

TOPIC PAGE

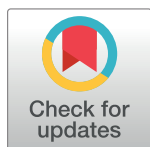
Micropeptide

Maria E. Sousa ^{1,2}, Michael H. Farkas ^{1,2,3*}

1 Ophthalmology, Jacobs School of Medicine and Biomedical Science, University of New York at Buffalo, Buffalo, NY, United States of America, **2** Research Service, Veterans Administration Western New York Healthcare System, Buffalo, NY, United States of America, **3** Department of Biochemistry, Jacobs School of Medicine and Biomedical Science, State University of New York at Buffalo, Buffalo, NY, United States of America

* mhfarkas@buffalo.edu

This is a Topic Page article for *PLOS Genetics*



Introduction

Micropeptides are commonly defined as [polypeptides](#) with a relatively arbitrary length of less than 100–150 [amino acids](#) (aa) in length [1–3]. They are distinguishable from bioactive peptides as the former is generated from short [open reading frames](#) (sORFs), whereas the latter is a cleavage product of a larger polypeptide [1,4]. They differ from canonical proteins which have an average length of 330 and 449 amino acids in prokaryotes and eukaryotes, respectively [5]. Micropeptides, also known as microproteins or sORF-encoded peptides, can also be named according to their genomic location. For example, the translated product of an upstream open reading frame (uORF) might be called a uORF-encoded peptide (uPEP) [6]. Micropeptides lack an N-terminal signaling sequence, suggesting they are likely to be localized to the cytoplasm after translation [1]. However, as more micropeptides are studied, they have been found in other cell compartments, as indicated by the existence of transmembrane micropeptides [7,8]. They are expressed in both prokaryotic and eukaryotic organisms. [1,9,10]. The sORFs from which micropeptides are translated can be encoded in a variety of genomic regions. These include 5' UTRs, small genes, polycistronic mRNAs, or genes originally characterized as long non-coding RNAs (lncRNAs) (Fig 1) [11].

Given their small size, sORFs were originally overlooked. However, hundreds of thousands of putative micropeptides have been identified through various techniques in a multitude of organisms. Only a small fraction of these with coding potential have had their expression and function confirmed. Those that have been functionally characterized, in general, have roles in cell signaling, organogenesis, and cellular physiology. As more micropeptides are discovered so are more of their functions. One regulatory function is that of peptoswitches, which inhibit expression of downstream coding sequences by stalling ribosomes, through their direct or indirect activation by small molecules [11].

Techniques for identifying potential micropeptides

Various experimental techniques exist for identifying sORFs and their translational products. It should be noted that these techniques are only useful for identification of sORF that may produce micropeptides and not for direct functional characterization.

OPEN ACCESS

Citation: Sousa ME, Farkas MH (2018) Micropeptide. *PLoS Genet* 14(12): e1007764. <https://doi.org/10.1371/journal.pgen.1007764>

Published: December 13, 2018

Copyright: © 2018 Sousa, Farkas. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The study was supported by funding from the National Eye Institute (R01EY028553) to MHF. The funder had no role in the preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Wikipedia Version: <https://en.wikipedia.org/wiki/Micropeptide>.

RNA-Sequencing

One method for finding potential sORFs, and therefore micropeptides, is through RNA-sequencing (RNA-Seq). RNA-Seq uses next-generation sequencing (NGS) to determine which RNAs are expressed in a given cell, tissue, or organism at a specific point in time. This collection of data, known as a **transcriptome**, can then be used as a resource for finding potential sORFs [1]. Because of the strong likelihood of sORFs less than 100 aa occurring by chance, further study is necessary to determine the validity of data obtained using this method [11].

Ribosome profiling (Ribo-Seq)

Ribosome profiling has been used to identify potential micropeptides in a growing number of organisms, including fruit flies, zebrafish, mice and humans [11]. One method uses compounds such as harringtonine, puromycin or lactimidomycin to stop ribosomes at translation initiation sites [12]. This indicates where active translation is taking place. Translation elongation inhibitors, such as emetine or cycloheximide, may also be used to obtain ribosome footprints which are more likely to result in a translated ORF [13]. If a ribosome is bound at or near a sORF, it putatively encodes a micropeptide [1,2,14].

