Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Found in translation: functions and evolution of a recently discovered alternative proteome

Christian R Landry[1,2,3], Xiangfu Zhong[1,2,3], Lou Nielly-Thibault[1,2,3] and Xavier Roucou[3,4]

A major goal in biology is to map entire proteomes to better understand the biology and evolution of cells. However, our current views of proteomes are conservative and biased against small proteins. Besides serendipitous discoveries of small proteins, it has been largely assumed that eukaryotic mature mRNAs contain a single ORF and that non-coding RNAs are not translated because their ORFs are too short to play a functional role. A flurry of recent studies brought to light an unexplored proteome that is mainly translated from short ORFs in non-coding regions and from alternative ORFs (AltORFs) in reference genes. The detection of these small proteins and the elucidation of their functions remain challenging and open a new dimension of eukaryotic proteomes, including the birth of novel genes and proteins.

Addresses
[1] Département de Biologie, Université Laval, Québec, Québec, Canada
[2] Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec, Canada
[3] PROTEO, Quebec Network for Research on Protein Function, Structure, and Engineering, Québec, Québec, Canada
[4] Département de Biochimie, Université de Sherbrooke, Sherbrooke, Québec, Canada

Corresponding author: Landry, Christian R (Christian.landry@bio.ulaval.ca)

Current Opinion in Structural Biology 2015, **32**:74–80

This review comes from a themed issue on **Sequences and topology**

Edited by **M Madan Babu** and **Anna R Panchenko**

http://dx.doi.org/10.1016/j.sbi.2015.02.017

Introduction

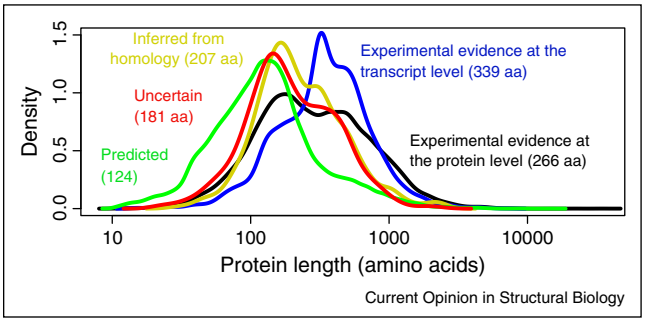
Understanding how information flows from the genome to cellular phenotypes is one of the grand challenges of biology. This challenge includes the need for the full description of the protein content of cells, the proteome. Studying the diversity of proteins produced from a given genome is difficult because one needs to know what to look for, as their discovery is largely based on mass spectrometry analyses in which peptides need to be predicted before they are detected. This approach is thus biased towards what we assume are *bona fide* protein-coding genes. Typically, computational methods that predict protein coding sequences use a cutoff of 100 codons as minimum length [1] and predict a single most probable protein per mRNA. Here we review recent work challenging these conventions [2••,3] and suggesting that a large fraction of eukaryotic proteomes has remained invisible. These studies have shown that functional proteins can be of small size, which means they can be encoded in parts of the genome that have been neglected so far in proteomic studies. This neglected proteome includes short translated ORFs from supposedly non-coding RNAs and untranslated regions (UTRs) of transcripts and located in alternative open reading frames within canonical mRNAs. We finally discuss the evolutionary and functional implications of these observations.

Big roles for small proteins

The current size criterion of 100 codons for the prediction of protein coding genes has been justified for practical reasons. First, a small fraction of known proteins are smaller than 100 amino acids (aa) (Figure 1), although there is no actual break in the distribution at this size [4]. Second, the fraction of functional ORFs diminishes with diminishing ORF size because spurious ORFs are expected to occur at high frequency as ORF sizes decrease. In addition, this small size makes other parameters that can be used to define real protein coding sequences [1] difficult to estimate with accuracy because of small sampling size. This stringent cutoff on ORF length has undoubtedly created a strong bias against the discovery of small proteins. Nevertheless, studies have provided evidence for key roles of small proteins [2••]. In mammals, chemokines, whose sizes vary between 8 and 10 kDa, play roles in leukocyte chemotaxis, development, angiogenesis, tumorigenesis, metastasis and immune response [5]. A 24 aa protein known as humanin has an anti-apoptotic function [6] and may have a neuroprotective role in Alzheimer's disease [7]. The 91 aa human protein ISD11 is essential for iron homeostasis and mutations in this gene result in serious oxidative phosphorylation deficiency [8]. Other conserved small proteins include 30–50 aa proteins that regulate the heart muscle contraction in insects and humans [9••], embryogenesis in zebrafish and *Drosophila* [10,11] and morphogenesis in plants [12••].

