discovery and widespread application of DNA-targeting CRISPR systems (e.g., CRISPR-Cas9), microbial (meta)genomic mining revealed a unique CRISPR-associated (Cas) protein—Cas13a (originally C2c2). Instead of DNA-nuclease domains (HNH and/or RuvC-like domains like Cas9 and Cas12a), Cas13a contains two RNA-nuclease domains (HEPN domains) that enable Cas13a to function as an autonomous, programmable (through its CRISPR (guide) RNA) RNA-targeting nuclease. More Cas13 protein families—denoted Cas13b, Cas13c, and Cas13d—were subsequently identified. The Cas13d family exhibits the smallest proteins (average length ~930 amino acids), and Cas13d from Ruminococcus flavefaciens produces potent mRNA knockdown in mammalian cell culture.

To overcome the limitations in targeting maternally provided mRNAs, we established the RNA-targeting CRISPR-Cas13d system in zebrafish where it autonomously and potently degrades target mRNAs through at least the first 48 hours of development (see system development in and detailed methods in). Because it targets both maternally provided and zygotically transcribed RNA (Figure 3C), CRISPR-Cas13d is a straightforward tool for loss-of-function studies of early developmental RNAs, including microprotein mRNAs and long non-coding RNAs. RT-qPCR offers a simple first-pass knockdown validation and a diagnostic tool for guide RNA design. For targets with >60% knockdown and a phenotype, it is simple to generate a high number of knockdown embryos to analyze the onset and the spectrum of the developmental effect. Further, this high offspring number is amenable to interrogating the molecular phenotype (e.g., genes and processes affected) and transcriptome-wide CRISPR-Cas13d targeting specificity with RNA-sequencing (Figure 4A).

Further, CRISPR-Cas13d guide RNAs are sequence-unconstrained, which enables effective targeting for any RNA of interest. Similarly, cross-validation of observed phenotypes with either multiple cocktails of guide RNAs or individual guide RNAs is feasible because their synthesis is simple and relatively inexpensive. Depending on the application, Cas13d guide RNAs can be co-injected with either Cas13d protein (quick knockdown, long production) or mRNA (long knockdown, quick production). Regardless of approach (see detailed methods in), CRISPR-Cas13d is an effective and efficient knockdown system. A CRISPR-Cas13d novice could reasonably perform a dozen unique CRISPR-Cas13d knockdowns in six months. Finally, although constitutive, heterologous expression systems in cell culture have revealed non-specific CRISPR-Cas13d RNA degradation when targeting abundant transcripts, the injection-based tools in zebrafish are powerful assets to interrogate any potential off-target effects. For example, co-injection of mRNA with the tbxta coding sequence rescues the “No-Tail” phenotype observed under CRISPR-Cas13d knockdown (targeting the endogenous 3’ UTR of tbxta), eliminating off-target concerns in this knockdown paradigm. CRISPR-Cas13d therefore opens the door to the interrogation of developmental functions for microproteins in zebrafish.

In addition to microprotein mRNAs, CRISPR-Cas13d enables loss-of-function studies for any maternally provided RNA of interest. For example, analyses of single zygotic mutants for many zebrafish long non-coding RNAs failed to find developmental functions. Although zygotic mutants enable maternally provided RNA to mask any developmental effect, CRISPR-Cas13d eliminates this maternal RNA and reduces the likelihood of false negatives in screens for non-coding RNA functions in development (Figure 3C). Outside of maternally provided RNA, double (or triple, etc.) mutants are required to uncover the impact of long non-coding RNAs with complementary functions. In contrast, CRISPR-Cas13d facilitates simple targeting of multiple RNAs to test for overlapping functions. Intriguingly, many microproteins are encoded on RNAs once described as non-coding, raising the possibility that a notable subset of RNAs have both coding
and non-coding functions or translation-dependent, non-coding functions (somewhat analogous to uORFs and dORFs). CRISPR-Cas13d, alongside established injection-based approaches in zebrafish, provides a truly feasible system in vertebrates to simultaneously interrogate both non-coding and coding functions within diverse RNAs.