Mass spectrometry

Mass Spectrometry (MS) is the gold standard for identifying and sequencing proteins. Using this technique, investigators are able to determine if polypeptides are, in fact, translated from a sORF.

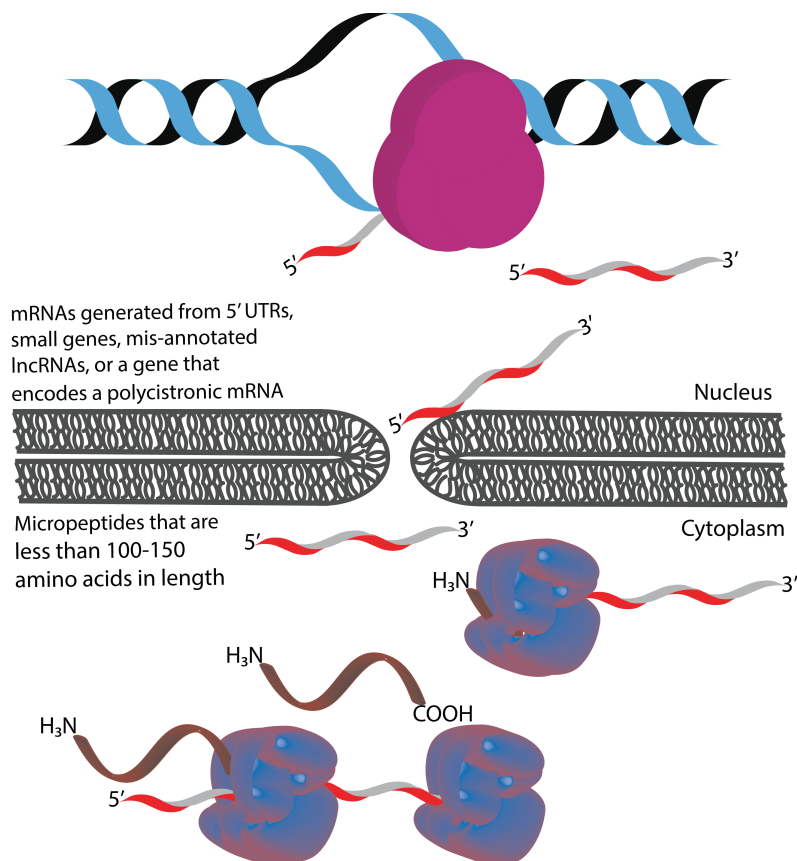


Fig 1. Genetic sources of micropeptides. Micropeptides can be transcribed from a small gene, a polycistronic mRNA, or from a lncRNA.

<https://doi.org/10.1371/journal.pgen.1007764.g001>

Proteogenomic applications

Proteogenomics combines proteomics, genomics, and transcriptomics. This is important when looking for potential micropeptides. One method of using proteogenomics entails using RNA-Seq data to create a custom database of all possible polypeptides. Liquid chromatography followed by tandem MS (LC-MS/MS) is performed to provide sequence information for translation products. Comparison of the transcriptomic and proteomics data can be used to confirm the presence of micropeptides [1,2].

Phylogenetic conservation

Phylogenetic conservation can be a useful tool, particularly when sifting through a large database of sORFs. The likelihood of a sORF resulting in a functional micropeptide is more likely if it is conserved across numerous species [11,12]. However, this will not work for all sORFs. For example, those that are encoded by lncRNAs are less likely to be conserved given lncRNAs themselves do not have high sequence conservation [2]. Further experimentation will be necessary to determine if a functional micropeptide is in fact produced.

Validating protein-coding potential

Antibodies

Custom **antibodies** targeted to the micropeptide of interest can be useful for quantifying expression or determining intracellular localization. As is the case with most proteins, low expression may make detection difficult. The small size of the micropeptide can also lead to difficulties in designing an epitope from which to target the antibody [2].

