Editorial

JOURNAL OF HEPATOLOGY

SMIM30, a tiny protein with a big role in liver cancer

Juan Pablo Unfried[1], Puri Fortes[1,2,3,*]

[1]University of Navarra (UNAV), Center for Applied Medical Research (CIMA), Program of Gene Therapy and Hepatology, Pamplona, Spain; [2]Navarra Institute for Health Research (IdiSNA), Pamplona, Spain; [3]Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Spain.

See Article, pages 1155–1169

One of the most remarkable discoveries from recent high-throughput transcriptomic analyses is that while most of our genome is transcribed, only a small percentage encodes for proteins. Consequently, compared to coding genes, there are an unexpectedly high number of genes that produce long non-coding RNAs (lncRNAs). Several lncRNAs have been described that interact with DNA, proteins or other RNAs to play essential functions in cell proliferation, differentiation, homeostasis, and signaling. Surprisingly, it was recently described that many lncRNAs are not as non-coding as previously thought, and some of them function by encoding small proteins. Although contradictory, the explanation is rather simple: small open reading frames (ORFs), which code for small proteins, were not considered as true ORFs for genomic annotations. Nowadays, in the genomics and transcriptomics era, it is still challenging to discriminate between true ORFs that code for proteins and those that occur randomly throughout the genome and that are not translated. This is especially arduous for short ORFs, as the likelihood that an ORF encodes a real protein increases with ORF length. These facts serve to underscore the relevance of the work by Pang and colleagues with the lncRNA-encoded micropeptide SMIM30, which contributes to hepatocellular carcinoma (HCC).[1]

To facilitate the discrimination between genuine and spurious ORFs, it was established that *bona fide* coding sequences should start with an AUG, be longer than 300 nucleotides, and be flanked by 5′ and 3′ untranslated regions (UTRs), leaving the ORF constrained into a monocistronic mRNA. These arbitrary criteria have governed protein identification for decades, causing the misannotation of many putative lncRNAs with ORFs shorter than the cut-off of 300 nucleotides. Recently, these rules have been questioned by ribosome profiling results.

Ribosome profiling is the deep sequencing of ribosome-protected RNA fragments and provides a genome-wide picture of active translation with single-nucleotide resolution.[2,3] The scrutiny of the frequency, period, and precise position of ribosomes on the RNA has revealed an unanticipated proteome complexity. Many annotated ORFs have amino-terminal extensions or truncations, and many non-annotated ribosome-bound ORFs are initiated with non-AUG codons or match with lncRNAs, as is the case of SMIM30, or with 5′ or 3′ UTR sequences, leading to polycistronic RNAs that may code for small proteins. However, not all translation produces stable and functional proteins. Instead, the simple act of translation may represent biological noise or may serve as a mechanism of regulation.[4,5]

As a result, additional ribosome profiling techniques are required to identify genuine short peptides. Mass spectrometry (MS) is one of them, as detection in MS data is considered a strong proof of micropeptide expression. On the contrary, since small and low abundant peptides are hard to detect by MS, the absence of a sequence in MS data is not evidence enough to conclude that the microprotein is not produced. In these cases, specific experiments such as *in vitro* translation, generation of specific antibodies, or tagging the endogenous peptide with CRISPR-Cas genome editing, can aid in validating the existence of the microprotein. Indeed, Pang *et al.*, used some of these methods to demonstrate that SMIM30 is a true microprotein. The combination of MS and transcriptomics (known as proteogenomics), with ribosome profiling and computational analyses that account for evolutionary conservation, signatures of purifying selection, and phylogenetic assessment of codon conservation patterns, have aided in the identification of thousands of microproteins in human cells.[6,7]

Subsequently, like microRNAs, microproteins have been overlooked until recently and now belong to a field of intensive research. Microproteins or micropeptides differ from canonical proteins in their length (less than 100–150 amino acids (aa) – SMIM30 is 59 aa long – instead of the average 450 aa of standard eukaryotic proteins). Microproteins do not result from the proteolytic cleavage of larger polypeptides. Instead, they are peptides translated from short ORFs located in sequences traditionally described as non-coding, including 5′ (upstream ORFs) and 3′ (downstream ORFs) UTRs, lncRNAs and circular RNAs.[8] In addition, they can also be generated from alternative frames within known coding sequences. Many are expected to be non-functional. Microproteins may be novel proteins, arising from recent evolutionary events that may not yet have a function, or may belong to the pervasive translation that generates the abundant raw material required for the generation of new

Received 23 June 2020; received in revised form 3 July 2020; accepted 6 July 2020; available online 22 August 2020

* DOI of original article: https://doi.org/10.1016/j.jhep.2020.05.028.

