UNIVERSITY of GUELPH

COLLEGE of BIOLOGICAL SCIENCE

DEPARTMENT OF MOLECULAR AND CELLULAR BIOLOGY

Seminar

SPEAKER SERIES

2022-2023

MAR **WED 1** **10 30 AM** | **SSC 2315**

DR. MARIE BRUNET

Assistant Professor,
Medical Genetics Service, Department of Pediatrics
Faculty of Medical and Health Services

UNIVERSITY OF SHERBROOKE

Topic: The coding potential of pseudogenes: between relics and new beginnings

Recent technological advances have revealed pervasive translation throughout the genome. These novel sequences are found in "non-coding" regions or overlapping canonical CDS in a different reading frame. We developed the first proteogenomics resource endorsing a multi-coding annotation of eukaryotic genomes, OpenProt. OpenProt cumulates experimental evidence for all predicted CDS from an exhaustive transcriptome. Large-scale mining of Ribo-seq and proteomics (MS) datasets highlighted 4,478 non-canonical CDS in the human genome detected in at least 3 independent Ribo-seq datasets or with at least 2 unique peptides within at least 2 independent MS datasets. 26% of these novel CDS are within pseudogenes.
Pseudogenes are widely accepted to be defective copies of protein-coding genes. Yet, we identified unique peptides for 9 proteins encoded by different pseudogenes of the GAPDH family within datasets of GAPDH affinity purification followed. Identification of each protein by MS was validated using synthetic peptides and folding and docking simulations further supported the interaction of each pseudogene-encoded protein with the parental protein GAPDH. Pseudogenes are usually identified as any genomic sequence similar to another gene but that is defective and often without introns. However, these criteria are ill-advised to predict the functionality of a gene or absence thereof. Using data cumulated with OpenProt, we developed CoP3E, a machine learning algorithm to predict the coding nature of pseudogenes. With an accuracy of 90%, our model largely outperforms existing models for ncRNAs. Furthermore, pseudogenization is ongoing in current population, and we developed a pipeline that can detect novel pseudogenes within whole exome sequencing data.
As pseudogenes are annotated as non-coding and non-functional, they are discarded from analyses and absent from protein databases. Our results call for reconsideration of the coding nature of pseudogenes, and their role in the evolution of genomes.

Pr. Brunet qualified as a Doctor of Veterinary Medicine (Paris, France) and later completed her PhD in Pharmacology at the University of Cambridge (Cambridge, UK) under a prestigious Gates Cambridge scholarship. She did her postdoctoral training in Biochemistry and Functional Genomics at the University of Sherbrooke (Sherbrooke, Canada) and, in 2021 she established her own research group at the Medical Genetics Service of the Department of Pediatrics at the University of Sherbrooke. During her postdoctoral studies she co-developed and managed the OpenProt resource, the first proteogenomics resource endorsing a polycistronic annotation of eukaryotic genomes. OpenProt is now co-led and co-developed by her lab and that of Pr Roucou. Beyond OpenProt, her research focuses on non-annotated coding sequences in our genomes, notably within pseudogenes. She uses both fundamental and computational research methods to better explore biological data and understand human diseases, focusing on pediatric cancers and rare diseases. She was awarded a FRQS Junior 1 career grant recognizing her leadership at the intersection of health sciences and artificial intelligence.

All welcome to attend
Light refreshments to be served

More information on MCB's website:
www.uoguelph.ca/MCB