National Cancer Institute • **Division of Cancer Prevention**

The Early Detection Research Network

Translational Research to Identify Early Cancer and Cancer Risk

Third Report • March 2005

Contents

Foreword

Remarkable strides in cancer research and technology in the late 20th Century have given way today to an opportunity for exponential progress against the disease. With this in mind, the National Cancer Institute (NCI) is pursuing a forward-looking goal to eliminate suffering and death due to cancer by 2015.

We believe in the next few years that new intervention strategies will allow us to prevent and/or eliminate many cancers and ultimately transform cancer into chronic, manageable diseases that patients live with-not die from.

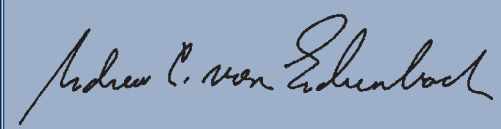
Research has led to the understanding that cancer, once mysterious, is actually a disease process whose mechanisms can be revealed and to some extent controlled. Our growing understanding of this process has exposed multiple chances to intervene in the process. These new intervention strategies include preventing initiation of the process; detecting it early when it is most amenable to elimination; and arresting the process to stop the metastasis.

Scientific advances and major discoveries from areas such as genomics, nanotechnology, proteomics, metabolomics, immunology, molecular imaging, and bioinformatics allow us to envision a future when a patient's genetics, lifestyle, and environmental risk for cancer can be combined with effective prevention and early intervention strategies, especially for those at high risk.

While vast amounts of information about the genetic basis of cancer has been produced, we are more critically learning that the functioning of normal and tumor cells is controlled by the proteins that are transcribed from abnormal genomes. These proteins along with genes and other indicators of the processes and pathways that distinguish cancer are true "biomarkers." Through the use of advanced transformational technologies NCI's Early Detection Research Network is developing innovative ways to discover and validate such biomarkers for use in clinical applications.

EDRN is a leader in establishing and using criteria for the validation of biomarkers; the creative use of information technology; and a multidisciplinary collaborative approach to discovery and development. Powerful tools for detecting cancer and cancer risk will support valuable partnerships across the cancer community, including other federal agencies, academic institutions, and private industry, to implement key elements of our strategic plan.

EDRN's discovery, development and delivery of cancer biomarkers will establish a strong foundation necessary to help make NCI's 2015 strategic goal happen.



Andrew C. von Eschenbach, M.D.
Director, National Cancer Institute

Introduction

Since 2000, the National Cancer Institute's Early Detection Research Network (EDRN) has fostered a highly collaborative, multidisciplinary research venue to improve the early detection of cancer. The Network is focused on translating new molecular knowledge into practical clinical tests that identify cancer at the earliest stages of a normal cell's transformation into a cancer cell, and to identify individuals at risk of developing cancer.

This scientific consortium of more than 30 institutions and their research teams from preeminent academic, private, and government institutions is a vital component of the core NCI mission to conduct and support research to decrease incidence, suffering, and death due to cancer.

Early detection of disease has played a pivotal role in patient survival and quality of life throughout the history of medicine. Found early, many lethal diseases can be prevented from progressing by using existing methods, such as vaccination, drug therapy, surgical, or lifestyle interventions. The inherent difficulty of this task, however, is to effectively detect conditions at the earliest possible stage with enough certainty that occurrences can be addressed successfully.

EDRN's overarching goal is to establish meticulously validated biological markers that are ready for large-scale clinical testing. Progress toward this end includes a strengthened Network infrastructure; increased usage of novel technologies and rigorous methods to prove the efficacy of biological markers of disease; and key research milestones in the development and validation of these biomarkers. Throughout this report certain terms appear in bold face. These are defined in the glossary found at the back inside cover.

EDRN is funded by NCI's Division of Cancer Prevention, headed by Peter Greenwald, M.D., Dr.P.H., and administered by the Cancer Biomarkers Research Group and its chief, Sudhir Srivastava, Ph.D., M.P.H. In addition to biomarker development and validation, the Division of Cancer Prevention plans and directs a broad extramural program of cancer prevention research focused on chemoprevention, nutritional science, genetic and infectious agents, and biometry; fosters training and career development; and advances community-based clinical research and the dissemination of effective practices.

Opportunities and Challenges

The potential of molecular techniques to significantly improve early detection of localized cancers provides an unprecedented opportunity to understand the biology, improve diagnosis, enhance treatment, and reduce mortality. Research in the past 30 years has deepened our insights into nearly all biological processes, particularly with the complete sequencing of the human genome in 2000. Yet, translating these discoveries into daily medical practice takes time.

EDRN was established to help expedite this translation of discoveries into medical practice. Instead of funding individual investigators in single laboratories, NCI funds groups of researchers in the form of a Network. Each competitively funded investigator within the group assumes specific responsibilities that relate to the overall success of the specific research objective. They are basic scientists, clinicians, and others, who agree to collaborate with each other to move new discoveries into the clinical validation process, that is, to prove that a discovery–be it a biological marker that can be measured or a specific test method–actually works in people.

The Network is based on the premise that integration of discovery, evaluation, and clinical validation phases of medical research are more likely to succeed when they are carried out in a concerted and systematic fashion. For this reason, all investigators in the Network have a stake in all of its broad research objectives.

Preferred Definitions

Biomarker - A characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic response to a therapeutic intervention.

Clinical Endpoint - A characteristic or variable that reflects how a patient feels, functions or survives.

Surrogate Endpoint - A biomarker intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm, or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic or other scientific evidence.

–NCI-Food and Drug Administration Working Group, 1999
(Atkinson, A.J. et al. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clin Pharmacol. Ther. 2001; 69:89-95.)

The challenge is to identify biomarkers and other technologies that provide an earlier indication of disease with a more reliable and precise predictive ability than current methods. The past decade has witnessed a revolution in biological science in which researchers can study thousands of different molecules in a single experiment. These so-called high-throughput studies allow a multitude of genes or proteins to be analyzed simultaneously. Genes, the functional and physical unit of heredity passed from parent to offspring, are pieces of DNA, with most genes containing the information for making a specific protein. Proteins are molecules made up of amino acids that are needed for the body to function properly and form the basis of body structures such as skin or hair, and of substances such as enzymes, cytokines, and antibodies.

High-throughput studies allow for the rapid comparison of samples from many different patients with the expectation of finding biological indicators that are directly related to the disease state. However, individual differences among people are a major hurdle when attempting to discover a disease-related biomarker in biological fluids, such as serum.

Nonetheless, the coupling of high-throughput technologies and protein science enables samples from hundreds of patients to be rapidly compared. These high-throughput technologies have greatly benefited proteomics (the study of the structure and function of proteins including the way they work and interact with each other inside cells), genomics (the study of the organization of genomes and the nucleotide sequences of the component genes), and transcriptomics (the study of genes transcribed from DNA within living cells to molecules of messenger RNA as the first step in protein synthesis).

Although proteomic approaches cannot yet remove the "needle-in-a-haystack" quality of discovering novel biomarkers, they do provide the capability to inventory components within the "haystack" at an unprecedented rate. In fact, such capabilities already have exponentially expanded our knowledge of the different types of proteins within serum, and opened the way for novel technologies for diagnosing cancers to emerge.

While many proteomic technologies and data management tools are in their infancy, the future of proteomics in disease diagnostics is promising. Molecular strategies involving imaging, proteomic and genomic analysis of tumors and other specimens may ultimately identify small and early lesions that to date have been inaccessible in conventional clinical practice. There are currently markers for diseases such as prostate (prostate-specific antigen, or PSA) and ovarian cancer (cancer-antigen 125, or CA125), that have not been shown to possess the sensitivity and specificity necessary to be used as a general screening tool in the diagnosis of early stage prostate and ovarian cancer.

Some have asked whether the amount of data accumulated in such studies will overwhelm researchers. This does not seem to be the case based on the leads already discovered using proteomics. It is likely that biomarkers with better sensitivity and specificity will be identified, and individuals will be treated using customized therapies based on their specific protein profile.

Timeline

The Early Detection Research Network

November 13, 1998	Early Detection Implementation Group proposes concept for Early Detection Research Network to NCI Board of Scientific Advisors. Concept is approved.
April 2000	Early Detection Research Network fully launched.
September 26-27, 2000	First EDRN Workshop in Chicago, IL
October 2000	Initial Report of Early Detection Research Network published.
June 26, 2001	Progress Report presented to the Board of Scientific Advisors.
October 13-15, 2001	Second EDRN Workshop convened in Seattle, WA
March 10-15, 2002	EDRN-Gordon Research Conference on New Frontiers in Cancer Detection and Diagnosis, Ventura, CA
November 14-15, 2002	Progress Report to NCI Board of Scientific Advisers
March 2003	EDRN Approved for Second 5-year cycle
July 2003	Validation Study launched: Profiling for Prostate Cancer
August 17-22, 2003	EDRN-Gordon Conference on New Frontiers in Cancer Detection and Diagnosis, Andover, NH
September 2003	Launching of the First Clinical Validation Study for Microsatellite Instability as a Biomarker for Bladder Cancer
March 17-19, 2004	Training workshop on the Analysis of Proteomic Spectral Data including SELDI/MALDI-TOF-MS Applications; Review of SELDI Phase 1, Seattle, WA
June 14 - 16, 2004	Third Annual Scientific Workshop, Bethesda, MD
July 28-29, 2004	Workshop on Research Strategies, Study Designs and Statistical Approaches to Biomarkers Validation for Cancer Diagnosis and Detection, Gaithersburg, MD
September 2004	EDRN Outreach Meetings Breast/GYN Collaborative Group Meeting, New York, NY; GI Collaborative Group Meeting, Norfolk, VA; GU Collaborative Group Meeting, Houston, TX; Lung Collaborative Group Meeting, Denver, CO
November 2-3, 2004	Investigators Training Meeting for EDRN's Validation of Serum Markers for Early Detection of Hepatocellular Carcinoma (DCP), Boston, MA
November 4, 2004	Biomarkers Developmental Laboratories Investigators' Training Meeting, Bethesda, MD
January 16-21, 2005	EDRN-Gordon Conference on New Frontiers in Cancer Detection and Diagnosis, Buellton, CA
March 21-23, 2005	Tenth Steering Committee Meeting, Rockville, MD

Early Detection Research: A Promising Approach

NCI considers collaboration as a prompt, effective, integrated response to harness the tremendous potential of projects that are critical to the NCI mission. Collaboration accelerates the response time to discoveries, and provides the unique opportunity for leveraging resources and extending research bases beyond the reach of an individual organization.

EDRN offers a fitting example of the collaborative model in which the substantive outcome and the relationship outcome are intertwined. This new approach depends on a distinctive organizational culture in which a set of important understandings held in common by its members guides individual and collective behavior. EDRN promotes a "vertical" approach for conducting biomarker research, whereby criteria for judging the roles and clinical significance of each newly discovered biomarker are provided, as are the criteria and strategies for judging biomarker relationships to one another. This is in contrast to the current culture in most academic environments in which researchers compete rather than collaborate, thereby creating redundancies in projects and reducing potential synergies across disciplines. Such a system fosters a "horizontal" approach, which may result in rapid discoveries of many biomarkers by participating laboratories, but limits elements to further advance the validation of biomarkers.

Laboratory Measures for Surrogate Endpoints

Reliability	repeatability, a high correlation between two measurements
Precision	the total error is zero
Accuracy	measure the true level, devoid of bias
Validity	measure the true change or effect of intervention on outcome

The Network is structured around four main components. EDRN comprises a group of Biomarkers Developmental Laboratories, which develop and characterize new biomarkers, or refine existing biomarkers; Biomarkers Reference Laboratories, which serve as a resource for clinical and laboratory validation of biomarkers, including technological development, standardization of assay methods, and refinement; Clinical Epidemiology and Validation Centers, which conduct and support the early phases of clinical and epidemiological research on the application of biomarkers; a Data Management and Coordinating Center which provides statistical, logistics, information support, and develops the theoretical statistical approaches to pattern analysis of multiple markers simultaneously; an Informatics Center led by investigators at the National Aeronautics and Space Administration's Jet Propulsion Laboratory serving as the lead for the informatics component, leveraging its experience in building science information systems.

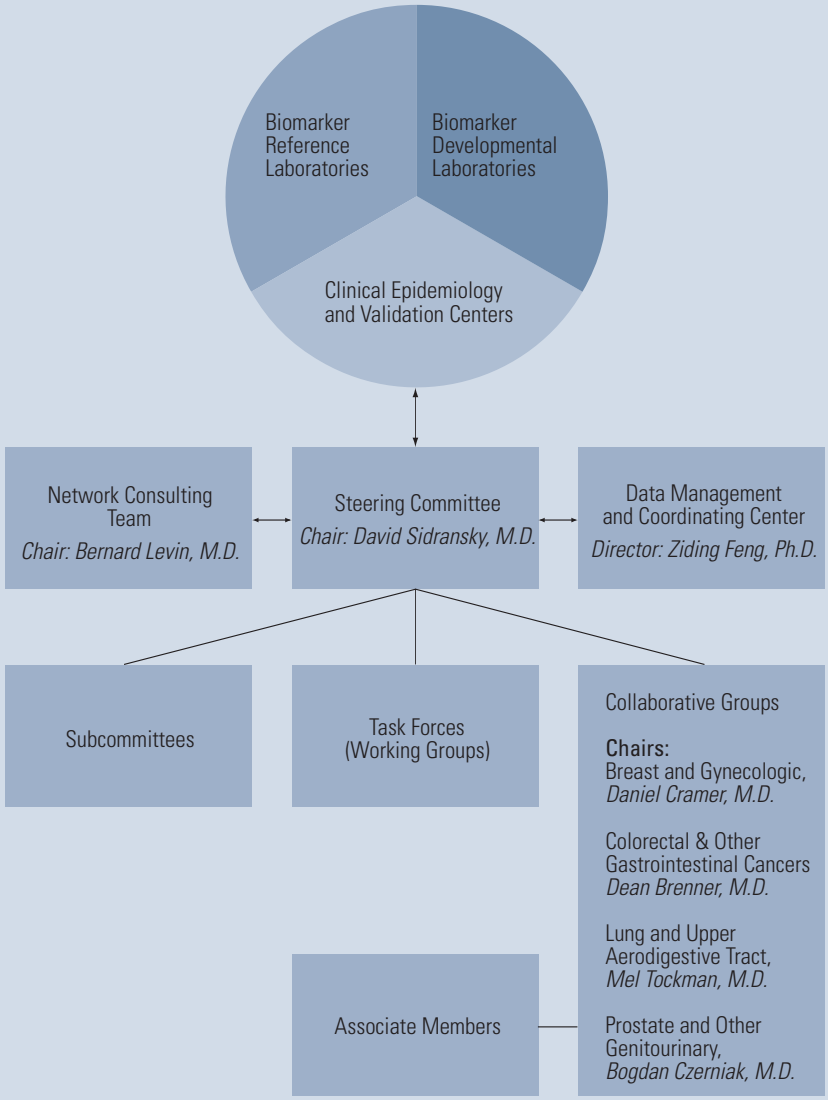
A Steering Committee, comprised of the Network's Principal Investigators and NCI staff, coordinates the work of the consortium and provides major scientific management oversight. The group is responsible for developing and implementing protocols, designs, and operations. An Executive Committee of the Steering Committee meets monthly; it is comprised of chairs for the Collaborative Groups, the NCI program director, the Steering Committee chair and co-chair.

Subcommittees and working groups report to the full committee.

Additional collaborations are encouraged through an Associate Membership program, which supports pilot and resource sharing projects, and provides open participation in meetings, workshops, and conferences by non-Network professionals with proposals focused on biomarker research.

Components of the Early Detection Research Network

This chart outlines the EDRN infrastructure for supporting collaborative research on molecular, genetic and other biomarkers in cancer detection and risk assessment.



Principal Investigators

Clinical Epidemiology and Validation Centers

Principal Investigator Institution (alphabetic by last name)	**Organ Focus**	**Funding Period** (by calendar year)
Dean E. Brenner, M.D. University of Michigan Ann Arbor, MI	Colon	2000-2010
Daniel W. Cramer, M.D. Brigham and Women's Hospital Boston, MA	Ovary	2000-2010
Paul Engstrom, M.D. Fox Chase Cancer Center Philadelphia, PA	Breast	2005-2010
Kathy Helzlsouer, M.D. Johns Hopkins University Baltimore, MD	Breast	2000-2005
Henry T. Lynch, M.D. Creighton University Omaha, NE	Pancreas	2000-2010
Alan Partin, M.D. Johns Hopkins University Baltimore, MD	Prostate	2000-2010
William N. Rom, M.D. New York University School of Medicine New York, NY	Lung	2000-2010
Martin Sanda, M.D. Beth Israel Deaconess Medical Center Boston, MA	Prostate	2005-2010
Margaret R. Spitz, M.D., M.P.H. University of Texas, M.D. Anderson Houston, TX	Head & Neck	2000-2005
Ian M. Thompson, M.D. University of Texas Health Science Center San Antonio, TX	Prostate	2000-2010
Elizabeth R. Unger, M.D., Ph.D. Centers for Disease Control and Prevention Atlanta, GA	Cervix	2000-2010

Biomarker Developmental Laboratories

Principal Investigator Institution (alphabetic by last name) *=new grantee	**Organ Focus**	**Technology/ Approach**	**Industry Collaboration**	**Funding Period**
William L. Bigbee, Ph.D. University of Pittsburgh Cancer Center Pittsburgh, PA	Colon	Proteomics	Predicant Biosciences	1999-2009
Timothy Block, Ph.D. Drexel University Philadelphia, PA	Liver	Proteomics, Glycomics	Xenomics, Inc., Immunotype, Inc.	1999-2009
*Paul Cairns, Ph.D. Fox Chase Cancer Center Philadelphia, PA	Kidney	Methylation, Proteomics	--	2005-2009
*Arul Chinnaiyan, M.D., Ph.D. University of Michigan Ann Arbor, MI	Prostate	Genomics, Proteomics, and Immune Response	GMP Companies, Inc.	2005-2009
Bogdan A.Czerniak, M.D., Ph.D. University of Texas M.D. Anderson Cancer Center Houston, TX	Bladder	Genomics	--	1999-2009
*Laura J. Esserman, M.D., M.B.A. University of California San Francisco, CA	Breast	Genomics and Proteomics	Sequenom Biotrue, BD Biosciences, Celera Diagnostics, Biospect, Chroma	2005-2009
Wilbur A. Franklin, M.D. University of Colorado Health Science Center Denver, CO	Lung	Genomics	--	1999-2009
Adi F. Gazdar, M.D. University of Texas Southwestern Medical Center Dallas, TX	Lung	Genomics and Proteomics	Rules-Based Medicine	1999-2009
Samir Hanash, M.D., Ph.D. Fred Hutchinson Cancer Research Center Seattle, WA	Lung and others	Proteomics and Auto anitbodies	--	1999-2009
*Michael Hollingsworth, Ph.D. University of Nebraska Medical Center Omaha, NE	Pancreas	Proteomics and Mice Models	--	2005-2009

*Anne M. Killary, Ph.D. University of Texas M.D. Anderson Cancer Center Houston, TX	Pancreas	Genomics	--	2005-2009
*Alvin Y. Liu, Ph.D. University of Washington Seattle, WA	Prostate and Bladder	Proteomics	MacroGenics Inc.	2005-2009
Jeffery R. Marks, Ph.D. Duke University Medical Center Durham, NC	Breast	Genomics (Gene Expression), Proteomics	Abbott Laboratories Diagnostic Division	1999-2009
Stephen J. Meltzer, Ph.D. University of Maryland School of Medicine Baltimore, MD	Esophagus	Genomic and Proteomics	--	1999-2009
*Hemant K. Roy, M.D. Evanston Northwestern Research Institute Evanston, IL	Colon	Spectral imaging	--	2005-2009
J. Oliver Semmes, Ph.D. Eastern Virginia Medical School Norfolk, VA	Multi-organ	Proteomics	Ciphergen	1999-2009
David Sidransky, Ph.D. Johns Hopkins University Baltimore, MD	Lung	Genomics	Oncomethylome Sciences, Affymetrix	1999-2009

Biomarker Reference Laboratories

Principal Investigator Institution	Organ Focus	Funding Period (by calendar year)
Peter E. Barker, Ph.D. National Institute of Standards and Technology Gaithersburg, MD	All sites	2001-2010
*Daniel W. Chan, Ph.D. Johns Hopkins University Baltimore, MD	All sites	2005-2010
David Chia, Ph.D. University of California Los Angeles, CA	All sites	2000-2010
William E. Grizzle, M.D., Ph.D. University of Alabama Birmingham, AL	All sites	2000-2010

*Sanford A. Stass, M.D. University of Maryland Baltimore, MD	All sites	2005-2010

Data Management and Coordinating Center

Principal Investigator Institution	Organ Focus	Funding Period
Ziding Feng, Ph.D. Fred Hutchinson Cancer Research Center Seattle, WA	All sites	2000-2010

Informatics Center

Principal Investigator Institution	Organ Focus	Funding Period
Dan Crichton, M.S. National Aeronautics and Space Administration California Institute of Technology Pasadena, CA	All sites	2001-2010

Network Consulting Team

Chair

Bernard Levin, M.D.
University of Texas M.D. Anderson Cancer Center
Houston, TX

Powell Brown, M.D., Ph.D.
Baylor College of Medicine
Houston, TX

Judy Ellen Garber, M.D., M.P.H.
Dana-Farber Cancer Institute
Boston, MA

Sylvan Green, M.D.
Arizona Cancer Center
Tucson, AZ

Steve Gutman, Ph.D.
U.S. Food and Drug Administration
Washington, DC

Leland Hartwell, Ph.D.
Fred Hutchinson Cancer Center,
Seattle, WA

Robert Jaffe, M.D.
University of California San Francisco
San Francisco, CA

Heidi Malm, Ph.D. (Ethicist)
Loyola University
Chicago, IL

Larry Norton, M.D.
Memorial Sloan-Kettering Cancer Center
New York, NY

Olufunmilayo I. Olopade, M.D.
University of Chicago
Chicago, IL

Anil K. Rustgi, M.D.
University of Pennsylvania
Villanova, PA

John Wagner, M.D., Ph.D.
Meck Research Laboratory
New York, NY

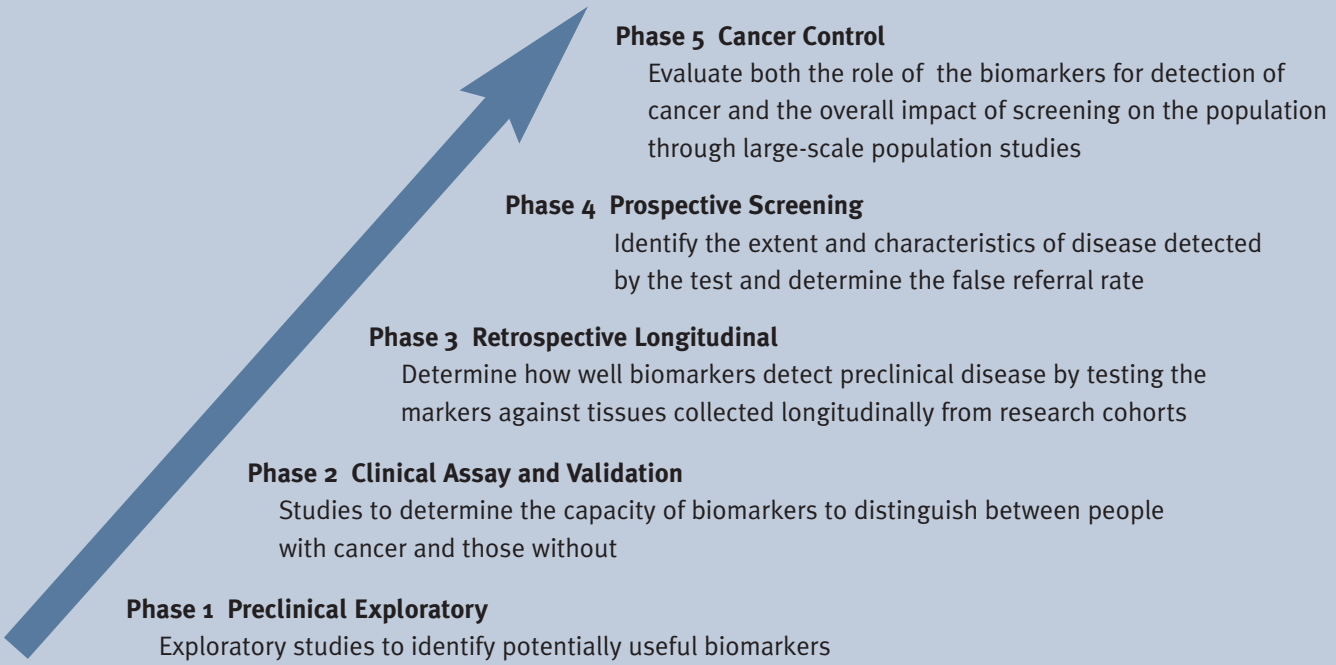
Jane Beth Williams (Cancer Survivor)
Houston, TX

Advancing Biomarker Discovery and Development: Progress from the EDRN Biomarker Development Laboratories

Through its established, integrated, and multidisciplinary environment, the Early Detection Research Network facilitates collaboration among basic scientists, clinicians, epidemiologists, biostatisticians, technology developers, and other health professionals. The initial challenge of encouraging researchers to work as collaborative teams rather than as competing individuals continued in 2003 and 2004 as the Network's institutional components and linkages were solidified.

The program focuses on speeding laboratory discoveries and their subsequent translation to clinical **biomarkers**. The effort to provide timely, cost-effective clinical tests for early detection of cancer and identification of high-risk individuals remains in high gear. Investigator-driven research is aimed at developing, testing, and evaluating promising biomarkers and technologies. At the same time, researchers are analyzing biomarkers and expression patterns to form the foundation for subsequent large definitive validation studies of cancer detection and screening methodologies.

Figure 1 **Guiding Principles for Biomarker Research: Phases of Early Detection Research**



Progress in Discovery by Organ Site

The Network is testing and evaluating predictive biomarkers for cancers in numerous organ sites. Research progress is presented here based on the formal organization of the EDRN Collaborative Groups: breast and gynecologic cancers; gastrointestinal and other associated cancers; lung and upper aerodigestive cancers; and prostate and urologic cancers.

Breast and Gynecologic Cancers

Some 30 investigators from 18 different institutions participating in EDRN are members of the Breast/Gynecologic Collaborative Group, whose shared goal is to advance the early detection of breast, ovarian, and cervical cancers.

BREAST CANCER

Biomarkers can play a critical role in decreasing both the number of women who are misdiagnosed with and those who will die from breast cancer. Breast cancer remains the most frequently diagnosed cancer in U.S. women, with over 211,000 new cases diagnosed each year, and is the cause of more than 40,000 deaths yearly.

Investigators at Eastern Virginia Medical Center are evaluating proteins from samples collected from the lining of the milk ducts by a technique known as ductal lavage. Little is known regarding this unique proteome source, so normal breast duct material was examined for identification of the major protein components. Employing a molecular profiling process using chips to capture proteins and then analyze them **(Surface Enhanced Laser Desorption/Ionization or SELDI)**, the researchers identified about 60 different proteins. About 15 of these proteins were common to every sample. Table 1 lists some of the identified proteins.

