Cancer Biomarkers 4 (2008) 129–178
IOS Press

Posters

Biorepository for validation of biomarkers predictive of the progression of Barrett's esophagus

Sharmila Anandasabapathy*, D. Kim Turgeon, Daniel Normolle, Melissa Tuck, John Poneros, Elena Stoffel, Sapna Syngal, Norman Marcon, John Baron, Mack Ruffin, Henry Appelman, Nadir Arber, Robert Bresalier and Dean Brenner
[a]*Great Lakes New England Clinical Epidemiology and Validation Center, Early Detection Research Network.*
[b]*MD Anderson Cancer Center, Houston TX, USA*
[c]*Dana Farber Cancer Center, Boston, MA, USA*
[d]*St. Michaels Hospital, Toronto, Canada, Dartmouth College, Hanover, NH, USA*
[e]*Sourasky Medical Center, Tel Aviv, Israel,*
[f]*University of Michigan Medical Center, Ann Arbor, MI, USA*

Corresponding author.

Background: Barrett's esophagus (BE) is characterized by replacement of the normal squamous epithelium with specialized columnar epithelium. Highly prevalent in the US, BE is the precursor to esophageal adenocarcinoma (EAC), one of the most rapidly rising cancers in the developed world. Fortunately, the majority of BE subjects do not progress to EAC. For those that do, diagnosis is usually made at a late, incurable stage. These factors have sparked great interest in the early detection of neoplasia via the identification of relevant biomarkers.

Aims:

(1) To create good clinical practice (GCP) quality reference sets of biosamples from Barrett's subjects for biomarker discovery and validation.
(2) In a *cross-sectional* study of subjects with varying grades of BE, to preliminarily estimate the variance of candidate biomarkers in cohorts defined by sex, race, age, and histologic diagnosis (GLNE 003).
(3) In a *longitudinal* cohort of subjects with intestinal metaplasia followed prospectively through the development of HGD/EAC, to estimate the variability of the candidate biomarkers in subjects who progress vs. those who do not (GLNE 008).

Methods: Subjects with Barrett's metaplasia, dysplasia and EAC, with no recent history of malignancy, or contraindication to endoscopy and biopsy were enrolled. Extensive questionnaires encompassing family, social, medical history, reflux symptoms, demographics, & medications were obtained as was urine, blood, biopsies, and brush cytology. Metaplastic patients were enrolled in the longitudinal protocol and agreed to repeat specimen collection on routine surveillance.

Results: Since November 2006, the GLNE Barrett's repository accrued 603 subjects across 5 centers. Table 1 describes the current repository of 246 patients reviewed to date and the demographic and clinicopathologic profile of these subjects. Metaplastics ($n = 750$) are actively being enrolled in a longitudinal trial which began in late-2007.

Conclusions: The current EDRN Barrett's database encompasses 603 subjects, one of the largest repositories to date. Efforts are underway to systematically collect samples in both a cross-sectional population of subjects with varying pathologic grades of Barrett's as well as in a longitudinal population of metaplastics to support biomarker validation.

Colon biomarker atlas: An integrated resource for EDRN GI cancer biomarker information

Kristin Anton[a,*], Daniel J. Crichton[b], Mark D. Thornquist[c], William E. Grizzle[d], Dean E. Brenner[e] and Donald Johnsey[f]
[a]*Dartmouth Medical School*
[b]*NASA Jet Propulsion Lab*
[c]*Fred Hutchinson Cancer Research Center*
[d]*University of Alabama at Birmingham*
[e]*University of Michigan*
[f]*National Cancer Institute*

Corresponding author.

ISSN 1574-0153/08/$17.00

EDRN Informatics has been focused on developing a software infrastructure allowing for distributed data systems to be integrated into a virtual Knowledge Environment. This knowledge environment includes enriching research data as well as information about specimens, biomarkers and studies. The "model-driven" architecture allows for the information content to change independent of the software core. At the center of the knowledge environment is a portal that allows access to the various information repositories. The "Biomarker Atlas" provides a visualization extension that operates compatibly with the EDRN core software infrastructure and allows researchers to access information captured in data warehouses based on anatomical and epidemiological information. Dan Crichton and the Jet Propulsion Laboratory has successfully developed a pilot "Atlas" site in collaboration with Dr. Wilbur Franklin at the University of Colorado Health Science Center using images and associated data generated from Lung Cancer studies. Dartmouth Medical School through Kristen Anton, the Jet Propulsion Laboratory through Dan Crichton, and the University of Alabama at Birmingham through Dr. William Grizzle propose a collaborative project to develop and implement, within the EDRN Knowledge Environment Atlas, an integrated resource for the GI Cancers modeled on the prototype developed with Lung Pathology data – a Colon Biomarker Atlas. The investigators will bring other EDRN GI researchers into the design process to create a most effective resource, with the intent of rapidly expanding the resource to include all EDRN GI centers.

GLNE-CEVC technical infrastructure

Kristen Anton[a,*], Mark R. Thomason[a], Robert L. Rheaume[a], Scott M. Gerlach[a], John A. Baron[a], Missy Tuck[b], Daniel Normolle[b] and Dean E. Brenner[b]
[a]*GLNE- CEVC, Dartmouth Medical School, VA, USA*
[b]*GLNE- CEVC, University of Michigan, MI, USA*

[*]*Corresponding author.*

The Great Lakes – New England (GLNE) CECV collaborates with other units of the EDRN to support the discovery and validation of biomarkers aimed at the early detection and risk assessment of two major GI malignancies–the colorectum and the esophagus. The CECV does not attempt to discover biomarkers and does not have an extensive laboratory effort. Rather, the CECV provides clinical translation of biomarkers that EDRN members and associate members discover. The CECV provides high quality, well-annotated samples consistent with EDRN Data Elements (CDEs), with the goal of developing panels of biomarkers that can be detected in human blood, serum, plasma, or urine that interrogate different aspects of known carcinogenesis mechanisms in colorectal cancer. The GLNE CEVC is supported by a robust technical infrastructure that is designed, implemented and maintained by the BioInformatics group at Dartmouth Medical School. All data is gathered and validated via Electronic Data Capture, and all biospecimens are tracked electronically with pre-organized kits and barcoded specimen containers that ensure that all specimens and specimen meta-data is controlled. Reports for all six of the CEVC protocols are available on-line, in real time, to authorized study staff. BioInformatics has collaborated with the DMCC and EDRN Informatics Team at JPL to make GLNE CEVC data available to the research community via the EDRN Resource Network Exchange (ERNE). This poster features the technical details of the GLNE CEVC bioinformatics system.

Improved methods and standards for telomerase detection: Quantitative histopathology using antibody staining

John P. Jakupciak, Nathan D. Gallant, Adrienne H. Smith, Matthew L. Becker, Alessandro Tona and Donald H. Atha[*]
Biochemical Science and Polymers Divisions, National Institute of Standards and Technology, Gaithersburg, MD, USA

[*]*Corresponding author.*

The use of telomerase as an early detection biomarker for cancer has been hindered by a lack of reliable *in situ* histochemical measurement methods and standards. Improved histochemical methods for measuring telomerase could expedite the acceptance and translation of telomerase as a biomarker for expanded use in diagnostic and clinical applications. The lack of a crystal structure for telomerase coupled with high variability in the available antibodies for immunohistochemical analysis has led to confusion in the literature regarding the binding specificity of these antibodies. To this end we have developed an automated fluorescence microscopy protocol to assess the specificity of three fluorescently-labeled telomerase antibodies

and to quantify telomerase in cultured human tumor cells and in human fibroblast cells as a control. Significant differences in the staining intensity distributions were observed. The fluorescence measurements in these cell lines were compared to telomerase measured by the telomerase repeat amplification protocol (TRAP), reverse transcription-polymerase chain reaction (RT-PCR) and flow cytometry. This combination of measurements ensured a more complete quantitation of telomerase levels in each of the respective cell lines and could be used as a model in the validation of this method for clinical use.

Sonic hedgehog promotes metastasis and desmoplasia in pancreatic cancer

Jennifer M. Bailey* and Michael A. Hollingsworth
Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE, USA

**Corresponding author.*

Sonic hedgehog (SHH) is a 19kD secreted protein, whose expression is tightly regulated throughout development. In the adult, aberrant expression of SHH is associated with the onset and progression of pancreatic cancer, as evidenced by increased levels of expression in premalignant and malignant lesions of the pancreas. We propose that SHH, secreted from adult pancreatic epithelia, functions in a paracrine manner to recruit and increase proliferation of cells of mesenchymal lineage. To address this hypothesis, we expressed SHH in a "normal" pancreatic epithelial cell line, hTert-HPNE, that was immortalized by stable expression of recombinant human telomerase. These primary cells were fully transformed after transduction with retroviruses that delivered the following genetic insults: mutated K-ras, human papillomavirus E6/E7 and the SV40 small T antigen. We transduced a retroviral vector into this transformed cell line that caused expression of SHH. We implanted transformed HPNE cells into the pancreas of athymic nude mice and discovered that SHH expression increased tumor size and incidence of metastasis to the spleen, peritoneum and liver in an orthotopic model of tumor challenge. Primary tumors expressing SHH displayed increased desmoplasia, and showed evidence of infiltration by myofibroblasts as detected by imuunohistochemical staining of tumor sections for two mesenchymal markers: vimentin and α-smooth muscle actin. Orthotopic challenge with another pancreatic tumor cell line expressing SHH, Capan-2, led to desmoplasia that was significantly inhibited by treatment once per week with 500ug of a SHH neutralizing antibody. This treatment also decreased the primary tumor growth rate and inhibited metastasis to the spleen, liver and lymph nodes. These data suggest that SHH, secreted from pancreatic epithelia, may be a critical player in establishing and regulating the tumor microenvironment through the recruitment and expansion of pancreatic myofibroblasts.

Distinct patterns of immunoreactivity from p53-specific autoantibodies in breast and ovarian cancer patient sera

Rodrigo Barderas[a,*], Niroshan Ramachandran[a], Genie Hainsworth[a], Jessica Wong[b], Jacob Raphael[a], Daniel Cramer[c], Joshua LaBaer[a] and Karen S. Anderson[b]
[a]*Harvard Institute of Proteomics, Harvard Medical School, Boston, MA, USA*
[b]*Cancer Vaccine Center, Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA*
[c]*Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*

**Corresponding author.*

Background: P53-specific gene mutations are frequent in multiple cancers, including 20% of breast cancers and 50% of ovarian cancers. The presence of mutated p53 in tumors is associated with the development of p53-specific autoantibodies. Their use as biomarkers for early cancer detection is limited because anti-p53 antibodies have limited sensitivities (10–20%) and are not specific for cancer type. We sought to map the immunogenic regions of p53 to determine the variability in epitope specificity in breast and ovarian cancer sera.

Methods: 447 sera from breast cancer, 15 sera from ovarian cancer, and 46 control female sera were screened by ELISA for anti-p53 antibodies. 30 breast cancer sera (7.4%) and 3 ovarian cancer sera (20%) were strongly positive and 8 sera were weakly positive. 16 of these sera were selected for p53 epitope mapping studies and probed on a nucleic acid programmable protein array (NAPPA) printed with 38 p53 deletions from the N-terminal and C-terminal end, 12 p53 tiled fragments and wild-type p53.

Results: All 16 sera had detectable antibodies to wild-type p53 protein expressed on the protein array. By clustering analysis of the deletion fragments, we identified i) a broad antibody reactivity throughout the molecule (6 patients), and ii) a specific terminal reactivity as well as core-region reactivity that was uncovered only with progressive deletions of the termini (10 patients).

Conclusions: These results demonstrate distinct patterns of p53-specific immunoreactivity in the sera of breast and ovarian cancer patients. The majority of patients have antibody reactivity to the core-region of the molecule. The association of the observed antibody specificities with p53 mutation status is being explored.

Ngal/Lipocalin-2 is a marker of early dysplastic changes in the pancreas and a possible diagnostic marker in pancreatitis and/or pancreatic adenocarcinoma

Subhankar Chakraborty[a], Nicolas Moniaux[a], Valerie K. Shostrom[d], Subodh M. Lele[c], Michel Ouellette[b], Aaron R. Sasson[e], Maneesh Jain[a], Michael A. Hollingsworth[b] and Surinder K. Batra[a,b,c,*]

[a]*Department of Biochemistry and Molecular Biology*

[b]*Eppley Institute for Cancer Research, Omaha, NE, USA*

[c]*Department of Pathology and Microbiology, Omaha, NE, USA*

[d]*Department of Societal and Preventive Medicine, Omaha, NE, USA*

[e]*Department of Surgery; University of Nebraska Medical Centre, Omaha, NE, USA*

[*]*Corresponding author.*

Pancreatic cancer (PC) is a highly lethal malignancy with a dismal 5-year survival of less than 5%. Neutrophil gelatinase associated lipocalin (NGAL) expression was reported to be upregulated nearly 27-fold in PC cells compared to normal pancreatic ductal cells in a microarray analysis. NGAL/Lipocalin-2 is a 24-kDa glycoprotein with multiple functional roles including acting as a bacteriostatic agent and a regulator of iron transport into cells. Further, NGAL has been shown to promote differentiation during embryonic development. Given our interest in biomarkers for the early diagnosis of PC, we hypothesized that NGAL would be overexpressed in the PC and that its expression would positively correlate with the degree of differentiation of the cancer cells. The objective of our study was to examine whether immunostaining for NGAL could be used to identify foci of PC, specifically early dysplasia in tissue sections. Further, as NGAL is a secreted glycoprotein, so we wanted to investigate whether there is an elevation in serum NGAL levels in PC cases. To investigate our hypothesis, we analyzed the expression of NGAL in pancreatic tissue samples by immunohistochemistry and measured serum NGAL levels by a quantitative sandwich ELISA assay. The composite scores for a given area (normal, pancreatitis, PanIN or cancer) were compared between the groups using the Mann-Whiteny U test. We observed that while NGAL expression was strongly upregulated in PC, and moderately in pancreatitis, only a weak expression could be detected in the healthy pancreas The average composite score of PC areas (4.26 ± 2.44) was significantly higher than that of normal pancreas (1.0) or areas of pancreatitis (1.0) ($p < 0.0001$). Further, while both well and moderately differentiated PC were positive for NGAL, areas of poorly differentiated adenocarcinoma were uniformly negative. Importantly, NGAL expression was detected as early as the PanIN-1 stage, suggesting that it could be a marker of early pre-malignant change in the pancreas. When serum NGAL levels were analyzed using the upper limit for healthy individuals (106 ng/ml) as the cut-off between normal and diseased cases, 94% of PC and 62.5% each of acute and chronic pancreatitis samples were above the cut-off. Non-parametric tests were used to assess the significance of any difference in mean NGAL levels between the four groups (normal, acute/chronic pancreatitis, PC). Although serum NGAL levels were significantly elevated in acute ($p = 0.035$), chronic pancreatitis ($p = 0.035$) and PC (0.004), the difference between serum levels in pancreatitis and PC was not significant. Upon analyzing the ability of the serum assay to distinguish PC from non-pancreatic cancer cases by plotting a receiver operating characteristic (ROC) curve, the area under the curve emerged to be 0.75, suggesting that the test is fairly accurate in distinguishing the two groups. Taken together, the results from the current study led us to conclude that NGAL/Lipocalin-2 is overexpressed in PC, specifically in the earliest pre-malignant lesion (PanIN-I). Also, NGAL expression positively correlates with the degree of differentiation in PC and serum NGAL levels could be explored in combination with other markers in the early diagnosis of acute/chronic pancreatitis and/or pancreatic adenocarcinoma.

Methylation markers for lung cancer risk prediction

Steven Belinsky
Lovelace Respiratory Research Institute, NM, USA

Lung cancer is the leading cause of cancer-related death in the U.S. and will soon reach epidemic levels worldwide. Mortality from this disease could be reduced through the development of an effective screening strategy for identifying persons with early stage disease and the implementation of chemopreventive strategies that can reverse or impede the progression of premalignant disease. Studies by our laboratory have evaluated genes inactivated by aberrant cytosine-guanosine (CpG) island methylation as candidate biomarkers for early detection of lung cancer. The specific hypothesis being evaluated is that methylation of genes detected in sputum can be used to identify early lung cancer in asymptomatic persons. We conducted the first study in collaboration with the Colorado Lung SPORE to prospectively evaluate a large panel of genes for their ability to predict lung cancer. This nested, case-control study of persons from the Colorado cohort revealed that a panel of genes could predict incident lung cancer 3–18 months prior to clinical diagnosis. Specifically, concomitant methylation of three or more of a six-gene panel was associated with a 6.5-fold risk and a sensitivity and specificity of 64%. We have extended our initial case-control studies with the Colorado cohort for the purpose of improving the sensitivity and specificity of the original gene panel. There were two goals for these studies: to increase power by increasing the number of cases and controls and to screen an additional 40 genes. Five genes (DAL1, PCDH20, KIF1A, P16, and DAPK) have now been identified that show significantly increased odds for methylation in cases compared to controls. In addition, we have identified 14 genes associated with a 2-fold or more increased lung cancer risk. Gene panels are being assembled to determine sensitivity and specificity. In addition, to better refine our gene panel for prospective studies, gene methylation is being assessed in sputum obtained from stage I lung cancer patients who are generally asymptomatic for disease. Results with an 8-gene panel revealed that the prevalence for methylation was similar (e.g. p16, MGMT) or strikingly exceeded (e.g., GATA4, GATA5) that seen in the Colorado lung cancer cases. This finding parallels our observation in the Colorado case-control study where the prevalence for methylation of several genes increased as the time between sputum collection and cancer diagnosis decreased.

Gene haplotypes associated with lung cancer could integrate with a gene methylation panel to improve the sensitivitiy and specificity for early lung cancer detection. We assessed the distribution of variants in 66 genes from the DNA repair, cell cycle, methylation, and apoptotic pathways in lung cancer cases and controls from the Colorado cohort by the Illumina Goldengate assay. Principal component analysis identified several genes that appear to be associated with lung cancer. Validation studies are planned that will lead to the identification of specific gene variant alleles and haplotypes that can be integrated into an early detection panel for screening high-risk smokers. (Supported by P50 CA 58184 and U01 CA09735697)

MicroRNA regulation of epidermal growth factor receptor in NSCLC

Lynne T. Bemis*, Glen J. Weiss, Michio Sugita, E. Nakajima, Diane Burks, William A Robinson, Marileila Varella-Garcia, Jerry Haney, Fred Hirsch, H. Kato, Paul A. Bunn Jr. and Wilbur A. Franklin
University of Colorado Denver School of Medicine, Aurora, CO, USA

**Corresponding author.*

We hypothesized that a microRNA might regulate the epidermal growth factor receptor (EGFR) thereby acting like a tumor suppressor gene and loss of such a microRNA would allow increased EGFR expression. In order to identify putative microRNA regulators of EGFR we queried the TargetScan database, which lists predicted microRNA regulators of genes of interest. miR-128 and miR-27 were listed as possible regulators of EGFR. Since miR-128b is located in the chromosomal location of 3p22, a region of common and early loss in lung cancer, we studied the ability of this microRNA to regulate EGFR expression in NSCLC.

Applying mimics and inhibitors of miR-128b in NSCLC cell lines we were able to show regulation of EGFR by miR-128b. These studies were complemented by reporter assays showing that the predicted binding sites for miR-128b in the 3'UTR of EGFR could be mutated and thereby block regulation by miR-128b. In addition, we examined the loss of miR-128b by qPCR of the genomic copy number in patient samples and found that loss of miR-128b correlates with overall survival in patients who had been treated with gefitinib.

We conclude that EGFR is regulated by miR-128b and furthermore that loss of miR-128b is frequent in tu-

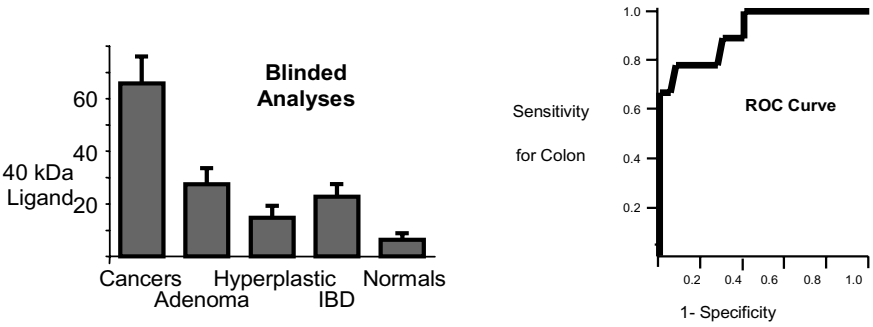

Fig. 1.

mor samples and correlates significantly with response and survival following gefitinib treatment. In contrast, EGFR mutation status did not correlate with survival.

Enrichment of serum phosphopeptide by nanoparticles to identify cancer biomarkers

A. Shah, V. Sachdev and A. Bharti*

**Corresponding author: Boston University Medical School, Boston, MA, USA.*

The role of kinases in transformation and cancer progression is established. We hypothesize that the role the specific phosphoproteins/phosphopeptides will be present in cancer patient serum as leak proteins and their isolation and identification will bear diagnostic/prognostic value. MS based methods are efficient for identification and characterization of phosphopeptides. However, low abundance serum phosphoprotein isolation and characterization still remains a challenge. The existing technology (micro-columns) to isolate and enrich phosphopeptide from biological samples lacks the specificity, efficiency and flexibility necessary to analyze larger sample volumes. To overcome these problems we have used nanoparticles to enrich phosphopeptide from serum and then analyze by mass spectrometry (MS). Nanoparticles provide a very large surface area (at least 1000 fold), is specific (amenable to functionalization) and can analyze larger serum samples (up to 10 ml). We have developed and characterized Fe_3O_4-SiO_2-TiO_2 core shell nanoparticles to specifically enrich phosphopeptides. Silica was first coated on 12 nm Fe_3O_4 magnetic core and subsequently TiO_2 was cross linked. The size of the Fe_3O_4-SiO_2-TiO_2 particles was determined to be 18nm and was uniform. Preliminary data indicates that particles bind specifically to phosphopeptides and these phosphopeptides bound to nanoparticles can also be analyzed directly by MALDI-MS. The ongoing work will determine the efficiency of phosphopeptide enrichment from serum.

A circulating ligand for galectin-3 is a haptoglobin-related glycoprotein elevated in the serum of individuals with colorectal neoplasia

Robert S. Bresalier*, Nachman Mazurek, James C. Byrd, David Hawke, Dean Brenner, Daniel Normolle, Sapna Syngal, Norman Marcon, Kim Turgeon, John Baron and the Great Lakes-New England Clinical and Epidemiology Center of the Early Detection Research Network

**Corresponding author: MD Anderson Cancer Center.*

Galectin-3 is a multifunctional beta-galactoside-binding protein involved in tumorigenesis, progression and metastasis of several cancer types including colorectal cancer. Post-translational modification through glycosylation is an important mechanism which confers heterogeneity and diversity to tumor-associated proteins. Since galectin-3 is a cell surface and secreted carbohydrate binding protein, we determined whether circulating galectin-3 ligands in serum are associated with colorectal neoplasia. In colon cancer sera (and in some patients with adenomas), the major circulating galectin-3 ligand is a 40 kDa glycoprotein distinct from mucin, CEA and Mac-2-binding protein. This was purified and identified by MALDI-MS as a haptoglobin-related protein. Further analysis suggested that this ligand represents an aberrantly glycosylated form of haptoglobin which may be produced by neoplastic cells in the colon. Preliminary validation using serum samples

across a variety of disease states suggested that this protein is present at levels 30- fold higher in patients with colorectal cancer compared to healthy subjects. Further validation using blinded reference samples from the Great Lakes-New England Clinical and Epidemiology Center (GLNE) of the EDRN confirmed these results. Using this reference set, levels of the 40kD ligand were significantly elevated in individuals with colon cancer (p <0.00001) compared to other groups, and in those with adenomas compared to healthy controls ($p = 0.0075$).

A constructed receiver operating characteristic curve of sensitivity versus (1-specificity) further indicated that levels of this ligand may successfully differentiate individuals with colon cancer from other groups. We have recently developed several ELISAs which may be useful for higher through-put analysis of serum samples for this glycoprotein. In preliminary validation of one of these assays, the sensitivity and specificity in differentiating sera from cancer patients versus normal controls was 91% and 82% respectively. We have also confirmed that colon cancer cells produce a form of haptoglobin distinct from that produced in the liver.

