smoked, and those who had never smoked (NS). We also examined the effect of central versus peripheral tumor location on gene inactivation. We found that p16 was methylated in 53% (36/68), 59% (16/27), and 49% (27/55) of peripheral tumors from smokers, former uranium miners, and those in the NS group. Central AdCs seem to have a higher frequency of p16 methylation (21/30; 70%) than peripheral tumors. This higher frequency of p16 methylation in central AdCs parallels our findings for squamous carcinoma, a predominantly central lesion. These studies also indicate that radon exposure does not interact synergistically with tobacco to target this gene for inactivation, but they do implicate p16 as a major pathway for AdC development in the NS group. The interaction of p19 with p16 methylation will also be discussed.

Dougherty, Edward

[49]

Clustering algorithms: can anything be concluded?

Edward Dougherty¹, Junior Barrera², Marcel Brun², Seungchan Kim¹, Roberto Marcondes², Yidong Chen³, Michael Bittner³ & Jeffrey Trent³

¹Department of Electrical Engineering, Texas A&M University, College Station, Texas, USA

²Center for Bioinformatics, University of São Paulo, São Paulo, Brazil

³Cancer Genetics Laboratory, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA

Clustering algorithms are used in various applications, including the clustering of genes by expression profiles. Clustering algorithms find clusters, and these are often visually satisfying. However, the worth of a clustering algorithm must be judged by the degree to which the clusters it produces agree with the true classes. For instance, when clustering discrete time series from expression data, genes whose expression profiles show similar behavior are clustered, indicating possible co-regulation. The ability of the clustering algorithms to identify truly co-regulated genes is limited by both the uniqueness of the transcription pattern and the extent of variance of the biological processes. Independently regulated processes that produce similar transcription profiles can be mistakenly clustered as a single class if the variance in the observed patterns is of the same magnitude as the difference in the patterns. We consider ways to test the precision of clustering for given sets of transcription patterns and given levels of pattern variance, analyzing this issue for various clustering algorithms, including K-means, fuzzy C-means, hierarchical correlation-based clustering and Euclidean-distance clustering. The analysis applies to clustering in general; however, we focus on expression-based time series. We postulate a model having gene classes whose expression profiles exhibit congruency. A profile is modeled as a time-expression template defining the congruency class, plus noise. Clusters are estimates of the congruency classes. We measure clustering precision by the expected number of misclassifications. Precision depends on several factors: class separation, experimental variability and the number of sample replications. Very poor results are obtained for a single replicate, but there can be vast improvement with as few as three replicates. Moreover, performance varies substantially among clustering algorithms. Application of the inference analysis to real data requires estimation of model parameters from the data. Once these parameters have been estimated, the inference analysis can be applied to predict the expected numbers of errors on the basis of various algorithms and the number of replications. Intuitively the analysis gives the number of errors one would expect given the data. Since the analysis requires means and variances for the congruency classes, we apply a clustering algorithm to the raw data to form congruency classes with which to seed the algorithm. Clustering algorithms are applied to the seed classes.

Duffy, Carol

[50]

Oligonucleotide array analysis of telomerase activation by the human papillomavirus type 16 E6 protein

Carol Duffy & Aloysius Klingelhutz

Department of Microbiology, University of Iowa, College of Medicine, Iowa City, Iowa 52242, USA

Human papillomavirus type 16 (HPV-16) is strongly associated with the development of cervical cancer. The E6 and E7 proteins from HPV-16 possess several functions that contribute to oncogenesis. HPV-16 E7 is known to bind and inactivate the retinoblastoma tumor suppressor protein. Expression of E7 alone extends the life span of newborn foreskin epithelial cells about twofold, whereas expression of E6 and E7 together can result in cellular immortalization without a significant crisis. HPV-16 E6 functions in the activation of telomerase and the targeting of p53 for ubiquitin-mediated degradation. E6 is known to alter the expression of several cellular genes by both p53-dependent and -independent mechanisms. Several cellular transcription factors (such as c-Myc, Sp1 and WT1) have been reported to be involved in telomerase regulation in either a cell-specific or -nonspecific manner. We propose that HPV-16 E6 activates telomerase through the upregulation of a cellular telomerase activator or the downregulation of a cellular telomerase repressor. To identify cellular genes involved in the activation of telomerase by E6 we have performed an oligonucleotide array analysis of messenger RNA from cells expressing E7 and either wild-type or mutant E6. The E6 mutants we are studying differentially activate telomerase and target p53. Array analysis of these mutants will help us identify changes in gene expression specifically associated with telomerase activation.

Duggan, David J.

[51]

Molecular profiling of prostate cancer progression

David J. Duggan¹, Jun Luo², Jurga Sauvageot², Michael Bittner¹, Yidong Chen¹, Jeff Trent¹ & William B. Isaacs²

¹Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA

²Department of Urology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Prostate cancer is the most commonly diagnosed malignancy and the second most common cause of cancer death in North American males. It is largely unknown why some tumors progress to aggressive, potentially life-threatening disease whereas others remain latent for decades. Although cancer classification based on gene expression has been successful for several types of malignancies, expression analysis in prostate cancer is complicated by the heterogeneity of the tumor (even within a single specimen) and by contaminating normal cells. We have developed an approach that combines trimming and sectioning of frozen samples with associated histopathological review to isolate and verify percent tumor and grade within a given specimen. Messenger RNA from more than 40 normal, benign prostatic hyperplasia, and low- and high-grade prostate tumor samples was isolated and compared with a single reference to generate over 300,000 data points. Gene expression profiles of all prostate samples were correlated with Gleason score, clinical stage and other phenotypic and pathologic descriptors. Using a series of statistical tools, including multidimensional scaling, we have recognized unique clusters of samples that seem to correlate with tumor aggressiveness. Statistical analy-

Abstracts: Session I

sis of these data also identified a subset of genes that distinguish tumor specimens from each other as well as from normal prostate tissue. Random permutation of the sample labels clearly indicates that the subset of genes separating tumor subgroups far exceeds that expected by chance alone ($P=0.002$). This work indicates that investigation of prostate cancer based on gene expression can identify genes that may help distinguish more aggressive forms of this disease.