**UNRAVELING MICROPROTEIN AND NON-CODING RNA FUNCTIONS WITH CRISPR-CAS13D**

On top of its potent loss-of-function effects, the flexibility of CRISPR-Cas13d guide RNAs enables routine injection-based approaches to provide key insights into developmental microprotein mRNA and/or
non-coding RNA functions. When CRISPR-Cas13d knockdown of a microprotein mRNA produces a developmental phenotype, the paramount question is whether the phenotype is a result of microprotein activity (Figure 4B). Because microproteins are short, a chemically synthesized microprotein can be co-injected in knockdown embryos to test its ability to rescue the phenotype. However, RNA-based approaches are more flexible and efficient. In-vitro transcription vectors in zebrafish enable the generation of a "rescue" mRNA with the microprotein coding sequence flanked by exogenous 5' and 3' UTRs, allowing guide RNAs targeting the microprotein UTRs to knockdown only the native zebrafish mRNA (Figure 4B). Microprotein activity is supported when a "rescue" mRNA alleviates the knockdown phenotype, but a frameshifted version fails to do so. Likewise, synonymous mutations in the microprotein sequence in the "rescue" mRNA should also alleviate the phenotype.

Rescue experiments can be leveraged for additional insights into conservation of function and critical amino acid residues. If microprotein function is conserved, homologous microprotein sequences (or syntenic non-coding RNAs) from other species in "rescue" mRNAs should similarly alleviate the developmental effect. In parallel, for any critical amino acid residue, a mutation in the "rescue" mRNA will fail to alleviate the phenotype. In a like manner, any developmentally critical non-coding RNA discovered with CRISPR-Cas13d can be mutated to assess for functional sequences and/or structures. Importantly, fluorescent or epitope tags can be added to microprotein coding sequences in the transcription vector alone, without the need for endogenous tagging. Although tagging has the potential to disrupt microprotein function, rescue of its CRISPR-Cas13d knockdown phenotype validates the activity of any tagged microprotein. Tag-based imaging can define microprotein subcellular localization whereas co-immunoprecipitation experiments based on the tag can identify microprotein interaction partners (Figure 4C).
NOVEL COMBINATIONS OF CRISPR-CAS13D WILL PROVIDE ADDITIONAL INSIGHTS

RNA-seq is very effective at defining the transcriptome-wide effects in CRISPR-Cas13d knockdowns. However, a plethora of techniques are eagerly waiting to work in concert with CRISPR-Cas13d to provide further insights into microprotein and/or non-coding RNA functions. For example, v3 HCR in situ hybridization in knockdown embryos can provide spatiotemporal readouts for transcripts of interest from RNA-seq to determine if their localization(s) is changing (Figure 5A). In-situ for cell-type specific markers (e.g., endoderm, ectoderm, and mesoderm) in knockdown embryos allow interrogation of potential cell fate decision and differentiation defects. Further, a key process during the maternal-to-zygotic transition is the activation of transcription from the zygotic genome (Figure 3A), which involves major changes to chromatin accessibility. CRISPR-Cas13d is amenable to genomics-based techniques (e.g., ChIP-seq, ATAC-seq) to interrogate chromatin states under microprotein knockdown because it is straightforward to generate many, homogeneous knockdown embryos (Figure 5B).

Although CRISPR-Cas13d enables the dissection of gene function during the maternal-to-zygotic transition without maternal/zygotic mutants, parallel targeted mutagenesis (e.g., CRISPR-Cas9) enables investigations into the individual phenotype contributions from maternally provided and/or zygotically transcribed RNA. In zebrafish, CRISPR-Cas9 is highly optimized and can produce over 90% biallelic deletions in F0 mutants, enabling rapid comparisons with CRISPR-Cas13d knockdowns. For instance, if zygotic RNA is the major contributor, zygotic F0 mutants will recapitulate the CRISPR-Cas13d knockdown phenotype. Alternatively, if maternal RNA is the major contributor, zygotic mutants will fail to recapitulate the phenotype (Figure 5C). In this case, a germline targeted Cas9 construct is available in zebrafish that can effectively generate maternal mutants in F0 (Figure 5C). Importantly, crossing F0 maternal mutants to wildtype males would validate the essential role of maternally provided RNA in a single generation (Figure 5C).

Further, simultaneous targeting of multiple microprotein or non-coding RNA genes is feasible with CRISPR-Cas13d. In fact, multiple targeting is especially useful in zebrafish where at least 26% of genes have a paralog from the teleost-specific genome duplication that could retain overlapping function. Moreover, CRISPR-Cas13d is effective in mouse, killifish, and medaka, enabling interrogation of conserved microprotein and non-coding RNA functions in development. Although this review has focused on development, adult zebrafish afford a large repertoire of tissue and cell-type specific drivers and human...
disease models. Cas13d could therefore be employed in a selective manner to interrogate microprotein or non-coding RNA functions in homeostasis or in zebrafish models of human disease.