[-2]proPSA improves prostate cancer detection: An EDRN validation study

Lori J. Sokoll[a,*], Yinghui Wang[b], Ziding Feng[b], Jacob Kagan[c], Alan W. Partin[a], Martin G. Sanda[d], Ian M. Thompson[e] and Daniel W. Chan[a]
[a]*Johns Hopkins University, Baltimore, MD, USA*
[b]*Fred Hutchinson Cancer Research Center, Seattle, WA, USA*
[c]*National Cancer Institute, Bethesda, MD, USA*
[d]*Beth Israel Deaconess Medical Center, Boston, MA, USA*
[e]*University of Texas Health Science Center at San Antonio, TX, USA*

[*]*Corresponding author.*

The molecular forms of free PSA, particularly proPSA, show promise in improving the early detection of prostate cancer compared to PSA and percent free PSA (%fPSA). In this study, the [-2]proPSA serum marker was evaluated using a blinded, archival, reference specimen set from three NCI Early Detection Research Network (EDRN) Clinical Epidemiology and Validation Centers (Beth Israel Deaconess Medical Center, Johns Hopkins University, University of Texas Health Science Center at San Antonio). Serum was collected pre-biopsy from 123 men with no prior biopsy or prostate cancer history. Specimens (cancer group: 51%; non-cancer group: 49%) were analyzed for PSA, free PSA, [-2]proPSA, BPSA, and testosterone on the Beckman Coulter ACCESS analyzer). Analytes were evaluated individually and in combination. Total PSA concentrations were similar between the two groups (non-cancer: 6.80 ± 5.20 ng/mL, cancer: 6.94 ± 5.12 ng/mL). Using ROC analysis in all subjects, %[-2]proPSA had the greatest area under the curve (AUC = 0.69) followed by %fPSA (AUC = 0.61). For %[-2]proPSA, maximal sensitivity was 60% and specificity was 70%. A logistic regression model combining PSA, BPSA, %fPSA, %[-2]proPSA, [-2]proPSA/BPSA, and testosterone had an AUC of 0.73. In the 2–10 ng/mL PSA range, %[-2]proPSA and the model had the largest AUC (0.73) with an AUC for %fPSA of 0.53. Specificities for %[-2]proPSA, the logistic regression model, and %fPSA at 95% sensitivity were 31%, 26%, and 16%, respectively. In summary, using a blinded EDRN reference set, %[-2]proPSA was the best predictor of prostate cancer detection compared to %fPSA in the 2–10 ng/mL total PSA range. Further studies are needed to validate these findings.

Discovery of gene fusions in prostate cancer and implications for cancer detection, diagnosis and prevention

Arul M. Chinnaiyan
University of Michigan, MI, USA

To date, the great majority of disease-specific, recurrent chromosomal rearrangements have been characterized in hematological malignancies and mesenchymal tumors and not in common epithelial tumors such as breast, lung, colon, or prostate cancer. Here, we employed a bioinformatics approach on a compendium of cancer gene expression data to discover candidate oncogenic chromosomal aberrations based on outlier gene expression. In addition to identifying many gene partners of characteristic rearrangements in human malignancies, this approach identified two members of the ETS family of transcription factors, *ERG* and *ETV1*, as outliers in prostate cancer. Either *ERG* or *ETV1* was over-expressed in the majority of prostate cancers (50–70%) and were mutually exclusive across several independent gene expression datasets, suggesting that they may be functionally redundant in prostate cancer development.

By RNA ligase-mediated rapid amplification of cDNA ends (RACE), we identified a recurring gene fusion of the 5' untranslated region of a prostate-specific, androgen-regulated gene *TMPRSS2* to *ERG* or *ETV1* in prostate cancer cases which over-expressed the respective ETS family member. These gene fusions were confirmed using quantitative PCR (QPCR) and sequencing of reverse transcription PCR products. In addition, using fluorescence in situ hybridization (FISH), we demonstrated that 23 of 29 (79%) prostate cancer samples harbor rearrangements in *ERG* or *ETV1*. Furthermore, in vitro cell line studies suggest that the androgen-responsive promoter elements of *TMPRSS2* mediate the aberrant over-expression of *ETS* family members in prostate cancer. Subsequently, we interrogated the expression of all *ETS* family members in prostate cancer profiling studies and identified outlier expression of ETV4 in two of 98 cases. In one such case, we confirmed the over-expression of *ETV4*, and by RACE, QPCR and FISH, we identified fusion of the *TMPRSS2* and *ETV4* loci.

Together, these results suggest a pathogenetically important role for recurrent chromosomal rearrangements in common epithelial tumors and have implications in the molecular diagnosis and treatment of prostate cancer. Importantly, these results identify three molecular subtypes of prostate cancer, *TMPRSS2:ERG*, *TMPRSS2:ETV1* and *TMPRSS2:ETV4*, and suggest that dysregulation of *ETS* family member expression through gene fusions with *TMPRSS2* may be a generalized mechanism for prostate cancer development.

In our most recent work, we explored the mechanism of *ETS* family over-expression in prostate tumors. Remarkably, we identified novel 5' fusion partners in prostate tumors with outlier expression of ETS family members, including untranslated regions from a prostate-specific androgen-induced gene and endogenous retroviral element, a prostate-specific androgen-repressed gene, and a strongly expressed housekeeping gene. As the commonality of these rearrangements is the aberrant over-expression of *ETS* genes, we recapitulated this event *in vitro*. We demonstrate that *ETS* over-expression in multiple benign prostate cells induces a marked increase in invasion, confirming the role of *ETS* gene rearrangements in prostate cancer development. Identification of distinct classes of *ETS* gene rearrangements demonstrates that dormant oncogenes can be activated in prostate cancer by juxtaposition to tissue-specific or ubiquitously active genomic loci. Subversion of active genomic regulatory elements may serve as a more generalized mechanism for carcinoma development. Furthermore, the identification of androgen-repressed and insensitive 5' fusion partners has important implications for the anti-androgen treatment of advanced prostate cancer.

References

[1] S.A. Tomlins, D.R. Rhodes, S. Perner, S.M. Dhanasekaran, R. Mehra, X.-W. Sun, S. Varambally, X. Cao, J. Tchinda, R. Kuefer, C. Lee, J.E. Montie, R.B. Shah, K.L. Pienta, M.A. Rubin and A.M. Chinnaiyan, Recurrent Fusion of TMPRSS2 and ETS Transcription Factor Genes in Prostate Cancer, *Science* **310**(5748) (2005), 644–648.

[2] S. Varambally, J. Yu, B. Laxman, D.R. Rhodes, R. Mehra, S.A. Tomlins, R.B. Shah, U. Chandran, F.A. Monzon, M.J. Becich, J.T. Wei, K.G. Pienta, D. Ghosh, M.A. Rubin, A.M. Chinnaiyan, Integrative Genomic and Proteomic Analysis of Prostate Cancer Reveals Signatures of Metastatic Progression, *Cancer Cell* **8**(5) (Nov 2005), 393–406.

[3] S.A. Tomlins, R. Mehra, D.R. Rhodes, R.B. Shah, M.A. Rubin, E. Bruening, V. Makarov and A.M. Chinnaiyan, Whole Transcriptome Amplification for Gene Expression Profiling and Development of Molecular Archives, *Neoplasia* **8**(2) (Feb 2006), 153–162.

[4] S.A. Tomlins, R. Mehra, D.R. Rhodes, L.R. Smith, D. Roulston, B.E. Helgeson, X.C. Cao, J.T. Wei, M.A. Rubin, R.B. Shah and A.M. Chinnaiyan, TMPRSS2:ETV4 Gene Fusions Define a Third Molecular Subtype of Prostate Cancer, *Cancer Res* **66**(7) (Aprl 2006), 3396–400.

[5] S. Perner, F. Demichelis, R. Beroukhim, F.H. Schmidt, J.M. Mosquera, S. Setlur, J. Tchinda, S.A. Tomlins, M.D. Hofer, K.G. Pienta, R. Kuefer, R. Vessella, X.W. Sun, M. Meyerson, C. Lee, W.R. Sellers, A.M. Chinnaiyan and M.A. Rubin, TMPRSS2:ERG Fusion-Associated Deletions Provide Insight into the Heterogeneity of Prostate Cancer, *Cancer Res* **66**(17) (Sep 2006), 8337–841.

[6] B. Laxman, S.A. Tomlins, R. Mehra, D.S. Morris, L. Wang, B.E. Helgeson, R.B. Shah, M.A. Rubin, J.T. Wei and A.M. Chinnaiyan, Noninvasive detection of TMPRSS2:ERG fusion transcripts in the urine of men with prostate cancer, *Neoplasia* **8**(10) (Oct 2006), 885–888.

[7] S.A. Tomlins, R. Mehra, D.R. Rhodes, X. Cao, L. Wang, S.M. Dhanasekaran, S. Kalyana-Sundaram, J.T. Wei, M.A. Rubin, K.J. Pienta, R.B. Shah and A.M. Chinnaiyan, Integrative molecular concept modeling of prostate cancer progression, *Nat Genet* **39**(1) (Jan 2007), 41–51.

[8] D.R. Rhodes, S. Kalyana-Sundaram, V. Mahavisno, R. Varambally, J. Yu, B.B. Briggs, T.R. Barrette, M. Anstet, C. Kincead-Beal, P. Kulkarni, S. Varambally, D. Ghosh and A.M. Chinnaiyan, Oncomine 3.0: Genes, Pathways and Networks in a Collection of 18,000+ Cancer Gene Expression Profiles, *Neoplasia*, In Press, 2007.

Methylation-based detection of kidney cancer

Paul Cairns* and Robert G. Uzzo
Fox Chase Cancer Center, Philadelphia, PA, USA

**Corresponding author.*

Kidney cancer confined by the renal capsule can be surgically cured in the majority of cases whereas the prognosis for patients with advanced disease at presentation remains poor. Novel strategies for early detection are therefore needed. Kidney tumors are heterogeneous in their histology, genetics and clinical behavior. We have found that promoter hypermethylation of tumor suppressor genes is common, can occur relatively early, may disrupt critical pathways, and thus likely plays a critical role in kidney tumorigenesis. We have also demonstrated sensitive and specific early detection of renal cancer in urine using a hypermethylated gene panel. We have generated, and communicated to the field, metrics for validation of methylation-based detection of cancer. To advance our goals, we have recently completed a comprehensive gene methylation profile to determine the timing of methylation in early stage, curable renal cancer and to identify the optimal methylated genes for inclusion in panel for early detection. Our study of urines obtained at follow-up of patients with no clinical evidence of disease several weeks or months after undergoing nephrectomy for organ-confined disease further supports the specificity of gene methylation for the presence of cancer and, importantly, may prove to be a useful procedure for molecular monitoring of renal cancer. We have performed a demethylating drug-based global epigenetic reactivation in renal carcinoma cells and thereby generated the first pass of the renal cancer cell methylome. Novel candidate tumor suppressor genes identified by this screen are profiled for methylation status in primary early stage renal tumors to determine utility for early detection and differential diagnosis of kidney cancer.

Early Detection Research Network (EDRN) eSIS project: Tracking study related research information and milestones using informatics tools

Jackie Dahlgren*(DMCC), Greg Warnick (DMCC), Deanna Stelling (DMCC), Don Johnsey (NCI), Dan Crichton (JPL), Cim Edelstein (DMCC), Steve Hughes (JPL), Sean Kelly (JPL), Peter Lin (DMCC), Royce Malnik (DMCC), Michelle Smerek (DMCC), Sudhir Srivastava (NCI), Suzanna Reid (DMCC) and Mark Thornquist (DMCC)

**Corresponding author: Fred Hutchinson Cancer Research Center*

EDRN's primary goal is to discover and validate biomarkers targeting early detection of cancer. This process, end to end, involves many steps from initial discovery to eventual large scale population studies prior to marketing and clinical use. In part, EDRN's success is evident by the number of investigators (>300) and biomarkers (>120) that are under investigation. The number of projects/protocols expanded to a point where an informatics solution to easily track study progress and view high level protocol summaries was necessitated. The EDRN Study Information System (eSIS) was created to deliver a high level protocol summary and progress overview.

Communication of project progression and accomplishments are best tracked using milestones. EDRN supports investigators by funding distinct biomarker related scientific projects and supplying a sound infrastructure including expert research coordination support, statistical analysis and informatics infrastructure. One focus of EDRN is promotion of a collaborative environment where information regarding EDRN funded studies is shared. Due to expansion in the EDRN network and increased volumes of information contained within EDRN databases eSIS was developed. Through the use of CDEs (Common Data Elements) project and protocol information as well as study milestones have been annotated with sufficient metadata for use throughout the EDRN ontology and knowledge environment. EDRN researchers are now provided a web based utility (through the EDRN secure site) by which to update milestones and list study related publications. EDRN members, NCI/DMCC administrative staff and the general public are able to view project/protocol summaries and milestones.

eSIS integrates summary level information for EDRN protocols on a single web page. Information is gleaned from *Site Manager* and *Protocol Manager* entries made through the EDRN secure site. This information is maintained by the DMCC, Lead PI and Involved PIs (at various sites) for each study. Additional fields were added to the *Site Manager* and *Protocol Manager* to complete the study basics and milestone information displayed in eSIS. eSIS simplified navigation to project/protocol summary and milestone information allowing for a quick status review of all EDRN related studies.

Changes in the stem/progenitor cell population can be used in cancer risk assessment and early detection

Gabriela Dontu[a,*], Christophe Ginestier[a], Suling Liu[a], Emmanuelle Charafe-Jauffret[b], Madhuri Kakarala[a] and Max S. Wicha[a]

[a]*Comprehensive Cancer Center, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA*
[b]*Centre de Recherche en Cancérologie de Marseille, Laboratoire d'Oncologie Moléculaire, UMR599 Inserm/Institut Paoli-Calmettes, France*

[*]*Corresponding author.*

Recently, increasing evidence accumulated to support the concept that the primary targets of transformation are the stem and progenitor cells. Furthermore it appears that tumors are organized in a hierarchy similar to that of normal tissues, where a "cancer stem cell" component is driving tumor progression. Based on this model, new strategies meant to target the cancer stem cells are being developed. These concepts may be applied to developing new methods of cancer risk reduction and early detection. One of the key events preceding transformation may be the de-regulation of the mechanisms that control the fate of normal stem and early progenitor cells, resulting in an initial expansion of this population. Any silent mutations harbored by the stem cells population would be propagated during this cellular expansion and eventually, additional oncogenic events will lead to transformation. Therefore, detection of an expanded stem/progenitor cell population in normal or pre-malignant lesions could be used in early detection.

Herein we present a set of experimental tools that can be used to assess expansion of the stem cell population in the normal human mammary epithelium, using in vitro and in vivo models. We have developed an in vitro experimental system based on suspension culture, in which normal mammary epithelial cells can be propagated as stem/progenitor cells, as spherical colonies (mammospheres). This experimental system allows for the in vitro study of normal mammary stem cell self-renewal. We also showed that aldehyde dehydrogenase 1 (ALDH1), identifies stem /progenitor cells in both the normal and the malignant human mammary epithelium and can be used to detect stem cell in situ, by immuno-staining. Supporting the concept that an expansion in the stem cell population precedes transformation, down-regulating BRCA1 in human normal mammary epithelial cells increases self-renewal in vitro, as shown by increase in mammosphere formation in suspension culture. Moreover, and expanded ALDH 1 positive stem/progenitor cell population can be detected in mammary biopsy samples from BRCA1 healthy carriers with apparently normal histopatholgical aspect. Allelic analysis at the BRCA1 locus performed on tissue collected by laser capture microdissection indicated that these ALDH1 positive epithelial cells represented indeed the pre-malignant lesion. LOH at the BRCA1 locus was detected in at least one of the BRCA1 polymorphic markers in ALDH1-positive but not in adjacent ALDH1 negative lobules.

These findings support our hypothesis that changes in the stem cell population can be used in cancer risk assessment or in early detection.

A prevalidation study for the evaluation of serum biomarkers of Barrett's esophagus and esophageal adenocarcinoma

Richard R. Drake[a,*], Takatsugu Kan[b], Izabela D. Karbassi[a], Marjan Gucek[b], Fumiaki Sato[b], Daniel P Normolle[c], Lisa H. Cazares[a], Sapna Syngal[d], Robert Bresalier[e], Norman Marcon[f], Robert Cole[b], O. John Semmes[a], Dean Brenner[c] and Stephen J. Meltzer[b]
[a]*Center for Biomedical Proteomics, Eastern Virginia Medical School, Norfolk, VA, USA*
[b]*GI Division, Dept. of Medicine, Johns Hopkins University, Baltimore, MD, USA*
[c]*Dept of Internal Medicine, University of Michigan, Ann Arbor, MI, USA*
[d]*Dept of Medicine, Dana Farber Cancer Center and Brigham and Womens Medical Center, Boston, MA, USA*
[e]*Dept of Gastroenterology, MD Anderson Cancer Center, Houston, TX, USA*
[f]*Dept of Medicine, St Michaels Hospital, Univ of Toronto, Toronto, Ont, Canada*

[*]*Corresponding author.*

Barrett's esophagus (BE), a premalignant metaplastic columnar mucosa due to chronic gastroesophageal reflux disease, is considered the obligate precursor lesion of esophageal adenocarcinoma (EAC). Detection and surveillance of BE is the best known means to successfully intervene in Barrett's-associated neoplastic progression. As endoscopy is unsuitable and impractical for population-based screening or detection of asymptomatic BE or EAC, a serum-based detection assay would have clinical utility. We applied a MALDI-TOF profiling workflow that incorporated trypsin digestion and identification of discriminating peaks to establish a serum protein panel distinguishing patients with BE or EAC from unaffected individuals (UI). Us-

ing linear discriminant analysis, we developed a panel demonstrating an area under the receiver-operator characteristic curve (AUROC) of 0.955 in discriminating patients with BE from those without this condition. Moreover, several of the proteins identified by this initial screen were validated using ELISA assays. iTRAQ comparing BE, EAC, and UI sera was also performed in a parallel study; some of the differentially abundant proteins identified by iTRAQ matched those from our MALDI-TOF screen. As a result of these findings, a blinded prevalidation study of this serum protein panel was designed to analyze fifty EA, fifty BE and fifty UI samples, collected and managed by the Great Lakes-New England (GLNE) consortium at the U. Michigan EDRN CEVC. Based on scores determined by the submission of MALDI-TOF-MS profiles to the LDF, samples will be classified as either EA, BE or UI, and these classifications will be sent to the GLNE CEC for unblinding. If successful, results from this prevalidation experiment will establish a foundation for a larger multicenter validation study.

Expressed prostatic secretion (EPS) collection for proteome characterization as an enriched source of prostatic disease biomarkers

Richard R. Drake*, Thomas W. Fuller, Lifang Yang, Mary Ann Clements, LiNing Qi, Paul F. Schellhammer, Robert W. Given, Donald F. Lynch, Raymond S. Lance and O. John Semmes
Center for Biomedical Proteomics and the Virginia Prostate Center, Eastern Virginia Medical School, Norfolk, VA, USA

**Corresponding author.*

Expressed prostate secretions (EPS) represent organ proximal fluids ideal for use in protein biomarker discovery strategies for prostatic disease diagnostics. We have initiated a clinical strategy for collection of EPS from men with benign disease and prostate cancers, and begun the comprehensive proteome characterization of the protein constituents present in EPS to identify potential prostatic disease biomarkers. The EPS urines were obtained following gentle prostate massage during digital rectal examination prior to scheduled prostate biopsy procedures. The massage consisted of three strokes on each side of the median sulcus of the prostate, forcing the expressed fluid directly into the urethra. Urine (10–20 ml) containing the EPS were collected, centrifuged to remove the cell sediment, and stored at -80°C. In a subset of donors, pre-DRE urines were collected prior to their EPS urine specimens. Undiluted EPS secreted from the penis following a pre-prostatectomy exam under anesthesia have also been collected. A combinatorial proteomic approach involving sample fractionation prior to mass spectrometry identification was used. Current procurement rates of EPS urines have been a consistent 15–20 samples/month per urologist, and 6–8 undiluted EPS/month. For proteomic characterization, the three separation methods were applied to representative EPS urines from benign disease and prostate cancers. Following tandem mass spectrometry sequencing of proteins within these fractions, over 500 individual proteins have been cumulatively identified, most functionally associated with extracellular and cell surface processes. Comparisons with proteins similarly analyzed in pre-DRE urines and undiluted EPS samples were also done. EPS is a rich source of potential protein biomarkers, and its collection can be easily incorporated into routine prostate examinations for benign and cancerous conditions, particularly at the time of biopsy.

Searching for 13q key players in esophageal squamous-cell carcinogenesis by qRT-PCR of microdissected tissues

Heidi S. Erickson[a,*], Jaime Rodriguez-Canales[a], Paul S. Albert[b], Sumana Mukherjee[a], Nan Hu[c], Alisa M. Goldstein[c], Rodrigo F. Chuaqui[a], Philip R. Taylor[c] and Michael R. Emmert-Buck[a]
[a]*Pathogenetics Unit, Laboratory of Pathology and Urologic Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA*
[b]*Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA*
[c]*Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA*

**Corresponding author.*

Esophageal squamous-cell carcinoma (ESCC) is exceptionally prevalent in the Shanxi Province of China. Frequent allelic loss on chromosome 13 has been shown in these cases. Furthermore, comprehensive genomic analysis of ESCC demonstrated more frequent loss of heterozygosity (LOH) on 13q associated

with family history positive ESCC than family history negative patients, indicating a gene or genes within this 13q region putatively plays an important role in ESCC etiology and progression. Previous linkage analysis revealed four common genomic deletion regions, including 13q12.11 and 13q12.3-q13.1 in family history positive patients. To identify key players dramatically up- or down-regulated in this region, we used qRT-PCR to analyze genes located between 13q11 and 13q14. Seventeen cases of matched frozen ESCC tumor and normal tissue sections were microdissected and RNA isolated. Twelve genes (LATS2, EFHA1, TNFRSF19, RNF6, KATNAL1, HSP1, BRCA2, RFC3, CCNA1, ELF1, TPT1, and RB1) were analyzed by qRT-PCR using appropriate normalization strategies. Statistical analysis was conducted by two-sided T-tests using a Bonferonni multiple comparisons correction. P-value$<0.05/11 = 0.0045$ was considered significant. Surprisingly, a region of transcriptomic up-regulation in tumor compared to normal was found between 13q12.3 (KATNAL1) and 13q12.3-q13 (CCNA1). Significant mean shifts were observed between EFHA1 and TNFRSF19 (up, P-value = 3.60E-03), RNF6 and KATNAL1 (down, P-value = 4.19E-03), RFC3 and CCNA1 (down, P-value = 1.24E-04) and CCNA1 and ELF1 (up, P-value = 0.19E-06). The mean shifts of RNF6/KATNAL1 and CCNA1/ELF1 demarcate the region of up-regulation. CCNA1 presented a 4+ up-regulated tumor mean fold change (71119.84 fold). Western blot analysis with CCNA1 MAb of two cases has putatively validated the expression patterns observed by qRT-PCR. CCNA1 is currently undergoing genomic validation analysis. These data demonstrate the discovery of a putative 13q clinical biomarker derived from microdissected samples of ESCC.

Array comparative genomic hybridization analysis reveals genomic aberrations in isolated circulating tumor cells (CTCs)

Mark Jesus Magbanua, Lauren Eisenbud, Janet Scott, Dan Pinkel, Michelle Melisko, Hope Rugo, Joe Gray, Laura Esserman* and John Park

**Corresponding author: University of California, San Francisco, CA, USA.*

Background: Molecular characterization of CTCs is technically challenging because they are rare and difficult to isolate. In addition, the picogram levels of genomic DNA (gDNA) isolated from these cells further confounds genomic profiling. Here we report a method for isolation of pure CTCs and amplification of gDNA for array comparative genomic hybridization (CGH) analysis. Our results confirm presence of genomic aberrations in isolated CTCs.

Methods: BT474 cells from spike-in experiments and CTCs from blood, bone marrow and cerebrospinal fluid of metastatic breast cancer patients were isolated via immunomagnetic enrichment followed by fluorescence activated cell sorting (IE/FC). Isolated gDNA was amplified using commercially available ligation-mediated PCR-based protocol. Amplified DNA and a reference male genomic DNA were labeled and co-hybridized to BAC arrays.

Results: Array CGH profiling of spiked BT474 cells revealed characteristic high-level gene amplifications at chromosome 17 (Her-2) and chromosome 20 (ZNF217). Pair-wise Pearson correlation analyses between unamplified BT474 gDNA and amplified gDNA from 20-500 BT474 cells showed an $R2>0.89$. Furthermore, CTCs isolated from blood, bone marrow and cerebrospinal fluid exhibited a wide range of genomic aberrations including gene amplifications, deletions and copy number changes.

Conclusion: We have developed a method to isolate CTCs and profile them via aCGH microarray analysis. Amplification of minute amounts of gDNA produced unbiased representation of the genome. Uncovering the genomic aberrations, such as gene amplifications, in CTCs may ultimately lead to the discovery of druggable biomarkers that will allow for specific targeting of these cells in patients who respond poorly to current aggressive therapies.

