



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-