

The Early Detection Research Network

Second Annual Scientific Workshop

14–16 October 2001, Seattle, Washington, USA

Abstracts

Biomarker discovery by comprehensive phenotyping

Nancy Grove

Chief Medical Officer, SurroMed, Inc., 2375 Garcia Ave., Mountain View, CA 94043, USA

The pharmaceutical and biotech industries are under extreme pressure to improve the processes by which new drugs are invented and converted to commercial products. It is widely believed that biomarkers, defined loosely as individual or combinations of molecular, cellular or clinical indicators that correlate with disease progression and/or response to therapeutic intervention, will play a critical role in improving efficiencies, lowering costs, and shortening timelines across the entire spectrum of drug discovery and development. While the future of biomarkers seems assured, there remains but a single problem in the present: how to discover them.

We have taken an integrated approach to biomarker discovery that builds on expertise in clinical medicine, bioanalytical chemistry, and bioinformatics. Specifically, we have focused on longitudinal, comprehensive phenotyping studies of well-defined patient populations in disease areas of interest. The foundations of this approach lie in the ability to design and implement clinical protocols that address key issues in pharmaceutical development (e.g. responders vs. non-responders), to perform state-of-the-art measurements on proteins, metabolites, and cell populations in a high-throughput clinical laboratory, and to efficiently mine broad, disparate data sets from numerous sources, including mass spectrometry, cytometry, immunoassays, electronic case report forms, electronic medical records, and patient-based questionnaires.

This presentation will address the current status of this biomarker discovery platform.

Oligonucleotide microarray analysis of gene expression profiles in lung adenocarcinomas and matched uninvolved lung tissue samples

Wilbur Franklin, Razvan Lapadat, Michio Sugita, Larry Hunter, Sonia Leach, Robert Jotte, Mark Geraci, York Miller and Gary Johnson

University of Colorado Health Sciences Center, Denver, CO, USA

Oligonucleotide microarray analysis of RNA extracted from clinical tumor samples offers a new avenue for biomarker discovery. In a preliminary trial of this technology, we have assayed RNAs from duplicate punch biopsies of 10 early stage adenocarcinomas and matched normal controls with the 12,625 gene Affymetrix Hu95a chip. We have also tested 2 SCLC lines, one adenocarcinoma line and one primary culture of non-malignant epithelium. Non-parametric statistical models with corrections for multiple testing were developed: to assess reproducibility of the assays; to define gene expression profiles that distinguish lung tumor from non-neoplastic lung; to identify patterns of expression that may define subgroups within the morphological category of adenocarcinoma; and to determine the expression status of specific genes and pathways. Use of the duplicate samples taken from each tissue was found to increase the statistical power of the analysis; we were able to identify smaller differences between normal and tumor expression levels in well replicated genes. Statistically significant differences were also found in genes that were not well replicated in the duplicate samples, but where the differences between tumor and normal lung were particularly large. Together, these two analyses showed significant differences in 913 genes, 422 over- and 491 under-expressed in tumor compared to autologous lung. The overexpression of two of the gene families was confirmed by immunohistochemical analysis or RT-PCR for MAGE

genes and for calcitonin related peptide 1. Inspection of the lists indicates that the genes fall into a wide variety of functional categories that are associated directly with tumor or with stromal response to tumor. One of the most consistently overexpressed clusters of genes was that associated with the proteasome. Cell cycle pathway genes, specific signal transduction genes and tumor antigens were also abnormally expressed in the tumor compartment. Considerable heterogeneity was observed within the adenocarcinoma category; however, the relatively small number of patients makes identification of statistically significant subgroups difficult. Tumors were divided into those with bronchioloalveolar features and those without; the expression levels of 19 genes were significantly different between the groups. In addition, two non-bronchioloalveolar tumors appeared similar with regard to their expression of MAGE and neuroendocrine associated genes. This work suggests that:

1. Gene expression array assays can be successfully applied to lung tumor homogenates.
2. Expression difference profiles can be created with non-parametric statistical models corrected for multiple testing, even with moderate sample sizes; duplicate samples increase statistical power of these tests.
3. The anatomical category of adenocarcinoma is heterogeneous.
4. Biomarkers can be identified by global analysis of gene expression that may be useful in a non-array setting.

IL-8 increases the metastatic potential of ovarian cancer cells

J. So, L.E. Graves, J.R. Navari and D.A. Fishman
Department of Obstetrics and Gynecology, Northwestern University Medical School, Chicago, IL 60611, USA

In the United States more women die annually from ovarian cancer than all other gynecologic malignancies combined. The primary reason for such high mortality is that 75% of women continue to be diagnosed with disseminated epithelial ovarian carcinoma and have a 5-year survival approximating 15%. Ovarian tumor growth and metastasis requires a coordinated series of events that include cellular adhesion, migration, extracellular matrix degradation, invasion into host parenchyma, proliferation, and neovascularization that are influenced by numerous regulatory molecules, such as IL-8, lysophospholipids (such as lysophosphatidic acid (LPA), urinary-type plasminogen activator (uPA) and receptor (uPAR), and matrix metalloproteinases (MMP). IL-8, LPA and MMPs have been detected at elevated levels in malignant ascites from ovarian cancer patients as well as established ovarian tumor cell lines. In melanoma cells IL-8 up-regulates MMP 2 expression, activity and invasiveness, and IL-8 expression directly correlates with metastatic potential. We presently are evaluating the effects of IL-8 on the individual components of the ovarian metastatic cascade. *In vitro*, ovarian cancer cells grown under hypoxic or acidic conditions or exposed to lysophosphatidic acid (LPA) increased IL-8 gene expression. We found that IL-8 treatment did not have any effect on cellular proliferation (Promega assay) or uPA activity (colorimetric assay) in ovarian carcinoma cells (DOV13). Acidic pH was found to have no effect on MMP or uPA activity and expression. However, exposure to IL-8 significantly increased the amount of pro-MMP-2 (72 kDa), which was confirmed by Western blotting and ELISA. IL-8 treatment alone did not up-regulate MMP-2 activity. We previously reported that LPA treated ovarian carcinoma cells (DOV13) resulted in a statistically significant increase in the active form of MMP-2 (62 kDa). A synergistic effect was observed when cells were treated with both LPA and IL-8, with the combination resulting in increased MMP-2 activation as determined by gelatin zymography. Cellular migration was unaffected by IL-8 as quantified by a modification of the Albrecht-Buehler colloidal gold assay. However, cellular invasion through a Matrigel membrane was significantly stimulated by IL-8.

Our findings indicate that IL-8 increases pro-MMP-2 expression and stimulates Matrigel invasion. Since LPA and IL-8 coexist within the ovarian carcinoma microenvironment with critical roles in metastatic dissemination, further studies will be aimed at understanding the mechanism of IL-8-induced MMP-2 expression and cellular invasion.

Our findings indicate that IL-8 increases pro-MMP-2 expression and stimulates Matrigel invasion. Since LPA and IL-8 coexist within the ovarian carcinoma microenvironment with critical roles in metastatic dissemination, further studies will be aimed at understanding the mechanism of IL-8-induced MMP-2 expression and cellular invasion.

Inhibition of PI3K prevents lysophosphatidic acid (LPA)-induced activation of matrix metalloproteinase-2 (MMP-2)

Laura E. Graves, John So, Heather J. Matzel, Jason R. Navari, M. Sharon Stack and David A. Fishman
Departments of Obstetrics and Gynecology and Cell and Molecular Biology and the R. H. Lurie Comprehensive Cancer Center, Northwestern University Medical School, Chicago, IL 60611, USA

Ovarian cancer is a highly metastatic disease that affected over 23, 000 women in the United States last

year and accounted for more deaths of women than any other gynecologic malignancy (Greenlee, et al., *A Journal for Clinicians* **50** (2000), 7–33). The pathogenesis of this disease involves the dissemination of cancer cells throughout the peritoneal cavity, and, eventually, metastasis to distant sites. Previous data from our lab and others has shown that invasion and metastasis of ovarian cancer cells are facilitated by the presence of the lipid mitogen lysophosphatidic acid (LPA). LPA levels are elevated in the majority of ovarian cancer patients, and this lipid, through an interaction with a family of G-protein-coupled receptors, stimulates the activity of various intracellular signaling molecules including focal adhesion kinases and mitogen-activated protein kinases. The effects of LPA-induced signaling in tumor cells include cellular proliferation, survival, invasion, and the upregulation of proteolytic enzymes (refs). The gelatinolytic enzyme matrix metalloproteinase-2 (MMP-2) is implicated in ovarian cancer cell invasion, and previous data from our lab has demonstrated that LPA induces peri-cellular MMP-2 activity. Here we investigate the mechanism of LPA-stimulated MMP-2 activation through inhibition of various signaling pathways. Preliminary results using the DOV13 ovarian cancer cell line suggest a major role of PI3-kinase in LPA-induced MMP-2 activation, with a minor contribution from p38 MAPK. Furthermore, the obstruction of these pathways results in partial inhibition of LPA-induced cellular invasion, supporting a role for MMP-2 activity in this process. As MMP-2 activation is dependent upon both MT1-MMP and β 1 integrin, we are also investigating the effects of LPA on the expression and localization of these molecules, as well as determining the signaling pathways through which LPA exerts these effects. The results of these studies will aid in understanding the role of LPA in ovarian cancer metastasis, and may provide targets for therapeutic intervention to prevent the activation of MMP-2 and thus inhibit cellular invasion.

A learning algorithm diagnoses prostate and ovarian cancer from human serum based on multiparametric proteomic patterns

E.F. Petricoin III, A. Ardekani, B.A. Hitt, P.S. Hackett, D.K. Ornstein, A. Velasco, C. Trucco, C.P. Paweletz, W.M. Linehan, M.R. Emmert-Buck, E.C. Kohn, R. Trost, G. Mills, D. Fishman and L.A. Liotta
Div. of Therapeutic Proteins/Center for Biologic Evaluation and Research, Food and Drug Administration, Bethesda, MD, USA; Correlogic Systems, Inc., 6701 Democracy Blvd, Suite 300, Bethesda, MD, USA;

Urologic Oncology Branch NCI, NIH Bethesda, MD, USA; Department of Chemistry, Georgetown University, Washington, DC, USA; Department of Urology, Catholic University of Chile, Santiago, Chile; Department of Surgery/Division of Urology, UNC at Chapel Hill, Chapel Hill, NC, USA; Dept of Molecular Therapeutics, MD Anderson Cancer Center; Laboratory of Pathology, DCS, NCI, NIH, Bethesda, MD, USA

An artificial intelligence-based algorithm was used to analyze multiparametric proteomic data streams to identify buried diagnostic patterns of expression for cancer in the serum of men and women for prostate and ovarian cancer receptively. Rapid, high-throughput molecular weight profiles consisting of tens of thousands of data points were generated by SELDI-TOF (surface enhanced laser desorption and ionization time-of-flight) and analyzed by a novel, heuristic pattern recognition algorithm. The trained algorithm was accurate in predicting the presence of biopsy-proven prostate cancer in 92% of the cohort serums (48/52) while maintaining an overall specificity of 100% (62/62). Of this cohort, the algorithm was successful in predicting the presence of prostate cancer from 19/21 men whose PSA fell within the diagnostic grey zone of 4–10 ng/ml. Moreover, after gaining experience with a further sample set, the optimized algorithm was able to increase its overall accuracy and sensitivity to 96%. Most importantly, the algorithm was able to accurately predict ovarian cancer in XXX of XXX women, including XXX of XXX from serums of women with stage 1 cancer. Since a reliable marker for ovarian cancer does not currently exist, this finding warrants further validation with much larger study sets. Validation of this new tool is ongoing in larger study sets of serums for diagnostic utility for early detection of prostate, ovarian, breast, and colon cancer. Artificial intelligence-based algorithms that learn to detect subtle disease signatures within complex molecular patterns may have broad utility as a new paradigm for cancer diagnostics and early detection in the future.

MT1-MMP and ovarian cancer

David A. Fishman and Laura Graves
Northwestern University Medical School, Chicago, IL 60611, USA

Objectives: Metastasis of epithelial ovarian carcinoma involves a cascade of events, including tumor cell adhesion, migration, degradation of extracellular matrix components, invasion, and proliferation. Previous studies have demonstrated roles for members of the

matrix metalloproteinase family (MMPs) in the processes of tumor cell invasion and metastasis. Several MMPs, including MMP-2, MMP-9, as well as urokinase-type plasminogen activator (uPA), have been detected at high levels in ascites from ovarian cancer patients. These findings suggest a potential role for these enzymes in the proteolytic activities that facilitate invasion and metastasis in ovarian carcinoma. Furthermore, MMP-2 and MMP-9 were found to be associated with shed membrane vesicles in the ascitic fluid, which may therefore serve as concentrated stores of proteinase activity (Dolo, et al., *Clin. Exp. Metastasis* **17**(2), 131–140). Our studies address a possible role for an additional metalloproteinase, membrane-type 1 MMP (MT1-MMP), in the ascites-associated MMP pool.

Methods: The presence and relative levels of MT1-MMP were detected using Western analysis and a slot blot technique. Blots were probed with multiple antibodies raised to the hinge, catalytic, hemopexin, or transmembrane domains of MT1-MMP. MMP activity in ascites samples was analyzed with gelatin zymography. Membrane preparations were used to isolate shed vesicles from ascites fluid. These preparations were then used for Western analysis and a collagenase assay to determine the presence and activity of MT1-MMP.

Results: Our analyses demonstrate the strong presence of a single MT1-MMP species, via Western analysis, in ascites from a panel of ovarian cancer patients. Interestingly, this particular MT1-MMP protein is detected by an antibody to a transmembrane antigen, suggesting that it may be associated with shed vesicles. Ongoing investigations in the laboratory are aimed at addressing whether this MT1-MMP species is active, and whether levels detected in ascitic fluids correlate with clinical parameters.

Conclusions: Our studies suggest that MT1-MMP in ascites fluid may contribute to the metastatic potential of ovarian carcinoma cells. Furthermore, this ascites-associated MT1-MMP may be a component of shed membrane vesicles, which have been shown to contain high levels of other metalloproteinases. While previous reports have indicated that MT1-MMP may be released in Con A-treated culture media (Harayama, et al., *Jpn. J. Cancer Res.* **90**(9), 942–950), our results point to a possible *in vivo* role for shed MT1-MMP in the metastatic progression.

VEGF, SEMA3F and their common receptors Neuropilin 1 and 2 expressions in preinvasive bronchial lesions and lung tumors

S. Lantuéjoul^{a,b}, B. Constantin^c, H. Drabkin^d, C. Brambilla^a, J. Roche^c and E. Brambilla^{a,b}

^aLung Cancer Research Group-EMI INSERM 9924, Institut A Bonniot, CHU-UJF, France

^bService de Pathologie Cellulaire, CHU A Michallon, BP 217 38043, Grenoble cedex 9, France

^cIBMIG ESA CNRS 6031, Université de Poitiers, Poitiers, France

^dDivision of Medical Oncology, University of Colorado Health Sciences Center, Denver, CO, USA

Two receptors, NP1 and NP2, bind class III secreted semaphorins including SEMA3F, a gene which we isolated from a recurrent 3p21.3 lung cancer deletion. In addition, NP1 and NP2 bind VEGF and enhance the effects of VEGF binding to KDR/FLK-1. Elevated levels of VEGF were associated with the loss of SEMA3F membrane staining in lung cancer suggesting that they competed for common binding sites on their NP receptors (Brambilla et al., 2000). To determine the timing when these changes occur, we compared the protein expression, by immunohistochemistry, of VEGF, SEMA3F, NP1 and NP2 in 112 lung tumors and 50 preneoplastic lesions. The primary lung carcinomas included squamous (42), small-cell (9), adenocarcinoma (34), large cell neuroendocrine (12), basaloid, and typical and atypical carcinoids (11). Preneoplastic lesions were obtained from 25 patients with concomitant carcinoma consisting of mild (10), moderate (14) and severe dysplasia (12), plus carcinoma in situ (CIS, 14). The results were also compared with 19 hyperplastic and 10 squamous bronchial metaplasias. In preneoplastic lesions, VEGF increased progressively from low-grade to high-grade dysplasia ($p = 0.001$). In contrast, SEMA3F levels were low in all preinvasive lesions; this included absent staining or cytoplasmic delocalization in most high-grade preneoplastic lesions. NP1 and NP2 levels increased progressively in dysplasia and microinvasive carcinoma ($p = 0.0001$) and correlated with VEGF expression ($p = 0.04$ and 0.0002 , respectively). In lung cancers, NSCLC overexpressed VEGF and NP1 and 2 significantly more often than NE tumors. Most lung tumors exhibited absent or low SEMA3F expression with cytoplasmic delocalization, whereas carcinoids and bronchioloalveolar adenocarcinoma exhibited moderate SEMA3F expression with a predominant membrane pattern. SEMA3F loss or delocalization was correlated with advanced tumor stages although migrating cells overexpressed VEGF, SEMA3F and NP1-2. In summary, VEGF, NP1 and 2 levels are elevated during the progression of pre-malignant lesions in conjunction with the loss or mislocalization of SEMA3F. These results also suggest that NP1 and 2 may preferentially bind VEGF thus impairing SEMA3F func-

tion. In addition, these results support trials of VEGF or receptor antagonists as chemopreventive agents.

Human kallikreins: Gene discovery and clinical applications in cancer

Eleftherios P. Diamandis, George M. Yousef and Liu-Ying Luo

Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada and Department of Laboratory Medicine and Pathobiology, University of Toronto, Ontario M5G 1L5, Canada

The human kallikrein gene family was, until recently, thought to consist of only 3 genes, namely KLK1 (encoding for pancreatic/renal kallikrein), KLK2 (encoding for human glandular kallikrein 2) and KLK3 (encoding for prostate-specific antigen, PSA). PSA and hK2 are currently used as biomarkers for diagnosis and monitoring of prostate adenocarcinoma. Through efforts of our laboratory and others, a number of new kallikrein genes have been recently cloned. The human kallikrein gene family now consists of 15 genes which all map to the chromosomal region 19q13.4, encode for secreted serine proteases of similar structure and have a high degree of sequence identity. We have recently described the complete organization of the human kallikrein gene locus and positioned all kallikrein genes in tandem, with high accuracy (Yousef and Diamandis, *Endocr Rev* **22** (2001), 184–204).