The functions of small proteins are often discovered fortuitously and elucidating their molecular functions

Figure 1



Distribution of protein length in the human proteome. In black are proteins detected experimentally (mass spec, antibodies, etc.); in blue are proteins predicted and for which a transcript has been detected (cDNA, Northern blots); in yellow are proteins predicted because an homolog exists in another species; in green are proteins for which there is no evidence at the level of the protein, transcript or homology. Uncertain cases are where the existence of the protein is not sure. Numbers in parenthesis represent the median length for each category. Data was obtained from Uniprot in November 2014 (ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/).

remains difficult. Investigating structure–function relationships and searching for protein domains are hampered by their small size. Although many small proteins may fold into specific structures, they are less likely to contain a specific domain, or have a catalytic activity. However, the absence of a stable structure does not indicate a lack of function, as observed for intrinsically disordered proteins, which do play important roles [13]. Despite these difficulties, a trend seems to have emerged from functional studies, as many appear to regulate other proteins and their interactions. For instance, the yeast small protein Hug1 is intrinsically disordered and regulates ribonucleotide reductase activity by modulating the assembly of the heterodimeric enzyme [14]. The human 24-aa humanin exerts its effect by binding to a heterotrimeric receptor [7] and regulates Bax during apoptosis by interacting and preventing Bax's translocation from the cytosol to mitochondria. The small protein MRI-2 also acts in a similar fashion in DNA repair [15]. Similar small proteins that contain protein–protein interaction domains and that regulate the assembly of functional protein complexes have been termed microproteins, by analogy with the action of microRNAs on mRNAs [16,17]. Microproteins may be as small as 34 aa long and block the assembly of functional transcription complexes [18]. Future functional studies will determine whether the regulatory role of small proteins in the assembly and function of protein complexes is a common mode of action.

Short protein-coding ORFs are abundant in eukaryotic genomes and are translated

If small proteins indeed play important roles, functionally important ORFs could be found almost anywhere within the transcribed genome [4]. Early RNA sequencing studies have revealed thousands of long RNAs (more than 200 nucleotide long, lncRNAs) transcribed in eukaryotic genomes that were initially considered as non-coding because they did not contain ORFs of significant enough length and of noticeable conservation to be considered as coding. However, many of those may actually be translated if size is not such an important criterion for function [19,20••]. In a recent large-scale study, Ruiz-Orera *et al.* [21••] analyzed ribosome profiling experiments from mouse, human, zebrafish, *Drosophila*, *Arabidopsis* and yeast. They found that most lncRNAs contained at least one ORF of 24 codons or longer and that these were associated with ribosomes in proportions ranging from 29% in yeast to 77% in mouse. Some of these RNAs showed higher ribosome occupancy than annotated coding RNAs. Other recent studies reached similar conclusions [19,22–24,25•,26,27]: many presumably non-coding RNAs are bound to ribosomes in patterns that are consistent with the production of proteins between 20 and 100 aa long, which were in many cases confirmed by biochemical methods [25•] or peptide evidence [21••]. However, because ribosome binding by RNAs may not necessarily lead to the production of a protein, let alone a functional one (see Section 'Biological noise or functions'), direct approaches such as those based on mass spectrometry are going to play a key role in directly exploring eukaryotic proteomes.