Table 1 Proteins in Breast Ductal Lavage of Normal Breasts

MW (kDa)	Protein Id
11	Histone H4
16	Prolactin-induced protein
24	Ig Lambda light chain
29	Ig Kappa light chain
35	Alpha-2-glycoprotein
52	Unknown human protein
55	Multidrug resistance protein
69	Human serum albumin
78	Lactoferrin precursor

In a follow-up study to discern which set of these proteins would most likely indicate cancer or cancer risk, fluid from ductal lavage of 18 women with breast cancer and 12 women at high risk of the disease was evaluated by investigators at Emory University. Using a statistical method known as CART (Classification and Regression Trees), which helps select the most reliable indicators from many possible combinations, they distinguished 3 protein peaks in samples indicative of cancer or high risk for cancer.

Cross validation studies showed that measurement of these three proteins allowed a correct cancer diagnosis with a **sensitivity** of 81% (identified 81% of breast cancers) and a **specificity** of 91% (correctly identified 91% of women without breast cancer). Future efforts will focus on purifying and identifying the different biomarker proteins observed in these ductal lavage samples, especially in lavage fluids associated with ductal tumors.

Researchers from Eastern Virginia Medical Center also profiled the proteins from the serum of women who have a mutation in the breast cancer gene *BRCA1* before any of these women were diagnosed with the disease. The *BRCA1* women were followed for 7 years or until they developed breast cancer and then were divided into two groups (*BRCA1* Cancer or Carrier) for analysis. In addition, a collection of serum samples from 15 women with breast cancer, but who do not have a mutation in *BRCA1* (called 'sporadic' breast cancers) was also included.

These samples were analyzed by **SELDI-TOF-MS** to reveal differentially expressed proteins between cancers with *BRCA1* mutations, *BRCA1* carriers, and sporadic breast cancer samples. Using the differences in proteins from the serum samples, the investigators were able to correctly identify 13 out of 15 women with *BRCA1* who had cancer versus women with *BRCA1* mutations but no disease and 14 out of 15 women with *BRCA1* who had cancers vs. women with sporadic cancers. Twenty-nine proteins were overexpressed in the women with *BRCA1* related cancer relative to the non-cancer *BRCA1* carriers.

Whether the proteins identified represent a very early detection of breast cancer or show cancer risk remains to be determined. Identification of the proteins is in progress. Follow-up studies are planned in women with *BRCA1* mutations and their sisters who do not have mutations.

In collaborations between Duke and Abbott Developmental labs, investigators are making progress on the banking of biological samples for research and on biomarker development for breast cancers. This group has now collected over 1,500 specimens from nearly 500 women undergoing definitive diagnosis and treatment for breast cancer. Continued efforts are focused on women who are in the process of being diagnosed. To this end, a protocol has been instituted to obtain blood specimens before sonographic-directed biopsy of suspicious lesions. These are the women who will most immediately benefit from additional tests that could discriminate benign from malignant conditions. Furthermore, radiologic information will be incorporated (both mammographic and sonographic) in future predictive models based upon circulating biomarkers.

Other promising biomarkers being investigated are:

1. Markers of **methylation**: The addition of a methyl group to specific sites on DNA is a fundamental process that can modulate gene expression and regulate the stability of chromosomes. Methylation markers in breast cancer includes Zyxin and alpha Catenin genes; and

2. Plasma-based expression markers: Proteins present in blood plasma that are the result of the expression of a gene are an area that is being actively mined for markers of cancer and cancer risk, including a focus on BS106, mammaglobin, and cytokeratin 19.

Table 2 Summary of the Developmental Status of Breast Cancer Biomarkers

Candidate Biomarker	Phase I (Discovery and Early Refinements)			Phase II	Phase III	
	Discovery	Predictive Analysis	Assay Refinement	Characterization (Blinded)	Retro-Longitudinal	Comments
SELDI -TOF-MS (Proteins from ductal lavage)	→	→				Promising
SELDI-TOF-MS (Proteins from serum)	→					Promising
Specific antigen detection-Luminex, ELISA (Serum, Nipple Aspirate Fluid)	→					Early Discovery
Chromosomal changes (Fine Needle Aspiration, Nipple Aspirate Fluid)	→	→				Early Discovery
Rare Cancer Cell Detection (Blood, Nipple Aspirate Fluid, Fine Needle Aspiration)	→	→	→			Pre-validation
Autoantibodies (Serum)	→					Early Discovery

OVARIAN CANCER

Ovarian cancer has the distinction of being the most fatal cancer of the female reproductive system. About 22,200 are diagnosed with this disease each year and more than 16,000 die of the disease yearly. Symptoms of the disease are often vague and can be ignored as gastrointestinal in origin. Early detection of ovarian cancer via biomarkers is a critical element in reducing deaths from this disease.

Advances have been made in describing new markers for ovarian cancer using different techniques. Research is under way using cDNA **microarray** analysis to identify genes that overproduce specific proteins (upregulated genes) in ovarian cancer cell lines, compared to normal ovarian epithelial cells. Investigators at Brigham and Women's Hospital in Boston reported on prostasin and osteopontin as two candidate biomarkers for ovarian cancer. A third gene identified in that study was epithelial cell adhesion molecule (Ep-CAM). Ep-CAM expression in ovarian cancer cells was correctly identified using an auto-antibody test with a **sensitivity** of 71% and a **specificity** of 81%. The fact that the body is producing antibodies to this protein suggests that it plays some role in a disease process. All three markers identified through these gene up-regulation studies could prove to be complementary

to CA125, a tumor marker for ovarian cancer that is already under study in a large-clinical trial for its value in detecting ovarian cancer.

Prevalidating Multiple Markers

As a prelude to full-scale validation studies, EDRN investigators are collaborating to perform a comparative study of several biomarkers on a set of specimens contributed by New York University. CA125, osteopontin, and prostasin have been tested through the system, and HK6, HE4 and haptoglobin will be run as soon as the assay is optimized. Samples from Northwestern University have been analyzed by **SELDI** with the Boston samples soon to follow.

Progress also continues on the isolation and identification of markers in both serum and urine originally identified through **SELDI**. Five proteins from serum were identified as likely to be able to predict ovarian cancer and four of these have been identified as known proteins. One protein was higher in controls than cases, making it a candidate biomarker by looking for its absence, and was identified as calmodulin-like skin protein. The remaining peaks/proteins were present at higher levels in women with ovarian cancer. One protein was identified as survival promoting peptide dcd-1; another as the alpha chain of haptoglobin, and a third as apolipoprotein A1.

A protein often found in the urine of ovarian cancer patients was purified, sequenced, and identified as non-secretory ribonuclease. Antibody assays qualitatively confirmed presence of the antigen in urine from ovarian cancer patients but not from women without the disease, making this another candidate biomarker.

In continuation of earlier work in which it was reported that low resolution **MS proteomic** pattern profiling was an effective tool to segregate cancer from non-cancer with a sensitivity, specificity and positive predictive value of 100%, 95% and 94%, respectively, investigators are following up analysis of sera from 248 women comparing spectra obtained from a newly developed high-resolution mass spectrometer versus the relatively low-resolution spectrometer currently in use. Results showed that the higher resolution spectra generated superior diagnostic models, several of which attained 100% sensitivity and specificity. Validation of this new tool is ongoing in larger study sets of serum.

The University of Texas M. D. Anderson Cancer Center is enrolling women from families at high risk for ovarian cancer into a prospective study to assess the value of longitudinal CA125 screening. In addition to information about CA125 generated from this study, thousands of prospective serum specimens are being generated for future use in validating other ovarian cancer screening tests.

Table 3 Summary of the Developmental Status of Ovarian Cancer Biomarkers

Candidate Biomarker	Phase I (Discovery and Early Refinements)			Phase II	Phase III	
	Discovery	Predictive Analysis	Assay Refinement	Characterization (Blinded)	Retro-Longitudinal	Comments
SELDI-TOF-MS profile DCD-1, alpha-haptoglobin, apolipoprotein-A1 (serum)	→					Early Discovery
Autoantibodies (serum) Ep-CAM	→	→				Early Discovery
Luminex panel for combining multiple markers	→	→	→			Pre-validation (Assay refinement)

ENDOMETRIAL AND CERVICAL CANCER

Cancer of the lining of the uterus, better known as endometrial cancer, occurs in nearly 41,000 women each year and causes 7,310 deaths. Cancer of the uterine cervix occurs in more than 10,000 women each year, despite the widespread use of Pap smear testing which can identify most cancers before they are invasive. More than 3,700 women will die of cervical cancer in 2005.

Investigators at the Creighton Epidemiology and Validation Center continue to offer EDRN researchers the resources of their Familial Cancer Clinic for collaborative studies, including extensive pedigrees and specimens. Potential biomarkers for endometrial cancer are being investigated in a project with investigators at The University of Texas M. D. Anderson Cancer Center using endometrial specimens from women from families with hereditary non-polyposis colon cancer (HNPCC) in which endometrial cancer also occurs more frequently.

Researchers at the Centers for Disease Control and Prevention have ongoing efforts in biomarker discovery based on gene expression profiling of RNA from exfoliated cervical cells. Based on results of a proof-of-principle study, using 15 samples and arrays of 3,800 genes, they completed a study on 30 samples using arrays that explored 30,000 genes. This larger study verified the effectiveness of sample labeling, image processing and data analysis. The most promising genes and others selected from literature will be tested in a "pre-validation" study using **qPCR**. (See box, previous section.) The test will include 50 samples from women with grade 3 cervical intraepithelial neoplasia (CIN) plus 50 matched samples randomly selected from women without disease. Results from the "pre-validation" study will determine the design of a full validation study.

The CDC has also initiated marker discovery using protein profiling in cervical mucous samples. Analysis of 60 age- and race-matched samples, including various permutations of CIN and/or HPV 16 positive cases, is in progress. Tissues from subjects enrolled in the CDC's study, as well as anonymous tissue blocks, will be used to construct a tissue **microarray** as a key tool for biomarker validation.

Other ongoing studies include serum samples from the biorepository being tested for EGRF and IGF-II, and Abbott/Vysis is investigating the performance of a panel of **FISH** chromosomal probes on exfoliated cells in the biorepository collection.

Table 4 Summary of the Developmental Status of Cervical Cancer Biomarkers

Candidate Biomarker	Phase I (Discovery and Early Refinements)			Phase II	Phase III	
	Discovery	Predictive Analysis	Assay Refinement	Characterization (Blinded)	Retro-Longitudinal	Comments
SELDI-TOF-MS profile (Cervical mucous or liquid cytology)	→	→	→			Pre-validation
RNA expression markers (Exfoliated cervical cell extracts)	→	→				Early validation
FISH chromosomal probes (Exfoliated cervical cells)	→	→				Early validation
EGFR, IGF-II (Serum)	→	→	→			Pre-validation
p16 or Pten histochemistry (Cervical and endometrial tissue)	→					Early Discovery
SELDI-TOF-MS proteomic patterns (Serum or liquid cytology)	→					Early Discovery

Colorectal and Other Gastrointestinal Cancers

There are more than 30 investigators representing more than 17 institutions within the EDRN Colorectal and Other Gastrointestinal Cancers Collaborative Group. To date, the bulk of their work has been discovery of biomarkers. Biomarkers have been identified and preliminarily characterized, and a summary of data for each of these biomarkers follows.

COLORECTAL CANCER

Investigators at Fred Hutchinson Cancer Center are investigating antibodies to ubiquitin C-terminal hydrolase L3 as a potential biomarker for colon cancer. The fact that the body is producing an antibody to the protein suggests it may be involved in a disease process (proteins that are 'normal' would not incite the body to this kind of immune response). Using sera from 15 newly diagnosed patients with colon cancer, 15 with lung cancer, and 15 normal controls, 39 proteins showed enhanced reactivity with sera from patients with colon cancer relative to controls. One protein that reacted with 9 of the 15 colon cancer sera was identified as ubiquitin C-terminal hydrolase isozyme 3.

Investigators have also found mutations in the *K-ras* gene in soluble urinary DNA from patients with colon cancer. The data from 66 patients are summarized in Table 5. Of interest is the high "false-positive" rate of approximately 20% with the DNA urine test. These samples were amplified and sequenced and although the *K-ras* mutation was found in these samples, the endoscopy was negative, possibly indicating cancer or cancer risk in a very early stage.

Table 5 Urine Detection of *K-ras* Mutations

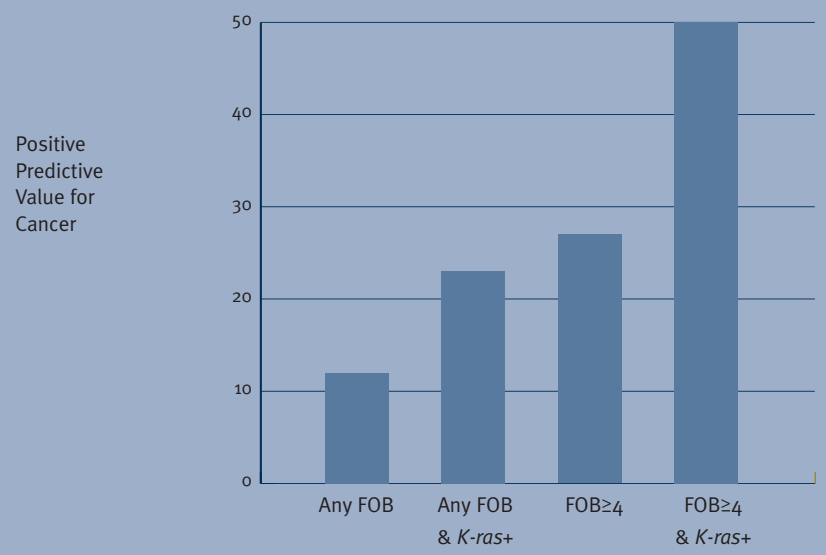
Group	N=	*Urine K-ras +*	*Tissue K-ras +*	Positive Rate
Healthy Controls	48	9	N/A	20%
Adenoma	13	10	13	77%
Cancer	5	5	5	100%

These data suggest that 1) there is a significant correlation between *K-ras* mutation detected in disease tissue and urine; and 2) a high incidence of *K-ras* mutations is detected in urine of patients with colorectal cancer and adenoma polyps. These studies support the concept that urine can be used as a source of DNA for early cancer detection. This work will be expanded in a larger Phase 2 biomarkers trial that studies 600 subjects with colon cancer, adenomas, and normal controls. Investigators at Drexel are expanding this assay to include probes for APC mutations and other common genetic mutations associated with colon cancer.

Another approach is to purify DNA from stool and measure the *K-ras* mutation levels. Current methods of fecal DNA purification and detection of cancer-associated genetic expression require the home collection and laboratory processing of large quantities of stool-an unpleasant task. Investigators at the Great Lakes-New England Consortium, in collaboration with investigators in Israel, have developed a method to isolate DNA from fecal occult blood test cards. This method requires no more stool than that used for standard fecal occult blood in stools screening tests. The stool on the card, once assessed for occult blood, may be reused for DNA extraction. Stool samples are excised from the cards, and DNA is isolated and processed to detect the presence of *K-ras* mutations from amplified DNA.

A total of 250 cards from fecal occult blood tests (FOBT) (211 positive and 39 negative) were studied. Twenty-six malignancies were detected among 211 positive cards using standard follow-up procedures and no malignancies were detected among 39 negative cards after a 2-year follow-up period. *K-ras* mutations were detected in 47 samples (38 in positive cards and 9 in negative cards). In a preliminary validation study, adding the *K-ras* test to a FOBT test doubled the positive predictive value for cancer detection (adenoma) (see Figure 2) while not changing false predictions from non-malignant colonic disease (hemorrhoids, diverticulitis, inflammatory bowel disease).

Figure 2 ***K-ras* Mutations Double the Predictive Value of a Positive Fecal Occult Blood Test (FOBT) for Colon Cancer**



These data suggest that sufficient quality and mass of DNA can be recovered from fecal occult blood test cards to enable amplification of key tumor-associated genes. Such a tool will be useful for multiple genes associated with colonic carcinogenesis and transformation.

Sufficient preliminary data are available from Phase I to consider Phase II biomarkers research. The preliminary data were drawn from a cohort of subjects with positive fecal occult blood tests. Phase II research will focus upon assessment of sensitivity and specificity for cancer, adenomas in subjects with and without a positive fecal occult blood test. The test will also be matched with urine sample assays for *K-ras* and tissue samples from colonic adenocarcinomas and adenomas to determine its detection sensitivity and specificity for prediction of *K-ras* mutations present in tissue samples.

Investigators at Creighton, in collaboration with Exact Science, continue to evaluate stool DNA for BAT-26, a site subject to **microsatellite instability** (MSA) in some cancers. It appears that stool BAT-26 testing may detect early-stage colorectal neoplasia in high-risk individuals with hereditary nonpolyposis colorectal cancer (HNPCC). In 52 patients with negative colonoscopies, there were no false-positives.

Aberrant DNA **methylation** is a common **epigenetic** alteration that contributes to colon cancer formation. Aberrant DNA methylation results in transcriptional silencing of genes and is a mechanism for inactivating tumor suppressor genes in colon cancer. The methylated tumor DNA can be detected using **methylation-specific PCR (MSP)** and thus has the potential to be used as a molecular marker for cancer. In colon cancer, the methylation of a number of genes occurs early in the adenoma-carcinoma sequence suggesting these alterations could be used for the early detection of colon cancer. Furthermore, DNA methylation and microsatellite instability appear to have a high concordance in colon cancer suggesting that methylated genes may be especially informative markers for neoplasms that occur in patients with HNPCC syndrome, who are at very high risk for developing colon cancer. Finally, aberrant DNA methylation of tumor suppressor genes may occur secondary to a genetic predisposition, such as HNPCC or to an organ-wide environmental exposure that increases risk and thus may be useful as a prognostic molecular marker for recurrent colon neoplasms in individuals with a history of colon adenomas or adenocarcinomas.

Methylation-specific PCR assays that assess the methylation status of *CDKN2A*, *MGMT*, and *MLH1* genes have been

Table 6 Decision Analysis of Methylation Gene Assays for Detection of Colonic Neoplastic Events

Gene	Analytical Sensitivity (95% CI)	Analytical Specificity (95% CI)	Clinical Sensitivity (95% CI)	Clinical Specificity (95% CI)
CDKN2A	50% (28-72%)	90% (60-98%)	33% (16-54%)	82% (63-94%)
MGMT	71% (47-87%)	82% (52-95%)	50% (31-69%)	70% (51-85%)
Composite			57% (37-75%)	59% (39-78%)

Table 7 Summary of the Developmental Status of Colorectal Cancer Biomarkers

Candidate Biomarker	Phase I (Discovery and Early Refinements)			Phase II	Phase III	
	Discovery	Predictive Analysis	Assay Refinement	Characterization (Blinded)	Retro-Longitudinal	Comments
Ubiquitin C-terminal L3 (Serum)	→	→				Assay Development, then Phase II
SELDI Profile (Serum)	→					Early Discovery, Plan Phase II 2005
Galectin 3-Ligand (serum)	→	→				In Phase II profile
K-ras (Urine)	→	→	→			Early Phase II
K-ras (Stool Guiac)	→	→	→			Early Phase II
Multigene (Stool-Exact)	→	→	→	→		Stool Biomarker "control"
Flat adenoma/Chromoendo	→	→				Phase I completion
GOS (Stool)	→	→	→	→		Stool Profile Phase II
MethylCDKN2A (Stool)	→	→	→	→		Stool Profile Phase II
MethylMGMT (Stool)	→	→	→	→		Stool Profile Phase II
Cox-2 RNA (Stool)	→	→	→	→		Stool Profile Phase II

developed. The *MLH1* MSP assay can detect tumor DNA in the sera of patients with sporadic colon cancer. A second-generation, highly-sensitive MSP assay that can detect DNA from colonic adenomas and colon adenocarcinomas in colonic lavage effluent has been developed. In a feasibility study on colon lavage effluent samples, the *MLH1* assay did not reach analytical sensitivity and specificity to be considered for insertion into an early diagnostic or risk assessment panel. Analytical sensitivity and specificity for CDKN2A and MGMT (accuracy of the assay) ranged from 50% to 90% (Table 6).

These markers will be studied in much larger cohorts to verify these findings in a multi-center environment. The data can be used to develop more complex algorithms to individualize detection and risk biomarkers from stool samples.

PANCREATIC CANCER

Adenocarcinoma of the pancreas will occur in more than 32,180 people in the United States this year and 31,800 will die of the disease. The 5-year survival rate for this highly lethal cancer is about 4%, but if found at an early stage it increases to 15%. This relative improvement in survival associated with earlier stage pancreatic cancer motivates the search for effective early detection and screening.

EDRN investigators have identified potential biomarkers from **genomic** and **proteomic** platforms, many of which have been preliminarily tested in serum and in cell culture systems. Genomic amplification tools with cells obtained from aspiration of pancreatic masses is yielding promising results. Other markers being discovered using protein arrays and

proteomics tools appear to be promising for use as predictive serum biomarkers. The collaborative plans are to proceed with both approaches– endoscopic ultrasound guided aspirates with genomic application of potential biomarkers for pancreatic transformation and ongoing protein discovery.

Investigators are working to identify proteins that commonly induce an immune response in pancreatic cancer. Sera from 36 newly diagnosed patients with pancreatic cancer, 18 patients with chronic pancreatitis, 33 patients with other cancers, and 15 healthy subjects were analyzed. Autoantibodies to two forms of the protein calreticulin were detected in 40-50% of patients with pancreatic cancers but seen only rarely in the other subjects. Further analysis suggests that these tumor-associated antigens may have utility as biomarkers for the early diagnosis of pancreatic cancer.

At the University of Pittsburgh, investigators analyzing sera from patients with pancreatic cancers and people without the disease, found five markers that exhibited levels significantly different between the groups. Future plans are to evaluate these five markers along with other markers under a Phase II biomarker validation trial for pancreatic cancer detection.

Investigators at the Van Andel Institute, MI are using a new detection method for antibody microarrays called two-color rolling circle amplification that allows high sensitivity detection of a wide range of proteins. The method measures the relative levels of proteins from two serum samples that have been captured on antibody microarrays. The samples came from five disease conditions: healthy, benign GI tract disease, pancreatitis, other cancers, and pancreatic adenocarcinoma. The preliminary data suggest that antibody panels identified using this method may be useful in the early detection of pancreatic cancer. These antibodies will be developed further.

Table 8 Summary of the Developmental Status of Pancreatic Cancer Biomarkers

Candidate Biomarker	Phase I			Phase II	Phase III	
	Discovery	Predictive Analysis	Assay Refinement	Characterization (Blinded)	Retro-Longitudinal	Comments
Maspin, ATDC, FXYD3 (Serum)	→					Validation needed
CEA1 (Serum)	→	→				Phase II candidate
Calreticulum (serum)	→	→				Phase II candidate
Antibody array (Serum)	→	→				Evaluation plan in process
EUS Aspirate	→	→				Large panel needed

HEPATOCELLULAR CARCINOMA

EDRN investigators are already validating des-gamma carboxyprothrombin (DCP) as a potential biomarker for hepatocellular cancer (see the next section, *Translating Discovery to Clinical Applications*). They also continue to investigate other biomarkers in hope that in the cohort being assembled for the validation study, newly discovered biomarkers could be tested quickly, efficiently and economically.

A collaborative translational project between the investigators from the University of Michigan and Drexel has identified a potential new biomarker for the early diagnosis of hepatocellular carcinoma. The protein GP73 was identified in serum samples from patients with hepatocellular carcinomas. GP73 was originally described as a Golgi type II transmembrane protein expressed primarily in epithelial cells of many human tissues. In human liver, expression was found in biliary epithelial cells, but was barely detectable in normal hepatocytes. However, expression of GP73 was found to be strongly up-regulated in hepatocytes from patients with advanced liver disease of both viral and non-viral origin. A collaborative EDRN phase I project is being performed to determine if 1) GP73 could be detected in the serum, 2) serum GP73 levels are higher in patients with liver disease, particularly those with cirrhosis and HCC, 3) serum GP73 levels are higher in patients with liver disease due to chronic HCV infection versus non-viral causes, and 4) GP73 is a better serum marker than alpha-feto protein (AFP) for differentiating HCC from non-malignant chronic liver disease.

Table 9 Summary of the Developmental Status of Hepatocelluar Carcinoma Cancer Biomarkers

Candidate Biomarker	Phase I			Phase II	Phase III	
	Discovery	Predictive Analysis	Assay Refinement	Characterization (Blinded)	Retro-Longitudinal	Comments
DCP (Serum)	→	→	→	→		Currently under network validation
GP73 (Serum)	→					Antibody being produced
SAP (Serum)	→					Antibody being produced
SELDI-MS-TOF Profile (Serum)	→	→				Assay refinement

ESOPHAGUS

Despite advances in surgical technique and multimodal therapy, the 5 year survival rate for esophageal cancer remains dismal at 5-15%. Most patients have advanced disease when they are diagnosed, and the disease often recurs, with both factors contributing to the low survival. About three times more common in males than females, this cancer occurs in more than 14,500 people each year and causes 11,550 deaths. Developing and refining methods for early cancer detection will be a means of improving survival from this deadly disease.

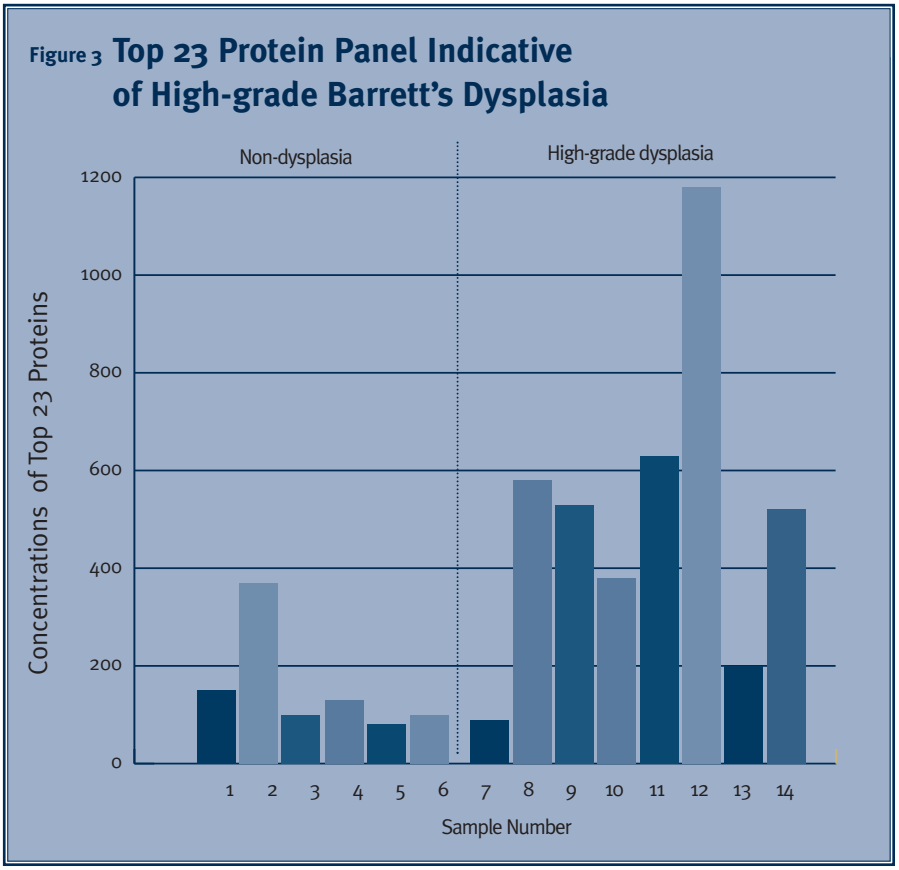
Because the condition known as Barrett's esophagus frequently precedes esophageal cancer, DNA promoter **methylation** in Barrett's esophagus tissue is under consideration as a potential cancer biomarker. Investigators at the University of Maryland performed methylation assays on tissue samples from a group of 14 Barrett's esophagus patients known to have progressed to develop dysplasia and/or adenocarcinoma along with a group of 17 Barrett's eosphagus patients who have not progressed to dysplasia and/or adenocarcinoma. These studies revealed a statistically significant trend toward more hypermethylation in tissues from patients with Barrett's esophagus who later progressed to develop dysplasia or adenocarcinoma.

