Combining Ribosome Profiling and Proteomics to Discover Micropeptides, Translation Products from Small Open Reading Frames.

Volodimir Olexiouk[a], Jeroen Crappé[a], Steven Verbruggen[a], Kenneth Verheggen[b], Gerben Menschaert[a], Wim Van Criekinge[a]

[a] Lab of Bioinformatics and Computational Genomics (BioBix), Department of Mathematical Modelling, Statistics and Bioinformatics, Faculty of Bioscience Engineering, Ghent University

[b] VIB Department of Medical Protein Research, UGent

SORFS.ORG

Overview

- Identification of small open reading frames (sORFs) through ribosome profiling
- Generation of additional in silico metrics to predict the coding potential of sORFs
- Automated search for mass spectrometry evidence using PRIDE and peptideshaker
- Providing a public repository for sORFs, facilitating research in the micropeptide field

Introduction

Even though no consensus has been reached, ***small open reading frames*** (sORFs) can be defined as open reading frames smaller than 100 amino acids.These “sORFs” are inherent to all genomes, but are historically ignored to have any coding potentia. A major contributor to this historical ignorance of functional sORFs is represented by the taken evolutionary trajectory of different tools in the field of bioinformatics/genomics/proteomics that were designed to reduce noise, neglecting the identification of these sORFs as a side effect. However, “recent” scientific breakthroughs[1][2] have discovered coding potential in several sORFs with clinical significance, indicating their importance. In particular, the invention of ***ribosome profiling***[3] (RIBO-seq), a next generation sequencing technique, providing a genome-wide snapshot of the translation machinery has provided great contributions to the discovery of sORFs.

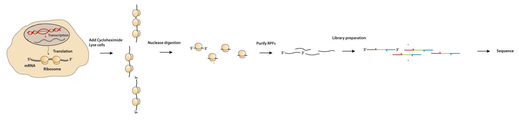
While RIBO-seq provides data on many putative translating sORFs, ribosomal activity does not always point to functionality of the peptides. Currently only a handful of sORFs are shown to produce functional micropeptides. To close the gap between sORFs and micropeptides several in silico tools have been designed measuring the ***coding potential*** of sORFs. For instance, analyzing the ribosome protected fragments, cross-species conservation and phenotype associated variations can provide an indication of the coding potential. The most straightforward way to measure the coding potential of sORFs is by acquiring ***MS-based evidence***. In order to facilitate micropeptide research, a ***public platform*** (**sorfs.org**) has been created combining ribosome profiling data, metrics depending the coding potential of sORFs and proteomics evidence.

Methods

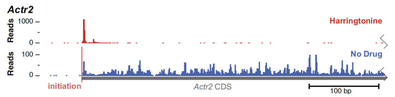
Ribosome profiling (RIBO-seq) captures and subsequently sequences the +/-30bp RNA- fragments captured within ribosomes (the protein translation machinery). This technique differs from a regular RNA-seq setup, as a ‘snap-shot’ is provided of what is currently being ***translated*** in a cell, rather than what is expressed in a cell. In this context, it provides the opportunity to detect small open reading frames (sORFs) that are being translated and possibly could encode functional peptides or small proteins [A].

Additionally ribosome profiling allow the identification of ***translation initiation sites*** (TIS). RIBO-seq can be performed in presence of different antibiotics, which can either accumulate ribosomes at the TIS or stall ribosome at their current position [B].

A

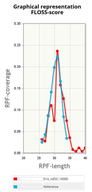


B



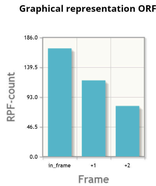
Results

C



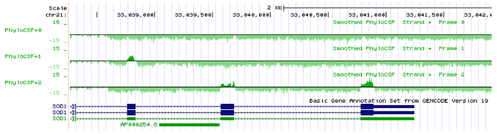
[C] FLOSS[4]: The FLOSS algorithm distinguises true coding from non-coding sequences based on the ***RPF-length distribution***.

D



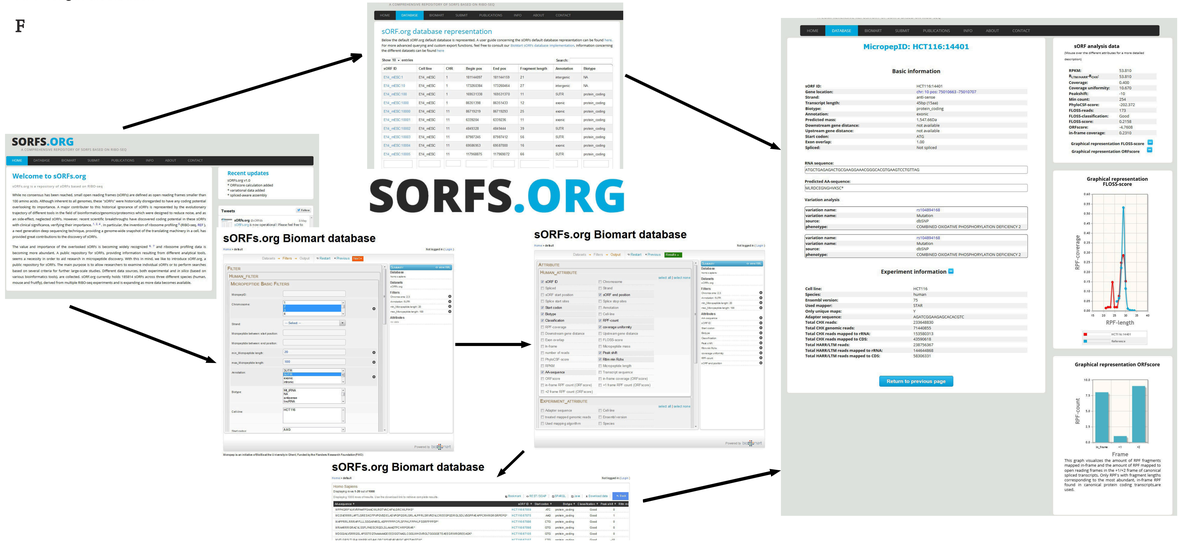
[D] ORFscore[5]: The ORFscore calculates the preference of RPFs to accumulate in the ***first frame*** of coding sequences.