FUTURE PERSPECTIVES FOR CRISPR-CAS13, MICROPROTEINS, AND NON-CODING RNAS

In all, CRISPR-Cas13d in zebrafish not only expands a solid vertebrate developmental model system but also opens novel angles for diverse developmental questions about microprotein and non-coding RNAs, particularly those that involve the maternal-to-zygotic transition. Alongside CRISPR-Cas13d, RNA and protein injection-based tools in zebrafish enable relatively simple dissection of diverse microproteins (e.g., encoded in overlapping ORFs, 2–10 codon ORFs) and non-coding RNAs (e.g., coordinate functions, and complementary roles). The utility of CRISPR-Cas13 across teleosts and mammals potentiates powerful comparative studies that can test conservation of function and provide insights into the evolution of both microproteins and non-coding RNAs. Similarly, implementation of CRISPR-Cas13d in plants could provide insights into microprotein and/or non-coding RNA function(s) during early plant development, which may involve contributions from both maternally and paternally provided RNAs. One caveat to the CRISPR-Cas13d system is that it degrades maternally provided microprotein mRNAs, not the microproteins themselves. Any maternally provided microproteins whose function depends entirely on their maternal microprotein pool would escape detection in a CRISPR-Cas13d loss-of-function screen.

Beyond its current role in RNA knockdown, the CRISPR-Cas13 system has exciting potential. For instance, catalytically dead Cas13d fused to translation initiation factors was recently shown to increase translation in E. coli. If established in zebrafish, this would enable over-expression studies for microproteins without the need for injection of in vitro transcribed RNA that may fail to recapitulate the native microprotein mRNA expression pattern. Furthermore, application of the dead Cas13d-based RNA imaging approaches (developed in mammalian cell culture) to the accessible live imaging provided by developing zebrafish would illuminate the spatiotemporal dynamics of microprotein mRNA and/or non-coding RNA in a living organism. Finally, implementing novel microprotein tagging approaches (e.g., MicroID, non-canonical amino acid labeling) in zebrafish would complement functional studies with CRISPR-Cas13d to better elucidate microprotein functions in their native contexts.

ACKNOWLEDGMENTS

We thank Mark Miller for assistance with illustrations, and Michay Diez, Josefina Doffo, Cameron Berry, Gabriel da Silva Pescador, and Danielson Baia Amaral for constructive feedback on the manuscript. This work was supported by the Stowers Institute for Medical Research. A.A.B was awarded a Pew Innovation Fund and the US National Institutes of Health (NIH-R01 GM136849). This was compiled as part of thesis research for A.J.T., Graduate School of the Stowers Institute for Medical Research.

AUTHOR CONTRIBUTIONS

Conceptualization, A.J.T. and A.A.B.; Writing—Original Draft, A.J.T.; Writing—Review and Editing, A.J.T. and A.A.B.; Visualization, A.J.T.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

REFERENCES

1. Basrai, M.A., Hieter, P., and Boeke, J.D. (1997). Small open reading frames: beautiful needles in the haystack. Genome Res. 7, 768–771. https://doi.org/10.1101/gr.7.8.768.

2. Mueller, P.P., and Hinnebusch, A.G. (1986). Multiple upstream AUG codons mediate translational control of GCN4. Cell 45, 201–207. https://doi.org/10.1016/0092-8674(86)90384-3.

3. Werner, M., Feller, A., Messenguy, F., and Piérad, A. (1987). The leader peptide of yeast gene CPA1 is essential for the translational repression of its expression. Cell 49, 805–813. https://doi.org/10.1016/0092-8674(87)90618-0.

4. Rohrig, H., Schmidt, J., Miklashevichs, E., Schell, J., and John, M. (2002). Soybean ENOD40 encodes two peptides that bind to sucrose synthase. Proc. Natl. Acad. Sci. USA 99, 1915–1920. https://doi.org/10.1073/pnas.022664799.
embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147, 789–802. https://doi.org/10.1016/j.cell.2011.10.002.

18. Brar, G.A., and Weissman, J.S. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. Nat. Rev. Mol. Cell Biol. 16, 651–664. https://doi.org/10.1038/nrm4069.