Defining new multi-gene classifiers and prognostic indicators for hormone receptor and triple negative breast cancers

Christina Yau, Dan Moore, Eugene Fan, Bill Chu, Marc Lenburg, John Sninsky, Fred Waldman, Laura Esserman*, John Park and Chris Benz

**Corresponding author: University of California, San Francisco, CA, USA.*

Background: While prognostic breast cancer gene expression profiles have been introduced into the clinic, to date there are no validated prognostic gene signatures identified for the clinically problematic subsets of

hormone receptor-negative (HRneg: ER, PR negative) and triple-negative (Tneg: ER, PR, HER2 negative) primary breast cancers at highest risk for early metastatic relapse. We developed multi-gene classifiers to improve the clinical management of node-negative (N0) HRneg and Tneg breast cancer patients by identifying those at highest and lowest risk for metastatic relapse independent of therapy.

Approach: A training set of 135 untreated, N0, HRneg primary breast cancers (108 identified as Tneg) was identified from published studies that used the Affymetrix U133A microarray platform (Wang et al., 2005, GSE2034; Minn et al., 2007, GSE5327). Candidate genes associated with metastasis-free survival were identified in the training set using PAM, iterative sampling, and multivariate Cox modeling, and were subsequently assigned into hierarchical prioritizations based on their biostatistical evaluation against another untreated N0 dataset of 64 HRneg and 46 Tneg cases similarly analyzed using the Affymetrix platform (TRANSBIG; Desmedt et al., GSE7390). High priority candidates were further validated against 37 HRneg NKI cases (van de Vijver et al., 2002) analyzed by an Agilent microarray platform.

Results: 18 HRneg and 10 Tneg genes were identified as prognostic candidates in the training set. These candidates were better able to predict metastasis-free survival as a summation index than as single gene predictors. 11/18 HRneg and 6/10 Tneg candidates were assigned higher priority following univariate Cox analysis against the TRANSBIG dataset, and 8/18 HRneg and 5/10 Tneg candidates were assigned highest priority following multivariate Cox modeling. The 6/11 higher priority HRneg candidates that are available on the Agilent platform showed significant prognostic value as a summation index in the NKI validation set.

Conclusions: Our prioritized HRneg and Tneg gene signatures show significant prognostic value. Prospective testing in new, untreated data sets with 25-year follow-up will further validate the expression signature.

Single nucleotide polymorphisms that modify breast cancer risk and disease phenotype: A validation study in a high-risk population

Yiwey Shieh, Salina Chan, Matthew R. Nelson, Dan H. Moore, Gareth L. Bond, Arnold J. Levine, Cecile-Rose Vibat, Charles R. Cantor, Jeffrey R. Marks, John W. Park, John L. Ziegler and Laura J. Esserman*

**Corresponding author: University of California, San Francisco, CA, USA.*

Background

Discovery efforts at Sequenom, Inc. have identified possible associations between four germ-line single nucleotide polymorphisms (SNPs) and breast cancer susceptibility. A model estimating the combined effect of these SNPs predicted that 4-SNP genotype combinations exist in the population at the same frequency, and confer comparable risk modification levels, as *BRCA1* and *BRCA2* mutations. The objective of our study is to validate the individual associations between the above SNPs and disease risk/phenotype, as well as the findings of the Sequenom 4-SNP model. In addition, we will genotype our study population for SNPs that have been associated with clinicopathologic features such as age of onset, nodal metastasis, and estrogen receptor positivity.

Methods

Study population: Participants were recruited from patients undergoing diagnostic workup for a mammographic abnormality, surgical treatment for invasive breast cancer, or genetic counseling for family history of breast cancer. At this time, we have accrued 1250 cases, women with Stage I-III breast cancer and no other active malignancies. The control group consists of 1000 women who have never been diagnosed with breast cancer.

SNP genotyping: DNA extracted from buffy coats will be sent to Sequenom, Inc. for genotyping using a multiplex MassARRAY system. A panel of 64 candidate SNPs will be assessed for presence of allelic variants.

Statistical analysis: A chi-squared test will be used to confirm Hardy-Weinberg equilibrium and calculate the p-value of associations between disease status and genotype. For each allele and genotype, the odds ratio of developing disease will be calculated and its mode of function will be determined by modeling. Logistic regression will be used to adjust for the effect of covariates including age of diagnosis and race. To investigate the effect of SNP genotypes on disease phenotype, a case-only association study will be conducted.

Proliferating macrophages predict worse outcome in breast cancer patients in US and West Africa using immunohistochemistry and peripheral blood analysis

A. Lin, L. Garwood, A. Kumar, D. Moore, A. Au, F. Baehner, M. Campbell, O. Olopade, M. McGrath and L. Esserman*

Corresponding author: University of California, San Francisco, CA, USA.

Background: Macrophages (M∅) may play a vital role in the progression and metastasis of some tumors. Tumor cells and M∅ interact through a CSF-1 paracrine signaling loop resulting in invasion of blood vessels and penetration of the extracellular matrix. We have found proliferating macrophages (*pm*) to be abundant in clinically palpable DCIS, and used retrospective data sets to determine their presence in invasive breast cancer (BC), and a prospective data set to determine the presence of M∅ in peripheral blood (PB).

Methods: 110 cases were selected from UCSF tumor bank and 43 cases were from West Africa (WA). Slides were double-stained for *pm* markers, CD 68 and PCNA using IHC techniques. A subset of pts (66) had tissue available for MAC387 staining. Stage, lymph node, grade, hormone status, and treatment modalities were assessed. Prospectively, a battery of monocyte markers including CSF-1R+ was used to evaluate M∅ from PB of 44 pts with stage I-III BC using FACS.

Results: Frequency of *pm* and MAC387+ M∅ correlate with poor outcome in patients with long-term follow-up. These markers are more prevalent in high grade disease, 54% with *pm* $\geqslant$ 5 per HPF versus 34% in low grade disease ($p = 0.04$) and in tumor from WA, where the tumors are predominantly ER negative (78%). Double staining with PCNA and CD 68 is not reliable or practical due to variable staining of PCNA and time-consuming aspect of counting PCNA+/CD68+ M∅ manually. However, MAC 387 is simpler. We are in the process of evaluating all candidate M∅ markers: CD 68, MAC 387, CSF-1R and CD11a.

Prospectively, we are assessing M∅ markers in the peripheral blood (PB). CSF-1R+ monocytes seem to best differentiate patients with invasive tumor, where the mean was 24.4% of total monocytes (range 0–85%). Higher levels of CSF-1R+ monocytes correlated with higher estimated 10 yr recurrence and 10 yr mortality ($p = 0.049$, $p = 0.004$, respectively) using Adjuvant!Online® program. The mean CSF-1R values were significantly lower for the control group, 7.9% in pts without invasive cancer ($p = 0.003$).

Conclusions: Proliferating macrophages are associated with higher recurrence and worse survival in BC and may be a target for therapy and early detection. We are prospectively evaluating and correlating levels of CSF-1R M∅ in both PB and tissue.

Enrichment of bronchial epithelial cells from sputum for lung cancer early detection[1]

Feng Jiang[a,*], Qi Qiu[a], Nevins W. Todd[b], Ruiyun Li[c], Zhenqiu Liu[d], Harris G. Yfantis[e], William H. Rodgers[a] and Sanford A. Stass[a]

[a]*Department of Pathology, University of Maryland School of Medicine, Baltimore, MD, USA*

[b]*Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA*

[c]*Department of Surgery, University of Maryland School of Medicine, Baltimore, MD, USA*

[d]*Department of Epidemiology and Preventive Medicine and Division of Biostatistics of The University of Maryland Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, MD, USA*

[e]*Department of Anatomic and Clinical Pathology, Baltimore VA Medical Health Care System, Baltimore, MD, USA*

Corresponding author.

We demonstrated in sputum using fluorescence *in situ* hybridization (FISH), that deletion of *FHIT*, a tumor suppressor gene, could be of value for detection of early lung cancer. Use of sputum is limited by its cellular heterogeneity with typically more than 95% macrophages and neutrophils and approximately 1% bronchial epithelial cells. In this study, bronchial epithelial cells were first enriched from sputum of 29 stage I lung cancer patients, 26 cancer-free smokers, and 28 healthy nonsmokers using magnetic-assisted cell sorting with antibody-coated beads. FISH analysis for detection of *FHIT* deletion and cytology were evaluated in the enriched specimens. The bronchial epithelial cells were concentrated to 40% purity from 1.1% of the start-

[1]This work was supported in part by Associate Membership Award from NCI-The Early Detection Research Network (EDRN) to Dr. Feng Jiang who is sponsored by Dr. Sanford A. Stass, an EDRN Principal Investigator.

ing population, yielding at least $2.3X10^5$ cells per sample. Detecting *FHIT* deletions by FISH for lung cancer diagnosis produced 58% sensitivity in the enriched sputum and 42% sensitivity in the unenriched samples ($P = 0.02$). Cytological examination of enriched sputum produced 53% sensitivity compared with 39% sensitivity in unenriched sputum ($P = 0.03$). Fewer slides were needed for the enriched sputum, compared to the unenriched sputum (2 versus 10). In conclusion, the enrichment of bronchial epithelial cells improves the diagnostic value of FISH and cytological analysis of sputum for detection of lung cancer. Future development of a probe panel using enriched sputum bronchial epithelial cells could be useful for early detection of lung cancer.

Familial hematologic cancers: Multiple myeloma and chronic lymphocytic leukemia

Henry T. Lynch*, Patrice Watson, Kelly Ferrara, Bart Barlogie, Elizabeth A. Coleman, Jane F. Lynch, Dennis Weisenburger, Warren Sanger, Henry Nipper, Vinetta Witt, Peter H. Wiernik and Stephan Thomé

**Corresponding author: Creighton University.*

Background: Multiple myeloma (MM) accounts for approximately 10% of all hematologic cancers and occurs frequently among the elderly, with only 2% of patients younger than 40 years of age [1–3]. The incidence of MM worldwide in 2002 was estimated to be 85,700 cases, with an annual mortality of 62,535 [4]. In the United States, the incidence for 2007 was estimated at 19,900 cases with a mortality of 10,790 [5]. MM accounts for approximately 1 percent of all forms of cancer, but approximately 2% of all cancer deaths [2]. Unfortunately, there are currently no primary preventive measures for this disease [3].

Our research plan has focused heavily upon extended families with putative integral hematologic and solid cancers of all anatomic sites. Pathology is verified whenever possible in accord with our hypothesis of phenotypic and genotypic heterogeneity in familial multiple myeloma (MM) and chronic lymphocytic leukemia (CLL). The pattern of tumor combinations within and between these families is investigated. The following five multiple myeloma pedigrees appear to reflect this heterogeneity. Note, for example, in Family 1, we describe the occurrence of MGUS, myeloma, and an increased frequency of carcinoma of the prostate.

In the family with CLL, we describe an autosomal dominant inherited pattern of CLL through three generations where the CLL appears to have an indolent survival pattern and wherein the downregulation of death associated protein kinase 1 (*DAPK1*) has segregated with CLL in the family [6].

We now have 109 families in our hematology resource. The most informative families are being considered for formal linkage analysis.

Summary and conclusions: Thanks to intense ongoing progress in molecular genetics, advances in cancer etiology, surveillance, and therapeutics through the identification of molecular-based designer drugs, have given further impetus to the importance of hereditary cancer studies. We believe that a subset of these hematologic families, inclusive of those with multiple myeloma and CLL, which we are investigating, will harbor DNA which hopefully will prove to be a source for biomarker advances, as well as, new drug development.

References

[1] J.S. Malpas, D.E. Bergsagel, R. Kyle and K. Anderson, *Multiple Myeloma: Biology and Management*, Oxford: Oxford University Press, 1998.
[2] G.J. Morgan, F.E. Davies and M. Linet, Myeloma aetiology and epidemiology, *Biomed Pharmacother* **56** (2002), 223–234.
[3] D.D. Alexander, P.J. Mink, H.-O. Adami et al., Multiple myeloma: A review of the epidemiologic literature, *Int J Cancer* **120**(Suppl 12) (2007), 40–61.
[4] International Agency for Research on Cancer, Globocan 2002, International Agency for Research on Cancer, 2004, 10-16-2006.
[5] A. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu and M.J. Thun, Cancer statistics, *CA Cancer J Clin* **57** (2007), 43–66.
[6] A. Raval, S.M. Tanner, J.C. Byrd et al., Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia, *Cell* **129** (2007), 879–890.

Glycans in the progression of HCV infection to hepatocellular carcinoma

R. Goldman[a,*], H.W. Ressom[a], S.A. Varghese[a], G. Bascug, C.A. Loffredo[a], M. Abdel-Hamid[b], Z. Kyselova[c], Y. Mechref[c] and M.V. Novotny[c]

[a]*Department of Oncology, Georgetown University, Washington, DC, USA*

[b]*National Hepatology and Tropical Medicine Research Institute, Cairo, Egypt*

[c]*Department of Chemistry, University of Indiana, Bloomington, IN, USA*

*Corresponding author.

Increasing incidence of hepatocellular carcinoma (HCC) in the US was associated with hepatitis C viral (HCV) infection. Our study of HCC in Egypt, a country with an epidemic of HCV, evaluated changes in glycans in the progression of HCV infection to cancer. Glycans enzymatically released from serum proteins were analyzed by MALDI-TOF/TOF following solid phase permethylation. Analysis of less than 0.02 ml of serum allowed relative quantification of 80 glycans. Even though some glycans changed with gender or age of the subjects, larger changes were associated with HCV infection. Further changes were observed in patients with chronic liver disease and HCC. Approximately half of the glycans changed significantly at $p <$ 0.01 in HCC patients ($n = 73$) compared to control with chronic liver disease ($n = 52$). Individual glycans classified HCC with accuracy comparable to alpha fetoprotein. A set of 3 glycans was sufficient to classify HCC with 90 % prediction accuracy in a blinded set of samples comparing cirrhosis and early stage HCC ($n = 43$). The changes of glycans with the progression of liver disease differ by structural glycan classes. In conclusion, our results show that mass spectrometric analysis of N-glycans in serum is useful in tracking the natural progression of HCV infection to HCC.

This work was supported in part by an Associate Membership from NCI's Early Detection Research Network and NIH grant R01CA115625-01A2 awarded to RG.

Identification of lung cancer cell surface proteins as potential biomarkers in the serum proteome

Vitor Faca, Lisa Newcomb, Veronika Glukhova, Doug Phanstiel, Alexei Krasnoselsky, Qing Zhang, Jason Struthers, Hong Wang and Samir Hanash*
Fred Hutchinson Cancer Research Center, Seattle, WA, USA

*Corresponding author.

Cell surface proteins play very important roles in many cellular processes, especially during cancer progression. A substantial proportion of proteins detected in plasma or serum represent membrane or surface associated proteins (>30%), making this class of proteins a source of disease biomarkers. In the present work, we compare proteins identified on the surface of lung cancer cells and proteins identified by in-depth analysis of serum from cancer patients and controls. Cell surface profiling was done for adherent and viable A549 cells using a biotinylation-capture approach (Shin et al., J. Biol. Chem. 278:7607–7616, 2003). Briefly, biotinylated proteins were isolated by immobilized neutravidin affinity chromatography, which provided 1% recovery from total protein extracts. Surface proteins were fractionated by reversed-phase chromatography followed by tryptic digestion and mass spectrometry identification by LC-MS/MS in a LTQ-FTICR mass spectrometer. A total of 990 proteins were confidently identified with FDR >2%. More than 40% of these proteins were annotated as membrane proteins. Lung cancer serum analysis was performed with a mixture of immunodepleted control and lung cancer serum samples differentially labeled with acrylamide isotopes, to obtain relative quantitation. 144 fractions were collected using an intact protein-based chromatographic approach. Individual fractions were analyzed by shotgun LC-MS/MS and a total of 1197 proteins across a wide dynamic range of protein concentrations were confidentially identified. Quantitative analysis indicated that 154 proteins were found up regulated in diseased serum. Correlation of the A549 lung cancer cell surface profile and the lung cancer serum profile resulted in 265 proteins identified in both datasets. Among these proteins, 100 had associated relative quantitation and 35 of them were up-regulated (<1.5 fold) in lung cancer patient serum compared to controls. Some of these proteins have already been described as up regulated at mRNA or protein level in human lung cancer tissue, supporting our findings. This panel of proteins will be the subject of further investigation and validation. With this combined profiling approach and in-depth serum and cell surface protein analysis we were able to identify and select candidate biomarkers relevant to cancer from an extensive list of proteins.

Baseline model to serve as a platform for evaluation of potential biomarkers for early detection and diagnosis of lung cancer

William Hazelton
Fred Hutchinson Cancer Research Center, Seattle, WA, USA

The goal of this project is to develop a baseline model to serve as a platform for evaluation of potential biomarkers for early detection and diagnosis of lung

cancer. Initial calibration is to 18,314 individuals from the Beta-Carotene and Retinal Efficacy Trial (CARET) cohort. I recently received access to CARET data, and have begun joint calibration to incidence and mortality for small- and non-small cell lung cancer using detailed histories of smoking and asbestos exposures (and beta-carotene and retinol in the treatment arm). The model has two pathways representing small- and non-small cell lung cancer. Each pathway has four stages that represent the joint distribution of normal, pre-malignant, malignant, and metastatic cells in each individual. Initial results suggest that smoking has different effects on clonal expansion and mutation rates for small- versus non-small-cell lung cancers. I plan further calibration to 1,268 heavy smokers and asbestos exposed workers with low-dose helical Computerized Tomography (CT) screening in the New York University (NYU) cohort, and 1,151 current and former heavy smokers in the Moffitt CT screening trial (MCTST). The lung cancer model will be validated on 3,755 current and former smokers in the Pittsburgh Lung Screening Study (PLuSS) cohort. The modeling should help clarify if the apparent high survival rates reported for CT arise from over-diagnosis of indolent nodules, from lead-time and length bias, or represent true mortality gains, or some combination of these explanations. This baseline model should be useful to evaluate the potential added benefit of biomarkers for early detection.

Molecular analysis of serum/plasma DNA for the detection of lung cancer

Mohammad Obaidul Hoque [a,*], Mariana Brait [a], Shahnaz Begum[b], Santanu Dasgupta[a], Marianna Zahurak[c], André L. Carvalho [a], Myoung Sook kim[a], Joseph A. Califano[a], William H. Westra[a,b] and David Sidransky[a]
[a] *Department of Otolaryngology-Head and Neck Surgery, The Johns Hopkins School of Medicine, Baltimore, MD, USA*
[b] *Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, MD, USA*
[c] *Department of Biostatistics/Oncology, The Johns Hopkins School of Medicine, Baltimore, MD, USA*

[*] *Corresponding author.*

Aberrant promoter hypermethylation of several known or putative tumor suppressor genes occurs frequently during the pathogenesis of lung cancers and is a promising marker for cancer detection. To determine the analytical sensitivity, we examined the tumor and the matched serum/plasma DNA for aberrant methylation of fifteen gene promoters from 10 patients with primary lung tumors by using quantitative methylation specific PCR (QMSP), a high-throughput DNA methylation assay. Additionally, we used a gene evaluation set to identify the most important DNA methylation changes in the serum of a limited number of lung cancer patients. Serum/plasma DNA from age-matched non-cancer patients was used as a control. In an independent set, we tested the six most promising genes (*APC, CDH1, MGMT, DCC, Rassf1A* and *AIM*) for further elucidation of the diagnostic application of this panel of markers.

In majority of cases, aberrant methylation in serum/plasma DNA was accompanied by methylation in the matched tumor samples. In the gene evaluation set, promoter hypermethylation in serum/plasma DNA from lung cancer patients was found in: 9/25(36%) *APC*, 4/17(24%) *AIM*, 4/17(24%) *CyclinD2*, 12/17(71%) *CALCA*, 10/17(59%) *CDH1*, 3/17(18%) *DCC*, 0/25(0%) p16, 5/25(20%) *MGMT*, 2/25(8%) *Rassf1A*, 0/20(0%) *MINT31*, 2/20(10%) *CyclinA1*, 1/20(5%) *ESR1*, 10/20(50%), 3/20(15%) *PGP9.5* and 10/20(50%) *TIMP3* with different specificity. By using 6 most promising genes high sensitivity and specificity was obtained in an independent set of samples. This approach needs to be evaluated in a larger test set to determine the role of this approach in early detection and surveillance of lung cancer.

High-throughput biomarker validation by automated magnetic bead-based chromatography and quantitative high-resolution MALDI-TOF-MS

Kazufumi Honda[*] and Tesshi Yamada
Chemotherapy Division and Cancer Proteomics Project, National Cancer Center Research Institute, Tokyo, Japan

[*] *Corresponding author.*

In order to improve the outcome of patients with pancreatic cancer, identification of a reliable serum/plasma biomarker that can be used for early detection of the disease would be highly advantageous. We previously identified a plasma biomarker set that we able to distinguish pancreatic cancer patients from healthy controls with high accuracy (Honda et al., Cancer Res., 65:10613–10622, 2005). Here we report the results of

an ongoing validation study using 332 plasma samples [pancreatic cancer ($n = 165$), benign pancreatic disorders ($n = 11$), and healthy controls ($n = 156$)] collected at two medical institutions. All the samples were blinded and processed randomly using a robotic magnetic bead-based hydrophobic interaction chromatography system. We confirmed that the intensities of two previously identified biomarker proteins (17252 m/z and 8765 m/z) differed between pancreatic cancer patients and healthy controls with high statistical significance [Mann-Whitney U test, P <1.0×10E-10 and AUC (area under curve) value >0.801]. In order to confirm their clinical utility, a collaborative study involving 8 medical institutions throughout Japan is now underway. So far, more than 800 serum/plasma samples of cancer patients and controls have been collected prospectively. We will report the preliminary result of the on-going multi-institutional validation study.

This collaborative study is part of the "Third-Term Comprehensive Control Research for Cancer" conducted by the Ministry of Health, Labor and Welfare of Japan.

Borrowing information across subpopulations in estimating positive and negative predictive values

Ying Huang* and Ziding Feng
Fred Hutchinson Cancer Research Center, Seattle, WA, USA

**Corresponding author.*

In biomarker research, a good study design with optimal utilization of information can lead to great savings of time and costs. In this work, we are interested in designing a study to validate a marker's risk prediction capability in terms of positive predictive value (PPV) and negative predictive value (NPV) in various subpopulations when the marker's classification accuracy as characterized by the ROC curve is invariant across subpopulations. A default strategy is to estimate PPV and NPV in different subpopulations separately. Here we propose estimators that incorporate the constant ROC curve assumption and borrow information across subpopulations. The new estimators are shown to be asymptotically more efficient than ordinary nonparametric estimators. Simulation studies are performed as well to assess these estimators' finite sample performance. We illustrate this methodology in a real dataset by evaluating PCA3 as a risk prediction marker for prostate cancer among subjects with or without initial biopsy.

An information model for biomarker research

Steve Hughes[a], Deanna Stelling[b], Greg Warnick[b], Dan Crichton[a,*], Chris Mattmann[a], Suzanna Reid[b], Sean Kelly[a], Don Johnsey[c]
[a]*JPL*
[b]*DMCC*
[c]*NCI*

[*]*Corresponding author.*

The Information Model is the foundation on which an Information System is built. It defines the entities to be processed, their attributes, and the relationships that add meaning. The development and subsequent management of the Information Model is the single most significant factor for the development of a successful information system.

The mission of the Early Detection Research Network (EDRN), a program managed by the National Cancer Institute, is the discovery and validation of biomarkers. In partnership with NASA's Jet Propulsion Laboratory (JPL) the program has developed the EDRN Knowledge Environment to provide integrated access to data across the EDRN enterprise that is captured during the biomarker discovery process. The infrastructure is a set of information system services that allow distributed data repositories to be integrated and accessed real-time. At the heart of the EDRN Knowledge Environment is an information model that describes the data entities involved in biomarker research and how they relate. The scope of this model ranges from tissue banking, to biomarker information, to validation studies, and scientific results. The information model integrates the data entities and enables researchers to navigate through the data repositories in the EDRN community using state-of-the-art search mechanisms.

Over the past year, the EDRN informatics team has made tremendous progress in developing the EDRN Information Model. Using a framework of data modeling tools, the information model has been captured from three sources, the EDRN registry of Common Data Elements (CDE), information specialists familiar with study design and processes, and the working subsystems. The model was captured in an ontology modeling tool and a specification document has been

written with the model expressed using object-oriented class notation. The information model continues to be refined in concert with systems development.

The goal of the information model is to unify biospecimen data, scientific data, study specific data, and biomarker data for the EDRN Knowledge Environment and enable the EDRN public portal to provide access to information in data repositories throughout the EDRN community.