For most newly discovered kallikreins, no immunological assays are available for their quantification in biological fluids. We have recently developed highly sensitive immunoassays for human kallikrein 6 (hK6) and human kallikrein 10 (hK10). These assays are suitable for quantification of these proteins in serum. We have screened large numbers of sera from patients with various malignancies and from normal control subjects. We found that the two kallikreins, hK6 and hK10, are highly elevated in serum of patients with ovarian cancer but not in serum of patients with other malignancies. Elevated levels of the two kallikreins were seen in approximately 70% of pre-surgical sera from ovarian cancer patients. These elevations can be used for diagnosis of the disease as well as for monitoring and prognosis. Serum elevations of the two kallikreins are dependent on ovarian cancer stage and grade, with higher stage and grade disease having higher positivity. We have identified subgroups of patients with low CA 125 levels who have elevated levels of hK6 and hK10. Consequently, it appears that the combination of CA 125 and

hK6 and hK10 is a more powerful biochemical marker profile for diagnosis and monitoring of ovarian cancer than the individual biomarkers alone.

We conclude that the two newly discovered kallikreins, hK6 and hK10, constitute novel ovarian cancer biomarkers which can be used in combination with the traditional biomarker, CA 125, for improved diagnosis and monitoring of patients with ovarian cancer.

Prostate cancer/benign prostate hypertrophy/normal classification based on high-dimensional ProteinChip[®] array mass spectrometry data: “Boosting” the power of the binary-marker combination approach

Y. Yasui, M.L. Thompson, M. Pepe, Z. Feng, Y. Qu, J. Potter, B.L. Adam, J.W. Davis, M.A. Clements, L. Cazares, P.F. Schellhammer, E.A. Dalmasso and G.L. Wright Jr.

Fred Hutchinson Cancer Research Center, Seattle, WA, USA, Virginia Prostate Center, Eastern Virginia Medical School, Norfolk, VA, USA and Ciphergen Biosystems, Inc., Fremont, CA, USA

At last year’s EDRN Scientific Workshop in Chicago, we presented two statistical approaches, Wavelets-based and Binary-Marker Combination (BMAC) methods, for analyzing high-dimensional protein expression data. The data were derived from serum samples analyzed by the EDRN Biomarker Developmental Lab at the Eastern Virginia Medical School (EVMS) using Ciphergen’s ProteinChip[®] System utilizing SELDI technology. The DMCC continued the collaboration with the EVMS team to identify protein profiles specific to each pathologic condition (prostate cancer, benign prostate hypertrophy (BPH), and normal). Here we extend the BMAC method with classification accuracy of 96.7% for prostate cancer cases, 100% for normals, and 40.0% for BPHs (note that a subset of “BPH” cases plausibly have cancerous lesions, a “misclassification” problem, which may explain the lower accuracy for the group). The method starts with pre-analysis processing of SELDI outputs, peak identification and calibration of the mass axis. We then use a simple algorithm called “boosting” to build an additive predictor of each pathologic condition using a set of linear univariate logistic classifiers. The boosting algorithm is applied to the classification of cancer/BPH vs. normal and then that of cancer vs. BPH. To deal with the misclassification problem of the “BPH” group, we identify the BPH cases that are likely to be cancer by a boosting analysis and then re-apply boosting excluding the “cancer-like”

BPH cases. Our stopping rule for boosting is based on Akaike's Information Criterion corrected for small-sample bias. The final classifier uses 20 protein mass peaks for distinguishing cancer/BPH vs. normal and 24 mass peaks for distinguishing cancer vs. BPH. The accuracy of the final classifier was estimated based on a test dataset ($N = 30$ cancer, $N = 15$ BPH, $N = 15$ normal), which is separate from the training dataset ($N = 167$ cancer, $N = 78$ BPH, $N = 81$ normal) used to develop the classifier.

Discovery of distinct protein patterns for normal, benign, pre-neoplastic and malignant prostate epithelial cells by SELDI mass spectrometry

Lisa H. Cazares^{a,d}, Bao-Ling Adam^{a,d}, Michael D. Ward^{a,d}, Suhail Nasim^{b,d}, Paul F. Schellhammer^{c,d} and George L. Wright, Jr.^{a,c,d}

Departments of ^aMicrobiology and Molecular Cell Biology, ^bPathology and Anatomy, ^cUrology, ^dVirginia Prostate Center, Eastern Virginia Medical School, and Sentara Cancer Institute, Norfolk, VA 23501, USA

SELDI mass spectrometry was used to detect protein alterations in cell lysates of pure populations of normal, benign (BPH), prostate intraepithelial neoplasia (PIN), and prostate carcinoma (PCA) cells procured by Laser Capture Microdissection (LCM). The objectives of this study were to discover potential biomarkers that could be used to differentiate malignant from non-malignant cell populations, especially early protein alterations that signal the initiation of a developing cancer. We hypothesize that these alterations in protein expression would be especially useful as potential markers for early detection and/or risk factors for development of prostate cancer. Cells were procured from each epithelial cell type (i.e., normal, BPH, PIN, PCA) identified in a prostatectomy specimen. Cell lysates were prepared from 40 samples procured from nine prostatectomy specimens. SELDI analyses was performed and the abundance level of each protein or peptide in the mass spectra was calculated and compared for each epithelial cell type (normal and diseased), including adjacent stromal cells. Several small molecular weight peptides or proteins (3000 to 5000 Da) were found over-expressed in the PIN and PCA cell lysates suggesting possible proteomic alterations occurring early in the disease process. Another peak, with an average mass of 5666 Da, was observed to be over-expressed in 86% of the BPH cell lysates, while only 22% of the PIN lysates and none of the PCA lysates over-expressed this protein. Expression differences were also found

for intracellular levels of prostate specific antigen (free-PSA), which was reduced in PIN and PCA cells when compared to matched normal epithelial cells. Although no single protein alteration was found in all PIN/PCA samples, combining two or more of the markers distinguished the benign cell types (N/BPH) from the diseased cell types (PIN/PCA), with sensitivities reaching 100%. Additional studies are in progress to validate these initial observations. Studies are also underway to isolate and identify these protein/peptides, and to determine if they are present in body fluids (i.e., serum, seminal plasma, urine) with the goal of developing a diagnostic test for the early detection of prostate cancer.

Assessing the diagnostic accuracy of repeated screening

Mary Lou Thompson

Department of Biostatistics, University of Washington, Seattle, WA, USA

Methods for characterizing the accuracy of a single diagnostic test are well-established. For a continuous test, for instance, the receiver operating characteristic (ROC) curve provides a summary which displays, for given diseased and non-diseased populations, the sensitivity and (1-) specificity associated with each possible test threshold. With repeated diagnostic tests as in, e.g., routine screening, each *individual* test is associated with an ROC curve but the accuracy of, say, k successive screens (or, put another way, T years of routine screening at intervals of t years) is not clearly defined. This issue has not received much attention in the literature.

We consider the definition of diagnostic accuracy in the context of repeated screening tests and explore factors which impact on this combined accuracy. Although our focus is repeated application of the same test, the general ideas developed here may also be applied to the setting of multiple tests at the same point in time. The complexity of diagnostic choices when two or more continuous tests are combined is illustrated and different approaches to reducing the dimensionality of the setting are discussed and evaluated. For instance, in practice when a single test is used repeatedly in routine screening, the same screening threshold is typically used at each screening visit. One possible alternative is to adjust the threshold at successive visits according to disease-specific or individual-specific characteristics.

We highlight some of the issues which should be considered (and exploited) in determining how multiple tests should be implemented and evaluated. These

ideas may be useful in designing a routine screening regimen. The ideas developed are illustrated using hypothetical examples and an application to screening for prostate cancer using total serum prostate-specific antigen (PSA), using data from the Beta-Carotene and Retinol Efficiency Trial, a randomized trial conducted at the Fred Hutchinson Cancer Research Center which included repeated serum samples from subjects enrolled in the trial.

Experiment design and power calculation for microarray experiment

Sue Li, Jeanette Bigler, Johanna Lampe, John Potter and Ziding Feng
Fred Hutchinson Cancer Research Center, Seattle, WA, USA

DNA microarrays are emerging as powerful and cost efficient tools to measure expression of thousands of genes simultaneously. In the past several years, microarray technologies have been used to identify genes that are differentially expressed between two types of tissues such as normal versus tumor or treated versus untreated, etc.. Choosing an efficient design is essential. A pertinent issue is how to decide on sample sizes to achieve adequate statistical power. A conventional sample-size calculation seems inappropriate here. Because the task is to select a number of genes (from among thousands) that are differentially expressed between two types of tissues, one criterion for choosing sample sizes should be that they are large enough to ensure that the differentially expressed genes have a high chance of being selected on the basis of the data. At the same time a good sample size should ensure a small risk of false positive findings.

Suppose that D is a set of truly differentially expressed genes of size K and D^c is the set of genes that are not differentially expressed of size $G - K$. The sample size should be chosen such that $Pr(\text{reject } H_i \text{ for all } i \in D) \geq \beta$ (say $\beta = 0.8$) (or $Pr(\text{reject } H_i \text{ for any } i \in D) \geq \beta$), where H_i is the null hypothesis that gene i is not differentially expressed, and $Pr(\text{reject } H_i \text{ for any } i \in D^c) \leq \alpha$ (say $\alpha = 0.05$). The first probability is an estimate of power and the second probability is an estimate of the false positive discovery rate. We propose to estimate the two probabilities by simulation. Suppose we observed two sets of gene expressions, $y_{1j} = (y_{1j1}, y_{1j2}, \mathbf{L}, y_{1jG})$, $j = 1, \mathbf{L}, n_1$, and $y_{2j} = (y_{2j1}, y_{2j2}, \mathbf{L}, y_{2jG})$, $j = 1, \mathbf{L}, n_2$, n_2 both from normal tissues in a pilot study. Suppose that we hypothesize that there are K expressed genes with b -

fold difference in expression. In the simulation, we chose one set of tissues (say the second set) as “tumor tissues” and then chose K out of G genes as “truly differentially expressed” genes and increased expression of these genes b fold. Then we created a further sample by re-sampling with replacement within the same type of tissues. We calculated a t-statistic for each gene to compare the differences in gene expression from tumor and normal tissues. These t-statistics may not follow a t-distribution; therefore, we used a resampling scheme (either permutation or bootstrap) to calculate p-values related to the t-statistics. We also used Westfall and Young’s step-down algorithm (Dudoit et al. 2000 applied to gene-expression data) to calculate p-values adjusted for multiple tests. We evaluated the validity of the two resampling methods. The performance of this procedure was examined. We applied the method to a crossover study of gene-expression profiles in colon tissues modulated by aspirin.

References

- P.H. Westfall and S.S. Young, *Resampling-Based Multiple Testing: Example and Methods for p-Value Adjustment*, Wiley Series in Probability and Mathematical Statistics, Wiley, 1993.
- S. Dudoit, Y.H. Yang, M.J. Callow and T. Speed, *Statistical Methods for Identifying Differentially Expressed Genes in Replicated cDNA Microarray Experiments*, Stanford technical report, 2000.

Identification of patients with ovarian cancer from SELDI protein spectra using mixtures of multivariate normal distributions

Steven J. Skates, Bin Ye, Sam Mok and Dan Cramer
Massachusetts General Hospital, Brigham and Women’s Hospital, Boston, MA, USA

Due to the low incidence of ovarian cancer, screening programs require an extremely high specificity and sensitivity so that a minimum positive predictive value of 10% is attained. Multiple potential biomarkers represented by peaks in surface enhanced laser desorption and ionization (SELDI) were measured in serum drawn pre-operatively from a group of ovarian cancer patients and compared to the same potential markers in serum from a group of control subjects to improve the operating characteristics. The standard approach to interpreting multiple markers is to declare a result positive if any marker is positive which may lead to an increase in sensitivity but with a decrease in specificity. We develop systematic methods for combining information

in multiple potential SELDI markers using mixtures of multivariate Normal distributions which increases simultaneously both sensitivity and specificity.

The SELDI serum markers were measured on 50 cases and 50 controls. The optimal method for combining information across markers is to calculate the ratio of the density for the cases to the density for controls at the point determined by the new subject's marker values. We estimate the two distributions from a sample of data using a mixtures of multivariate Normal distributions. We are developing an application of a method for estimating the parameters in the mixture models developed by Tibshirani and Hastie (1995) to the data described above. This process involves searching the peaks for an optimal panel which will produce the maximum sensitivity for a given specificities of 98% and 95%. The sensitivity will be obtained from an operating characteristic curve based on the approximation to the optimal decision method. Computational methods are in development for more rapidly exploring the space of panel subsets of the SELDI peaks to search for the optimal panel.

Artificial neural networks distinguish subtypes of neoplastic colorectal lesions

Florin Selaru^a, Yuriko Mori^a, Thomas C. Liu^a, Theresa Xu^a, Jing Yin^a, Tong Zou^a, Noam Harpaz^b, Anatoly Leytin^b, John M. Abraham^a, Fumiaki Sato^a, Suna Wang^a, Charlie Twigg^a, Andreea Olaru^a, David Shibata^a and Stephen J. Meltzer^{a,*}

^a*Department of Medicine, Division of Gastroenterology and Greenebaum Cancer Center, University of Maryland School of Medicine; Baltimore VA Hospital, 8-009 Bressler Research Building, 655 West Baltimore Street, Baltimore, MD 21201, USA*

^b*The Mt. Sinai School of Medicine, 1 Gustave L. Levy Place, New York, NY 10021, USA*

Colorectal dysplasias can occur either as sporadic adenomas or specifically in the setting of inflammatory bowel disease (IBD). The diagnosis, management and follow-up of patients with sporadic colorectal adenomas are well standardized, posing little clinical debate. Sporadic adenomas are diagnosed histologically, removed colonoscopically, and followed up with repeat surveillance colonoscopy at regular intervals. In contrast, IBD-related dysplasias mandate colectomy if

they are high-grade. It is difficult to distinguish between these two types of dysplasia on a histopathological basis alone, but this distinction is very important. In order to improve the ability to discriminate between the two types of lesion, we hybridized microarrays containing 8,000 cDNA clones to 39 colorectal neoplastic specimens composed of 31 sporadic adenomas or carcinomas (SACs) and 8 IBD-related dysplasias or cancers (IBDNs). An artificial neural network (ANN) was constructed and trained on a set of 5 IBDNs and 22 SACs. After training, this ANN correctly diagnosed 12 of 12 blinded samples in a test set (3 IBDNs and 9 SACs). Our proprietary software, GeneFinder, was used to decrease the number of genes used for classification. Through an iterative process involving GeneFinder, Cluster and Neural Networks, the number of clones used for the diagnosis was reduced from 8000 to 97. The ANN retained its capacity for correct diagnosis even after the number of clones was reduced to 97. These data demonstrate the ability of ANNs to distinguish IBDNs from SACs based on global gene expression data. We conclude that ANNs have the potential to discriminate among subtly different clinical entities, as well as to reduce the number of genes necessary to provide this level of discrimination.

Finding markers with a decision tree classifier for ProteinChip[®] data

Y. Qu, Z. Feng, Y. Yasui, M. Thornquist, J. Potter, B.L. Adam, J.W. Davis, P.F. Schellhammer, M.A. Clements, E. Dalmaso and G.L. Wright Jr.

Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Virginia Prostate Center, Eastern Virginia Medical School, Norfolk, VA, USA

Ciphergen Biosystems, Inc., Fremont, CA, USA

We used wavelet algorithm to classify prostate cancer, BPH, and normal sample. Though it works very well, the classifier is based on the wavelet coefficients and lack direct interpretation of predictive proteins. To address this issue, we propose a decision tree classification (DTC) method for identifying the potential protein markers for serum mass spectra data. The key features for this method include (1) dividing the spectra into N intervals with lengths equal to 0.4% of the molecular masses; (2) for each interval, determining whether there is a peak by using wavelet transform, and if there is, whether the peak is high (exceeding a threshold); (3) the locations and intensities of the peaks (features) are used for classification; (4) mutual information be-

*Address for correspondence: Stephen J Meltzer, M.D., 8-009 BRB, 655 West Baltimore Street, Baltimore, MD 21201, USA; E-mail: smeltzer@medicine.umaryland.edu.

tween the features and classes is used for feature selection and threshold determination. Some advantages of this method are: (1) it is easy to interpret the proteins predictive for classes; (2) it is a nonparametric method, no distribution assumption of the data is needed; (3) the computation is more efficient than the Fisher Discriminant analysis. The DTC method can be applied to microarray data. The performance of DTC is evaluated on both training dataset and validation dataset.

Selecting differentially expressed genes from microarray experiments

Margaret Sullivan Pepe^{a,b}, Gary Longton^b, Garnet Anderson^b and Michel Schummer^c
Dept. of Biostatistics, University of Washington, Seattle, WA, USA

High throughput technologies, such as gene expression arrays and protein mass spectrometry, allow one to simultaneously evaluate thousands of potential biomarkers that distinguish different tissue types. Of particular interest here are cancer versus normal organ tissues. We consider statistical methods to rank genes (or proteins) in regard to differential expression between tissues. Various statistical measures are considered and we argue that two measures related to the Receiver Operating Characteristic Curve are particularly suitable for this purpose. We also propose that sampling variability in the gene rankings be quantified and suggest using the 'selection probability function', the probability distribution of rankings for each gene. This is estimated via the bootstrap. A real data set derived from gene expression arrays of 23 normal and 30 ovarian cancer tissues is analyzed. Simulation studies are also used to assess the relative performance of different statistical gene ranking measures and our quantification of sampling variability. Our approach leads naturally to a procedure for sample size calculations appropriate for exploratory studies that seek to identify differentially expressed genes.

Application of support vector machines to the classification of ProteinChip[®] system mass spectral data of prostate cancer serum samples

Asa Ben-Hur^a, Tony Barnes^a, Peter Bartlett^a, Olivier Chapelle^a, André Elisseeff^a, Herb Fritsche^a, Isabelle Guyon^a, Bernhard Schölkopf^a, Jason Weston^a, Eric Fung^b, Cynthia Enderwick^b, Enrique A. Dalmasso^b, Bao-Ling Adam^c, John W. Davis^c, Antonia Vlahou^c, Lisa Cazares^c, Michael Ward^c, Paul F. Schellhammer^c, John Semmes^c and George L. Wright Jr.^c

^a*BIOwulf Technologies*

^b*Ciphergen Biosystems, Inc.*

^c*Eastern Virginia Medical School, Norfolk, VA 23507-1696, USA*

Support Vectors Machines (SVMs) are a recently introduced technique of pattern classification, which has proved successful in the recent years in many areas of application, including medical diagnosis. Successors to other well-established discriminant function techniques and neural networks, SVMs have consistently outperformed their predecessors in many benchmarks and are solidly grounded in theory. The key properties of SVMs are derived from the use of kernel similarity measures and support vectors, which are the most informative examples.

