# *Visit of Dr AU Kin Fai*
## *University of Iowa*
## *27 May 2014*

**Bio:**

Dr Au is the assistant Professor in Department of Internal Medicine, University of Iowa. He received his BS degree in Biological Sciences from Tsinghua University in 2004. In 2009, Dr Au received my PhD degree in Structural Biology from Oxford University. From 2007 to 2009, he studied in Professor Wing H Wong's group and received his MS degree in Statistics at Stanford University in 2009. Since 2009, Dr Au had postdoctoral training in the Department of Statistics at Stanford University to expand his research on Next Generation Sequencing and stem cell biology. In 2013, he started his faculty position in University of Iowa. The research interest in Dr Au's group is methodology development of Second Generation Sequencing and Third Generation Sequencing.

## Seminar:

**Date:** **27 May 2014 at 10am – 12 noon**

**Venue:** **Seminar Room 4, Laboratory Block**, Faculty of Medicine Building.

**Title:** **Gene Isoform Identification of Human ESC Transcriptome by Second/Third Generation**

**Abstract:** Although transcriptional and post-transcriptional events are detected in RNA-seq data from second-generation sequencing (SGS), full-length mRNA isoforms are not captured. On the other hand, third generation sequencing (TGS), which yields much longer reads, has current limitations of lower raw accuracy and throughput. Here, we combine SGS and TGS with a custom-designed method for isoform identification and quantification to generate a high confidence isoform data set for human embryonic stem cells (hESC). We report 8,084 RefSeq-annotated isoforms detected as full length, and additional

5,459 isoforms predicted through statistical inference. Over one-third of these are novel isoforms, including 273 RNAs from gene loci that have not previously been identified. Further characterization of the novel loci indicates that a subset is expressed in pluripotent cells but not in diverse fetal and adult tissues; moreover, their reduced expression perturbs the network of pluripotency-associated genes. Results suggest that gene identification, even in well-characterized human cell lines and tissues, is likely far from complete.