Identifying molecular markers in non-small cell lung cancer using computer-aided scoring and analysis (CASA)

Abbas Shakoori[a,b], Junya Fukuoka[a,b], Tatiana Dracheva[a,b], Joanna H. Shih[c], Konstantin Shilo[e], Uchenna Nwosu[a], Henry Zhang[b], Rajbir Gill[a], Hyosung Jeon[a], Robert Clifford[b], Teri J.Franks[e], Stephen Hewitt[d], William Travis[f] and Jin Jen[a,*]

[a]*Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA*

[b]*Lab of Population Genetics, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA*

[c]*Biometrics Research Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA*

[d]*Tissue Array Research Program, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA*

[e]*Department of Pulmonary & Mediastinal Pathology, Armed Forces Institute of Pathology, Washington, DC, USA*

[f]*Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA*

[*]*Corresponding author.*

Immunohistochemical analysis of tissue microarray (TMA) has enabled simultaneous examination of various antigens for a large number of clinical samples. However, manual evaluation of IHC results on TMA is time consuming and subjective. To overcome these challenges, we developed a computer-aided scoring and analysis (CASA) method to allow facile and reliable scoring of immunohistochemistry (IHC) staining in TMA samples. We applied CASA analysis using a lung cancer TMA containing 150 adenocarcinoma (AD) and 150 squamous cell carcinoma (SCC) cases. After IHC staining using optimized conditions, TMA slides were digitally scanned, dearrayed, and then analyzed based on adjusted intensity values. After normalization by size and overall intensity, association of CASA scores for each antibody were correlated with associated clinical data including age, gender, differentiation, pathologic stage, histology types. A total of 87 antibodies were examined and predictive values to survival were examined using Cox-multivariate analysis for factors that belonged to the same molecular pathway or have shared expression patterns.

Based on CASA scores, we observed several protein markers with significant correlation with stage and histology in lung cancer ($p < 0.01$). Also, using pairwised analysis using the CASA scores identified five combination of protein expression patterns that was significantly correlated with overall survival in non-small cell lung cancer patients ($p < 0.01$). These proteins included TGFBR2-BAF170, cMyc-p16, MSIN3A-p16, cMyc-MSIN3A and HER2-MASPIN. The overall survival remained significant when the status of these protein combinations were examined using manual scores of IHC staining for the same factors and after stratification to stage and tissue type ($P < 0.05$).

Our study demonstrates that computer-aided scoring analysis is comparable to manually scoring and can be potentially used as a tool for robust and systematic evaluation of IHC staining involving TMA and for the identification of clinically significant molecular markers.

A functional genomic approach to biomarker discovery for the early detection of pancreatic cancer

Ann Killary[*], Marsha Frazier and Subrata Sen

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

[*]*Corresponding author.*

Our biomarker discovery laboratory associated with the NCI Early Detection Research Network is taking a functional genomic approach to discover biomarkers for the early detection of pancreatic cancer by targeting a tumor suppressor pathway associated with loss of chromosome 3p, and an oncogenic pathway associated with amplification of chromosome 20q. Using functional complementation assays, we previously identified a chromosome 3p12 tumor suppressor locus proximal to the most common fragile site in the human genome, and we hypothesize that loss of this interval is an initiating event in a cytogenetic pathway associated

with smoking related tumors. In order to identify a chromosome 3p pathway, 880 partial cDNAs were obtained from a suppression subtractive hybridization library (SSH) constructed to identify chromosome 3p12 pathway genes. cDNAs differentially expressed from the SSH library were compared to expression profiles from an Affymetrix GeneChip array interrogated with RNAs derived from the same starting materials used for SSH library construction to identify a subset of genes across these two platforms that were differentially expressed. Affymetrix arrays were then interrogated with tumor/normal pancreatic samples to further stratify genes based on a third expression platform. Results have identified 8 genes that are consistently differentially expressed across the three expression platforms and verified by quantitative RT-PCR. Two/eight genes (KSF-1 and KSF-2) have been shown to be upregulated and secreted in pancreatic cancers. One of the three genes identified (KSF-3) is expressed abundantly only in the normal pancreas and downregulated in pancreatic cancers. Genetic profiling of the 20q amplicon in four well characterized pancreatic cancer cell lines (BxPC-3, Capan-2, MIA PaCa-2, PANC-1) utilizing Agilent CGH and expression microarray platforms identified a list of candidates which were overexpressed and amplified and mapped into 20q. This list was narrowed by screening expression arrays from pancreatic normal/tumor samples and candidates verified by quantitative RT-PCR Candidates are being screened for differential expression in a large collection of pancreatic tumor/normal samples as well as plasma samples from pancreatic cancer, pancreatitis and normal individuals. Additionally, we have examined associations between single nucleotide polymorphisms (SNPs) in 167 patients with pancreatic cancer. SNPs in the Aurora Kinase A gene predicted early onset pancreatic cancer and in combination with p16 polymorphism, increased the risk of pancreatic cancer in a multiplicative manner. Patients with mutant genotypes for both genes had an approximately 3-fold increased age-associated risk for diagnosis of pancreatic cancer compared with wild-type genotypes. The overall goal of this BDL is to develop a panel of biomarkers for the early detection and risk assessment of pancreatic cancer.

Biomarker results variability

James T. Snyder II[a], Randall Grimes[b], Robert V. Woods[a], Yu Min Shi[a], Ping Su[a], Donald Weber,[b], Emelie G. Gunter[a], Steven A. Chase[a] and Louis Kirby[b,*]

[a]*Provista Diagnostics, a division of Provista Life Sciences, Phoenix, AZ, USA*
[b]*Provista Life Sciences, Phoenix, AZ, USA*

[*]*Corresponding author: Dr. Kirby Provista Life Sciences, 6225 N. 24th Street, Suite 150, Phoenix, AZ, USA.*

Background: In the process of refining a serum biomarker panel for breast cancer, we encountered significant variability in biomarker assay results. Investigation of the causes yielded a number of unexpected parameters that adversely affected the reproducibility of the results.

Methods and results: Nine biomarkers were investigated in a clinical protocol for breast cancer. Early results showed considerable run to run variability. Analysis demonstrated four factors principally responsible for this observation: 1) Lot to lot variability from same supplier was up to $>$ 50%. 2) intra-assay variability (Drift across a 96 well plate) yielded a consistent variance of $>$30% from the first well to the last. 3) multiplexed biomarker panels yielded a determinant shift in results compared to singlet reagent panels (possible enhancements or interference) and 4) variability between different vendors of assay kits for the same biomarker demonstrated up to a 10-fold difference and some assay kits were unable to separate disease from controls. Differing strategies were utilized to address each factor to improve reproducibility.

Conclusions: Biomarker analysis results were adversely affected by unexpected parameters that may be due in part to the difference among vendors' development processes and to their differing methods for quality control in a research setting compared to that required for clinical use. Application of appropriate quality control measures for our specific application resolved these adequately to achieve reproducibility in a high-throughput laboratory setting.

Retrospective study of select biomarkers for breast cancer screening

Louis Kirby[a,*], Linda Greer[b], Christa Corn[c], Caroline Hardy[a], Deborah S. Gelbspan[d], Randall Grimes[a], Robert V. Woods[e], Yu Min Shi[e], Ping Su[e], Donald Weber[a] and James T. Snyder II[e]
[a]*Provista Life Sciences*
[b]*SimonMed Medical Imaging*
[c]*PLLC*

[d]*Surgical Pathology Consultants*
[e]*Provista Diagnostics, a division of Provista Life Sciences*

[*]*Corresponding author: Dr. Kirby, Provista Life Sciences 6225 N. 24th Street, Suite 150, Phoenix, AZ 85016, USA.*

Background: Studies on cancer biomarkers have reported an ability to identify the presence of breast cancer in small and retrospective studies. We explored a breast cancer detection test utilizing an array of protein biomarkers in an attempt to define a reliably reproducible model.

Methods: 376 women, 35 to 75 years old, were enrolled treatment naive from community breast biopsy ($n = 244$) and screening mammography referrals ($N = 132$). Biopsy patients were further divided into breast cancer positive ($n = 35$) and benign breast conditions ($n = 209$) groups based on subsequent tissue pathology. Serum was analyzed for IL-2, IL6, IL8, IL12, TNF alpha, HGF, EGF, FGF, and VEGF utilizing either bead-based multiplex or ELISA methodology. Marker data were combined with each subject's demographic and clinical information and analyzed with a proprietary iterative regression analysis.

Results: A subset of the original biomarkers was shown to be statistically significant ($P < 0.05$) and used to create the final score predictive of the presence or absence of breast cancer. This score yielded an AUROC of 0.830 with a specified sensitivity of 63% and specificity of 86%.

Conclusion: These findings demonstrate in a clinical trial the ability of a select set of biomarkers to correlate with the presence or absence of breast cancer in an enriched population of patients. A prospective clinical trial is underway to validate the numbers presented here. Additional trials of differing study design are warranted to quantify the clinical utility in an unselected population.

Generation of an antibody microarray for the early detection ovarian cancer markers

Rachel Kravitz[*], Deven McGlenn, Patricia Nason, Todd Meyer, Nick Wiegert, Jack Miller and Uwe Müller
NeoClone Biotechnology, Madison, WI, USA

[*]*Corresponding author.*

Neoclone has a Phase II SBIR contract to develop a high-throughput monoclonal antibody microarray for the diagnosis of early stage ovarian cancers. These arrays will be produced using the glass slide based platform of GenTel Biosciences, and after development and validation of the assay system with fluorescence detection these arrays will be provided to EDRN investigators for testing with human samples.

NeoClone Biotechnology develops monoclonal antibodies (mAbs) using a novel ABL-MYC retroviral technology. Our process targets antigen-specific B cells and transforms them into stable plasmacytomas, secreting high levels of immunoglobulins. We can rapidly and efficiently produce large numbers of antigen-specific antibody producing cell lines. This process is ideal for the generation of antibodies to be used in diagnostic applications where high clonal diversity is important for selecting antibodies with specific characteristics, such as high affinity and specificity as well as stability and compatibility with different detection platforms. NeoClone is in the process of developing a panel of mococlonal antibodies to several ovarian cancer markers, including CA-125, CA 15-3, CA 19-9 and CA 72-4. In collaboration with GenTel Biosciences antibodies specific to a minimum of 16 cancer markers will be printed in a microarray for multiplex detection of these biomarkers in human serum. Preliminary data suggest that such microarrays will detect these biomarkers with a linear dose response over at least two logs of target concentration.

An efficient algorithm for biomarker identification

Jiang Li[a,*], Rick McKenzie[a], Lisa Cazares[b], Richard Drake[b] and John Semmens[b]
[a]*Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, USA*
[b]*George L. Wright Jr., Center for Biomedical Proteomics, Department of Pathology and Anatomy, Eastern Virginia Medical School, Norfolk, VA, USA*

[*]*Corresponding author.*

One challenge for biomarker identification is how to handle high dimensional mass spectral data. Various feature selection algorithms have been applied, including the Wilcoxon test, Area under the ROC curve, Fisher score, J5 test, Simple separability criterion,t-test score, Weighted separability criterion, principal component analysis (PCA), Wavelet-based algorithms and

the genetic algorithm, to name a few. However, each of them has its own limitations. We present an efficient feature selection algorithm, recently developed by the authors, for the biomarker identification task. Advantages of the proposed algorithm are as follows. 1) It selects peaks rather than a combination of all available peaks such as those selected by transformation based methods (PCA, Wavelet), 2) It considers interactions among peaks and measures the correlations in terms of amount of explained variances by the peaks, 3) It is computationally efficient, 4) It automatically handles extremely unbalanced data sets where the number of instances in some classes are significantly more than those in other classes, and 5) The algorithm produces a list of near-optimal combinations for all possible number of peaks with sensitivity and specificity calculated for each of the combinations. Users can then choose the best peak combination based on its sensitivity and specificity. We applied the proposed algorithm to MALDI-MSI mass spectra to identify biomarkers for prostate cancer and the effectiveness of the proposed method is clearly demonstrated.

Autoantibody approach for serum-based detection of head and neck cancer

Ho-Sheng Lin
Wayne State University and Karmanos Cancer Institute, Detroit, MI, USA

Currently, no effective tool exists for screening or early diagnosis of head and neck squamous cell carcinoma (HNSCC). Here we describe an approach for cancer detection based on analysis of patterns of serum immunoreactivity against a panel of biomarkers selected using microarray-based serological profiling and specialized bioinformatics. We biopanned phage-display libraries derived from 3 different HNSCC tissues to generate 5,133 selectively cloned tumor antigens. Based on their differential immunoreactivity on protein microarrays against sera from 39 cancer and 41 control patients, we reduced the number of clones to 1,021. The performance of a neural network model (Multilayer Perceptron) for cancer classification on a dataset of 80 HNSCC and 78 control samples was assessed using ten-fold cross-validation repeated 100 times. A panel of 130 clones was found to be adequate for building a classifier with sufficient sensitivity and specificity. Using these 130 markers on a completely new and independent set of 80 samples, an accuracy of 84.9% with sensitivity of 79.8% and specificity of 90.1% was achieved. Similar performance was achieved by reshuffling of the dataset and by using other classification models. The performance of this classification approach represents a significant improvement over current diagnostic accuracy (sensitivity of 37% to 46% and specificity of 24%) in the primary care setting. The results shown here are promising and demonstrate the potential use of this approach toward eventual development of diagnostic assay with sufficient sensitivity and specificity suitable for detection of early stage HNSCC in high risk populations.

Harnessing DNA repair enzyme activity for use as biomarkers for risk assessment and early detection of lung cancer

Tamar Paz-Elizur[a], Ziv Sevilya[a], Yael Leitner-Dagan[a], Edna Schechtman[b], Ofra Barnett-Griness[c], Ronit Almog[c], Gad Rennert[c] and Zvi Livneh[a,*]
[a]*Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel*
[b]*Department of Industrial Engineering and Management, Ben Gurion University of the Negev, Beer Sheva, Israel*
[c]*Department of Community Medicine & Epidemiology, Carmel Medical Center and Technion Faculty of Medicine, and CHS National Israeli Cancer Control Center, Haifa, Israel*

[*]*Corresponding author.*

DNA repair provides a major defense against cancer in humans as indicated by the high cancer predisposition of individuals with hereditary germ-line mutations in DNA repair genes. Several studies have shown that reduced DNA repair plays a similarly important role in sporadic cancers. However the scarcity of functional specific DNA repair assays that are suitable for epidemiological studies slows down the progress in this field.

Our goal is to develop a series of DNA repair biomarkers for cancer risk assessment and early detection, and apply them to large-scale screening directed towards cancer prevention in general, and lung cancer prevention in particular. Our approach is based on functional DNA repair assays, and specifically enzymatic DNA repair activities. We have previously developed an enzymatic activity assay for the repair of the oxidative DNA lesion 8-oxoguanine in extracts from human

peripheral blood mononuclear cells (PBMC) [1]. Using this assay evidence was obtained to indicate that reduced activity of the enzyme OGG (8-oxoguanine DNA glycosylase), which removes 8-oxoguanine from DNA, is a risk factor in lung [3] and head and neck cancer [2].

Under the EDRN program we are developing additional blood tests for DNA repair enzymes involved in repair of oxidative DNA damage. These include the AP endonuclease (APE1; HAP1; APEX), an AP endonuclease that is common to most base excision repair sub-pathways acting on an abasic site, and AAG (also termed ANPG, or MPG), a DNA glycosylase that removes from DNA etheno-adenine (ethenoA). Since AAG removes also hypoxanthine (the spontaneous deamination product of adenine) from the DNA, we are developing in parallel AAG assays for both ethenoA and hypoxanthine, to examine whether there might be differential expression of its repair specificity in cancer risk. Emphasis is given to the development of robust, reproducible, and quantitative assays. These DNA repair biomarkers are expected to be useful for a variety of cancers in addition to lung cancer.

References

[1] T. Paz-Elizur, D. Elinger, Y. Leitner-Dagan, S. Blumenstein, M. Krupsky, A. Berrebi, E. Schechtman and Z. Livneh, Development of an enzymatic DNA repair assay for molecular epidemiology studies: Distribution of OGG activity in healthy individuals, *DNA Repair* **6** (2007), 45–60.

[2] T. Paz-Elizur, R. Ben-Yosef, D. Elinger, A. Vexler, M. Krupsky, A. Berrebi, A. Shani, E. Schechtman, L. Freedman and Z. Livneh, Reduced repair of the oxidative 8-oxoguanine DNA damage and the risk of head and neck cancer, *Cancer Res* **66** (2006), 11683–11689.

[3] T. Paz-Elizur, M. Krupsky, S. Blumenstein, D. Elinger, E. Schechtman, Z. Livneh, Reduced DNA repair activity for oxidative damage and the risk of lung cancer, *J Natl Cancer Inst* **95** (2003), 1312–1319.

Mitochondrial genome deletion aids in the identification of false- and true-negative prostate needle core biopsy specimens

Samantha Maragh
National Institutes of Standards and Technology, Gaithersburg, MD, USA

We report the usefulness of a 3.4-kb mitochondrial genome deletion (3.4mtΔ) for molecular definition of benign, malignant, and proximal to malignant (PTM) prostate needle biopsy specimens. The 3.4mtΔ was identified through long-extension polymerase chain reaction (PCR) analysis of frozen prostate cancer samples. A quantitative PCR assay was developed to measure the levels of the 3.4mtΔ in clinical samples. For normalization, amplifications of a nuclear target and total mitochondrial DNA were included. Cycle threshold data from these targets were used to calculate a score for each biopsy sample. In a pilot study of 38 benign, 29 malignant, and 41 PTM biopsy specimens, the difference between benign and malignant core biopsy specimens was well differentiated ($P < 0.0001$), with PTM indistinguishable from malignant samples ($P = 0.833$). Results of a larger study were identical. In comparison with histopathologic examination for benign and malignant samples, the sensitivity and specificity were 80% and 71%, respectively, and the area under a receiver operating characteristic (ROC) curve was 0.83 for the deletion. In a blinded external validation study, the sensitivity and specificity were 83% and 79%, and the area under the ROC curve was 0.87. The 3.4mtΔ may be useful in defining malignant, benign, and PTM prostate tissues.

A distributed biomarker atlas for lung research aiding the discovery and early detection of cancer biomarkers

Wilbur Franklin (UCHSC), Dan Crichton (JPL), Mary Reid (Roswell Park), Chris Mattmann* (JPL), Andrew Hart (JPL), Dayi Deng (UCHSC), Brian Logue (UCHSC), John Hayes and Deanna Stelling (FHCRC)

**Corresponding author.*

The EDRN Biomarker Atlas Working Group recently has developed a Biomarker Atlas software system that allows a researcher to correlate lung cancer patients with similar characteristics around regions of the lung in which their sample specimens including multiple types of bronchoscopies (e.g., white light, fluorescent), slide images of cancerous lesions, and other data products were collected. Researchers can observe characteristics of collected specimens over time and use the Biomarker Atlas to identify patient trends based on family characteristics, epidemiology, and the site at which the specimen were collected. Patient specimen data is annotated with a common set of data elements that allow the Biomarker Atlas system to properly filter through and co-locate the available data in the system.

Recent technological advances in location-based data discovery and AJAX-based user interfaces, e.g., Google Maps, have served user communities well, allowing even novice users to find businesses, restaurants and sites of interest using a few simple keywords and the selection of a "region" on the map. Not only is the learning curve for these user interfaces significantly smaller, but also the navigation and general interaction patterns between user and application are becoming pervasive in modern search engine technology.

Our "Biomarker Atlas" capability is predicated upon the above recent technical advances and provides a means for lung cancer researchers to browse lung cancer specimens collected at sites participating in the distributed specimen sharing network using a location-based map of a human lung. The backend distributed data system of the Biomarker Atlas is built upon the data grid middleware framework called OODT, the principal enabling technology of the EDRN Resource Network Exchange (ERNE).

The initial pilot sites for the Biomarker Atlas include University of Colorado Health Sciences Center (UCHSC) and Roswell Park Cancer Institute (RPCI). Future work entails collaborating with more EDRN sites to include their specimens in the distributed Biomarker Atlas system, and increase the changes of cancer biomarker discovery and observation.

A reusable web-based CAT (CT) scan data management system for temporally characterizing solid nodules and ground glass opacities in lung cancer patients

Chris Mattmann* (JPL), Michael Khilkin (NYU), William Rom (NYU), Dan Crichton (JPL), Sean Kelly (JPL), Pedro Rivera (NYU), Jane Ko (NYU), Brendan Phalan (NYU), Steve Sotero (NYU) and Ellen Eylers (NYU)

Corresponding author.

NYU and the EDRN Informatics team at JPL have been developing a general, reusable CAT (CT) scan web application and associated data management infrastructure for use by radiologists studying lung cancer to record observations regarding patient solid nodules and ground glass opacities as they evolve over time. We have developed a data model, and associated prototype web front end that allows a user to input metadata information (e.g., demographic information, ATS respiratory symptoms, occupational, family, medical, surgical and female-only history and underwent PFTs and MDCT scanning) and store that information to a backend relational database (MySQL). The resultant web-based data system tracks the evolution of a nodule and integrates this information with clinical data to afford a researcher the ability to better define the follow-up interval in specific situations and identify earlier nodules with a greater malignant potential.

To construct the database, we are leveraging several emerging Java-based open source technologies, including: (1) Apache Tapestry for building a compact, easily maintainable web-based user interface to use for data input and for free-text and forms-based search, and (2) an object-relational persistence technology called Hibernate that affords us the ability to rapidly effect database design changes to our backend CT scan data model and collected data. Using these technologies, we are able to easily provide a means for persisting our collected data into MySQL, an open source relational database management system for further analysis by the radiologists and cancer researchers.

The CAT (CT) scan web database system and associated data model are also being developed generically so that they can be deployed and reused at other EDRN sites that are recording information about CAT scans. The information recorded by the CAT (CT) scan database has the potential to feed into other EDRN software activities, including the development of the EDRN Biomarker Atlas application for sharing patient image specimens (currently bronchoscopies, fluorescence, but with the potential for CT scans) identifying cancerous areas within the lung.

An information model for biomarker research

Steve Hughes (JPL), Deanna Stelling (DMCC), Dan Crichton (JPL), Chris Mattmann* (JPL), Greg Warnick (DMCC) and Suzanna Reid (DMCC)

Corresponding author.

The Information Model is the foundation on which an Information System is built. It defines the entities to be processed, their attributes, and the relationships that add meaning. The development and subsequent management of the Information Model is the single most significant factor for the development of a successful information system.

The mission of the Early Detection Research Network (EDRN), a program managed by the Nation-

al Cancer Institute, is the discovery and validation of biomarkers. In partnership with NASA's Jet Propulsion Laboratory (JPL) the program has developed the EDRN Knowledge Environment to provide integrated access to data across the EDRN enterprise that is captured during the biomarker discovery process. The infrastructure is a set of information system services that allow distributed data repositories to be integrated and accessed real-time. At the heart of the EDRN Knowledge Environment is an information model that describes the data entities involved in biomarker research and how they relate. The scope of this model ranges from tissue banking, to biomarker information, to validation studies, and scientific results. The information model integrates the data entities and enables researchers to navigate through the data repositories in the EDRN community using state-of-the-art search mechanisms.

Over the past year, the EDRN informatics team has made tremendous progress in developing the EDRN Information Model. Using a framework of data modeling tools, the information model has been captured from three sources, the EDRN registry of Common Data Elements (CDE), information specialists familiar with study design and processes, and the working subsystems. The model was captured in an ontology modeling tool and a specification document has been written with the model expressed using object-oriented class notation. The information model continues to be refined in concert with systems development.

The goal of the information model is to unify biospecimen data, scientific data, study specific data, and biomarker data for the EDRN Knowledge Environment and enable the EDRN public portal to provide access to information in data repositories throughout the EDRN community.

A distributed informatics knowledge environment for biomarker research

Dan Crichton (JPL), Mark Thornquist (DMCC), Sean Kelly (JPL), Chris Mattmann* (JPL), Don Johnsey (JPL), Jackie Dahlgren (DMCC), Deanna Steling (DMCC), Greg Warnick (DMCC), Suzanna Reid (DMCC), Cim Edelstein (DMCC), Andrew Hart (JPL) and Heather Kincaid (JPL)

**Corresponding author.*

Informatics in biomedicine is becoming increasingly interconnected via distributed information services, interdisciplinary correlation, and cross institutional collaboration. The Early Detection Research Network (EDRN), a program managed by the National Cancer Institute, has been a pioneer at developing the "data grid" concept for bioinformatics having partnered with NASA's Jet Propulsion Laboratory to deploy JPL's data grid software tools for virtualized access to biospecimen data repositories distributed at several EDRN sites back in 2002.

EDRN's core mission is the discovery and validation of biomarkers. Critical to that mission is having an informatics infrastructure that is "biomarker-centric". The EDRN Knowledge Environment provides integrated access to data across the EDRN enterprise that is captured during the biomarker discovery process. The EDRN Knowledge Environment is built on the notion of a "data grid" allowing loosely related items across a highly heterogeneous, distributed environment to be linked together. The infrastructure of the data grid is a set of information system services that allow distributed databases to be integrated and accessed real-time. At the heart of the EDRN Knowledge Environment is a core "ontology model" that describes the concepts of biomarker research and how they relate to existing data from tissue banking, to managing information about proposed biomarkers, to validation studies and scientific results. Defining this model allows for the data to be integrated into a unified science portal allowing researchers to navigate through the EDRN data using state-of-the-art search mechanisms.

