2013 JMNMF RSA Summary & Photo

Andrew Sedgewick, University of Pittsburgh Cancer Institute and Department of Computational & Systems Biology, Pittsburgh Cancer Institute

Pictured Left to Right: Takis Benos, PhD, Associate Professor, Department of Computational & Systems Biology; Andrew Sedgewick, JMNMF RSA Recipient; and, Hussein Tawbi, MD, PhD, Assistant Professor of Medicine, University of Pittsburgh Cancer Institute.

Research Summary: <u>Finding Biomarkers in Melanoma by Integrated Analysis of Genomic Data</u>

My research is focused on vertical integration of many types of genomic data taken from the same patients. Recent projects including ENCODE and The Cancer Genome Atlas have shown that it is essential to consider a variety of functional genomic data to produce an accurate molecular picture of a cell. This type of analysis is particularly important in the study of melanoma and cancer in general where understanding the changes in regulatory pathways can produce both targets for therapy and biomarkers that accurately predict survival and disease progression. How to integrate these heterogeneous datasets is an active research field in machine learning with many applications. I propose to develop new methods and algorithms for probabilistic network structure learning to integrate these different types of data without the benefit of complete prior knowledge of the relationships between the various data features. By learning the network structure of functional genomic data from melanoma, we will be able to both find informative biomarkers and to produce hypotheses about how pathway perturbations are linked to clinical variables such as disease state or response to chemotherapy.

A number of studies have had success integrating genomic data such as copy number, expression and methylation, but these studies depended on prior knowledge about the structure of the connections between the datasets and were usually limited to 2 different data types. Other works have shown that methods that are agnostic to the underlying structure of genomic data can produce useful results. My proposed algorithm will learn the statistical connections between data features with and without such prior knowledge. I will use the learned network to draw conclusions about the connections between the datasets and to make prognostic predictions. My algorithms will be applicable to many large heterogeneous datasets that lack well understood links between data types. It will be possible to extend my method to include other types of data including clinical and image traits. I will primarily be working with methylation, miRNA, mRNA and SNP data collected by the University of Pittsburgh Cancer Institute's Melanoma Program, and I will use data from The Cancer Genome Atlas for validation of my findings.