Investigators have now begun measuring methylation of circulating genes using plasma from patients with Barrett's esophagus, dysplasia and adenocarcinoma. Among 24 patients with adenocarcinoma in an initial study, 19 (79%) showed detectable specific gene methylation in their plasma, while none of 8 healthy control subjects showed this finding. The high frequency of specific gene methylation in plasma among patients with adenocarcinoma suggests that this event may be an important prognostic biomarker of esophageal adenocarcinoma. It is also conceivable that the specific gene methylation may be an early event in esophageal carcinogenesis.

Studies being conducted at the University of Michigan have shown that Cathepsin B (CTSB) along with *GATA4* are overexpressed in esophageal adenocarcinomas. CTSB is a secreted protein involved in extracellular matrix degradation and tumor cell invasion. Serum protein levels from patients with CTSB amplified tumors, non-amplified tumors, or high-grade dysplasia were then examined and CTSB serum protein levels varied tremendously between serum samples. This indicates that CTSB would not provide a useful alternative to measuring *GATA4* DNA in serum samples.

Investigators have also tried to identify protein biomarkers that might be useful for detection of dysplastic Barrett's. Relative concentrations of the top 23 proteins that showed significant differences between non-dysplastic Barrett's metaplasia samples and the six samples of Barrett's with high-grade dysplasia are shown in Figure 3. The spots that were observed to be optimal candidates were picked and subjected to analysis. Proteins that already have commercial antibodies available are currently being chosen for further validation using cryosections or formalin-fixed sections of Barrett's mucosa, high-grade dysplasia and adenocarcinomas.

Figure 3 **Top 23 Protein Panel Indicative of High-grade Barrett's Dysplasia**



Lung and Upper Aerodigestive Cancers

Lung cancer continues to be the most common cause of cancer death, with more than 163,000 lung cancer deaths expected in 2005, and the search for lung cancer biomarkers continues to challenge EDRN investigators. Unfortunately, the disease is so heterogeneous that a single marker often fails to meet the expectations of desirable **sensitivity** and **specificity**. EDRN investigators are trying to create an ensemble of biomarkers that may collectively offer better performance characteristics. Some of the biomarkers towards this panel are described below.

Johns Hopkins University has explored aberrant DNA **methylation** patterns in bronchoalveolar lavage (BAL) samples from lung cancer patients. BAL from 31 patients with primary lung tumors and matching samples from 10 healthy individuals were examined for aberrant methylation of eight gene promoters. Hypermethylation of at least one of the genes was detected in all 31 lung primary tumors and BAL samples from the same patients showed a similar profile in hypermethylation as that found in the tumors. In contrast, BAL samples from the 10 control subjects without evidence of cancer revealed no methylation for five genes of this panel and low levels of methylation for the remaining three genes. These findings suggest that promoter hypermethylation in BAL can be detected in the majority of lung cancer patients prompting further development as a promising lung cancer biomarker panel.

To expand the panel of useful markers, a number of other EDRN labs are pursuing additional genes that appear to be hypermethylated in lung cancer. Death-associated

Table 10 Summary of the Developmental Status of Barrett's Progression/Esophageal Cancer Biomarkers

Candidate Biomarker	Phase I			Phase II	Phase III	
	Discovery	Predictive Analysis	Assay Refinement	Characterization (Blinded)	Retro-Longitudinal	Comments
2D GelProteomics (Serum)	→					Will result in new proteins
Methylation Panel (Serum)	→					Requires more Phase I
Gene Amplicons (Tissue)	→	→				Requires more Phase I
Ploidy (Tissue)	→	→	→			Preparing for Phase II
CyclinD1A870GSNP (Tissue)	→					New data from MD Anderson

protein kinase *(DAPK)* is often aberrantly methylated and linked with a concomitant down-regulation in expression. Investigators at the Moffitt Cancer Center are studying silencing via methylation in non-small cell lung cancers of both facilitative and inhibitory genes regulating the TGF-ß signal pathway. In a preliminary study, it was demonstrated that in cell lines, silencing of these genes is associated with promoter methylation and reactivation occurs after treatment with a demethylating agent. In primary lung cancers, methylation frequencies were rather high ranging from 23% to 62% and at least one gene from this set was methylated in 81% of tumors. These findings indicate that the TGF-ß signaling pathway is frequently deregulated in lung cancers via epigenetic changes of facilitative and inhibitory genes.

Recent reports of mutations in the epidermal growth factor receptor gene *(EGFR)* in lung cancers have generated considerable interest because they predict sensitivity to EGFR kinase inhibitors. To further understand the role of *EGFR* mutations in the pathogenesis of lung cancers, 653 lung cancers and 243 other epithelial cancers were sequenced and compared. *EGFR* mutations were present in 20% of non-small cell lung cancers and were absent in other types of carcinomas. *EGFR* mutations were somatic (nonhereditary) in origin, found in cancers showing an adenocarcinoma histology, were significantly more frequent in never smokers, people from Oriental countries (Japan and Taiwan), and in women. Mutations were not related to patient age, clinical stage, bronchioloalveolar histologic features, or patient survival. Mutations of *K-ras* were present in 8% of lung cancers and while they also targeted adenocarcinoma histology, mutations in both genes were never present in individual tumors. These findings demonstrate that geographic origin, absence of smoke exposure, gender and histological type influence the frequencies of *EGFR* mutations in lung cancers. Furthermore, the pathogenesis of *EGFR* mutant and *K-ras* mutant adenocarcinomas are different and unidentified carcinogen(s) contribute to the origin of lung cancers arising in never smokers.

Transforming growth factor ß (TGFß) regulates growth and differentiation in normal squamous epithelium via the interaction with specific receptors and intracellular signaling molecules (Smads). A decrease in the TGFß type II receptor (TßR-II) expression is believed to be partly responsible for the resistance of tumor cell lines and 85% of malignant tumors to the anti-proliferative effects of TGFß. As previously shown for breast and head and neck tumors, the expression of TßR-II is greatly reduced in all types of lung carcinomas examined. In contrast, the reduction of Smad2 was dependent on tumor type, with only undifferentiated carcinomas showing a statistically significant reduction. In this series of lung carcinomas, a **sensitivity** of 80% and a **specificity** of 69% were observed for the TßR-II antibody. However, when other antibodies examined in this study are also included, both sensitivity and specificity increase, to 91% and 78%, respectively. These data suggest that defects in the TGFß signaling pathway are common in lung carcinomas and could be exploited as potential detection and diagnostic markers.

Advances in protein tagging, fractionation, and mass spectrometry have made it possible to detect, analyze, and identify proteins in tumor tissue, serum and plasma at an unprecedented level of sensitivity. **SELDI-TOF-MS** was used to generate protein profiles in malignant lung tumors, and premalignant airway epithelium showing neoplastic transformation. Lung tumor specimens taken from patients participating in a lung cancer screening study at the H. Lee Moffitt Cancer Center were laser capture microdissected to obtain pure cell populations from frozen sections of normal lung, atypical adenomatous hyperplasia and malignant tumors. Three proteins in tumor samples were markedly increased compared with normal cells and one was not detected in any of the normal cells. Although additional study is ongoing to validate these patterns as unique diagnostic tools, these “malignant” protein signatures lend themselves to identification of populations at high-risk for lung cancer and for monitoring response to lung cancer chemopreventive agents.

As part of its ongoing research activity, investigators at the Moffit Cancer Center are also conducting a prospective, longitudinal, single-arm cohort screening trial, with the hypothesis that screening with CT and sputum molecular markers will increase the proportion of stage I cancers to over 60% of total cohort lung cancers while reducing the advanced stage cancers to less than 40% (stage shift). Samples collected at the Center have the potential to assess the risk of developing lung cancer and might inform who should be screened to detect lung cancer early, at a localized stage when current therapies might lead to cure.

The group at the Fred Hutchinson Cancer Center is pursuing several **proteomic** approaches for developing markers for early detection. The humoral immunogenic response against tumor antigens that occurs in cancer is being exploited at the University of Michigan to develop a screening test for early detection. Several tumor antigens that induce an antibody response in lung cancer have been identified that are currently going through EDRN validation. If successfully validated, these markers would be highly beneficial for developing strategies for early lung cancer detection, either as a stand alone screening procedure, or to complement other modalities such as CT screening.

Recently, an innovative strategy that allows **microarray**-based display of tumor lysate proteins has been implemented. The microarray-based approach has a much higher **throughput** than **2-D** gels with improved quantitation of antigen/antibody reactions. Proteins are spotted on microarrays and interrogated with subject sera. Illustrative of this approach is the identification of anti-UCHL3 auto-antibodies in sera from colon cancer patients. This establishes the potential of natural protein microarrays in a high throughput approach to uncover cancer antigens that induce an antibody response. Such a paradigm is being pursued to allow rapid screening of large numbers of sera from lung cancer patients.

Investigators at the University of Colorado have focused on enhancing sputum cytology for detection of precancerous growth and cancer by **FISH**. They have expanded on conventional morphology by applying new multi-color FISH to the analysis of malignant and dysplastic sputa from high risk patients. Two findings emerge from these studies. One is the high-frequency (26%) of **aneuploidy** in the sputa of individuals at high-risk for lung carcinoma. A second finding is that aneuploidy is present in the sputum of 40-50% of individuals who subsequently develop carcinoma within 12 months of analysis. In combination with abnormal sputum cytology, the predictive power of these tests is 83%. Future plans include the evaluation of automated image analysis for the detection of aneuploidy in sputum cytology samples. They will also be testing additional probes from the Vysis Corporation for their sensitivity and specificity in the early detection of lung carcinoma in high-risk patients.

Table 11 Summary of the Developmental Status of Lung Cancer Biomarkers

Candidate Biomarker	Phase I			Phase II	Phase III	
	Discovery	Predictive Analysis	Assay Refinement	Characterization (Blinded)	Retro-Longitudinal	Comments
Autoantibody screening (Annexins I,II, PGP 9.5)	→	→	→			Developing Validation Proposal
MALDI-TOF-MS Profile (Serum)	→	→	→	→		Developing Validation Proposal
SMRP (Mesothelioma marker)	→	→	→	→		Developing Validation Proposal
Hypermethylation (4-5 gene panel in serum)	→	→	→			In Pre-validation
Loss of TGFß Receptor-II	→	→				Promising in Tumor, Transitioning to Sputum
Sputum cytology/FISH	→					Early Discovery
Telomerase	→					Early Discovery

Prostate and Other Urologic Cancers

EDRN prostate cancer investigators are developing biomarkers for early detection of prostate cancer, to predict which tumor is likely to progress and become aggressive, as well as biomarkers for assessment of risk for prostate cancer. In research focused on biomarkers to one day replace prostate-specific antigen (PSA), EDRN investigators are developing and/or testing the following biomarkers.

Efforts are under way to identify patterns of prostate cancer serum immunoreactivity to characterize the immune responses elicited by tumors. **Microarrays** of tumor-derived proteins are being used to profile the antibody repertoire in sera of prostate cancer patients and controls. Proteins from a prostate cancer cell line were separated into 1760 fractions, each of which were screened against serum samples from 25 men with prostate cancer and 25 male controls. Statistical analysis revealed that 38 of the fractions had significantly higher levels of antibody binding in the prostate cancer samples compared to the controls and two fractions showed higher binding in the control samples. The significantly higher antibody reactivity may reflect a strong immune response to the tumors in the prostate cancer patients. Analysis was used to classify the samples as either prostate cancer or control with 84% accuracy. Adding a decision tree with two levels of partitioning helped to classify the samples with 98% accuracy. These results suggest that microarrays of fractionated proteins could be a powerful tool for tumor antigen discovery and cancer diagnosis.

Investigators at the Eastern Virginia Medical School are employing **mass spectrometry**-assisted immunoassay to optimize the utility of protein expression profiling. The approach builds upon the concept that disease-specific actions cause the cleavage of proteins that, in turn, amplify the signal to cause more cleavage of proteins. These signals will be fragments of whole proteins and cannot be distinguished by standard **ELISA** immunoassays. The EVMS researchers have shown that fragments of relatively abundant whole proteins, such as apoplipoproteins, can be used to discriminate the presence of cancer, especially in men with PSA levels lower than 4 ng/ml. A multiplexed immunoassay consisting of several families of amplified protein pieces is currently being examined for the ability to distinguish cancer from benign disease in men with marginal clinical status.

In collaboration with Ciphergen Biosystems, investigators have formed a collaborative network, the EDRN Prostate Cancer SELDI Investigational Collaborative (EPSIC), to evaluate the utility of protein expression profiling for the early detection of prostate cancer. The investigators have completed the first phase of the study, which evaluated the analytical reproducibility of the diagnostic platform. The second phase will evaluate the clinical utility of the system in 1,000 men. In parallel, a study lead by investigators as Johns Hopkins is identifying a panel of serum proteins to discriminate men with prostate cancer confined to the prostate gland from men with benign prostate disease. A set of 345 men who had a stored serum sample available were included in this study. The most optimal panel of biomarkers for maximum separation of the prostate cancer and the benign prostate disease cohorts was revealed using protein chip arrays. A panel of 3 proteins was selected and tested separately and in combination; all three proteins exhibited greater **specificity** with prostate cancer than PSA and at a specificity range of 30% to 80%. The combination of the three chips showed significant improvement over PSA.

In collaboration with Matritech, Inc., EDRN investigators have identified in the serum of an individual with prostate cancer a protein previously called NMP48, which is related to vitamin D-binding protein. To further investigate the possible relationship of this serum protein to prostate cancer, serum samples were obtained from men with biopsy-confirmed prostate cancer, high-grade prostatic intraepithelial neoplasia, and benign prostatic hyperplasia. In 52 samples, the NMP48 protein was found in 96% of the sera from individuals with prostate cancer including 11 of 12 specimens that exhibited prostate-specific antigen values of less than 4 ng/ml. NMP48 was found in 10 of 19 samples from men with prostatic intraepithelial neoplasia. By comparison, it was not detected in over 70% of sera obtained from men with benign prostatic hyperplasia, in 80% of patients who had previously undergone radical prostatectomy, or in 96% of specimens from healthy controls.

Using three Dunning rat prostate cancer cell lines, investigators identified a protein that is overexpressed in the types of prostate cancer most likely to spread throughout the body (metastatic). After this protein has been sequenced, it will be characterized using stored serum from patients with varying stages of disease to determine if this novel protein can serve as a clinically useful biomarker.

Research aimed at improving PSA includes:

- EDRN centers exploring the clinical utility of proenzyme PSA, in collaboration with Hybritech Beckman Coulter, Inc., recently discovered that free, uncomplexed PSA in serum is more complicated than originally thought. Free PSA (fPSA) is now known to be comprised of at least three distinct forms: (1) a proenzyme or precursor form (pPSA) that is associated with cancer, (2) an internally cleaved or degraded form of PSA, termed BPSA, that is more highly associated with benign prostate hyperplasia, and (3) intact, enzymatically inactive forms. Truncated forms of pPSA are not detected in seminal plasma, but are found to be elevated in peripheral zone cancer tissues and can be detected in serum using specifically designed immunoassays. The specificity to detect cancer was significantly greater for pPSA than for fPSA. Thus, in the 2.5 to 4.0 ng/ml total PSA range, 75% of cancers can potentially be detected with 59% of unnecessary biopsies being spared by determining the fraction of pPSA as compared to sparing only 33% of unnecessary biopsies when considering the fraction of fPSA.
- A second study examined the role of pPSA in the detection of prostate cancer in 93 men (44% cancer, 56% non-cancer) with a total PSA in the range of 4-10 ng/ml. Using multivariate logistic regression, a model with total PSA, fPSA and all pPSA forms was predictive for prostate cancer detection at 90% **sensitivity** and 44% specificity. This model proved significantly more predictive of prostate cancer than that determined by analysis of any individual PSA species. In summary, these initial studies suggest pPSA has promise as a biomarker to improve prostate cancer detection in the 2.5-10 ng/mL total PSA range.

BLADDER CANCER

A major accomplishment of EDRN investigators working in bladder cancer is the 3-year study to validate a test to detect the recurrence of bladder cancer that has been initiated by NCI at 13 centers across the United States and Canada. By examining genetic changes in DNA obtained through urine samples, the test, if successfully validated, will provide a sensitive and non-invasive method of screening for bladder cancer recurrence.

Table 12 Summary of the Developmental Status of Prostate Cancer Biomarkers

Candidate Biomarker	Phase I (Discovery and Early Refinements)			Phase II	Phase III	
	Discovery	Predictive Analysis	Assay Refinement	Characterization (Blinded)	Retro-Longitudinal	Comments
GSTP1, Methylation (tissue)	→	→	→			Promising prevalidation
SELDI-MS-TOF profile (serum)	→	→	→	→		Promising prevalidation
Nkx3.1	→					Promising Negative Predictive Value (NPV)
pPSA (serum)	→	→				Promising NPV
HK2	→	→				Promising NPV
DD3/PCA3	→	→				Promising NPV
uPM3	→	→				Promising NPV
N-Methylacyl-CoA Racemase (AMACR)	→					Overexpressed in prostate cancer and detection of autoantibodies against the protein
Hepsin	→					Overexpressed in prostate cancer
pim-l	→					Overexpressed in prostate cancer
EZH2	→					(advanced metastasis prostate and breast cancers)
hKLK5 (splice variant 1)	→					

In addition, EDRN investigators have developed a model of human urinary bladder cancer progression from in situ precursor lesions to invasive carcinoma using whole-organ histologic and genetic mapping. The model is based on the analysis of chromosomes 1-22 in over 50,000 tests and was developed by extensive studies performed over the last decade.

Investigations of the deletion patterns of genes in the 13q14 chromosomal region of the cell have implicated the involvement of a number of genes (*P2RY5, ITM2B, CHC1L*) in the earliest stages of carcinogenesis, but prior to the expression of the much analyzed retinoblastoma gene (*RB1*). The term "forerunner genes" is used in referring to them because they may be involved in cancer progression prior to the involvement of their neighboring tumor suppressor *RB1* gene. Since the loss of just one *RB1* allele is also an early event associated with the beginning of cancer, the *P2RY5* gene located within an intron of *RB1* may represent a candidate forerunner gene. The *ITM2B* gene encodes a mitochondrial membrane protein that induces apoptosis (programmed cell death). The *CHC1L* gene encodes a ras-related protein, while *P2RY5* encodes a receptor which may affect the proliferation rate of cells.

Inactivation or reduced expression of these genes may increase cell survival and provide a growth advantage that drives the initial expansion of the first cancer cells.

Overall, the data indicate that three forerunner genes, *ITM2B*, *P2RY5*, and *CHC1L*, located in the immediate vicinity of *RB1*, are inactivated by various mechanisms in an early phase of bladder carcinogenesis and that the loss of their function drives the initial growth of cells with preneoplastic changes. It also shows that inherited mutations in forerunner genes may represent a novel risk factor for cancer development. These observations are confirmed by studies showing that down-regulation or loss of forerunner genes provides a growth advantage for cells lining the bladder. In contrast, reintroduction of active forerunner genes into cells with inactivated forerunners, significantly reduced their proliferation rate.

Another potential biomarker is a novel oncogene (cancer promoting gene), Aurora-A, that is frequently over expressed in different human cancers, including bladder cancer. The over expression of a protein that acts as a regulatory switch to this gene, called Aurora-A kinase, induces **aneuploidy** and other chromosomal abnormalities in human cells. These results demonstrate that Aurora-A is a key intermediary in multiple pathways that regulate many different cellular characteristics affected during the transformation of healthy cells into tumor cells. Published reports have identified several Aurora-A-interacting proteins, which appear to be involved in the regulation of these pathways. It is hypothesized that each Aurora-A-interacting protein has the potential of being an early detection biomarker.

Investigators at the University of Texas M. D. Anderson Cancer Center have identified a set of Aurora-A kinase substrate proteins (proteins that are also regulated by this kinase) based on their involvement in Aurora-A kinase-related pathways regulating cell division processes. The study demonstrated that when Aurora-A kinase is over expressed in human bladder tumors, there is a loss of the activity of the tumor suppressor gene *p53*.

FISH analysis demonstrated that most **diploid** and near-diploid bladder cancers had 3-4 copies of Aurora-A (normal cells contain 2 copies), while aneuploid bladder cancers frequently contain more than 4 copies of Aurora-A per cell.

Immunohistochemical analysis demonstrated that overexpression of Aurora-A was associated with decreased metastasis-free interval and survival. These preliminary data suggest that Aurora-A has increased expression in bladder cancer and high (>4) copy numbers are associated with high-grade aneuploid aggressive bladder cancers. In contrast, low level of Aurora-A amplification (3-4 copies) is associated with low-grade superficial bladder neoplasia. Recent analysis on exfoliated cells in voided urine specimens confirms that amplification of Aurora-A is ubiquitous in bladder cancer. In addition, FISH analysis of 20 samples of voided urine from healthy individuals showed the normal 2 copies of the gene. These data demonstrate that FISH analysis of Aurora-A in voided urine is an extremely promising molecular marker for detection of bladder cancer and is a candidate for validation by the EDRN.

In collaboration with EDRN researchers at the University of Texas Southwestern Medical Center, investigators are also studying methylation status of tumor suppressor genes in bladder cancer. The **methylation** status of 10 genes was determined in 98 bladder tumors and the methylation index (MI), a reflection of the methylated fraction of the genes, was determined. The MIs of this panel ranged from 2% to 36% and correlated significantly with several parameters of poor prognosis (tumor grade, growth pattern, muscle invasion and aneuploidy). Methylation of the genes *CDH1* and *FHIT* and a high MI were associated with shortened survival. *CDH1* methylation positive status was independently associated with poor survival. They also examined 54 samples of voided urine from patients with bladder cancers. The methylation profile of the urine samples closely mimicked that of bladder cancers. These results suggest that the methylation profile of voided urine may be a potential biomarker of risk prediction and early detection of bladder cancer.

In a recent study, methylation of laminin-5 genes correctly distinguished invasive from non-invasive tumors, and their methylation frequencies in urine samples (from cancer patients) mimicked the tumor patterns, indicating that urine examination was an excellent method to detect invasive cancers. Methylation of *LAMA3* was most frequent in both cancers and urines. Methylation of *LAMA3* and *LAMB3* were significantly associated with growth pattern, muscle invasion, tumor grade, tumor stage, tumor ploidy but not with survival. While methylation of *LAMC2* was relatively infrequent, it correlated with survival. Thus, the methylation status of *LAMA3* and *LAMB3* correlate with multiple important clinico-pathological features, while methylation of *LAMC2* is an independent marker for prognosis.

GSTP1 Methylation Detects Prostate Cancer with Greater Accuracy

Investigators at Johns Hopkins have evaluated the hypermethylation of the glutathione S-transferase P1 gene *(GSTP1)* that occurs at a very high frequency in prostate adenocarcinoma. In a blinded study, the histologic review of biopsy samples from 72 excised prostates were compared with those obtained using **methylation-specific PCR (MSP)** for *GSTP1*. Formal surgical pathologic review of the resected prostates was used to determine the number of patients with (n = 61) and without (n = 11) prostate cancer. Histology alone detected prostate carcinoma with 64% sensitivity and 100% specificity, whereas the combination of histology and *GSTP1* MSP at an assay threshold greater than 10 improved detection of prostate carcinoma to 75% sensitivity and 100% specificity. The combination of histology and *GSTP1* MSP at an assay threshold greater than 5 detected prostate adenocarcinoma with 79% sensitivity over histology alone. Thus, *GSTP1* MSP improved the sensitivity of histologic review of random needle biopsies for prostate cancer diagnosis. Further studies should determine whether detection of *GSTP1* hypermethylation in a biopsy sample with normal histology indicates the need for an early repeat biopsy at the same site.

Table 13 Summary of the Developmental Status of Bladder Cancer Biomarkers

Candidate Biomarker	Phase I (Discovery and Early Refinements)			Phase II	Phase III	
	Discovery	Predictive Analysis	Assay Refinement	Characterization (Blinded)	Retro-Longitudinal	Comments
Aurora A	→	→				Promising prevalidation
Aurora B & C	→					Discovery Phase
FR Genes around RB1	→	→				Promising preclinical/clinical
FR Genes around p53	→					Discovery Phase
Urine Proteomics	→	→				Promising prevalidation
Urine (methylation profile)	→	→				Promising prevalidation
E-cadherin	→	→				Promising prognostic
Zyxin	→	→				Promising prognostic
Moesin	→	→				Promising prognostic
Caveolin	→	→				Promising prognostic
Kiss-1	→	→				Promising prognostic Loss of expression is associated with progression
P33ING1	→	→				Promising prognostic
MSA	→	→	→	→	→	Validation phase

Translating Discovery to Clinical Applications: Delivery

Since its inception, the Network has initiated validation studies and developed concepts for several more in the quest to narrow the vast field of potential and promising biological markers for the early detection of cancer. These efforts underscore the Network's successful steps forward in fulfilling its mission to deliver translational studies necessary to meet NCI's 2015 goal of eliminating the suffering and death due to cancer.

This section highlights the Network paradigm for validating biomarkers; ongoing Network validation studies utilizing this approach; and current studies exploring emerging technologies for discovery, development and validation of cancer biomarkers.

Validation Paradigm and Standards

The Network developed a new validation paradigm for setting standards for reagents and study designs. The approach is intended to expedite methods to evaluate and validate biomarkers for clinical application during the early stages of investigation.

The concept of validation asks:

- Whether a test is clearly described;
- If the true presence or absence of disease can be established for all individuals;
- Whether the spectrum of patients with and without disease is adequate;
- Whether the assessment of test and disease status is conducted in an unbiased manner; and
- If the test performance is summarized by the important terms of **sensitivity** and **specificity**.

Adapted from the National Aeronautics and Space Administration's Technology Readiness Levels, the Network paradigm is designed to evaluate and measure a biomarker's readiness level (BRL) for validation.

In brief, these levels are:

1) BRL1: This is the lowest level of evidence at which scientific research or characteristics are identified and translational research is proposed. The evidence may derive from experiments in animal models, cell cultures, or human specimens.