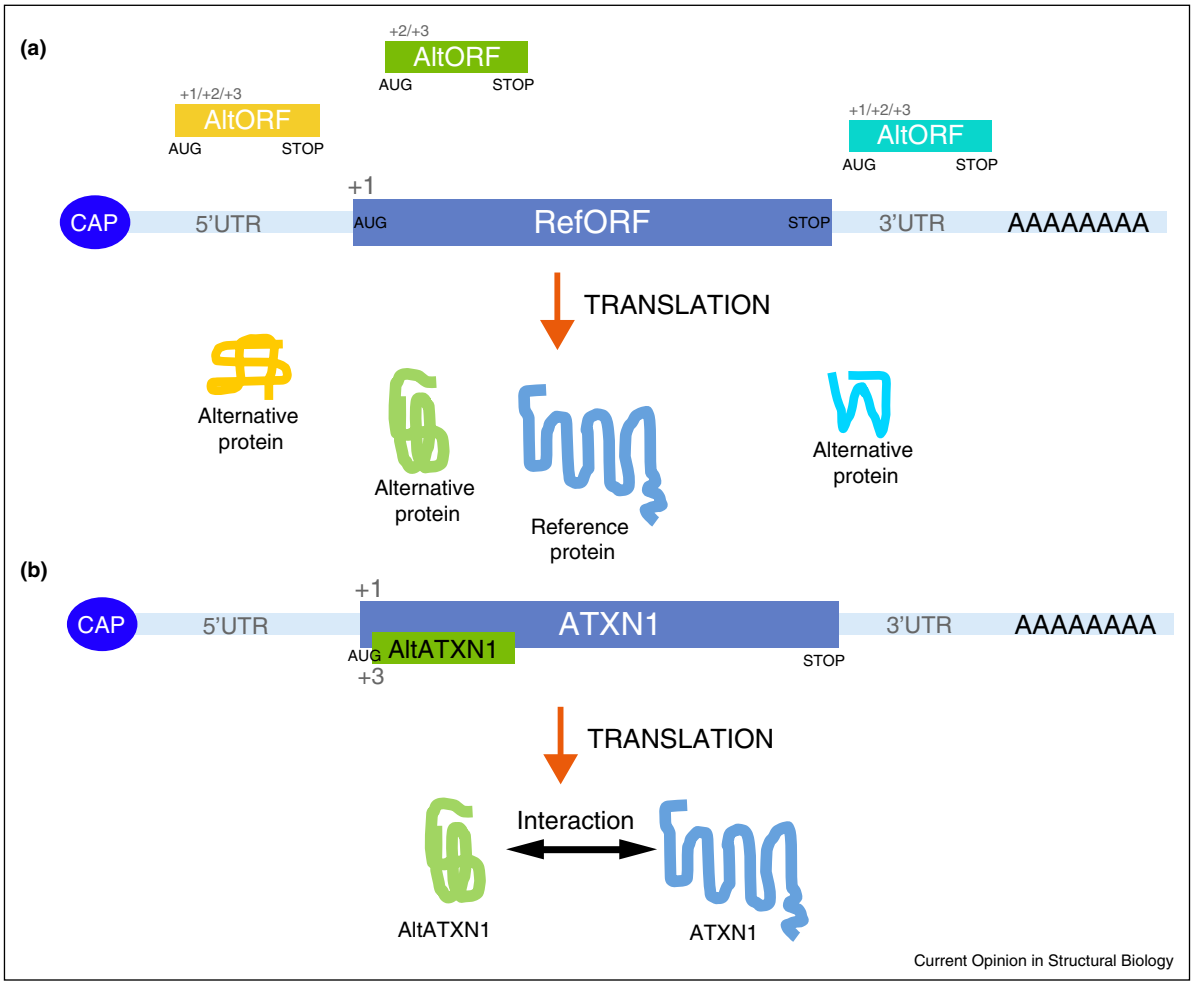
Small proteins are difficult to detect by standard methods for several reasons. First, low molecular weight molecules are lost and/or underrepresented during protein isolation [28]. Second, a small number of unique peptides are expected to be detected for any given small protein. However, one major limitation up until recently was that peptides that would result from the translation of non-coding RNAs were absent from reference databases because these typically focus on annotated transcripts [3]. Relaxing these criteria did in fact lead to the discovery of more small proteins. For instance, Slavoff *et al.* [28] looked for small proteins that could be translated from the entire human transcriptome, irrespective of whether the transcripts were previously known to encode proteins. They identified 37 translated small ORFs located in 5′ and 3′ UTRs, non-coding RNA or antisense transcripts, and 11 located inside an annotated protein coding sequence but in a different reading frame. The majority of these proteins were less than 100 aa long but surprisingly, their concentration was within the range of typical cellular proteins. The observations were further substantiated by Kim *et al.* [29•], who produced a map of the human proteome using more than 30 different tissues. Sixteen million MS/MS spectra that did not initially match annotated genes were analyzed using a comprehensive database that included exhaustive ORF prediction from the genome. More than 800 novel annotations were identified, including the translation of 140 pseudogenes and

44 novel ORFs. An additional 430 peptides were identified from intergenic lncRNAs by Wilhelm *et al.* [30•], none of which having sequence homology with UniProt sequences. Again, the distribution of the abundance for these peptides covered the same range as those of canonical proteins.

