

**Early  
Detection  
Research  
Network**



**Statistical and Informatics  
Infrastructure:  
EDRN DMCC and JPL**

*Ziding Feng, DMCC*  
*Mark Thornquist, DMCC*  
*Dan Crichton, JPL*

**Early  
Detection  
Research  
Network**



**Statistical Tools and  
Research Management**

*Ziding Feng, Ph.D.*

*Mark Thornquist, Ph.D.*

*Margaret Pepe, Ph.D.*

*EDRN Data Management and Coordinating Center*

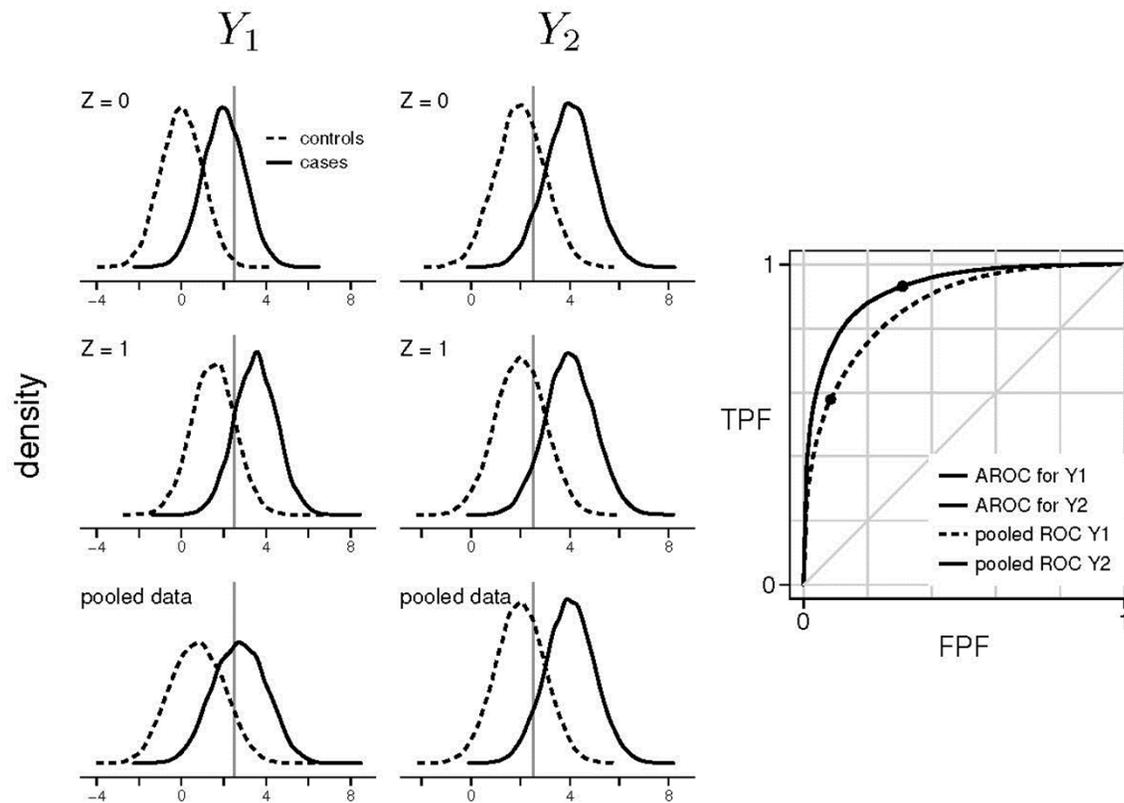
## **Areas of Emphasis**

1. Design of validation studies
2. Analysis of validation studies
3. Censored event time outcomes
4. Issues in risk prediction
5. Quantitative proteomics methods
6. Biomarker discovery