* Corresponding author. Address: Pio XII 55, 31008 Pamplona, Spain. Tel.: 34-948-194700-814025, fax: 34-948-194700.
E-mail address: pfortes@unav.es (P. Fortes).

https://doi.org/10.1016/j.jhep.2020.07.015

proteins. Microproteins may also be non-functional by-products of translation events required for the regulation of gene expression. However, their small size does not prevent many microproteins from playing relevant roles in cellular processes, including development, myoblast fusion, or calcium homeostasis, and in key features of tumor initiation and progression such as cell proliferation and death, response to stress, cell metabolism, DNA repair, inflammation, angiogenesis, gene expression regulation, or epithelial-to-mesenchymal transition.[9] These functions likely represent just the tip of the microproteome iceberg, since only a low percentage of the expected total number of microproteins has been well studied to date. This repertoire of functions is exerted by microproteins acting independently, as ligands or signaling molecules, or by interacting with and regulating large protein complexes. SMIM30 is an excellent prototype of this last group.

Microprotein translation is prominent in the liver, and there is great expectation that hepatic microproteins will play relevant roles in healthy liver homeostasis and the development of liver disease.[6] Preliminary studies on lncRNA-encoded micropeptides in human liver have shown that many classical lncRNAs can give rise to small peptides and that most localize to the mitochondrial membrane in different cell lines, potentially impacting energy metabolism and therefore, the development of HCC. PC3-secreted microprotein (PSMP), is one of the few well-characterized liver microproteins.[10] PSMP is expressed to high levels in patients and murine models of liver fibrosis and binds the chemokine receptor CCR2 to induce inflammation and fibrosis. Interestingly, as PSMP is secreted, it can be neutralized with specific antibodies to reduce liver fibrosis development *in vivo*. Pang and colleagues now shed light on the function of hepatic microproteins related to liver cancer.[1] The authors use RNA co-immunoprecipitation with the 40s ribosomal protein RPS6 to identify hundreds of ribosome-bound lncRNAs in liver, prostate, colon, and pancreatic cancer cells. Over 100 were identified in all cancer cell lines tested. Among them, the authors selected LINC00998 and showed that it is upregulated in HCC and cholangiocarcinoma compared to healthy tissue. Expression levels of LINC00998 correlate significantly with overall survival and disease-free survival in HCC. Indeed, upregulation may result from transcription induction by Myc, which binds to the promoter of LINC00998 (Fig. 1). Unlike most lncRNAs, LINC00998 is highly and ubiquitously expressed, accumulates in the cytoplasm, is conserved in mice, and has a single small ORF with positive coding potential. This was enough to convince Uniprot to reannotate LINC00998 as a new coding gene called SMIM30, from small integral membrane protein 30. Pang and colleagues had to follow all the required steps to demonstrate that SMIM30 encodes for a microprotein: *in vitro* translation experiments, generation and evaluation of antibodies and tandem MS analyses. Antibodies were used to show that the microprotein is highly expressed in primary HCC compared to healthy tissue. Knock-down experiments showed that LINC00998/SMIM30 induces cell cycle progression, proliferation, and migration in culture, and promotes HCC growth and metastasis in animal models. Pang *et al.* also show that the microprotein is functional while the former LINC00998 is not. This is not obvious, as several lncRNAs are bifunctional and possess both coding and non-coding roles.

Sequence analysis and localization experiments showed that SMIM30 is a transmembrane protein (Fig. 1). Remarkably, a large proportion of the microproteins studied to date are predicted to contain hydrophobic domains that drive their localization to the

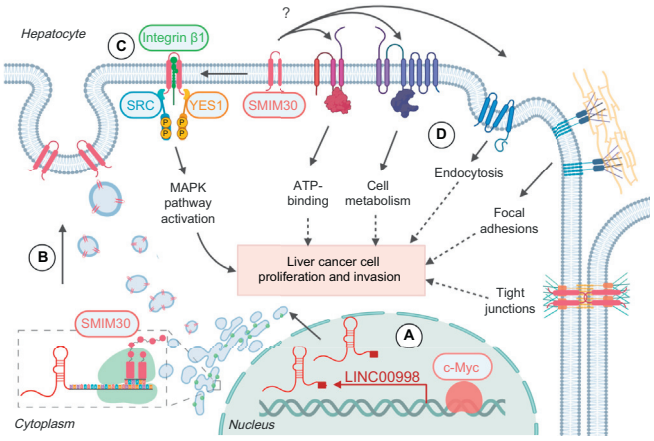

Fig. 1. Hypothetical model of SMIM30 function. (A) LINC00998/SMIM30 levels increase in response to Myc activation. (B) SMIM30 microprotein is translated and transported to the cell membrane. Note that SMIM30 processing is currently unknown, and it cannot be discarded that the mature form is a single-pass transmembrane microprotein. (C) SMIM30 binds SRC/YES1 and contributes to their interaction with integrin β1, membrane anchoring and the activation of the pathway, leading to MAPK signaling and hepatocyte proliferation. (D) SMIM30 may also interact with other protein complexes involved in cell metabolism, ATP regulation, endocytosis, and cell adhesion, which could also contribute to the protumoral role of SMIM30.

plasma membrane or the surface of different organelles, including mitochondria or endosomes.[11] It has been suggested that membrane localization makes microproteins especially suited to engage large complexes and to regulate complex biological systems. This could well be the case of SMIM30.