We have applied SVMs to the problem of classifying patients with prostate disease using the Ciphergen SELDI system. Our samples were extracted from the serum of 385 patients from 4 groups (group A: 93 samples, group B: 98 samples, group C, 98 samples, group D: 96 samples). The patients of group A have Benign Prostate Hyperplasia; B and C patients represent two grades of cancer; D is a set of reference healthy patients. The data consist of 385 time-of-flight mass spectra averaged over two repeated trials obtained by Ciphergen's ProteinChip[®] System utilizing SELDI technology. They were split into 325 examples for training and 60 for testing (15 from each group). In our experiments, we selected the kernel parameters and other hyper-parameters by cross-validation, using the training data only. The results reported are the performance obtained by the best system on the separate test set that was not used for training.

We report performance on the multiclass problem A, B, C, D, a difficult task that other techniques address usually indirectly. We use a particular SVM kernel (a polynomial kernel of degree 2) and some simple normalization on peaks manually detected. We obtain 85% percent correct classification, a significant improvement over the baseline method that was previously used. SVMs also provide a means of selecting a subset of relevant peaks. In our experiments, no significant degradation in performance is observed by removing up to 80% of the manually selected peaks. Additionally, we have devised a method for automatically detecting peaks that works in conjunction with SVMs and shows performance comparable to the manual peak detection. Our initial results show that mass spectra provide data that is particularly well suited to SVM analysis and our methods show promises in a wide range of ProteinChip System-based mass spec-

trometry applications, including medical diagnosis and biomarker discovery.

Clinically meaningful combinations of biomarkers for cancer screening: A prostate cancer case study

Ruth Etzioni^a, Charles Kooperberg^a, Robert Smith^a, Ian M. Thompson^b and Peter H. Gann^c

^a*Fred Hutchinson Cancer Research Center, Seattle, WA, USA*

^b*University of Texas Health Sciences Center at San Antonio, TX, USA*

^c*Northwestern University, Chicago, IL, USA*

Background. Use of the PSA test for prostate cancer screening has become relatively commonplace in the US. However, the test is prone to low specificity in certain subpopulations. To reduce false positive rates, researchers have suggested using percent free PSA to discriminate prostate cancer cases from healthy men when PSA levels are suspicious but moderate. Efforts have focused on identifying a threshold for percent free PSA when total PSA levels are in the reflex range between 4 and 10 ng/ml. However, some studies have suggested extending the reflex range to include lower PSA levels. The absolute level of complexed PSA (total minus free PSA) has also been proposed. A consensus about how best to combine information on free and total PSA has not been reached.

Objectives. The objectives of this study were (a) to systematically search the space of clinically meaningful rules combining total and percent free PSA to identify those with optimal sensitivity and specificity and (b) to evaluate whether the combination rules represented a significant advantage over the standard PSA-based rule. Clinically meaningful rules were considered to be the class of *logic* rules, namely those consisting of and-or combinations of threshold rules (*marker > c*) in each biomarker.

Methods. We used *logic regression*, a technique for identifying logic combinations of variables to predict an outcome. These methods were applied to PSA data from 429 prostate cancer cases and 1640 matched controls who had had serum stored at enrollment in the Physicians' Health Study. We searched for logic rules minimizing the overall misclassification rate (false positives plus false negatives). By differentially weighting the different types of misclassification errors we identified a sequence of logic rules with different true- and false-positive rates. Plotting the true- versus false-positive rates yielded the logic rule ROC curve which we compared with the ROC curves for rules based on

total PSA and complexed PSA. We then re-ran the logic regression algorithm entering age as a factor to determine whether age-specific combination rules provided any diagnostic advantage.

Results. The ROC curves for the combination rules identified several logic rules that provided better discrimination than PSA-based rules and rules based on complexed PSA, particularly for low false-positive rates. In this dataset the true and false positive rates for PSA > 4.0 ng/ml were 34.5% and 8.8%. The best combination rule with comparable sensitivity was PSA > 2.0 ng/ml and percent free PSA < 15, and yielded a false positive rate of 5.4%. Adding age as a factor led the logic regression to select rules that depended on age, but these did not appear to substantially improve diagnostic performance.

Conclusions. Logic regression provides a method for systematically and objectively identifying clinically meaningful combinations of markers to be used in prostate cancer screening. The results confirm the utility of percent free PSA in prostate cancer screening and support extending the reflex range to include lower PSA levels.

EDRN bioinformatics pilot project: Supporting cancer biomarker research

Heather Kincaid^a, Dan Crichton^b, Mark Thornquist^a, Steve Hughes^b, Lynn Anderson^c, Betsy Higgins^d, Robert Negm^e, Sudhir Srivastava^e and Greg Downing^f

^a*Fred Hutchinson Cancer Research Center, Seattle, WA, USA*

^b*California Institute of Technology, Jet Propulsion Laboratory, NASA, Pasadena, CA, USA*

^c*Moffitt Cancer Research Center, Tampa, FL, USA*

^d*University of Texas-San Antonio, San Antonio, TX, USA*

^e*Division of Cancer Prevention, National Cancer Institute, Bethesda, MD, USA*

^f*National Institutes of Health, Office of Science Policy, Office of the Director, Bethesda, MD, USA*

Knowledge discovery and data correlation are essential elements of cancer biomarker research that can be augmented by a unified approach to data management. The Early Detection Research Network (EDRN) comprises a very diverse informatics infrastructure across the various laboratories that constitute the network. The EDRN Knowledge System Working Group recognized the diversity of the EDRN and the need to build a collaborative informatics system to aid the network in the discovery of cancer disease biomarkers. As a result,

they undertook a pilot project with the objective of facilitating data sharing between EDRN centers through a web-based knowledge environment. The pilot project explores the feasibility of interconnecting laboratory databases while retaining their geographic and heterogeneous implementations. Increasing the ability for scientists to locate, share, and interpret data across laboratories without disturbing the traditional method for data collection was recognized as the principal goal of the project. In addition, four specific informatics goals were identified as follows: 1) understanding where the data resources reside, 2) understanding how data is accessed in each system, 3) interpreting and defining the underlying data models for each system, and 4) building an interface that demonstrates enterprise-wide query capabilities. The working group identified the sharing of bio-specimen data as an initial data set that could validate the feasibility implementation of the informatics system and aide scientists in the collaboration of specimens across the EDRN. In addition, it would support the goal of creating an integrated research network by deploying an informatics system that combines research databases and supports location transparency. The pilot identified Moffitt Cancer Research Center in Tampa, Florida and the University of Texas, San Antonio in San Antonio, Texas as initial sites to be integrated. Moffitt Cancer Research Center, a Biomarker Development Laboratory, provided access to the following data: Moffitt Cancer Registry, the Moffitt Research Specimen Banking System, and a Lung Cancer Cohort dataset and sputum bank. The University of Texas, San Antonio, a Clinical Epidemiological Center, provided access to the following: the San Antonio Center of Biomarkers of Risk for Prostate Cancer (SABOR) database and the Prostate Cancer Tissue Bank. The Data Management and Coordinating Center provided overall project management and has worked with the Jet Propulsion Laboratory in development of the overall informatics data system. This presentation will provide an update on the project and its current status, a demonstration of data sharing across the centers, and an example of queries that can be used to support cancer research. The presentation will also discuss the long-term vision for the data architecture and its plan to continue to expand to new sites and data sets supporting scientists in the discovery and validation of disease biomarkers within the network.

Computer assisted image analysis of cyclin D1 in patients with non-dysplastic Barrett's esophagus

L.A. Kresty, G.D. Stoner, D. Normolle, D.G. Beer, H.Appelman, T. Eaton, J. Fromkes, U. Sundaram, R.H.

Moseley, S. Spechler, A. Rustgi, T. Nostrant, S. Habib, S. Jewell, E. Hawk and D.E. Brenner

Ohio State University College of Medicine and Public Health, Columbus, OH, USA; Dallas VA Medical Center, Dallas, TX, USA; University of Pennsylvania, Philadelphia, PA, USA; The National Cancer Institute, Bethesda, MD, USA; University of Michigan Medical School, Ann Arbor, MI, USA

Barrett's mucosa is considered main precursor of adenocarcinoma of the distal esophagus and gastroesophageal junction. The specific molecular events in the progression of Barrett's esophagus to adenocarcinoma are incompletely understood, but presumably include alterations in genes controlling cellular proliferation, differentiation and apoptosis. Immunohistochemical techniques were employed to evaluate 68 biopsies from 38 baseline patients for the oncogenic marker Cyclin D1. Barrett's patients with cyclin D1 positive biopsy specimens have previously been reported to be at increased risk of progression to adenocarcinoma. Thus, improving our understanding of this biomarker may facilitate improved surveillance for Barrett's patients at greatest risk of malignant transformation to adenocarcinoma. Sources of Cyclin D1 variation were assessed between participating clinical sites and within sites at the patient, biopsy, slide, and region of interest levels. Each biopsy was scanned (> 50 glands on average) utilizing a brightfield microscope and computer-assisted image analysis. Positive nuclear staining for Cyclin D1 was found among 71% of the patients analyzed. Approximately 40% of the glands within a biopsy were D1 positive with mean cyclin D1 levels of 5.7% in positive glands (range 2.3 to 16.1%). Variation in Cyclin D1 levels by study site was homogenous and very small supporting consistency of sampling and processing methods among participating sites. Statistical analysis based on mean changes in raw Cyclin D1 values appeared inappropriate and uninformative due to the highly skewed distribution of the D1 data. Cyclin D1 staining indices showed extreme hot spots with large between biopsy variation within patients necessitating extensive harvesting of biopsies. A Logit transformation was performed to account for the skewed data distribution. Post-transformation the greatest variation in Cyclin D1 values were between biopsies within a patient (38.9%), followed by between patient variation (33.4%), variation between slides within a block (21.5%), and a small amount of variation between fields within a slide (6.2%). Our data support that study protocols utilizing cyclin D1 as a biomarker in Barrett's esophagus patients need to thoroughly sample

the esophagus, analyze multiple slides within a patient and employ appropriate statistical methods to account for the non-normal distribution of this biomarker in this patient population.

A simple method to improve probe set estimates from oligonucleotide arrays

Emmanuel N. Lazaridis, Dominic Sinibaldi, Gregory Bloom, Shrikant Mane and Richard Jove
H. Lee Moffitt Cancer Center & Research Institute at the University of South Florida, 12902 Magnolia Drive, MRC-CANCONT, Tampa, FL 33612, USA

A popular commercially available oligonucleotide microarray technology employs sets of 25 base pair (bp) oligonucleotide probes for measurement of gene expression levels. A mathematical algorithm is required to compute an estimate of gene expression from the multiple probes. Previously proposed methods for summarizing gene expression data have either been substantially ad hoc or have relied on model assumptions that may be easily violated. Here we present a new algorithm for calculating gene expression from probe sets. Our approach is functionally related to the “leave-one-out bootstrap”, a non-parametric statistical technique that is often applied in limited data situations. We illustrate this approach using data from our study seeking a molecular fingerprint of STAT3 regulated genes for early detection of human cancer.

A genetic variant of DNMT3B promoter and risk of lung cancer – A case-control study

Hongbing Shen^{a,*}, Luo Wang^{b,*}, Margaret R. Spitz^a, Waun K. Hong^b, Li Mao^b and Qingyi Wei^a
Departments of ^aEpidemiology and ^bThoracic and Head and Neck Medical Oncology, The University of M. D. Anderson Cancer Center, Houston, TX, USA

In previous molecular epidemiological studies, we have demonstrated that reduced DNA repair capacity (DRC), increased DNA adduct levels, and reduced expression levels of several nucleotide excision repair (NER) genes are associated with increased risk of lung cancer. It is biologically plausible that epigenetic alterations in expression of NER genes may be responsible for reduced DRC and inefficiency of removal of DNA adducts induced by smoking-related carcinogens, such as benzo[a]pyrene diol epoxide. Because methylation of promoter regions is one of the major regu-

latory mechanisms of gene expression and most promoters of NER genes are not fully described, we hypothesized that genetic variants of genes that are responsible for regulating the methylation status of other genes are associated with risk of lung cancer. To test this hypothesis, we investigated the association between a novel genetic variant of *de novo* cytosin DNA-methyltransferase (DNMT)-3B and risk of lung cancer incident lung cancer cases and 340 healthy controls. This was a hospital-based case-control study with frequency matching by age (± 5 years), sex, ethnicity and smoking status. All subjects were Caucasians. The mean age was 59.6 and 59.8 years for the cases and the controls, respectively. The proportion of male subjects was 53.5% for the cases and 55.9% for the controls. The variant was a C to T transition at -149 nucleotide of the promoter of *DNMT3B*, which causes a twofold increase in the promoter activity. This variant creates a AvrII site and was detected by the PCR-based restriction fragment length polymorphism method, which resulted in two recognizable alleles T (380 bp; uncut) and C (207 bp; cut) and three corresponding genotypes: CC, CT, and TT. The distribution of these three genotypes of the controls is in agreement with Hardy-Weinberg disequilibrium ($P = 0.125$). The variant T allele frequency was 0.494 for the cases and 0.441 for the controls, and the difference is borderline significant ($P = 0.056$). Compared with the CC genotype, the CT heterozygotes were associated with more than twofold increased risk (adjusted odds ratio (OR) = 2.13; 95% confidence interval (CI) = 1.47–3.08), whereas the TT homozygotes were associated with slightly increased risk (adjusted OR = 1.42; 95% CI = 0.91–2.21). The combined variant genotypes (CT + TT) was associated with nearly twofold increased risk (adjusted odds ratio = 1.88; 95% confidence interval = 1.32–2.66). The risk associated with the combined variant genotypes was more pronounced in subgroups of the subjects who were younger (< 60 years) (OR = 2.34), female (OR = 2.72) or never smokers (OR = 2.71). The results suggest that this novel variant (C \rightarrow T) of *DNMT3B* resulting in a twofold increase in promoter activity is associated with increased risk of lung cancer. This finding further suggests that this variant is of biological relevance, which, once confirmed in prospective studies, has a strong implication in identifying individuals genetically susceptible to tobacco-induced cancers for targeted cancer prevention.

*Contributed equally.

Proposal to develop a cancer specific proliferating cell nuclear antigen (csPCNA) tumor marker immunosensor

Cha-Mei Tang^a, Platte T. Amstutz^a, Gary Siebert^b, David A. Yost^c and Richard G. Saul^c

^a*Creativ MicroTech, Inc. Rockville, MD 20850, USA*

^b*Minerva Pharmaceuticals, Inc., Baltimore, MD 21230, USA*

^c*Axo Diagnostics, LLC, Rockville, MD 20850, USA*

We propose to develop a new cancer screening and diagnostic based on the cancer specific proliferating cell nuclear antigen (csPCNA) using a new sensitive immunosensor technology.

Proliferating cell nuclear antigen, or PCNA, is one of the components of the DNA replication complex (DNA Synthesome). DNA replication/repair in cancer cells is found to be significantly more error-prone than that in non-malignant cells. This decrease in replication fidelity can be tracked by following a structural alteration of PCNA. Non-malignant cells express only the normal form of PCNA, which exhibits a basic isoelectric point (pI), while cancer cells express a second form of the protein, cancer specific PCNA, or csPCNA, which has an acidic pI. This difference in isoelectric point is not due to a mutation in the gene for PCNA, but appears to be epigenetic in nature. The two forms of PCNA can be isolated from malignant cell lines.

csPCNA is potentially an early detection cancer marker. It has been shown to be present exclusively in 100% of the cancer cell lines and cancer tissues tested, and has not been found in any matched normal cell lines, normal tissues, or benign tumors.

Dr. Linda Malkas (UMAB) demonstrated a binding assay capable of discriminating between the two forms of PCNA using the xeroderma pigmentosum (XP) G protein. Preliminary data was presented in abstract 2507 from the AACR meeting in March, 2001 titled "Detection of the Cancer Specific Form of PCNA by Elisa Assay". An assay based wholly on antibodies is not yet practical since all commercially available antibodies have been unable to differentiate between the malignant and the non-malignant forms of the antigen. This new csPCNA specific assay will allow detection of csPCNA and serve as the foundation of the proposed cancer diagnostic test.

Creativ MicroTech is developing an immunosensor. This immunosensor has high selectivity, high sensitivity, low limit of determination, and fast response time. The technology can be miniaturized and integrated with microfluidics and other microfabricated components.

The immunosensor technology will be applied to the detection of csPCNA.

Gastrin-Releasing Peptide Receptor (GRPR) expression is associated with lung cancer risk

Jill M. Siegfried, William E. Gooding, Autumn Gaither Davis, Toni Hopkins, Joseph Pilewski, York E. Miller and Wilber A. Franklin

Lung Cancer SPORE and EDNR Program, University of Pittsburgh Cancer Institute, PA, USA and Lung Cancer SPORE and EDNR Program, University of Colorado-Denver, CO, USA

Women make up an increasing proportion of new lung cancer cases, and also comprise the majority of nonsmokers diagnosed with lung cancer. Evidence is conflicting regarding excess lung cancer risk in women who smoke; while some studies have documented increased odds ratios in women compared to men after adjustment for degree of tobacco exposure, others have not. Over 80% of lung cancer in non-smokers occurs in females, thus baseline rates of lung cancer in female vs. male nonsmokers may affect these odds ratio calculations, when comparisons are made only among individuals of the same gender. We examined lung cancer risk in a case-control study of individuals who underwent bronchoscopy or lung resection (either for lung cancer or a non-neoplastic condition), and related risk to the joint effect of gender, smoking history, and/or expression of the gene GRPR. GRPR is an X-linked gene that is known to escape X-inactivation and shows more frequent expression in female nonsmokers compared to males. Of 106 lung cancer cases and 57 controls, odds ratios for smoking were computed for the entire cohort and showed significance ($p < 0.0001$) for trend related to smoking alone, as expected. When gender specific calculations were made, odds ratios for extent of smoking in females were not significantly different from the odds ratios for extent of smoking in males. However, there were three times as many female never-smoking lung cancer cases as male never-smoking lung cancer cases in the cohort (9 female cases and 3 male cases). This skewed distribution has been commonly observed. The joint effect of smoking and gender was then examined in computing odds ratios. Using never-smoking males as a baseline reference, the female odds ratio for never smokers was 4.3. Odds ratios for smokers using never-smoking males as the reference were: 1–25 pack-years (PY), 3.1 for males vs. 7.6 for females; 26–50 PY, 6.5 for males vs. 20.6 for females; 51+ PY, 10.7 for males vs. 45.5 for females. When

never-smoking females were used as the baseline reference, we found significantly reduced odds ratios for males (0.2 for never-smoking males and similarly decreased odds ratios for males in the different smoking categories). These findings indicate that females are at increased risk for lung cancer compared to males, regardless of smoking history, and that female smokers show higher odds ratios than male smokers at all smoking levels if baseline gender differences are taken into account. We further examined the joint effects of gender, smoking and GRPR status. We found that expression of GRPR reduced the magnitude of gender differences in odds ratios. In comparing males and females positive for GRPR expression, the increase in female odds ratios at each smoking level was approximately 1.2 to 2-fold, compared to 1.5 to 4-fold in individuals negative for GRPR expression. Joint odds ratio calculations taking into account gender, smoking history and GRPR status showed that females who smoke and are GRPR positive are at the highest lung cancer risk. For example, the odds ratio for male smokers of 26–50 PY who are GRPR positive was 27.0 (CI 2.1-1278, $p = 0.037$) compared to 66.0 for females (CI 4.1-3064, $p = 0.003$). Similar differences were seen in all smoking categories. However, expression of GRPR accounts for only part of the increased risk for lung cancer observed in females compared to males, since GRPR negative females also had higher odds ratios than GRPR negative males. (Supported in part by EDNR/BDL grant U01 CA84968.)