Over the past year, the EDRN informatics team has made tremendous progress in unifying biospecimens, scientific data, study specific data, and biomarker data into the EDRN public portal, providing virtual access to information repositories through the EDRN grid infrastructure and realizing the vision of building the EDRN Knowledge Environment.

A web-based data management infrastructure for curation, annotation and dissemination of biomarker research results for the early detection of cancer

Andrew Hart (JPL), Dan Crichton (JPL), Don Johnsey (NCI), Chris Mattmann* (JPL), Christos Patriotis (NCI), Heather Kincaid (JPL), Sudhir Srivastava (NCI) and Mark Thornquist (DMCC)

**Corresponding author.*

The Early Detection Research Network (EDRN) is an NCI program focused on the discovery and vali-

dation of biomarkers. Bioinformatics plays a critical role in defining, capturing and managing the biomarker information that is produced within the EDRN. In 2006, the EDRN began developing plans for capturing and managing information about cancer biomarkers as part of the information captured by informatics systems within the EDRN enterprise. The specific goal of capturing this information is to provide the research community access to information about biomarkers that have been studied along with their results. This includes access to annotations about the biomarkers studied, organized by organ-site, as well as links to the studies and publications.

The Biomarker Database (BMDB) is a web-based data management tool designed to allow a cancer researcher to quickly grasp the current state of EDRN (and non-EDRN) research pertaining to a particular biomarker. Its features include the ability to aggregate and present heterogeneous biomarker information that is captured from a number of sources including publications (available via PubMed), study information, as well as organ-specific information -including pre-computed and dynamically computed sensitivity, specificity and predictive values.

Included in the Biomarker Database is a curation interface which provides tools for managing and aggregating data from multiple sources into an integrated view that allows several of these sources to remain distributed, but related to a biomarker. The curator builds a representation for each biomarker by indicating connections and associations between organ sites, studies, related publications, and external resources. While the current Biomarker Database is deployed in a prototype, it is being extended to support a peer review process. The peer review will screen biomarkers to validate the quality of the annotated data prior to making it available for access. In addition, the database will support specific security measures to ensure there is restricted access to biomarker information until it released for public view.

Long term, the biomarker database will be fully integrated into the EDRN Knowledge Environment. This will provide scientific access to a variety of information, including biomarkers, allowing scientists to locate and access data across the EDRN enterprise from a single web-based portal.

Identifying putative ovarian cancer serum markers using antibody arrays combined with tandem mass spectrometry

Martin McIntosh*, Wendy Law, Christian Loch, Arturo Ramirez and Paul Lampe
Fred Hutchinson Cancer Research Center, Seattle, WA, USA

**Corresponding author.*

We identify plasma proteins that might classify ovarian cancer using information provided from both high-density antibody arrays and in-depth plasma proteome profiling using mass spectrometry. Each of these platforms has their own advantages and disadvantages. Antibody arrays provide high resolving power and throughput sufficient to characterize individual level-behavior, but due to potential off target binding in a single antibody format one cannot be certain the protein which is captured is the protein to which the antibody was produced. Mass spectrometry has greater fidelity to identify a protein's amino-acid sequence, but is low throughput requiring pooling to provide depth of coverage. Here we use both methods to select high quality candidates based on their combined information to identify biomarkers.

Towards breast cancer diagnostics based on interphase spatial genome positioning

Karen J. Meaburn[a,*], Prabhakar R. Gudla[b], Kaustav Nandy[b], Stephen J. Lockett[b] and Tom Misteli[a]
[a]*National Cancer Institute, NIH, Cell Biology of Genomes Group, Bethesda, MD, USA*
[b]*Image Analysis Laboratory, NCI/SAIC-Frederick, Frederick, MD, USA*

**Corresponding author.*

The genome is non-randomly organized within the three-dimensional space of the cell nucleus. The nuclear position of many genes and genomic regions changes during physiological processes such as proliferation, differentiation and disease. We hypothesize that we can exploit the changes in gene positioning patterns as indicators of disease and disease potential to develop a novel tool in the detection of breast cancer and for the identification of pre-malignant cells with tumorigenic potential in normal tissue. To this end, we have analyzed

the spatial position of a defined set of cancer-associated genes in an established mammary epithelial 3D cell culture model of early stages of breast cancer. We find that the genome is globally reorganized during normal and tumorigenic epithelial differentiation. Systematic mapping of changes in spatial positioning of cancer-associated genes reveals gene specific positioning behavior and we identify several genes which are specifically repositioned during tumorigenesis. Alterations of spatial positioning patterns during differentiation and tumorigenesis were unrelated to gene activity. Our results demonstrate the existence of activity-independent genome repositioning events in early stages of tumor formation. To test the feasibility of these genes for diagnostic purposes, we are comparing their positioning patterns in normal and tumor breast tissues and we are defining a set of marker genes which exhibit differential positioning in malignant tissues.

Analysis of GP73 in combination with fucosylated alpha-1-antitrypsin and fucosylated kininogen as a biomarker of primary hepatocellular carcinoma

Mengjun Wang[a], Ronald E. Long[a], Mary Ann Comunale[a], Omer Junaidi[b], Jorge Marrero[c], Adrian M. Di Bisceglie[b], Timothy M. Block[a] and Anand S. Mehta[a,*]

[a]*Drexel University College of Medicine, and Department of Microbiology and Immunology and Drexel Institute for Biotechnology and Virology, Doylestown, PA, USA*

[b]*Saint Louis University School of Medicine, St. Louis, MO, USA*

[c]*Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA*

[*]*Corresponding author.*

We have previously reported changes in N-linked glycosylation that occur with the development of liver cancer and through the use of glycoproteomics, identified many of those proteins. The altered glycosylation of two proteins identified, fucosylated kininogen and fucosylated alpha-1-anti-trypsin (Fc-AAT) were analyzed individually or in combination with a previously identified biomarker, Golgi protein 73 (GP73). This analysis was performed in two separate patient cohorts consisting of a total of 277 patients with cirrhosis or cirrhosis plus hepatocellular carcinoma (HCC). In both patient cohorts the levels of fucosylated kininogen and Fc-AAT were significantly higher in patients with HCC compared to those with cirrhosis ($p < 0.0001$). The performance of these two markers was determined for each marker individually or in combination with two previously identified proteins, golgi protein 73 (GP73) and alpha-fetoprotein (AFP). Greatest performance was achieved through the combination of fucosylated kininogen, AFP and GP73 giving an optimal sensitivity of 95%, a specificity of 70% and an AUROC of 0.94. In conclusion, the altered glycosylation of serum glycoproteins can act as potential biomarkers of primary hepatocellular carcinoma when used independently or in combination with other makers of liver cancer.

Challenges and potential solutions in the design of biomarker prevalidation trials

Daniel Normolle[*] and Dean Brenner

[*]*Corresponding author: University of Michigan.*

While popular as a prototype for development, no biomarker for early detection of cancer has ever completed the five-phase validation process described by Sullivan Pepe et al [2001]. Many potential markers never complete the early phases, because: the application of the marker is not fixed at the outset of the process; the range of feasible applications for a marker is not properly assessed; the number of samples in biorepositories, including appropriate controls, is grossly inadequate for rigorous validation. We present examples that demonstrate these challenges, and discuss possible alternative study designs in the early part of the validation paradigm that may address these issues.

The American-Australian Mesothelioma Consortium: DRN Mesothelioma Biomarker Discovery Laboratory

Harvey I. Pass[*], Amanda Beck[*], Alla Ivanova[*], Sergey Ivanov[*], Chandra Goparaju, Jessica Donington, Jenette Creaney and Bruce Robinson

[*]*Corresponding author: NYU School of Medicine, New York; University of Western Australia, Perth, Australia.*

Mesothelioma (MM) is an orphan disease which is asbestos related, presently has a median survival of

8-18 months from diagnosis and for which therapeutic options remain problematic. Nevertheless, there is justification for pursuing early biomarkers for MM in well-characterized, asbestos-exposed subjects since (1) cytotoxic chemotherapy is associated with a 41% partial response rate (2) multimodal approaches involving surgery, chemotherapy with or without radiation therapy can have median survivals of 32 months or greater for Stage I patients. The American Australian Mesothelioma Consortium, centered at NYU, has been the funded EDRN Biomarker Discovery Laboratory for Mesothelioma with Fujirebio Diagnostics, Inc. as an industrial partner since 2006. Preliminary data regarding two specific biomarkers, SMRP and Osteopontin (OPN) were originally validated at NYU using sera and plasma from specimens which originated from clinical trials at the National Cancer Institute and the Karmanos Cancer Institute. This abstract summarizes the status of studies since September 2006.

SMRP

At NYU, we evaluated SMRP in serum from MM patients ($n = 90$), lung cancer (LC) patients ($n = 174$), age and tobacco-matched AE individuals ($n = 66$), and in MM pleural effusions ($n = 45$), benign effusions ($n = 30$), and non-MM effusions ($n = 20$) using the MesoMark$^{\mathrm{TM}}$ ELISA kit (Fujirebio Diagnostics). Receiver operating characteristic curves (ROC) were used to define true and false positive rates at various cut-offs. *RESULTS:* Mean serum SMRP levels were higher in MM compared to LC (9.47 $\pm$ 3.39 nM [mean $+/-$ SEM] vs 1.95 $\pm$ 0.44 nM, $p = 0.029$), and Stage I MM SMRP levels ($n = 12$; 2.09 $\pm$ 0.41 nM) were significantly higher than those in AE individuals (0.99 $\pm$ 0.09 nM, $p = 0.02$, respectively). Stage 2–4 SMRP serum levels were significantly higher (10.61 $\pm$ 3.89 nM, $p = 0.03$) than those for Stage 1. The area under the ROC (AUC) for serum SMRP was 0.805 for differentiating MM and AE, cut-off = 1.2 nM (sensitivity = 76.7%, specificity = 72.7%). The positive predictive value was 69% and negative predictive value was 79.8% for serum. MM pleural effusion SMRP was significantly higher than benign or other non-MM pleural effusions (65.57 $\pm$ 11.33 nM vs 18.99 $\pm$ 7.48 nM [$p = 0.001$] and 27.46 $\pm$ 11.25 nM [$p = 0.021$] respectively). *These SMRP data are compatible with results from other smaller cohorts of MM and AE patients, and confirmed the data of Robinson et al in a North American cohort.*

Osteopontin (*OPN*)

Our group previously revealed that OPN could also discriminate AE from MM, and that serum OPN levels were influenced by asbestos exposure and degree of radiographic changes. We demonstrated that in the same cohort of patients, the sensitivities of serum SMRP and serum OPN were improved when both markers were used. We also investigated the plasma concentrations of OPN to see if there was improvement in differentiating AE from MM and whether plasma OPN was a prognostic marker for MM. Plasma OPN from 39 MM (mean age 63+ 8.4 years;9 females, 30 males; 11 Stage I/II, 28 Stage III/IV; 21 having surgical cytoreduction) and from 79 asbestos-exposed (AE) individuals (mean age 63+ 10.6 years; 9 females, 70 males) was measured with the Research and Diagnostics (R&D, Minneapolis MN) and Immuno- Biological laboratories (IBL, Minneapolis, MN) kits. Differences in OPN levels in MM and AE individuals were compared using ROC curves and sensitivity and false positive rates based on logistic models for each of the test kits. Survival for MM was estimated using Kaplan-Meier curves; comparisons between groups are based on log rank chi-square tests. Hazard ratios and 95% confidence intervals were estimated from Cox proportional hazards models. A formal cut point analysis was performed using the maximum chi-square with p value adjustment method to determine the OPN values that were most strongly associated with survival. *RESULTS:* The area under the ROC curve was 0.93 (R&D, cutpoint corresponding to a sensitivity of 0.91 and false positive rate of 0.23 = 59.6 ng/ml) and 0.96 (IBL, cutpoint corresponding to a sensitivity of 0.91 and false positive rate of 0.10 = 132.6 ng/ml). The median overall survival for all 39 MM was 11 months (95% CI: 5, 13) with a 37 month median overall follow up for survivors. Patients with OPN levels greater than or equal to 212.6 ng/ml had 5.7 (95% CI: 2.4, 13.3) times the mortality risk of patients with lower levels (adjusted p value = 0.007). Additional multivariable analyses indicate that lower stage (HR: 5.0; 95% CI: 1.7, 15.0; $p = 0.004$) and OPN level from the R&D kit less than the cutpoint of 212.6 ng/ml (HR: 3.5; 95% CI: 1.5, 8.4; $p = 0.004$) were significantly associated with improved survival. *These data confirm that plasma OPN may be a sensitive discriminator for the development of MM in high risk AE cohorts, and also that OPN levels in MM patients may be of prognostic importance.*

Our group has further characterized OPN as having three distinct isoforms in MM, and that each isoforms has different functional characteristics with regard to proliferation, migration, and invasion.

Validation trials for SMRP and OPN

In collaboration with the DMCC, a validation trial to establish ranges for measuring SMRP and OPN has been written. Using cohorts from Karmanos, NCI, Libby Montana, Australia, and Mt. Sinai Hospital of NYC, over 680 serum specimens representing either high risk AE or MM patients are presently housed at the EDRN Biomarker Facility at Bellevue Hospital in NYC. Blinded validation of these specimens will be performed at three validation sites. Pending the results of this validation, a prospective trial examining these markers in the serum of villagers in the epidemic sites for MM in Cappadochia Turkey will be performed. Preliminary data from these villages regarding the sensitivity and specificity of SMRP and OPN are gratifyingly consistent with the data from the EDRN Biomarker Discovery Laboratory.

New potential biomarkers

MMP9 was found in preliminary experiments to have increased expression in mesothelial cells subjected to asbestos. We measured plasma MMP9 from MM, AE, LC, and BD individuals and found that MMP9 was elevated significantly in AE but decreased in MM. Paradoxically, this was not the case with BD and LC where BD individuals had significantly lower levels of plasma MMP9. Further investigations regarding the utility of MMP9 to differentiate LC from MM as part of a combinatorial analysis of markers are ongoing.

Hyalluronic Acid Proteoglycan Link Protein 1 (*HAPLN1, CRTL1*)*:* originally was predicted to be 23 fold elevated in the extracellular matrix of MM compared to AE cohorts from our genomic pathway analyses. Our laboratory has validated these data in matched normal peritoneum and MM specimens. Due to the lack of reagents for protein measurement, we generated a rabbit polyclonal antibody which revealed that (1) MM cell lines stained for the antibody as opposed to mesothelial cell lines or mesothelial short term cultures (2) paraffin embedded MM of any histology had cellular but not stromal staining (3)transfection of HAPLN1 into low expressing MM cell lines remarkedly increased invasion, proliferation, and migration of those cell lines compared to empty vector transfection. Attempts at the production of an ELISA are ongoing at this time.

Identification of biomarkers from glycans released from serum glycoproteins using a printed glycan array (*Collaboration with Cellexicon, Inc.*)*:* is a new initiative that was started in 2007. Using a preliminary set of 20 specimens each representing AE, benign disease but smoker (BD), adenocarcinoma of the lung (AD), MM, and squamous cell carcinoma of lung (SC), comparative analyses of the serum glycan status was determined for AE versus MM, BD versus AD, and BD versus SC. Using six ranked glycans, an AUC of 0.93 was determined for AE versus MM. similarly, an AUC of 0.96 was recorded comparing BD to AD using 4 ranked glycans, and an AUC of 0.94 determined between BD and SC. These numbers were increased to approximately 70 AE and 70 MM in a subsequent analysis. Using 20 glycans, an AUC of 0.94 ws determined for AE versus MM with a sensitivity of 0.87 and a specificity of 0.94. These investigations are continuing.

MicroRNA profiles between MM and normal peritoneum, and between MM and AD (*Collaboration with Rosetta Genomics*)*:* is another new initiative begun in 2007. 36 snap frozen MM (8F,28M) with 20 matching peritoneum from the time of their resections were used for mir analysis. There were 10 Stage I/II (MS $=$ 24 months) and 26 Stage III/IV patients (MS $=$ 8 months). Samples were hybridized to Rosetta Genomics microRNA microarray. Stepwise Cox regression allowed the evaluation of the significance of individual and combined mirs on patient survival, alone or in tandem with clinical risk factors, such as stage, age, smoking and gender. Kaplan Meier plots and logrank analysis were used to compare survival and time to progress profiles of discrete groups. *RESULTS:* For the 20 matched tumor/normal pairs, there were 136 miRs mirs expressed in at least one set of samples. Of these, 66 mirs were significantly differentially expressed ($p < 0.005$) between MM and normal abdominal mesothelium. 34 mirs were upregulated in tumor vs. normal, while 32 were downregulated respectively. No differences in mir expression were seen with regards to gender, age, or the histology of the MM. In a univariate analysis, lower stage, epithelial histology, absence of lymph node involvement, minimal asbestos exposure and female gender were predictive of longer survival. The presence of 4 mirs significantly improved ($p = 9 \times 10^{-6}$) survival in MM patients (MS$>$38 months) compared to those who did not express the mirs (MS $=$ 5 months). These data represent the first use of mirs for the discovery of potentially novel biomarkers as well as the prognostication of MM. In a separate series of experiments, Rosetta also profiled 7 MM formalin-fixed paraffin-embedded (FFPE) samples and 85 adenocarcinoma FFPE samples from multiple origins, including breast, colon, pancreas and others using the Rosetta Genomics microRNA microarray. Expression levels of over 700 microRNAs

were measured on the microarray and compared between the two sample groups. Differentially expressed microRNAs were identified based on their expression on the microarray. The company further developed a microRNA-based assay using qRT-PCR and tested it on 22 MM samples and 44 adenocarcinoma samples from the following origins: lung, colon, pancreas, bladder, kidney, ovary and breast. *RESULTS*: Three microRNAs were significantly differentially expressed between MM and adenocarcinoma. One microRNA was over-expressed and two microRNAs were under-expressed in MM relative to adenocarcinoma. MM primary tumors could be separated from lung adenocarcinoma tumors with sensitivity and specificity exceeding 95%. MM could be separated from the full set of adenocarcinoma samples with sensitivity and specificity above 90%. These data imply that a small number of microRNAs is sufficient to discriminate MM from adenocarcinoma from multiple origins, and mir expression profiles in serum of MM and AE patients is beginning.

Discovery and verification of head-and-neck precancer and cancer biomarkers by differential protein expression analysis using iTRAQ-labeling and multidimensional liquid chromatography and tandem mass spectrometry

Ranju Ralhan[a,b,d,*,1], Leroi V. DeSouza[a,b], Ajay Matta[d], Satyendra Chandra Tripathi[d], Shaun Ghanny[b,c], Siddartha Datta Gupta[e], Sudhir Bahadur[f] and K.W. Michael Siu[a,b]

[a]*Department of Chemistry, York University, Toronto, Ontario, Canada*

[b]*Centre for Research in Mass Spectrometry, York University, Toronto, Ontario, Canada*

[c]*Department of Mathematics and Statistics, York University, Toronto, Ontario, Canada*

[d]*Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India*

[e]*Department of Pathology, All India Institute of Medical Sciences, New Delhi, India*

[f]*Department of Otorhinolaryngology, All India Institute of Medical Sciences, New Delhi, India*

[*]*Corresponding author: Tel.: +91 416 736 2100 40048; E-mail: ralhanr@yorku.ca.*

[1]RR is a recipient of the U.S. National Cancer Institute(NCI)-Novartis Translational Cancer Research Fellowship Award of the International Union against Cancer (UICC) at York University.

Currently, there are no clinically established tumor markers to facilitate the diagnosis or prognosis of head-and-neck cancer. Proteomics combined with mass spectrometry (MS) is a powerful paradigm for the examination of proteins in a global manner in the post-genomics era for the discovery of cancer biomarkers. Multidimensional liquid chromatography – mass spectrometry (LC-MS/MS) has been used for the analysis of biological samples labeled with isobaric mass tags (iTRAQ) to identify proteins that are differentially expressed in human head-and-neck/oral squamous-cell carcinomas (HNOSCCs) and Oral premalignant lesions (OPLs) in relation to non-cancerous head-and-neck tissues (control) for cancer biomarker and drug target discovery. We identified over 800 non-redundant proteins in HNOSCCs, including structural proteins, signaling components, enzymes, receptors, transcription factors and chaperones. The panels of proteins showing consistent differential expression in HNOSCCs and OPLs relative to the non-cancerous controls were discovered. A panel of the three best-performing biomarkers – achieved a sensitivity of 0.92, specificity of 0.91 and predictive value of 0.95 in discriminating cancerous from non-cancerous head-and-neck tissues. Verification of differential expression of proteins in clinical samples of HNOSCCs and paired and non-paired non-cancerous tissues by immunohistochemistry, immunoblotting, and RT-PCR confirmed their overexpression in head-and-neck cancer. Verification of the panel of potential biomarkers in an independent set of HNOSCCs achieved a sensitivity of 0.92, specificity of 0.87, and predictive value of 0.83 in discriminating cancerous from non-cancerous head-and-neck tissues, thereby confirming their overexpression and utility as credible cancer biomarkers. Large scale validation of these biomarkers in underway for ultimate goal of translation into clinics.

Computational methods for MALDI-TOF MS data analysis and their application for discovering peptide and glycan biomarkers of hepatocellular carcinoma

Habtom W. Ressom[a,*], Rency S. Varghese[a], Lenka Goldman[a], Yanming An[a], Christopher A. Loffredo[a], Mohamed Abdel-Hamid[b], Zuzana Kyselova[c], Yehia Mechref[c], Milos Novotny[c], Steven K. Drake[d] and Radoslav Goldman[a]

[a]*Georgetown University, Lombardi Comprehensive Cancer Center, Washington, DC, USA*

[b]*Minia University and Viral Hepatitis Research Laboratory, NHTMRI, Cairo, Egypt*
[c]*National Center for Glycomics and Glycoproteomics, Department of Chemistry, Bloomington, IN, USA*
[d]*Clinical Chemistry Service, Department of Laboratory Medicine, NIH, Bethesda, MD, USA*

[*]*Corresponding author.*

We propose computational methods for quantitative comparison of peptides and glycans in serum using matrix-assisted laser desorption/ionization–time of flight mass spectrometry (MALDI-TOF MS). The methods are applied to identify candidate biomarkers in serum samples of 203 participants from Egypt; 73 hepatocellular carcinoma (HCC) cases, 52 patients with chronic liver disease (CLD) consisting of cirrhosis and fibrosis cases, and 78 population controls. Two complementary sample preparation methods were applied prior to generating mass spectra: (1) low molecular weight (LMW) enrichment of each serum sample was carried out for MALDI-TOF quantification of peptides, and (2) glycans were enzymatically released from proteins in each serum sample and permethylated for MALDI-TOF quantification of glycans. A peak selection algorithm was applied to identify the most useful peptide and glycan peaks for accurate detection of HCC cases from high-risk population of patients with CLD. The peak selection process was preceded by peak screening, where we eliminated peaks that have significant association with covariates such as age, gender, and viral infection based on the peptide and glycan spectra from population controls. The performance of the selected peptide and glycan peaks was evaluated in terms of their ability in detecting HCC cases from patients with CLD in a blinded validation set. Finally, we investigated the possibility of using both peptides and glycans in a panel to enhance the diagnostic capability of these candidate markers. Further evaluation is needed to examine the potential clinical utility of the candidate peptide and glycan markers identified in this study.

Using high throughput resequencing microarrays to detect mutations in genes involved in lung cancer

Liachun Piao[a], Albert Gunnison[a], Arthur Nadas[a], Wen-chi Chen[a], Daisuke Nonaka[a], Simon Spivack[b], Harvey Pass[a], William N. Rom[a,*] and Moon-shong Tang[a]

[a]*Departments of Environmental Medicine, Pathology, Medicine and Cardiothoracic Surgery, New York University School of Medicine, New York, NY, USA*
[b]*Department of Medicine, Albert Einstein College of Medicine, New York, NY, USA*

[*]*Corresponding author.*

Cancers are diseases with mutations in multiple genes that are involved in carcinogenesis and metastasis. Identification of genetic abnormalities in these genes may enhance understanding of the pathological process of cancer and the diagnosis of cancer. Identification of mutations in multiple genes by conventional capillary DNA sequencing method is costly and time consuming. Over the past decade, the development of high density microarray-chip technologies provide us with a promising approach that potentially will allow comprehensive analysis of genetic changes in any number of genes, limited only by the density of the chip, in a high resolution and high throughput fashion. Using Affymetrix custom high density high throughput resequencing microarray technology we have developed a microarray-chip that can detect mutations in the exon regions of 14 genes that are frequently mutated in lung cancer. These genes are: tumor suppressor genes, p53, Rb, p15, p16, PPP2R1B, and proto-oncogenes, K-, N- and H-ras, DNA repair-related genes MSH2, MLH1, XPC and BRCA1, the signal transduction gene PTEN, and the apoptotic gene BAX. A total of 70 microarrays were hybridized with genomic DNA from lung tissue from individuals with lung cancer (tumor tissues and their marginal tissues of 28 patients including 22 matched pairs) and from healthy individuals ($n = 8$ including PBMC). We found that GDAS software provided by Affymetrix could not identify 10–20% of the total sequence, therefore, mutations in these "no call" regions were undiscovered. Furthermore, this software failed to detect insertion, deletion or heterozygous mutations. To overcome this problem, we have designed an algorithm and software that is able to identify 99–99.5% of the total sequence. Using the combination of Affymetrix and our sequence software we detected 20 point mutations in tumor tissues, 19/20 identical point mutations in tumor marginal tissues; 170 SNP sites in 28 lung cancer patients and 8 healthy individuals were not significantly different. We also found 1 heterozygous mutation in the XPC gene and 3 deletion mutations in exon 11 of the p53 gene in lung tumor tissues.