## **Dissemination**

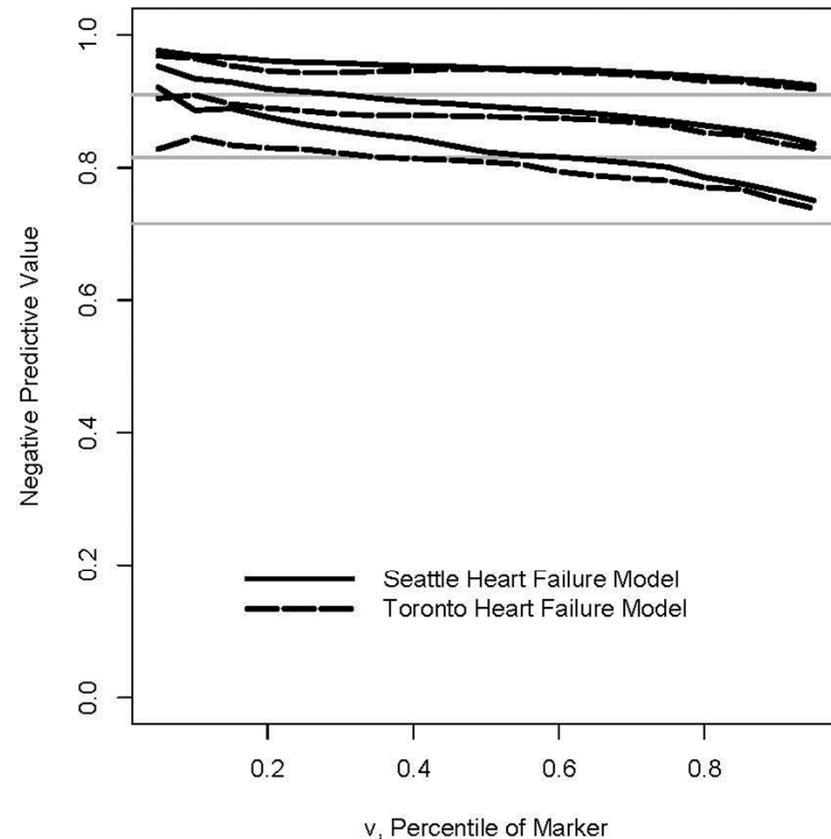
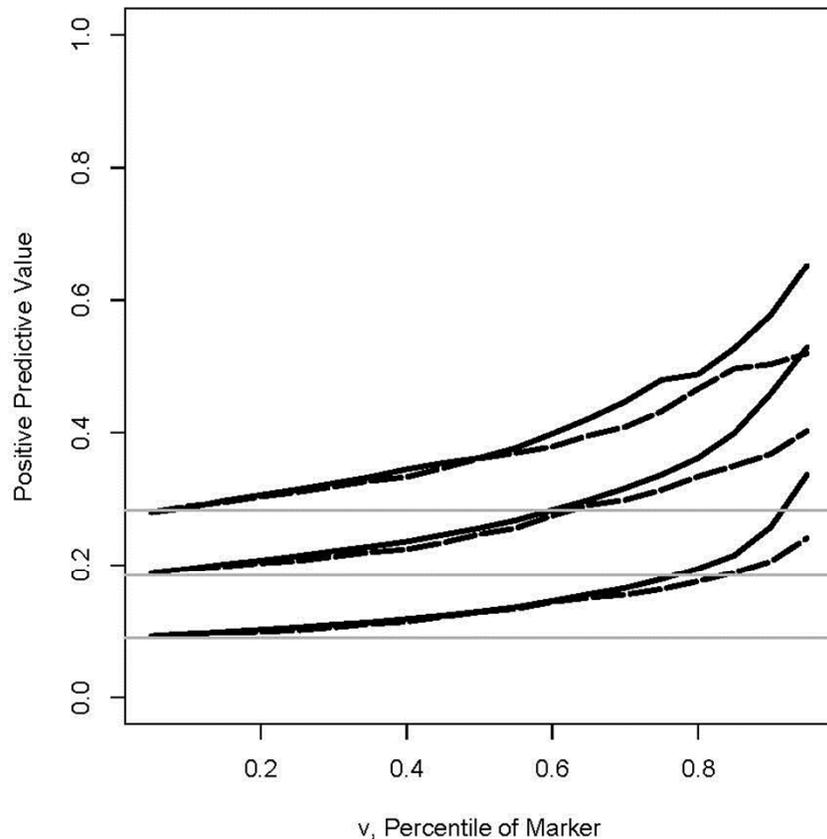
# Example: Analysis of Validation Studies

$Y_1$  and  $Y_2$  have the same performance,  $Y_1$  varies with  $Z$ , but  $Y_2$  does not. The unadjusted ROC curve for  $Y_1$  is attenuated while that for  $Y_2$  is not.