Alternative ORFs (AltORFs) within mRNAs or multiple proteins encoded in a single transcript

As illustrated above, non-coding RNAs are a significant source of small ORFs and thus of small proteins. However, canonical coding mRNAs also contain a large number of small AltORFs (Figure 2a), which are different from isoforms translated from alternative transcripts. These AltORFs are encoded partially or entirely outside of the canonical coding sequence within the transcript and/or in a different reading frame. Vanderperre *et al.* [31••,32] generated a new database for mass spectrometry that contains AltORFs present in NCBI's RefSeq human mRNA database. This database led to the description of 83,886 putative AltORFs with a minimum length of 40 codons [31••] and roughly 4 AltORFs per mRNA. Vanderperre *et al.* and others identified more than 1200 proteins corresponding to these AltORFs in different human cell lines and tissues [28,31••,33,34•,35], nine of which were subsequently confirmed in an independent study [29•]. A large proportion of such AltORFs were previously predicted by ribosomal profiling, further supporting their translation

Figure 2



(a) Typical mature eukaryotic mRNA and its possible AltORFs. The reference ORF (RefORF) is defined as the coding ORF in the canonical +1 reading frame, and is annotated in current nucleotide databases. An alternative ORF (AltORF) is defined as a nucleotide region comprised between a start and a stop codon (different from that of the RefORF) and is predicted to encode an alternative protein. AltORFs may be localized inside the 5′ UTR, inside the RefORF, inside the 3′ UTR. AltORFs may also partially overlap an UTR and the RefORF (not shown here for clarity purposes). In contrast to reference proteins, alternative proteins are not annotated in current protein sequence databases used in proteomics studies and therefore not search for. **(b)** ATXN1 is a dual-coding transcript. The AltATXN1 coding sequence (distinct reading frame) overlaps the ATXN1 N-terminal coding sequence and interacts with the ATXN1 N-terminal domain. ATXN1 also controls the subcellular distribution of AltATXN1 [35].

[36,37]. More than 40% of predicted and detected AltORFs are encoded inside RefORFs but in a different reading frame [31••]. These AltORFs would go unnoticed in protein expression studies with plasmids containing RefORFs, unless antibodies were developed to detect the co-expression of the corresponding alternative proteins. This also means that the knockdown of the RefORFs potentially diminishes the function of a second gene product simultaneously.

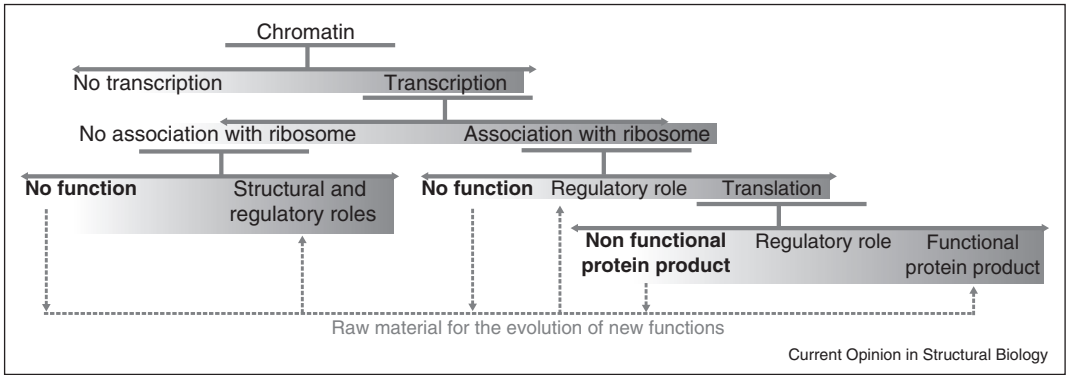
As for small proteins in general, alternative proteins are challenging to study experimentally. However, experimental studies have been conducted on few of these AltORFs and they suggest specific molecular activities, such as the physical interactions between the reference protein and an alternative protein of the same gene. For instance, the XXLαs transcript produced from the GNAS1 gene encodes XXLαs, an N-terminally extended G-protein α-subunit [38,39]. The XXL exon contains two overlapping ORFs encoding both the XXL domain of XXLαs (+1 reading frame) and a second protein termed AlexX (+2 reading frame). Surprisingly, not only is AlexX co-expressed with XXLαs, but it also interacts with the XXL domain [38]. A similar case was observed for ATXN1, which encodes the ataxin-1 (ATXN1) protein implicated in spinocerebellar ataxia type 1 [35,40] (Figure 2b). These observations suggest that functional interactions between proteins encoded by different ORFs in the same mRNA may be a widespread mechanism for effective regulation of biological pathways, although more data is necessary to test this prediction.