19. Ma, J., Ward, C.C., Jungreis, I., Slavoff, S.A., Schwaid, A.G., Neveu, J., Budnik, B.A., Kelis, A., and Saghatelian, A. (2014). Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. J. Proteome Res. 13, 1757–1765. https://doi.org/10.1021/pr401290w.

20. Schwaid, A.G., Shannon, D.A., Ma, J., Slavoff, S.A., Levin, J.Z., Weerapana, E., and Saghatelian, A. (2013). Chemoproteomic discovery of cysteine-containing human short open reading frames. J. Am. Chem. Soc. 135, 1650–1653. https://doi.org/10.1021/ja406606x.

21. 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, 59–64. https://doi.org/10.1038/nchembio.1120.

22. Tinoco, A.D., Tagore, D.M., and Saghatelian, A. (2010). Expanding the dipeptidyl peptidase 4-regulated peptidome via an optimized peptidomics platform. J. Am. Chem. Soc. 132, 3819–3830. https://doi.org/10.1021/ja909524e.

23. Khtun, A., and Slavoff, S.A. (2019). Proteomic detection and validation of translated small open reading frames. Curr. Protoc. Chem. Biol. 11, e77. https://doi.org/10.1002/cphc.201905054.

24. Haendel, C.A., and Bosco, G. (2008). Comparative genomic analysis of novel conserved uORF homologous upstream open reading frames in Drosophila melanogaster and other insect species. BMC Genom. 9, 61. https://doi.org/10.1186/1471-2164-9-61.

25. Haendel, C.A., and Jorgensen, R.A. (2007). Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription-factor-encoding genes. BMC Genom. 5, 32. https://doi.org/10.1186/1471-2164-9-52.

26. Neafsey, D.E., and Galagan, J.E. (2007). Dual modes of natural selection on upstream open reading frames. Mol. Biol. Evol. 24, 1744–1751. https://doi.org/10.1093/molbev/msm093.

27. 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, 218–223. https://doi.org/10.1126/science.1168978.

28. Anderson, D.M., Makarewich, C.A., Anderson, K.M., Karouzas, J.M., Bezpuzko, A., Kiser, D.L., Olson, E.N. (2016). Widespread control of calcium signaling by a family of SERCA-inhibiting micropeptides. Sci. Signal. 9, ra119. https://doi.org/10.1126/scisignal.aaj4160.

29. Bi, P., McAnally, J.R., Shelton, J.M., Sanchez-Ortiz, E., Bassel-Duby, R., and Olson, E.N. (2018). Fusogenic Myomixer is essential for satellite cell fusion and muscle regeneration. Proc. Natl. Acad. Sci. USA 115, 3864–3869. https://doi.org/10.1073/pnas.1800521115.

30. Bi, P., Ramirez-Martinez, A., Li, H., Cannavino, J., McNally, J.R., Shelton, J.M., Sanchez-Ortiz, E., Bassel-Duby, R., and Olson, E.N. (2017). Control of muscle formation by the fusogenic micropeptide Myomixer. Science 356, 323–327. https://doi.org/10.1126/science.aam9361.

31. Ladoukakis, E., Pereira, V., Magny, E.G., Eyre-Walker, A., and Couso, J.P. (2011). Hundreds of putatively functional small open reading frames in Drosophila. Genome Biol. 12, R118. https://doi.org/10.1186/gb-2011-12-11-r118.

32. Bazzini, A.A., Johnstone, T.G., Christiano, R., Mackowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C., and Giraldez, A.J. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J. 33, 981–993. https://doi.org/10.1002/embj.20148411.

33. Chen, J., Brunner, A.D., Cogan, J.Z., Nunez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., and Weissman, J.S. (2020). Pervasive functional translation of noncanonical human open reading frames. Science 367, 1140–1146. https://doi.org/10.1126/science.aay0262.

34. van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J.F., Adami, E., Faber, A.B., Kirchner, M., Maatz, H., Blachut, S., Sandmann, C.L., et al. (2019). The translational landscape of the human heart. Cell 178, 242–260.e29. https://doi.org/10.1016/j.cell.2019.05.010.