2) BRL2: At this level, practical application characteristics including technology assessment can be identified and evaluated.

3) BRL3: At this level of maturation, active research and development is initiated. This includes proof of principle in support of addressing the clinical question: detection, diagnosis, or prognosis.

4) BRL4: At this level, the system or components of the system, i.e., biomarker and associated assay, must be integrated to establish that the "pieces" will work together in the laboratory environment.

5) BRL5: The fidelity of the system or component has to be tested in appropriate samples or cohorts.

6) BRL6: This level represents a model or prototype that is ready to be tested in a controlled setting under rigorous experimental protocols.

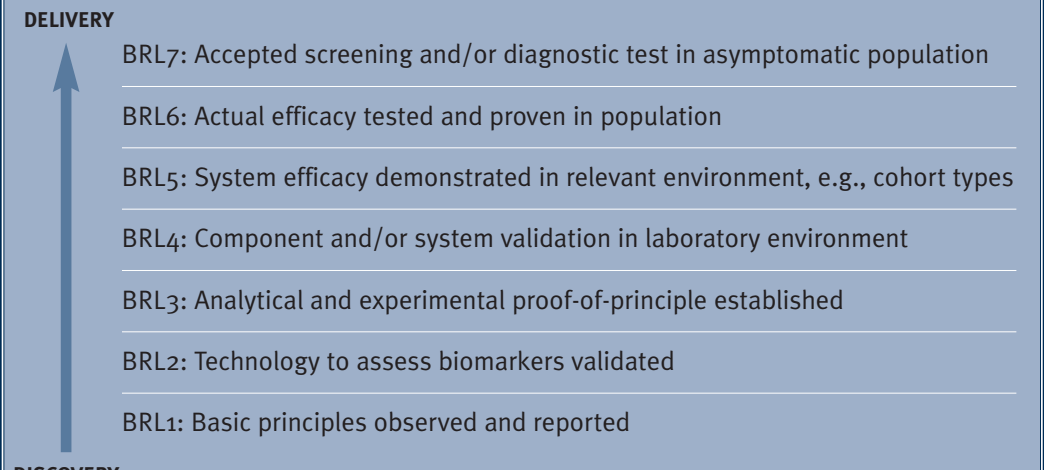
7) BRL7: By now, the system is fully validated and ready for use in the population.

FDA Regulations for Computerized Systems in Clinical Trials

The Network developed a robust framework called the Validation Study Information Management System (VSIMS) to allow multiple studies to be administered efficiently-minimizing both development time and standardization of information and data management across studies and multiple research sites.

VSIMS (discussed in more detail in the informatics section of this report) is a secure web-based system that consists of a toolkit of key modules for tasks such as forms creation, data entry, and specimen tracking. These tools are customized for each instantiation for a particular study.

Figure 1 Biomarkers Readiness Levels for Validation



Adapted from NASA

Development of VSIMS occurred simultaneously with the planning stages of the Network's first clinical validation study, Detection of Bladder Cancer by **Microsatellite Analysis (MSA).** Described in detail later in this section, this study required that MSA's instance of VSIMS (called VSIMS-MSA) be validated for compliance with FDA regulations (21 CFR Part 11) for computerized systems in clinical trials as the results may be used to apply for FDA approval.

VSIMS-MSA is a category-5 custom software package, and the FDA regulation requires extensive testing and documentation. A risk assessment was performed first to determine the scope of validation required for each component of VSIMS-MSA. Steps required for the FDA validation process for VSIMS-MSA are shown below. (Documentation of all steps is on file at the Network's Data Management and Coordinating Center.) With initial validation complete, any changes to the validated system must follow the Change Control Standard Operating Procedure, in which Steps 2 through 4 are repeated. Any future instance of VSIMS used in a Network validation study that may support an FDA application will need to be similarly validated.

While validating regulatory compliance for future instances of VSIMS may be easier because templates were developed for use in the various steps in this study, it would still require the same level of testing and documentation to validate the study-specific components.

Figure 2 Steps Required for FDA Validation of VSIMS-MSA

Step 1 Validation Master Plan and Project Plan

- Develop master plan to complete all deliverables with description of all items and procedures to create, execute, and approve each deliverable

Step 2 Develop Validation Materials (see table below)

- Perform a dry run of the Operational Qualification test cases as the System Test
- Migrate the system to the Validation Environment

Step 3 Execute Validation Test Scripts

- Execute the Installation Qualification Test Cases in the Validation Environment
- Prepare the Installation Qualification Summary Report
- Execute the Operational Qualification Test Cases in the Validation Environment
- Prepare the Operational Qualification Summary Report
- Execute the Installation Qualification Test Cases on the Production Environment
- Prepare the Installation Qualification Summary Report
- Execute the Performance Qualification Test Cases in the Production Environment
- Prepare the Performance Qualification Summary Report

Step 4 Prepare Validation Final Report

- Prepare the Validation Summary Report
- Collate and review all of the Validation Deliverables

Figure 3 Validation Materials to Meet FDA Approval

Deliverable	Purpose
Section 1: Validation Master Plan and Project Plan	Overall plan to create all deliverables with description of all items, and procedures to create, execute, and approve each deliverable. Includes project plan with tasks, time, milestones
Section 2: Validation Materials Developer/Validation Team Training Plan	Training requirements and plan to train development team. Includes training material in appendix for: • General FDA/regulatory background • Good documentation practices • Good document management practices • 21 CFR 11 regulation • Computer system validation (general) • Validation Plan and all associated procedures • FDA guidance for clinical computer systems
Risk Assessment	Identification and evaluation of specific risks and corresponding design or mitigation controls. Overall determination of criticality of system sections or modules, with matching to required degree of testing for each
Issues Log	Identification, evaluation, and resolution of compliance-related issues during development and testing
System Description	Overall description of system, with summaries of business functions and technical infrastructure specification
User Requirements	Written statements of business-level requirements and system rules
Functional Specification System	System-level description of system design, including program structure, abstracts, database design, and all system design requirements
Program/Object Specifications	Narrative of each program or system object
Infrastructure Specification	Description of infrastructure requirements including servers, storage, network, workstations, communications, operating system(s), and support utilities. Includes infrastructure schematic
Programming/ Development Standards	Standards and rules for technical development and coding
Code Review	Independent review of each program/object source or structure against standards and specifications
Validation Protocol	Detailed strategy and plan for system testing, including appendix with detailed procedures for writing, approving, executing, documenting, reviewing tests, including deviation management

Figure 3 Validation Materials (continued)

Test Cases/Scripts -Installation Test (IQ) -Operational Test (OQ) -Performance Test (OQ)	Detailed testing scripts for installation of infrastructure (IQ), testing system functionality (OQ), and user-acceptance (PQ) Test case structure and outline
Trace Matrix	Mapping of each URS requirement to functional specification(s), technical specifications, to test case(s) that proves each requirement
Section 3: **Execution of Validation Materials** Developer/Validation Training records	Training or qualification records for both technical and system areas for development team. Training in appendices for both teams. Training will be conducted, documented, and records added to validation package
Test Execution/Report -Installation -Operational -Performance	Execution documents (completed test cases) with supporting evidence (screen prints, reports)
Incident Reports	Explanation of unexpected test results or deviation from test protocol/case, with documentation of investigation, root cause determination, resolution, and retesting.
Summary Reports -IQ -OQ -PQ	Summary of system testing, with conclusions (with supporting rationale) as to acceptance of each stage and system in total
System Manual(s)	Documentation of system design, installation, and use (user manuals), tested to assure correctness
Section 4: **Validation Final Report** Validation Report/ System Release	Summary of entire validation process and documentation, with conclusion (with supporting rationale) of acceptance of system validation team. Validation package complete set of all above deliverables, organized into binders, with TOC and supporting notations
Section 5: **Training and** **Interpretation Guidance** 21 CFR 11 Part Assessment	Explanation of how system complies with 21 CFR 11, including validation requirements, for each phase of the regulation.

Current Studies Applying the Standards

Key studies applying the standards of the Network validation paradigm are in progress. These efforts investigate multiple biomarkers to determine their capacity for distinguishing between people with cancer and those without.

As described below, Network investigators are examining a comprehensive program to validate early detection of prostate cancer with novel protein identification techniques; validate serum markers for the early detection of hepatocellular carcinoma; detect bladder cancer by microsatellite analysis (MSA) of urinary sediment in a multi-institutional study; a new assay for telomerase; standard reagents for measuring telomerase activity (TA); quality control for mitochondrial DNA sequencing; collaborating with the human proteome organization on technology, plasma **proteomics**, and data analysis; RNA archiving; and development of model protocols for protein profiling.

Validating Protein Expression for Early Detection of Prostate Cancer

This three-phase study is a comprehensive program to validate a novel approach for early detection of prostate cancer based on protein identification techniques. It builds on protein expression profiling of body fluids by **Surface-Enhanced Laser Desorption Ionization time-of-flight mass spectrometry (SELDI-TOF-MS),** a process that enables exploring and mapping proteins. It also uses artificial intelligence algorithms.

The goal of Phase I was to assess the reproducibility and portability of SELDI-TOF-MS using protein profiles generated from serum. Phase I was successfully completed in February 2004 at six institutions using a single source of pooled sera.

Testing Technology Reproducibility and Portability for Cancer Diagnostics

Using the Network-developed staged approach for testing technology for accuracy and portability, EDRN has made strides to cross-validate six sites using **SELDI-TOF-MS** to analyze a reference set of sera from prostate cancer patients and normal subjects.

The instruments at all sites were calibrated and standardized in parallel. Each site was then presented with the same set of 14 normal sera and 14 case sera. Based on these "known" samples, all six sites were able to discriminate between normal versus cancer when applying certain classifier algorithms. Then all sites were presented with a different set of 28 "blinded" samples and challenged to determine which samples were normal and which were cancer.

Four sites classified all 28 correctly, one site called 26 of 28 correctly, and one site did not pass after correctly classifying just 19 samples. The results from this multi-institutional study demonstrate that validation is in fact feasible for protein profiling where screening and assessment of cancer can be performed in a reproducible manner by a multitude of clinical centers in a standardized manner. To this end, Phase II validation studies are continuing the development of serum protein-profiling for prostate cancer.

(*ClinChem* 2005; 51:102-12.)

The overall goal of Phase II is to develop and evaluate an algorithm for classifying cases and controls using protein profiles produced from SELDI-TOF-MS using serum collected from prostate cancer cases and non-cancer controls.

The objectives of the study are:

1. Identify serum biorepositories with clinically-characterized, appropriately collected and stored samples from five groups of patients: controls with negative biopsy, "other" cancers and inflammatory disease, cases with aggressive or advanced prostate cancer, and cases with intermediate and low-risk prostate cancer.

2. Generate protein profiles from two independent laboratories using SELDI-TOF-MS and IMAC Proteinchips™, using serum from 500 prostate cancer cases and 250 prostate cancer controls. Another group of 50 patients with other cancers but no evidence of prostate cancer and 50 patients with various inflammatory diseases will also be examined.

3. The Data Management and Coordinating Center (DMCC) will construct classifiers and develop algorithms. Two research issues loom:

   - To create a classifier that has adequate **sensitivity** (95% or greater) and **specificity** (65% or greater) to distinguish, among men who underwent prostate biopsy, those with and without prostate cancer; and

   - (Secondary) To create a classifier that has adequate sensitivity (95% or greater) and specificity (65% or greater) to distinguish, among men who underwent prostate biopsy, those with and without high-grade prostate cancer.

Validating Protein Expression for Early Detection of Prostate Cancer

Participating Institutions

Eastern Virginia Medical School, Norfolk, VA
EDRN Data Management and Coordinating Center, Fred Hutchinson Cancer Research Center, Seattle, WA
University of Alabama, Birmingham, AL
University of Texas Health Science Center, San Antonio, TX
University of Pittsburgh Cancer Institute Hillman Cancer Center, Pittsburgh, PA
Johns Hopkins Medical Institutions, Baltimore, MD
Center for Prostate Disease Research Walter Reed Army Medical Center, Rockville, MD
National Cancer Institute, SAIC Frederick, MD
National Cancer Institute, Division of Cancer Prevention, Bethesda, MD

Milestones

February 2003:	Investigators meeting and protocol development
May 2003:	Protocol approved
December 2004:	Initial results; successful completion of concordance among institutions on SELDI profile data on various machines
February 2005:	Completion of analytical validation
November 2005:	Completion of Phase II clinical validation
September 2006:	Completion of Phase III clinical validation

Validating Serum Markers for Early Detection of Hepatocellular Carcinoma

Developing more sensitive and specific serum markers for early detection of primary liver cancer, hepatocellular carcinoma (HCC), could lead to better tracking of the disease and improve patient survival.

This study aims to determine the sensitivity and specificity of the serum marker des-gamma carboxyprothrombin (DCP); compare the accuracy of DCP with alpha-fetoprotein (AFP), the only serum marker currently available for HCC; and determine whether demographic or other factors of underlying liver disease alter the expression of DCP or AFP. Cirrhosis is the major risk factor for HCC.

To achieve these aims, patients will be enrolled at six liver centers to perform a case control study of those with cirrhosis and those with early HCC. Demographics, medical history, history of liver disease, social history (attention to lifetime smoking and alcohol), etiology of liver disease, family history, and clinical and laboratory data of patients with cirrhosis and early HCC will be obtained. In addition, serum, plasma and DNA for the evaluation of markers for HCC will be obtained. All the data will be maintained in a web-based database.

The study calls for enrolling 450 HCC patients and 450 cirrhosis controls in order to target 190 early-stage HCC cases to achieve 90% power. This is an important validation study of DCP, which may lead to the development of a clinically needed marker for HCC.

Validating Serum Markers for Early Detection of Hepatocellular Carcinoma

Participating Institutions

University of Michigan Medical Center, Ann Arbor, MI
University of Pennsylvania, Philadelphia, PA
Mount Sinai Hospital, New York, NY
Mayo Clinic, Rochester, MN
Stanford University, Palo Alto, CA
St. Louis University, St. Louis, MO
National Cancer Institute, Division of Cancer Prevention, Bethesda, MD

Milestones

November 2004:	Protocol completed
February 2005:	Begin enrollment
April 2005:	Begin assays
February 2006:	End enrollment
May 2006:	Finish assays and transfer sera to NCI
May 2006:	Data analysis
November 2006:	Finish analysis

Detecting Bladder Cancer by Microsatellite Analysis (MSA) of Urinary Sediment

This prospective multi-center study looks at a promising new molecular diagnostic test for bladder cancer called **microsatellite analysis (MSA).** In clinical sites throughout the United States and Canada, the study will determine the usefulness of MSA in early detection and monitoring of superficial bladder cancer. It will also consider the utility of MSA in studies as a substitute for standard of care monitoring procedures.

Study objectives are to determine:

1. Sensitivity and specificity of MSA of urine sediment, using a panel of 15 microsatellite markers, in detecting bladder cancer in participants requiring cystoscopy; this technique will be compared to the diagnostic standard of cystoscopy, as well as to urine cytology;

2. Timeliness characteristics of MSA of urine sediment; and

3. The most predictive individual markers or combination of markers that make up the MSA test.

The study will test the:

- Specificity of MSA in two groups of participants without bladder cancer both healthy, normal controls as well as participants without bladder cancer who have conditions known to confound the performance of previous urinary tests for bladder cancer;

- Sensitivity of MSA in the presence of clinically evident, pathologically confirmed bladder cancer;

- Sensitivity and specificity of MSA for the detection of bladder cancer among participants undergoing surveillance for bladder cancer; and

- The value of MSA to predict subsequent development of bladder cancer among participants with a history of bladder cancer.

Three groups will be enrolled in the trial. Two of them will include 200 participants without bladder cancer who will serve as control groups. These two control groups will include 100 participants without a history of or current genitourinary (GU) diseases and 100 participants with one of four disease processes requiring cystoscopy [benign prostatic hypertrophy (BPH), foreign body (urinary stones, stents, and catheters), infection, and hematuria]. These conditions have historically led to false positive urinary bladder cancer screening studies.

The third group will include 300 participants with incident or recurrent superficial bladder tumors who are followed every 3 months for recurrence of their bladder tumors. MSA results will be compared with these standard examinations for recurrent disease.

Investigators will be blinded to the MSA assay results. No clinical decision regarding medical care or management will be based upon the MSA assay results alone. The contract laboratory responsible for performing the MSA assay will be blinded to all participant urinalysis, cytology, cystoscopic and pathology results to minimize bias.

Detecting Bladder Cancer by Microsatellite Analysis of Urinary Sediment

Participating Institutions

Johns Hopkins Medical Institution, Baltimore, MD
Baylor College of Medicine, Houston, TX
MD LURN (Linked Urology Research Network), Atlanta, GA
M.D. Anderson Cancer Center, Houston, TX
Memorial Sloan Kettering, New York, NY
Stanford University, Stanford, CA
University of Alabama, Birmingham, AL
University of Chicago Urology Center, Chicago, IL
University of Michigan, Ann Arbor, MI
University of Rochester Medical Center, Rochester, NY
University of Texas, San Antonio, TX
University of Toronto, Toronto, ON, Canada
Washington University, St. Louis, MO
National Cancer Institute, Division of Cancer Prevention, Bethesda, MD

Milestones

December 2003:	Protocol Approved
January 2004:	First Meeting of Investigators
October 2004:	Enrollment began
June 2005:	Interim Data Analysis begins
September 2007:	Final Results Expected

New Assays and Standard Reagents for Measuring Telomerase Activity

A number of studies have demonstrated a close association between telomerase activity (TA), hTERT mRNA expression, circulating telomerase DNA, and telomere length with multiple cancers, such as bladder, lung, esophagus, prostate, and breast. These associations have provided strong evidence and scientific rationale for using telomerase as a biomarker for cancer detection and prediction.

Challenges remain, however, in applying telomerase for clinical use. For example, in many cases only a single marker, such as TA, has been measured independently instead of testing multiple related markers, such as hTERT mRNA, circulating telomerase DNA, and telomere length from the same samples. In addition, TA results vary from different assays and different laboratories, possibly indicating its sensitivity to sample preparation and testing methods.

To overcome these problems and to validate telomerase as a biomarker for cancer early detection, scientists from academia, NIST, and NCI have been performing analytical validation of telomerase as a biomarker and developed a prototype of automatic high throughput assays/technologies for testing telomerase activity. This model uses a robot-assisted telomerase repeat amplification protocol capillary electrophoresis (RApidTRAP CE) to handle samples, perform **quantitative PCR (qPCR),** and run CE followed by data analysis.

This assay has been developed and tested in a serially diluted human lung carcinoma cell line, A549, with an average 50-500 cells/per reaction to measure TA activity using CE and qPCR. In addition, hTERT mRNA expression has also been measured using specific primers and probes on a LightCycler™ in the range of 10-3500 cells/per reaction in cultured RPE-28 cells. (*J Mol Diagn* 2004; 6:157-65)

Comparison of TA using the RApidTrap and those performed manually were found reproducible and consistent. The reproducibility of these assays among investigators and laboratories has been conducted. A series of diluted A549 cells have been used as positive controls for testing clinical samples to detect any possible errors that could be caused by people and equipment, as well as to detect variability among different assays.

The correlation of TA with hTERT mRNA expression was demonstrated in these clinical samples. Further testing of the assay for multiple types of clinical samples will provide more important information for the usefulness and efficiency of the assay for clinical usage. Studies are planned to evaluate the assays' usefulness for clinical samples, using esophagus cytology cells that were collected from a previously NCI-supported screening trial of a Chinese population. Additional measurements of telomere length and methylation of the telomerase gene are also being considered to measure their impact on the **sensitivity** and **specificity** of telomerase assessment.

Since measurements of TA and the hTERT RNA expression vary among laboratories, it is sometimes difficult to compare results across different experiments. To overcome this problem, standard reagents have been developed as candidates for measuring TA.

Cross-validating Hypermethylation Assay

Abnormal DNA **methylation** patterns are characteristic of most cancers. For example, gene specific methylation changes in tumor cells and sputum are being evaluated as promising markers of lung cancer in several Network laboratories.

The Network performed an assay validation study comparing three platforms for measuring hypermethylation in tissue: standard **MSP** (analyzed at three laboratories), nested MSP, and **qPCR**. Each of the five laboratories involved in the study received thirty specimens of lung tissue: 6 frozen adenocarcinoma, 6 frozen squamous cell carcinoma, 12 frozen adjacent normal tissue, and 6 samples from cultured tissue cell lines. Laboratories were masked to the identity of the specimens. The loci at which hypermethylation was assessed were *p16, MGMT, RAR-β, DAPK,* and *RASSF-1*. No laboratory analyzed all of these loci; laboratories analyzed only those loci for which they had current experience and protocols.

The standard MSP and qPCR assays were very specific (91%-100%) at all loci examined, but their sensitivities were low (18%-41%). Nested MSP was more sensitive than standard MSP (41%-61%) but less specific (59%-83%). While more work needs to be done selecting a common platform for use in Network studies and in establishing a consistent protocol for assessing hypermethylation using the selected platform, the study results indicated that the analysis platforms had different performance characteristics but were all capable of detecting methylation in tissue.

New Quality Control for Mitochondrial DNA Sequencing

Mutations of mitochondrial DNA (mtDNA) are reported in many cancers, such as colorectal, breast, liver, prostate, pancreatic, and lung cancers, as well as in pre-neoplastic lesions.

A robotic-assisted whole mitochondrial genomic sequencing is an advanced approach to detect mutations in mitochondrial **genome**. Because DNA sequencing is a major component in this approach, the quality of the sequencing results is essential to the accuracy of the detected mutations. The rigorous quality control of the sequencing involves several parts, in which **PCR** quality is a key step since unqualified PCR product, which includes nonspecifically amplified products could cause DNA sequencing failure, or generate unreadable DNA sequencing data. Sometimes the PCR quality has been checked in gel-based electrophoreses, which is not an automatic and efficient assay.

To overcome this problem, scientists at NCI and NIST's DNA Technologies Group developed a new method of quality control for the whole mitochondrial DNA sequencing to detect mutations and validate the mitochondrial mutations for cancer detection. The PCR Quality and Quantity Control Procedure for DNA Sequencing (QCPS) has been integrated into a reliable, **high-throughput** assay to determine the role of sequence variation in the mitochondrial genome. (*J Mol Diagn* 2005; in press)

At least 99.8% of mitochondrial genome sequencing data has been successfully covered. Reproducible results have been obtained with 10-20 ng of DNA obtained from the clinical laboratory. More importantly, the QCPS method could also be applied to genomic DNA sequencing-based assays for detecting variations and chip-based microarray hybridization that uses PCR product as probes.

Cross Technology Platform Comparisons, Reagents, and Standard Reference Materials

Public health considerations demand that diagnostic assays and reagents be rigorously tested and standardized for consistency, reproducibility and accuracy. Molecular diagnostic assays are subject to a variety of inconsistencies arising from sample preparation, drifts in instrument calibration and precision, inter-operator variations, inter-laboratory variations, and the lack of quality assay reagents.

Standard Reference Materials (SRMs) provide a means to minimize these variabilities and lend appropriate precision and accuracy characteristics for assay development and standardization. Cross-technology platform comparisons are performed on SRMs to weed out measurement noise, select appropriate measurement standards for a particular technology, and evaluate the comparative performance of each technology.

Standard Reference Material

Standard Reference Material (SRM) is a certified reference material issued by the National Institute of Standards and Technology (NIST), United States Department of Commerce. SRM is a well-characterized material produced in quantity to improve measurement science. It is certified for specific chemical or physical properties, and is issued by NIST with a certificate that reports the results of the characterization and indicates the intended use of the material. SRM is prepared and used for three main purposes:

1. To help develop accurate methods of analysis;
2. To calibrate measurement systems used to facilitate exchange of goods, institute quality control, determine performance characteristics, or measure a property at the state-of-the-art limit; and
3. To assure the long-term adequacy and integrity of measurement quality assurance programs.

From NIST Technology Services, SRM Web Site (http://ts.nist.gov/ts/htdocs/230/232/ABOUT/definitions.htm)

Collaborating with HUPO on Technology, Plasma Proteomics, and Data Analysis

The Network has an active, ongoing collaboration with the Plasma Proteome Project Initiative of the Human Proteome Organization (HUPO) to evaluate multiple technology platforms, develop bioinformatic tools and standards for protein identification, and create a database of the plasma proteome. Results from these studies will pave the way for using serum or plasma for diagnostic assays and help identify proteins for molecular targeted detection and treatment.

The specific objectives of the initiative are:

- Comprehensive analysis of plasma and serum, including physiological: age, sex/menstrual cycle, exercise; pathological: selected diseases/cohorts; and pharmacological: common medication;
- Determination of the extent of variation across populations and within a population; and
- Identification of biological sources of variation within individuals over time, with validation of biomarkers.

See the web site at: http://www.hupo.org.

RNA Archiving

In collaboration with the Centers for Disease Control and Prevention, the Network is setting up standards for reference genes for **microarray** analysis. Twenty-one genes have been identified as potential "reference" genes for cervical samples.

CDC is also developing RNA archiving by developing strategies and optimizing the protocol to expand the utility of sense RNA amplification for: (1) partially degraded RNA, such as that from cervical exfoliated cells and other clinical sources; and (2) application to large scale qPCR studies as part of biomarker validation studies.

Development of Model Protocols for Protein Profiling

Based on the study described earlier in this section ("Validating Protein Expression for Early Detection of Prostate Cancer"), the Network developed a study design for a systematic evaluation of protein

profiling, in this case **SELDI-TOF**, for cancer diagnosis. This published model, which can be applied to any other profile-based **proteomics** platforms, has been extensively discussed and accepted. (See Semmes OJ, et al, Evaluation of serum protein profiling by surface-enhanced laser Desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin Chem* 2005; 51:102-12.) This activity represents the Network's goal to provide the scientific community with its experience and findings to help accelerate diagnostic research.

The protocol discusses a step-wise evaluation of protein profiling as follows:

1) *Stage 1:* Examine to see if the technique can be executed at different Network sites and replicated based on the discrimination with algorithms. This stage is divided into the following substages: *Stage 1a:* Standardize SELDI methodology and synchronize SELDI among seven participating institutions using a single source of pooled normal sera with verified presence of diagnostic peaks *Stage 1b:* Each site is blinded and sent 14 cases of normal and prostate cancer sera and asked to run the SELDI and submit data to the EDRN Data Management Coordinating Center. The data are analyzed to see if the algorithms correctly classify cases from controls. *Stage 1c:* Sera from each center are sent to all the other centers and SELDI runs are made. Data are then analyzed to see if the cases can be discriminated from controls.

2) *Stage 2:* Prove that the same results can be achieved with a number of prostate cancer cases and controls at the various Network sites.

3) *Stage 3:* This stage is concerned with measuring the sensitivity and specificity of SELDI in clinically well characterized cases drawn from prospectively collected retrospective samples. Discussions with the Prostate Cancer Prevention Trial leadership for samples are underway.