Amplification/overexpression of a mitotic kinase gene is associated with aneuploidy and aggressive phenotype of human bladder cancer

Subrata Sen^a, Hongyi Zhou^a, Ruo-Dan Zhang^a, Dong SupYoon^a, Funda Vakar-Lopez^a, Shigemi Ito^a, Feng Jiang^a, Dennis Johnston^a, H. Barton Grossman^a, Arnout C. Ruifrok^a, Ruth L. Katz^a, William Brinkley^b and Bogdan Czerniak^a

^aUniversity of Texas M.D. Anderson Cancer Center, Houston, TX, USA and ^bBaylor College of Medicine, Houston, TX, USA

Background: We have shown that amplification and over-expression of a recently cloned gene STK15/BTAK/*aurora2* causes polyploidization of centrosomes and abnormal segregation of chromosomes resulting in the aneuploid transformed phenotype. We hypothesized that STK15 amplification and over-expression may correlate with clinically aggressive aneuploid variant of bladder cancer.

Design: STK15 expression levels were tested by immunohistochemistry and correlated with the gene amplification revealed by fluorescence in situ hybridization (FISH) and Southern blot analyses. Semiquantitative protein expression levels demonstrated by immunohistochemistry were related to various pathological parameters of the tumors, their DNA ploidy, and follow-up data in 160 patients with transitional cell carcinoma (TCC) of the bladder.

Results: Immunohistochemical studies revealed that nearly 70% of high-grade invasive aneuploid tumors showed over-expression of STK15 while no or minimal levels of its expression were detected in normal urothelia or superficial low grade diploid tumors. Over-expression of STK15 was strongly associated with high histological grade, aneuploidy, propensity for invasion and metastasis, as well as decreased overall survival of patients with TCC. FISH and Southern blot studies documented that the high levels of STK15 amplifications (> 4 copies) were associated with pronounced aneuploidy. Tumors with no or minimal amplification of STK15 (< than 4 copies) were diploid or near-diploid.

Conclusion: Amplification of STK15 and its over-expression is strongly associated with pronounced aneuploidy and clinically aggressive behavior of bladder cancer.

Preliminary data on association between pre-diagnostic serum levels of fatty acid synthase (FAS) and breast cancer risk: A nested-case control study in the CLUE cohort

Craig J. Newschaffer^a, Hong Lai^a, Frank Kuhajda^b and Kathy Helzlsouer^a

^aThe Johns Hopkins Bloomberg School of Public Health, Department of Epidemiology, Baltimore, MD, USA

^bThe Johns Hopkins University School of Medicine, Department of Pathology, Baltimore, MD, USA

Background: Fatty acid synthase (FAS) is a multi-enzyme protein consisting of two identical subunits which catalyzes the NADPH dependant condensation of malonyl-CoA and acetyl-CoA to produce the 16-carbon saturated fatty acid, palmitate. With a normal diet, the expression of FAS in normal, adult human tissues (except for the cycling endometrium and lactating breast) is low. However, cells from a variety of different solid tumor types, including breast tumors, have been found to constitutively express FAS. Moreover, FAS expression in pre-cancerous breast tissue at early stages of malignant transformation is also markedly

higher than that in normal breast tissue suggesting that activation of the fatty acid synthesis genetic program occurs early in the carcinogenic process. FAS is detectable in human sera with ELISA technology and circulating levels of FAS have been found to be higher among breast cancer patients (also, prostate, colon, and ovarian) than controls. However, it is yet to be determined whether the pre-diagnostic serum FAS levels predict subsequent development of cancer.

Specific aim: To examine the association between prediagnostic serum FAS levels and the development of breast cancer using data and biosamples available in the CLUE cohorts.

Study design and methods: Nested case-control study of 500 breast cancer cases and matched non-cancer controls with archived serum samples available from 1974 and 1989 (approximately 150 pairs with samples available in both years). Case and control FAS levels will be compared. Analyses will be stratified by time from blood draw to diagnosis, case disease stage, and case vital status. Potential confounders not accounted for by matching (e.g. BMI) will be controlled for in the analyses. Circulating FAS will be assayed based on monoclonal-mono-clonal and monoclonal-polyclonal anti-human ELISA assays developed by Dr. Kuhajda.

Results: At this point, FAS assays have been performed on 50 case-control pairs where the case was diagnosed between 1990 and 1998 and each member of the pair had two available biosamples. Results from the monoclonal-mono-clonal assays are as follows. Mean FAS levels (ng/ml) in 1974 were 12.9 (sd 14.6) and 10.7 (sd 8.0), respectively. For the 1989 samples, mean levels were 17.7 (sd 30.1) and 14.7 (sd 15.7) for cases and controls respectively. The p-values for the paired t-tests of case-control differences were 0.36 for the earlier and 0.53 for the latter samples. The mean change 1974-to-1989 among cases was 7.0 ng/ml (sd 29.5) and among controls was 1.7 (sd 16.4). The p-value for the paired t-test of the case-control difference here was 0.28.

Conclusion: Mean 1989 serum FAS levels and 1974-to-1989 serum FAS level increases for breast cancer cases diagnosed 1990-98 were higher than those for matched controls. However, at the current sample size, these differences were not large enough to attain statistical significance at conventional alpha error tolerances. Moreover, there was substantial overlap of the serum FAS distributions across case and controls. We hope to have data available on 50 additional case-control pairs by the end of October. This additional sample size

should allow for the estimation of more stable estimates of case-control differences as well as analyses stratifying on the potentially important covariates of time from blood-draw to diagnosis and stage of disease at diagnosis.

Determination of biomarkers of prostate cancer risk in the U.S. White of Hispanic origin population as determined by acycloprime FP SNP analyses

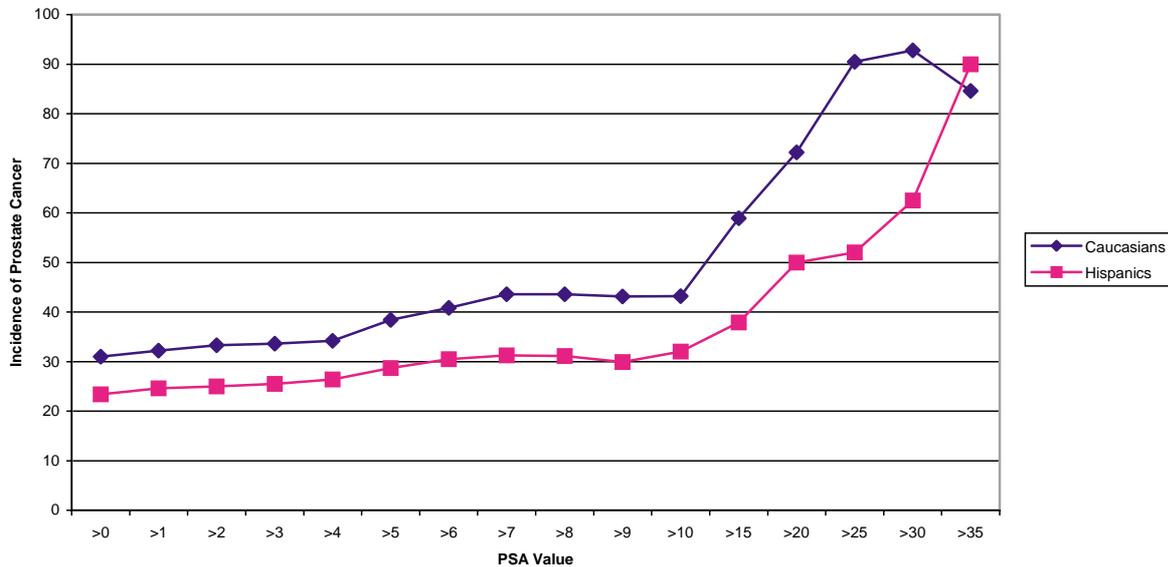
Ivana Balic^a, Robin J. Leach^{a,b}, Dean A. Troyer^c, Ian M. Thompson^d and Teresa L. Johnson-Pais^b

Departments of ^a*Cellular and Structural Biology,* ^b*Pediatrics,* ^c*Pathology,* ^d*Surgery University of Texas Health Science Center-San Antonio, San Antonio, TX 78229, USA*

An increased risk of developing prostate cancer (CaP) has been shown to be associated with single nucleotide polymorphisms (SNPs) within certain prostate-related genes. Two of these loci are key genes involved in the biosynthesis of testosterone. The CYP17 gene encodes the cytochrome P450c17 alpha enzyme, which mediates two key steps in testosterone synthesis. A SNP has been identified in the 5' untranslated region of the gene that may influence transcriptional activity. A 2.57-fold increase in the risk of CaP is found in men homozygous for the A1 allele in the CYP17 gene. The SRD5A2 gene product converts testosterone to the active form of dihydrotestosterone. Two SNPs have been identified in the coding region of the SRD5A2 gene. A 3.6-fold increase in the risk of CaP in Hispanic men is associated with the A49T missense mutation. Other prostate-related genes with CaP associated SNPs are the prostate-specific antigen (PSA) gene and the HPC2 gene. PSA is one gene regulated by testosterone and has been shown to have a SNP in one of the androgen response elements in the promoter region. Men with the PSA GG genotype were shown to have a 2-fold increase in their risk for developing CaP. HPC2 is a hereditary prostate cancer susceptibility locus. Two common missense SNPs cause a Ser to Leu change at amino acid 217 and an Ala to Thr change at amino acid 541. The risk of CaP is significantly increased in men with the Leu217/Thr441 variants.

Genotyping of these SNPs has not previously been performed in large cohorts of White Hispanic origin men, and it is well documented that White men of Hispanic origin show a decreased risk of prostate cancer as compared to African American men. As part of the San Antonio prostate cancer study of biomarkers of risk (SABOR) project, we have obtained 103 blood

Detection of Prostate Cancer by Increasing PSA



samples from White men of Hispanic origin at the age of 50 years or older with a PSA value of less than two and a negative digital rectal examination. We also have access to a prostate cancer tumor bank of 100 samples from White men of Hispanic origin. Using the technique of acycloprime FP (Perkin Elmer) SNP analysis, we are rapidly genotyping the SNP alleles for SRD5A, CYP17, PSA, and HPC2 in DNA from White Hispanic men with no evidence of prostate cancer and comparing that data to the allele frequency found in DNA from the prostate tumor specimens. The SABOR study provides us with this unique opportunity to genotype these markers in a large cohort of White Hispanic men. Identification of polymorphisms associated with increased risk of CaP may be useful for the early detection of disease.

Prostate cancer in the Hispanic population

Carlos E. Bermejo, Jeffrey A. Leslie and Ian M. Thompson

University of Texas Health Science Center at San Antonio, TX, USA

Introduction: The Hispanic population is the fastest growing minority in the United States. While the predictive value of PSA in African American and Caucasian populations have been compared there are few data that describe the performance of this test among Hispanic men. We sought to address this question in a population of men from San Antonio, TX.

Materials and method: A review of 877 patients who underwent prostate biopsy between January 1999 and October 2000 was conducted. Patients were categorized by ethnicity as Hispanics or Caucasian. The predictors of prostate cancer (PSA value, age, prostate volume) were described and the performance of PSA in the two groups compared.

Results: No significant difference between Hispanic and Caucasian men was noted with relationship to age, PSA value and prostate volume ($p > 0.01$). Incidence of prostate cancer in these two groups, however, was significantly different: 23.4% and 31%, respectively ($p < 0.01$). The figure illustrates the likelihood of a positive biopsy based upon PSA value.

Conclusions: In this population, Hispanic men had a lower incidence of prostate cancer and, for a given PSA value, the likelihood of a positive biopsy was consistently less until very high PSA levels were reached. These data provide further evidence of the importance of studies analyzing biomarkers of disease in the three major ethnic populations in the US and the conduct of hypothesis-generating studies to seek an explanation for these differences.

The frequency of malignant biopsies in patients with a positive digital rectal examination and a prostate specific antigen < 4

Aldo A. Ghobriel, Carlos E. Bermejo, Jeffrey A. Leslie, Dean Troyer, Betsy Higgins and Ian M. Thompson

*The University of Texas Health Science Center at San Antonio, San Antonio, TX, USA
(Presentation to be made by Aldo A. Ghobriel, MD.)*

Introduction and objectives: The utility of Prostate Specific Antigen (PSA) for prostate cancer detection is unquestionable with approximately 25% of patients with a PSA > 4.0 ng/ml having disease. At issue is whether digital rectal examination (DRE) is also necessary in the group with a normal PSA. Thus, whether PSA can obviate the necessity of DRE in this population. As a part of the San Antonio center for Biomarkers of Risk, an Early Detection Research Network Clinical and Epidemiologic Center, the DRE is an annual early detection test. We sought to determine the likelihood of a positive prostate biopsy (i.e., the utility of DRE as a marker of the disease) in patients with an abnormal DRE but normal PSA.

Materials and methods: A review of 971 patients who had undergone a prostate biopsy between Jan 1999 and Oct 2000 was performed. Patients were grouped by PSA values < 1, < 2.5, < 4 and > 4. The number of patients in each category with a) an abnormal DRE and b) a positive biopsy was determined. We then segregated those with positive biopsies to poorly differentiated (Gleason's score 7 or more) and well- or moderately-differentiated (Gleason's score 6 or less).

Results: The groups of PSA < 1 (50 patients), PSA < 2.5 (101 patients), PSA < 4 (170 patients) and PSA > 4 (801) patients were found to have abnormal DREs in 45, 92, 130 and 286 cases, respectively. Prostate cancer was detected in 2, 10, 20 and 255, respectively. The percentage of patients with positive DREs who were found to have cancer on biopsy was 4.4%, 10.87%, 15.38% and 96.5%, respectively. The positive biopsies were found to be poorly differentiated in 0%, 50%, 40% and 52.53%.

Conclusions: In our patient population, those with positive DRE and a PSA < 2.5 have a 10.87% chance of having a malignant biopsy. Of this group, there is a 50% chance of having a poorly differentiated tumor. Those with a positive DRE and a PSA < 4 were found to have a 15.38% chance of a malignant biopsy, of which 40% are poorly differentiated. The only population for whom a DRE could have been omitted might be those men with a PSA < 1.0 ng/ml in whom biopsy was positive in only 4.4% and, in all of these, the tumor was of the lowest biologic aggressiveness.

Race and age-dependent alterations in global methylation of DNA in squamous cell carcinoma of the lung

C.J. Piyathilake, O. Henao, M. Macaluso, A.R. Frost,

G.L. Johannig, W.C. Bell, D.C. Heimburger, A. Niveleau and W.E. Grizzle

The University of Alabama at Birmingham, AL 35294, USA and University Joseph Fourier of Grenoble, 38706 La Tronche, France

We previously reported that alteration in global DNA methylation is an epigenetic difference in susceptibility for the development of squamous cell cancer (SCC) of the lung. The current study investigated the race and age-dependent alterations in global DNA methylation on the development and progression of SCCs of the lung. Global methylation status was evaluated in SCC and in the associated uninvolved bronchial mucosa and epithelial hyperplasia of 53 Whites and 23 African Americans by using an antibody specific for 5-methyl cytosine (5-mc). A low 5-mc score indicates global hypomethylation of DNA. 5-mc scores of SCC (0.59 ± 0.06) were significantly lower compared to 5-mc scores of uninvolved bronchial mucosa (UBM) (0.87 ± 0.07) and epithelial hyperplasia (EH) (0.82 ± 0.07) in Whites ($p < 0.05$). In African Americans, 5-mc scores of SCC (0.55 ± 0.09) were not significantly different from 5-mc scores of UBM (0.60 ± 0.09) and EH (0.54 ± 0.14), suggesting an involvement of methylation in the development of SCCs in Whites, but not in African Americans. 5-mc scores were lower in younger (< 65-years) subjects compared to older (> 65-years) subjects in Whites. Since cancers in younger subjects tend to be more aggressive than cancers in older subjects, these observations may suggest that hypomethylation may have contributed to aggressiveness cancers of younger Whites. Hypomethylation of SCCs in White men was associated with shorter survival from the disease. These preliminary results suggest that the methylation status of DNA may affect the development, aggressiveness and prognosis of SCCs in Whites. It is unclear whether the inconsistencies across race and gender subgroups are an effect of selection and size of the study groups. Replication of this study in other populations is necessary to increase the scientific credibility of the observed results.

ErbB-2/c-neu, Epidermal Growth Factor Receptor (EGFR), Bcl-2, and p53 in breast carcinomas in pre- and post-menopausal women

Andra R. Frost^a, Walter A. Bell^a, Santosh Niwas^b, Lynya I. Talley^b, David Chhieng^a, Georgia Moore^a and William E. Grizzle^a

^aUniversity of Alabama at Birmingham, Department of Pathology, Birmingham, AL, USA

^b*Biostatistics Unit of the University of Alabama at Birmingham Comprehensive Cancer Center, Birmingham, AL, USA*

Introduction: Breast carcinomas in young, pre-menopausal women have more adverse biological characteristics and an overall worse prognosis than those in older, post-menopausal women. It has been hypothesized that these differences are a result of the higher histologic grade of pre-menopausal breast carcinomas.