Biological noise or functions

Ribosomal profiling and footprinting suggest the presence of more small proteins from non-coding RNAs and cryptic proteins than what mass spectrometry and other methods can detect. This discrepancy can be explained. First, tools to detect small and cryptic proteins have not been developed, including antibodies and updated protein databases containing these cryptic protein sequences. Second, mass spectrometry experiments often lack the sensitivity or technical details that would allow detecting all proteins in a sample [28]. Third, the association of ribosomes with RNA may represent no function or lead to the production of small non-functional proteins that are rapidly degraded and thus not detectable. This possibility demands that we formally and quantitatively investigate the potential role of small ORFs in non-coding RNAs or AltORFs in mRNAs.

Short ORFs are likely to appear often by chance alone given a random DNA sequence, with a small fraction only becoming associated with a function and being retained by natural selection. We therefore expect many short ORFs in a genome at mutation-drift equilibrium, i.e. even in conditions where natural selection plays a minor role. The analysis of small proteins and short ORFs therefore needs to consider different scenarios in the interpretation of large-scale data (Figure 3). We argue

Figure 3



Functional roles of transcripts and proteins. Each of the processes from transcription to translation may include non-functional products. The simplest scenario is that short and AltORFs encode peptides that fulfill important functions. The second possibility is that short and AltORFs do not encode functional peptides but their translation plays a role. For instance, it was shown that the translation of short ORFs in 5′ UTR (uORF) may regulate protein expression by triggering mRNA decay or by regulating the translation of the major ORF [47•]. The association with ribosomes may reflect the general function of the ribosome but no actual function for a given RNA molecule. For instance ribosome association with capped transcripts could be a default status and in the absence of selection for the use of specific start codon, translation is initiated at any site in the 5′ region [48]. Finally, short and AltORFs may derive from evolutionary noise that is inherent to biological systems [49], that is, the accumulation of ORFs in transcribed regions and that have features that leads to their translation [48]. Ribosomes are extremely abundant in the cell and could thus bind and actively translate ORFs even if they do not have the optimal sequences for initiating translation. In addition, ORFs with start and stop codons can appear by mutation and drift alone in the gene regions that are transcribed and since a large fraction of the genome is transcribed, these offer a template for translation and production of non-functional peptides. This biological noise may however serve as raw material for the evolution of novel functions (arrows feeding back to the process) [20••]. Not discussed here are false positive identifications in proteomics experiments whereby translation products that do not exist are detected by the instruments. This is another layer of concern that could be added to this complex picture.

that, in the absence of additional data, the non-functionality of short translation products should be considered as the null hypothesis. In all cases, evolutionary comparative analyses can be used to try to estimate the functional constraints acting on the short ORFs, as is done for other proteomic features [41,42]. Because functional elements tend to be preserved throughout evolution by negative selection against alternative sequences, they are expected to be more ancient than neutral ones appearing by mutation and drift alone. One limitation of this type of analysis is that short ORFs that only recently evolved functions show a very limited phylogenetic distribution [43•,44] and will thus fail at this test, in which case the study of within species polymorphism may be more useful to detect purifying selection. The analysis of conservation is also particularly complex for AltORFs because the overlap with a functional ORF itself may create an evolutionary constraint that will maintain the AltORFs despite the fact that it may have no function. It is therefore important to consider other possible sources of constraint in these analyses. Nevertheless, although non-functional translated ORFs and proteins could be considered as a nuisance for biologists, it may represent a potential transition towards the evolution of novel protein coding-genes. Indeed, the origin of genes and proteins *de novo* may require to start with the translation of a RNA molecule before it can be acted upon by natural selection and further improved [20••,45] (Figure 3).