Making Resources Available for Validation Research

A number of technologies being evaluated and clinical specimens collected are available for collaborative research. In addition to reagents and standards, Network validation laboratories are evaluating other emerging technologies for discovery, development and validation of biomarkers. These include:

- Generating a breast cancer tissue microarray. At the UCLA Validation Laboratory, validation of the array by immunostaining with p53, ER, PR, and Her-2 was done, as was an examination of the expression of BS106, BU101, and mammaglobin from Duke University in collaboration with Abbott Laboratories Diagnostics Division. The expression of BS106, and mammaglobin were similar, while the expression of mammaglobin is higher than BS106 in breast tissue. The expression was higher in benign tissues as compared to malignant tissues. This was reverse for BU101 expression, where expression was higher in malignant cells than benign cells.
- Exploring methods for constructing an immunoassay "platform" to assess multiple cancer markers. At UCLA in collaboration with Genefluidics of Monterey Park, CA, an electrochemical

method detects the amount of antigen present in sera by first coating an electrode on a chip with specific antibody. Assays for prostasin, osteopontin, haptoglobin alpha, protease M, and CA125 are in progress.

- Generating reference specimen sets. Although the Network is not a biospecimen consortium, a great number of unique biospecimen resources have been created by investigators at the various centers and laboratories. For example, the specimen collection at the Boston Clinical Epidemiology Laboratory includes pre-operative serum and plasma from 225 women with ovarian cancer, 343 women with benign gynecologic diseases, and 431 normal women selected from the general population. The specimen bank also received 20,000 specimens from London consisting of serial serum collections in 1,000 women previously enrolled in prospective studies of screening for ovarian cancer. At the same time, the Northwestern Developmental Laboratory reports collection of pre-operative serum and plasma from over 5,600 women with ovarian cancer, 1,000 women with benign gynecologic diseases, and 7,500 normal women from the high-risk population who have been cancer-free for the past 6 years. In addition, over 140 breast cancer positive women, who previously received the ovarian Pap test, have subsequently had prophylactic bilateral salpingo-oophorectomies.

  The Network's specimen collection effort was discussed in the report, *Human Tissue Repositories: Best Practices for Biospecimen Resource for the Genomic and Proteomic Era,* published by RAND Corporation (2004) and in the *National Biospecimen Network Blueprint,* by Constella Group, Inc. (2004). The informatics infrastructure supporting sharing and dissemination of information on specimen availability through the EDRN Network Exchange System (ERNE) was highlighted, and many aspects of organization and procedures associated with specimen collection were deemed "Best Practices" in the field.

- Investigating stability of stored biological fluids. With high dimensional data, bias could also be inducted into experimental settings by the types of samples, storage, duration and handling of samples. A wide range of population and sampling conditions might affect the evaluation of a molecular biomarker. In relation to population diversity, consider a signature composed of A,B --- through F molecular biomarkers. Biomarker A may be subject to degradation after more than 1 year of storage at -80ºC, after more than 3 freeze thaw cycles, upon setting at room temperature for over 2 hours, and upon exposure to light for more than 3 hours. Biomarker B may only be important in Caucasians, may vary greatly with acute glucose levels at sampling, and may be stable in EDTA-plasma but not heparin plasma.

  It is critical to match samples of disease with samples of controls as closely as possible for all known parameters. Other factors that may add bias to a study include the stress of patients prior to sampling; binding of the biomarker to large proteins or to the glass of some but not all sampling containers; and acute and chronic dietary effects. The bias would occur when the effects of these parameters are different in the disease samples than in the control samples so that the experimental method actually separates disease from non-disease based on the bias rather than differences in the disease from non-disease. Bias also can extend to the methodology if samples are not analyzed randomly. Statistical approaches do not identify bias. Only a careful separation and evaluation of samples based on, for example, the time of storage at -80ºC can identify bias secondary to differences in storage of samples.

Leading the Knowledge Base

Bringing scientists together to share their findings is a hallmark of the Network. EDRN researchers collaborate across the scientific spectrum through a variety of professional vehicles. These include: NCI partnerships with other federal agencies; a comprehensive annual workshop and periodic working group meetings; biennial Gordon Research Conferences; organ-specific collaborative group meetings held in a "town hall" format; and infrastructure focused public-private partnerships.

NCI Partnerships with Federal Agencies

NCI has formed interagency agreements with several federal agencies to enhance the Network's capability. These include the Centers for Disease Control and Prevention; National Institutes of Standards and Technology; Pacific Northwest National Laboratories of the Department of Energy; and the National Aeronautics and Space Administration's Jet Propulsion Laboratory.

Centers for Disease Control and Prevention (CDC)

This interagency agreement capitalizes on CDC's resources for accruing high-risk cohorts and leveraging their experience with the state registries for infectious diseases and for cancer.

The effort features:

- Establishing infrastructure for specimen and data collection to detect and validate molecular markers of cervical cancer in a high-risk population;
- Characterizing molecular features of HPV 16 and the host immune response in histologically defined grades of incipient cervical neoplasias in a high-risk urban population and determining changes in cellular gene expression in the same samples; and
- Analyzing clinical, epidemiological and laboratory data to identify molecular fingerprints of neoplastic progression using appropriate epidemiological methods to perform initial validation studies of promising molecular markers identified in this and other Network studies, as potential signatures of early cervical neoplasia.

CDC is also setting up standards for reference genes for **microarray** analysis and has identified 21 genes as potential reference genes for cervical cancer samples. Another task has been developing RNA archiving. Strategies are in development for optimizing the protocol to expand the utility of sense RNA amplification for partially degraded RNA, such as that from cervical exfoliated cells and other clinical sources, and application to large scale **quantitative polymerase chain reaction (qPCR)** studies as part of biomarker validation studies.

NCI-CDC Interagency Agreement Supports Sampling Technique

The interagency agreement between NCI and CDC led to collection of epidemiology data for 1,666 subjects and colposcopy data for 1,659 subjects. The samples obtained to generate this data included mucosal wicks, serum, plasma, cervical cells and cervical total nucleic acids. Current studies focus on interpreting the results of microarray studies, and verifying that samples containing cells shed or exfoliated from cervical linings are adequate for biomarker discovery.

Extensive headway has been made in the evaluation of genes expressed in exfoliated cervical cells. This analysis has revealed that gene expression patterns in these samples parallels closely with that of intact cervical tissue, although the number of genes expressed in exfoliated cells was about 20% less. It concluded that exfoliated cells include only a subset of gene transcripts from cervical epithelium, but retain the same general functional profile.

Similar to comparison of cytology and histology, gene expression profiles from exfoliated cells can be expected to partially reflect that of the underlying tissue. The diversity of the gene profiles of the exfoliated cells support use of this sampling technique for continued biomarker discovery.

National Institutes of Standards and Technology (NIST)

This agreement funds and hosts a Biomarker Reference Laboratory which validates assays for early cancer detection. NIST investigators are conducting two biomarker assay validation projects.

The first project concerns mitochondrial DNA mutations as biomarkers for early cancer detection. Mitochondrial DNA mutations have been reported in a large variety of cancers. In collaboration with NCI and the Johns Hopkins University, NIST investigators are determining whether an oligonucleotide array (Mito Chip™) can be used to rapidly and accurately sequence mitochondrial DNA from samples obtained noninvasively. If successful, this assay may be useful for detecting early stage cancers.

The second project concerns telomerase as a biomarker for early cancer detection. Most postnatal somatic cells lack detectable telomerase, but telomerase is expressed in approximately 85% of cancers.

NIST investigators have modified and improved two methods to detect telomerase. The most commonly used method for the detection and quantification of telomerase enzyme activity is the telomerase repeat amplification protocol (TRAP) assay. NIST developed an automated capillary electrophoresis method to analyze the TRAP products that is more sensitive and reproducible than the commonly used slab-gel methods. NIST also developed a real-time reverse transcriptase polymerase chain reaction (PCR) assay to measure the mRNA for telomerase. Both of these assays were developed and tested using cell lines. NIST and NCI collaborators continue testing whether these assays can be used in human cancer specimens.

National Aeronautics and Space Administration's Jet Propulsion Laboratory (JPL)

Since April 2002, when NCI formed an interagency agreement with JPL (operated by the California Institute of Technology), JPL has lead informatics technology and architecture development for the Network. (See the Informatics section of this report for more details.) The agreement focuses on research and development of emerging informatics technologies to leverage and adapt existing work performed by JPL for planetary science.

The informatics research group at JPL has significant experience in developing scientific data systems for highly distributed scientific projects, working with agencies worldwide to connect such systems together. The software developed by JPL has formed the basis for the Network's knowledge system, enabling capture and exchange of scientific data sets between research institutes.

Food and Drug Administration (FDA)

EDRN has been collaborating with the NCI-FDA Clinical Proteomics Program investigators in technology platform and reagents standardization, collection and storage of biological materials, particularly blood and serum. In addition to formal collaborations, investigators from the NCI and FDA are collaborating with a number of EDRN investigators in research designs, sample collection, and cross-validation of **proteomic** assays. This ongoing collaboration culminated in organizing a Joint NCI-FDA Workshop on Research Strategies: Research Designs and Statistical Approaches to Biomarker Validation, held on July 28-29, 2004.

Trans-NIH Collaboration

EDRN program staff coordinated a one-year trans-NIH grant involving six institutes (NCI, National Institute on Aging, National Institute on Alcohol Abuse and Alcoholism, National Institute on Diabetes and Digestive and Kidney Diseases, National Institute on Environmental Health Sciences, and National Institute of Neurological Disorders and Stroke) for the Human Proteome Organization (HUPO) Plasma Proteome Project (PPP) pilot phase.

The aims of this global collaboration with more than 30 laboratories, led by the University of Michigan, are to: (1) evaluate the performance of a range of technology platforms for separation and identification of proteins; (2) compare serum versus plasma using different anticoagulants as part of pre-analytical variables arising in collection, handling, and storage of specimens; (3) assess the need for and complications of depletion of the most abundant plasma proteins; (4) create a major database with parsimonious and high-confidence peptide and protein identifiers; and (5) lay the groundwork for large-scale **biomarker** studies.

At the 3rd World Congress on Proteomics October 23-27, 2004, multiple presentations on the Plasma Proteome Project referred to it as a new basis for further developments. The core dataset has 3,020 distinct International Protein Index proteins detected and identified with two or more peptides in serum or plasma or both. A special issue of Proteomics early in 2005 is expected to contain 25 papers from the PPP, including extensive annotations and biological insights. All data and protocols will be made accessible by internet at the University of Michigan and at the European Bioinformatics Institute. Collaboration data will be available through the Network web site.

NCI Inter-Divisional Collaboration

NCI heavily invested in building infrastructures to facilitate discovery, development and evaluation of biomarkers across many programs. The Network maintains active dialogues and interactions with programs, like the Specialized Programs of Research Excellence (SPORE) and the Mouse Model of Human Cancers Consortium (MMHCC). These programs share common interests in developing biomarkers for cancer detection, diagnosis and prognosis. NCI program directors are invited to Network meetings, workshops and conferences, and staff meets periodically to discuss emerging leads on biomarker discovery deriving from their respective programs.

The Working Group on Lung Cancer Proteomics was formed in 2004 to support and coordinate protein-based biomarkers being developed through these programs toward their clinical validation. The group is comprised of Network and SPORE members. In addition, a number of Network liaisons have been appointed to SPORE and MMHCC programs and to the professional societies to update the community on the latest relevant research activities.

Annual Workshops

EDRN holds workshops annually to provide a diverse forum for discussion on timely, relevant scientific discoveries arising from the Network and elsewhere. Speakers represent a broad area of expertise in biomarker discovery and validation.

During the 3rd EDRN Annual Workshop, held June 14-16, 2004 in Bethesda, Maryland, over 300 researchers from the United States and other countries participated in discussions on topics such as: biology of early cancer; novel enabling technologies for detection of early cancer; molecular approaches to screening; analysis of high **throughput** biologic data for prediction and marker discovery; biology of hereditary cancers; impact on sporadic cancer detection; validation of biomarkers; and organ-specific translational research.

Periodically convened scientific workshops and working group meetings are adjuncts to the larger events. NCI's Division of Cancer Prevention Cancer Biomarkers Research Group holds specialized gatherings to facilitate scientific progress in discovery, development, and delivery of biomarkers in cancer detection and diagnosis.

Sampling of Workshops, 2002 to 2004

Linking Haplotypes and Genetic Variations with Cancer Assessment, Detection, Prevention and Treatment
December 4-5, 2003, Bethesda, MD

Analysis of Proteomic Spectral Data Including SELDI-MALDI-TOF-MS Applications
March 17-19, 2004, Fred Hutchinson, Cancer Center, Seattle, WA

NCI-Food and Drug Administration joint workshop on Study Designs and Statistical Approaches to Biomarkers Validation for Cancer Diagnosis and Detection
July 28-29, 2004, Gaithersburg, MD

Gordon Conferences

A formal partnership with the Gordon Research Conferences, a private conference organizer, results in a major meeting every 18 to 24 months to assess the topic, "New Frontiers in Cancer Detection and Diagnosis." The objective is to share progress made by EDRN investigators; exchange research ideas; develop collaborations; and identify significant research issues that could be addressed by the Network.

EDRN Gordon Conferences Chairs

2005	Bernard Levin, M.D., University of Texas, M. D. Anderson Cancer Center
2003	Sudhir Srivastava, Ph.D., Chief, Biomarkers Research Group, NCI Division of Cancer Prevention
2001	David Fishman, M.D., Northwestern University School of Medicine

This successful collaboration has helped EDRN disseminate research findings to investigators with relevant expertise who might otherwise not have the opportunity to meet. Examples of topics addressed include:

- Molecular basis for early detection;
- Animal and other pre-clinical models;
- Molecular detection of tumors;
- Biomarker validation methodology;
- Molecular detection of tumors;
- Epigenomics;
- High throughput technology;
- Proteomics trials and tribulations; and
- Novel adjuncts to early detection.

Collaborative Group Town Hall Meetings

In 2004, EDRN launched collaborative group meetings in a town hall format to bring diverse groups of scientists together. The format allows ample opportunity for attendees and invited guests to engage in discussions with Network investigators and share their experiences with accelerating biomarker discovery and evaluation.

Expected outcomes of the meetings are to:

1. Reach a consensus on a biomarker or panel of biomarkers that might be ready for further evaluation using Network resources; and
2. Review state-of-the-science in the selected area of research and identify gaps and barriers to be redressed through the Network mechanism.

Breast and Gynecologic Collaborative Group Meeting

Held in New York City September 27-28, 2004, this workshop primarily discussed the current status of early detection markers for women's cancers (breast, cervical, ovarian and endometrial) and prioritized them in the context of validation. The workshop

also explored possible collaborative efforts among SPORE and EDRN investigators. Topics included issues in early detection of breast cancer from the perspective of the researchers and patients; clinical assessment of breast cancer risk and use of random periareolar fine needle aspiration, ductal lavage, and nipple aspirate fluid for acquisition of breast epithelial cells; and complexity of ductal carcinoma in situ.

Roundtable discussions addressed biomarkers and methodologies useful for discovery and the populations of women to be targeted; types of samples that can be reasonably obtained and inventoried; and possible available resources through collaboration with other NCI funded programs, such as SPORE. Sessions also focused on endometrial and cervical cancers.

Colorectal and Other Gastrointestinal-related Collaborative Group Meeting

Focused on biomarkers for pancreatic cancer, this meeting was held September 12-13, 2004 in Norfolk, Virginia. It was attended by 30 scientists, including some SPORE investigators, and one patient advocate. Participants evaluated the status of development of genomic and proteomic biomarkers for early detection of pancreatic cancer and risk assessment. They determined whether these biomarkers were ready for Phase II/III validation trials and, if not, what additional data and resources would be required. Biomarkers were prioritized and study designs discussed in preparation for EDRN-supported Phase II/III validation studies.

The only protocol that appeared nearly ready for validation was a mutational load distribution analysis to monitor individuals known to be at high risk of developing pancreatic cancer. Promising biomarkers that may require validation in the near future include a panel of DNA **methylation** markers, a panel of antibodies, serum profiling, MUC1 protein, and mucin-related proteins. There was general agreement that a major barrier to biomarker validation for early detection of pancreatic cancer was a lack of well annotated specimens. The participants suggested that EDRN consider funding a repository that could be used to validate a number of the potential biomarkers for the early detection of pancreatic cancer.

Lung and Upper Aerodigestive Collaborative Group Meeting

Held September 19-20, 2004, this meeting was attended by about 25 scientists, including SPORE investigators. Participants evaluated the status of development of **genomic** and proteomic biomarkers for early detection of lung cancer and risk assessment. They determined whether these biomarkers were ready for Phase II/III validation trials and, if not, what additional data and resources would be required. Biomarkers were prioritized and study designs discussed in preparation for EDRN-supported Phase II/III validation studies.

Most of the discussed biomarkers were determined to be in the developmental stage, but several were recommended for further validation. A panel of DNA methylation markers (detected in sera) that appears to give 50-70% sensitivity is ready for analytical validation. A serum proteomic profiling protocol that appears to give 65% **sensitivity** and 93% **specificity** is also ready for analysis. A single protein marker, serum mesothelin-related protein, shows 84% sensitivity and near perfect specificity for early stage mesothelioma.

Genitourinary Collaborative Group Meeting

On September 14, 2004 this meeting was held in Houston, Texas and attended by the respective EDRN working group investigators, several members of the genitourinary and prostate SPOREs, and selected investigators with interests in genitourinary cancer biomarkers.

Potential prostate and bladder biomarkers were presented at this meeting. After multiple presentations on a wide range of biomarkers that may prove useful for the early detection of prostate cancer, the working group convened a discussion session in an attempt to prioritize markers in the validation process.

The consensus of the group was that it would be necessary to develop uniform criteria and standard materials (cases and controls) for systematic analysis (pre-validation study) of promising biomarkers in a blinded, reproducible fashion. This approach will permit comparisons on the performances of individual markers or panel of biomarkers studied in different laboratories. Other collaborative groups are gleaning from this insight as they are now instituting equivalent criteria and standards for prioritization.

A secondary, more imminent aim derived from this meeting was to develop markers from serum, plasma or urine, which when used in conjunction with the current prostate specific antigen (PSA) test, should reduce the high number of unnecessary biopsies performed due to high PSA readings.

The progress presented for biomarkers in bladder cancer showed great promise as a new urine-based test identifying **microsatellite instability** in 20 marker genes is now underway in Phase II validation studies. Furthermore, a number of other potential biomarkers pertinent to different stages of bladder cancer are also in the pipeline for validation in the near future.

Public-Private Partnerships

The Network has established numerous partnerships with biotechnology companies to facilitate cancer research. The Network and industry benefit from a mechanism that allows for formal partnering in the development of the necessary infrastructure.

Partnership strengthens these collaborations by creating a vehicle for industry to interact with and, where appropriate, co-fund the development of the required infrastructure with NCI. Industry benefits by having the opportunity to assist in the identification and defining of infrastructure objectives through the NCI planning process; leveraging their investments in infrastructure by contributing to a greater resource base; having access to new resources, expertise, databases and reagents that result from NCI-coordinated infrastructure investments; and having experienced management of their infrastructure investments by the NCI.

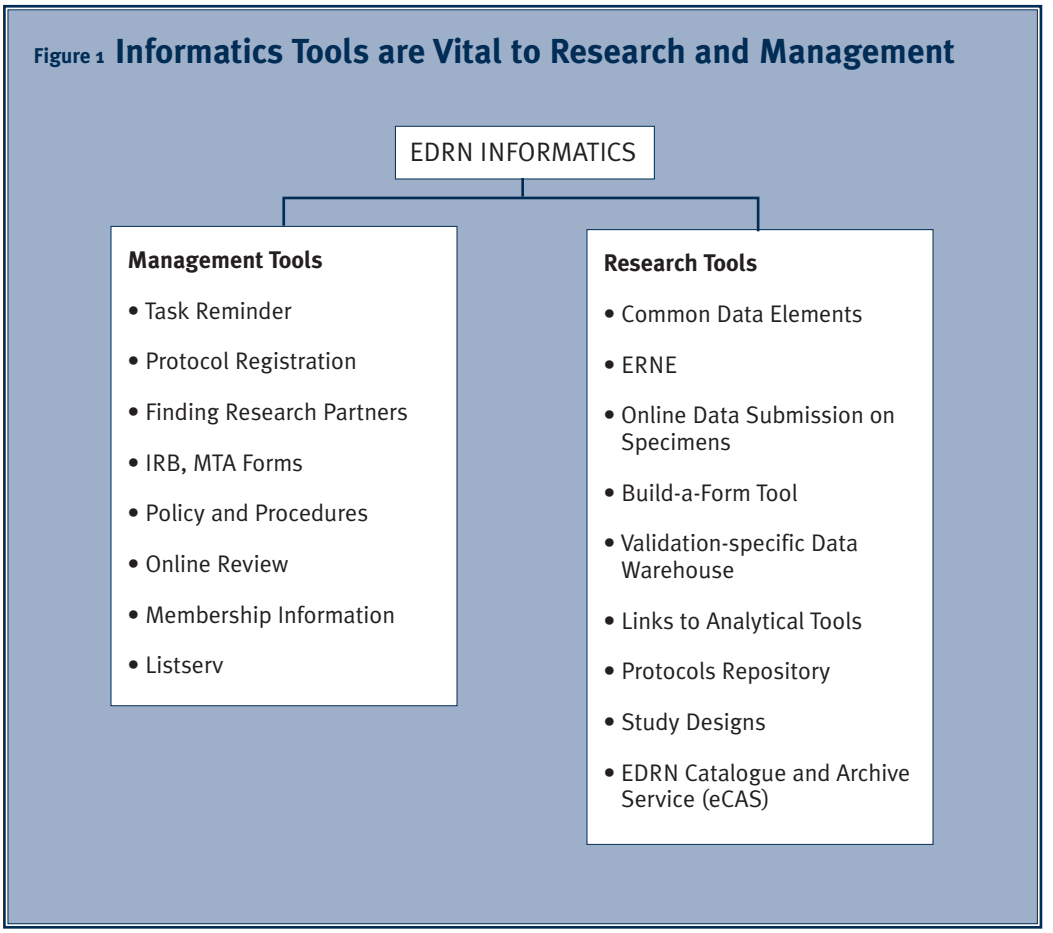
The NCI benefits by having access to industry expertise and technology, availability of an expanded investment base to build the critical infrastructure for cancer research, the expanded scientific scope industry brings to the scientific priority setting process, and the strengthened capabilities of our industrial partners which is essential for bringing the products of our research investments to the American public. A successful example of such collaboration is the Microsatellite Analysis (MSA) study for bladder cancer detection, in which CANGEN, a privately owned company that owns the license for MSA testing, and Commonwealth Biotechnology Inc., a publicly traded company, are collaborating with EDRN on this study.

Building Informatics for Effective Collaboration

Informatics–the collection, classification, storage, and analysis of recorded knowledge using computers–plays a key role in supporting the Network's scientific discovery, validation, and collaboration processes. By connecting research institutions together into a virtual knowledge system, informatics is making it possible for multiple organizations to access, share and analyze data regardless of its location or format.

Both the informatics infrastructure and data mining tools have advanced the Network's research efforts. (See Figure 1.) Portals with data located at principal investigator institutions across the country are being linked in a grid-like infrastructure developed and operated by the Network's Data Management and Coordinating Center, NCI, and NASA's Jet Propulsion Laboratory. The Network's investment in informatics has helped make it a leader in applying new technology for NCI.

Figure 1 Informatics Tools are Vital to Research and Management



Supporting Scientific Endeavors for Discovery, Development and Delivery

Defining the associated scientific processes that support the discovery and validation of cancer biomarkers is central to developing this virtual knowledge system. Multidisciplinary working groups representing basic scientists, clinicians, epidemiologists, biostatisticians and computer scientists, determine which discovery and validation methods will be used across the Network to perform the data collection and analysis.

With the completion of its virtual specimen bank in late 2002, the EDRN Resource Network Exchange (ERNE) accomplished initial efforts to build a national informatics infrastructure. ERNE enables scientists to search distributed, heterogeneous biospecimen repositories across participating institutions. Informatics experts and scientists worked together to understand how scientists would search and use biospecimen data. This approach was critical to developing a tool that would support the scientists in their discovery process.

The Network completed implementation of the Validation Study Information Management System (VSIMS) in 2004. VSIMS provides a national infrastructure for conducting collaborative validation studies providing the ability to capture and archive information in a multi-study manner. Generic tools provide forms management, issue tracking, specimen tracking and data access.

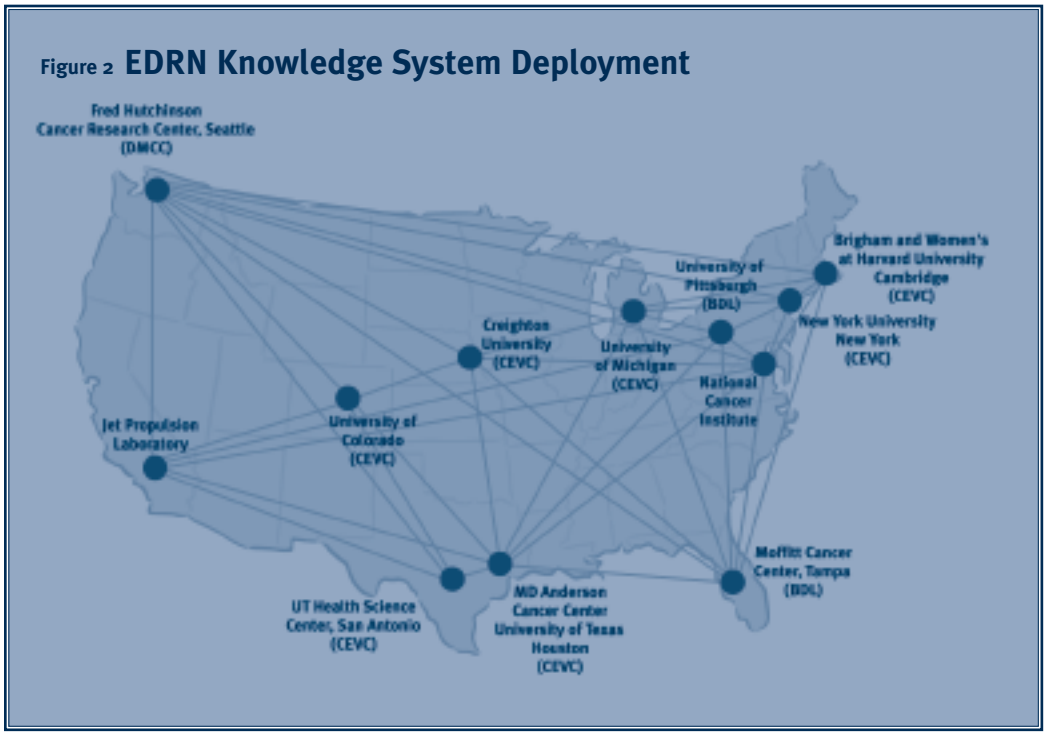
Working groups of scientists were also established to standardize the content collected in current studies using common data elements (CDE), a set of standard data terms and associated values (http://ncicb.nci.nih.gov/CDEBrowser). This provides a common language to be used in future studies and enables consistency across institutions collecting data. Any data shared is compliant with federal privacy and security regulations, including the Health Insurance Portability and Accountability Act. The law requires that certain identifiers be removed from research data to protect the confidentiality of patients. Careful attention also has been applied to ensure secure data transmission.