Methods: To address this issue, we compared the expression of several prognostic indicators (ErbB-2/c-neu, EGFR, and Bcl-2 expression and nuclear accumulation of p53) in infiltrating ductal carcinomas (IDC) of similar histologic grade in pre-menopausal women ≤ 45 and post-menopausal women ≥ 65 . One hundred ninety six invasive ductal carcinomas (IDC) and 113 concomitant ductal carcinoma in situ lesions (DCIS) from pre-menopausal ($n = 99$), and post-menopausal ($n = 97$) women were studied. Formalin-fixed, paraffin-embedded tissue sections were immunohistochemically stained for ErbB-2/c-neu, EGFR, Bcl-2 and p53. ErbB-2/c-neu, EGFR and Bcl-2 stains were given a semi-quantitative immunoscore from 1 to 4. p53 was assessed both as a percentage of stained cells (%+) and an immunoscore. Staining above (high) and below (low) the median immunoscore or median %+ (for p53) of all cases was compared between groups.

Results: No differences in expression of ErbB-2/c-neu or EGFR (cytoplasmic or membrane staining) or Bcl-2 in IDC from pre- versus post-menopausal women were found. IDC in pre-menopausal women were more likely to express p53 (above vs. below the median) than those in post-menopausal women ($p = 0.02$). There was increasing expression of cytoplasmic and membrane EGFR and cytoplasmic ErbB-2/c-neu in the progression from histologically normal epithelium (NL) to DCIS to IDC, whereas membrane expression of ErbB-2/c-neu was highest in DCIS, followed by IDC then NL. Bcl-2 expression decreased from NL to DCIS to IDC. p53 nuclear accumulation increased from NL to DCIS, but was similar in DCIS and IDC. These changes with progression were similar in pre- and post-menopausal women. When comparing high and low grade IDC, cytoplasmic and membrane ErbB-2/c-neu and membrane EGFR expression were not significantly different, whereas cytoplasmic EGFR and Bcl-2 expression were stronger in low than high grade IDC ($p = 0.04$ and $p = 0.001$, respectively) and p53 was more frequently expressed in high than low grade IDC ($p = 0.003$).

Conclusions: After adjusting for histologic grade, p53 overexpression may be more frequent in pre- than

post-menopausal women; however, there were no statistically significant differences in expression of ErbB-2/c-neu, EGFR, and Bcl-2 in pre- and post-menopausal breast cancers. Furthermore, alterations in expression of ErbB-2/c-neu, EGFR, Bcl-2 and p53 in the progression from normal epithelium to invasive carcinoma did not differ in pre- vs. post-menopausal women.

Ethnic differences in the mutational spectra of the p53 gene in colorectal adenocarcinomas

Uppender Manne, Bernard Gary and William E. Grizzle
Department of Pathology, University of Alabama at Birmingham, AL 35294, USA

Previous studies from our laboratory showed that the prognostic value of abnormal p53 expression, assessed by immunohistochemistry, is a strong predictor of poor survival only in Caucasian patients with colorectal adenocarcinoma (CRC) [Cancer (1998) 83: 2456]. To assess the prevalence and types of genetic alterations in the p53 gene, we analyzed the status of p53 in CRC and matching benign colonic epithelial tissue samples prospectively collected from 11 African-American and 18 Caucasian patients. Mutational analysis of p53 was determined by direct sequencing of the p53 cDNAs amplified in RT-PCR. The primers used in this study amplify the coding region covering 40-319 codons of exons 4 through 9. This analysis will detect both single nucleotide substitutions and insertions or larger deletions in the expressed transcript of p53.

Analysis of the current study indicated that 6 out of 11 (54%) CRCs from African-Americans exhibited point mutations. Four of these point mutations were located at codon 72 (4 of 11, 36%) and the additional two mutations were present at codon 286. The mutation at codon 72 has been described as an allelic polymorphism in the p53 gene in several cancers including CRC. In Caucasian patients only 6 out of 18 (33%) CRCs exhibited point mutations. All these point mutations were located at 'hot spot' codons [175, 238, 248, 273, and 282 (in two samples)]. Contrary to African-Americans none of the CRCs from Caucasians exhibited mutations at polymorphic site codon 72.

These preliminary results suggest that the observed ethnic differences in the prevalence and the pattern of point mutations in the p53 gene may contribute to different biologic consequences of phenotypic expression of p53. These findings may also indicate distinct genetic pathways involving p53 in the development of colorectal neoplasia in African-American and Caucasian patients.

Second primary tumors in patients with upper aerodigestive tract cancers: Joint effects of smoking and alcohol

Kim-Anh Do, Dorota A. Doherty, Fadlo Khuri, J. Jack Lee, Waun K. Hong and Margaret R. Spitz
The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030, USA

Background: Tobacco smoking and alcohol consumption have long been identified as the two most important risk factors for the development of cancer of the upper aerodigestive tract. The development of second primary tumors (SPT) has also been associated with these risk factors. We evaluated the effects of baseline tobacco smoking and alcohol consumption on the incidence of SPT in patients with previous squamous cell carcinoma of the oral cavity, pharynx, and larynx. We also present methodology for longitudinal analysis of the relevant data.

Method: A randomized, placebo-controlled chemoprevention trial of the efficacy of low-dose 13-cis retinoic acid (13cRA) in patients with head and neck cancer commenced in 1991, and currently 1181 eligible randomized patients are being followed for survival and development of SPTs. Data on smoking history and alcohol consumption were collected at baseline and during the follow-up. The effects of smoking and alcohol on development of SPTs were examined using logrank test, and the Cox proportional hazards model was applied to analyze the effects of single and multiple risk factors in association with the development of SPTs.

Conclusions: Cigarette smoking and alcohol consumption emerged as influential risk factors for the development of SPTs in this patient population; however, tobacco appeared to be a stronger risk factor compared with alcohol. Smoking cessation and avoidance of alcohol is likely to reduce the risk of a new cancer diagnosis in patients with prior head and neck cancers.

The development of a Pancreatic Cancer (PC) surveillance program in families with pancreatic cancer-Familial Atypical Multiple Mole Melanoma (FAMMM) syndrome

Randall E. Brand^a, Joseph Anderson^a, Candace Grier^a, Aurelio Matamoros^a, Carolyn Deters^b, Christine Young^a and Henry Lynch^b

^a*University of Nebraska Medical Center, Omaha, NE, USA*

^b*Creighton University Medical Center, Omaha, NE, USA*

Hereditary factors account for 10% of PC cases. Healthy members of PC-prone families should be considered for surveillance studies with the hope of detecting the disease at a premalignant stage. The FAMMM syndrome is an autosomal dominant inherited disorder characterized by the occurrence of multiple atypical nevi and malignant melanomas. An increase in PC has been observed in a subset of FAMMM families with a p16 germline mutation. This provides a unique opportunity to develop an intensive cancer surveillance program for both melanoma and PC in a high-risk population in which individuals have been stratified according to their mutation status.

Methods: In a research setting, endoscopic ultrasound (EUS) is performed in healthy members of PC-FAMMM families that harbor the p16 mutation, with the goal of either identification of changes in the pancreatic parenchyma, which have been reported to possibly predict the presence of dysplastic changes, or the early detection of a small (< 1–2 cm) PC. If the EUS is abnormal, an ERCP, CT and MRI are done. Also, pancreatic juice is collected in the duodenum by a secretin stimulation, and along with serum, is being stored for investigation with various biological assays.

Results: 7 members: 2F, 5M (median age 32, range 28–43; 5 affected, 2 unaffected) from 3 PC-FAMMM families have participated in this study. A total of 11 EUS have been performed. One affected member has undergone 4 EUS over a year. His first EUS showed parenchymal changes consonant with dysplasia. After having the member abstain from alcohol, a repeat EUS 6 months later had reverted to normal. The other imaging studies (MRI, CT, ERCP) obtained at this time were also normal. On a subsequent EUS six months later, the member had returned to drinking against medical advice and again had concerning parenchymal changes. The Patient again abstained from alcohol and again a repeat EUS obtained six months later had reverted back to normal. Concurrent MRI and ERCP were also normal. The EUS have been normal in all other members. We are counseling 5 additional families for inclusion in this study.

Summary: These results emphasize the importance of obtaining a meticulous alcohol history on patients undergoing surveillance EUS. It appears feasible to develop an intensive pancreatic cancer surveillance program, featuring genetic PC-FAMMM risk although it remains to be proven whether this will lead to improved outcome. Future work will focus on investigating different biological assays, such as altered high molecular weight DNA, telomerase activation, K-ras mutations or proteomics from the collected pancreatic juice.

Tumor suppressor gene loci on chromosome 16 involved in urinary bladder cancer progression

D.S. Yoon, L. Li, J.Y. Ro, D.A. Johnston, B. Grossman and B. Czerniak

The University of Texas, M. D. Anderson Cancer Center, Houston, TX, USA

Background: Putative tumor suppressor gene loci on chromosome 16 were implicated to play a role in the development and progression of several common human malignancies. In this study, we have shown for the first time that distinct loci on chromosome 16 are involved in the development and progression of urinary bladder neoplasia.

Design: Whole organ superimposed histologic and genetic mapping with 30 hypervariable markers mapped to chromosome 16 was performed on 234 DNA samples of 5 cystectomies with transitional cell carcinoma (TCC). Loss of heterozygosity (LOH) was related to microscopically identified precursor conditions and TCC using whole organ histologic and genetic mapping. The significance of alterations in the individual loci for the development or progression of urinary bladder cancer was initially analyzed by the nearest neighbor algorithm and was ultimately tested by LOD score analysis.

Results: Of 11 markers with LOH, 7 showed statistically significant alterations in relation to development or progression of intraurothelial neoplasia. The markers with significant LOD score linking the allelic losses to different phases of urothelial neoplasia clustered in 5 distinct chromosomal regions: p13.2-3(D16S513), p13.1(D16S500), q11.2-12.1(D16S541, D16S415), q22.1(D16S512), q24(D16S505, D16S520). Of these regions, loci p13.1 and q24 were shown to be involved in early phases of urothelial neoplasia. Moreover locus q22.1 which was involved in progression to invasive phenotype contains E-cadherin gene implicated to play a role in invasion and metastasis.

Conclusions: In this study we documented an involvement of several distinct tumor suppressor gene loci on chromosome 16 in the development and progression of urothelial neoplasia from precursor intraurothelial lesions to invasive bladder cancer.

Results of a DNA-based stool assay in people with HNPCC

Steven J. Laken^a, Henry T. Lynch^b, Joe Urbanowski^a, Carolyn Deters^b, Anthony P. Shuber^a and Patrice Watson^b

^a*Exact Sciences, Maynard, MA 01754, USA*

^b*Department of Preventative Medicine, Creighton University, Omaha, NE 68178, USA*

Colorectal cancer (CRC) is the second leading cause of cancer related deaths in men and women. Despite the availability of CRC screening tests, morbidity and mortality from this disease has remained virtually unchanged. As the molecular basis of CRC is becoming well understood, it is possible to create a non-invasive screening test for CRC which assays stool specimen for the somatic mutations that occur in the genesis of this cancer. Two publications have recently presented data using a molecular panel of markers consisting of *p53*, *Apc*, *K-Ras*, and BAT-26 (Gastroenterology 2000;119:1219–1227 and JNCI 2001;93:858–865) as a viable screening modality in patients with average-risk of developing CRC. Despite the fact that many CRCs occur in people at average-risk for developing this disease, at least five percent of all colorectal cancers occur in patients with an inherited condition called hereditary non-polyposis colorectal cancer (HNPCC). People with HNPCC have an 80% lifetime risk of developing cancer. In over 90% of the HNPCC associated colorectal cancers and adenomas that develop, somatic mutations in a marker called BAT-26 can be identified. As our studies in the average risk population showed that the presence of CRC with BAT-26 mutations was associated with detectable mutated DNA in stools, we wanted to evaluate the effectiveness of this approach in people with HNPCC having routine colonoscopy. A total of 47 patients have participated in this study to date. Three patients were excluded because they were determined to be mutation negative after their colonoscopy (all three had normal colonoscopies and negative stool tests). Of the remaining 44 patients, three were found to have CRC at colonoscopy (2 Dukes' A and 1 Dukes' C) and one patient had an advanced adenoma as determined by size (a 2-3 cm tubular adenoma and a 1 cm hyperplastic polyp). The stool test was positive in one of the patients with CRC (Dukes' A) and in the one patient with polyps. The two cancers from patients that were negative by the stool assay were tested for microsatellite instability, and mutations in BAT-26 were present in one of these tumors. All 40 patients with negative colonoscopies were negative by the stool assay. This stool assay has shown the ability to detect early stage disease with 100% positive predictive value, and may have use as an adjunct to colonoscopy screening in people with HNPCC.

High throughput automated microarray hybridization

Catherine E. Campbell, Daisy R. Lee, Angelique Habis, Irina Dimulescu and Elizabeth R. Unger
Centers for Disease Control and Prevention, Atlanta, GA, USA

Introduction: Glass microarrays are increasingly used to monitor the simultaneous expression of thousands of genes in a single hybridization. Application of this technology to samples from population-based epidemiologic studies has the potential to identify markers for the early detection of cancer. However this approach requires high-throughput reproducible hybridizations to allow the results from multiple samples to be compared. It has been recognized that methods of RNA extraction and sample labeling influence the representation and reproducibility gene profiles, however, the impact of hybridization conditions has not been considered. We investigated the use of a commercially available platform for automated hybridization of glass slides (Ventana Discovery, Ventana Medical Systems, Tucson, AZ) to increase the throughput and reproducibility of microarray hybridizations.

Methods: The hybridization platform accommodates 20 slides per run and uses a liquid coverslip to control reagent application and mixing over the slide. Reagent application is controlled by bar coding on the slide and each slide has individual temperature control throughout all steps of the reaction. With the exception of the probe application, all steps are automated and the instrument is completely "walk-away". We tested the automated system using commercially available oligonucleotide microarrays (Clontech 1.0 and 3.1 Microarrays, Clontech, Somewhere, CA). Total RNA extracted from peripheral blood lymphocytes was labeled with Cy-3 using three different methods: Direct incorporation with reverse transcription (5 μ g), *in vitro* transcription (2 μ g), and SMARTTM RT-PCR with random prime labeling (1 μ g). Each labeling was performed in duplication and hybridizations were performed on the instrument and manually. Slides were scanned (Axon Scanner) and single channel analysis performed with GenePix Software (Company, City, State).

Results and conclusion: Automated hybridization of glass microarrays gives consistent results and requires no more probe than the manual method. The automated results are comparable or superior to those obtained manually in terms of both signal and background. Automation minimizes hands-on time and reduces the chance of human error. The Ventana sys-

tem increases the throughput and minimizes the experimental variation between assays by automation and the simultaneous processing of up to 20 samples per run.

Identification of potential markers for early detection of cervical neoplasia with fluorescent differential display PCR

Daya G. Ranamukhaarachchi, Mangalathu S. Rajeevan, Daisy R. Lee, Deon K. Williams, Suzanne D. Vernon and Elizabeth R. Unger
Centers for Disease Control and Prevention, Atlanta, GA, USA

Introduction: Epithelial dysplasias show abnormal differentiation and are recognized as precursors of cancers. Neoplastic progression is associated with further loss of differentiation. Genes differentially regulated during dysplastic differentiation are potential biomarkers for early detection of cancers.

Methods: We screened for genes regulated during dysplastic differentiation of cervical epithelial cells using the 20861 subclone of the HPV-16 positive W12 cells (gift of Dr. P. Lambert, University of Wisconsin). The cells were grown as monolayers (undifferentiated) and as differentiating raft cultures. Differentiation was verified by histologic evaluation. Total RNA was extracted using guanidinium isothiocyanate and evaluated by denaturing gel electrophoresis and spectrophotometry. Fluorescent differential display PCR technology (Beckman Coulter, Inc.) was used to analyze the RNA to search for differential gene expression in this system.

Results: We used 29 primer combinations and excised 273 bands representing differences in expression levels between differentiated and undifferentiated cells. Thirty nine percent of excised bands corresponded to genes expressed at moderate to strong levels with > 4-fold differences in intensity. Ten percent of the bands were unique to RNA from differentiated or undifferentiated cells. We sequenced 263 bands and used the BLAST program to search for sequence similarities in the GenBank database. We identified 90 genes as regulated during dysplastic differentiation (49 with previously known and 41 with no known functions). Most (84%) of these genes were down-regulated with differentiation. The known genes included transcription factors (Zn-finger proteins), DNA replication and repair genes (superoxide dismutase and DNA polymerase Epsilon P2 subunit), cell cycle control genes (GTP binding proteins), calcium binding proteins (calpactin light chain), cytoskeletal (actin binding and sequestering), apoptosis/oncogenesis (tumor promoters) and riboso-

mal proteins (ribosomal protein S10). The majority of the known genes regulated by differentiation fall into the COG's (Clusters of Orthologous Groups, NCBI) functional classification of cellular processes (63%), with information storage and processing (28%), and metabolism (9%) comprising the remainder.

Conclusions: Differential display PCR is an effective tool to identify both known and unknown genes of potential importance as markers of early detection. We are evaluating these genes for their roles in dysplastic differentiation and as potential biomarkers for the onset or progression of cancer.

Analysis of gene expression in prostate cancer

D.A. Troyer^a, R. Leach^b, I.M. Thompson^c AND G.L. Shen^d

University of Texas Health Science Center, Departments of ^aPathology, ^bCellular and Structural Biology, and ^cUrology, San Antonio, TX, USA

^d*Gene Logic Inc., Pharmacogenomics, Gaithersburg, MD, USA*

The Early Detection Research Network (EDRN) is a recently initiated consortium of the National Cancer Institute (web site: <http://edrn.nci.nih.gov/>). This project is an extension of the San Antonio Biomarkers of Early Detection and Research (SABOR), a participating group in the EDRN. Gene Logic and The University of Texas Health Science Center at San Antonio (UTHSCSA) have established a collaborative agreement to determine gene expression in prostate cancers. Microarray analysis is performed on the Human U95 Affymetrix GeneChip[®] high density oligonucleotide microarray platform for simultaneous analysis of 60,000 fragments, with 12,000 fragments covering Full Length Genes and 48,000 fragments covering Unigene ESTs. This technology does not require competition with another cell type, as is required for the cDNA microarray assay, thereby allowing comparison of absolute levels of gene expression of all samples acquired and processed.

