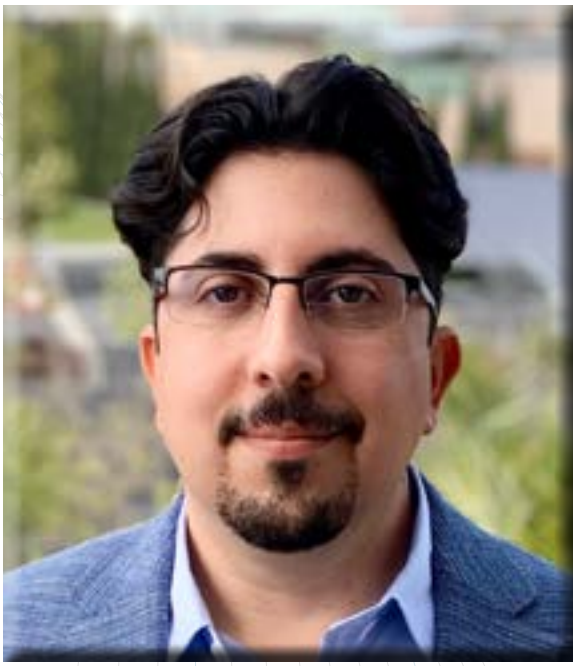
Monday, May 17, 2021 4:00 PM EST

Zoom Link

Annotation and Characterization of Human Protein-coding Small Open Reading Frames



Thomas Martinez, PhD
Postdoctoral Fellow
Salk Institute for Biological Studies

Abstract

Functional protein-coding small open reading frames (smORFs) are emerging as an important class of genes. Several smORF-encoded microproteins have been characterized and implicated in a variety of critical processes, including regulation of mRNA decay, DNA repair, and muscle formation. Thus, rigorous and comprehensive annotation of protein-coding smORFs is critical to our understanding of basic biology and physiology, as well as disease. We recently developed an improved workflow that integrates de novo transcriptome assembly and ribosome profiling to overcome obstacles with previous methods to more confidently annotate thousands of novel smORFs across multiple human cell lines, including hundreds encoded on putative non-coding RNAs. Over 1,500 smORFs are found in two or more cell lines, and ~40% lack a canonical AUG start codon. Evolutionary conservation analyses suggest that hundreds of smORF-encoded microproteins are likely functional. We also find that smORF-derived peptides are detectable on human leukocyte antigen complexes, positioning smORFs as a source of novel antigens. The annotation of protein-coding smORFs radically alters the current view of the human genome's coding capacity and will provide a rich pool of unexplored, functional human genes.

CENTER FOR RNA BIOMEDICINE