# Example: Analysis of Markers' Prognostic Value

PPV (left panel) and NPV (right panel) curves at  $t = 1$ ,  $t = 2$  and  $t = 3$  years (from top to bottom) after enrollment. Solid lines, estimates from the Seattle heart failure model; dashed lines, estimates from the Toronto heart failure model; Gray horizontal lines are for  $P(T < 1\text{year})$ ,  $P(T < 2\text{years})$  and  $P(T < 3\text{years})$  in the PPV plot and  $P(T \geq 1\text{year})$ ,  $P(T \geq 2\text{years})$  and  $P(T \geq 3\text{years})$  in the NPV plot.



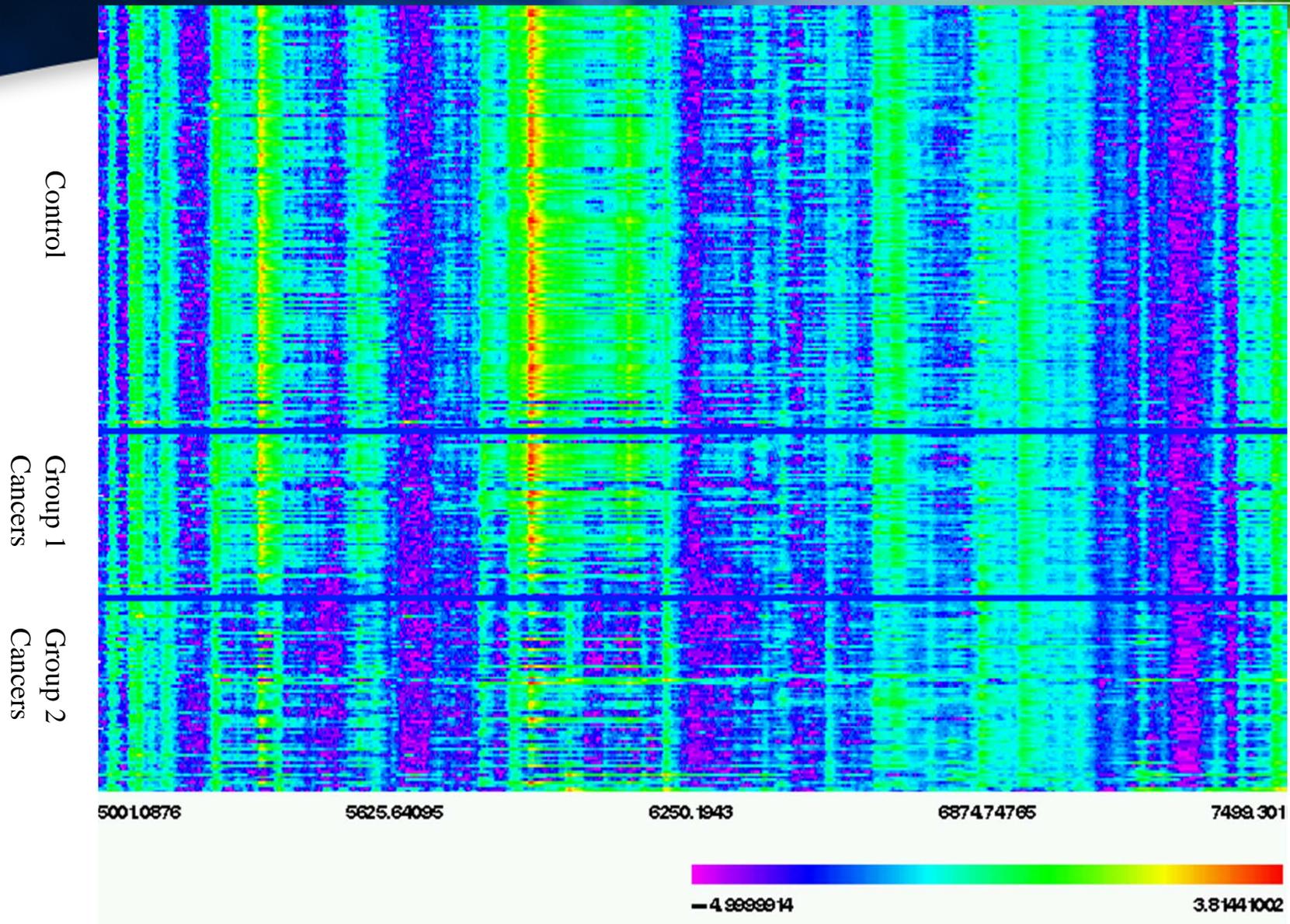
## Short Courses

- For statistical audiences
- 2007: FDA (September), JSM (August), SISG (June)
- Plans to expand to other audiences

## Software

- DABS website accessible from EDRN website
- Stata commands for ROC analysis to do estimation, comparisons and regression with and without covariate adjustment
- Plans to translate to R, which is free
- Stata commands for power calculations also
- Proteomics data and code on EDRN website. Code incorporated into the Insightful **Splus Proteome** tool

# Cases and Controls Ordered by Blood Draw Date



# Pepe et al.: Design of Phase II Biomarker Validation Studies Check Lists

## Clinical Context

- Define the population; Enrollment criteria and process; Setting for specimen collection
- Is generality adequate? Consider multiple institutions, heterogeneous population, simple protocol
- Specify procedures for outcome ascertainment; Prospective specimen collection prior to outcome ascertainment?
- Define case (seek biomarker positive in cases) and subsets of interest; Define control (seek biomarker negative in controls) and subsets of interest; Do all subjects in the population fit into a case or control category?
- Random selection of cases and controls?
- Matching of controls to cases on factors related to the biomarker? Consider limitations on questions that can be addressed by the study

## Marker Performance

- Define TPR= Proportion of cases positive; FPR= Proportion of controls positive
