Abstracts: Session III

Sutphin, Patrick D. [41]

Application of supervised machine learning to identify genes associated with the hypoxia response

Patrick D. Sutphin[1], Soumya Raychaudhuri[2], Nicholas C. Denko[1], Russ B. Altman[2] & Amato J. Giaccia[1]

[1]Department of Radiation Oncology, Stanford University, Stanford, California, USA
[2]Stanford Biomedical Informatics, Department of Medicine, Stanford University, Stanford, California, USA

We demonstrate the potential use of machine learning approaches in elucidating new insights into a biological response. Supervised machine learning methods are methods that group expression profiles into one of several known categories through comparison with known examples. Examples of supervised machine learning include logistic regression, neural networks, decision trees and linear discriminant analysis. The method we employ here is a version of linear discriminant analysis modified to cope with high-dimensional data. We used the publicly available NCBI60 cancer cell line microarray expression database as reported by Schref and colleagues[1]; this diverse data set includes samples from nine different tissue types. These cell lines were separated into high and low expressers of the Von Hippel Lindau gene (*VHL*). The *VHL* protein product is a critical regulator of the hypoxic response; it targets the hypoxia inducible factor 1-alpha (HIF-1a) transcription factor for proteasome-mediated degradation. We examined the features that the supervised machine learning algorithm used for classification and conjectured that they might be genes that are regulated at the level of transcription by HIF-1a. Many of these genes were already known HIF-1a target genes; others were genes known to be expressed under hypoxic conditions. Supervised machine learning methods succeed here where traditional unsupervised clustering methods may fail. The phenomenon studied here is restricted to a handful of genes, whose expression patterns may be masked by the tissue-specific profiles that will dominate in clustering.

1. Scherf, U. *et al. Nature Genet.* **24**, 236–244 (2000).

Suxing Liu [42]

Genome-wide gene expression analysis of human breast cancer cells

Suxing Liu[1], Qun Wu[1], Paul Kirschmeier[1], Terri McClanahan[2], Wei Ding[1], Ferdous Gheyas[1], Luquan Wang[1], Marco Hernandez[1] & Jonathan Greene[1]

[1]Tumor Biology Department, Bioinformatics Group, and Biostatistics Group, Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, New Jersey 07033, USA
[2]DNAX Research Institute, 901 California Avenue, Palo Alto, California 94304, USA

We studied differential gene expression between normal breast epithelial cells and their malignant counterparts with a microarray technique, using GeneAlbum GEM (Incyte), which contains approximately 60,000 complementary DNA clones. We searched for different gene expression patterns among individuals, ages and races using four normal human mammary epithelial cells (HMEC). In addition, we used eight well-established breast cancer cell lines as probes against a common normal HMEC. Comparison between normal HMEC and malignant HMEC verified several known HMEC molecular markers: laminin-5 beta 3 and alpha 3, and cytokeratins 5, 15 and 17. We also observed some known breast-cancer-related genes, such as *HER2, HER3* and *pS2*, which were expressed as expected. The microarray quantitation was verified independently using Taqman analysis. The

results of flip-flop experiments are discussed, as are the distribution of differentially expressed genes in 11 probe pairs and their cluster analyses.

Takahashi, Yasuo [43]

Thymidylate synthase expression correlates closely with *E2F1* expression in primary colon cancer

Yasuo Takahashi, Toshihito Nagata, Yukimoto Ishii, Satoshi Asai & Koichi Ishikawa

Nihon University School of Medicine, Tokyo, Japan

Thymidylate synthase (TS) is a key enzyme in the synthesis of DNA and the target enzyme of 5-fluorouracil, the most widely used chemotherapeutic agent for colon cancer. The expression of TS messenger RNA or protein reportedly predicts overall survival for colon cancer patients and correlates with resistance to 5-fluorouracil. Because the gene encoding TS is thought to be one of the target genes bound to and regulated by the E2F1 transcription factor, it was of interest to measure TS and *E2F1* gene expressions in primary colon cancer specimens. We found that TS expression correlates closely with *E2F1* expression in colon cancer specimens. Moreover, this correlation was observed in all tumors, regardless of clinical stage. These results suggest that the ability of a tumor to increase TS expression may be due to an over-expression of *E2F1*, and they provide new insights into the molecular mechanisms underlying the regulation of TS expression in colon cancers.

Tanenbaum, David [44]

High-throughput genomic analysis of small-cell lung cancer: allelotyping using Affymetrix HuSNP arrays

David Tanenbaum[1], Kerstin Lindblad-Toh[2], Eric Lander[2] & Matthew Meyerson[1]

[1]Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA
[2]Whitehead Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

An understanding of the global patterns of loss of heterozygosity in a neoplasm is achieved by analysis of polymorphic markers, or allelotyping, which is essential for both identification of new tumor suppressor genes and cancer classification. Although analysis of simple sequence length polymorphisms is a reliable method for loss of heterozygosity analysis, detailed allelotyping studies require more efficient methods and a denser set of genetic markers. We performed high-throughput allelotyping on small-cell lung cancer tissue using Affymetrix single-nucleotide polymorphism (SNP) microarrays. Our proof-of-principle experiments demonstrated that SNP array hybridization is an accurate (approximately 95.4% on the basis of validation experiments) method for rapidly evaluating genome-wide tumor allelic imbalance (primarily loss of heterozygosity) at high marker density. The SNP array data were consistent both with previous analyses based on simple sequence length polymorphisms and with comparative genomic hybridization data for the same samples. Taken together, these experiments have demonstrated that SNP array hybridization is an especially promising technique for a wide array of tumor allelotyping applications. Increased numbers of mapped SNPs and technical improvements to extend the range of samples that can be analyzed should further increase the utility of this already powerful method. Current technological modifications include reducing required sample input to 5 ng for a full allelotyping and utilization of archival paraffin-embedded tissue samples, for which clinical outcome data are available.