Knowledge System Deployed: Network Exchange

EDRN deployed the ERNE knowledge system to 10 institutions in early 2003, providing a common web-based client interface. (See Figure 2.) The system unifies search and retrieval of biospecimen data from all institutions regardless of where it is located, how it is stored, or the differences in the underlying data models. This enables a scientist, for example, to locate tissue specimens for breast cancer by searching data catalogs at participating institutions across the country.

As the knowledge system evolves, the governing cancer common data elements model and the use-cases derived in the working groups will be used to drive the relationships between the data sets enabling discovery through data mining. Scientists, for example, will be able to query an assay result from a validation study and then find the associated specimens that were collected as part of that assay.

Figure 2 EDRN Knowledge System Deployment



Validation Studies Information Management System (VSIMS)

The Validation Studies Information Management System (VSIMS) is a major component of the EDRN knowledge system. Critical to any knowledge system is its ability to capture data as part of the science data processing and analysis infrastructure. Within the Network, this occurs as part of the process to identify and validate cancer biomarkers.

EDRN designed a secure, web-based system that includes the main components needed for capturing and preserving the necessary metadata and data objects that integrate into the overall knowledge system architecture. These components include protocol management tools, communication tools, a data collection and processing system, and a specimen tracking system. All are based on having a robust data architecture as described earlier in this report.

Information maintained in the system is secure and stored separately for each multi-site study, allowing multiple protocols to be coordinated centrally through the same data management system.

VSIMS is data-driven, enabling adaptability to various validation study requirements. Each multi-site study or protocol has its own procedures defined in the system that allow components that need to be modified to meet a protocol's specifications. This allows for protocols to be implemented quickly and modified easily. A central data management and coordinating center handles several protocols simultaneously or consecutively.

Multiple levels of security make VSIMS secure. VSIMS uses 128-bit encryption (the de facto standard for data encryption over the public Internet for all data transfers) and requires all users to be authenticated prior to entering the unified portal. Each VSIMS user must complete an access application to obtain an account. The accounts are assigned protocol-specific access that identifies the protocol and associated permissions. Additional security measures include auditing, connection time-outs, and deactivation of inactive accounts.

The VSIMS web portal, shown in Figure 3, dynamically configures itself based on the user's permissions, and displays only those protocols and capabilities provided to the user. Each protocol has its own unique data-driven home page, which consists of a vertical list of buttons representing the major system components on the left-hand side of the screen. The study update section is a place to post important study information; the documents section has links to various study documents, minutes from conference calls, or other documents deemed important and allows users to have easy access to study communications. A message board provides additional communications space.

The Network developed a metadata-driven forms entry system that takes advantage of the Network's CDE metadata repository. Metadata is generally considered "data about data." In this case, it includes attributes such as data element name, wording of question, definition, data type, permissible value list, form instruction, etc. The metadata repository enables uniformity in data collection across multiple studies. Study-specific forms are created and the appropriate CDEs linked to the forms.

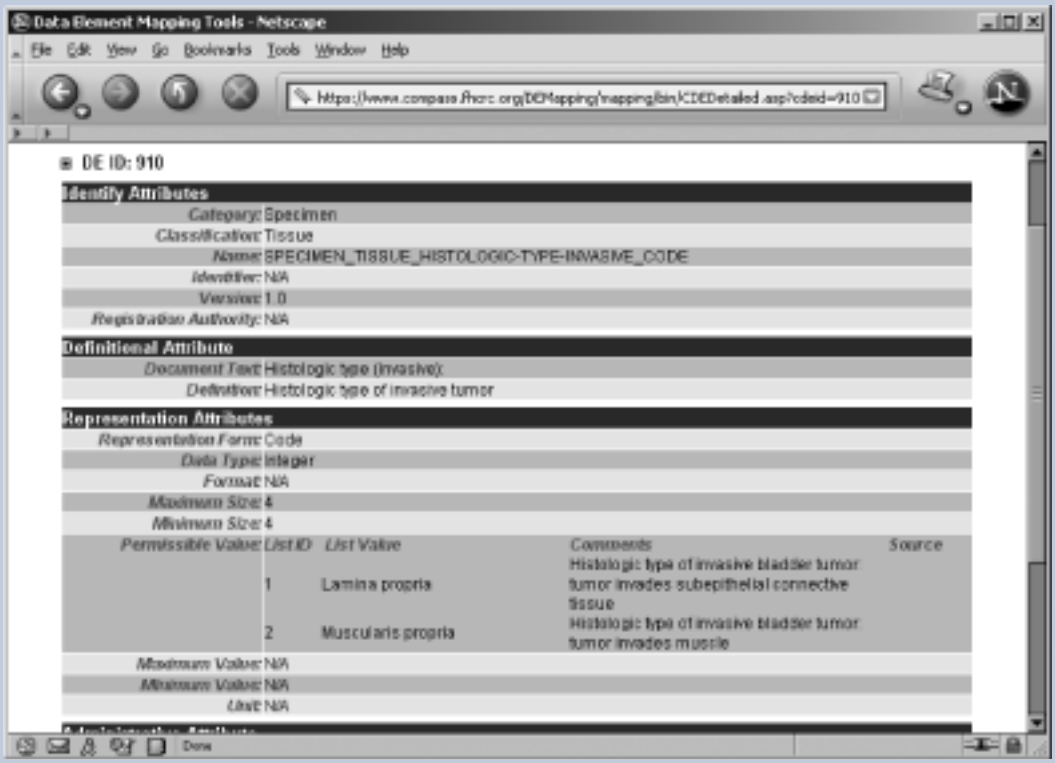
The system provides an automated data audit trail by automatically inserting the name of the data entry person, time and date when data are entered or modified, and prompts a required field to describe the reason for any data changes.

The system links to a specific VSIMS defined protocol such study-based criteria as: skip patterns; order of questions; choice of the display of the valid values of a question as check boxes or pull down lists; deletion, addition, and modification of questions and their valid values; range and logic checks; and optional double data entry verification. Data-driven forms enables flexibility and adaptability in collecting data across a variety of validation studies, and allows online data entry forms to be created quickly.

The Network is building a secure data transfer and processing infrastructure that permit data collected at remote locations to be ingested into VSIMS using a secure communications infrastructure. The infrastructure will catalog all data transfers using the CDEs. (See Figure 4.)

Figure 3 VSIMS Web Portal

Figure 4 Common Data Elements Tool



Communication tracking tools have been developed using an open source product called Scarab. Scarab, developed by Tigris.org, is a tool geared towards tracking software development-related issues, such as defect, enhancement, and requirement. The Network modified Scarab to track questions to and from the coordinating center along with data clarifications. For example, a specific question about a procedure in the manual of operations could be sent electronically through this system from a study site and routed to the appropriate person at the coordinating center. This question would then be answered and tracked appropriately, and likely rerouted to all other sites.

Tracking capabilities lets the coordinating center follow data clarifications within the database and document this information. For example, to clarify a data collection question, the coordinating center would send a question to a study site that would be tracked in the system and then be cross-referenced in the data clarification tracking module.

VSIMS includes a specimen tracking system to trace shipments and receipts of specimens between sites for validation studies. It integrates with the ERNE application by providing distributed interfaces to its online specimen catalog. Using bar coding technology, each specimen is labeled at the site that collects the specimen with a unique specimen identification number. Shipping and receiving sites then just scan the specimen containers and the system automatically enters the date, time, specimen ID, participant ID, shipping, and receiving site into the database, minimizing error due to manual data entry.

Infrastructure and Common Data Elements Link the System Architecture

The informatics infrastructure for the Network knowledge system is based on a distributed software framework developed at NASA's Jet Propulsion Laboratory (JPL), called the Object Oriented Data Technology (OODT) framework. OODT was selected in 2003 as NASA's runner-up for agency software of the year. The framework, used to support NASA's planetary and earth science missions, provides a set of software tools capable of connecting heterogeneous databases together and building new databases capable of archiving data.

Combining common data elements with the OODT software enables shared mechanisms for searching databases located at Network-funded institutions. The software is intelligent enough to handle mappings between different database implementations. Scientists and other researchers can make discoveries using different data sets produced by different organizations with different meanings, as if they are a single, large repository of knowledge. As such, the infrastructure can be configured to fit into different domains critical to scientific research, such as biomedicine and space science. As a result, disparate databases and systems can be connected over the Internet without requiring those systems to be re-implemented or modified.

Structuring and organizing the data, known as data architecture, is vital to effectively searching heterogeneous distributed data systems and enabling correlative science. Data architecture defines the common data elements and their relationships within the knowledge space. It allows interoperability between distributed institutions by providing a common language for communication. The data architecture for the Network's virtual knowledge system was developed to provide an over-arching model for describing critical cancer data objects.

Industry Standards

Industry standards support the definition of data architecture. EDRN uses the ISO/IEC 11179 standard in conjunction with the Dublin Core Standard to develop the minimal set of data elements that must be provided in any data architecture. ISO/IEC 11179 provides a standard definition for describing data elements to facilitate consistency in developing data dictionaries. The ISO/IEC standard recommends that a data element consist of attributes for four primary categories: identification, definitional, representational and administrative.

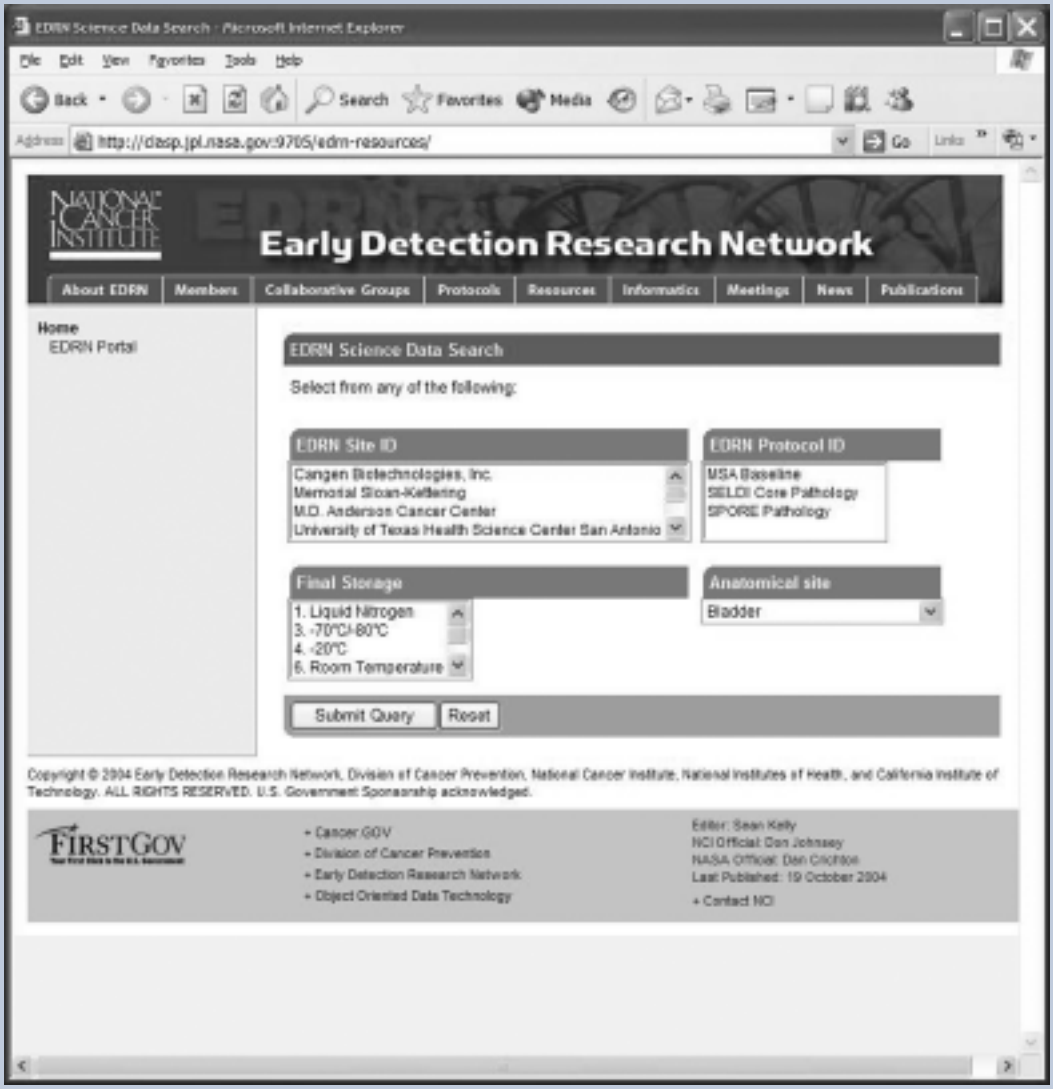
Once the common data elements are established, each participating institution can map its local data models to the knowledge system model to provide semantic consistency. Mapping tools were developed to allow the Network's informatics experts to capture the mapping of local site data models to the knowledge system model. Attributes of the data element, including permissible values, units, format, and data type, were captured and mapped to one another to provide the mapping at the informatics level. This enabled the informatics infrastructure software to run a translation function as part of the process of querying and retrieving data from the distributed institutions.

Capturing and Sharing Science Data: eCAS

The collaborative nature of EDRN represents a modern challenge for building major bioinformatics systems, namely that the capture and distribution of science and ancillary data produced within biomedical research is highly distributed across multiple cancer research centers.

The next challenge for the Network's Informatics Working Group focuses on another major component of the knowledge system, the EDRN Catalog and Archive Service (eCAS). It is anticipated that eCAS could become a national asset by providing the basic infrastructure to protect and share published data through a secure, long term, reliable and robust environment for managing scientific results. (See Figure 5.)

Figure 5 EDRN Catalog and Archive Service (eCAS)



This part of the system supports the capture and distribution of published science data acquired during a validation study. The group is working to ensure that the CDEs used throughout the infrastructure also fit into a common information model able to describe the science data produced during a study. This enables effective annotation of the science data sets and to search and retrieve the information through the Network's grid-based middleware.

The eCAS infrastructure will satisfy several Network informatics goals, including

- Acquisition of published science data from multiple EDRN validation studies;
- Software interfaces for sharing of data across biomedical applications;
- Secure transfer and distribution of data to the science community;
- A common information model for describing EDRN science data; and
- A web-based interface to search and download EDRN science products.

The Informatics Working Group is prototyping eCAS to validate these goals. The prototype demonstrates how data acquired locally at institutions can be archived using the EDRN common information model and then distributed to scientists nationally through a common science portal. Participating validation sites will be able to catalog their science results remotely via the public Internet. The common science portal will provide access to all published Network science data enabling distribution to a broad range of users using a well-defined security access policy.

The eCAS capability will greatly enhance both the amount and type of data available to the science community. It will also support the Network's expansion by providing a system capable of capturing and distributing the science data for all of its validation studies.

EDRN Portals

Informatics plays a key role in supporting the scientific discovery process by building the infrastructure and tools that connect the EDRN research institutions together into a virtual knowledge system. At the same time, the public web site disseminates the growing body of knowledge to all interested parties.

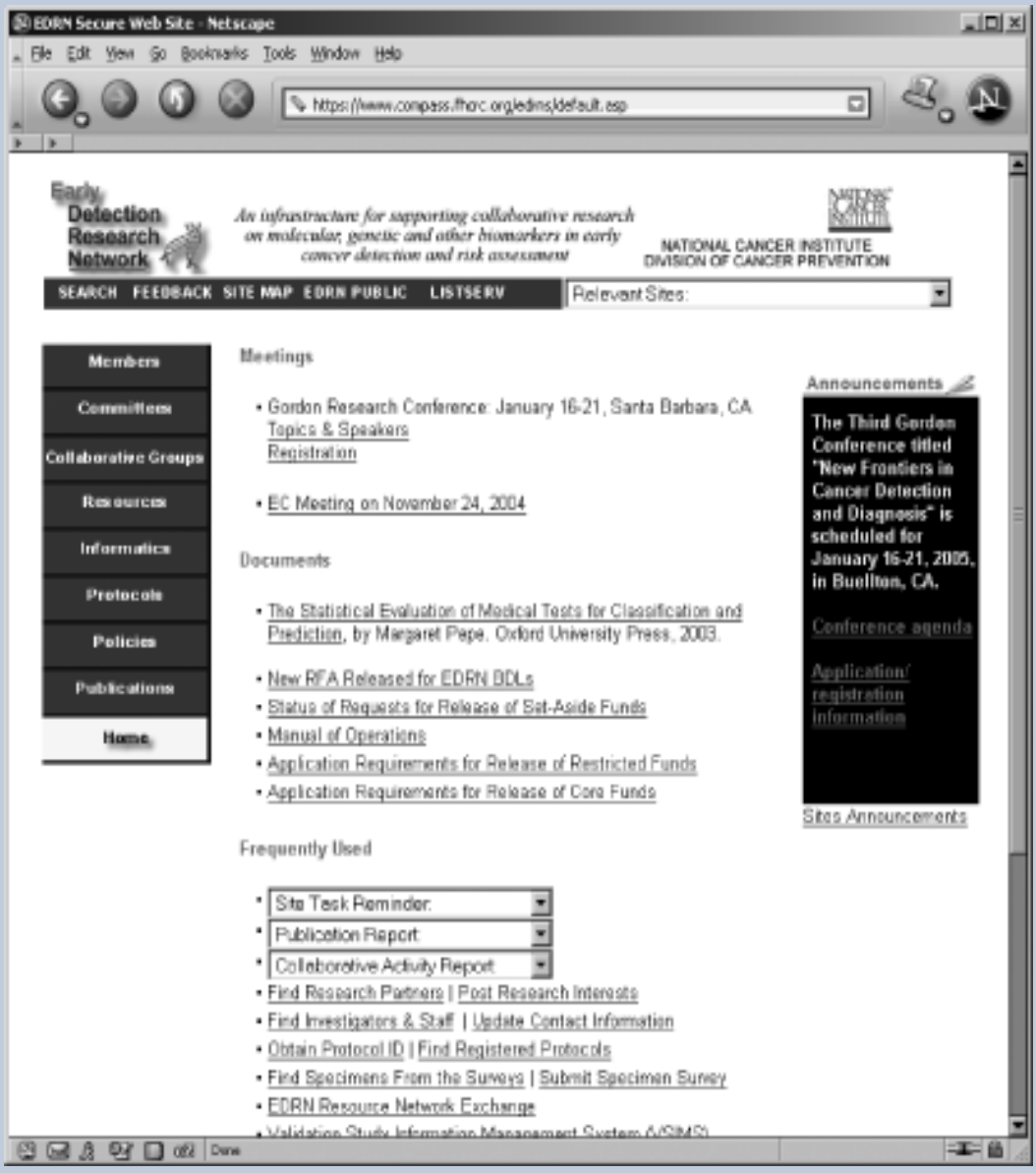
Secure Web Site

The Network's secure web site is designed to be the central hub for communication for the Network, making it possible for multiple organizations to access, share and analyze data regardless of its location or format. (See Figure 6.) The secure site is implemented and maintained by the Data Management and Coordinating Center (DMCC), and is accessible to approved applicants.

Specialized online tools include:

- A tool to register and track the Network's research protocols that allows investigators to enter, update, or search protocol title, abstract, IRB approval, and specimens collected;

Figure 6 EDRN Catalog and Archive Service (eCAS)



- A CDE tool to manage all of the Network's metadata, to allow users to search for CDEs, to allow DMCC users create and manage the CDEs and to assign them to study protocols that drive the online data collection tools;
- A survey building tool to let users quickly construct online surveys by entering, updating and reading their information; and
- A publication tool to permit investigators to enter all publications and print a report to send in with their annual report to NCI.

Other features of the site provide:

- Contact information for all Network sites that can be searched or updated;
- Utilities to subscribe, unsubscribe, and view all subscribers of an email list;
- Search capability for collaborators by research interests, keywords, and Network site;

- A review system for the Standing Review Group to oversee proposals for associate members and other reviews;
- Registration for steering committee and other scientific meetings;
- A reminder system for completing necessary tasks such as entering publication citations, protocols, research interests, or relevant surveys; and
- Administrative functions to ease maintenance, such as posting documents or announcements through a web interface that archives the documents for future reference and automatically posts them to the site.

The Network is an NCI leader in developing a research informatics infrastructure for scientists. This scalable infrastructure will enable it to expand data and tools, thus providing a long term platform for cancer research. As newer tools are developed to aid the discovery process, scientists will continue to mine and correlate data across multiple data sets and studies. This will include the introduction of data understanding software and algorithms that are capable of using the existing knowledge system infrastructure to construct knowledge bases of metadata using automatic feature detection. This metadata will augment existing metadata used to describe existing data products, and will enhance the informatics infrastructure for more sophisticated search and correlation capabilities.

User, Beware!

Bias: refers to any systematic error in the design, conduct, or analysis of a study that results in a mistaken estimate of experimental outcome, e.g., diagnostic test.

Chance: refers to erroneous conclusions drawn due type I and type II errors. Type I error refers to false positive conclusions that there is a difference between compared groups when no difference exists. Type II error refers to false negative conclusions that there is no difference when a difference does exist.

Public Web Site

The overall goal of the public portal implementation effort is to develop a single entry point for NCI that publicizes information about the Network. This includes synthesizing data currently collected by variety of sources about programmatic and Network resources. It will link to existing and future Network informatics tools and data, and it will integrate with the cancer.gov domain and be the principal interface for Network efforts.

The public portal (www.cancer.gov/edrn) enables wider access to the information available within the Network. This includes facts about investigators, on-going studies, meetings, funding opportunities, working groups, scientific discoveries and release of public data sets, publicly available informatics tools, and news.

The EDRN Data Management and Coordinating Center, NCI and NASA's Jet Propulsion Laboratory each play a critical role in developing and operating the informatics systems. Each of these partners, along with other EDRN institutions, needs the ability to share data, tools and information with both the Network and broader scientific communities.

Longer term, scientific data generated as part of a validation study needs to be made available to the science community. The public portal plays a pivotal role in providing access to all of the Network's science data as it is generated. Ultimately, the development, hosting and publication for the portal will be divided into different roles to facilitate the development, operations and information publishing for the portal by different groups.

Bringing Fresh Perspectives, New Ideas and Research Platforms

Novel approaches to enhance collaboration have been established as part of the Network's standard operations. These include reward systems for collaboration, an associate membership program that casts a wide net to interested participants, and new funding opportunities through program initiated projects.

Collaboration-led Competition

EDRN is funded through a Cooperative Agreement mechanism, which has to be renewed, and the investigators will have to compete for funding every 5 years. While the mechanism does impact the continuity of investigators, it fosters a healthy competition among investigators, many of whom are already collaborating on a range of projects. The continuing infusion of new ideas and investigators make the Network dynamic and vibrant. This is evident from the response to the most recent cycle of Request for Applications for EDRN: more than 140 applications were received for the various components of EDRN. A number of new laboratories or centers were awarded.

Reward System for Collaboration

The Network builds in incentives in its funding mechanism to promote collaboration within and outside the Network. Twenty percent of each individual grant is restricted for collaboration on Network-wide validation, intra-group (within the same organ site) or inter-group (across organ sites) studies. Proposals for such collaborations are reviewed and approved by the Network's Executive Committee and NCI program directors, and the Network's Data Management and Coordinating Center monitors their progress.

Associate Membership Programs

For investigators outside of the Network, opportunities for collaboration exist as Associate Members in research efforts related to **biomarker** development, validation and clinical testing. These non-Network investigators can apply for supplemental funds or for the use of shared resources through the Network, such as novel technologies, specimens, high-risk registries, and cohorts.

Three categories of associate membership exist:

1. Category A members propose basic or translational research consistent with EDRN priorities. The Network provides seed money for two-year pilot studies.

2. Category B members contribute to Network priorities by sharing available technologies, contributing specimens, providing high-risk registries, cohorts, and other resources. Funding can be applied for annually.

3. Category C members include scientists, clinicians, patient advocates, and ethicists, who participate in EDRN workshops and conferences, and Collaborative Group meetings, but do not receive EDRN funding or support for travel expenses.

There are currently more than 40 Associate Members to date. Once selected, they keep their status for the duration of the Network's existence. These members include distinguished scientists from both academic and industrial settings. The first Associate Member meeting was held June 13, 2004 to enhance collaborative interactions among Associate Members and showcase their accomplishments. Their presentations will be published as proceedings in *Disease Markers*.

Associate Members are eminent researchers involved in a spectrum of activities. For example, Associate Member David S. Alberts, M.D., recognized as a pioneer in the field of chemoprevention, was the recipient of the 2004 American Association for Cancer Research and Prevention Foundation Award for Excellence in Cancer Prevention Research. He is Regents Professor of Medicine, Pharmacology, Nutritional Science, and Public Health at the University of Arizona College of Medicine, and Director of the Cancer Prevention and Control Program at the Arizona Cancer Center.

Highlights of associate member activities include:

- Hepatocellular carcinoma (HCC) is a life threatening cancer with poor prognosis. Most patients are diagnosed at an advanced stage when curative surgical intervention (liver transplantation or tumor resection) cannot be performed. Dr. Jorge Marrero has launched a study to establish a specimen bank of patients with HCC for the development and validation of biomarkers for early diagnosis of HCC namely des-gamma carboxy prothrombin and alpha feto protein, in a limited patient study for their performance characteristics. These biomarkers have been approved for an elaborate prospective case control validation study by the Network's Executive Committee.