Tagging with CRISPR-Cas9

Genome editing can be used to add FLAG/MYC or other small peptide tags to an endogenous sORF, thus creating fusion proteins. In most cases, this method is beneficial in that it can be performed more quickly than developing a custom antibody. It is also useful for micropeptides for which no epitope can be targeted [2].

In vitro translation

This process entails cloning the full-length micropeptide cDNA into a plasmid containing a T7 or SP6 promoter. This method utilizes a cell-free protein-synthesizing system in the presence of ³⁵S-methionine to produce the peptide of interest. The products can then be analyzed by gel electrophoresis and the ³⁵S-labeled peptide is visualized using autoradiography [2].

Databases and repositories

There are several repositories and databases that have been created for both sORFs and micropeptides.

A repository for of small ORFs discovered by ribosome profiling can be found at www.sorfs.org

A repository of putative sORF-encoded peptides in *Arabidopsis thaliana* can be found at ARA-PEPs

A database of small proteins, especially encoded by non-coding RNAs can be found at SmProt

Prokaryotic examples

To date, most micropeptides have been identified in [prokaryotic](#) organisms. While most have yet to be fully characterized, of those that have been studied, many appear to be critical to the survival of these organisms. Because of their small size, prokaryotes are particularly susceptible to changes in their environment, and as such have developed methods to ensure their existence.

Escherichia coli (*E. coli*)

Micropeptides expressed in *E. coli* exemplify bacterial environmental adaptations. Most of these have been classified into three groups: leader peptides, ribosomal proteins, and toxic proteins. Leader proteins regulate transcription and/or translation of proteins involved in amino acid metabolism when amino acids are scarce. Ribosomal proteins include L36 (*rpmJ*) and L34 (*rpmH*), two components of the 50S ribosomal subunit. Toxic proteins, such as *ldrD*, are toxic at high levels and can kill cells or inhibit growth, which functions to reduce the host cell's viability [15].

Salmonella enterica (*S. enterica*)

In *S. enterica*, the MgtC virulence factor is involved in adaptation to low magnesium environments. The hydrophobic peptide MgrR, binds to MgtC, causing its degradation by the FtsH protease [9].

Bacillus subtilis (*B. subtilis*)

The 46 aa Sda micropeptide, expressed by *B. subtilis*, represses sporulation when replication initiation is impaired. By inhibiting the histidine Kinase KinA, Sda prevents the activation of the transcription factor Spo0A, which is required for sporulation [10].

Staphylococcus aureus (*S. aureus*)

In *S. aureus*, there are a group of micropeptides, 20–22 aa, that are excreted during host infection to disrupt neutrophil membranes, causing cell lysis. These micropeptides allow the bacterium to avoid degradation by the human immune systems' main defenses [16,17].

Eukaryotic examples

Micropeptides have been discovered in [eukaryotic](#) organisms from *Arabidopsis thaliana* to humans. They play diverse roles in tissue and organ development, as well as maintenance and function once fully developed. While many are yet to be functionally characterized, and likely more remain to be discovered, below is a summary of recently identified eukaryotic micropeptide functions.

Arabidopsis thaliana (*A. thaliana*)

The *POLARIS* (*PLS*) gene encodes a 36 aa micropeptide. It is necessary for proper vascular leaf patterning and cell expansion in the root. This micropeptide interacts with developmental PIN proteins to form a critical network for hormonal crosstalk between auxin, ethylene, and cytokinin [18,19,20].

ROTUNDIFOLIA (*ROT4*) in *A. thaliana* encodes a 53 aa peptide, which localizes to the plasma membrane of leaf cells. The mechanism of *ROT4* function is not well understood, but

mutants have short rounded leaves, indicating that this peptide may be important in leaf morphogenesis [21].

Zea mays (Z. mays)

Brick1 (Brk1) encodes a 76 aa micropeptide, which is highly conserved in both plants and animals. In *Z. mays*, it was found to be involved in morphogenesis of leaf epithelia, by promoting multiple actin-dependent cell polarization events in the developing leaf epidermis [22]. Zm401p10 is an 89 aa micropeptide, which plays a role in normal pollen development in the tapetum. After mitosis it also is essential in the degradation of the tapetum [23]. Zm908p11 is a micropeptide 97 aa in length, encoded by the *Zm908* gene that is expressed in mature pollen grains. It localizes to the cytoplasm of pollen tubes, where it aids in their growth and development [24].