35. Prensner, J.R., Enache, O.M., Luria, V., Krug, K., Clauser, K.R., Dempster, J.M., Karger, A., Wang, L., Stumbraite, K., Wang, V.M., et al. (2021). Noncanonical open reading frames encode functional proteins essential for cancer cell survival. Nat. Biotechnol. 39, 697–704. https://doi.org/10.1038/s41587-020-00866-2.

36. Hanada, K., Higuchi-Takeuchi, M., Okamoto, M., Yoshizumi, T., Shimizu, M., Nakaminami, K., Nishi, R., Ohashi, C., Iida, K., Tanaka, M., et al. (2013). Small open reading frames associated with morphogenesis are hidden in plant genomes. Proc. Natl. Acad. Sci. USA 110, 2395–2400. https://doi.org/10.1073/pnas.1219358110.
49. Cao, X., Khutun, A., Harold, C.M., Bryant, C.J., Zheng, S.J., Baserga, S.J., and Slavoff, S.A. (2022). Nascent alt-protein chemoproteomics reveals a pre-60S assembly checkpoint inhibitor. Nat. Chem. Biol. 18, 643–651. https://doi.org/10.1038/s41589-022-01003-9.

50. Cao, X., Khutun, A., Luo, Y., Na, Z., Phoondokmai, T., Sagaputhan, K., Olatunji, E., Uttamapinant, C., and Slavoff, S.A. (2021). Alt-RPL36 downregulates the PI3K-AKT-mTOR signaling pathway by interacting with TMEM62. Nat. Commun. 12, 598. https://doi.org/10.1038/s41467-020-20841-6.

51. D’Lima, N.G., Ma, J., Winkler, L., Chu, Q., Loh, Y., et al. (2017). A human microprotein that interacts with the mRNA decapping complex. Nat. Chem. Biol. 13, 174–180. https://doi.org/10.1038/nchembio.2249.

52. Zhang, Q., Vashishth, A.A., O’Rourke, J., Corbel, S.Y., Moran, R., Romero, A., Miriglia, L., Zhang, J., Durrant, E., Schmedt, J., et al. (2017). The microprotein Minion controls cell fusion and muscle formation. Nat. Commun. 8, 16644. https://doi.org/10.1038/ncomms16644.

53. Chng, S.C., Ho, L., Tian, J., and Reversade, B. (2013). ELABE: a hormone essential for heart development signals via the apelin receptor. Dev. Cell 27, 672–680. https://doi.org/10.1016/j.devcel.2013.11.002.

54. Pauli, A., Norris, M.L., Valen, E., Chew, G.L., Gagnon, J.A., Zimmerman, S., Mitchell, A., Ma, J., Dubrulle, J., Reyon, D., et al. (2016). Toddler: an embryonic signal that promotes cell fusion. Science 343, 1248636. https://doi.org/10.1126/science.1248636.

55. Chu, Q., Martinez, T.F., Novak, S.W., Donaldson, C.J., Tan, D., Vaughan, J.M., Chang, T., Diedrich, J.K., Andrade, L., Kim, A., et al. (2019). Regulation of the ER stress response by a mitochondrial microprotein. Nat. Commun. 10, 4883. https://doi.org/10.1038/s41467-019-12816-z.

56. Koh, M., Ahmad, I., Ko, Y., Zhang, Y., Martinez, T.F., Diedrich, J.K., Chu, Q., MoreSCO, J.J., Perlman, A., Saghathalian, A., et al. (2021). A short ORF-encoded transcriptional regulator. Proc. Natl. Acad. Sci. USA 118. e2021943118. https://doi.org/10.1073/pnas.2021943118.

57. Rathore, A., Chu, Q., Tan, D., Martinez, T.F., Donaldson, C.J., Diedrich, J.K., Yates, J.R., et al. (2018). MIEF1 microprotein regulates mitochondrial translation. Biochemistry 57, 5566–5575. https://doi.org/10.1021/acs.biochem.8b00726.

58. Slavoff, S.A., Heo, J., BudiK, B.A., HanakaHi, L.A., and Saghathalian, A. (2014). A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining. J. Biol. Chem. 289, 10950–10957. https://doi.org/10.1074/jbc.C113.53996.

59. Senis, E., Eglesias, M., Najas, S., Jiménez-SáBado, V., Bertani, C., Giménez-Alejand, M., Escoche, A., Ruiz-Oreja, J., HerquetA-Redondo, L., Jiménez, M., et al. (2021). TUNAR IncRNA encodes a microprotein that regulates neural differentiation and neurite formation by modulating calcium dynamics. Front. Cell Dev. Biol. 9, 747667. https://doi.org/10.3389/fcell.2021.747667.

