

CIMIT Forum Summary – July 25, 2006

Jill Mesirov, PhD, MIT Broad Institute, described “Knowledge-Based Methods for Gene Expression Analysis.” Over the last ten years, the use of microarray technology has grown tremendously and now allows researchers to study expression levels for every gene in the body and to attempt to correlate gene expression with phenotype.

Traditionally, genes have been considered individually. Many analyses of microarray data focus on building class predictors that can distinguish one phenotype from another. To build such a predictor, one must use a relatively small number of samples to infer a function that will send 20,000 to 30,000 features (genes) to one of two values – a difficult task. To reduce the number of dimensions being considered, researchers select marker genes that individually are correlated with phenotype and then build a predictor based on multiple such markers. These predictors can help clinicians make more accurate diagnoses and can shed light on the molecular taxonomy of cancer. Dr. Mesirov’s team successfully built a class predictor to distinguish acute myeloid leukemia from acute lymphoid leukemia. In general, however, building a class predictor based on marker genes can be problematic. Sometimes, signature differences can be too subtle to distinguish between related phenotypes, as was the case when Dr. Mesirov’s team compared diabetic to normal muscle cells. In other cases, there is a lack of consensus and reproducibility. For example, three different research groups studied lung adenocarcinomas, but they only agreed on one marker gene. Finally, it is difficult to understand the underlying biology of a system by looking only at a random mix of marker genes. To better analyze microarray data, Dr. Mesirov’s team looked at sets of functionally related genes instead of at individual genes, a choice that made their results

more sensitive, more reproducible, and more easily interpreted. Calling their method Gene Set Enrichment Analysis (GSEA), they discovered that they could distinguish diabetic from normal muscle based on a set of genes associated with a known oxidative phosphorylation pathway, suggesting that this pathway may be involved with the disease. They applied GSEA to the gene expression data obtained by the three different groups studying lung adenocarcinomas, and they found that although the results of the different groups showed little overlap in terms of individual genes, the results did agree on sets of genes, shedding light on the pathways involved in this cancer. To help others use their methods, they have created a database of sets of related genes collected from sources such as pathway databases, perturbation experiments, and the literature. The use of these knowledge-based sets of genes increases consistency, boosts subtle signals, and helps researchers gain insight into underlying cellular processes.

Pamela Silver, PhD, Harvard Medical School, discussed “Designing Biological Systems.” A combination of engineering and biology, synthetic biology focuses on redesigning biological systems in order to better understand how those systems work and to explore new engineering possibilities. Unlike biological design, in which a structure or mechanism must evolve through a series of functioning intermediates, human design can proceed through informative intermediates that do not necessarily function.

Biological systems are like those engineered by humans in that they are capable of sensation, communication, and duplication, but they can also repair themselves and evolve. The modularity of biological systems, when coupled with improving technology, is beginning to allow humans to engineer a variety of genes, promoters, and repressors.

Dr. Silver's team is exploring cellular memory and logic. They hope to make yeast cells that are biological counters, capable of counting the number of times that they have divided. To start, Dr. Silver's team is attempting to create many simple devices that could eventually be components of the counting system. One such component is a daughter-specific device that they made by putting a daughter-specific promoter in front of the genetic code for a fluorescent yellow protein. They created a "localizer" device that regulates nuclear transport depending on a yeast cell's position in the cell cycle, and they also made a tuneable synthetic-repressor that can suppress the production of fluorescent proteins. Finally, they created a cell-cycle-dependent degradation device that consists of a cell-cycle-dependent degradation tag followed by the code for a fluorescent protein. At a given stage in the cell cycle, the fluorescent protein is degraded, allowing researchers to determine a cell's stage in the cell cycle. Currently, biological counters can count to two, but Dr. Silver's team hopes to use their relatively simple devices to create a system that can count higher by using multiple negative feedback loops. Dr. Silver's team also studied the transcription factor forkhead, which regulates apoptosis, and the phosphatase PTEN, which is mutated in many types of tumors. Forkhead and PTEN can interact to create a two-state spatially-based oscillator that could shed light on how cancer drugs work. In terms of education, Dr. Silver is involved in the International Genetically Engineered Machines competition, a competition in which teams of undergraduates compete to engineer biological machines such as bacterial cameras.