- Does 'time between obtaining specimen and occurrence of outcome' enter into definition of (TPR,FPR)?
- Are there subgroups of cases and controls for whom (TPR,FPR) will be calculated separately?
- What are minimally acceptable values (or ranges) for (TPR,FPR) in the clinical application?
- For a prediction marker: 'positive' is defined as the prediction probability exceeding a threshold. Provide the threshold and rationale for its choice. Similarly justify a threshold if identification of a low risk subgroup is the study objective.
- For a prediction marker: define adequate calibration of the risk model
- Does a classification method currently exist? What is its performance? How will comparative performance be quantified? What are target levels?

## Marker Measurement

- Specify procedures for specimen collection, processing, storage and retrieval
- Specify assay procedures and how results are reported
- Are mechanisms in place to blind specimen handling, assay and reporting of results to outcome status?
- Are the biomarker data to be combined with other information on the patient in the intended clinical application? This could include other clinical information, other markers, previous measurements of the biomarker in the patient. Specific algorithm for calculating the combination must be defined (it cannot be developed within the study).
- If other biomarkers or predictors will be combined or compared: Describe in detail protocols and procedures for obtaining these data; Provide assurance that procurement of these items is blinded to patient outcomes.

## Sample Size and Power

- Recall minimally acceptable performance criteria
- Define and provide rationale (data evidence) for anticipated performance levels.
- Calculate case and control sample sizes
- Plan for prospective collection until sufficient numbers of cases and controls are enrolled.
- Plan for early termination of the study if appropriate.

