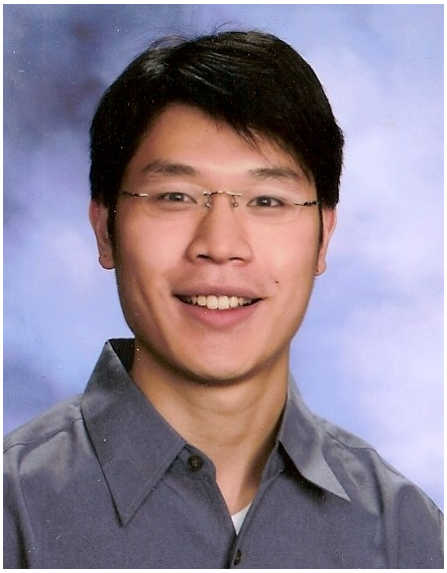
NUS National University of Singapore | Department of Biological Sciences, Faculty of Science

SEMINAR

Tues, 26 Nov 2019 | 10 am | DBS Conference Room 1

Hosted by Prof Ding Jeak Ling



Illuminating the dark proteome: Pervasive functional micropeptides encoded by the "unannotated genome"

By **Jin Chen**

Department of Cellular and Molecular Pharmacology, University of California, San Francisco, Jonathan Weissman Lab

More than 15 years after the completion of the Human Genome Project, our understanding of the "annotated" genome is still incomplete. Recent advances in functional genomics and bioinformatics have revealed numerous putative protein-coding regions (open reading frames, ORFs) outside of canonical annotations, but only a few of the encoded novel peptides have been characterized. Hindered by a lack of systematic methods to functionally characterize novel ORFs, the full protein-coding capacity of the genome remained unexplored. Here, I used ribosome profiling, mass spectrometry, and CRISPR-based screens to map the functions of these non-canonical ORFs, transcriptome-wide. I revealed pervasive translation outside of annotated protein-coding regions, many encoding "micropeptides" upstream of canonical ORFs (upstream ORFs, uORFs), or on what previously were thought to be long non-coding RNAs (lncRNAs). Functional characterization revealed multiple examples of lncRNA-encoded peptides that form functional complexes with distinct cellular localizations, challenging the conventional view that these RNAs are "non-coding". I also identified peptides encoded by uORFs, which traditionally have been viewed as translational modulators. Intriguingly, I uncovered uORF peptides binding to the downstream-encoded protein on the same mRNA. This unanticipated result highlights a diverse use of functional bicistronic operons in mammals. Together, my results uncover a large family of functional human micropeptides, encoded on supposedly "non-coding" regions, that play critical and diverse cellular roles, revealing a whole new level of complexity for the functional mammalian proteome.