Prostate specimens are drawn from tissue samples banked from patients undergoing prostatectomies at University Hospital and the Audie Murphy Veterans Hospital, San Antonio, Texas. The banking of these samples is conducted under an IRB approved protocol commencing in 1994. Prostate specimens are dissected in the fresh state, suspicious areas sampled and immersed in isopentane (−20°). Hematoxylin and eosin stained frozen sections are used to identify areas highly enriched in tumor (> 90%) and similarly to identify

areas from the same prostates that are free of tumor ("normal"). Normal and tumor samples from each prostate are submitted to and are processed by Gene Logic Inc (GLGC). Analysis of gene expression data will be performed jointly by scientists from both institutions. Results from the GeneChip[®] analysis include relative expression results; primary filtering using existing GeneExpress[®]2000 software (Gene Signature[™]), Gene Signature Differential, Fold Change analysis, E-Northern[™]). Using this approach, we have identified 9,142 genes/EST fragments commonly expressed in a set of 22 prostate cancer tumor samples, and 10,195 gene/EST fragments commonly expressed in a set of 13 "normal" prostate samples were identified. Principal component analysis (PCA) was performed on samples from each tissue type to investigate global differences in expression profiles between tumor and normal samples. The normal and tumor-derived samples could be separated by PCA into 2 distinct populations based upon their respective gene expression patterns. Fold change analysis of the prostate cancer and "normal" prostate resulted in the identification of 211 fragments which have 3-fold or more changes in expression level between the two sample sets. We plan to further subdivide the prostate cancer sample sets based on clinical criteria such as invasiveness of tumors and pre- and post-surgery PSA levels to identify potential markers for prostate cancer.

Quantitation of GSTP1 hypermethylation distinguishes between non-neoplastic prostatic tissue and organ confined prostate adenocarcinoma

Carmen Jerónimo^a, Henning Usadel^a, Rui Henrique^b, Jorge Oliveira^c, Carlos Lopes^b, William G. Nelson^d and David Sidransky^{a,d}

^a*Departments of Otolaryngology-Head and Neck Surgery, Head and Neck Cancer Research Division, ^dDepartment of Urology, Johns Hopkins University School of Medicine, 818 Ross, 720 Rutland Avenue, Baltimore, MD 21205, USA*

^b*Unit of Molecular Pathology-Department of Pathology and ^cDepartment of Urology, Instituto Português de Oncologia de Francisco Gentil – Centro Regional do Porto, Portugal*

Methylation of regulatory sequences near the pi-class glutathione S-transferase (GSTP1) gene is the single most common (> 90%) reported epigenetic alteration found in prostate cancer. We compared quantitative GSTP1 methylation to standard histopathologic assessment of prostate tissues.

Tissue samples from 69 patients with early stage prostatic adenocarcinoma, 28 prostatic intraepithelial neoplasia lesions, and 31 patients with benign prostatic hyperplasia were tested for GSTP1 hypermethylation by quantitative fluorogenic real-time methylation specific PCR. To further verify the clinical applicability of this assay we performed a blinded investigation of prospectively collected prostate sextant biopsies of 21 patients with raised serum prostate-specific antigen (PSA) levels (11 with histologically identified adenocarcinoma, and 10 with no morphological evidence of adenocarcinoma).

The median ratios (methylated GSTP1/MYOD1) found in resected hyperplastic prostatic tissue, intraepithelial neoplasia, and adenocarcinoma were 0.0, 1.4, and 250.8, respectively ($P < 0.00001$). The median GSTP1 methylation ratios found in adenocarcinomas and normal prostate tissue in sextant biopsies from the 21 prospective patients with high PSA levels also differed significantly (410.6 and 0.0, respectively; $P = 0.0007$).

Quantitation of GSTP1 hypermethylation may augment standard pathology by accurately discriminating between normal hyperplastic tissue and prostatic carcinoma within a small tissue sample.

Identification of a mononucleotide repeat as a major target for mitochondrial DNA alterations in human tumors

Montserrat Sanchez-Cespedes, Paola Parrella and David Sidransky
Department of Otolaryngology-Head and Neck Surgery, Head and Neck Cancer Research Division, 720 Rutland Av., 818 Ross Research Building, Baltimore, MD 21205-2196, USA
E-mail: dsidrans@jhmi.edu

Mitochondrial DNA (mtDNA) mutations scattered through coding and non-coding regions have been reported in cancer. The mechanisms that generate such mutations and the importance of mtDNA mutations in tumor development are still not clear. Here we present the identification of a specific and highly polymorphic homopolymeric C stretch (D310), located within the displacement (D) loop, as a mutational hotspot in primary tumors. Twenty-one percent of the 293 primary tumors analyzed harbored somatic deletions/insertions at this mononucleotide repeat. Moreover, these alterations were also present in head and neck preneoplastic lesions. We studied the use of D310 mutations as clonal marker for early detection and for evaluation of

metastases in lymph nodes in breast cancer patients. We detected identical changes in matched fine needle aspirates and in metastases positive lymph nodes. We further characterized the D310 variants that appeared in the lung and head and neck tumors. Most of the somatic alterations found in tumors showed deletion/insertions of 1 or 2-bp generating D310 variants identical to previously described constitutive polymorphisms. Sequencing analysis of individual clones from lymphocytes revealed that patients with D310 mutations in the tumors had statistically significant higher levels of D310 heteroplasmy (more than one length variant) in the lymphocyte mtDNA as compared to the patients without D310 mutations in the tumor mtDNA. Based on our observations, we propose a model in which D310 alterations are already present in normal cells and achieve homoplasmy in the tumor through a restriction/amplification event due to random genetic drift and clonal expansion. The high frequency of D310 alterations in primary cancers combined with the evidence these alterations are already present in preneoplastic lesions provide a new molecular tool for cancer detection.

Validation of mitochondrial homoplasmic changes in cancer

Catherine O'Connell^a, Maura Markowitz^a, deLoise Gambrell-Hocker^a and David Sidransky^b

^a*Biotechnology Division of NIST, Gaithersburg, MD 20899-8311, USA*

^b*Johns Hopkins Medical Institutions, Baltimore, MD, USA*

In the past several years, homoplasmic changes in the tumor-specific DNA sequence of the mitochondrial genome have been reported in tumors from a series of cancer patients. These involve mitochondrial DNA (mtDNA) sequence comparisons of tumor with normal tissue from the same patient. In lung cancers, the majority (70%) of homoplasmic sequence changes have been localized to the mitochondrial D-loop sequence, a region critical for mtDNA replication and transcription. To determine the practicality of mitochondrial homoplasmy as an early indicator of lung tumors, reliable, highly sensitive and clinically robust assays are needed. Thus, pilot studies of sensitivity and reproducibility of methods for rapid, full genome mtDNA sequencing are underway at NIST in collaboration with JHMI as part of the biomarker validation work of the EDRN. When methodological analyses are complete, this pilot study will be expanded to sequence sufficient paired tumor

and normal DNA samples obtained from lung cancer patients to determine whether specific regions might detect the majority of homoplasmic changes as an early detection test for lung cancer.

NIST has developed a national Standard Reference Material (SRM) for the detection of sequence changes in the human mitochondrial genome for forensic applications. This reference standard was previously validated using 58 PCR primer pairs for DNA amplification and sequencing. However, the large number of primer pairs may not be optimal for high-throughput clinical assays needed in the EDNRN. As limited tumor tissue will be available for this project and for diagnostic purposes, we have established robust conditions for analysis of 20 ng samples of cell line DNA for PCR amplification with subsets of these primers that encompass the entire mtDNA genome. Two sets of PCR primers have been validated with human cell line HL-60. One set consists of twenty primer pairs that generate twenty 1 kb (721 bp to 1366 bp) segments of the mitochondrial genome. A second set of nine primer pairs generates nine 2 kb (1831 bp to 2532 bp) segments, also covering the entire mtDNA sequence. These two PCR primer sets will be developed for sequence analysis of mtDNA subtypes from human cell lines that model the genetic diversity of US populations. (Supported by NIST-NCI (EDNRN) Interagency Agreement #CN-0103-02.)

Instabilotyping: Comprehensive identification of frameshift mutations caused by coding region microsatellite instability

Yuriko Mori^a, Jing Yin^a, Florin Selaru^a, Asma Rashid^b, Barbara A. Leggett^c, Joanne Young^c, Peter M. Kuehl^b, Patricia Langenberg^d, John M. Abraham^a, Stephen J. Meltzer^a and O. Colin Stine^{b,d}

^aDepartment of Medicine, Division of Gastroenterology and Greenebaum Cancer Center, University of Maryland School of Medicine; Baltimore VA Hospital, 8-009 Bressler Research Building, 655 West Baltimore Street, Baltimore, MD 21201, USA

^bHuman Genetics Program, University of Maryland School of Medicine, 3-119 108 North Green Street, Baltimore, MD 21201, USA

^cConjoint Gastroenterology Lab, Royal Brisbane Hospital Foundation, Clinical Research Centre, Bancroft Centre, 300 Herston Road, Herston, Queensland 4029, Australia

^dDepartment of Epidemiology, University of Maryland School of Medicine, 102A Howard Hall, 660 West Redwood Street, Baltimore, MD 21201, USA

Microsatellite instability (MSI) is a common abnormality among cancers of the stomach, colon, and endometrium that can cause frameshift mutation in genes with microsatellites in their protein encoding regions. Coding region MSI in tumor-related genes is one mechanism contributing to tumorigenesis in cancers with frequent MSI (MSI-high cancers). Mutation of the *TGF β 1 type II receptor (TGFBR2)* gene in MSI-high colorectal cancers is a widely known example of this process. In order to identify additional candidates in this type of process, a large-scale mutational screening of coding region microsatellites was conducted. To discover coding region microsatellites, 21,000 mononucleotide homopolymer registered in an on-line genetic database (Unigene) were examined, and 300 coding region loci with eight or more nucleotides were identified. Mutational screening was performed at 152 of these loci in 46 MSI-high colorectal cancers. Nine loci were mutated in at least 20% more of tumors, 10 loci were mutated in 10–20%, 24 loci in 5–10%, 43 loci in less than 5%, and 66 loci were not mutated in any tumors. The most frequently mutated novel loci were a member of the TGF- β receptor superfamily, the *activin type II receptor* gene (58.1%); an endoplasmic reticulum gene, *SEC63* (48.8%); an interferon-inducible gene, *absent in melanoma 2* (47.6%); a gene encoding a subunit of the NADH-ubiquinone oxidoreductase complex (27.9%); a probable human homologue of the mouse embryonal protein gene *cordon-bleu* (23.8%); and a cell-cycle related gene, *EBP1/PA2G4* (20.9%). Loci previously shown to undergo coding region MSI were also mutated, including *TGFBR2*, *BAX*, and *hMSH3*.

This genome-wide approach identifies coding region MSI in genes or pathways not previously implicated in colorectal tumorigenesis. Genes with frequent coding region MSI may be involved in the process of tumorigenesis or progression, and thus may merit functional study or other further analysis.

Identification of secreted biomarkers from breast carcinoma

L. Molz, K. Grove and D. Beach
Genetica, Inc., One Kendall Square, Building 600,
Cambridge, MA 02139, USA
E-mail: dbeach@genetica.cc

The broad goal of this program is the discovery of secreted and cell surface proteins that are specifically overexpressed in tumor cells of the breast. By definition, secreted and cell surface proteins are accessible to molecular probes, and thus have high prognostic and

diagnostic potential. Our SPaRX screening technology can rapidly isolate expressed signal sequence tags (*secrets*) from any tissue source. We used a secreted reporter, CD8, to trap gene fragments with signal sequence activity from a breast tumor library. Of the proteins identified to date, 53% are extracellular, 28% are tethered to the plasma membrane, and 20% are intracellular. We were pleased to see a high percentage of extracellular and plasma membrane proteins, since these proteins have characteristics of useful biomarkers. We are currently analyzing additional clones from the 'trapped' breast carcinoma library, and we believe that we will identify a several hundred additional secreted sequence tags. These biomarkers are now ready for the next step in the EDRN process: validation.

The signal sequence trap reduces a complex genome to a subset of genes with high potential diagnostic utility. We will identify biomarkers from this subset by determining the profiles of expression of these genes. Both quantitative PCR and *in situ* hybridization will be used. We intend to perform these experiments in collaboration with an EDRN validation lab.

Evaluation of new expression based markers for detection of breast cancer cells

J.R. Marks^a, N.M. Brown^a, T.T. Stenzel^a, L. Roberts^b, J. Henslee^b, and P. Friedman^b

^aDuke University Medical Center, Durham, NC, USA

^bAbbott Laboratories Diagnostics Division, Abbott Park, IL, USA

Genes that are expressed in a highly tissue or disease-specific manner provide possible targets for therapeutics, early detection of cancer, and monitoring of disease burden during and after treatment. Further, genes of this type that code for secreted or shed proteins may allow for serum detection of the product facilitating our ability to specifically detect the cancer in all circumstances. To this end, we are working towards identification and characterization of such genes that are specifically expressed in breast epithelium. Two new genes that are highly restricted in their expression to the breast epithelium emerged from a directed screen of the Incyte LifeSeq Database: one of these is a newly discovered uteroglobin termed BU101, while the other is a novel gene with mucin-like properties termed BS106. The expression of both of these genes is largely limited to normal and neoplastic breast epithelium. In the current study, we have measured the expression of these two markers in comparison with two other candidate markers for detecting breast epithelium, mammaglobin and

cytokeratin 19 (CK). Mammaglobin has been proposed as a breast-specific gene while cytokeratin has been used extensively to detect epithelium in axillary lymph nodes, bone marrow, and in peripheral blood. Using both semi-quantitative end-point PCR and quantitative real-time PCR, we compared the expression of these four genes in a series of primary breast cancers and uninvolved lymph nodes. As anticipated, CK was highly sensitive in detecting all breast cancers however amplification of specific products was also seen from lymph nodes from non-cancer patients. The three breast markers, mammaglobin, BU101, and BS106 were more specific than CK, however each of these markers also failed to detect a small partially overlapping subset of breast cancers. Therefore, while no one of these markers efficiently detects all breast cancers, a combination of two or more may achieve a very high sensitivity in assaying for circulating or occult breast cancer cells. The development of these markers for the detection of tumor cells in the bone marrow and in circulation will be further discussed.

Utilization of stool samples from Hemocult Sensa cards for isolation of amplifiable DNA amenable for mutation detection

Zeev Lev^a, Aaron Lerner^b and Gad Rennert^c

^aDepartment of Biology, ^bDepartment of Pediatrics,

^cDepartment of Community Medicine and Epidemiology and CHS National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Technion, Israel and Institute of Technology, Haifa, Israel

Mutations in several genes can be found in DNA isolated from stool of colorectal cancer patients. These mutations are being used as biomarkers in experimental protocols aimed at the early detection of colorectal adenomas or the follow up of chemo-prevention trials. However, while advanced techniques for obtaining amenable mutation detection show promising advance toward sensitive and specific mutation detection, their utilization in large-scale screening is hampered by severe obstacles in stool sampling, delivery, and storage.

We have found that stool samples from processed Hemocult Sensa cards can be used for isolating amplifiable DNA. These cards are used routinely for the fecal occult blood test (FOBT) in stool of populations at risk as a marker for early detection of colorectal cancer. This finding provides access to a vast number of samples, essentially solving the problem of sampling and delivery. We also solved the problems of long

storage of stool samples, efficient DNA isolation, and reproducible yields of amplified DNA, amenable for convenient mutation detection. Our protocol is simple, robotics compatible, and thus suitable for cost-effective, large-scale mutation screening. Once this protocol was established, we were able to process more than 500 stool samples with only about 10% unsuccessful amplifications. In one experiment described herein in details DNA was extracted from 264 samples and amplified successfully from 238 samples (90% efficiency). K-ras mutations were detected in 20 cases.

Mapping of tumor suppressor gene loci on chromosome 5 in development and progression of urothelial neoplasia

A. Kram, L. Li, J.Y. Ro, D. Johnston, B. Grossman and B. Czerniak
The University of Texas, M.D. Anderson Cancer Center, Houston, TX, USA

Background: Involvement of several putative tumor suppressor gene (TSG) loci mapped to chromosome 5 has been implicated in the development of some common human malignancies, including urinary bladder cancer. In this study we addressed the issue of evolution of allelic losses on chromosome 5 in progression of urothelial neoplasia from occult preneoplastic conditions to invasive transitional cell carcinoma (TCC) of the bladder.

Design: Sequential alterations on chromosome 5 were studied by whole organ histologic and genetic mapping in cystectomies with invasive TCC and microscopically recognizable preneoplastic changes in adjacent urothelium. 39 hypervariable markers mapped to chromosome 5 were studied on 234 urothelial samples of 5 cystectomy specimens. The significance of alterations in individual loci was tested by nearest neighbor and LOD score analyses. Markers showing allelic losses (LOH) with statistically significant relationship to progression of urothelial neoplasia were subsequently tested on 32 bladder tumors and voided urine samples of 27 patients with TCC.

Results: Markers with statistically significant LOH in relation to progression of urothelial neoplasia clustered in 5 distinct loci of chromosome 5: q14-22 (48.4cM; D5S424-D5S656); q23.1-23.3 (18.5cM; D5S656-D5S808), q31.1 (0.3cM; D5S808-D5S816), q31.3 (25.7cM; D5S816-SPARC) and q35.2 (1.4cM, IG22-D5S1456). Loci q14-22, q23.1-23.3, q31.1 and q31.3 were involved in early phases of urothelial neoplasia. On the other hand q35.2 region showed LOH

associated with progression to an invasive phenotype. Testing of markers with statistically significant LOH on larger number of tumors and voided urine samples of patients with TCC disclosed that minimally deleted regions in q14-22 and q23.1-23.3 showed LOH in nearly 20% of TCC while the remaining loci were involved in less than 10% of samples.

Conclusion: This study documented that loci 5q14-22 and 5q21.1-23.3 containing known TSG (MCC, APC) may play a role in the development of urothelial neoplasia.

Characterization of genes expressed in pancreatic adenocarcinomas

Randy S. Haun and Sarah K. Johnson
Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Pancreatic cancer is the fourth and fifth most common cause of cancer-related deaths in men and women, respectively, in the United States. Detection is normally observed only in late stages of the disease; thus, prognosis of patients upon diagnosis is extremely poor. As the only curative intervention is early surgery, with delayed detection the one-year survival rate of patients is less than 10% after the disease has been diagnosed and the 5-year survival rate is negligible. Currently, there is a lack of any specific or sensitive diagnostic test for early stages of pancreatic cancer. Although a wide variety of tumor-associated antigens have been evaluated as markers for screening and diagnosing pancreatic cancer, most have proven ineffective due to their low sensitivity and cross reactivity with other tumors. The characterization of genes differentially expressed in pancreatic adenocarcinomas may provide tumor markers that are both selective and sensitive for the early detection of pancreatic tumors. The long-term goal of this project is to identify genes that are expressed selectively in pancreatic adenocarcinomas compared with normal or other diseased pancreas (e.g., pancreatitis). These transcripts may serve as targets for the development of specific diagnostic tests or new treatment modalities for pancreatic cancer. To achieve this goal, we have utilized two PCR-based techniques to identify genes differentially expressed in pancreatic adenocarcinomas.