CA19-9 decorated mucin proteins in pancreatic cancer

Christy A. Rothermund-Franklin*, Thomas C. Caffrey and Michael A. Hollingsworth
University Nebraska Medical Center, Eppley Institute for Research in Cancer and Allied Diseases, Omaha, NE, USA

**Corresponding author.*

The carbohydrate antigen 19-9 (CA19-9) is a tumor-associated carbohydrate antigen expressed on mucins in numerous cancers. Mucins are extensively glycosylated, high molecular weight glycoproteins whose normal functions are to protect and lubricate luminal epithelial surfaces. In cancers, mucin expression is deregulated and the mucin core proteins aberrantly glycosylated. Measurement of CA19-9 decorated mucins has long been used as a marker for the diagnosis and prognosis of cancers, however; a careful analysis of which mucin core proteins are CA19-9 decorated has not been performed. In pancreatic cancer, CA19-9 is the only serum based clinical assay in use, but it is well known that the measurement of CA19-9 as a diagnostic tool can be variable. In this study, we sought to determine which set of mucin core proteins carried the CA19-9 carbohydrate antigen in pancreatic cancer. To this end, we performed deglycosylation of samples immunoprecipitated from cancer cell lines and pancreatic cancer patient tissue specimens. This allowed us to unmask and assay for different mucin core proteins. Our studies show that the mucin core proteins MUC1, MUC4 and MUC5AC are CA19-9 decorated in pancreatic cancer. These findings suggest that the use of a combination detection strategy of both the CA19-9 tumor antigen along with the mucin core protein(s) may help to more efficiently diagnose and better differentiate pancreatic cancer from other diseases.

MicroRNA expression in colon adenocarcinoma is associated with prognosis and therapeutic outcome

Aaron J. Schetter[a,*], Suet Yi Leung[b], Jane J. Sohn[a], Krista A. Zanetti[a], Elise D. Bowman[a], Chang-gong Liu[c], Carlo M. Croce[c] and Curtis C. Harris[a]
[a]*Laboratory of Human Carcinogenesis, CCR, NCI, MD, USA*
[b]*University of Hong Kong, Pokfulam, China*
[c]*Ohio State University Comprehensive Cancer Center, Columbus, OH, USA*

[*]*Corresponding author.*

MicroRNAs are small, noncoding RNAs that have potential as both diagnostic biomarkers and therapeutic targets in cancer. MicroRNA expression profiling of colon tumors and paired nontumor tissues was performed on a US cohort of 84 subjects. We evaluated associations with tumor status, TNM staging, survival prognosis and response to adjuvant chemotherapy. Associations were validated in a Chinese cohort of 113 subjects. Thirty-seven microRNAs were differentially expressed in tumors from the US cohort. miR-20a, miR-21,miR-106a, miR-181b and miR-203 were selected for validation and all were enriched in tumors from the Chinese cohort ($p < 0.001$). Adenomas express high miR-21 levels ($p = 0.006$) and may represent an early event in colon carcinogenesis. More advanced tumors also express higher miR-21 ($p < 0.001$). MiR-21 was expressed at high levels in colonic carcinoma cells by *in situ* hybridization, indicating that aberrant expression of miR-21 is likely due to increased expression in tumor cells rather than stromal cells. High miR-21 expression predicts a poor survival in both the training ($p = 0.005$) and validation cohorts ($p = 0.001$) and are independent of clinical covariates, including TNM staging. High miR-21 expression predicts a poor response to fluorouracil-based, adjuvant chemotherapy as well. These results were found in two independent cohorts with different racial compositions arguing that these results are likely applicable to the majority of sporadic colon adenocarcinoma cases and are not limited to the study population. Our results suggest that miR-21 expression in tumors is a useful prognostic biomarker for colon adenocarcinomas and can predict a poor therapeutic outcome. If high levels of miR-21 expression are causal to poor survival, then anti-miR21 therapeutics may have the potential to improve survival outcomes of colon cancer patients to reduce the burden of this disease.

Trace DNA in exhaled breath condensate (EBC)

Simon Spivack
Albert Einstein College of Medicine, NY, USA

Trace DNA in exhaled breath condensate (EBC) could potentially support CpG methylation analyses. We analyzed promoter methylation status of *DAPK*,

RASSF1A, PAX5β, P16, and *CDH1* by a previously-developed, tag-modified bisulfite genomic DNA sequencing (tBGS) assay, in EBC samples from 37 consecutive individuals [controls: current ($n = 16$), former ($n = 9$), never-smokers ($n = 8$); cases: current and former ($n = 4$)]. The tBGS results showed: (1) Wide inter-individual variability in both methylation density and spatial patterns, only partially explained by smoking, for the three promoters showing methylation (*DAPK, PAX5B, RASSF1A*); (2) Methylation patterns from paired EBC and mouth rinse specimens did not overlap; (3) For controls, higher average methylation rates were generally observed in current- and former- *versus* never-smokers (0.0030 and 0.0045, resp.); (4) Methylation density classified subjects as lung cancer cases or controls by ROC analysis by *DAPK* alone ($p = 0.025$), and by *DAPK, PAX5B, RASSF1A* combined ($p = 0.014$); (5) The tBGS-generated spatial patterns also classified cases as more methylated than controls on a site- and gene-specific basis ($p = 0.0000$ to 0.0060). A second method, real-time quantitative methylation-specific probing (qMSP) was employed at two separate probe sites [-158bp upstream (higher methylation) and -99bp downstream (less methylation)] in the *DAPK* promoter, confirming DNA in exhaled breath. For the upstream site, pack-years and age correlated with degree of methylation; results diverged between the two probe sites. Overall, our results suggest that: (a) methylation analyses are feasible from exhaled breath; (b) the anatomic source of that DNA appears to be the lower airway; (c) methylation correlates with tobacco exposure and case-control status; (d) spatial pattern of promoter methylation is not easily inferred by limited sampling techniques such as MSP; and (e) EBC-DNA methylation assays warrant further development.

A potential prediction model for the recurrence of colorectal adenomas based on epigenetic changes in an index polyp

Bogdan Paun[a], Justin Price[a], Zhe Jin[a], Yuriko Mori[a], Stephen J. Meltzer[a] and Sanford A. Stass[b,*]
[a]*GI Division, Johns Hopkins University, Baltimore, MD, USA*
[b]*Department of Pathology, University of Maryland, Baltimore, MD, USA*

[*]*Corresponding author.*

Background: Colorectal cancer (CRC) is one of the leading causes of death in the Western world. CRC develops from premalignant lesions, chiefly colorectal adenomas. Today, the most accurate screening method for finding colon adenomas is colonoscopy for subjects over age 50. However, the cost and risk associated with this procedure are relatively high. This is especially true for patients who already have a polyp discovered.

Aim: The purpose of the current study was to correlate epigenetic changes in the index polyp, histological characteristics and age with the recurrence of another polyp on repeat colonoscopy within one to three years, to evaluate the potential validity of index polyp methylation as an assay of the need for repeat colonoscopy within 3 years.

Materials & Methods: 39 paraffin-embedded polyps from 39 patients were selected for this study (20 progressors and 19 non-progressors). DNA was extracted, modified with bisulfite using Qiagen's Epitect kit, and subjected to real-time quantitative methylation-specific PCR for 5 genes using an ABI 7900HT sequence detector. Data was analyzed using Sequence Detector Software, Excel and Analyse-It.

Results: Using a variable student's t test we identified one gene (p-value < 0.008) as a potential marker of polyp recurrence. The remaining 4 genes lacked statistical significance on univariate analysis. However, using linear discriminant analysis based on 3 of the 5 genes, we obtained a combined prediction model with a p-value < 0.006.

Conclusions: We identified 1 gene as a possible prediction marker for the recurrence of polyps. This gene may be involved in polyp development and deserves further study as a clinical biomarker. The small number of samples and genes studied thus far suggests that these results can improve further by studying additional samples and genes. The need for repeat colonoscopy may ultimately be narrowed down to subjects with positive biomarker findings.

Development of a multiplexed immunoassay panel for ten cancer markers: AFP, Ca125, Ca19-9, CEA, cKit, E-Cadherin, EGFR, ErbB2, MMP-9, & Osteopontin

Martin Stengelin[*], Jessica Levy, Eli N. Glezer and Jacob N. Wohlstadter

[*]*Corresponding author: Meso Scale Diagnostics.*

Table 1

Analyte	Assay range	Analyte concentration in ∼20 normal serum & plasma samples	
		Median	Range
AFP	0.15–500 ng/ml	3 ng/ml	1–14 ng/ml
Ca125	3–5,000 U/ml	8 U/ml	$<$ 3–90 U/ml
Ca19-9	5–5,000 U/ml	26 U/ml	$<$5–170 U/ml
CEA	0.1–500 ng/ml	4 ng/ml	1–6 ng/ml
cKit	3–2,000 ng/ml	130 ng/ml	60–210 ng/ml
E-Cadherin	0.1–500 ng/ml	18 ng/ml	10–48 ng/ml
EGFR	0.6–1,000 ng/ml	67 ng/ml	48–82 ng/ml
$ErbB_2$ (Her_2/neu)	0.2–500 ng/ml	3 ng/ml	2–4 ng/ml
MMP-9	0.2–3,000 ng/ml	50 ng/ml	25–1,600 ng/ml
Osteopontin	0.05–100 ng/ml	3 ng/ml	0.8–17 ng/ml

An electrochemiluminescence-based multiplexed immunoassay panel was developed and analytically validated for simultaneous measurement of ten classical and emerging cancer markers per well in a 96-well format. The assay requires only 25 ul of a fivefold diluted sample, i.e. 5 ul for all ten assays combined.

The assay format is simple: assay diluent and diluted sample or calibrator is added to blocked and washed plates, and after a two-hour incubation with agitation, plates are washed, and detection antibody reagent is added. After a second two-hour incubation, plates are washed and read on a MSD SECTOR® Imager 6000 instrument (throughput of one plate per minute).

The assays are sensitive enough to measure these biomarkers in normal samples, and the dynamic range is sufficient for measurement of the elevated levels expected in disease states without additional serial sample dilution.

Spike recovery and dilution linearity were in the range of 80% to 120%. Intra-plate CVs were approximately 5–15%. Each analyte in the multiplex panel is measured accurately even in the presence of a high abundance of other analytes, as demonstrated in an experiment where an elevated concentration of one analyte at a time was measured.

On-chip diagnostic immunoassays of panels of bioselected tumor antigens: Binding to serum autoantibodies for the early detection of cancer

Michael A. Tainsky[a,*] and Sorin Draghici[b]
[a]*Barbara & Fred Erb Professor of Cancer Genetics, Director, Program in Molecular Biology and Genetics Karmanos Cancer Institute, Department of Pathology, Wayne State University School of Medicine, Detroit, MI, USA*
[b]*Director of the Bioinformatics Core, Karmanos Cancer Institute, Associate Professor, Department of Computer Science, Wayne State University, Detroit, MI, USA*

[*]*Corresponding author.*

The humoral immune response is an exquisite biosensor of novel proteins expressed by tumor cells. Panels of tumor antigens could provide a sensitive and specific multianalyte immunoassay for the presymptomatic of cancer. The development of early detection tests for cancers has previously depended on single biomarker molecules. Using a high-throughput cloning method, panels of epitopes/antigens that react with autoantibodies to tumor proteins in the serum of patients with cancer have been isolated. By cloning a large set of epitopes or tumor antigens, discovery of biomarker panels was directed in an unbiased fashion without a previous notion of their function rather than choosing individual candidate biomarkers. The binding properties of these serum antitumor antibodies on microarrays and advanced bioinformatics tools led to a panel of diagnostic antigens. We call this process Epitomics. The process has been successful for ovarian, breast, oral, lung and colon cancer. The sequences that were identified using high-throughput phage display cloning technology have led to the discovery of novel disease-related proteins also useful as tissue biomarkers. There are numerous advantages of employing serum antibodies as the analytes, not the least of which is the ability to readily adapt these assays to standard clinical platforms.

Signalyte™-II, ultra-sensitive fluorometer for detection of low-concentration markers

Shuhong Li, Yunqi Zhang, Daniel L. Adams, Cha-Mei Tang[*] and Platte T. Amstutz

Creatv MicroTech, Inc., Potomac, MD, USA

Corresponding author: Cha-Mei Tang, Creatv MicroTech, Inc., 11609 Lake Potomac Drive, Potomac, MD 20854, USA. Tel.: +1 301 983 1650; Fax: +1 301 983 6264; E-mail: cmtang@creatvmicrotech.com.

Creatv MicroTech introduces its ultra-sensitive fluorometer/luminomer, Signalyte™-II, which is one to three orders of magnitude more sensitive than commercial plate readers. Sample size is 1 to 35 microliters. The instrument simultaneously tests up to eight samples, plus a reference.

The instrument incorporates Creatv's proprietary Integrating Waveguide technology to measure fluorescence signals. Sample is placed in a specialized glass cuvette that acts as an optical waveguide. Excitation light impinges at a 90 degree angle to the side of the cuvette. Emitted light is efficiently collected at the end, where it passes through lenses and optical filters to a spectrometer.

Signalyte™-II comes standard with four excitation wavelengths at 470nm, 530nm, 590nm, and 630nm to allow multiplex detection of analytes. Shorter and longer wavelengths are available on a custom basis. Bandpass filters are selected to match the absorption and emission spectra for popular organic dyes and quantum dots. Detection spectra are displayed from 400 to 750 nm.

Side-by-side comparisons were performed with three brand-name plate readers, BMG FLUOstar Omega, Perkin Elmer VICTOR^{3}V, and Molecular Devices SpectraMax M5. Four popular dyes were prepared in serial dilutions spanning 12 orders of magnitude: Cy™-5 (GE Healthcare), FITC (Fisher), SPHERO™ Fluorescent Purple, and SPHERO™ Fluorescent Sky Blue Particles (Spherotech). Sample size was 200 microliters for plate readers and 35 microliters for Signalyte™-II. Results showed that Signalyte™-II is one to three orders of magnitude more sensitive than the plate readers, depending on the particular dye.

Signalyte™-II's sensitivity, sample size and flexibility are ideally suited for cancer biomarker research.

Lung cancer DSA: A platform for discovery of biomarkers in lung cancer

Austin Tanney
Almac Diagnostics, Craigavon, UK

Non-small cell lung cancer (NSCLC) is the leading cause of cancer mortality worldwide with poor differential diagnosis of the disease and with low response rates to standard chemotherapy treatment. It is therefore a subject of extensive research focused on identification of reliable genomics biomarkers to aid in accurate classification of the disease, predicting its progression and patients' response to both available therapies and those in development. Powerful genomics tools used in this research are however lacking disease focus and thus are likely to miss potentially vital information contained in patients' tissue samples.

Through a combination of large-scale in-house sequencing, gene expression profiling and public sequence and gene expression data mining we have characterised the transcriptome of NSC lung cancer and used this information to create a unique disease focused microarray – Lung Cancer DSA™ research tool. Built on the Affymetrix GeneChip™ platform the tool allows for interrogation of ∼60,000 transcripts relevant to Lung Cancer, tens of thousands of which are unavailable on leading commercial microarrays. Presented here are the array design process and the results of experiments carried out to demonstrate the array's utility for use in biomarker discovery projects with using NSCLC and normal samples.

Predicting prostate cancer risk through incorporation of a molecular biomarker: PCA3

Donna Pauler Ankerst[a,b], Jack Groskopf[c], John R. Day[c], Amy Blase[c], Harry Rittenhouse[c], Brad H. Pollock[b], Cathy Tangen[d], Dipen Parekh[a], Robin Leach[a,e,*] and Ian Thompson[a]
[a]*Department of Urology, University of Texas Health Sciences Center, San Antonio, TX, USA*
[b]*Department of Epidemiology and Biostatistics, University of Texas Health Sciences Center, San Antonio, TX, USA*
[c]*Gen-Probe, Incorporated, San Diego, CA, USA*
[d]*Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA*
[e]*Department of Cellular and Structural Biology, University of Texas Health Sciences Center, San Antonio, TX, USA*

Corresponding author.

Purpose: The online Prostate Cancer Prevention Trial (PCPT) Risk Calculator combines prostate-specific

antigen (PSA), digital rectal examination (DRE), family and prior biopsy history, age and race to determine the risk of prostate cancer. This work incorporates the biomarker Prostate Cancer Gene 3 (PCA3) into the PCPT Risk Calculator.

Materials and methods: Methodology was developed to incorporate new markers for prostate cancer into the PCPT Risk Calculator based on likelihood ratios calculated from separate case control or cohort studies. The methodology was applied to incorporate the marker PCA3 into the Risk Calculator based on a cohort of 521 men who underwent prostate biopsy with measurements on urinary PCA3, serum PSA, DRE and prior biopsy history. Updated cancer risks incorporating PCA3 (posterior risks) were compared to PCPT risks, PSA and PCA3 by cross-validated operating characteristics: area underneath the receiver operating characteristic curve (AUC), sensitivity, specificity, positive and negative predictive value (PPV, NPV).

Results: The AUC of posterior risks (AUC = 0.703, 95% confidence interval [CI] 0.654–0.753) outperformed PSA (AUC = 0.554, 95% CI = 0.500–0.609) and PCPT risks (AUC = 0.618, 95% CI 0.561–0.676) (both $p < 0.025$), but did not statistically significantly outperform PCA3 alone (AUC = 0.678, 95% CI 0.624–0.732; $p > 0.05$). Sensitivities, PPV and NPV of posterior risks were higher than PCA3, PSA and PCPT risks.

Conclusions: New markers for prostate cancer can be incorporated into the PCPT Risk Calculator by a novel approach. Incorporation of PCA3 improved the diagnostic accuracy of the PCPT Risk Calculator.

Quantitation of TMPRSS2: ERG fusion transcripts in normal prostate and prostate cancers (PC)

Dean A Troyer[a,d,*], Hongyin Bu[a], Yao Wang[a], Augusto Rojas-Martinez[b], Margarita Martinez-Fierro[b], Mario Hernandez[b], Rosario Mendez-Meza[a], Roble G Bedolla[a], Ian M. Thompson[c] and Hongxin Fan[a]
[a]*Department of Pathology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA*
[b]*Department of Biochemistry, Department of Forensic Medicine, Universidad Autonoma de Nuevo Leon, Monterrey, Mexico*
[c]*Department of Urology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA*
[d]*Lakewood Pathology Associates, Lakewood, NJ, USA*

[*]*Corresponding author.*

Introduction: Our objective was to determine if quantitative real time RT-PCR (QR-PCR) is a practical method for detecting TMPRSS2: ERG fusion/translocation in human prostate cancers. ERG gene expression was also assessed.

Methods: 17 tumors and controls were compared. 14 additional controls were from the Medical Examiner's Office in Monterrey, Mexico. Prostate RNA pool (Clontech) was also assayed. Transcripts were detected by either TaqMan (T1/E4 [B. Luxman et al., 2006]) or SYBR Green QR-PCR (T1/E6, [Tomlins et al., 2005]). SYBR Green QR-PCR was used to detect ERG expression (E5-6/E6, [Tomlins et al., 2005]). Sensitivity, accuracy and linearity were determined using serial dilutions of DuCap RNA (Dr. Kenneth J. Pienta; U. of Michigan).

Results: Assays were linear across five orders of magnitude and sensitive to 10^{-6} tumor cells. Using TaqMan method, fusion was detected in 15/17 tumors (fusion/β2M ratios 0.002 to 11.54). 17 benign prostate samples showed fusion/β2M ratios (0.0001–0.014). No fusion was detected in the 14 control prostates. Using SYBR Green method, fusion was detected in tumors (0.0002–15.1 vs. 0.0003–0.08 for benign. Ratios for 14 controls were 0.0002–0.02. ERG expression in tumors was (ratio 0.56–30.2) vs. benign (0.2–1.9; $p = 0.0003$) vs. (0.23–1.43; $p = 0.0003$) in 14 controls.

Conclusions: All 3 QR-PCR methods are rapid, sensitive, and linear. The TaqMan method may be preferable due to its better specificity. A cut off value will be required for the SYBR Green method.

Colon and esophagus reference sets

Melissa K. Tuck[*], Daniel P. Normolle, Sharmila Anandasabapathy, D. Kim Turgeon, John Poneros, Elena Stoffel, Norman Marcon, Henry Appelman, Mack Ruffin, Nadir Arber, Robert Bresalier, Kristen Anton, Mark Thomason, John H. Baron, Sapna Syngal and Dean E. Brenner
Great Lakes New England Clinical Epidemiology and Validation Center
MD Anderson Cancer Center, Houston, TX, USA
Dartmouth Medical College, Hanover, NH, USA
Dana Farber Cancer Center, Boston, MA, USA
St. Michaels Medical Center, Toronto, Ont, Canada
Sourasky Medical Center, Tel Aviv, Israel
University of Michigan, Ann Arbor, MI, USA

[*]*Corresponding author.*

Table 1
Samples in the Colon Reference Set

Diagnosis	Colorectal Adenocarcinoma	Normal control-colonoscopy	Colorectal Adenoma
# of subjects	50	50	50
Total SERUM aliquots	812	918	929
Mean # Serum aliquots $\pm$ SD per subject	16.9 ± 3.7	18.4 ± 2.8	18.6 ± 2.8
Total Plasma Aliquots	749	844	852
Mean # plasma aliquots $\pm$ SD per subject	15.6 ± 5.1	16.9 ± 4.6	17.0 ± 5.2
Total Urine Aliquots	462	523	337
Mean # Urine aliquots $\pm$ SD per subject	11.8 ± 6.4	10.5 ± 4.5	10.5 ± 7.3

Background: A reference set may be defined as a collection of human biosamples that have been collected and documented in a reproducible manner from multiple sites. Reference sets are chosen so that aliquots within the set from individuals can be compared across different assays. Such an approach is powerful because one can compare data from assays from different laboratories. Since biosamples from different subjects are managed according to rigid protocol, data across multiple subjects can be analyzed for a single or multiple analytes. Such methodology permits the dynamic development and predictive assessment of biomarker panels. We describe our strategy for assembling and disseminating reference sets for colorectal adenocarcinoma and esophageal adenocarcinoma risk assessment and early detection.

Objectives: 1. To create good clinical practice (GCP) quality reference sets of biosamples from subjects for colorectal and esophageal adenocarcinoma biomarker discovery and validation; and 2. To provide high-quality well-annotated colon and esophagus sample sets to EDRN researchers for use in pre-validation analyses.

Methods: Each reference set contains a single aliquot (300–500 μl) of a biosample from the same subject. A colon reference set consists of 50 aliquots from 50 different subjects with colorectal adenocarcinoma, colorectal adenomas, or normal colons after colonoscopy. An additional set of samples from 30 subjects with IBD (Crohn's or ulcerative colitis) and 50 subjects with hyperplastic polyps only are available, but are not explicitly part of the reference sets. Samples chosen for the reference set had to meet strict criteria including maximum sample age, subject-specific diagnosis, integrity and completeness of data collection and specimen handling, and minimum number of aliquots available. Secondary considerations included a balance of subjects from sites and by gender. Esophageal reference sets have 50 subjects each from Barrett's metaplasia without dysplasia, adenocarcinoma, or normal after endoscopy. An additional 25 samples are available from subjects with Barrett's with high or low-grade dysplasia. Pre-validation studies proposed using the Colon and Esophagus Reference set are subject to approval by the EDRN GI Collaborative Group.

Results: Table 1 provides the numbers of subjects and total available aliquots for replicate reference sets. There are samples available 15 complete colon reference sets for serum with 150 aliquots, 10 colon reference sets for plasma with 150 aliquots, and 10 colon reference sets for urine with 150 aliquots. DNA and tissue are not part of the reference sets, but may be available from some subjects. Colon reference sets are being assayed for the following biomarkers: CCSA-3, CCSA-4 (Getzenberg); 6 novel antibodies (Milagen, Inc), TRAIL-Based markers (Diadexus); and GOS in stool (Pre-MD). Esophagus reference set is currently being assay for proteomics based biomarkers (Drake/Meltzer).

