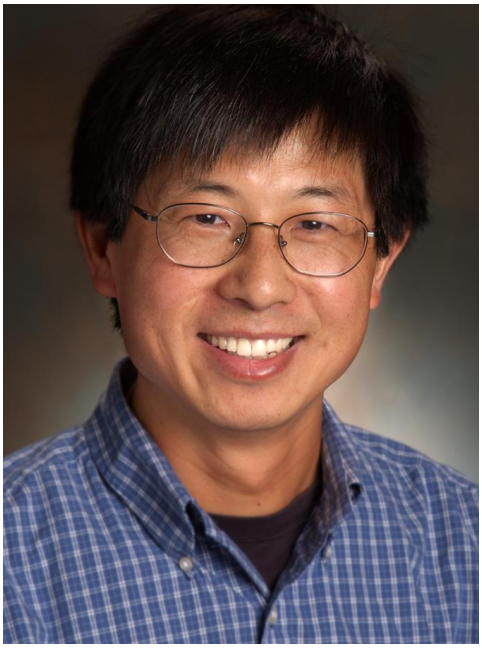
NUS National University of Singapore | Department of Biological Sciences, Faculty of Science

BIOLOGY COLLOQUIUM

Friday , 13 April 2018 | 4pm | DBS Conference Room 1

Hosted by Professor Gong Zhiyuan

Defining and evaluating Coding Capacity of *C. elegans* genome



By **Jack Chen**

Professor, Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, Canada

Since the completion of the first animal genome sequencing project 20 years ago in 1998, which reconstructed the complete genome sequence of the nematode *Caenorhabditis elegans*, many genomes have been successfully reconstructed. An outstanding challenge in the field of genomics is to annotate the complete set of protein-coding transcripts in genomes for subsequent functional studies, using innovative computational and computational methods. A frequently asked question is: have we successfully defined the complete set of protein-coding transcripts (i.e., coding capacity) of a genome? Addressing this question is rather challenging because there is no objective "yard stick" measuring the completeness of "coding capacity" of a genome. In this work, we undertake to address this question using *C. elegans* as a model. Our general assumption is that a complete set of protein-coding transcripts should include all introns and exons, which are relatively straightforward to define. I will present our work in four specific aims. First, we define introns and exons in *C. elegans* using RNA-Seq libraries. Second, we will evaluate the completeness of the protein-coding transcripts annotated in WormBase, a database that curate all transcripts in *C. elegans*. We predict that our current annotation of *C. elegans* coding capacity is incomplete. Third, we will define new transcripts using RNA-Seq libraries and various computational methods, followed by evaluating and identifying novel protein-coding transcripts in *C. elegans*. Finally, we will analyze transcripts that are rare globally (relative to all transcripts per gene), but non-rare temporally or spatially. We will explore the relationship of these transcripts in common pathways which may help define their importance in *C. elegans* development. We will also discuss the utility of Iso-Seq in defining full-length transcripts. This research may shed light on how to evaluate the completeness of coding capacity in general.

Artwork by Ann Nee, DBS, NUS