THE FIELDS INSTITUTE

DISTINGUISHED LECTURE SERIES IN STATISTICAL SCIENCE

TERRY SPEED

Walter & Eliza Hall Institute of Medical Research, Melbourne

APRIL 9 & 10, 2015 • THE FIELDS INSTITUTE, ROOM 230

GENERAL LECTURE: APRIL 9, 3:30 P.M.

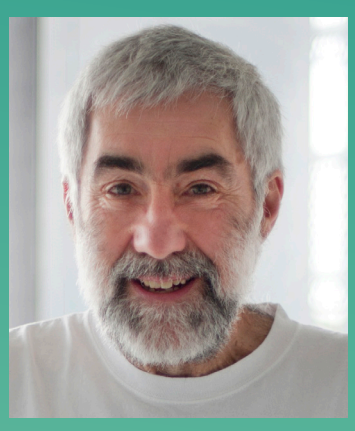
Epigenetics: A New Frontier

Scientists have now mapped the human genome - the next frontier is understanding human epigenomes; the 'instructions' which tell the DNA whether to make skin cells or blood cells or other body parts. Apart from a few exceptions, the DNA sequence of an organism is the same whatever cell is considered. So why are the blood, nerve, skin and muscle cells so different and what mechanism is employed to create this difference? The answer lies in epigenetics. If we compare the genome sequence to text, the epigenome is the punctuation and shows how the DNA should be read. Advances in DNA sequencing in the last 5-8 years have allowed large amounts of DNA sequence data to be compiled. For every single reference human genome, there will be literally hundreds of reference epigenomes, and their analysis will occupy biologists, bioinformaticians and biostatisticians for some time to come. In this talk I will introduce the topic and the data, and outline some of the challenges.

SPECIALIZED LECTURE: APRIL 10, 11 A.M.

Normalization of omic data after 2007

For over a decade now, normalization of transcriptomic, genomic and more recently metabolomic and proteomic data has been something you do to "raw" data to remove biases, technical artifacts and other systematic non-biological features. These features could be due to sample preparation and storage, reagents, equipment, people and so on. It was a "one-off" fix to what I'm going to call removing unwanted variation. Since around 2007, a more nuanced approach has been available, due to JT Leek and J Storey (SVA) and O Stegle et al (PEER). These new approaches do two things differently. The first is that they do not assume the sources of unwanted variation are known in advance, they are inferred from the data. And secondly, they deal with the unwanted variation in a model-based way, not "up front." That is, they do it in a problem-specific manner, where different inference problems warrant different model-based solutions. For example, the solution for removing unwanted variation in estimation not necessarily being the same as doing for prediction. Over the last few years, I have been working with Johann Gagnon-Bartsch and Laurent Jacob on these same problems through making use of positive and negative controls, a strategy which we think has some advantages. In this talk I'll review the area, and highlight some of the advantages of working with controls. Illustrations will be from microarray, mass spec and RNA-seq data.



Terry Speed is currently a Senior Principal Research Scientist at the Walter and Eliza Hall Institute of Medical Research. His lab has a particular focus on molecular data collected by cancer researchers, but also works with scientists studying immune and infectious diseases, and those who do research in basic biomedical science. His research interests are broad, but include the statistical and bioinformatic analysis of microarray, DNA sequence and mass spectrometry data from genetics, genomics, proteomics, and metabolomics. The lab works with molecular data at several different levels, from the lowest level where the data come directly from the instruments that generate it, up to the tasks of data integration, and of relating molecular to clinical data. Speed has been recognized by the Prime Minister's Prize for Science (Australia) and Eureka Prize for Scientific Leadership among other awards.

For more information, please visit:
www.fields.utoronto.ca/programs/scientific/14-15/DLSS/

NSERC CRSNG

Ontario