

Into Uncharted Water: The Small Open Reading Frame-encoded Peptidome

Qian Zhao

State Key Laboratory of Chemical Biology and Drug Discovery, Department of Applied Biology and
Chemical Technology, The Hong Kong Polytechnic University, Hong Kong SAR, P. R. China

Email: q.zhao@polyu.edu.hk

Small open reading frames (sORFs) are novel coding DNA sequences that are shorter than 100 codons. They used to be considered non-coding or even junk. In recent years, accumulating evidence suggests that sORFs can encode microproteins or peptides with important functions. Given 98% of the human genome is still defined as non-coding regions, sORFs-encoded peptides (SEPs) are foreseeably an underexplored territory and gold mine. However, large-scale, and confident identification of SEPs have been technically challenging.

We have established a systematic approach to discover, quantify, and characterize novel SEPs. First, we analyzed samples with ribosome profiling to predict thousands of sORFs hidden in 5'UTR, 3'UTR and lncRNA, which showed high temporal and spatial specificity. Next, we detected SEPs by mass spectrometry with the following optimizations to improve the identification number and data reproducibility: 1) Peptide enrichment; 2) Tailored data acquisition and searching; 3) Customized sORF database. With our approach, thousands of novel SEPs were identified and quantified from cells and tissues. Some of the SEPs are associated with mouse embryonic development, while others play important functions in cancer progression, drug resistance, and immunotherapy. Our work not only provided a mass spectrometry platform to identify SEPs, but also, reported novel peptides with important biological functions.







	RNA type	sORF location
	uORF	5'UTR of CDS
	uORF extension	5'UTR and parts of CDS
	Frameshift	Multiple parts of CDS
	dORF extension	3'UTR and parts of CDS
	dORF	3'UTR of CDS
	lncRNAs	Non-coding region

Figure 1. Classification of sORFs that encode SEPs.

References:

[1] Nucleic Acids Research. 2021; 49(11): 6165-80

[2] *Molecular & Cellular Proteomics*. 2023; 22(1), 100480