Category A Associate Members

Laura Beretta, Ph.D.	Fred Hutchinson Cancer Research Center
David Bowtell, Ph.D.	Peter MacCAllum Cancer Institute
Randall E. Brand, M.D.	Evanston Northwestern Healthcare Research Institute
Eleftherios P. Diamandis, Ph.D.	Mount Sinai Hospital
Hany Elsaleh, Ph.D.	University of California, Los Angeles
William Grady, Ph.D.	Vanderbilt University Medical Center
Brian Haab, Ph.D.	Van Andel Research Institute
Moncef Jendoubi, Ph.D.	Milagen, Inc.
Seema A. Khan, M.D.	Northwestern University, Feinberg School of Medicine
Bonnie L. King, Ph.D.	Yale University School of Medicine
Zvi Livneh, Ph.D.	Weizmann Institute of Science
Craig D. Logsdon, Ph.D.	University of Texas MD Anderson Cancer Center
Anna Lokshin, Ph.D.	University of Pittsburgh Cancer Institute
Richard F. Luduena, Ph.D.	University of Texas Health Science Center San Antonio
Sanford Markowitz, M.D., Ph.D.	Case Western Reserve University
Leonard Stephen Marks, M.D., M.A.	Urological Sciences Research Foundation
Carolyn Muller, M.D.	University of New Mexico
George L. Mutter, M.D.	Brigham and Women's Hospital
Susan L. Naylor, Ph.D.	University of Texas at San Antonio
Jong Park, Dr.P.H.	H. Lee Moffitt Cancer Center
Harvey Ira Pass, M.D.	Wayne State University
Edward Patz, M.D.	Duke University Medical Center
Diane M. Simeone, M.D.	University of Michigan
Karen Smith-McCune, M.D., Ph.D.	University of California, San Francisco
Ying Hsiu Su, Ph.D.	Drexel University College of Medicine
Samuil Umansky, Ph.D.	Diagen Corporation
Robert W. Veltri, Ph.D.	Johns Hopkins University

Category B Associate Members

Bao-Ling Adam, Ph.D.	Medical College of Georgia
Gerald L. Andrioli, M.D.	Washington University School of Medicine
Joseph W. Basler, M.D., Ph.D.	University of Texas Health Science Center San Antonio
Alex Befeler, M.D.	Saint Louis University
Mai N. Brooks, Ph.D.	University of California, Los Angeles
Allen D. Cooper, M.D.	Stanford University
Guido Dalbagni, M.D.	Memorial Sloan-Kettering
Colin P. Dinney, M.D.	University of Texas M.D. Anderson Cancer Center

Kathleen Groover	National Cancer Institute at Frederick
Robert Harris	Commonwealth Biotechnologies, Inc.
Robert Kalik	Cangen Biotechnologies, Inc.
Adam Kibel, M.D.	Washington University
Laurence Klotz, M.D.	University of Toronto
Cheryl T. Lee, M.D.	University of Michigan
Seth P. Lerner, M.D.	Baylor College of Medicine
Daniel W. Lin, M.D.	University of Washington
Jorge Marrero, M.D.	University of Michigan
Edward Messing, M.D.	University of Rochester Medical Center
Joseph C. Presti, M.D.	Stanford University
Rajender Reddy, M.D.	University of Pennsylvania
Lewis Roberts, M.D., Ph.D.	Mayo Clinic
Mark P. Schoenberg, M.D.	Johns Hopkins University
Myron Schwartz, M.D.	Mount Sinai Hospital
Caroline Shepherd	BRT Laboratories, Inc.
Neal Shore, M.D.	CURC/Carolina Urologic Research Center
Shiv Srivastava, Ph.D.	Walter Reed Army Medical Center
Gary D. Steinberg, M.D.	University of Chicago Urology Center
Donald Urban, M.D.	University of Alabama at Birmingham

Category C Associate Members

David S. Alberts, M.D.	Arizona Cancer Center
David Beach, Ph.D.	Genetica, Inc.
Joseph Bigley	OncoMethylome Sciences, Inc.
Jose Costa, M.D.	Yale University School of Medicine
Edward Highsmith, Ph.D.	University of Maryland
Nancy B. Kiviat, M.D.	University of Washington
Subbi P. Mathur, Ph.D.	Medical University of South Carolina
David F. Ransohoff, M.D.	University of North Carolina at Chapel Hill
Brian J. Reid, M.D., Ph.D.	Fred Hutchinson Cancer Research Center
Roger S. Rittmaster, M.D.	GlaxoSmithKline
Steven Smith, Ph.D.	Beckman Research Institute of the City of Hope
Victoria L. Stevens, Ph.D.	Emory University
Steinunn Thorlacius, Ph.D.	Iceland Genomics Corporation
Bruce Trock, Ph.D.	Johns Hopkins Medical Institutions
Yingming Zhao, Ph.D.	UT Southwestern Medical Center

- Mesothelioma is an aggressive tumor typically associated with individuals exposed to asbestos. Dr. Harvey Pass is coordinating a study in conjunction with the Network's Data Management and Coordinating Center (DMCC) to plan Phase II validation studies on serum mesothelin-related protein (SMRP), a biomarker exhibiting high specificity and sensitivity for malignant mesothelioma.
- Lung Cancer - Early lung carcinoma detection strategies involving imaging have yet to demonstrate a reduction in mortality. Identification of serum biomarkers that could complement radiologic studies and facilitate earlier diagnosis of lung carcinoma would be of significant benefit to patients. In an excellent pilot study, Dr. Edward Patz and his group evaluated two overexpressed proteins in lung carcinoma, serum amyloid A and macrophage migration inhibitory factor, as potential diagnostic serum biomarkers for this malignancy.
- Colorectal Cancer - Human urine has been shown to possess slight amounts of DNA. Dr. Ying Hsiu Su's group at Drexel University College of Medicine compared the mutated *K-ras* proto-oncogene sequences present in DNA isolated from tumor, blood, and urine from an individual with colorectal carcinoma. Results showed that the low molecular weight class of urine DNA is derived from the circulation. This finding has broader implications for early detection and risk assessment of other cancers.
- Ovarian Cancer - Recent evidence suggests that many members of the human kallikrein (*KLK*) gene family are differentially regulated in ovarian cancer and have potential as diagnostic and/or prognostic markers. Using serial analysis of gene expression and expressed sequence tag databases of NCI's Cancer Genome Anatomy Project, Dr. Eleftherios Diamandis's group from Mount Sinai Hospital in Toronto, Ontario, Canada performed computer analyses of the expression pattern of the 15 human *KLK* genes in normal and cancerous ovarian tissues and cell lines. They found that seven *KLK* genes are up-regulated in ovarian cancer.

New Funding Opportunities through Program-Initiated Projects

NCI's Division of Cancer Prevention, Biomarkers Research Group, has initiated a number of research projects funded through the Small Business Innovation Research (SBIR) program and the Small Business Technology Transfer (STTR) program mechanism on topics that support the mission of the Network. Described below, these include virtual microscopy; antibody arrays; bioinformatics research; metabolomics; circulating cells; and proteomics for infectious agents associated cancers.

Virtual Microscopy for the Early Detection of Cancer

NCI hopes to facilitate development and application of novel digital microscopic imaging modalities that are applicable to early detection and screening.

Virtual microscopy is a novel means of detecting biological abnormalities, but it still needs to be refined and improved. Potential research and clinical applications have yet to be explored. Virtual microscopy will offer new advantages in the visualization of biological tissues, enabling accurate determination of pre-malignant and malignant tumors.

Virtual microscopy provides high-resolution images in three dimensions and facilitates the visualization of tissue biopsy specimens that can be manipulated by the computer. Using this technique in combination with differential staining characteristics, it is possible to visualize only the specific anatomical details of interest without having to look at complex tissue structures.

As a new form of microscopy, virtual microscopic imaging requires the development of new methods of tissue processing and visualization. Potentially this technology can be applied to clinical specimens for the detection of biomarkers and the determination of their spatial localization in the biological samples for accurate differentiation of pre-neoplastic lesions from malignant tumors.

Metabolomics for Early Cancer Detection

NCI is seeking is to stimulate research to determine whether metabolomics can be used to distinguish patients with cancer from healthy individuals; to develop and optimize metabolomics methods for use with body fluids from cancer patients; and to assess the potential usefulness of these technologies for early cancer detection and risk assessment.

Metabolomics is the study of small molecules, or metabolites in cells, tissues, and body fluids. Representative small molecules include compounds like glucose, cholesterol, ATP, and lipid signaling molecules. The identities, concentrations, and fluxes of these molecules are the final products of interactions between gene expression, protein expression, and the cellular environment. When compared to inherent complexity of DNA, RNA, and proteins, the limited numbers of small molecules make them suitable for analysis by high **throughput** methods.

Metabolomics researchers concentrate on biofluids, including blood, urine, and cerebrospinal fluid, and attempt to identify and quantify all the small molecules within a sample to find new markers for disease or drug toxicity, or indicators of nutritional status. Researchers are using high throughput metabolomic technologies to identify early signatures of diseases, especially central nervous system disorders. While numerous cancers have been shown to alter metabolite levels in body fluids, there is little research on using this technology for early cancer detection or risk assessment.

Antibody Arrays for Cancer Detection

No single marker or combination of a limited number of biomarkers has a sufficient **sensitivity** and **specificity** to diagnose asymptomatic cancer or early stage cancer at present. However, recent developments in gene and **proteomic** profiling of precancerous and cancerous lesions suggest that a panel of markers may be used to distinguish cancer and non-cancer with high sensitivity and specificity.

The purpose of this initiative is to develop an antibody array in collaboration with the Network. This methodology is likely to become the next step in conversion of **mass spectrometry** proteomic patterns to well-defined diagnostic targets.

It is anticipated that the collaboration will reveal sets of biomarkers that permit the development, production and dissemination of antibody **microarray** technologies for the scientific community engaged in research on early cancer detection and risk assessment.

Specific objectives are:

- Prepare and purify biomarker-specific antibodies in the form of recombinant antibodies or monoclonal antibodies (mAb);
- Develop and/or improve methodologies for quantitative measurements of the bound antigens on antibody microarrays; and
- Perform initial validation studies in collaboration with EDRN using the antibody microarrays.

Antibody microarrays will provide a fast, reliable, high-throughput, sensitive, and quantitative detection tool of multiple differentially expressed antigens (annotated proteins and post-translationally modified proteins) from a limited amount of sample (e.g. 20 µl of serum) obtained through a minimally invasive method.

Early Detection Research Network Bioinformatics Research Program

This initiative supports development of software for analysis and evaluation of cellular signatures for earlier cancer detection in prevention research. The objectives are:

- Develop analytical methods for proteomic and **genomic** data analysis, pre-analytical data processing algorithms for time-of-flight (TOF) mass spectrometric (MS) data and genomic expression data protein biomarker identification via innovative data mining and pattern recognition methods, such as the classification tree, boosting, support vector machines, artificial neural networks, and cluster analysis.
- Develop algorithms to improve diagnostics by applying: algorithms for longitudinal or cross-sectional data to classify patients according to the relevant disease states using **surface-enhanced laser/matrix-assisted laser desorption/ ionization (SELDI/MALDI)** profiling data; gene expression analysis; algorithms to patient data for early detection of cancer; validating the clinical utility of algorithms to differentiate cancer types; and validation of the algorithms through the analysis of simulated data and comparison with well established results.
- Develop bio-computational approaches to automated calibration, normalization and synchronization of SELDI/MALDI instruments. Develop methods to assess and quantify the reproducibility of high-throughput proteomic and genomic technologies.

Circulating Cells in Cancer Detection

This initiative will develop novel technologies for capturing, enriching, and preserving exfoliated abnormal cells and macromolecules in body fluids or effusions, and develop methods for harvesting the cells for biomarker studies.

This exfoliation includes cellular materials and subcellular materials, such as DNA and proteins. In body fluids, such as sputum, the number of exfoliated tumor cells is often small compared to the number of non-neoplastic cells. The detection of exfoliated abnormal cells by routine cytopathology is often limited because few atypical cells may be present in the specimen.

There may be difficulty in separating dysplastic cells from non-specific reactive changes and degenerating cells or variation in diagnostic criteria. Exfoliated cells are also frequently contaminated with normal cells, bacteria, and other cellular debris, which makes molecular analysis difficult without physical separation of the neoplastic cells.

Development of enrichment methods is critical for the routine detection of small numbers of exfoliated cells and small amounts of subcellular materials in biological fluids for molecular analysis. Enrichment will allow exfoliated cells and subcellular molecules, for example from urine, to be used for genomic, proteomic, and **epigenomic** analyses that may lead to improvements in the detection of bladder cancer through measurements of alterations in expressed genes, peptide profiles, and epigenetic markers.

Proteomic Portraits of Infectious Agents-Associated Cancer

Through this initiative, NCI will encourage research to identify protein markers for risk assessment and early detection in individuals exposed to infectious agents that have been linked to cancer.

Areas needing support to develop proteomic signatures for infectious agent-associated cancers include:

1) Establishment of proteomic profiles of normal, precancerous, and cancerous lesions following infection and of body fluids from infected individuals; and

2) Evaluation of these proteomic profiles for use in early detection, risk assessment, and prevention of specific cancers.

The proteomic profiles can be analyzed to determine whether a single protein biomarker, panel of protein biomarkers, or proteomic patterns can be used to ascertain which infected individuals are at risk of developing cancer. It can also be used to determine the transition from chronic infection to the initiation of cancer. These studies may provide future targets for cancer prevention and therapeutic vaccine development.

Future Directions

Cancer is a complex disease. Its understanding will require a systems biology approach and leverages from diverse fields such as biology, chemistry, engineering, informatics and computational modeling. Clinicians and health professionals need to be involved at the outset so that developments in proteomics and genomics are tailored to address specific clinical questions.

For proteomics and genomics to be successful in public health, approaches should be compared for their benefits over other medically accepted detection and screening modalities. It is also desirable to integrate proteomics with imaging to enhance the special localization and temporal view of the disease. The development of both proteomics and imaging should proceed in tandem for molecular-based assays not requiring biopsy or tissue samples.

Implementation of the Early Detection Research Network has met the need for a multi-disciplinary effort by the National Cancer Institute that emphasizes collaboration, discovery, development, and delivery of biomarkers for detection, diagnosis, and risk assessment.

Genome-wide global profiling coupled with proteomics is expected to lead to a molecular taxonomy of cancer that goes beyond organ and tissue types. These technologies will supercede or enhance the classifications based on histopathology or based on the patterns of expression of genes of unknown biological significance uncovered by global transcriptomic profiling.

Yet, strategies must be put into the most appropriate manner for integrating clinical and genomic data with proteomic data for tailored therapy in the future. In pursuit of this goal, EDRN will:

1. Accelerate discovery and validation of technologies and newly discovered markers.
2. Expand the Network focus to include rare tumors such as nasopharyngeal carcinoma and mesothelioma.
3. Expand public-private partnership for validation studies by providing access to EDRN resources.
4. Develop comprehensive protein fingerprint “proteomics” for pre-malignant and pre-invasive cancer in collaboration with other NCI-supported programs.

5. Develop a cancer biomarker knowledge base.

6. Develop bioinformatics tools to integrate data in partnership with the Cancer Bioinformatics Grid (CaBIG).

7. Promote study on signatures of cancer cells and the microenvironment.

8. Leverage experience of and work with new NCI programs in nanotechnology, proteomics, and specimen biorepository to accelerate biomarker discovery and validation.

The validation of biomarkers is imperative for molecular screening, detection and diagnosis. Such validated biomarkers may, in turn, increase the efficiency of drug development by showing the need for treatment despite the absence of disease or, distinguishing those patients likely to benefit from treatment from those not likely to respond to drug therapy. Biomarkers and diagnostics, therefore, go hand in hand in alleviating suffering and death due to cancer. In the interest of public heath, it is important to ensure that biomarkers provide accurate, convincing evidence as to the presence or absence of the disease and conform to the regulatory requirements.

The Network has established a robust validation criteria based on the NASA approach to readiness levels, and a computer-assisted framework that simultaneously helps multiple investigators through each step of the FDA regulatory compliance process. Using these measures early on will ensure that biomarkers validated through EDRN are brought to practical use in cancer detection and treatment in a timely fashion. The potential benefits of biomarkers in cancer are great, and their application requires careful, judicious consideration.

Appendix

I Key Publications by Investigators in the Early Detection Research Network

Adam B-L, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, et al. Serum protein fingerprinting coupled with a pattern matching algorithm that distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* 2002; 62:3609-14.

Alonzo TA, Pepe MS. Distribution-free ROC analysis using binary regression techniques. *Biostatistics* 2002; 3: 421-432.

Andrews GA, Xi S, Pomerantz RG, Lin CJ, Gooding WE, Wentzel AL, et al. Mutation of p53 in head and neck squamous cell carcinoma correlates with Bcl-2 expression and increased susceptibility to cisplatin-induced apoptosis. *Head Neck* 2004; 26:870-7.

Becker S, Cazares LH, Watson P, Lynch H, Semmes OJ, Drake RR, Laronga C. Surfaced-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) differentiation of serum protein profiles of BRCA-1 and sporadic breast cancer. *Ann Surg Oncol* 2004; 11:907-14.

Bepler G, Goodridge Carney D, Djulbegovic B, Clark RA Tockman MA. Systematic review and lessons learned from early lung cancer detection trials using low-dose computed tomography of the chest. *Cancer Control* 2003; 10:306-14.

Binkley CE, Zhang L, Greenson JK, Giordano TJ, Kuick R, Misek D, et al. The molecular basis of pancreatic fibrosis: common stromal gene expression in chronic pancreatitis and pancreatic adenocarcinoma. *Pancreas* 2004; 29:254-63.

Block TM, Mehta AS, Fimmel CJ, Jordan R. Molecular viral oncology of hepatocellular carcinoma. *Oncogene* 2003; 22:5093-107.

Bouwman K, Qiu J, Schotanus M, Mangold LA, Vogt R, Erlandson E, et al. Microarrays of tumor cell derived proteins uncover a distinct pattern of prostate cancer serum immunoreactivity. *Proteomics* 2003; 3:2200-7.

Bresalier R, Byrd J, Tessler D, Lebel J, Koonen J, Hawke D, et al. A circulating ligand for galectin-3 is a haptoglobin-related glycoprotein elevated in individuals with colon cancer. *Gastroenterology* 2004; 127:741-8.

Cazeres LH, Adam BL, Ward MD, Nasim S, Schellhammer PF, Semmes OJ, Wright GL. Normal, benign, pre-neoplastic and malignant prostate cells have distinct protein expression patterns resolved by SELDI mass spectrometry. *Clin Cancer Res* 2002; 8:2541-52.

Comunale MA, Mattu TS, Lowman M, Evans AA, London WT, Semmes OJ, et al. Comparative proteomic analysis of de-N-glycosylated serum from hepatitis B carriers reveals polypeptides that correlate with disease status. *Proteomics* 2004; 4:826-38.

Crosio C, Maria Fimia G, Loury R, Kimura M, Okano Y, Zhou H, et al. Mitotic phosphorylation of histone H3: spatio-temporal regulation by mammalian aurora kinases. *Mol Cellular Biol* 2002; 22:874-885.

Ding YS, Vernon SD, Unger ER, Karen KL. A simplified approach for analysis of SELDI-TOF mass spectrometry data. *Appl Genomics Proteomics* 2003; 2:71-7.

Dinney CP, McConkey DJ, Millikan RE, Wu X, Bar-Eli M, Adam L, et al. Focus on bladder cancer. *Cancer Cell* 2004; 6:111-6.

Feng Z, Hu W, Chen JX, Pao A, Li H, Rom W, et al. Preferential DNA damage and poor repair determine ras gene mutational hotspot in human cancer. *J Natl Cancer Inst* 2002; 94:1527-36.

Feng Z, Hu W, Rom WN, Costa M, Tang MS. Chromium(VI) exposure enhances polycyclic aromatic hydrocarbon-DNA binding at the p53 gene in human lung cells. *Carcinogenesis* 2003; 24:771-8.

Gasperino J, Rom WN. Gender and lung cancer. *Clin Lung Cancer* 2004; 5:353-9.

Gazdar AF, Shigematsu H, Herz J, Minna JD. Mutations and addiction to EGFR - the Achilles "heal" of lung cancers? *Trends Mol Med* 2004; 10:481-86.

Gelmann EP, Semmes OJ. Expression of genes and proteins specific for prostate cancer. *J Urol* 2004; 172:S23-6.

Gharib TG, Chen G, Huang CC, Misek DE, Iannettoni MD, Hanash SM, et al. Genomic and proteomic analyses of vascular endothelial growth factor and insulin-like growth factor-binding protein 3 in lung adenocarcinomas. *Clin Lung Cancer* 2004; 5:307-12.

Goldenberg D, Benoit NE, Begum S, Westra WH, Cohen Y, Koch WM, et al. Epstein-Barr virus in head and neck cancer assessed by quantitative polymerase chain reaction. *Laryngoscope.* 2004; 114:1027-31.

Greenberg AK, Yee H, Rom WN. Preneoplastic lesions of the lung. *Respir Res* 2002; 3:20-46.

Greenberg AK, Basu S, Hu J, Yie TA, Tchou-Wong KM, Rom WN, and Lee TC. Selective p38 activation in human non-small cell lung cancer. *Am J Respir Cell Mol Biol* 2002; 26:558-64.

Gretzer MB, Chan DW, Van Rootselaar CL, Rosenzweig JH, Dalrymple S, Mangold LA, et al. Proteomic analysis of Dunning prostate cancer cells with variable metastatic potential using SELDI-TOF. *Prostate* 2004; 60:325-31.

Grizzle WE, Semmes OJ, Basler J, Izbicka E, Feng Z, Kagan J, et al. The early detection research network surface-enhanced laser desorption and ionization prostate cancer detection study: a study in biomarker validation in genitourinary oncology. *Urol Oncol* 2004; 22:337-43.

Habis AH, Vernon SD, Lee DR, Verma M, Unger ER. Molecular quality of exfoliated cervical cells: implications for molecular epidemiology and biomarker discovery. *Cancer Epidemiol Biomark Prev* 2004; 13:492-6.

Hanash S. Integrated global profiling of cancer. *Nat Rev Cancer* 2004; 4:638-44.

Harden SV, Sanderson H, Goodman SN, Partin AA, Walsh PC, Epstein JI, Sidransky D. Quantitative GSTP1 methylation and the detection of prostate adenocarcinoma in sextant biopsies. *J Natl Cancer Inst* 2003; 95:1634-7.

Hirsch FR, Varella-Garcia M, Bunn PA Jr, Di Maria MV, Veve R., Bremmes RM, et al. Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J Clin Oncol* 2003; 21:3798-807.

Hlavaty JJ, Partin AW, Shue MJ, Mangold LA, Derby J, Javier T, et al. Identification and preliminary clinical evaluation of a 50.8-kDa serum marker for prostate cancer. *Urology* 2003; 61:1261-5.

Hong S-H, Misek DE, Wang H, Puravs E, Giordano TJ, Greenson JK, et al. An autoantibody-mediated immune response to calreticulin isoforms in pancreatic cancer. *Cancer Res* 2004; 64:5504-10.

Hoque MO, Begum S, Topaloglu O, Jeronimo C, Mambo E, Westra WH, et al. Quantitative detection of promoter hypermethylation of multiple genes in the tumor, urine, and serum DNA of patients with renal cancer. *Cancer Res* 2004; 64:5511-7.

Imafuku Y, Omenn GS, Hanash S. Proteomics approaches to identify tumor antigen directed autoantibodies as cancer biomarkers. *Dis Markers* 2004; 20:149-53.

Jeronimo C, Henrique R, Hoque MO, Ribeiro FR, Oliveira J, Fonseca D, et al. Quantitative RARbeta2 hypermethylation: a promising prostate cancer marker. *Clin Cancer Res* 2004; 10:4010-4.

Katayama H, Sasai K, Kawai H, Yuan ZM, Bondaruk J, Suzuki F, et al. Phosphorylation by aurora kinase A induces Mdm2-mediated destabilization and inhibition of p53. *Nat Genet* 2004; 6:55-62.

Khan MA, Partin AW, Rittenhouse HG, Mikolajczyk SD, Sokoll LJ, Chan DW, Veltri RW. Pro-PSA for early detection of prostate cancer in men with a total PSA range of 4.0-10.0 ng/ml. *J Urol* 2003; 170:723-36.

Kim JH, Herlyn D, Wong KK, Park DC, Schorge JO, Lu KH, et al. Identification of epithelial cell adhesion molecule autoantibody in patients with ovarian cancer. *Clin Cancer Res* 2003; 9:4782-91.

Kim J, Skates S, Uede T, Kwong-Kwok W, Schjorge JO, Feltmate CM, Berkowtiz RS, et al. Osteopontin as a potential marker for ovarian cancer. *JAMA* 2002; 287:1671-9.

King BL, Tsai SC, Gryga ME, D'Aquila TG, Seelig SA, Morrison LE, et al. Detection of chromosomal instability in paired breast surgery and ductal lavage specimens by interphase fluorescence in situ hybridization. *Clin Cancer Res* 2003; 9:1509-16.

Laronga C, Becker S, Watson P, Gregory B, Cazares L, Lynch H, et al. SELDI-TOF serum profiling for prognostic and diagnostic classification of breast cancers. *Dis Markers* 2003-2004; 19:229-38.

Li L, Tang H, Wu Z, Gong J, Gruidl M, Zhou J, et al. Data mining techniques for cancer detection using serum proteomic profiling. *Artif Intell Med* 2004; 32:71-83.

Li J, White N, Zhang Z, Rosenweig J, Mangold L, Partin AW, Chan DW. Detection of prostate cancer using serum proteomic pattern in a histologically confirmed population. *J Urol* 2004; 171:1782-7.

Lu X, Weiss P, Block TM. A phage with high affinity for hepatitis B surface antigen for detection of HBsAg. *J Vir Met* 2004; 119:51-4.

Lynch HT, Brand RE, Hogg D, Deters CA, Fusaro RM, Lynch JF, et al. Phenotypic variation in eight extended CDKN2A germline mutation familial atypical multiple mole melanoma-pancreatic carcinoma-prone families: the familial atypical multiple mole melanoma-pancreatic carcinoma syndrome. *Cancer* 2002; 94:84-96.

Lynch HT, Coronel SM, Okimoto R, Hampel H, Sweet K, Lynch JF, et al. A founder mutation of the MSH2 gene and hereditary nonpolyposis colorectal cancer in the United States. *JAMA* 2004; 291:718-24.

Lynch HT, de la Chapelle A. Genomic medicine: hereditary colorectal cancer. *New England J Med* 2003; 348:919-32.

Lynch HT, Deters CA, Hogg D, Lynch JF, Kinarsky Y, Gatalica Z. Familial sarcoma/challenging pedigrees. *Cancer* 2003; 98:1947-57.

Malyarenko DI, Cooke WE, Adam BL, Malik G, Chen H, Tracy ER, et al. Enhancement of sensitivity and resolution of surface-enhanced laser desorption/ionization time-of-flight mass spectrometric records for serum peptides using time-series analysis techniques. *Clin Chem* 2005; 51:65-74.

Marcus B, Arenberg D, Lee J, Kleer C, Chepeha DB, Schmalbach CE, et al. Prognostic factors in oral cavity and oropharyngeal squamous cell carcinoma. *Cancer* 2004; 101:2779-87.

Marrero JA, Su GL, Wei W, Emick D, Conjeevaram HS, Fontana RJ, et al. Des-gamma carboxyprothrombin can differentiate hepatocellular carcinoma from nonmalignant chronic liver disease in American patients. *Hepatology* 2003; 37:1114-21.

Miller CT, Moy JR, Lin L, Schipper M, Normolle D, Brenner DE, et al. Gene amplification in esophageal adenocarcinomas and Barrett's with high-grade dysplasia. *Clin Cancer Res* 2003; 9:4819-25.

Misek DE, Imafuku Y, Hanash SM. Application of proteomic technologies to tumor analysis. *Pharmacogenomics* 2004; 5:1129-37.

Misek DE, Kuick R, Hanash SM, Logsdon CD. Oligonucleotide-directed microarray gene profiling of pancreatic adenocarcinoma. *Methods Mol Med* 2004; 103:175-88.

Mori Y, Selaru M, Olaru A, Perry K, Kimos MC, Tamura G, et al. Instabilotyping reveals novel unique mutational spectra in microsatellite-unstable gastric cancers. *Cancer Res* 2002; 62:3641-5.

Nam MJ, Madoz-Gurpide J, Wang H, Lescure P, Schmalbach CE, Zhao R, et al. Molecular profiling of the immune response in colon cancer using protein microarrays: occurrence of autoantibodies to ubiquitin C-terminal hydrolase L3. *Proteomics* 2003; 3:2108-15.

Nawroz-Danish H, Eisenberger CF, Yoo GH, Wu L, Koch W, Black C, et al. Microsatellite analysis of serum DNA in patients with head and neck cancer. *Int J Cancer* 2004; 111:96-100.