# Study Coordination

## Administration

- Conduct planning meeting
- Hold regular conference calls
- Create mailing lists
- Develop timeline
- Track IRB approvals and MTAs
- Develop Manual of Operations
- Resolve intellectual property issues
- Resolve commercial/industrial issues
- Accommodate FDA requirements

## Development

- Calculate sample sizes
- Determine matching criteria
- Blind samples
- Perform inter-lab comparisons
- Determine data elements
- Standardize specimen collection, processing and shipping guidelines

## Analysis and Dissemination

- Statistical Analysis
- Publication
- Dissemination of Data

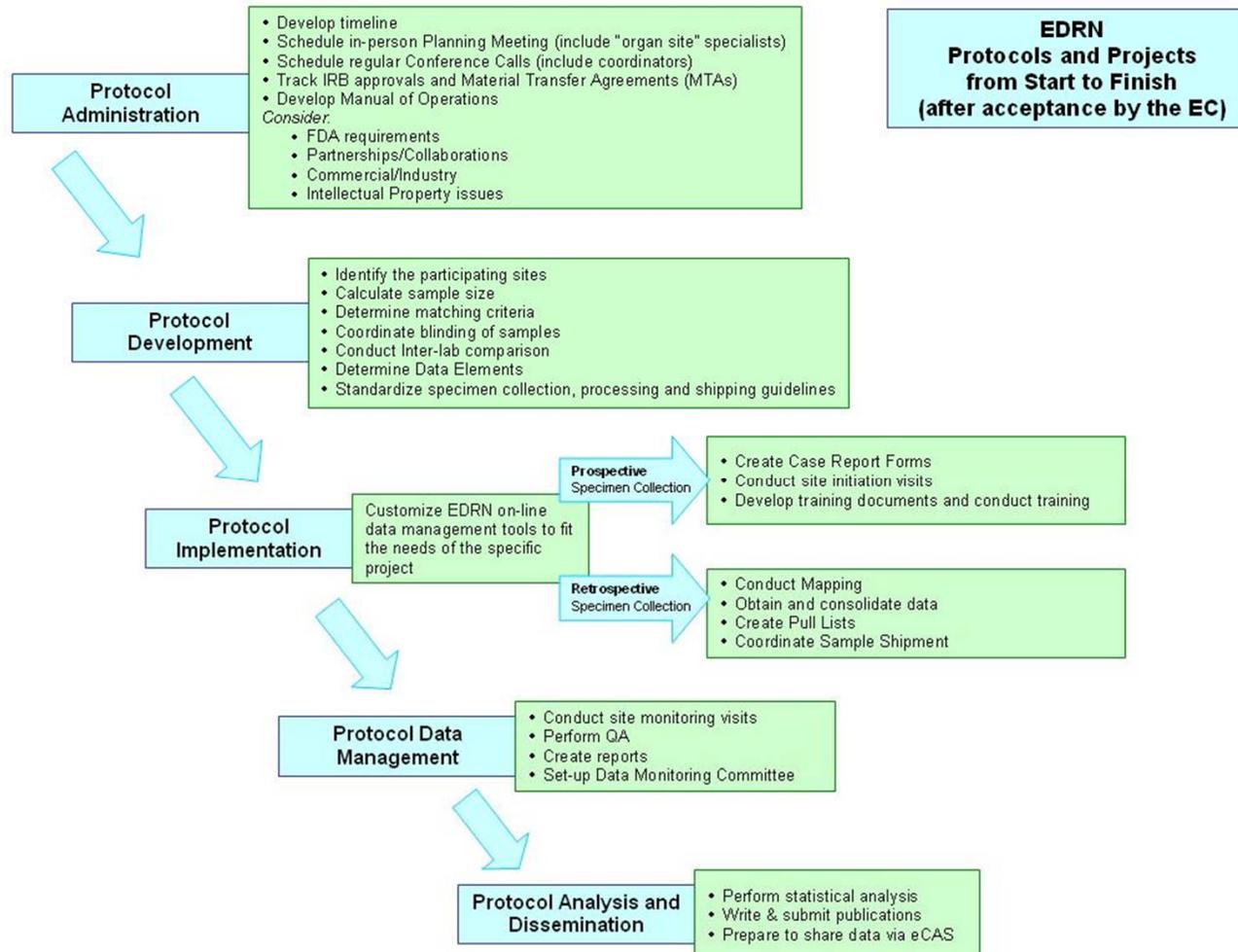
## Implementation

- Create case report forms
- Customize data management tools
- Conduct site initiation visits & training
- Obtain data
- Select matched cases and controls
- Perform sample assay assignment

## Data Management

- Conduct site monitoring visits
- Prepare reports
- