



The University of Chicago
Department of Statistics

Seminar Series

GIOVANNI PARMIGIANI

Departments of Oncology, Biostatistics and Pathology
Johns Hopkins University

“Mining and Modeling Multiple Microarray Datasets.”

MONDAY, May 1, 2006 at 4:00 PM
133 Eckhart Hall, 5734 S. University Avenue
Refreshments following the seminar in Eckhart 110.

ABSTRACT

Investigations of transcript levels on a genomic scale using hybridization-based arrays led to formidable advances in our understanding of the biology of many human illnesses. At the same time, these investigations have generated controversy, because of the probabilistic nature of the conclusions, and the surfacing of noticeable discrepancies between the results of studies addressing the same biological question. In this lecture I will present approaches for mining and formal modeling of multiple microarray datasets. I will start with simple exploratory data analysis tools for gauging the degree to which the findings of one study are reproduced by others, and for integrating multiple studies in a single analysis. I will describe these approaches in the context of studies of both lung and breast cancer. I will then present a multilevel bayesian model for formal inference and data integration. The main conclusion of our work to date is that it is possible to identify a substantial, biologically relevant, subset of the human genome within which hybridization results are reproducible. The subset generally varies with the platform used, the tissues studied, and the populations being sampled. Despite important differences, it is also possible to develop measures that allow comparison across platforms, studies, labs and populations. While these are not perfect, important biological signal is often preserved or enhanced.