Conclusion

Contrary to widespread assumptions, proteins can be of very small size and yet play important molecular, cellular and physiological roles. This vision, along with the recent findings described above, opens new avenues in the study of proteome complexity, function and evolution. We identify three important challenges in this new field. A first challenge will be to determine if general rules can be established in terms of the molecular function of small proteins, which often lack defined structures. Another will be to determine how translation is regulated in different ORFs from the same transcript [46], which may in many cases have independent functions. Finally, another key challenge will be to build better models of how these short ORFs and AltORFs come about during evolution. This will allow to identify the functional ORFs more rapidly and to determine how they could contribute to the emergence of *de novo* genes and thus of novel functions. These challenges will be met by a better integration of cell and evolutionary biology.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgements

We thank D. Hunting, N. Aubin-Horth, J.B. Leducq, A.E. Chrétien, M. Filteau and G. Charron for comments on the manuscript. CRL's research is funded by CIHR Grants 324265 and 299432 and a NSERC Discovery grant. CRL is a FRSQ Junior II investigator. XZ is partially supported by a PROTEO Fellowship. CRL's and XR's work on alternative proteins is supported by a FRQNT Team Grant. XR's research is supported by CIHR Grants MOP-137056 and MOP-136962.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Basrai MA, Hieter P, Boeke JD: **Small open reading frames: beautiful needles in the haystack**. *Genome Res* 1997, **7**:768-771.

2. •• Andrews SJ, Rothnagel JA: **Emerging evidence for functional peptides encoded by short open reading frames**. *Nat Rev Genet* 2014, **15**:193-204.
Thorough and insightful review of the role and function of small proteins across model organisms. Provides a list of functional small proteins that have well-established roles.

3. Nesvizhskii AI: **Proteogenomics: concepts, applications and computational strategies**. *Nat Methods* 2014, **11**:1114-1125.

4. Frith MC, Forrest AR, Nourbakhsh E, Pang KC, Kai C, Kawai J, Carninci P, Hayashizaki Y, Bailey TL, Grimmond SM: **The abundance of short proteins in the mammalian proteome**. *PLoS Genet* 2006, **2**:e52.

5. Le Y, Zhou Y, Iribarren P, Wang J: **Chemokines and chemokine receptors: their manifold roles in homeostasis and disease**. *Cell Mol Immunol* 2004, **1**:95-104.

6. 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, **423**:456-461.

7. Takeshita Y, Hashimoto Y, Nawa M, Uchino H, Matsuoka M: **SH3-binding protein 5 mediates the neuroprotective effect of the secreted bioactive peptide humanin by inhibiting c-Jun NH2-terminal kinase**. *J Biol Chem* 2013, **288**:24691-24704.

8. Lim SC, Friemel M, Marum JE, Tucker EJ, Bruno DL, Riley LG, Christodoulou J, Kirk EP, Boneh A, DeGennaro CM *et al.*: **Mutations in LYRM4, encoding iron-sulfur cluster biogenesis factor ISD11, cause deficiency of multiple respiratory chain complexes**. *Hum Mol Genet* 2013, **22**:4460-4473.

9. •• Magny EG, Pueyo JI, Pearl FM, Cespedes MA, Niven JE, Bishop SA, Couso JP: **Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames**. *Science* 2013 http://dx.doi.org/10.1126/science.1238802.
Identification of two proteins of less than 30 aa that regulate the *Drosophila* heart through the regulation of calcium transport. These proteins appear to be conserved in humans.

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

11. Pauli A, Norris ML, Valen E, Chew GL, Gagnon JA, Zimmerman S, Mitchell A, Ma J, Dubrulle J, Reyon D *et al.*: **Toddler: an embryonic signal that promotes cell movement via Apelin receptors**. *Science* 2014, **343**:1248636.

12. •• Hanada K, Higuchi-Takeuchi M, Okamoto M, Yoshizumi T, Shimizu M, Nakaminami K, Nishi R, Ohashi C, Iida K, Tanaka M *et al.*: **Small open reading frames associated with morphogenesis are hidden in plant genomes**. *Proc Natl Acad Sci U S A* 2013, **110**:2395-2400.
Comprehensive analysis of the transcription profile of small ORFs (30–100 codons) in the Arabidopsis genome and transgenic analysis to show than 10% induce visible phenotypes when overexpressed.