Drosophila melanogaster (D. melanogaster)

The evolutionarily conserved polished rice (*pri*) gene, known as *tarsal-less (tal)* in *D. melanogaster*, is involved in epidermal differentiation. This polycistronic transcript encodes four similar peptides, which range between 11–32 aa in length. They function to truncate the transcription factor Shavenbaby (Svb). This converts Svb into an activator that directly regulates the expression of target effectors, including *miniature (m)* and *shavenoid (sha)*, which are together responsible for trichome formation [25].

Danio rerio (D. rerio)

The *Toddler (tdl)* gene is believed to be important for embryogenesis, and is specifically expressed during late blastula and gastrula stages. During gastrulation, it is critical in promoting the internalization and animal-pole directed movement of mesendodermal cells. After gastrulation, it is expressed in the lateral mesoderm, endoderm, as well as the anterior, and posterior, notochord. Although it is annotated as a lncRNA in zebrafish, mouse, and human, the 58 aa sORF was found to be highly conserved in vertebrates [26,27].

Mus musculus (M. musculus)

Myoregulin (Mln) is encoded by a gene originally annotated as a lncRNA. Mln is expressed in all 3 types of skeletal muscle, and works similarly to the micropeptides phospholamban (Pln) in the cardiac muscle and sarcolipin (Sln) in slow (Type I) skeletal muscle. These micropeptides interact with sarcoplasmic reticulum Ca^{2+} -ATPase (SERCA), a membrane pump responsible for regulating Ca^{2+} uptake into the sarcoplasmic reticulum (SR). By inhibiting Ca^{2+} uptake into the SR, they cause muscle relaxation. Similarly, the endoregulin (ELN) and another-regulin (ALN) genes code for transmembrane micropeptides that contain the SERCA binding motif, and are conserved in mammals [7].

Myomixer (Mymx) is encoded by the gene *Gm7325*, a muscle-specific peptide, 84 aa in length, which plays a role during embryogenesis in fusion and skeletal muscle formation. It localizes to the plasma membrane, associating with a fusogenic membrane protein, Myomaker (Mymk). In humans, the gene encoding Mymx is annotated as uncharacterized *LOC101929726*. Orthologs are found in the turtle, frog and fish genomes as well [8].

Homo sapiens (H. sapiens)

In humans, NoBody (non-annotated P-body dissociating polypeptide), a 68 aa micropeptide, was discovered in the long intervening noncoding RNA (lincRNA) *LINC01420*. It has high

sequence conservation among mammals, and localizes to P-bodies. It enriches proteins associated with 5' mRNA decapping. It is thought to interact directly with Enhancer of mRNA Decapping 4 (EDC4) [28].

The *C7orf49* gene, conserved in mammals, when alternatively spliced is predicted to produce three micropeptides. MRI-1 was previously found to be a modulator of retrovirus infection. The second predicted micropeptide, MRI-2, may be important in non-homologous end joining (NHEJ) of DNA double strand breaks. In Co-Immunoprecipitation experiments, MRI-2 bound to Ku70 and Ku80, two subunits of Ku, which play a major role in the NHEJ pathway [29].

The 24 amino acid micropeptide, Humanin (HN), interacts with the apoptosis-inducing protein Bax (Bcl2-associated X protein). In its active state, Bax undergoes a conformational change which exposes membrane-targeting domains. This causes it to move from the cytosol to the mitochondrial membrane, where it inserts and releases apoptogenic proteins such as cytochrome c. By interacting with Bax, HN prevents Bax targeting of the mitochondria, thereby blocking apoptosis [30].

A micropeptide of 90aa, 'Small Regulatory Polypeptide of Amino Acid Response' or SPAAR, was found to be encoded in the lncRNA *LINC00961*. It is conserved between human and mouse, and localizes to the late endosome/lysosome. SPAR interacts with four subunits of the v-ATPase complex, inhibiting mTORC1 translocation to the lysosomal surface where it is activated. Down-regulation of this micropeptide enables mTORC1 activation by amino acid stimulation, promoting muscle regeneration [31].