E



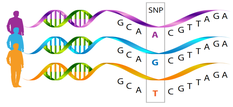
[E] PhyloCSF[6] :Cross species ***conservation*** is a general adopted technique in order to acquire evidence for genomic important regions.

F



[F] sORFs.org: A ***public repository*** for sORFs based on RIBO-seq. Providing a platform for scientist to inspect, query and retrieve information regarding sORFs.

G



[G] Variation analysis: Information embedded in phenotype related ***mutations, insertions and deletions*** provide an additional source for evidence in favor of the coding potential of sORFs.

H

[H] Automated PRIDE resprocessing[7][8][9]: The PRIDE[10] database is explored to find ***Mass spectrometry*** evidence for the translation of sORFs into functional micropeptides.

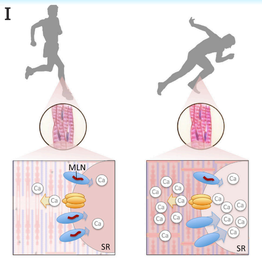
Conclusion & future perspective

The micropeptide research field has grown significantly, but still remains in its infancy. In order to take micropeptide research to the next level a public platform has been created, ***combining genomics and proteomics*** in order to facilitate micropeptide discovery.

Both RIBO-seq and micropeptide research shows an evident increase in interest, represented by the exponential increase in publications. As a result we expect to ***ellaborate*** on the number of datasets represented as well as the number of species supported as more data becomes available. ***Additional metrics*** concerning the coding potential of sORFs are currently being development alongside ***visual tools*** in order to facilitate manual inspection of sORFs. Undoubtedly more micropeptides will be discovered in the near function and we are confident that our public repository will contribute significantly.

Example micropeptide

I



[I] Myoregulin (MLN), a novel tissue-specific ***micropeptide*** has been discovered in long noncoding RNA by Anderson et. al. 2015[11]. Myoregulin influences the skeletal muscle calcium handling pathway in order to regulate muscle contraction.

References

[1] Kondo T., Plaza S., Zanet J., Benrabah E., Valenti P., Hashimoto Y., Kobayashi S., Payre F., Kageyama Y. (2010). Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. Science Jul 16 329(5989):336-9.

[2] Magny E., Pueyo J., Pearl F., Cespedes M., Niven J., Bishop S., Couso J. (2013). Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. Science. Sep 6 341(6150):1116-20.

[3] Ingolia N., Ghaemmaghami S., Newman J., Weissman J. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling.
Science Apr 10 324(5924):218-23

[4] Ingolia N., Brar G., Stern-Ginossar N., Harris M., Talhouarne G., Jackson S., Wills M., Weissman J. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. Cell Rep. Sep 11 8(5):1365-79

[5] Ariel B., Timothy J., Romain C., Sebastian M., Benedikt O., Elizabeth F., Charles V., Miler L., Nikolaus R., Tobias W., Antonio G. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. The EMBO Journal 33, 981-993

[6] Michael L., Irwin J., Manolis K., (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regionts.
Bioinformatics 27 (13): i275-i282.

[7] Hulstaert N., Reisinger F., Rameseder J., Barsnes H., Vizcaíno J., Martens L. (2013). Pride-asap: automatic fragment ion annotation of identified PRIDE spectra. Proteomics. Dec 16;95:89-92

[8] Vaudel M., Barsnes H., Berven F., Sickmann A., Martens L. (2011). SearchGUI: An open-source graphical user interface for simultaneous OMSSA and XTandem searches. Proteomics Mar;11(5):996-9.

[9] Marc V., Julia B., René Z., Eystein O., Frode B., Albert S., Lennart M., Harald B. (2015). PeptideShaker enables reanalysis of MS-derived proteomics data sets.
Nature Biotechnology 33, 22-24

[10] Juan A.V., Richard C., Attila C., José D., Antonio F., Joseph F., Johannes G., Emanuele A., Melih B., Javier C., Gavin O., Andreas S., David O., Yasset P., Florian R., Daniel R., Rui W. and Henning H. (2013). The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. Nucl. Acids Res. (1 January 2013) 41 (D1): D1063-D1069.

[11] Anderson D., Anderson K., Chang C., Makarewich C., Nelson B., McAnally J., Kasaragod P., Shelton J., Liou J., Bassel-Duby R., Olson E. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell. 2015 Feb 12;160(4):595-606.

Contact

Volodimir.Olexiouk@Ugent.be , Gerben.Menschaert@Ugent.be