Co-immunoprecipitation experiments followed by MS and the required validations, show that SMIM30 interacts with the amino-terminal domain of protein kinases SRC/YES1. This domain associates with the cell membrane, where SMIM30 locates, suggesting that SMIM30 could recruit SRC/YES1 to the membrane and contribute to the interaction between SRC/YES1 and integrin β1. In agreement with this possibility, binding between SRC/YES1 and integrin β1 is weaker in the absence of SMIM30. Besides, SMIM30 contributes to the phosphorylation and activation of SRC/YES1 and the downstream induction of MAPK signaling, leading to cell proliferation. Given that SRC/YES1 are also essential liver regulators of bile flow and glucose production, it is reasonable to suspect that SMIM30 will also play a key role in liver homeostasis.[12,13]

Interestingly, the proteome bound by SMIM30 is larger than expected for such a small protein. Although further validation is required, many of the proteins identified by MS may contribute to the role of SMIM30 in liver homeostasis, cell proliferation and migration. Remarkably, several interacting proteins are from focal adhesion and cell tight junctions. Others are involved in endocytosis, metabolism, or ATP binding. It is currently unclear whether most of these proteins bind to SMIM30 as independent complexes, if the interaction is direct or if binding also contributes to membrane targeting. Increased levels of SMIM30 in HCC could modify the binding of these proteins to the cell membrane and alter protein functionality, contributing to the development and progression of HCC.

SMIM30 interactions should reach clinical relevance. Indeed, microproteins are prominent targets for therapeutic interventions. Neutralizing antibodies may efficiently block

secreted micropeptides, and their function may be enhanced with synthetic peptide mimics. Inside the cell, artificial microproteins may be introduced with cell-penetrating peptides. The binding domain of micropeptides such as SMIM30, a short domain that may interact with a large proteome, seems an excellent target for the development of small molecules with potent antitumor capacities. New studies are urgently required to decipher the role played by microproteins in liver disease and other pathologies and exploit the potential of these molecules towards clinical translation.

Financial support

Our research is funded by European FEDER funding and grants from the Ministry of Science and Innovation, Spain (SAF2015-70971-R; RTI2018-101759-B-I00); and Gobierno de Navarra (33/2015). CIBERehd is funded by the Instituto de Salud Carlos III.

Conflict of interest

The authors declare no conflicts of interest.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contributions

JPU and PF discussed the original draft, JPU designed the figure (using Biorender.com) and reviewed the manuscript, PF reviewed the figure and wrote the manuscript.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2020.07.015.

References

Author names in bold designate shared co-first authorship

[1] **Pang Y**, **Liu Z**, Han H, Wang B, Li W, Mao C, et al. Peptide SMIM30 promotes HCC development by inducing SRC/YES1 membrane anchoring and MAPK pathway activation. J Hepatol 2020;73(5):1155–1169.
[2] Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 2009;324:218–223.
[3] 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.
[4] Morris DR, Geballe AP. Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol 2000;20:8635–8642.
[5] Makarewich CA, Olson EN. Mining for micropeptides. Trends Cell Biol 2017;27:685–696.
[6] van Heesch S, Witte F, Schneider-Lunitz V, Schulz JF, Adami E, Faber AB, et al. The translational landscape of the human heart. Cell 2019;178:242–260.e29.
[7] Chen J, Brunner AD, Cogan JZ, Nuñez JK, Fields AP, Adamson B, et al. Pervasive functional translation of noncanonical human open reading frames. Science 2020;367:1140–1146.
[8] Ruiz-Orera J, Albà MM. Translation of small open reading frames: roles in regulation and evolutionary innovation. Trends Genet 2019;35:186–198.
[9] Merino-Valverde I, Greco E, Abad M. The microproteome of cancer: from invisibility to relevance. Exp Cell Res 2020;392:111997.
[10] **She S**, **Wu X**, Zheng D, Pei X, Ma J, Sun Y, et al. PSMP/MSMP promotes hepatic fibrosis through CCR2 and represents a novel therapeutic target. J Hepatol 2020;72:506–518.
[11] Makarewich CA. The hidden world of membrane microproteins. Exp Cell Res 2020;388:10–13.
[12] Byun S, Kim D-H, Ryerson D, Kim Y-C, Sun H, Kong B, et al. Postprandial FGF19-induced phosphorylation by Src is critical for FXR function in bile acid homeostasis. Nat Commun 2018;9:2590.
[13] **Louet J-F**, **Chopra AR**, Sagen JV, An J, York B, Tannour-Louet M, et al. The coactivator SRC-1 is an essential coordinator of hepatic glucose production. Cell Metab 2010;12:606–618.