60. Lee, C.Q.E., Keroaunant, B., Chothani, S., Zhang, S., Chen, Y., Mantri, C.K., Hock, D.H., Lim, R., Nakdwani, R., Huyhn, V.T., et al. (2021). Coding and non-coding roles of MOCCI (C15ORF48) coordinate to regulate host inflammation and immunity. Nat. Commun. 12, 2130. https://doi.org/10.1038/s41467-021-22997-5.

61. Zhang, S., Reijii, B., Liang, C., Keroaunant, B., Francisco, J.C., Peh, J.H., Mary, C., Jagannath, N.S., Olesiuk, V., Tang, C., et al. (2020). Mitochondrial peptide BRAWNI is essential for vertebrate respiratory complex III assembly. Nat. Commun. 11, 1312. https://doi.org/10.1038/s41467-020-14499-2.

62. Bhatta, A., Atianand, M., Jiang, Z., Crabtree, J., Bui, J., Lin, J., and Fitzgerald, K.A. (2020). A mitochondrial micropeptide is required for activation of the Nlrp3 inflammasome. J. Immunol. 204, 428–437. https://doi.org/10.4049/jimmunol.1900791.

63. Lafranchi, L., Schlesinger, D., Klemier, K.J., and Elssasser, S.J. (2020). Universal single-residue terminal labels for fluorescent live cell imaging of microproteins. J. Am. Chem. Soc. 142, 20080–20086. https://doi.org/10.1021/jacs.0c09574.

64. Tollis, S., Singh, J., Palau, R., Thattikota, Y., Ghazal, G., Coloumb-Huntington, J., Tang, X., Moore, S., Blake, D., Bonnel, E., et al. (2022). The microprotein Nr1 rewires the G1/S transcriptional machinery during nitrogen limitation in budding yeast. PLoS Biol. 20, e3001548. https://doi.org/10.1371/journal.pbio.3001548.

65. Asayan, J.J., Raina, M., Zhang, A., and Storz, G. (2022). Dual-function site of an RNA encodes a 15-amino acid protein that regulates the CRP transcription factor. Proc. Natl. Acad. Sci. USA 119. e2119866119. https://doi.org/10.1073/pnas.2119866119.

66. Wang, H., Yin, X., Wu Orr, M., Dambach, M., Curtis, R., and Storz, G. (2017). Increasing intracellular magnesium levels with the 31-amino acid MgsT protein. Proc. Natl. Acad. Sci. USA 114, 5689–5694. https://doi.org/10.1073/pnas.1703151114.

67. Sberro, H., Fremin, B.J., Zilni, S., Eders, F., Greenfield, N., Snyder, M.P., Pavlopoulos, G.A., Kyripcs, N.C., and Bhatt, A.S. (2019). Large-scale Analyses of human microbiomes reveal thousands of small. Cell 178, 1245–1259.e14. https://doi.org/10.1016/j.cell.2019.07.016.

68. Gray, T., Storz, G., and Papenfort, K. (2022). Small proteins; big questions. J. Bacteriol. 204, e0034121. https://doi.org/10.1128/JB.00341-21.

69. Graef, M., Straub, D., Eggen, T., Dole, U., Rodrigues, V., Brandt, R., and Wenkel, S. (2016). MicroProtein-mediated recruitment
of CONSTANS into a TOPLESS trimer complex represses flowering in arabidopsis. PLoS Genet. 12, e1005959. https://doi.org/10.1371/journal.pgen.1005959.

70. Hong, S.Y., Sun, B., Straub, D., Blaksemmer, A., Mineri, L., Koch, J., Brinch-Pedersen, H., Holme, I.B., Burow, M., Lyngs Jørgensen, H.-J., et al. (2020). Heterologous microprotein expression identifies LITTLE NINJA, a dominant regulator of jasmonic acid signaling. Proc. Natl. Acad. Sci. USA 117, 26197–26205. https://doi.org/10.1073/pnas.2005198117.

71. Bhati, K.K., Dolde, U., and Wenkel, S. (2021). Microproteins: expanding functions and novel modes of regulation. Mol. Plant 14, 709–728. https://doi.org/10.1016/j.molp.2021.01.006.

72. 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. https://doi.org/10.1126/science.1188158.

73. Ray, S., Rosenberg, M.I., Chanut-Delalande, H., Decaras, A., Schwertner, B., Toubiana, W., Auman, T., Schnellhammer, J., Teuscher, M., Valenti, P., et al. (2019). The mlpt/UBR3/Sivb module comprises an ancient developmental switch for embryonic patterning. Elife 8, e39768. https://doi.org/10.7554/eLife.39768.

74. Ho, L., Tan, S.Y.X., Wee, S., Wu, Y., Tan, S.J.C., Ramakrishna, N.B., Chng, S.C., Nama, S., Szcerbinska, I., Szcerbinski, I., et al. (2015). ELABELA is an endogenous growth factor that sustains hESC self-renewal via the PI3K/AKT pathway. Cell Stem Cell 17, 435–447. https://doi.org/10.1016/j.stem.2015.08.010.

75. Fu, H., Wang, T., Kong, X., Yan, K., Yang, Y., Cao, J., Yuan, Y., Wang, N., Kee, L., Z.J., and Xi, Q. (2022). A Nodal enhanced patterning. Elife 2021, 105547, December 22, 2022.

76. Yue, H., Chen, C., Li, H., Cheng, C., Xie, M., Yang, Y., and Xie, Z. (2021). Tissue- and stage-specific landscape of the mouse translatome. Nucleic Acids Res. 49, 6156–6180. https://doi.org/10.1093/nar/gkaa482.

77. Jain, S., Gentsch, G.E., Spruce, T., Monteiro, R.S., Owens, N.D.L., Martin, S.R., and Smith, J.C. (2018). Innate immune response and off-target mis-spacing are common morpholino-induced side effects in Xenopus. Dev. Cell 44, 597–610.e10. https://doi.org/10.1016/j.devcel.2018.01.022.

78. Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., et al. (2015). Discovery and functional characterization of diverse class 2 CRISPR-cas systems. Mol. Cell 59, 385–397. https://doi.org/10.1016/j.molcel.2015.10.008.

79. Abudayyeh, O.O., Gootenberg, J.S., Joung, J., Slak, M., Cox, D.B.T., Smith, S., Makarova, K.S., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 353, aaf5733. https://doi.org/10.1126/science.aaf5733.

80. Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdel, V., Cox, D.B.T., Keller, M.J., Regev, A., et al. (2017). RNA targeting with CRISPR-Cas13. Nature 550, 260–264. https://doi.org/10.1038/nature24049.

81. East-Selletsky, A., O’Connell, M.R., Burstein, D., Knott, G.J., and Doudna, J.A. (2017). RNA targeting by functionally orthogonal type VI-A CRISPR-cas enzymes. Mol. Cell 66, 373–383.e3. https://doi.org/10.1016/j.molcel.2017.04.008.

82. East-Selletsky, A., O’Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H.D., Tjian, R., and Doudna, J.A. (2016). Two distinct RNA activities of CRISPR-Cas22 enable guide RNA processing and RNA detection. Nature 538, 270–273. https://doi.org/10.1038/nature19802.

83. Smargon, A.A., Cox, D.B.T., Pyzocha, N.K., Zheng, K., Slak, M., Gootenberg, J.S., Abudayyeh, O.A., Essletzbichler, P., Shmakov, S., Makarova, K.S., et al. (2017). Cas13a is a type VI-B CRISPR-associated RNA-guided RNAse differentially regulated by accessory proteins Cas27 and Cas28. Mol. Cell 65, 618–630.e7. https://doi.org/10.1016/j.molcel.2016.12.023.

84. Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., et al. (2017). Diversity and evolution of class
2 CRISPR-Cas systems. Nat. Rev. Microbiol. 15, 169–182. https://doi.org/10.1038/nrmicro.2016.184.

99. Yan, W.X., Chong, S., Zhang, H., Makarova, K.S., Koonin, E.V., Cheng, D.R., and Scott, D.A. (2018). Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. Mol. Cell 70, 327–339.e5. https://doi.org/10.1016/j.molcel.2018.02.028.

100. Konermann, S., Lofty, P., Brideau, N.J., Oki, J., Shokhirev, M.N., and Hsu, P.D. (2018). Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. Cell 173, 665–676.e14. https://doi.org/10.1016/j.cell.2018.02.033.