In the first approach, we employed differential display, a PCR-based method of differential expression cloning. Using this method, we have characterized two PCR products that were consistently overexpressed in pancreatic tumors relative to normal pancreatic tissues.

Sequence analysis of these cDNAs revealed that one encoded a portion of the recently cloned GABA_A receptor π subunit and the other encoded the anion transporter down-regulated in adenoma (*DRA*), a gene expressed primarily in colon and small intestine.

Pancreatic cancer is a highly aggressive disease characterized by local invasion of adjacent structures, perineural invasion, early metastases to lymph nodes and liver, and an intense desmoplastic stromal reaction. Local proteolysis facilitates tumor growth and metastasis. Proteases have been associated with invasion and metastasis of many cancers due to their ability to degrade extracellular matrix proteins and to activate other proteases. To investigate the proteases that may be selectively expressed in pancreatic cancer, a PCR screen was performed using degenerate oligonucleotides targeted to the conserved catalytic domain of various families of proteases. Proteases were identified and an expression profile generated by sequencing and cataloging clones prepared from both normal and tumor pancreatic tissues. Using this approach, two proteases were identified that displayed tumor-specific expression: stratum corneum chymotryptic enzyme (SCCE), a serine protease previously reported to be skin-specific, and meprin β , an astacin-family metalloproteinase that are normally only expressed at high levels in mammalian renal and intestinal brush-border membranes.

RT-PCR analysis of a panel of RNAs prepared from pancreatic ducts isolated from normal and chronic pancreas and pancreatic tumors confirmed that these transcripts were significantly overexpressed in pancreatic carcinomas. The results of these studies have demonstrated the validity of screening normal and carcinoma tissues for differentially expressed genes in an attempt to identify potential tumor marker genes. These findings combined with other techniques for identifying genes differentially expressed in pancreatic tumors (e.g., DNA arrays) may provide valuable tools for the early detection of this devastating disease.

Alterations in global DNA methylation in buccal mucosal cells reflect methylation status in cancerous tissues of the lung

Chandrika J. Piyathilake, Robert J. Cerfolio, Martin A. Whiteside, Sreelatha Meleth, Gary L. Johannang, Douglas C. Heimbürger and William E. Grizzle
University of Alabama at Birmingham, Birmingham, AL, USA

There is increasing interest in investigating the status of global DNA methylation in peripheral leuko-

cytes in relation to risk of developing cancers at other sites. Whether global DNA methylation in peripheral leukocytes or any other easily accessible tissue reflects methylation changes in tissues at risk of developing cancer is unknown. To address this issue, we investigated the status of global DNA methylation in matched samples of peripheral leukocytes, buccal mucosal cells, cancerous and non-cancerous tissues of 14 subjects who had developed lung cancer (nine primary non-small cell and five metastatic lung cancer) by using an *in-vitro* radio-labeled methyl incorporation assay. The radio-labeled methyl incorporation (inversely related to the degree of *in-vivo* DNA methylation) was significantly higher in DNA from cancerous tissues and buccal mucosal cells compared to non-cancerous lung DNA and leukocyte DNA. There was no significant difference in radio-labeled methyl incorporation between cancerous tissues and buccal mucosal cells. In primary non-small cell lung cancer subjects, but not in metastatic lung cancer subjects, the status of global DNA methylation in buccal mucosal cells was positively associated with methylation status in cancerous tissues ($r = 0.7$, $p = 0.06$, 95% CI = 0.19–1.0). The association between methylation status in leukocyte DNA and cancerous cells was non-significant in both primary and metastatic lung cancer subjects. These results suggested that the changes in global methylation in buccal mucosal cells reflect changes in tissues at risk of developing non-small cell lung cancer, and may serve as a useful marker for risk of developing the disease.

Aberrant promoter methylation profile of cells in voided urine as a potential method for the early detection of bladder cancer

Riichiroh Maruyama, Bogdan Czerniak and Adi F. Gazdar

University of Texas Southwestern Medical Center, Dallas, TX and University of Texas, M. D. Anderson Cancer Center, Houston, TX, USA

We investigated the aberrant promoter methylation profile of bladder cancers and correlated the data with clinico-pathological findings. The methylation status of ten genes was determined in 98 surgically resected bladder cancers and we calculated the methylation index (MI), a reflection of the methylated fraction of the genes tested. Methylation frequencies of the genes tested in bladder cancers were 36% for *CDH1*, 35% for *RASSF1A* and *APC*, 29% for *CDH13*, 16% for *FHIT*, 15% for *RAR α* , 11% for *GSTP1*, 7% for

p16^{INK4A}, 4% for *DAPK* and 2% for *MGMT*. Methylation of four of the individual genes (*CDH1*, *RASSF1A*, *APC*, and *CDH13*) and the MI were significantly correlated with several parameters of poor prognosis (tumor grade, growth pattern, muscle invasion and aneuploidy). Methylation of *CDH1*, *FHIT* and a high MI were associated with shortened survival. *CDH1* methylation positive status was independently associated with poor survival in multivariate analyses.

We examined the voided urine of 54 patients with untreated or previously treated bladder cancer. The methylation profile of the urine sediments were very similar or identical to the tumor specimens. Our results suggest that the methylation profile may be a potential new biomarker of risk prediction, and that examination of voided urine may aid early detection of bladder cancer.

Chromosomal instability is commonly observed in bronchial cells from smokers

Marileila Varella-Garcia, Lin Chen, Roger Powell and Wilbur Franklin

Lung Cancer Program, University of Colorado Health Sciences Center, Denver, CO, USA

Chromosomal abnormalities critical in the initiation and progression of lung cancer are incompletely defined. The recent development of multicolor karyotyping techniques has significantly enhanced the ability to detect and comprehensively identify chromosomal abnormalities that are not accessible through conventional cytogenetic methods. We applied the spectral karyotyping (SKY) technique to primary cultures of 59 bronchial specimens biopsied from 42 individuals (28 males and 14 females) at high risk for lung cancer (smoking > 30 pack/year, obstructive airway disease, and abnormal sputum cytology). Cultures were performed in coverlips for 7–20 days and karyotype analysis was carried out in approximately 17 cells per specimen. An elevated fraction of abnormal cells was found in 16 specimens from 15 individuals, including a number of sites with normal results by fluorescence bronchoscopy and pathology analyses. With only one exception, the abnormalities were present in a minority of mitotic cells. The major abnormalities included reciprocal translocations, derivative chromosomes, isochromosomes, pericentric inversions, deletions, and trisomies. Chromosomal losses, polyploidy and chromosomal breaks, which are more likely to be related to technical variables, were not included in the significant anomalies. Chromosomal abnormalities

were found on average in 8.8% of cells in the high-risk individuals and in 2.5% of cells in lung epithelia from non-smokers used as controls. Cells displaying chromosomal anomalies were more frequent in males (9.9%) than females (6.4%).

SKY results in metaphase cells were validated in interphase FISH assays with selected probes. In 10 specimens displaying extranumerary copies of either apparently normal or rearranged chromosomes, we confirmed the presence of an equivalent population of interphase nuclei using centromere or locus-specific probes. Conversely, chromosomal losses identified in metaphases were not confirmed in interphase assays, and polyploidy was generally found in 2–10% of nuclei in every assayed specimen. Trisomies 7, 8, and 18 were the most frequent numerical abnormality, and breakpoints at 1q11-13, 3p14-21, 5q12, and 9q12 were prevalent among the structural abnormalities. The majority of translocations was balanced and involved material originating from two chromosomes. However, balanced translocations and derivative chromosomes from 3-way translocations were also observed. Recurrent translocations were not detected, either among the individuals or among the specimens from distinct sites in a given individual. Homogeneously staining regions were absent. Double minutes originated from chromosome 3 were observed in a single cell.

We are currently analyzing the breakpoints involved in rearrangements in these short-term, primary cultures to identify relevant sequences potentially related with biologically significant mutations. These sequences will be used in a multicolor, multiprobe probe configuration in FISH assays in interphase cells of pre-invasive epithelial lesions and sputum, in a search for subtle rearrangements that may be important in early stages of lung carcinogenesis.

Genome-wide map of putative tumor suppressor gene loci involved in bladder cancer progression

B. Czerniak, L. Li, R.D. Zhang, D.S. Yoon, A. Kram, J.Z. Li, J.Y. Ro, V. Chaturvedi and D.A. Johnston
The University of Texas, M. D. Anderson Cancer Center, Houston, TX, USA

Background: A prerequisite to understanding the mechanisms of carcinogenesis is the determination of which chromosomal regions are structurally altered in the process of cancer progression on the whole genome scale.

Design: Whole organ histologic and genetic mapping with 771 hypervariable markers mapped to chro-

mosome 1–22 (chromosomes X and Y were not included in this analysis) was performed on 234 DNA samples of 5 cystectomies with transitional cell carcinoma (TCC). Loss of heterozygosity (LOH) was related to microscopically identified precursor conditions and TCC. The patterns of chromosomal alterations in the progression of neoplasia were used to assemble a genome-wide model of cancer progression.

Results: Of 200 markers with LOH, 131 showed statistically significant alterations in relation to development or progression of intraurothelial neoplasia. The markers with significant LOD score linking the allelic losses to different phases of urothelial neoplasia clustered in 87 distinct chromosomal regions, identifying these regions as positions of putative tumor suppressor genes. Some of the markers with statistically significant allelic losses mapped to the regions containing well-characterized tumor suppressor genes but many were located in previously unknown loci. Overall, 43(32.8%) of the markers exhibited statistically significant LOH in association with the development of precursor or intraurothelial conditions, and 26(19.8%) of the alterations could be related to the development of the invasive phenotype.

Conclusions: Markers exhibiting allelic losses in early phases of urothelial neoplasia could be used as powerful tools to monitor the preclinical and even pre-microscopic phases of urothelial neoplasia in histologic samples and voided urine sediments. The approach used in this study provides important chromosomal landmarks for more specific studies on molecular mechanisms of multistep urinary bladder carcinogenesis.

Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH

Jose M. Lage^a, John Leamon^a, Tanya Pejovic^a, Stefan Hamann^a, Deborah Dillon^a, Bettina Vossbrinck^a, Antonio Gonzalez^b, Jose Costa^a and Paul M. Lizardi^a

^a*Department of Pathology, Yale University School of Medicine, 310 Cedar St., New Haven, CT, USA*

^b*Instituto de Parasitologia y Biomedicina, Granada, Spain*

Structural genetic alterations often involve gene loss or gene amplification. With the advent of microarray approaches, scanning for gene dosage alterations is limited only by issues of microarray density. However, samples of clinical interest often comprise small clusters of just a few hundred cells, which do not pro-

vide sufficient DNA for array comparative genomic hybridization (a-CGH) analysis. We describe a method that permits amplification of genomic DNA with very limited sequence representation bias. The method, based on random priming and hyperbranched strand displacement, generates thousands of copies of the genome in a few hours. Reactions catalyzed by Phi29 give high DNA replication fidelity, while Bst DNA polymerase produces demonstrably superior sequence representation. Using whole genome isothermal amplification (WeGi), in combination with a-CGH, we demonstrate the capability for detection of gene gains and losses in yeast and in human tumor cell lines, which high precision and high resolution. This capability for quantitative whole genome analysis using DNA samples from a few hundred cells opens exciting new avenues for studies in cancer genetics.

Serum protein profiling by SELDI mass spectrometry coupled with a learning algorithm distinguishes prostate cancer from non-cancer

Bao-Ling Adam, Michel D. Ward, MaryAnn Clements, Lisa H. Cazares, John W. Davis, Paul F. Schellhammer, O. John Semmes and George L. Wright, Jr., Enrique Dalmaso, Christine Yip, Yinsheug Qu, Yutaka Yasui and Ziding Feng

Departments of Microbiology and Molecular Cell Biology, and Urology, Virginia Prostate Center, Eastern Virginia Medical School, Norfolk, VA, USA; Ciphergen Biosystems, Inc., Fremont, CA, USA; and Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Prostate specific antigen (PSA) is regarded as one of the best molecular diagnostic cancer biomarkers presently available. However, the PSA serum test lacks specificity, especially in discriminating benign prostate hyperplasia (BPH) from prostate cancer (PSA) in the so-called "grey-zone" PSA range of 4–10 ng/ml. The poor specificity (i.e., 25–30%) in this range has resulted in the performance of unnecessary prostate biopsies. Attempts to improve the PSA test have led to a number of modifications including the calculation of free- to total or complexed PSA ratios, PSA density, and PSA velocity, with only minor improvement. The weaknesses of the PSA test, and the robust molecular and cellular heterogeneity of PCA, have led us to believe a combination or panel of biomarkers will be required to improve the early detection/diagnosis of PCA. Therefore, we are exploring the use of the ProteinChip[®] SELDI mass spectrometer to profile the proteins in sera in an effort to identify prostate cancer-associated "finger-

prints" that would more accurately discriminate PCA from non-cancer groups, i.e. BPH and normal. For this study, 500 serum samples consisting of 5 groups, were obtained from the Virginia Prostate Center Tissue and Body Fluid Bank: (1) Normal young ($N = 98$); (2) Normal age-matched ($N = 96$); (3) Cancer early-stage ($N = 99$); (4) Cancer late stage ($N = 98$); and (5) BPH ($N = 93$). The serum samples were applied to an IMAC3-Cu binding protein chip array and subjected to SELDI analysis. The pre-processed time-of-flight (TOF) data was then analyzed by the Wavelet learning algorithm. Phase I analyses was performed comparing the age-matched normal group with the cancer groups (both early and late stage). 167 PCA and 81 normal spectra were used to train the algorithm, and 30 PCA and 15 normal were used as the test set. A sensitivity/specificity of 99.42%/98.72% and 96.67%/100% was obtained for the training and test sets, respectively. A Phase II analyses was then performed comparing the normal, BPH, and cancer groups. This resulted in 97.53% (79/81) of the normal, 91.03% (71/78) of BPH, and 88.02% (147/167) of the PCA being identified correctly in the training set. For the test set, 86.67% (13/15) of normal, 80.00% (12/15) of BPH, and 73.33% (22/30) of PCA were correctly predicted. Further training of the Wavelet learning algorithm is in progress in an effort to achieve higher prediction rates, especially for the PCA group. These initial results suggest that SELDI protein profiling has potential for the development of a novel, rapid and high throughput early detection assay for prostate cancer. (Supported by grants from the National Cancer Institute (CA85067- Early Detection Research Network) and Virginia Prostate Center).

Proteome analysis of patient serum for the early detection of hepatocellular carcinoma

Laura F. Steel, Taj S. Mattu and Timothy Block
Jefferson Center for Biomedical Research, Thomas Jefferson University, Doylestown, PA, USA

Chronic infection with hepatitis B or C virus (HBV or HCV) is a major risk factor in the development of hepatocellular carcinoma (HCC). We are using two-dimensional gel electrophoresis (2DE) to compare serum proteins of healthy patients, those at high risk (i.e. with chronic hepatitis infection), and those diagnosed with HCC to identify changes that correlate with disease progression and may serve as early detection markers for HCC. Serum samples collected from 8–10 patients (matched for age, sex, and ethnicity) in each

of our diagnostic groups have been combined to generate composite gels for each group. A comparison of these gels, aided by computerized spot detection and gel matching, has identified a number of protein spots that change in relative intensity by more than a factor of two between the healthy and HCC groups. These spots are now being evaluated for the degree to which they vary among the individuals that comprise the composite groups. We are asking whether the variation seen between the healthy and HCC composite gels is supported by a consistently higher or lower level of each of the proteins among the individuals that contributed to the composite. Spots that are showing promising differences are being identified by mass spectrometry. We are also examining composite and individual gels generated from the chronically infected patient groups. Appearance of a potential HCC marker in one of the chronically infected groups could indicate a true early detection marker for HCC in these high risk groups, or could indicate that the marker only reflects liver inflammation or some general aspect of liver disease. Further evaluation of this issue will be carried out using samples taken serially from a single patient over the many years it takes chronic HBV infection to progress, or not, to HCC. We are also continuing to explore methods to improve our detection of lower abundance proteins, including removal of abundant proteins from the serum and removal of post-translational modifications (e.g. N-linked glycosylation) so as to consolidate numerous spots into a single feature on the gel. Databases of clinical and experimental sample information, that form the basis for our pseudo-LIMS, are being integrated into our web-based interface that is being compiled with 2-DE maps of normal human serum and the de-N-glycosylated serum with hyperlinks to genomic (Genbank) and carbohydrate (Carbank, Glycominds etc.) databases.

Proteomic analysis of human lung adenocarcinomas

Guoan Chen^a, Tarek G. Gharib^a, Michael S. Prescott^a, Chiang-Ching Huang^b, Kerby A. Shedden^b, Jeremy M.G. Taylor^b, Dafydd G. Thomas^c, Thomas J. Giordano^c, Rork D. Kuick^d, Melissa C. Krause^d, Christopher Wood^d, Mark D. Iannettoni^e, Mark B. Orringer^a, Samir Hanash^d and David G. Beer^a
Section of General Thoracic Surgery, Departments of ^aSurgery, ^bBiostatistics, ^cPathology, ^dPediatrics and ^eEpidemiology, University of Michigan, Ann Arbor, MI 48109, USA

Lung cancer is the leading cause of cancer death for both men and women in the United States, with over

170,000 new cases diagnosed each year. The 5-year overall survival rate remains only 10–15%, and has not significantly improved over the last 20 years. Non-small cell lung cancer (NSCLC) accounts for almost 80% of lung cancers and adenocarcinomas comprise approximately 40% of all new cases of NSCLC. Although patients diagnosed with stage I adenocarcinoma have an overall 5-year survival rate of 63%, nearly 35% will relapse following surgical resection and exhibit a poor prognosis. Identification of these high-risk patients with resectable early stage disease and providing further adjuvant therapy may help increase survival. The goals of this study were to identify proteins that may be useful for early detection of lung cancer as well as proteins associated with patient prognosis. Analysis of the same samples for mRNA expression was performed using Affymetrix oligonucleotide arrays to assess potential transcriptional mechanisms for changes in protein expression. A series of 90 lung adenocarcinomas (62 stage I and 28 stage III) and ten uninvolved lung samples were examined for protein expression using 2-D polyacrylamide gel electrophoresis (2-D PAGE) and candidate proteins analyzed using matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). A total of 820 individual protein spots were quantified for each sample. Kaplan-Meier analysis and F-test statistic methods were used to determine associations between different clinical-pathologic variables and patient survival. Three hundred proteins were found to be significantly different between normal lung and adenocarcinomas and 51 differed significantly between stage I and III tumors. Fifty protein spots were found to be significantly ($p < 0.05$) correlated with patient survival using Cox proportional hazards regression analysis. Expression profiles for the 20 proteins most significantly correlated with survival were used to create a risk index using a leave-one-out, cross-validation method. A separate high and low risk group were identified among patients with stage I lung adenocarcinoma which differed significantly ($p = 0.0017$) in survival. Eleven of the 50 proteins significantly associated with survival ($p = < 0.05$) and 94 associated with normal-tumor or tumor stage differences were identified by MALDI-MS. Candidate proteins were confirmed as expressed in lung tumor samples using immunohistochemistry on tissue microarrays made from the same tumors from this study. Correlations within the same samples between the protein and mRNA expression were made for proteins found to be overexpressed in lung adenocarcinomas as well as those found to be associated with survival. A significant correlation

of the protein with their respective mRNA was found for a subset of proteins, indicating potential transcriptional regulation. The data from these studies suggest that 2D protein expression analyses can identify proteins that show quantitative differences between normal lung and tumors as well as identify proteins associated with survival. Use of the expression profile of proteins associated with patient survival may help identify a high-risk subgroup among stage I lung adenocarcinomas as well as identify potential candidates useful for early detection.

