

## AWARDS & ANNOUNCEMENTS

Dr. Lee Dicker

### **Dense and Sparse Methods in High-Dimensional Data Analysis**

Many statistical methods for high-dimensional data analysis begin with the assumption that the parameter of interest is, in some sense, sparse. Broadly speaking, sparsity measures the degree to which a specified outcome may be described by relatively few features (e.g. when the occurrence of a given disease is determined by very few genes). Sparse methods for high-dimensional data analysis attempt to leverage sparsity and have proven to be very effective in many applications. However, the performance of sparse methods hinges on the sparsity of the underlying parameters and statistical methods for checking sparsity assumptions are lacking from the literature. The driving goal of this project is to develop practical statistical tools for identifying situations where the relevant parameters are in fact sparse, or where sparse methods for high-dimensional data analysis may be applied effectively. A rich statistical framework for sparsity testing will be developed and specific sparsity testing procedures will be proposed and analyzed. These methods, when applied to a given dataset, will help researchers determine the potential benefits of using sparse methods for high-dimensional data analysis. This research is likely to have significant implications for understanding reproducibility in high-dimensional data analysis and broad applications in the analysis of genomic data.