References

1. Crappé J, Van Crielinge W, Menschaert G. Little things make big things happen: a summary of micropeptide encoding genes. *EuPA Open Proteomics*. 2014 Jun 1; 3:128–37.
2. Makarewich CA, Olson EN. Mining for Micropeptides. *Trends in cell biology*. 2017 Sep 1; 27(9):685–96. <https://doi.org/10.1016/j.tcb.2017.04.006> PMID: 28528987
3. Guillén G, Díaz-Camino C, Loyola-Torres CA, Aparicio-Fabre R, Hernández-López A, Díaz-Sánchez M, Sanchez F. Detailed analysis of putative genes encoding small proteins in legume genomes. *Frontiers in plant science*. 2013 Jun 20; 4:208. <https://doi.org/10.3389/fpls.2013.00208> PMID: 23802007
4. Hashimoto Y, Kondo T, Kageyama Y. Lilliputians get into the limelight: novel class of small peptide genes in morphogenesis. *Development, growth & differentiation*. 2008 Jun; 50:S269–76.
5. Zhang J. Protein-length distributions for the three domains of life. *Trends in Genetics*. 2000 Mar 1; 16(3):107–9. PMID: 10689349
6. Rothnagel J, Menschaert G. Short Open Reading Frames and Their Encoded Peptides. *Proteomics*. 2018 May; 18(10):1700035.
7. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*. 2015 Feb 12; 160(4):595–606. <https://doi.org/10.1016/j.cell.2015.01.009> PMID: 25640239
8. Bi P, Ramirez-Martinez A, Li H, Cannavino J, McAnally JR, Shelton JM, et al. Control of muscle formation by the fusogenic micropeptide myomixer. *Science*. 2017 Apr 21; 356(6335):323–7. <https://doi.org/10.1126/science.aam9361> PMID: 28386024
9. Alix E, Blanc-Potard AB. Peptide-assisted degradation of the Salmonella MgtC virulence factor. *The EMBO journal*. 2008 Feb 6; 27(3):546–57. <https://doi.org/10.1038/sj.emboj.7601983> PMID: 18200043
10. Burkholder WF, Kurtser I, Grossman AD. Replication initiation proteins regulate a developmental checkpoint in *Bacillus subtilis*. *Cell*. 2001 Jan 26; 104(2):269–79. PMID: 11207367
11. Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. *Nature Reviews Genetics*. 2014 Mar; 15(3):193. <https://doi.org/10.1038/nrg3520> PMID: 24514441
12. Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *The EMBO Journal*. 2014 Apr 4; e201488411.

13. Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJ, Jackson SE, Wills MR, Weissman JS. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell reports*. 2014 Sep 11; 8(5):1365–79. <https://doi.org/10.1016/j.celrep.2014.07.045> PMID: 25159147
14. Stern-Ginossar N, Ingolia NT. Ribosome Profiling as a Tool to Decipher Viral Complexity. *Annual Review of virology*. 2015 Nov 9; 2:335–49.
15. Hemm MR, Paul BJ, Schneider TD, Storz G, Rudd KE. Small membrane proteins found by comparative genomics and ribosome binding site models. *Molecular microbiology*. 2008 Dec; 70(6):1487–501. <https://doi.org/10.1111/j.1365-2958.2008.06495.x> PMID: 19121005
16. Wang R, Braughton KR, Kretschmer D, Bach TH, Queck SY, Li M, et al. Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. *Nature medicine*. 2007 Dec; 13(12):1510. <https://doi.org/10.1038/nm1656> PMID: 17994102
17. Hemm MR, Paul BJ, Miranda-Ríos J, Zhang A, Soltanzad N, Storz G. Small stress response proteins in *Escherichia coli*: proteins missed by classical proteomic studies. *Journal of bacteriology*. 2010 Jan 1; 192(1):46–58. <https://doi.org/10.1128/JB.00872-09> PMID: 19734316
18. Casson SA, Chilley PM, Topping JF, Evans IM, Souter MA, Lindsey K. The POLARIS gene of *Arabidopsis* encodes a predicted peptide required for correct root growth and leaf vascular patterning. *The Plant Cell*. 2002 Aug 1; 14(8):1705–21. <https://doi.org/10.1105/tpc.002618> PMID: 12172017
19. Chilley PM, Casson SA, Tarkowski P, Hawkins N, Wang KL, Hussey PJ, et al. The POLARIS peptide of *Arabidopsis* regulates auxin transport and root growth via effects on ethylene signaling. *The Plant Cell*. 2006 Nov 1; 18(11):3058–72. <https://doi.org/10.1105/tpc.106.040790> PMID: 17138700
20. Liu J, Mehdi S, Topping J, Friml J, Lindsey K. Interaction of PLS and PIN and hormonal crosstalk in *Arabidopsis* root development. *Frontiers in plant science*. 2013 Apr 5; 4:75. <https://doi.org/10.3389/fpls.2013.00075> PMID: 23577016
21. Narita NN, Moore S, Horiguchi G, Kubo M, Demura T, Fukuda H, et al. Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in *Arabidopsis thaliana*. *The Plant Journal*. 2004 May; 38(4):699–713. <https://doi.org/10.1111/j.1365-313X.2004.02078.x> PMID: 15125775
22. Frank MJ, Smith LG. A small, novel protein highly conserved in plants and animals promotes the polarized growth and division of maize leaf epidermal cells. *Current Biology*. 2002 May 14; 12(10):849–53. PMID: 12015123
23. Wang D, Li C, Zhao Q, Zhao L, Wang M, Zhu D, et al. Zm401p10, encoded by an anther-specific gene with short open reading frames, is essential for tapetum degeneration and anther development in maize. *Functional Plant Biology*. 2009 Jan 30; 36(1):73–85.
24. Dong X, Wang D, Liu P, Li C, Zhao Q, Zhu D, et al. Zm908p11, encoded by a short open reading frame (sORF) gene, functions in pollen tube growth as a profilin ligand in maize. *Journal of experimental botany*. 2013 May 1; 64(8):2359–72. <https://doi.org/10.1093/jxb/ert093> PMID: 23676884
25. Kondo T, Plaza S, Zanet J, Benrabah E, Valenti P, Hashimoto Y, et al. Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science*. 2010 Jul 16; 329(5989):336–9. <https://doi.org/10.1126/science.1188158> PMID: 20647469
26. Pauli A, Norris ML, Valen E, Chew GL, Gagnon JA, Zimmerman S, et al. Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science*. 2014 Feb 14; 343(6172):1248636 <https://doi.org/10.1126/science.1248636> PMID: 24407481
27. Chng SC, Ho L, Tian J, Reversade B. ELABELA: a hormone essential for heart development signals via the apelin receptor. *Developmental cell*. 2013 Dec 23; 27(6): 672–680. <https://doi.org/10.1016/j.devcel.2013.11.002> PMID: 24316148
28. D'Lima NG, Ma J, Winkler L, Chu Q, Loh KH, Corpuz EO, et al. A human microprotein that interacts with the mRNA decapping complex. *Nature chemical biology*. 2017 Feb; 13(2):174–80. <https://doi.org/10.1038/nchembio.2249> PMID: 27918561
29. Slavoff SA, Heo J, Budnik BA, Hanakahi LA, Saghatelian A. A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining. *Journal of Biological Chemistry*. 2014 Mar 7; jbc-C113.
30. Guo B, Zhai D, Cabezas E, Welsh K, Nouraini S, Satterthwait AC, Reed JC. Humanin peptide suppresses apoptosis by interfering with Bax activation. *Nature*. 2003 May; 423(6938):456. <https://doi.org/10.1038/nature01627> PMID: 12732850
31. Matsumoto A, Pasut A, Matsumoto M, Yamashita R, Fung J, Monteleone E, et al. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature*. 2017 Jan; 541(7636):228. <https://doi.org/10.1038/nature21034> PMID: 28024296