13. Oldfield CJ, Dunker AK: **Intrinsically disordered proteins and intrinsically disordered protein regions**. *Annu Rev Biochem* 2014, **83**:553-584.

14. Meurisse J, Bacquin A, Richet N, Charbonnier JB, Ochsenbein F, Peyroche A: **Hug1 is an intrinsically disordered protein that inhibits ribonucleotide reductase activity by directly binding**

Rnr2 subunit. *Nucleic Acids Res* 2014 http://dx.doi.org/10.1093/nar/gku1095.

15. Slavoff SA, Heo J, Budnik BA, Hanakahi LA, Saghatelian A: **A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining**. *J Biol Chem* 2014, **289**:10950-10957.

16. Staudt AC, Wenkel S: **Regulation of protein function by 'microProteins'**. *EMBO Rep* 2011, **12**:35-42.

17. Fricker LD: **Analysis of mouse brain peptides using mass spectrometry-based peptidomics: implications for novel functions ranging from non-classical neuropeptides to microproteins**. *Mol Biosyst* 2010, **6**:1355-1365.

18. Magnani E, de Klein N, Nam HI, Kim JG, Pham K, Fiume E, Mudgett MB, Rhee SY: **A comprehensive analysis of microProteins reveals their potentially widespread mechanism of transcriptional regulation**. *Plant Physiol* 2014, **165**:149-159.

19. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS: **Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling**. *Science* 2009, **324**:218-223.

20. •• Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charloteaux B, Hidalgo CA, Barbette J, Santhanam B *et al.*: **Proto-genes and de novo gene birth**. *Nature* 2012, **487**:370-374.
Provides a thorough analysis of *de novo* evolved protein coding genes and proposes a model by which there is a continuum between non-coding to coding sequences in which evolution exploits what may be considered as dispensable sequences to created new protein functions.

21. •• Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM: **Long non-coding RNAs as a source of new peptides**. *Elife* 2014, **3**:e03523.
Comprehensive analysis of ribosome association with lncRNAs in six model organisms that suggests the translation of many short peptides.

22. Crappe J, Van Criekinge W, Trooskens G, Hayakawa E, Luyten W, Baggerman G, Menschaert G: **Combining in silico prediction and ribosome profiling in a genome-wide search for novel putatively coding sORFs**. *BMC Genomics* 2013, **14**:648.

23. Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC *et al.*: **Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation**. *EMBO J* 2014, **33**:981-993.

24. Prabakaran S, Hemberg M, Chauhan R, Winter D, Tweedie-Cullen RY, Dittrich C, Hong E, Gunawardena J, Steen H, Kreiman G *et al.*: **Quantitative profiling of peptides from RNAs classified as noncoding**. *Nat Commun* 2014, **5**:5429.

25. • Smith JE, Alvarez-Dominguez JR, Kline N, Huynh NJ, Geisler S, Hu W, Coller J, Baker KE: **Translation of small open reading frames within unannotated RNA transcripts in *Saccharomyces cerevisiae***. *Cell Rep* 2014, **7**:1858-1866.
Comprehensive demonstration that many unannotated RNA stranscripts produce small proteins in the budding yeast and that many are conserved.

26. Juntawong P, Girke T, Bazin J, Bailey-Serres J: **Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis**. *Proc Natl Acad Sci U S A* 2014, **111**:E203-E212.

27. Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MA, Brocard M, Couso JP: **Extensive translation of small open reading frames revealed by Poly-Ribo-Seq**. *Elife* 2014, **3**:e03528.

28. Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A: **Peptidomic discovery of short open reading frame-encoded peptides in human cells**. *Nat Chem Biol* 2013, **9**:59-64.

29. • Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S *et al.*: **A draft map of the human proteome**. *Nature* 2014, **509**:575-581.
The authors produce a comprehensive map of the human protein using a proteogenomic analysis and identify several unknown proteins from several tissues, allowing for a better annotation of the human genome.

30. • Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H *et al.*: **Mass-spectrometry-based draft of the human proteome**. *Nature* 2014, **509**:582-587.
This study reports a massive database of protein expression in human tissues and several proteins expressed from non-coding parts of the genome.

31. •• Vanderperre B, Lucier JF, Bissonnette C, Motard J, Tremblay G, Vanderperre S, Wisztorski M, Salzet M, Boisvert FM, Roucou X: **Direct detection of alternative open reading frames translation products in human significanly expands the proteome**. *PLoS One* 2013.
The authors predict new protein-coding sequences termed alternative open reading frames in different genomes, show significant evolutionary conservation in vertebrates and invertebrates, and detect these novel proteins by mass spectrometry.

32. Vanderperre B, Lucier JF, Roucou X: **HAltORF: a database of predicted out-of-frame alternative open reading frames in human**. *Database (Oxford)* 2012, **2012**:bas025.

33. Menschaert G, Van Criekinge W, Notelaers T, Koch A, Crappe J, Gevaert K, Van Damme P: **Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events**. *Mol Cell Proteomics* 2013, **12**:1780-1790.

34. • Vanderperre B, Staskevicius AB, Tremblay G, McCoy M, O'Neill MA, Cashman NR, Roucou X: **An overlapping reading frame in the PRNP gene encodes a novel polypeptide distinct from the prion protein**. *FASEB J* 2011, **25**:2373-2386.
The first example of an overlapping open reading frame encoding a novel protein in a gene linked to neurodegenerative disorders.

35. Bergeron D, Lapointe C, Bissonnette C, Tremblay G, Motard J, Roucou X: **An out-of-frame overlapping reading frame in the Ataxin-1 coding sequence encodes a novel Ataxin-1 interacting protein**. *J Biol Chem* 2013, **288**:21824-21835.

36. Ingolia NT, Lareau LF, Weissman JS: **Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes**. *Cell* 2011, **147**:789-802.

37. Michel AM, Choudhury KR, Firth AE, Ingolia NT, Atkins JF, Baranov PV: **Observation of dually decoded regions of the human genome using ribosome profiling data**. *Genome Res* 2012, **22**:2219-2229.

38. Klemke M, Kehlenbach RH, Huttner WB: **Two overlapping reading frames in a single exon encode interacting proteins — a novel way of gene usage**. *EMBO J* 2001, **20**:3849-3860.

39. Abramowitz J, Grenet D, Birnbaumer M, Torres HN, Birnbaumer L: **XLalphas, the extra-long form of the alpha-subunit of the Gs G protein, is significantly longer than suspected, and so is its companion Alex**. *Proc Natl Acad Sci U S A* 2004, **101**:8366-8371.

40. Zoghbi HY, Orr HT: **Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, spinocerebellar ataxia type 1**. *J Biol Chem* 2009, **284**:7425-7429.

41. Levy ED, Michnick SW, Landry CR: **Protein abundance is key to distinguish promiscuous from functional phosphorylation based on evolutionary information**. *Philos Trans R Soc Lond B Biol Sci* 2012, **367**:2594-2606.

42. Freschi L, Osseni M, Landry CR: **Functional divergence and evolutionary turnover in mammalian phosphoproteomes**. *PLoS Genet* 2014, **10**:e1004062.

43. • Zhao L, Saelao P, Jones CD, Begun DJ: **Origin and spread of de novo genes in *Drosophila melanogaster* populations**. *Science* 2014, **343**:769-772.
The authors identify de novo genes as they are originating and spreading in natural populations.

44. Tautz D, Domazet-Loso T: **The evolutionary origin of orphan genes**. *Nat Rev Genet* 2011, **12**:692-702.

45. Cai J, Zhao R, Jiang H, Wang W: **De novo origination of a new protein-coding gene in *Saccharomyces cerevisiae***. *Genetics* 2008, **179**:487-496.

46. Xu H, Wang P, Fu Y, Zheng Y, Tang Q, Si L, You J, Zhang Z, Zhu Y, Zhou L *et al.*: **Length of the ORF, position of the first AUG and the Kozak motif are important factors in potential dual-coding transcripts**. *Cell Res* 2010, **20**:445-457.

47. • Barbosa C, Peixeiro I, Romao L: **Gene expression regulation by upstream open reading frames and human disease**. *PLoS Genet* 2013, **9**:e1003529.

The authors revisit ribosomal profiling data and analyses to show that ribosome association with long non-coding RNAs may not produce functional proteins for the most part.

48. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES: **Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins**. *Cell* 2013, **154**:240-251.

49. Landry CR, Levy ED, Abd Rabbo D, Tarassov K, Michnick SW: **Extracting insight from noisy cellular networks**. *Cell* 2013, **155**:983-989.