Conclusions: The GLNE CECV provides high quality, well annotated samples and sufficient data and information about the specimens to ensure detailed interpretation of pre-validation analyses.

Enroute to protein biomarker discovery for early detection of cervical cancer

Gitika Panicker, Vincent A. Emanuele II, Brian M. Gurbaxani, Daisy R. Lee and Elizabeth R. Unger*
Division of Viral and Rickettsial Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA

**Corresponding author.*

Introduction: Cervical mucous may be a good choice for cervical neoplasia biomarker discovery as it is produced in the area of disease. In this study, surface-enhanced laser desorption/ionization time of flight (SELDI-TOF) analysis was used to evaluate parameters for protein profiling of mucous.

Methods: For optimization, total protein was extracted from cervical mucous weck-cels ($n = 6$) collected from women matched for age and race. Experimental parameters included protein extraction reagent, protein concentration, matrix, bind/wash parameters and fractionation. A larger pilot study ($n = 54$) with samples also matched for HPV16 status at a 2:1 ratio for control vs. diseased samples was analyzed using the selected parameters. SELDI-TOF was performed using the Ciphergen PBS-IIc mass spectrometer on different protein chip types.

Results: Analysis of the SELDI spectra of the unfractionated QC sample on varying array types using sinnapinic acid as matrix revealed an average of 30 peaks in the 2.5–30 kDa mass range. The quality control spectra were reproducible with intra- and inter-assay CV for CM10, H50 and Q10 arrays being less than 17% and 28% respectively. IMAC30 chips had a higher intra- and inter-assay CV's at 25% and 35%. The spectra on the H50 chip could be segregated by the presence of blood in the sample. Accounting for the impact of blood, spectra stratified by disease exhibit significant differences in protein profiles on the H50 chip type in the 10-12 and 21-23 kDa range.

Conclusions: Cervical mucous can be used for protein biomarker discovery on the SELDI platform.

Using multiple-reaction monitoring triple quadruple mass spectrometry for the confirmation of putative biomarkers

Keith Waddell*, Ning Tang, Christine Miller and Hongfeng Yin
Agilent Technologies, Santa Clara, CA, USA

**Corresponding author.*

Biomarker discovery using mass spectrometry is now well established. As a consequence of this, it is logical that mass spectrometry should also be used for confirmation of candidate biomarkers. Multiple-reaction monitoring (MRM) on a triple quadruple (QQQ) mass spectrometer provides superior sensitivity and selectivity for targeted compounds in a complex sample. MRM also offers high precision in quantitation and fast scan speed, which makes it an ideal technology for validating biomarkers in a high-throughput fashion. In this study, we tested the confirmation methodology using peptides from proteins present in immunodepleted human serum.

The selected proteins were digested *in silico* using Spectrum Mill Peptide Selector to predict the peptides and their optimum MS/MS product ions. These predicted results were then compared to experimental results from digests of the standard proteins and the lists of MRM transitions were then created. The plasma samples were analyzed by robust and reproducible nanoflow LC/MS using the HPLC-Chip/MS interfaced to a high performance QQQ mass spectrometer. Multiple peptides from each protein were used for the MRM acquisition parameters. The accuracy and precision for the quantitation will be reported to demonstrate the utility of this workflow for validation of putative biomarkers.

A novel approach for cancer diagnosis and prognosis using omics data

Jin Wang[a,*] and Peter He[b]
[a]*Department of Chemical Engineering, Auburn University, AL, USA*
[b]*Department of Chemical Engineering, Tuskegee University, AL, USA*

**Corresponding author.*

Currently abundant omics data (microarray or proteomic profiles) have been made available for cancer diagnosis and prognosis. However, how to extract useful information accurately and consistently remains challenging due to the unique characteristics of the omics data, namely large within-group variations, limited number of samples and modest between-group differences. Although various statistical analysis and pattern classification methods have been applied to analyzing omics data, these challenges have not been fully addressed.

This work develops a novel approach that helps addressing the above-mentioned challenges and provides consistent performance in cancer diagnosis and prognosis using omics data. The developed method is based on the recently developed theory that the robustness of the biological organization is implemented through network structure and complex feedback control loops, rather than through the use of precision components. Based on the theory, we hypothesize that although gene expression or protein levels may vary significantly between different individuals, the correlations among different mRNA or protein levels are tightly controlled due to the robustness of biological systems. Therefore,

this work proposes to identify the significant between-group correlation changes, instead of level changes, caused by cancer as the criteria for cancer diagnosis and prognosis.

In this study, we show that the variations of the correlations among different mRNAs/proteins are much smaller compared to the variations of mRNAs/proteins levels themselves, which supports our hypothesis. In addition, based on the hypothesis, we develop algorithms to identify significant correlation changes and classification algorithms for cancer diagnosis and prognosis. The performance of proposed methods is evaluated using several serum proteomic datasets. Sensitivity, specificity and reproducibility of the developed algorithms are compared with other published microarray/proteomic analysis methods.

Use of urinary metabolomics to identify biomarkers for kidney cancer

Robert H Weiss[a,d,*], Kyoungmi Kim[b] and Vladimir Tolstikov[c]
[a]*Nephrology Center, Metabolomics Core, University of California, Davis, CA, USA*
[b]*Biostatistics Center, Metabolomics Core, University of California, Davis, CA, USA*
[c]*Genome Center, Metabolomics Core, University of California, Davis, CA, USA*
[d]*Medicine, VA Medical Center, Sacramento, CA, USA*

[*]*Corresponding author.*

The diagnosis of cancer by examination of the urine has the potential to improve patient outcomes by means of earlier detection of this disease prior to symptomatic or imaging diagnosis. Due to the fact that the urine contains metabolic signatures of many biochemical pathways, this biofluid is ideally suited for metabolomic analysis, especially involving diseases of the kidney and urinary system. We utilized 32 pre-operative urine samples from two institutions as a training set and 15 samples from one of the institutions as controls. The samples were analyzed by GC- and LC-MS and significant features were identified by MarkerView 1.1 software. The GC-MS data were analyzed using two pattern recognition methods, principal component analysis (PCA), and partial least square discriminant analysis (PLS-DA), to identify metabolites responsible for the classification. 28 metabolites were putatively identified by LC-MS/MS. Current work is ongoing to confirm identities of metabolites and employ a blinded test set of 9 RCC samples and 9 controls to ascertain whether RCC can be differentiated from control patients using the putative biomarkers, either singly or in combination. In future work, these potential biomarkers will be further validated with a larger patient cohort and will include controls with non-cancer renal disease and non-renal cancer to increase specificity; such investigation will likely lead to clinically applicable assays for earlier diagnosis of RCC as well as other malignancies, and thus improved patient prognosis.

NCI-EDRN validation of salivary oral cancer biomarkers

Jianghua Wang[a], Jean Reiss[b], Bernhard Zimmermann[a], Hui Zhou[a], David Elashoff[c], Daryl Morris[d], Tim Randolph[d], Ziding Feng[d], David Chia[b] and David T. Wong[a,*]
[a]*School of Dentistry and Dental Research Institute, University of California, Los Angeles, CA, USA*
[b]*David Geffen School of Medicine, University of California, Los Angeles, CA, USA*
[c]*School of Public Health, University of California, Los Angeles, CA, USA*
[d]*Fred Hutchinson Cancer Center, Seattle, WA, USA*

[*]*Corresponding author.*

Background: We have previously identified a panel of salivary mRNA biomarkers (SAT, DUSP1, IL1B, IL8, S100p, H3F3A, OAZ1) for oral cancer detection and have demonstrated high specificity and sensitivity of these biomarkers in the training set. Additionally, the clinical behavior have been consistently demonstrated in three independent validations involving over 220 subjects, including a Phase 2 biomarker validation project with the Biomarker Reference Laboratory (BRL) of EDRN at UCLA reported here.

Methods: The levels of 7 salivary biomarkers for oral cancer detection were simultaneously quantified by EDRN BRL and our lab in anonymized aliquots of saliva samples ($n = 60$) with a novel multiplex RT-PCR and real-time PCR protocol. Data were analyzed by statisticians at EDRN Data Management and Coordinating Center. Areas under the curves (AUC) for the combined markers were generated using a logistic regression model.

Results: 1) The expression pattern and ranking of prediction power of these oral cancer markers were

almost identical between these two labs. Three genes (IL8, IL1B and SAT) for EDRN-BRL and all 7 from the UCLA lab, were significantly elevated in oral cancer samples ($p < 0.05$). 2) AUC (7 marker combined) of 0.85 and 0.86 were achieved by EDRN-BRL and the UCLA lab, respectively.

Conclusion: The prediction power of 3 markers was validated by EDRN lab, and standardized assay of these salivary oral cancer biomarkers were consistent and reproducible between labs. A second step EDRN validation ($n = 100$) is currently underway.

Supported by PHS NIH Grant R01 DE15970 (D.T.W.) and NIH/National Cancer Institute 5U24 CA08636607 (D.C.).

Quantitation of HER2 and telomerase biomarkers in solid tumors with IgY antibodies and nanocrystal detection

Yan Xiao[a,*], Xiugong Gao[b], Gallya Gannot[c], Michael R. Emmert-Buck[c], Sudhir Srivastava[d], Paul D. Wagner[d], Michael D. Amos[e] and Peter E. Barker[e]
[a]*NIST-NCI EDRN Cancer Biomarker Reference Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA*
[b]*Translabion, Clarksburg, MD, USA*
[c]*Pathogenetics Unit, Laboratory of Pathology and Urologic Oncology Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA*
[d]*Cancer Biomarker Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD, USA*
[e]*Chemical Science and Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA*

[*]*Corresponding author: E-mail: yan.xiao@nist.gov.*

In an effort to improve affinity biomarker validation in fixed patient tissue specimens, we have developed a novel quantum dot-based bioimaging system that utilizes chicken IgY antibody for high sensitivity and specificity relative quantitation of cancer proteins. Monospecific, polyclonal IgYs were generated against human HER2 and telomerase [1], and analytically validated for specificity by western blot and immunohistochemistry on tumor and normal cells and for relative affinity by layered peptide array (LPA). IgYs bound desired targets in cell lines and fixed tissues and showed greater affinity than commercial mammalian antibodies for both HER2 and telomerase proteins. In tissue microarray experiments, HER2 quantitation with IgY antibody and quantum dot imaging correlated well with chromogenic *in situ* hybridization (CISH), whereas telomerase quantitation suggested a trend toward correlation with prostate cancer Gleason Grade and differentiation. Although patient numbers were small, these findings demonstrate the feasibility of relative quantitation of cancer biomarkers with IgY and quantum dot fluorophores, and show promise for rigorous clinical validation in large patient cohorts.

Reference

[1] Y. Xiao, IgY antibodies to human telomerase reverse transcriptase, Patent application filed, June 26, 2006, attorney docket number 000479.00163.

Removal of high molecular weight DNA by carboxylated magnetic beads enhances the detection of mutated K-ras DNA in urine

Ying-Hsiu Su[a,*], Janet Song[a], Zhili Wang[b], Xiao-He Wang[a], Mengjun Wang[a], Dean E. Brenner[c] and Timothy M. Block[a]
[a]*Department of Microbiology and Immunology, College of Medicine, Drexel University, Doylestown, PA, USA*
[b]*Nucleic Acid/Protein Research Core Facility, The Children's Hospital of Philadelphia, Philadelphia, PA, USA*
[c]*Departments of Internal Medicine and Pharmacology, 2150B Cancer and Geriatrics Center, University of Michigan Medical Center, Ann Arbor, MI, USA*

[*]*Corresponding author.*

We have previously demonstrated that mutated DNA derived from the circulation can be detected in urine and predominantly exists as <1 kb DNA fragments. To preferentially isolate the trans-renal DNA from urine, we developed a method using carboxylated beads to separate high MW (1 kb or larger) from low MW urine DNA. A primer set for 18s rRNA (generating a PCR product of 872 bp) was designed and optimized for real-time PCR quantification of high MW DNA templates. To evaluate the method, urine samples from 5 volunteers with no known diseases and 36 patients with various colorectal diseases were collected and tested. It was found that the average removal efficiency of high MW

DNA from total urine DNA using carboxylated beads is 92.72% $\pm$ 1.42%. Furthermore, compared with using total urine DNA, our method provides a greater ability to detect mutated *K-ras* in the urine of colorectal cancer patients. The concurrence of *K-ras* mutations detected in disease tissue and the corresponding urine specimen is significantly higher ($p = 0.0015$) when the samples were enriched in low MW DNA.

Genome-wide microarray analysis of CpG island methylation

A. Bhattacharjee*, K. Pandit, A. Wong, S. Giles and D. Roberts,
Agilent Technologies, Santa Clara, CA, USA

**Corresponding author.*

CpG islands are stretches of high GC content DNA containing multiple CpG dinucleotides. When CpG dinucleotides within these islands are methylated, especially in promoter regions, expression of the corresponding downstream genes is often repressed. Aberrant CpG island methylation is implicated in cancer. We have refined a method for methylated DNA immunoprecipitation (mDIP) and coupled it with microarray detection. DNA isolated by mDIP is fluorescently labeled and hybridized to an oligonucleotide microarray that specifically represents the unique CpG islands in the human genome. This microarray contains $\sim$237,000 oligo probes tiling $\sim$20,000 CpG islands, with an average spacing between probes of 95 base pairs. As proof of concept we performed mDIP and microarray analysis of human genomic DNA samples from normal tissue. We compare our mDIP array data with bisulfite sequencing data from the human epigenome project. We demonstrate the ability of our array based methylation assay to distinguish probes corresponding to methylated regions from probes corresponding to unmethylated regions. We extend these observations to the island level and identify methylated CpG islands. We then applied the whole-genome assay to cancer cell lines and tumor DNA. We describe differential methylation patterns in these tissues and identify previously known and novel methylation markers.

Serum mesothelin for early detection of the asbestos-induced cancer malignant mesothelioma

Jenette Creaney[a,b,c,*], Bruce W.S. Robinson[a,b,c], Amanda Segal[a,d], Arthur W. Musk[a,c], Nick de Klerk[a,e], Harvey I. Pass[f] and Steven J. Skates[g]
[a]*National Research Centre for Asbestos Related Disease, Western Australian Institute of Medical Research, WA, Australia*
[b]*Australian Mesothelioma Tissue Bank, Sir Charles Gairdner Hospital, Perth, WA, Australia*
[c]*School of Medicine and Pharmacology, University of Western Australia, Queen Elizabeth, Medical Centre, Perth, WA, Australia*
[d] *PathWest Laboratory Medicine WA, Queen Elizabeth II Medical Centre, Perth, WA, Australia*
[e]*Biostatistics and Epidemiology Unit, Telethon Institue for Child Health Research, Hay St, Subiaco, WA, Australia*
[f]*NYU School of Medicine and Comprehensive Cancer Center, New York, NY, USA*
[g]*Biostatistics Center, Massachusetts General Hospital, Boston, MA, USA*

**Corresponding author.*

Malignant mesothelioma is an aggressive, asbestos-induced tumour which is uniformly fatal. Generally it presents at relatively late stages, so an early diagnostic test is crucial if effective therapy is to be found. Individuals exposed to asbestos through their employment or residences in Wittenoom in the north-west of Western Australia have taken part in a cancer prevention program since 1994. These individuals were reviewed annually and a blood sample collected. In a subset of this cohort we have measured serum levels of the biomarker mesothelin using the MESOMARKTM assay kindly provided by Fujirebio Diagnostic Incorporated (Malvern PA).

Mesothelin levels were determined from the archived serum samples of 106 individuals who developed malignant mesothelioma. There were seven women in this group, the average age at mesothelioma diagnosis was 66 years (range 32–84 years) and the median survival following diagnosis was 9 months (range 0–122 months). On average 8 annual pre-diagnosis serum samples were available from the individuals who subsequently developed mesothelioma; with the penultimate C sample being less than 18 months before diagnosis in 80% of these 106 individuals.

As a control group mesothelin levels were determined in the serum of 99 randomly-selected individu-

als with asbestos-exposure who after approximately 10 years of follow-up have not developed any malignancy; on average of 8 annual samples were available from each individual in this group.

Mesothelin levels were statistically higher in the penultimate pre-morbid sample of people who subsequently developed mesothelioma compared to the asbestos-exposed controls ($p < 0.0001$). Using the historically defined cut-off of 2.5 nM mesothelin, 14 of the 106 individuals who developed mesothelioma had positive samples before diagnosis. We are currently evaluating changes in longitudinal mesothelin levels to determine if the sensitivity of this biomarker in a screening setting can be improved.

A multi-center, double-blinded pre-validation study of methylation biomarkers for progression prediction in Barrett's esophagus

Zhe Jin*, Yingye Zheng, Jesse Gu, Yulan Cheng, Alexandru Olaru, Bogdan Paun, Herbert Wolfsen, Nicholas Shaheen, Kay Washington, Yvonne Romero, Kenneth K. Wang, Ziding Feng, Richard D. Sampliner and Stephen J. Meltzer
GI Division, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA
Department of Biostatistics, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
GI Division, Department of Medicine, University of Arizona, Tucson, AZ, USA
GI Division, Department of Medicine, Mayo Clinic Foundation, Rochester, MN, USA
Department of Pathology, Vanderbilt University, Nashville, TN, USA
GI Division, Department of Medicine, University of North Carolina, Chapel Hill, NC, USA
GI Division, Mayo Clinic Foundation, Jacksonville, FL, USA

Corresponding author.

Adenocarcinoma risk in Barrett's esophagus (BE) is increased 30-fold over the general population, but progression in BE is relatively rare. Biomarker-based prediction models would be useful in stratifying patients for more efficient surveillance endoscopy. We performed a multi-center, double-blinded prevalidation study of a BE progression prediction model based on 8 methylation biomarkers. Progressors were categorized into two tiers: progression within 2 (tier 1) or 4 years (tier 2). Pre-progression BE biopsies were obtained from 5 participating institutions. Methylation was assayed in 124 nonprogressors (NP) and 32 progressors (P) by real-time quantitative methylation-specific PCR. P *vs.* NP did not differ significantly in terms of BMI, length of BE segment, gender, percentage with LGD, family history of LGD/HGD/EAC, cigarette smoking, alcohol use, or use of acid-inhibitory drugs; however, P were significantly older than NP (68.9 *vs.* 61.9 years, $p = 0.004$). We then evaluated a linear combination of the 8 markers, using coefficients from a multivariate logistic regression model. Area under the ROC curve (AUC) was high in the 2-tier model (0.865 and 0.862; $p < 0.001$ and $p < 0.001$, respectively). Even after rigorous correction for potential overfitting, based on 3-fold cross-validation repeated 200-fold, AUCs remained high and AUC shrinkage was minimal (AUC $= 0.798$ and 0.788; Δ-AUC $= 0.074$ and 0.067, respectively). Since age was not balanced in P *vs.* NP, we explored the incremental AUC value contributed by (8-marker panel plus age) *vs.* age alone. Even after correcting for overfitting, AUC values remained high in the 2-tiered model (0.800 and 0.819), and incremental AUC values contributed by the 8-biomarker+age panel *vs.* age alone were substantial (Δ-AUC $= 0.182$ and 0.162). A reduced panel containing only 5 markers selected by logistic regression with stepwise forward selection yielded only slightly smaller AUCs and Δ-AUCs. These findings suggest that a large-scale validation study of a methylation biomarker-based Barrett's neoplastic progression prediction model is warranted. A reduced biomarker subset can be considered for this large-scale validation study.

A sequential fractionation workflow with immobilized trypsin paramagnetic beads for expression profiling and identification of serum proteins by LIFT-MALDI-TOF/TOF

Izabela D. Karbassi, Lisa Cazares, O. John Semmes and Richard R. Drake*
Center for Biomedical Proteomics, Eastern Virginia Medical School, Norfolk, VA, USA

Corresponding author.

Matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry offers the capability of higher throughput profiling analysis of clinical samples like serum. Many profiling strategies typi-

cally employ single chemical affinity beads or surfaces to decrease sample complexity. However, a majority of the intact proteins fractionated in this manner are not detected in the lower mass ranges where MALDI-TOF mass spectrometers are most effective. We describe an expression profiling workflow that incorporates front-end affinity bead capture in tandem with immobilized trypsin paramagnetic bead digestions. This approach allows proteins bound to the capture resin to be reduced to peptides that fall into the ideal range for profiling and direct LIFT-MALDI-TOF/TOF sequence determinations. The bead-based trypsinization method was highly reproducible and efficient in digesting large serum protein fractions with short incubation times, and the resulting peptides were readily identifiable by LIFT-MALDI-TOF/TOF. Additionally, since the fractionation and trypsinization steps are performed via paramagnetic particles, this method can be automated using a robotic liquid handling platform. As proof-of-concept for clinical applications, the method was used in a serum profiling study with the goal of detecting differences between individuals diagnosed with benign prostatic hyperplasia (BPH) and those with prostate cancer, stratified to PSA levels. Using weak cationic magnetic beads in tandem with the bound trypsin beads, we were able to identify Apolipoprotein A-IV as elevated in individuals with BPH. This result was reproducible and additionally confirmed by a workflow using weak anionic beads.

Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot

Yinghua Qiu, Tasneem H. Patwa, Li Xu, Kerby Shedden, David E. Misek, Mack T. Ruffin, Dean E. Brenner and David M. Lubman*
The University of Michigan Medical Center, Ann Arbor, MI, USA

*Corresponding author.

Colorectal cancer (CRC) remains a major worldwide cause of cancer-related morbidity and mortality largely due to the insidious onset of the disease. The current clinical procedures utilized for disease diagnosis are invasive, unpleasant and inconvenient; hence the need for simple blood tests that could be used for the early detection of CRC. In this work, we have developed methods for glycoproteomics analysis to identify plasma markers with utility to assist in the detection of colorectal cancer (CRC). Following immunodepletion of the most abundant plasma proteins, the plasma $N-$linked glycoproteins were enriched using lectin affinity chromatography and subsequently further separated by nonporous silica reverse phase (NPS-RP)-HPLC. Individual RP-HPLC fractions were printed on nitrocellulose coated slides which were then probed with lectins to determine glycan patterns in plasma samples from 9 normal, 5 adenoma, and 6 colorectal cancer patients. Statistical tools, including principal components analysis, hierarchical clustering, and Z-statistic analysis, were employed to identify distinctive glycosylation patterns. Patients diagnosed with colorectal cancer or adenomas were shown to have dramatically higher levels of sialylation and fucosylation as compared to normal controls. Plasma glycoproteins with aberrant glycosylation were identified by nano-LC MS/MS, while a lectin blotting methodology was used to validate proteins with significantly altered glycosylation as a function of cancer progression. The potential markers identified in this study for diagnosis to distinguish colorectal cancer from adenoma and normal include elevated sialylation and fucosylation in complement C3, histidine-rich glycoprotein, and kininogen-1. These potential markers of colorectal cancer were subsequently validated by lectin blotting in an independent set of plasma samples obtained from 10 CRC patients, 10 patients with adenomas and 10 normal subjects. These results demonstrate the utility of this strategy for the identification of $N-$linked glycan patterns as potential markers of CRC in human plasma, and may have the utility to distinguish different disease states.

Fox Chase Cancer Center breast cancer reference set construction

P. Engstrom*, M. Daly and A. Godwin
Fox Chase Cancer Center (*FCCC*), *Philadelphia, PA, USA*

*Corresponding author.

The objective of this study is to assemble a well-characterized panel of blood specimens from subjects with different risks for invasive breast cancer. The panel of blood specimens will be used to validate the sensitivity and specificity of potentially useful breast cancer biomarkers. The samples will be obtained from healthy women who are registered for a screening mammogram

at FCCC. The goal is to collect blood samples from women prior to a tissue diagnosis of invasive cancer, ductal cancer in situ, lobular cancer in situ, and benign breast conditions. Women with no cancer or precancer diagnosis and normal mammograms will serve as controls. All subjects will be at least 18 years of age but have no prior biopsy diagnosis of breast cancer or benign breast disease, no prior breast surgery, radiotherapy or chemotherapy, and not pregnant or nursing. Subjects complete a survey which details demographics, breast health, reproductive history, family history, and personal cancer history. Blood samples (28 ml.) are collected prior to surgery or anesthesiology and are separated into serum, plasma, and white blood cells for storage at–80c. Between October 2006 and February 2008; 2,057 women were contacted and 604 eligible women consented for the protocol. There have been 5 cases of invasive cancer, 1 DCIS, 1 LCIS, 11 non-proliferative histologies, and 2 proliferative histologies.

Discovery of antibody based biomarkers for diagnosis of non small-cell lung cancer

D.L. Blum[a,*], T.V. Pedchenko[b], D.V. Parekh[a], P.P. Massion[b] and R.L. Mernaugh[a]
[a]*Department of Biochemistry Vanderbilt University Medical Center, Nashville, TN, USA*
[b]*Department of Allergy, Pulmonary and Critical Care Medicine Vanderbilt University Medical Center, Nashville, TN, USA*

[*]*Corresponding author.*

Non small-cell lung cancer (NSCLC) is the most prevalent form of lung cancer and, if not detected early, has a low cure rate. Autoantibodies can be produced as part of an immune network antibody response to NSCLC. As such, the autoantibodies can serve as biomarkers in immunoassays for use in disease diagnosis or prognosis. A large phage-displayed scFv recombinant antibody library was used to obtain 50 soluble scFv antibodies that bound to IgG obtained from pooled NSCLC adenocarcinoma or squamous carcinoma patient samples, but bound only slightly or not at all to IgG obtained from pooled normal patient samples. High throughput antibody printing was then used to assay individual scFv against 200 individual NSCLC and normal patient samples to determine scFv specificity for lung cancer immune network IgG. Twelve scFv, identified by antibody printing, discriminated IgG obtained from individual NSCLC and normal patients. The scFv, when assayed by ELISA, yielded similar results. By ELISA, we also identified 3 scFv that preferentially bound to normal patient serum proteins and not cancer patient serum proteins. The methodologies developed during this study will serve as a platform for future biomarker discovery studies using greater numbers of scFv on a large cohort of NSCLC patient samples.

Ultra-sensitive detection of cancer biomarkers by the liposome polymerase chain reaction

Junkun He[a], Timothy J. O'Leary[b] and Jeffrey T. Mason[a,*]
[a]*Armed Forces Institute of Pathology, Rockville, MD, USA*
[b]*Veterans Health Administration, Washington, DC, USA*

[*]*Corresponding author.*

The early detection of biomarkers for diseases, such as cancer, remains a significant challenge in clinical diagnostics. Although the polymerase chain reaction (PCR) can detect trace copies of nucleic acids, there is no equivalent clinical assay for proteins. Attempts to couple PCR to protein detection using immuno-PCR methods have failed, in large part, because of the unacceptably high levels of false positives that result from contamination by the antibody-coupled DNA template (amplicon) used as the PCR substrate in the immuo-PCR assay. Here we report the development of an ultra-sensitive diagnostic method, called liposome polymerase chain reaction (LPCR), which has been designed to circumvent this problem. In LPCR, the detection reagent is a unilamellar liposome with amplicons encapsulated inside the liposomes and N-biotin labeled polyethylene glycol (PEG) polymers incorporated into the outer bilayer leaflet of the liposome. As in a conventional enzyme-linked immunosorbent assay (ELISA) the protein target is immobilized inside a microplate well by a capture antibody followed by the addition of a biotinylated second antibody. The biotin-labeled liposome detection reagent is then coupled to the second antibody through a neutravidin bridge. Any amplicons located outside the liposomes, which are responsible for the background signal, are then degraded by treatment with DNase I followed by inactivation of the nuclease with heat. Throughout this process,

the amplicons inside the liposomes are protected because the DNase I cannot permeate the bilayer. The liposomes are then lysed with Triton X-100 to release the amplicons, which are quantified by real-time PCR. We have developed an LPCR assay to quantify carcinoembryonic antigen (CEA) spiked into human serum and have achieved a detection threshold of 10^{-16}M (15 fg/ml or $\sim$6,000 molecules for a 100μl sample), which is 10,000 more sensitive than the current clinical assay for this cancer biomarker. Interestingly, the assay sensitivity was limited by the binding properties of the antibody pair and not the assay noise threshold, which would have allowed for a theoretical sensitivity of 10^{-18} M. The LPCR assay can be used to detect other cancer biomarkers by simply changing the capture and biotinylated antibodies. Accordingly, we believe that the LPCR assay method shows great promise as a clinical diagnostic tool for the early detection of cancer biomarkers.

A visual programming platform for diagnostic workflows based on a combination of protein and gene expression profiles

Maciek Sasinowski
INCOGEN, Inc.

The use of mass spectrometry (MS) and gene expression microarray (MA) technologies for clinical applications has extraordinary potential for accurate, early, and minimally invasive diagnoses of complex diseases, such as cancer. The project team is developing an integrated, modular environment for combining MA and MS data that provides interaction between the three major components for data classification: signal processing, profile construction, and classification and validation. The software is based on INCOGEN's VIBE application and provides users with a graphical environment to construct, combine, and optimize MS and MA data analysis pipelines. The application is written in Java and is comprised of algorithms and wrappers for algorithms written in other languages such as R or C. The data can be loaded from a variety of sources and formats and subjected to several types of signal processing algorithms (e.g., background subtraction, smoothing, peak picking, and/or peak alignment). Following the signal processing routines, the data can be analyzed with variable selection and classification algorithms (e.g., LDA, tree, K-nearest-neighbor) and the results can be viewed graphically or exported as text-based files.

The goal of the software is to significantly enhance the effectiveness of both MS and MA profiling research by allowing researchers to: (1) quantitatively characterize instrumental noise and determine optimal experimental protocols that enhance diagnostic sensitivity, (2) identify diagnostically significant features (biomarkers) in the data and allow unequivocal association of those features with organ or cell-related disease outputs, and (3) select and cross-validate a classification scheme that optimizes diagnostic classification rates.

This effort is funded by the NIH National Cancer Institute Early Detection Research Network (EDRN) and Biomedical Information Science and Technology Initiative (BISTI) programs.

Biomarker source for advanced prostate cancer: The LuCaP xenografts

Laura E. Pascal, Ricardo Z. Vêncio, Robert L. Vessella and Alvin Y. Liu*
University of Washington and the Institute for Systems Biology, Seattle, WA, USA

**Corresponding author.*

Introduction: The University of Washington series of LuCaP xenografts display a wide range of characteristics and maintain a histological appearance consistent with the tumor implants. Their histology appeared unchanged in serial passages. Cluster designation (CD) cell surface antigen phenotyping shows that they could represent the different CD cancer cell types in prostate tumors. They also differ from the predominant cancer cell type found in a majority of primary tumors. Secreted proteins made by these cancer cells, when identified, are promising biomarkers for lethal prostate cancer.

Methods: Xenografts representative of either adenocarcinoma or small cell cancer were harvested from mice. Tumor samples were analyzed by Affymetrix DNA arrays for their transcriptomes. A robust data analysis software tool, HTself, was used to identify differentially expressed genes. The mouse blood of the sacrificed animals that might contain proteins released by the implanted human tumors was saved for proteomics.

Results: The over 20 established LuCaP xenografts all produced ample amounts of PSA, detectable in the mouse sera. LuCaP 35 and LuCaP 49 were harvested at 5 different passages and profiled by HG-U133 GeneChips. LuCaP 35 was positive for AR and

KLK3/PSA, and MME/CD10 in agreement with other types of gene expression analysis. LuCaP 49, a small cell cancer, had low expression of these genes but was positive for CD133/PROM1, a stem cell marker. Overall, the expression pattern appeared to be maintained for the different passages analyzed. A large degree of similarity in expression was seen between LuCaP 35 and LNCaP, reflecting that both were derived from node metastasis and were $CD10^+$, which is the predominant CD cancer cell type found in positive nodes. An example of gene expression difference between LuCaP 35/LNCaP and $CD26^+/CD10^-$ cancer cells of a Gleason 3 primary tumor was the gene family encoding cancer/testis antigens MAGE and GAGE. MAGE expression in circulating epithelial cells was detected significantly more in patients carrying a higher risk of disease recurrence than those with a lower risk. Patients with metastasis had MAGE expression compared to those with localized disease (Kufer et al., *Cancer Res* 2002,62:251).

Conclusions: Cancer CD phenotyping and gene expression analysis showed that the LuCaP xenografts were representative of progressed cancer and none were representative of the primary cancer cell type. This was not surprising as these xenografts were established from metastatic lesions of lymph node, bone, fat, liver, and intestine. Genes associated with stem cells such as CD44 and CD133 were also detected in these xenografts. Expression of these two genes is infrequent in cells of primary tumors. A useful byproduct of the study is the blood of the sacrificed animals. Secreted proteins of the xenografts (e.g., PSA) may enter the mouse circulation and be detected by quantitative proteomics. These proteins are to be identified by array data analysis.

Mass spectrometric analysis of a reference material for cancer diagnostics

Yan Li*, Lori Sokoll, Daniel W. Chan and Hui Zhang
Department of Pathology, Johns Hopkins University, Baltimore, MD, USA

**Corresponding author.*

Recently, the early detection research network (EDRN) of the National Cancer Institute (NCI), developed a serum reference material intended to be used by laboratories for cancer proteomics research, for providing standardization across the proteomics research community, and for the validation of analytical instrumentation. This pilot reference material was spiked with several proteins (FDA approved biomarkers) to simulate the cancer disease state. Using this reference material, we report here the mass spectrometric analysis of five clinical tumor markers: prostate-specific antigen (PSA), carcinoembryonic antigen (CEA), Her-2, chorionic gonadotropin (HCG), and CA125. The five tumor markers were first digested using trypsin and identified by LC-MS/MS. In addition, N-linked glycopeptides of each protein were isolated using Solid Phase Extraction of N-linked glycopeptides (SPEG) since the glycopeptide isolation largely reduced the sample complexity and significantly improved the analytical sensitivity The extracted glycopeptides from each protein detected by LC-MS/MS and N-linked glycopeptides were identified from five tumor markers. The results of these experiments provided a list of candidate peptides for targeted detection of these tumor markers. In addition, heavy-isotope-labeled-peptide standards can be synthesized using the identified peptide sequences from this study and used for development of quantitative analysis of marker proteins in clinical specimens using a highly specific, sensitive and high-throughout MS-based analysis method. Therefore, this pilot reference material could be used by the cancer research community in the standardization of proteomics technologies.

Bayesian and information theories applied to feature selection and classification of MS data

Karl W. Kuschner
Department of Physics, College of William and Mary, Williamsburg, VA, USA

Our group at the College of William and Mary, in collaboration with Eastern Virginia Medical School, wishes to enhance high throughput mass spectrometry (MS) methods for identifying mass values (and hence protein molecules) in human tissue that are discriminative of some particular disease. We have examined wrapper techniques for finding diagnostic variables, selecting specific mass value "features" with a Naïve Bayesian Classifier (NBC). The resulting feature sets allowed the NBC to classify disease groups to above 90% accuracy. However, this type of feature selection, as well as more common statistical methods, have problems such as small sample effects that can cause incorrect feature selection. In addition, the physical characteristics of

the MS instrument causes single proteins to show up in highly correlated, but separate, features.

These features, if included, cause instabilities in classifiers; if excluded, we lose important physical information about the system. Our current work uses information theory-based techniques such as mutual information and conditional mutual information to discover correlated features and diagnostic groups of features that we will eventually combine and use as "meta-features" in our analysis. With these "meta-features," we intend to implement a more robust classification method called a Bayesian Network (BN) that allows these features to be linked not only to the disease but to each other and easily handles the problem of the uncertainties in measurements. Our final goal is to create a BN which reflects the biologic processes, has incorporated both measurement uncertainties and system limitations into the feature selection and network structure, and is a stable classifier for the specific disease in question.

Development of a multimarker assay for early detection of ovarian cancer

Zoya Yurkovetsky, Steven Skates, Aleksey Lomakin, Francesmary Modugno, Jeffrey Marks, Andrew Godwin, Elieser Gorelik, Ian Jacobs, Usha Menon, Karen Lu, Donna Badgwell, Robert C. Bast Jr and Anna Lokshin*

**Corresponding author: University of Pittsburgh, Pittsburgh, PA, USA.*

Clinical outcome in ovarian cancer is very likely to be improved by early detection. Eighty four biomarkers putatively associated with epithelial ovarian cancers were analyzed in sera from healthy women and from patients with ovarian cancers (stage I-II and III-IV), benign pelvic tumors, and breast and lung cancers, using multiplex xMAP^{TM} bead-based immunoassays. A training set including sera from ovarian cancer patients (34 stage IA and 43 stage IB-II) and 153 healthy women was analyzed with logistic regression and cross-validation to identify an optimal panel of biomarkers for discriminating cancer cases from healthy controls. The multimarker panel that provided the highest diagnostic power, 92% sensitivity (SN) at 98% specificity (SP) was comprised of 4 biomarkers: CA 125, EGFR, HE4, and sVCAM-1. This model was applied to an independent blinded validation set consisting of sera from 34 patients with stage IA and 52 patients with stage IB-II ovarian cancer, 105 patients with stage III-IV ovarian cancer, and 103 healthy women providing unbiased estimates of 90% SN for stage I-II and 91% SN for stage III-IV cases at 98% SP. This panel was selective for ovarian cancer showing SN $=$ 34% for cases with benign pelvic disease, SN $=$ 5% for breast cancer, and SN $=$ 36% for lung cancer. A panel of 4 biomarkers exhibits sufficient sensitivity and specificity to serve as an initial stage in a screening strategy for epithelial ovarian cancer.

Stem cell marker CD133 (PROMININ-1) is epigenetically regulated in ovarian cancer

Tsukasa Baba[a,b], Patricia Convey[a], Noriomi Matsumura[b], Regina S. Whitaker[a], Tiffany Perry[a], Zhiqing Huang[a], Seiichi Mori[c], Eiji Kondoh[a,b], Rex C. Bentley[d], Shingo Fujii[b,e], Andrew Berchuck[a,c], Susan K. Murphy[a,c] and Jeffrey R. Marks[c,f,*]

[a]*Division of Gynecologic Oncology, Duke University, Durham, NC, USA*

[b]*Kyoto University, Japan*

[c]*Institute for Genome Sciences and Policy, Duke University, Durham, NC, USA*

[d]*Department of Pathology, Duke University, Durham, NC, USA*

[e]*National Hospital Organization, Kyoto Medical Center, Japan*

[f]*Department of Surgery, Duke University, Durham, NC, USA*

**Corresponding author.*

Objectives: The cell surface molecule CD133 (PROM1) distinguishes a number of cancer stem cells *in vitro* and *in vivo*. Increasing evidence suggests that CD133(+) cancer stem cells are involved in chemotherapeutic resistance and/or recurrence, yet little is known about the expression status and clinical significance of CD133 expression in ovarian cancer (OVCA). Our objective was to characterize CD133 expression in serous epithelial OVCA.

Methods: CD133 expression in 25 stage III-IV OVCAs was determined by immunohistochemistry. Cell surface CD133 from 11 OVCA patient's ascitic fluid and for 40 OVCA cell lines was detected by flow cytometry. Flow activated cell (FAC) sorting for cell lines with heterogeneous CD133 expression was used to purify the CD133(+) and CD133(−) cell fractions.

Quantitative bisulfite sequencing was used to interrogate the methylation status of the ovarian tissue-active P2 promoter of CD133.

Results: CD133(+) cancer cells were detected in 5 of 11 ascitic fluid specimens and in 10 of 25 primary OVCAs. Of 40 OVCA cell lines, seven (17.5%) were uniformly CD133(+), eight (20%) were heterogeneous for CD133 expression, and 25 (62.5%) were CD133(−). *CD133* transcription and cell surface CD133 expression are positively correlated in these cell lines ($R2 = 0.63$; $p < 0.0001$). Cell lines with decreased or no cell surface CD133 exhibited DNA methylation for 15 CpG sites at the *CD133* P2 promoter. Increased methylation was associated with decreased *CD133* transcription, indicating epigenetic regulation of *CD133* (Spearman $r = -0.63$, 95% CI, -0.79 to -0.37; $p < 0.0001$). Intriguingly, single FAC-sorted CD133(+) PEO1 OVCA cells undergo asymmetric division to generate CD133(+) and CD133(−) progeny. Daughter CD133(−) PEO1 cells exhibit increased methylation relative to the parental CD133(+) cells (avg. 62% vs. 39%), suggesting active acquisition of *CD133* methylation and transcriptional silencing accompanying cell division. Analysis of 4 primary OVCAs also showed methylation (2–78%) that is inversely correlated with *CD133* transcription ($R2 = 0.68$; $p = 0.17$).

Conclusions: This is the first demonstration that the stem cell marker CD133 is regulated by promoter methylation, and the first comprehensive analysis of CD133 in epithelial ovarian cancer. Further investigations are warranted to define the role of CD133 expression in OVCA and to determine if CD133(+) cells represent ovarian cancer stem cells.

Development and validation of highly sensitive and specific test for diagnosis of endometrial cancer

G.L. Maxwell, Z. Yurkovetsky, A. Lomakin, K. Lu and A. Lokshin*

Corresponding author: University of Pittsburgh, Pittsburgh, PA, USA.

Introduction: It is estimated that 39,080 new cases of endometrial cancer were diagnosed in the US in 2007, making it the most common gynecologic malignancy. At present, there are no early detection tests for endometrial cancer in women who present without symptoms that are typical to endometrial cancer. A sensitive and specific blood-based test could be particularly appealing screening mechanism because it is less invasive and more cost effective than endometrial biopsy for application in populations at varying levels of risk for endometrial cancer, including patients with morbid obesity and genetic predisposition.

Methods: Twenty eight bead-based xMAP^{TM} immunoassays for endometrial cancer serum biomarkers were utilized in this study, TNFRI, IL-6, eotaxin, IL-2R, Cyfra 21-1, sFas, sFasL, CA 125, CA 19-9, CA 72-4, EGFR, IGFBP-1, HE4, VCAM-1, E-selectin, MPO, tPAI 1, MIF, FSH, LH, TSH, prolactin, GH, ACTH, CD40L, adiponectin, MMP-2, MMP-3, and MMP-9. The optimal scoring function (SF) for a given set of markers was constructed using Monte Carlo (MC) optimization based on Metropolis-Hastings algorithm as a linear combination of logarithms of fluorescence intensities. A diagnostic model was created based on the training cross-validation set consisting of 108 endometrial cancer patients and 150.

Results: We have identified a combination of 4 serum proteins, PRL, HE4, eotaxin, and FasL that offered 100% sensitivity at 98% specificity for endometrial cancer in the training set, and 100 sensitivity at 93% specificity in the solely evaluated independent blinded validation set obtained from different sources than the training set indicating the high robustness of the test.

Conclusions: Our data have demonstrated that patients with endometrial cancer have significantly different expression patterns of several serum biomarkers as compared to healthy controls. If this performance of the panel identified in this study holds in asymptomatic subjects, the panel could be used as the first line test in a two-step strategy for early detection where positivity in blood test triggers endometrial biopsy.

Specific profiling of sialylated glycoproteins from breast cancer

Yuan Tian
Johns Hopkins School of Medicine, Baltimore, MD, USA

Glycosylation, one of the most important modifications to proteins, is known of heterogeneity and being varied associated with health state. The heterogeneity of glycosylation was presented by the modification of various sugars and a number of potential glycosylation sites on proteins. The question of which specific glycosylation related to a specific disease becomes urgent to be addressed for early-stage diagnosis and therapy of

diseases. To determine the breast cancer related glycosylation, Lectin microarray was employed in this study. Sialic acid glycoproteins showed consistent changes in breast cancer and normal tissues. A specific isolation of sialylated glycoproteins was developed and reported here to identify and quantify sialylated glycoproteins in breast cancer tissues. The N-linked sialylated glycopeptides were profiled from three paired breast cancer and normal tissues. The patterns generated from N-linked sialylated glycopeptides were compared with the patterns generated from total N-linked glycopeptides and total tryptic peptides from the same specimens. The specific sialylated glycopeptides associated with cancer were determined.

Performance of cancer/testis antigens and EEF1A2 mRNAs as tissue biomarkers for non-small cell lung cancer

W.A. Franklin[a,*], J. Haney[a], M. Sugita[a], J.D. Mitchell[b], M.J. Weyant[b], L. Bemis[c], J. Gu[d] and Z. Feng[d]
[a]*Department of Pathology, University of Colorado Health Sciences Center, Research Center-1 South, Aurora, CO, USA*
[b]*Department of Surgery, University of Colorado Health Sciences Center, Research Center-1 South, Aurora, CO, USA*
[c]*Department of Medicine, University of Colorado Health Sciences Center, Research Center-1 South, Aurora, CO, USA*
[d]*Department of Statistics, Fred Hutchinson Cancer Research Center, Seattle, WA, USA*

[*]*Corresponding author.*

We have interrogated microarray data from lung cancer cell lines to identify genes that are highly expressed in lung cancer cells but are silent in normal bronchial cells. Candidate genes were cloned into TOPO TA plasmids (Invitrogen) to serve as standards in a qRT-PCR assay in which copy numbers for the candidate biomarker genes were normalized to 10^6 copies of β-actin. A panel of biomarkers was selected to directly test sensitivity and specificity in frozen tissue samples from 103 NSCLC, 60 matched non-malignant tissue samples from the same resected lung specimens and 18 normal lung tissue samples from never smokers. Mean copy numbers/10^6 copies β-actin are shown below.

By ROC analysis, maximum AUC for an individual marker was obtained for MAGE A (0.712 tumor vs adjacent lung and 0.784 tumor vs never smoker lung). AUC for the combined five gene panel was 0.885 for tumor versus adjacent lung, 0.957 for tumor vs never smoker lung and 0.945 for adjacent lung vs never smoker lung. A logistic model using percentiles had an AUC of 0.814 for tumor vs adjacent lung. Rigorous marker by marker testing in tumor tissue may yield a specific RNA biomarkers panel that is suitable for validation in early detection studies.

Table 1
Mean Copy #/million b-actin copies (SD)

Marker	Tumor	Adjacent	Never smokers
MAGE A	830,000 (5E6)	15.02 (38.27)	0.61 (1.24)
Xage	5,000,000 (4E7)	400 (1743)	2.56 (6.71)
NY-ESO	200 (800)	28.85 (83.43)	1.61 (2.70)
TEX15	2100 (950)	15.68 (17.17)	4.39 (2.62)
EEF1A2	5300 (2E4)	130 (605)	18.39 (11.05)

Heterogeneous TMPRSS2-ERG expression in human prostate cancers: Insights from a prospective prostate biopsy series

Sandra M. Gaston[a,*], Michael C. Kearney[a], Andrew L. Guerra[a], Heather Conroe[a], Gerardo Trejo[a], Joanna Jung[a], Gerri Bueti[a], Jihad Hayek[b], Juan-Miguel Mosquera[c], Sven Perner[d], Mark A. Rubin[d], Gary P. Kearney[b] and Martin G. Sanda[a]
[a]*Beth Israel Deaconess Medical Center, Boston, MA, USA*
[b]*New England Baptist Hospital, Boston, MA, USA*
[c]*Roger Williams Medical Center, RI, USA*
[d]*Weill Cornell Medical Center, NY, USA*

[*]*Corresponding author.*

Introduction and objective: The discovery that the majority of prostate cancers (PrCa) contain a recurrent genomic rearrangement involving one of the ETS family oncogenes has produced a new class of biomarkers for this disease. Most commonly, a chromosome 21q22.2-3 rearrangement places ERG expression under the control of a prostate-specific, androgen regulated TMPRSS2 promoter. In PrCa, TMPRSS2-ERG fusions show considerable variability, and it's been proposed that specific TMPRSS2-ERG transcript structure(s) may be directly associated with tumor aggressiveness. We analyzed TMPRSS2-ERG at the mRNA and genomic levels in PrCa patients from a prospectively collected needle biopsy cohort. Our goal is to assess the relative contribution of clonal vs transcription-

al heterogeneity in this pivotal molecular event. Methods: Prostate biopsy patients were enrolled in a protocol in which tissue print micropeel "touch preps" were collected from each biopsy core. Tissue cores were processed as usual for pathology; after a diagnosis had been established, additional sections were obtained for FISH analysis of chr 21q22-3 rearrangements. Tissue print micropeels were snap frozen upon collection; purified RNA and DNA fractions were prepared from each print for molecular analysis. Rt-PCR and direct sequencing were used to determine the detailed molecular characteristics of specific TMPRSS2-ERG transcripts. Results: In our biopsy series, approx 35% of the PrCa patients who were positive for TMPRSS2-ERG transcripts showed 2 or more major patterns of expression of this gene fusion at biopsy. Alternate mRNA splicing probably accounts for some of this heterogeneity, but in some cases sequence analysis showed more than one chr 21q22-3 rearrangement, consistent with multiple PrCa initiations (multi-clonal cancer) or with progressive genetic changes within a single PrCa focus (clonal evolution). Interestingly, approx 70% of patients with 2 or more TMPRSS2-ERG transcript patterns showed Gleason 7 or higher PrCa on biopsy, suggesting a predisposition for more aggressive tumors among patients with heterogeneous TMPRSS2-ERG mutations. Conclusions: The characteristics of the TMPRSS2-ERG fusions found in a patient's PrCa at biopsy may provide insight into the molecular etiology of that tumor(s). Moreover, our data suggest that patients with a high degree of TMPRSS2-ERG heterogeneity are more likely to have higher Gleason score tumors, suggesting that instability of this genetic lesion may be a marker for more aggressive disease.