101. Kushawah, G., Hernandez-Huertas, L., Albugattas-Nuñez Del Prado, J., Martinez-Morales, J.R., DeVore, M.L., Hassan, H., Moreno-Sanchez, I., Tomas-Gallardo, L., Diaz-Moscoso, A., Monges, D.E., et al. (2020). CRISPR-Cas13d induces efficient mRNA knockdown in animal embryos. Dev. Cell 54, 805–817.e7. https://doi.org/10.1016/j.devcel.2020.07.013.

102. Hernandez-Huertas, L., Kushawah, G., Diaz-Moscoso, A., Tomas-Gallardo, L., Moreno-Sanchez, I., da Silva Pescador, G., Bazzini, A.A., and Moreno-Mateos, M.A. (2022). Optimized CRISPR-RxiCas13d system for RNA targeting in zebrafish embryos. STAR Protoc. 3, 101058. https://doi.org/10.1016/j.xpro.2021.101058.

103. Ai, Y., Liang, D., and Wilusz, J.E. (2022). CRISPR/Cas13 effectors have differing extents of off-target effects that limit their utility in eukaryotic cells. Nucleic Acids Res. 50, e65. https://doi.org/10.1093/nar/gkac159.

104. Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A., and Schier, A.F. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res. 22, 577–591. https://doi.org/10.1101/gr.133009.111.

105. Hyatt, T.M., and Ekker, S.C. (1999). Vectors and techniques for ectopic gene expression in zebrafish. Methods Cell Biol. 59, 117–126. https://doi.org/10.1016/s0091-679x(08)61823-3.

106. Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development (Cambridge, U. K.) 145, dev165753. https://doi.org/10.1242/dev.165753.

107. Bogdanovic, O., Fernandez-Mihlan, A., Tena, J.J., de la Calle-Musientes, E., and Gomez-Skarmeta, J.L. (2013). The developmental epigenomics toolbox: ChIP-seq and MethylCap-seq profiling of early zebrafish embryos. Methods 62, 207–215. https://doi.org/10.1016/j.ymer.2013.04.011.

108. Doganli, C., Sandoval, M., Thomas, S., and Hart, D. (2017). Assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-Seq) protocol for zebrafish embryos. Methods Mol. Biol. 1507, 59–66. https://doi.org/10.1007/978-1-4939-6518-2_5.

109. Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.D., Fernandez, J.P., Mis, E.K., Khokha, M.K., and Giraldez, A.J. (2015). CRISPRscans: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat. Methods 12, 982–988. https://doi.org/10.1038/nmeth.3543.

110. Vejnar, C.E., Moreno-Mateos, M.A., Cifuentes, D., Bazzini, A.A., and Giraldez, A.J. (2016). Optimization strategies for the CRISPR-cas9 genome-editing system. Cold Spring Harb. Protoc. 2016 pdb.top09894. https://doi.org/10.1101/pdb.top09894.

111. Autran, D., Baroux, C., Raisig, M.T, Lenormand, T., Wittig, M., Grob, S., Steimer, A., Barann, M., Klostermeier, U.C., Leblanc, O., et al. (2011). Maternal epigenetic pathways control parental contributions to Arabidopsis early embryogenesis. Cell 145, 707–719. https://doi.org/10.1016/j.cell.2011.04.014.

112. Nodine, M.D., and Bartel, D.P. (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature 482, 94–97. https://doi.org/10.1038/nature10756.

113. Zhao, P., Zhou, X., Shen, K., Liu, Z., Cheng, T., Liu, D., Cheng, Y., Peng, X., and Sun, M.X. (2019). Two-step maternal-to-zygotic transition with two-phase parental genome contributions. Dev. Cell 49, 882–893.e5. https://doi.org/10.1016/j.devcel.2019.04.016.

114. Otoupal, P.B., Cress, B.F., Doudna, J.A., and Schoeniger, J.S. (2022). CRISPR-RNAa: targeted activation of translation using dCas13 fusions to translation initiation factors. Nucleic Acids Res. 50, 8986–8998. https://doi.org/10.1093/nar/gkac680.

115. Yang, L.Z., Wang, Y., Li, S.Q., Yao, R.W., Luan, P.F., Wu, H., Carmichael, G.G., and Chen, L.L. (2019). Dynamic imaging of RNA in living cells by CRISPR-cas13 systems. Mol. Cell 78, 981–997.e7. https://doi.org/10.1016/j.molcel.2019.10.024.