

STATISTICS COLLOQUIUM

MONDAY, APRIL 15, 2013

TALK: 4:15 PM — SCIENCE CENTER RM. 705

RECEPTION: 3:50 PM — SCIENCE CENTER, 7TH FLOOR

“Models and Statistics for Detection of Genome Structural Variation”

Nancy Zhang

Department of Statistics

The Wharton School, University of Pennsylvania

ABSTRACT

Structural variation, which includes deletion, insertion, and inversion of stretches of DNA, comprise an important class of genome variation in the human population, and are implicated in many diseases. High throughput paired end short read sequencing allows for genome-wide detection of a wide spectrum of structural variation. We develop a general model for this data, based on a Poisson random field, under which signals that are characteristic for each type of structural change can be modeled using a likelihood based framework. Statistics derived from the model integrate information from coverage, insert length, and other aspects of the data, and thus has improved sensitivity over methods that only utilize any single feature. We also describe how to control the false discovery rate for scan statistics of Poisson random fields, and illustrate our methods on 1000 genomes data. This is joint work with David Siegmund and Benjamin Yakir.