Dunn, Sandra, E.

[52]

Gene expression profiling of normal breast epithelial cells following treatment with insulin-like growth factor I

Jennifer S. Oh¹, Pierre Buchel², Karla Martin², Jill E. Kucab³, Tomo Oshimura², Lee Bennett², Kelli Reynolds², David M. Cykert¹, J. Carl Barrett⁴, Richard P. DiAugustine¹, Cynthia Afshari² & Sandra E. Dunn³

¹Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

²Microarray Group, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

³Functional Genomics Program, North Carolina State University, Raleigh, North Carolina, USA

⁴Laboratory of Biosystems and Cancer, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

Clinical and experimental data support a role for insulin-like growth factor I (IGF-I) in breast cancers. Our laboratory previously showed that IGF-I is a stimulus for mammary gland branching morphogenesis as well as breast tumor growth and metastasis. Recent epidemiological studies indicate that elevated serum IGF-I levels in premenopausal women are linked with increased mammographic density, ductal carcinoma *in situ* and breast cancer risk. The IGF-I receptor is reportedly overexpressed in 30–40% of early-stage breast cancers. It is unclear why activation of IGF-I receptor signaling is associated with the development of cancer. We are using complementary DNA microarrays to profile gene expression in normal breast epithelial cells and to investigate the mechanism behind cancer risk related to IGF-I. We profiled the gene expression of normal breast epithelial cells in the absence or presence of IGF-I at 0, 0.5, 1, 2, 4, 6, 8 and 24 h on 2K cDNA microarray chips and observed inhibition of pro-apoptotic genes, including *fas* ligand. We also found increased expression of invasion and neovascularization genes. For example, the potent vascular mitogen VEGF was induced by IGF-I at the 2-h time point. To confirm differential gene expression, messenger RNA was validated by a real-time quantitative polymerase chain reaction, and the protein products were confirmed by enzyme-linked immunosorbent assay. Cellular trafficking was studied to illustrate that the cells were responding to IGF-I through a PI3K/AKT pathway. We demonstrate that the activated forms of AKT, GSK, CREB and HIF-1 α translocated into the nucleus in response to IGF-I treatment. The differential gene profiling of normal breast epithelial cells to growth factors has not been reported to date, and this investigation provides insight into the expression of gene products that may have fundamental importance to the development and progression of breast cancer.

Eckart, Meese

[53]

Genes overexpressed in the transition zone from normal to tumor tissue in larynx and pharynx carcinoma

Eckart Meese, Jörg Bohlender, Helmut Steinhart & Ulrike Fischer

Department of Human Genetics, University of Saarland, 66421 Homburg/Saar, Germany

To identify genes relevant in the development and progression of larynx and pharynx carcinoma, we investigated samples derived from the transition zone between normal tissue and tumor. For larynx and pharynx carcinoma this transition zone offers an ideal starting point to identify genes involved in early tumor development. Using microarray expression analysis we found overexpression of a group of genes, including the gene for PA-FABP. Overexpression of PA-FABP in the transition zone was confirmed by northern blot hybridization using RNA from normal tissue as control. Northern blot expression analysis also revealed overexpression of PA-FABP in the corresponding tumor samples. A second group of genes was predominantly overexpressed in the transition zone from larynx carcinoma to normal tissue, whereas no overexpression was detectable in the transition zone from pharynx carcinoma to normal tissue and in all corresponding tumor samples. These genes may be relevant in the early development of larynx carcinoma and may serve as a valuable diagnostic marker.

Einspahr, Janine

[54]

Mutation and overexpression of p53 as intermediate biomarkers in the chemoprevention of skin cancer

Janine Einspahr, David Alberts, Nancy Hart, Nelson Mark & Bowden Timothy

University of Arizona Cancer Center, Tucson, Arizona, USA

The incidence of skin cancer is increasing in the United States, and effective primary prevention and chemoprevention strategies are needed to reduce its incidence, morbidity and mortality. Squamous cell carcinoma, in contrast to most internal cancers, is characterized by a high rate of p53 mutations and p53 overexpression in premalignant actinic keratosis (AK). Similarly, p53 mutations are frequently found in normal-appearing, chronically sun-exposed skin. We have been using p53 mutations and p53 overexpression as a potential biomarker of response to chemoprevention agents. Difluoromethylornithine (DFMO), an irreversible inhibitor of ornithine decarboxylase, inhibits the synthesis of polyamines, which have an important role in the process of skin carcinogenesis. In a randomized, placebo-controlled study of topical DFMO for six months in 48 subjects with AK, topical DFMO was effective in significantly reducing AK number (23.5%; $P=0.001$), spermidine levels (26%; $P=0.04$) and p53 protein expression (22%; $P=0.04$), but not cell proliferation. There were no significant differences in the frequency of p53 mutations (exons 5–8) (23.8% at baseline, 28.6% after placebo and 26.2% after DFMO). DFMO seemed to be more effective in reducing AK, spermidine and potentially p53 mutations on right arms (33% = one p53 mutation) as compared with left arms (67% = one p53 mutation), suggesting that left arms are more resistant to the effects of topical DFMO. The addition of biomarkers, developed on the basis of the molecular biology of the disease, to chemoprevention studies is essential.