

CIMIT Forum Summary – July 18, 2006

Zhiping Weng, PhD, Boston University, discussed “Studying Transcriptional Regulation in the Human Genome.” Looking for novel regulatory regions, her team analyzed 129 datasets from the ENCODE consortium, a multi-center project seeking to extensively annotate segments of the human genome. The datasets were obtained via experiments in which chromatin immunoprecipitation (ChIP) was followed by genomic sequencing or by tiling array (chip) analysis. The experiments located transcription-factor binding sites and histones (proteins associated with active transcription) in the genetic sequence of the ENCODE segments. Dr. Weng’s team found that histones correlated extensively with each other and that their concentrations were highest near transcriptional start sites. Many transcription-factor binding sites were also concentrated around transcriptional start sites, although some transcription factors showed broader distributions or did not show any peaks. Dr. Weng’s team developed four computational methods to integrate over the datasets, thus magnifying signals that may have been too faint to notice in any individual experiment. Based on the distributions of histones and transcription-factor binding sites, the four computational methods predicted 1393 regulatory regions, 53 % of which were shared by more than one method. The predicted regions overlapped significantly with known 5’-ends of genes, and they also overlapped significantly with known 3’-ends of genes, suggesting that some of the predicted regions could play a role in antisense regulation. Using transient transfection assays and 5’-RACE experiments, Dr. Weng’s team tested 163 novel regions for promoter activity and found that 25 % of these regions showed enhanced transcriptional activity in at least one cell line. These

results suggest that a number of the novel regions may be active promoters and that many functional promoters in the human genome remain undiscovered.

Jerome Brody, MD, Boston University School of Medicine, described “The ‘Field of Injury’ in Lung Disease.” Smoking cigarettes has a large impact on gene expression in the epithelial cells that line the airway. Some changes are reversible. In one study, Dr. Brody’s team found 305 genes that were differentially expressed in current smokers and never-smokers and that were similarly expressed in former smokers and never-smokers. A number of these genes are associated with protection against oxidant stress. Many other genes, however, have an expression pattern that is altered in both current and former smokers, compared to never-smokers, and these genes include putative oncogenes. In a disproportionate number of the former smokers who go on to develop lung cancer, many of the genes with expression that usually normalizes after smoking cessation do not return to normal, or return to normal very slowly, potentially helping to identify former smokers who are at high risk for lung cancer. Airway gene expression follows characteristic patterns for different diseases. Dr. Brody’s team studied gene expression in patients with lung cancer and/or COPD. They found 202 genes that were associated with both diseases, including cell cycle genes and DNA repair genes. They also found 103 genes that were associated just with COPD, including genes related to the inflammatory response. These studies involved gene expression in bronchial epithelial cells obtained via bronchoscopy, but other studies of the field of injury have focused on nasal epithelial cells, which are easier to obtain. Comparing small samples of current smokers and non-smokers, Dr. Brody’s team found 67 genes that were differentially

expressed in the nose. Of the 48 of these genes that were also differentially expressed in the bronchi, 41 changed in the same direction and 6 changed in the opposite direction. These results suggest that smoking-induced changes in gene expression in nasal epithelial cells may be indicative of expression patterns deep within the lungs. In another study of gene expression in nasal epithelium, 15 patients with pulmonary sarcoidosis were compared to 15 people without the disease, and 37 genes were found to be differentially expressed. Thus, even pulmonary sarcoidosis, a disease without clear clinical manifestations in the nose, is associated with altered gene expression in nasal mucosa. To be most effective, gene expression profiles must be integrated with proteomics, imaging, genomic data, and clinical data. Gene expression profiles may eventually be used to help clinicians identify people at high risk of developing a disease, to improve the accuracy of prognoses, and to predict a patient's response to a certain drug.