



Dr. Sunduz Keles

Associate Professor
Departments of Statistics and of
Biostatistics and Medical Informatics
University of Wisconsin-Madison

Multi-reads enable detection of novel peaks in ChIP-seq analysis

Friday
December 3, 2010
12:00—1:00 p.m.

The Hospital for Sick Children
CDIU Multimedia Theatre
Room 4132, 4th Floor, Elm Elevators
555 University Ave., Toronto, ON

About Dr. Keles: Sunduz Keles is an Associate Professor in the Department of Statistics and of Biostatistics and Medical Informatics at the University of Wisconsin, Madison. She received her undergraduate degree from Bilkent University in Turkey and her Ph.D. from the University of California at Berkeley. Dr. Keles's research concerns development and application of statistical and computational methods for problems that arise in genome research. Her contributions can be broadly grouped in two areas as the statistical and computational methods for studying transcription regulation via ChIP-chip and ChIP-Seq experiments and efficient dimension reduction and variable selection methods for high dimensional genomic data. She received the David Byar Young Investigator award of the The Biometrics Section of the American Statistical Association in 2006 for her model-based approach for analyzing ChIP-chip data. She is also the recipient of the 2007 PhRMA Foundation Award in Informatics. Her research is supported by funds from NIH and NSF.

About the Talk: Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) is well on its way to replace chromatin immunoprecipitation followed by genome-wide tiling array analysis (ChIP-chip) as the preferred approach for mapping transcription-factor binding sites and chromatin modifications. Current state of the art for analyzing ChIP-seq data relies on using only reads that map uniquely to relevant reference genome (uni-reads). This can lead to omission of up to 20% of alignable reads. We describe a general approach for utilizing reads that map to multiple locations (multi-reads) on the reference genome. Our approach is based on allocating multi-reads as fractional counts using a weighted alignment scheme. Using human STAT1 and mouse GATA1 ChIP-seq datasets, we illustrate that incorporation of multi-reads significantly increases sequencing depths of ChIP-seq data, leads to detection of novel peaks that are not otherwise identifiable with uni-reads, and improves detection of peaks in mappable regions. We investigate various genome-wide characteristics of peaks detected only by utilization of multi-reads via computational experiments. Overall, these peaks have similar characteristics to peaks that are identified by uni-reads except that majority of them reside in segmental duplications. We further validate a number of GATA1 multi-read only peaks by quantitative ChIP assays and real-time PCR.

Housed at the University of Toronto Dalla Lana School of Public Health, **CIHR STAGE** is a training program in genetic epidemiology and statistical genetics funded by the Canadian Institutes of Health Research through the Strategic Training Initiative in Health Research program. Seminars are sponsored by The Hospital for Sick Children, the Samuel Lunenfeld Research Institute of Mount Sinai Hospital, the Ontario Institute for Cancer Research, the Department of Statistics of the University of Toronto, the Ontario Cancer Institute of the University Health Network, and the CIHR Institute of Genetics.

SEMINAR
International Speaker