Ojaniemi H, Evengard B, Lee DR, Unger ER, Vernon SD. Impact of RNA extraction from limited samples on microarray results. *BioTechniques* 2003; 35:968-73.

Padar A, Sathyanarayana UG, Suzuki M, Maruyama R, Hsieh JT, Frenkel EP, et al. Inactivation of cyclin D2 gene in prostate cancers by aberrant promoter methylation. *Clin Cancer Res* 2003; 9: 4730-4.

Park J, Chen L, Tockman MS, Elahi A, Lazarus P. The human 8-oxoguanine DNA N-glycosylase 1 (hOGG1) DNA repair enzyme and its association with lung cancer risk. *Pharmacogenetics* 2004; 14:103-9.

Phillips M, Cataneo RN, Cummin AR, Gagliardi AJ, Gleeson K, Greenberg J, et al. Detection of lung cancer with volatile markers in the breath. *Chest* 2003; 123:2115-23.

Prindiville SA, Byers T, Hirsch FR, Franklin WA, Miller YE, Vu, KO, et al. Sputum cytological atypia as a predictor of incident lung cancer in a cohort of heavy smokers with airflow obstruction. *Cancer Epidemiol Biomarkers Prev* 2003; 12:987-93.

Pusztai L, Gregory BW, Baggerly KA, Peng B, Koomen J, Kuerer HM, et al. Pharmacoproteomic analysis of prechemotherapy and postchemotherapy plasma samples from patients receiving neoadjuvant or adjuvant chemotherapy for breast carcinoma. *Cancer* 2004; 100:1814-22.

Qiu J, Madoz-Gurpide J, Misek DE, Kuick R, Brenner DE, Michailidis G, et al. Development of natural protein microarrays for diagnosing cancer based on an antibody response to tumor antigens. *J Proteome Res* 2004; 3:261-7.

Rajeevan MS, Dimulescu IM, Vernon SD, Verma M, Unger ER. Global amplification of sense RNA: a novel method to replicate and archive mRNA for gene expression analysis. *Genomics* 2003; 82:491-7.

Ratnasinghe DL, Yao SX, Forman M, Qiao YL, Andersen MR, Giffen CA, et al. Gene-environment interactions between the codon 194 polymorphism of XRCC1 and antioxidants influence lung cancer risk. *Anticancer Res* 2003; 23:627-32.

Reesink-Peters N, Wisman GB, Jeronimo C, Tokumaru CY, Cohen Y, Dong SM, et al. Detecting cervical cancer by quantitative promoter hypermethylation assay on cervical scrapings: a feasibility study. *Mol Cancer Res* 2004; 2:289-95.

Rom WN, Tchou-Wong KM. Functional genomics in lung cancer and biomarker detection. *Am J Respir Cell Mol Biol* 2003; 29:153-6.

Rom WN, Tchou-Wong KM. Molecular and genetic aspects of lung cancer. *Met Mol Med* 2003; 75:3-26.

Romeo MS, Sokolova IA, Morrison L E, Zeng C, Baron AE, Hirsch F R, et al. Chromosomal abnormalities in non-small cell lung carcinomas and in bronchial epithelia of high-risk smokers detected by multi-target interphase fluorescence in situ hybridization. *J Mol Diagn* 2003; 5:103-12.

Sathyanarayana UG, Maruyama R, Padar A, Suzuki M, Bondaruk J, Sagalowsky A, et al. Molecular detection of noninvasive and invasive bladder tumor tissues and exfoliated cells by aberrant promoter methylation of laminin-5 encoding genes. *Cancer Res* 2004; 64:1425-30.

Sauthoff H, Pipiya T, Heitner S, Chen S, Norman RG, Rom WN, Hay JG. Late expression of p53 from a replicating adenovirus improves tumor cell killing and is more tumor cell specific than expression of the adenoviral death protein. *Hum Gene Ther* 2002; 13:1859-71.

Semmes OJ. Defining the role of mass spectrometry in cancer diagnostics. *Cancer Epidemiol Biomarkers Prev* 2004; 13:1555-7.

Semmes OJ, Feng Z, Adam B-L, Banez LL, Bigbee WL, Campos D, et al. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin Chem* 2005; 51:102-12.

Sen S, Zhou H, Zhang RD, Yoon DS, Vakar-Lopez F, Ito S, et al. Amplification/overexpression of a mitotic kinase gene in human bladder cancer. *J Natl Cancer Inst* 2002; 94:1320-9.

Shao L, Lerner SL, Bondaruk J, Czerniak BA, Zeng X, Grossman HB, et al. Specific chromosome aberrations in peripheral blood lymphocytes are associated with risk of bladder cancer. *Genes Chromosomes Cancer* 2004; 41:379-89.

Shigematsu H, Suzuki M, Takahashi T, Miyajima K, Toyooka S, Shivapurkar N, et al. Aberrant methylation of HIN-1 (high in normal-1) is a frequent event in many human malignancies. *Int J Cancer* 2005; 113:600-4.

Shivapurkar N, Reddy JL, Matta H, Sathyanarayana UG, Huang CX, Toyooka S, et al. Loss of expression of death–inducing signaling complex (DISC) components in lung cancer cell lines and the influence of MYC amplification. *Oncogene* 2002; 55:8510-14.

Shivapurkar N, Toyooka S, Toyooka KO, Reddy J, Miyajima K, Suzuki M, et al. Aberrant methylation of TRAIL decoy receptors is frequent in multiple tumor types. *Int J Cancer* 2004; 109:786-92.

Slaton JW, Millikan R, Inoue K, Karashima T, Czerniak B, Shen Y, et al. Correlation of metastasis related gene expression and relapse-free survival in patients with locally advanced bladder cancer treated with cystectomy and chemotherapy. *J Urol* 2004; 171:570-4.

Sokoll LJ, Chan DW, Mikolajczyk SD, Rittenhouse HG, Evans CL, Linton HJ, et al. Proenzyme PSA for the early detection of prostate cancer in the 2.5-4.0 ng/mL total PSA range: preliminary analysis. *Urology* 2003; 61:274-6.

Steel LF, Shumpert D, Trotter M, Seeholzer SH, Evans AA, London WT, et al. A strategy for the comparative analysis of serum proteomes for the discovery of biomarkers for hepatocellular carcinoma. *Proteomics* 2003; 3:601-9.

Sugita M, Geraci M, Gao B, Powell RL, Hirsch FR, Johnson G, et al. Combined use of oligonucleotide and tissue microarrays identifies cancer/testis antigens as biomarkers in lung carcinoma. *Cancer Res* 2002; 62:3971-9.

Takahashi T, Shivapurkar N, Reddy J, Shigematsu H, Miyajima K, Suzuki M, et al. DNA methylation profiles of lymphoid and hematopoietic malignancies. *Clin Cancer Res* 2004; 10:2928-35.

Tchou-Wong KM, Rom WN. Lung-specific expression of mutant p53 as mouse model for lung cancer. *Met Mol Med* 2003; 74:465-80.

Tebbitt CL, Bentley RC, Olson JA Jr, Marks JR. Estrogen receptor alpha (ESR1) mutant A908G is not a common feature in benign and malignant proliferations of the breast. *Genes, Chromosomes & Cancer* 2004; 40:51-4.

Terry KL, Sluss PM, Skates SJ, Mok SC, Ye B, Vitonis AF, Cramer DW. Blood and urine markers for ovarian cancer: a comprehensive review. *Dis. Markers* 2004; 20:53-70.

Tokumaru Y, Yamashita K, Osada M, Nomoto S, Sun DI, Xiao Y, et al. Inverse correlation between cyclin A1 hypermethylation and p53 mutation in head and neck cancer identified by reversal of epigenetic silencing. *Cancer Res* 2004; 64:5982-7.

Tokumaru Y, Harden SV, Sun DI, Yamashita K, Epstein JI, Sidransky D. Optimal use of a panel of methylation markers with GSTP1 hypermethylation in the diagnosis of prostate adenocarcinoma. *Clin Cancer Res* 2004;10:5518-22.

Topaloglu O, Hoque MO, Tokumaru Y, Lee J, Ratovitski E, Sidransky D, Moon CS. Detection of promoter hypermethylation of multiple genes in the tumor and bronchoalveolar lavage of patients with lung cancer. *Clin Cancer Res* 2004; 10:2284-8.

Toyooka S, Maruyama R, Toyooka KO, McLerran D, Feng Z, Fukuyama Y, et al. Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer. *Int J Cancer* 2003; 103:153-60.

Toyooka S, Toyooka KO, Harada K, Miyajima K, Makarla P, Sathyanarayana UG, et al. Aberrant methylation of the CDH13 (H-cadherin) promoter region in colorectal cancers and adenomas. *Cancer Res* 2002; 62:3382-6.

Toyooka S, Toyooka KO, Miyajima K, Reddy JL, Toyota M, Sathyanarayana UG, et al. Epigenetic down-regulation of death-associated protein kinase in lung cancers. *Clin Cancer Res* 2003; 9:3034-41.

Unger ER, Steinau M, Rajeevan MS, Swan D, Lee DR, Vernon SD. Molecular markers for early detection of cervical neoplasia. *Dis Markers* 2004; 129:193-7.

Varella-Garcia M, Kittelson J, Schulte AP, Vu KO, Wolf HJ, Zeng C, et al. Multi-target interphase fluorescence in situ hybridization assay increases sensitivity of sputum cytology as a predictor of lung cancer. *Cancer Detect Prev* 2004; 28:244-51.

Vathesatogkit P, Harkin TJ, Addrizzo-Harris DJ, Bodkin M, Crane M, Rom WN. Clinical correlation of asbestos bodies in BAL fluid. *Chest* 2004; 126:966-71.

Vernon SD, Farkas DH, Unger ER, Chan V, Miller DL, Chen Y-P, et al. Bioelectronic DNA detection of human papillomaviruses using eSensor™: a model system for detection of multiple pathogens. *BMC Infect Dis* 2003; 3:12.

Virmani A, Rathi A, Sugio K, Sathyanarayana UG, Toyooka S, Kischel FC, et al. Aberrant methylation of TMS1 in small cell, non small cell lung cancer and breast cancer. *Int J Cancer* 2003; 106:198-204.

Vlahou A, Giannopoulos A, Gregory BW, Manousakas T, Kondylis FI, Wilson LL, et al. Protein profiling in urine for the diagnosis of bladder cancer. *Clin Chem* 2004; 50:1438-41.

Vlahou A, Laronga C, Wilson L, Gregory B, Fournier K, McGaughey D, et al. A novel approach towards development of a rapid blood test for breast cancer. *Clinical Breast Cancer* 2003; 4:203-9.

Wadsworth JT, Somers KD, Cazares LH, Malik G, Adam BL, Stack BC Jr, et al. Serum protein profiles to identify head and neck cancer. *Clin Cancer Res* 2004; 10:1625-32.

Wadsworth JT, Somers KD, Stack BC Jr, Cazares L, Malik G, Adam BL, et al. Identification of patients with head and neck cancer using serum protein profiles. *Arch Otolaryngol Head Neck Surg* 2004; 130:98-104.

Wagner M, Naik DN, Pothen A, Kasukurti S, Devineni RR, Adam BL, et al. Computational protein biomarker prediction: a case study for prostate cancer. *BMC Bioinformatics* 2004; 5:26.

Wehrli BM, Huang W, de Crombrugghe B, Ayala AG, Czerniak, B. Sox9, a master regulator of chondrogenesis, distinguishes mesenchymal chondrosarcoma from other small blue round cell tumors. *Hum Pathol* 2003; 34:263-9.

Yamashita K, Upadhyay S, Osada M, Hoque M, Xiao Y, Mori M, et al. Pharmacologic unmasking of epigenetically silenced tumor suppressor genes in esophageal squamous cell carcinoma. *Cancer Cell* 2002; 2:485-95.

Ye B, Cramer DW, Skates SJ, Gygi SP, Pratomo V, Fu L, et al. Haptoglobin-alpha subunit as potential serum biomarker in ovarian cancer: identification and characterization using proteomic profiling and mass spectrometry. *Clin Cancer Res* 2003; 9:2904-11.

Zhukov TA, Johanson RA, Cantor AB, Clark RA, Tockman MS. Discovery of distinct protein profiles specific for lung tumors and pre-malignant lung lesions by SELDI mass spectrometry. *Lung Cancer* 2003; 40:267-79.

II Publications Coauthored by NCI Program Staff

Atha DH, Miller K, Sanow AD, Xu J, Hess JL, Wu OC, et al. High-throughput analysis of telomerase by capillary electrophoresis. *Electrophoresis* 2003; 24:109-14.

Barker PE, Wang W, Wagner PD, Pinsky P. Inter-rater agreement on chromosome 5 breakage in FISH-based mutagen sensitivity assays (MSAs). *Mutat Res* 2004; 562:133-42.

Grizzle WE, Adam B-L, Bigbee WL, Conrads TP, Carroll C, Feng Z, et al. Serum protein expression profiling for cancer detection: validation of a SELDI-based approach for prostate cancer. *Dis Markers* 2003-2004; 19:185-95.

Grizzle WE, Semmes OJ, Basler J, Izbicka E, Feng Z, Kagan J, et al. The early detection research network surface-enhanced laser desorption and ionization prostate cancer detection study: a study in biomarker validation in genitourinary oncology. *Urol Oncol* 2004; 22:337-43.

Habis AH, Vernon SD, Lee DR, Verma M, Unger ER. Molecular quality of exfoliated cervical cells: implications for molecular epidemiology and biomarker discovery. *Cancer Epidemiol Biomarkers Prev* 2004; 13:492-6.

Jakupciak JP, Wang W, Barker PE, Srivastava S, Atha DH. Analytical validation of telomerase activity for cancer early detection: TRAP/PCR-CE and hTERT mRNA quantification assay for high-throughput screening of tumor cells. *J Mol Diagn* 2004; 6:157-65.

Maruvada P, Srivastava S. Biomarkers for cancer diagnosis: implications for nutritional research. *J Nutr* 2004; 134:1640S-45S.

Rajeevan MS, Dimulescu IM, Vernon SD, Verma M, Unger ER. Global amplification of sense RNA: a novel method to replicate and archive mRNA for gene expression analysis. *Genomics* 2003; 82:491-7.

Ravichandran R, Vasquez GB, Srivastava S, Verma M, Petricoin E, Lubell J, et al. Data standards for proteomics: mitochondrial two-dimensional polyacrylamide gel electrophoresis data as a model system. *Mitochondrion* 2004; 3:327-36.

Verma M. Viral genes and methylation. *Ann NY Acad Sci* 2003; 983:1-11.

Verma M, Dunn BD, Ross S, Jain P, Wang W, Hayes R, Umar A. Proceedings of the workshop on epigenetics in cancer prevention: early detection and risk assessment. *Ann NY Acad Sci.* 2003; 983:298-319.

Verma M, Maruvada P, Srivastava S. Current approaches in mitochondrial proteomics. *MitoMatters* 2004; 3(3-5).

Verma M, Kagan J, Sidransky D, Srivastava S. Proteomic analysis of cancer-cell mitochondria. *Nat Rev Cancer* 2003; 3:789-95.

Verma M, Srivastava S. Epigenetics in cancer: implications for early detection and prevention. *Lancet Oncol* 2002; 3:755-63.

Verma M, Srivastava S. New cancer biomarkers deriving from NCI early detection research. *Recent Results Cancer Res* 2003; 163:72-84.

Wagner PD, Maruvada P, Srivastava S. Molecular diagnostics: a new frontier in cancer prevention. *Expert Rev Mol Diagn* 2004; 4:503-11.

Wagner PD, Verma M, Srivastava S. Challenges for biomarkers in cancer detection. *Ann NY Acad Sci* 2004; 1022:9-16.

III Metrics for Programmatic Evaluation

Excerpted from the Manual of Operations of the Early Detection Research Network
December 21, 2004

It is the responsibility of the awarding agency, in this case the National Cancer Institute (NCI), National Institutes of Health, to review progress achieved towards scientific goals in original grant applications over specified grant periods and to provide scientific and logistical input to grantees to enhance the quality of their scientific efforts. For details, see HHS 45 CFR, Part 74. To review progress towards achieving the objectives of the Early Detection Research Network (EDRN) and its investigators, it is imperative for EDRN program officials to gather information on the functioning of the network in order to update the NCI leadership. This document describes metrics, rationale, and standards for evaluating the overall success of the EDRN.

Introduction

Fair, rigorous peer review of investigator-initiated scientific applications remains the cornerstone of scientific progress in the United States. Peer review has ensured that the best science is supported. The EDRN was initiated with this concept in mind. By selecting scientific collaborators for the EDRN on the basis of rigorous peer review and fully funding the best applications, the NCI has successfully obtained strong participation from the scientific community.

The EDRN represents a major pioneering effort in collaborative translational research. It departs from prior Cancer Cooperative Group models in many important ways - through empowering investigators by funding their Centers directly and by placing the burden of scientific leadership, research agenda, and collaboration upon these directly funded Centers. Basic scientists with robust bench research records have been funded to pool their ideas, resources, and tools. Translational and epidemiologic investigators with strong tools and publication track records are directly funded with a mandate to translate concepts arising from basic science labs. Analytical tools, laboratories, statistical methods, and informatics are also supported directly with a collaborative mandate. Leadership of this collaborative must emanate from the grass-root investigators, and the Executive Leadership must communicate with a highly knowledgeable group of scientists in a manner that enhances collaboration and productivity. This Network represents a new paradigm of Cooperative research.

NCI Charge to the EDRN

At the opening meeting of all of the funded EDRN units, NCI leadership and Program Staff provided the following charge for this collaborative enterprise:

- Establish criteria for the discovery and validation of biomarkers at all points of the integrated research scheme;
- Establish a rigorous quality assurance/quality control program for biomarkers;

- Establish and deal with issues of biorepositories-how the samples will be obtained, stored and most importantly, allocated;
- Support Translational Research Projects-both within and outside the EDRN-and establish policies and procedures that are inclusive of investigators who wish to utilize the infrastructure and facilities of the EDRN;
- Establish and foster industrial collaborations which will be crucial to the ability to rapidly translate the research effort into products and to test innovative biomarkers being developed by industry;
- Establish and maintain effective and efficient communications, including the use of EDRN websites (public and private), listservs, email, and regularly scheduled meetings;
- Develop and maintain an effective, efficient, and productive management domain with minimal committee structure and maximal collaboration, with financial rewards for collaboration;
- Encourage inclusiveness by ensuring that scientists with promising research ideas get the opportunity to collaborate constructively with the EDRN.

Evaluation Metrics

Since there are no prior models of such a cooperative research enterprise, it is very important to carefully monitor and assess progress from both macro and micro perspectives. This review will be particularly important during the first grant period in which substantial administrative effort should be expended in order to build the new infrastructure. The following evaluation metrics are suggested.

For the Individual Laboratory and Center

1. Scientific Excellence

Quality of Questions: Has the EDRN site clearly defined their objectives, hypotheses, and scientific plan?

Scientific Progress to Date: Has the EDRN site made progress towards meeting these objectives as specified in their originally funded research plan? What pitfalls have been encountered and how have they been managed?

Innovation: How has the EDRN site used innovation to overcome obstacles? Is the site aware of new methods or approaches that might be useful to or portable into the EDRN environment?

Future Plans: What does the site plan to do over the coming two years? How will these plans meet the original grant objectives?

2. Productivity Metrics

Publication productivity: Has the site published papers on the objectives funded by the EDRN? How many and in what Journals? If not, are there problems that need to be addressed or require assistance?

Grant funding: Has the site applied for additional grant or contract funding? Has the site team been successful in gaining additional funds? Has the EDRN been helpful to the success of funding these new grants or contracts?

Biomarkers identified (BDLs): Number of new biomarkers pursued for evaluation? Number of biomarkers sent forward to CEVCs or BRLs for validation? Number of biomarkers added to early detection or risk assessment panels? Number of biomarkers used in chemoprevention clinical trials?

Assays performed (BRLs): Numbers of assays developed for EDRN projects? Numbers of samples processed? Types of samples processed? Results reported? Quality control of samples assayed? Number and type of development projects approved? Use of CDEs?

DMCC: Standards of informatics support? Type of informatics, QC procedures, patient privacy protection measures, data storage, and retrieval systems for Validation Studies? Development of Network-wide communication systems? Development of Network-wide systems to promote data and specimen sharing? Development of statistical methodology to meet the needs of EDRN?

Samples collected and provided (CEVCs): Numbers of samples collected? Types of samples collected? Sources of samples collected? Numbers of samples provided to EDRN BDLs or BRLs? Use of CDEs? How many CEVCs have had their set-aside funds released? How many CEVCs have requested the release of developmental funds?

3. Collaborative Metrics

EDRN collaborations: With whom is the EDRN site collaborating? How many projects are collaborative? How many joint papers have been published? Use of EDRN resources: Has the EDRN site collaborated with CEVCs, a BRL or BDL site? If so, how many? Joint publications? Joint grants? How many BDLs have requested release of their restricted funds for Network Collaborative Studies?

Participation in EDRN Activities: Attendance from the site at EDRN meetings. Participation on Committees, working groups, and task forces? Special EDRN projects completed. Did EDRN site participate in developing the CDEs? Did EDRN site help to standardize/streamline the IRB approval process? Did EDRN site help develop systems for streamlining data sharing and/or specimen sharing? Did EDRN site help develop systems to standardize/streamline technology transfer issues?

EDRN outreach: Number of new Associate Members from the outside? Amount of Chair's funds allocated to new Associate Members? The number of applications for Chair's funding? Other outreach activities?

Process for Evaluating Metrics

1. Annual written progress report

Reviews should be based upon the yearly progress report required for non-competitive renewal. Instructions for preparation of the non-competitive renewal should be specific and emphasize progress towards scientific goals of the original grant application and progress towards addressing EDRN's mission. While scientific quality and progress need to be recorded and addressed, primarily, metrics should be required to allow NCI staff to report data to NCI leadership.

The review process should assess the progress of each of the funded units towards meeting the specific aims of their funded grant application and their progress and contributions in meeting the above-described charges for the entire group. While the review is structured to provide NCI leadership and staff with data to track the progress of the EDRN and its components, equally important goals are to provide constructive feedback to EDRN Principal Investigators and their collabora-

tors. Reviews may be used by EDRN leadership, NCI staff, and the Network Consulting Team to make mid-course changes or to encourage constructive changes in individual scientific direction or focus. Initial reviews might assist in building collaborations among investigators and their groups. Reviews may also be used to assess administrative progress, to quantify publications and grants, and to quantify numbers of subjects studied.

2. Site Visits:

Each Center/Laboratory should be site visited by a panel comprised of external consultants (individual members of the Network Consulting Team), NCI staff and other experts on an as needed basis. The site visit should be brief (preferably a half day or less) but enable a thorough review of scientific progress, future scientific plans, performance metrics, facilities and staff in support of the EDRN charge. The site Principal Investigator would provide a 2-3 hour presentation period to review scientific progress, spell out new scientific initiatives for EDRN research, and address required metrics. The Principal Investigator should be encouraged to share problems, concerns, and questions to the site visit team so that the process is interactive and collegial. While an agenda and presentation should be necessary, no scoring should be used.

3. Frequency of Site Visits

The frequency of the site visits will be determined by the NCI. However, it is anticipated that one initial site visit by NCI program officials, in year one will occur, and one mid-grant site visit (for a five-year grant, it will be the between the year 2 and year 3). Additional site visits may be required when deemed necessary by the NCI.

Deficient performance and remedies will be conducted in accordance with HHS 45 CFR, Part 74 and other pertinent regulations.

4. At time of Type 5 renewal each site must submit their Site Task Reminder List generated from the EDRN Secure Website. In addition each site must submit a copy of each of the individual reports (publications, specimen survey, research interests, and registered protocols).

Overall Evaluation of Early Detection Research Network

It is the intention of the NCI that the members of the Network Consulting Team and Chairs and Co-Chairs of the EDRN Steering Committee will discuss the overall performance of the EDRN using the metrics presented in this document and suggest changes/ modifications in the working structure of EDRN for the next five-year cycle.

Acknowledments

Dr. Srivastava would like to acknowledge the following individuals for their assistance in the creation of this Third Report:

Paul Wagner, Ph.D.
Program Director
Division of Cancer Prevention
National Cancer Institute

Jacob Kagan, M.Sc., Ph.D.
Program Director
Division of Cancer Prevention
National Cancer Institute

Kara Smigel Croker, M.S.
Communications Manager
Division of Cancer Prevention
National Cancer Institute

Gwen Moulton
Editorial Consultant
GM Authoring Services

Tish Holbrook
Graphic Design

Glossary

The entries defined here are highlighted in bold type at the first occurrence in each section of this report.

Aneuploidy - Possessing an abnormal number of chromosomes.

Biomarker - A characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic response to a therapeutic intervention.

Bioinformatics - Computational analysis and management of biomedical information.

Clinical Endpoint - A characteristic or variable that reflects how a patient feels, functions, or survives.

Diploid - Possessing the correct number of chromosomes (two sets).

Enzyme-linked immunosorbent assay (ELISA) - A method where antibodies are used to quantify levels of a biological marker.

Epigenetics, epigenomics - The study of events affecting the functional state of DNA and gene expression without changing its sequence or linear arrangement.

Fluorescence in situ hybridization (FISH) - A technique using fluorescent probes to visualize locations of specific gene sequences on chromosomes. Often used for gene mapping and identifying chromosomal abnormalities.

Genomics - Characterization of the entire DNA and gene expression within a cell, tissue, or organism.

Mass spectrometry (MS) - A method using sophisticated instruments to detect molecules based on their sizes.

Matrix-assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF) - A platform for profiling a population of proteins by mass spectrometery according to the size and net charge of individual proteins. The peaks identified by this method require further analysis to determine the identitiy of the corresponding proteins.

Methylation - The addition of a methyl group to specific sites on DNA. The methylation of a gene can change its expression.

Methylation specific PCR (MSP) - An adaptation of PCR to identify and quantitate relative levels of methylated genes in DNA.

Microarray - A system of printing large numbers of DNA sequences, proteins, antibodies, or tissue lysates on a slide which can then be analyzed in a high-throughput fashion.

Microsatellite (Instability) Analysis (MSA) - Microsatellites are short sequences of DNA, usually 1 to 4 base pairs in length, repeated any number of times in various locations of DNA. Microsatellite instability analysis is a test to determine if the number of repeating units has changed at any specific location(s).

Polymerase chain reaction (PCR) - A technique to amplify, or produce multiple copies, of a defined span of DNA.

Proteomics - Characterization of all proteins from a biological source.

Quantitative PCR (qPCR) - An adaptation of PCR to quantify levels of defined mRNA transcripts.

Sensitivity - The proportion of individuals with a disease who test positive.

Specificity - The proportion of individuals without a disease who test negative.

Surface-enhanced Laser Desorption-Time of Flight (SELDI-TOF) - A modification of MALDI-TOF where some selectivity of proteins can be achieved prior to analysis.

Surrogate Endpoint - A biomarker intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm, or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic or other scientific evidence.

Throughput - The number of samples that can be processed in a defined time period.

Transcriptomics - Characterization of all genes transcribed to mRNA within a cell or tissue.

Two-Dimensional electrophoresis (2-DE) - A technique to separate proteins in two dimensions by two different criteria.