Lung cancer is a disease with dismal prognosis; only 15% of newly-diagnosed patients survive for five years. Our understanding of how to diagnose, stage, and treat it is based largely on macroscopic or cellular phenomena. A molecular understanding of the disease may provide improved clinical management and new therapeutic options.

My group focuses on predicting the survival of lung cancer patients. In particular, we develop algorithms to exploit microarray datasets to develop biomarkers of survival, called prognostic markers. In this talk I will describe three recent results: an algorithm, a database, and an empirical finding.

First, I describe a new feature selection algorithm, called modified steepest descent (mSD). This algorithm couples gradient-descent with unsupervised machine-learning. Through greedy forward-selection it generates a six-gene prognostic marker for lung cancer that is validated in over 500 patient samples.

Second, I describe a meta-analytic database that compiles the data from nine transcriptomic studies of lung cancer. These studies were integrated using a novel normalization approach and then subject to meta-analysis. For each gene present in the analysis (16,391 in total), the univariate prognostic capacity was calculated. I show that this database increases our statistical power sufficiently to allow separate analysis of different histological subtypes of lung cancer.

Third, I describe an analysis of biomarker plurality. From an empirical study of biomarker-space we found that the number of effective markers is very large. The inter-relationship amongst these markers contains information about gene-gene interactions, and may provide an avenue for understanding the specific pathways dysregulated in lung cancer.

Lung cancer incidence remains high and survival remains low. The development of prognostic markers may improve this situation by allowing personalized therapy. The computational approaches described here may be applicable beyond this one disease, and may provide insight into the types of methodologies that will work well for other problem-domains.