Profiling and imaging of proteins in tissue sections using mass spectrometry as a discovery tool in cancer research

Richard M. Caprioli and Pierre Chaurand
Vanderbilt University, School of Medicine, Nashville, Tennessee, TN, USA

Imaging Mass Spectrometry is a relatively new technology that takes advantage of the methodology and instrumentation of matrix-assisted laser desorption ionization (MALDI) mass spectrometry. It can be used to locate specific molecules such as peptides and proteins up to about 80,000 Daltons directly from fresh frozen tissue sections or blots of a tissue. The data from such an analysis is a pictograph, in real x, y dimensions, of the location of a signal at any given molecular weight. Using a raster of mass spectra over a given area of a section, images of samples are produced in specific mass-to-charge (m/z) values, or ranges of values. Each spot on the sample irradiated by the laser is approximately 25 microns in diameter and the resulting spectrum contains hundreds of protein signals. Individual m/z values can then be assembled from these mass spectra to produce selected m/z images. Tissue slices from mouse prostate, colon and both human and mouse brain have been imaged to locate and map tissue specific peptides and proteins. Tumors present in these tissues have been profiled and imaged, comparing new proteins highly expressed in the tumor but not the corresponding normal tissue, and those expressed at low levels in the tumor that are highly expressed in normal tissue. Imaging Mass Spectrometry is an effective discovery tool for the analysis of proteins and peptides in these tissues. It is extremely useful for the comparison of molecular weight based protein patterns in tumors versus normal tissues and in helping identify and image potential tumor markers in precursor lesions and various stages of tumor progression.

A novel highly specific and sensitive urine-based assay for the detection of bladder cancer

Thu-Suong T. Nguyen^a, Tracy Davido^a, Badrinath R. Konety^b and Robert H. Getzenberg^a

^a*Departments of Urology, Pathology and Pharmacology and the University of Pittsburgh Cancer Institute, University of Pittsburgh, Pittsburgh, PA, USA*

^b*Department of Urology, University of Iowa, Iowa City, IA, USA*

There is a need to develop novel markers for bladder cancer that can be utilized to replace cytology to identify bladder cancer with high specificity and sensitivity. We have identified proteins which can differentiate human bladder tumors from normal bladder, and which are not found in other types of cancers. These proteins represent alterations to the cancer cell that are hallmarks of the neoplastic process. Immunoblot analysis of BLCA-4, the first of these proteins to be characterized, reveals that it can differentiate individuals with bladder cancer from disease free subjects. Furthermore, BLCA-4 is expressed throughout the bladder, even in normal appearing areas, in individuals with bladder cancer. Utilizing an ELISA, we have been successful in detecting BLCA-4 in the urine and determining that the levels are significantly higher in patients with bladder cancer compared to healthy subjects ($p = 2.4 \times 10^{-6}$). All normal individuals ($n = 51$) had urinary BLCA-4 levels *below* the prospectively utilized cutoff of 13 O.D. units per μg of protein with a mean value of 4.02 ± 4.21 , whereas 52 of the 54 individuals with bladder cancer had urinary BLCA-4 levels *above* this cutoff with an average value of 43.36 ± 49.52 . Analysis of these results reveals a test specificity of 100% and a sensitivity of 96.4%. Investigations in animal models of bladder cancer indicate that the expression of this protein appears well before the observance of grossly visible tumors in the bladder suggesting that the expression of this protein may be useful as an early detection biomarker of bladder cancer. Analysis of the cDNA sequence of BLCA-4 reveals a high homology with the ETS domain of the ELK3 oncogene for a portion of the protein. This homology suggests that BLCA-4 may serve as a regulator of gene expression in bladder cancer and we are now exploring the functional role of BLCA-4 in bladder cancer. BLCA-4 appears to be the first bladder cancer specific marker to discriminate patients with bladder cancer from those without the disease and which may play an important role in the regulation of bladder gene expression. The assay that we have developed can detect bladder cancer with

high specificity and sensitivity and may permit early detection and improved prognosis of subjects at risk. A large national trial is currently underway to provide further validation of this marker.

Supported by NIH Grant R01 CA82522.

Osteopontin is a potential plasma marker for epithelial ovarian cancer

Jae-Hoon Kim^{a,b}, Toshimitsu Uede^c, Kwong-kwok Wong^d, Gary K. Yiu^a, John O. Schorge^e, Karen H. Lu^f, Ross S. Berkowitz^{a,g}, Daniel W. Cramer^{a,g} and Samuel C. Mok^{a,g}

^a*Department of Obstetrics, Gynecology and Reproductive Biology, Division of Gynecologic Oncology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*

^b*Department of Obstetrics and Gynecology, Saint Vincent Hospital, The Catholic University of Korea, Suwon, Korea*

^c*Section of Immunopathogenesis, Institute of Immunological Science, Hokkaido University, Kita-ku, Sapporo, Hokkaido, USA*

^d*Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA*

^e*Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX, USA*

^f*Department of Gynecologic Oncology, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA*

^g*Dana-Farber Harvard Cancer Center, Boston, MA, USA*

At the Boston EDRN site, we have used a variety of techniques to identify new markers for ovarian cancer. Using the MICROMAX cDNA microarray system and RNA isolated from ovarian cancer cell lines and normal ovarian surface epithelial cells (HOSE), we identified a gene called osteopontin that exhibited an ovarian cancer-to-HOSE ratio of 184. Real time quantitative PCR analysis revealed significant overexpression of osteopontin mRNA in 14 ovarian cancer cell lines and 27 microdissected epithelial cancers compared to that in five cultured normal HOSE lines and microdissected germinal epithelium from two normal ovaries. Western blotting also revealed overexpression of osteopontin protein in ovarian cancer tissues compared to normal ovarian tissue and benign ovarian tissue. Immunolocalization of osteopontin in paraffin block sections (3 normal ovaries, 6 benign ovarian tumors, 29 borderline ovarian cancers, and 61 invasive ovarian

cancers) showed that immunoreactivity of osteopontin was overexpressed in borderline and invasive cancer cases than benign tumors and normal germinal epithelia cases ($p < 0.01$). Immuno-histochemical staining suggested that overexpression of osteopontin was especially apparent in mucinous subtypes compared to other types. To evaluate the potential of osteopontin as a tumor marker, we examined the amount of osteopontin in subjects' plasma by enzyme-linked immunosorbent assay (ELISA). Osteopontin levels were significantly higher ($p < 0.05$) in 59 patients with epithelial ovarian cancer (77.8 ng/ml) compared to 45 normal controls (25.7 ng/ml), 28 patients with benign ovarian diseases (36.7 ng/ml), and 27 patients with other gynecologic cancers (48.5 ng/ml). When used a single cut off value of 34.9 ng/ml was used, osteopontin showed a sensitivity of 84.7% and a specificity of 84.4%. Although osteopontin is known to be associated with carcinoma of various origins, its association with ovarian carcinoma has not been previously reported. Because of its RGD (arginine-glycine-aspartic acid) tripeptide and adhesive properties, it has been proposed that osteopontin plays a role in cancer progression by the interaction with integrin receptors and CD44 as a ligand. This study suggests that osteopontin may be a tumor marker with clinical usefulness.

Ep-CAM auto-antibody is a potential serum marker for epithelial ovarian cancer

Jae-Hoon Kim^{a,b}, Dorothee Herlyn^c, Kwong-kwok Wong^d, Gary K. Yiu^a, John O. Schorge^e, Karen H. Lu^f, Ross S. Berkowitz^{a,g}, Daniel W. Cramer^{a,g} and Samuel C. Mok^{a,g}

^a*Department of Obstetrics, Gynecology and Reproductive Biology, Division of Gynecologic Oncology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*

^b*Department of Obstetrics and Gynecology, Saint Vincent Hospital, The Catholic University of Korea, Suwon, Korea*

^c*The Wistar Institute, Philadelphia, PA, USA*

^d*Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA*

^e*Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, TX, USA*

^f*Department of Gynecologic Oncology, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA*

^g*Dana-Farber Harvard Cancer Center, Boston, MA, USA*

At the Boston EDRN site, we have used a variety of techniques to identify new markers for ovarian cancer. Using the MICROMAX cDNA microarray system and RNA isolated from ovarian cancer cell lines and normal ovarian surface epithelial cells (HOSE), we identified a gene called the epithelial cell adhesion molecule (Ep-CAM) that exhibited a cancer-to-HOSE ratio of 444. Real time quantitative PCR analysis revealed significant overexpression of Ep-CAM mRNA in cancer cell lines ($P < 0.001$) and microdissected cancer tissues ($p = 0.035$), compared to that in cultured normal HOSE and microdissected germinal epithelium, respectively. Immuno-histochemical staining of paraffin block sections revealed that Ep-CAM expression was absent in stromal areas of normal ovaries or those with benign disease or cancer. In contrast, a gradient of expression was found in the germinal epithelium with ovaries from women with borderline or invasive cancer displaying the greatest level of expression, normal ovaries the least, and ovaries from women with benign tumors intermediate expression ($p < 0.05$). No significant differences in Ep-CAM immuno-histochemical staining were observed among ovarian cancer samples with different histologic types and grades. Because Ep-CAM auto-antibody levels have been shown to be elevated in other cancers, such as colon, we examined levels of auto-antibody against Ep-CAM in patients with epithelial ovarian cancer and controls by enzyme-linked immunosorbent assay (ELISA). Ep-CAM auto-antibody levels (measured in units of absorbance at 450 nm) were: 0.132 in 52 patients with ovarian cancer, 0.098 in 26 cases with benign gynecologic disease, and 0.090 in 26 normal women ($p < 0.05$). When a cut-off value of 0.115 was used, the Ep-CAM auto-antibody assay showed a sensitivity of 71.2% and a specificity of 80.8% whereas the sensitivity and specificity of CA 125 measured in 52% of the same subjects were 84.6% and 88.5% with a CA 125 cut-off of 35 U/ml. However, the Ep-CAM auto-antibody assay may be complementary to CA125, as indicated by the fact that combining the test with CA 125 increased the sensitivity to 94.2% and specificity to 100.0%. This investigation has demonstrated the potential value of cDNA microarray analysis in identifying overexpressed genes in ovarian cancer, and suggests that the Ep-CAM auto-antibody may offer a biomarker for ovarian cancer with clinical usefulness.

A novel ovarian cancer biomarker: Haptoglobin α chain, identification and characterization with mass spectroscopy and liquid chromatography

Bin Ye^a, Daniel W. Cramer^{a,b}, Vanessa Pratomo^a,

Steven Skates^c, Sau-Mei Leung^d, Ross S. Berkowitz^{a,b} and Samuel C. Mok^a

^a*Department of Obstetrics, Gynecology, and Reproductive Biology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*

^b*Gillette Center for Women's Health, Dana-Farber Cancer Center, Boston, MA, USA*

^c*Department of Biostatistics, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA*

^d*Ciphergen Biosystems, Inc., Palo Alto, California, CA, USA*

An objective of the Boston EDRN site is to identify potential biomarkers for the early detection of ovarian cancer in serum employing Surface Enhanced Laser Desorption/Ionization (SELDI)-Mass Spectroscopy (MS). Using four different types of surface specific ProteinChip arrays, we screened for protein markers of molecular weight less than 50 kDa in a total of 108 age-matched serum samples (58 cases and 50 normal controls). Comparing mass spectra profiles generated from these samples, several protein peaks were identified as potential markers for different histologic subtypes of ovarian cancer. One protein peak at about 11,700 Da was identified from the profiles using the copper-surfaced (IMAC3) chip. This protein commonly appeared in sera from patients with all histologic types of ovarian cancer cases, but less so in sera from controls. At a concentration intensity of greater than 0.2 from the SELDI output, the sensitivity and specificity of this biomarker for ovarian cancer was 83% and 70%, respectively. The 11,700 DA protein was then further purified by affinity chromatography and sequenced by Ion Trap Tandem Mass Spectrometry. It was identified as the α chain of Haptoglobin. Haptoglobin is composed of α and β subunits and is generally secreted by the liver. It is involved in binding free hemoglobin and preventing loss of iron from the body, but it is also recognized as an acute-phase reactant protein involved in immune regulation. The elevated level of Haptoglobin α chain in sera from ovarian cancer patients, suggested by SELDI, was confirmed by western blotting with an anti-human Haptoglobin 1-1 polyclonal antibody, which can cross react with both α and β chains. Of interest, complete Haptoglobin has previously been found to be elevated in ovarian cancer patients. Ours is the first report to suggest that the α chain may be a better marker. More detailed studies are necessary, including use of a more specific and quantitative assay to validate Haptoglobin α chain as a marker for ovarian cancer detection. In addition, the mechanism(s) which might cause the elevated serum

Haptoglobin α subunit levels with ovarian cancer also need to be understood.

Identification of specific nuclear matrix protein alterations in human colon cancer

Gisela Bruenagel^b, Anthony J. Bauer^b, Robert E. Schoen^b and Robert H. Getzenberg^a

Departments of ^aUrology, ^bMedicine, University of Pittsburgh School of Medicine and University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA

Background: The early diagnosis of colorectal cancer (CRC) and the early detection of recurrence are central to the effective treatment of this disease. Because the prognosis of later stage CRC is poor and treatment options are limited, early detection of CRC is essential to improve patient outcomes. While available screening tests such as fecal occult blood testing, sigmoidoscopy or colonoscopy are effective, they are cumbersome, complicated to implement, and considerable barriers to their use remain. Thus, non-invasive and highly sensitive and specific early detection molecular markers of CRC are urgently needed. The nuclear matrix is the structural scaffolding of the nucleus. We have demonstrated that high resolution two-dimensional (2-D) gel analysis of nuclear matrix proteins (NMPs) in bladder, renal and prostate cancers demonstrate a specific oncological "fingerprint". An example of this approach has been our recent studies in bladder cancer in which we have successfully developed a urine immunoassay that is able to detect bladder cancer with a sensitivity of 96.4% and a specificity of 100%. Development of a tumor marker, which can detect CRC at an early stage, would greatly aid in the diagnosis of this disease.

Aim: The immediate objective of the present study was to identify the existence of a specific NMP fingerprint for human colon cancer and, there by identify unique human colon cancer nuclear matrix proteins. Ultimately the goal is to develop a molecular assay to detect early lesions.

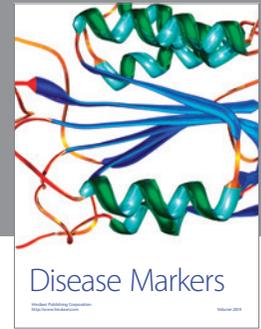
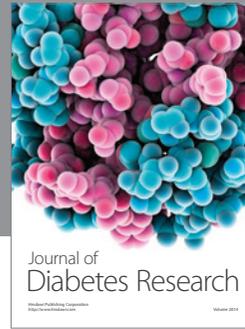
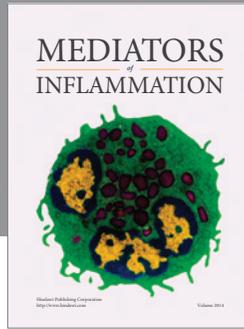
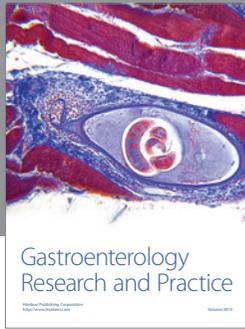
Material and methods: Using 2-D gel analysis we have identified proteins to differentiate human colon cancer tissue from normal donor and normal adjacent colon tissue. These studies included the examination of the NMP composition of ten matched colon cancer samples and adjacent tissue, and four normal donor samples, and one colon polyp, which was histologically a carcinoma *in situ*.

Results: The analysis resulted in the identification of four proteins present in all tumor samples that were not present in the matched normal adjacent and donor tis-

sues (CC2, CC3, CC4, CC5) and five proteins that are uniquely expressed in the normal adjacent and donor tissues (N2, N3, N4, N5, N6). Two proteins (CC6a/b) were also found in all cancer and donor tissues but not in the normal adjacent areas. These proteins were also identified in the colonic polyp. The polyp tissue contained two (CC3, CC4) of the four proteins, which are specific for colon cancer, and all five proteins, which are specific to the normal adjacent and donor colon tissues. Since tissue samples are complex mixtures of cells and to verify the use of cell lines as a means to characterize these proteins, two colon cancer cell lines were examined for expression of the above-described proteins. Two of the four colon cancer-associated proteins, CC3 and CC4, were verified in the human colon cancer cell line (CaCo2) but none were found in a second human colon cancer cell line (SW480). All of the proteins unique for normal adjacent and donor tis-

sue were not expressed in either of the human colon cancer cell lines. Data provided here demonstrate that examination of the nuclear matrix composition is able to differentiate colon cancer tissue from normal adjacent and donor tissue. Preliminary data from polyp tissue suggests that two specific NMPs are expressed in the early stage of colon cancer development.

Conclusion: Developing an assay with which to detect these specific NMPs is a promising modality for early detection of colorectal cancer by examining tissue, serum or stool specimens. The functional identification of these proteins and their early detection through the generation of NMP-specific antibodies could significantly impact on the understanding of colon cancer progression and greatly assist in early detection efforts. (Supported by EDRN/BDL grant UO1 CA84968 and G.B. is supported in part by the A.v.-Humboldt Foundation Fellowship.)



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

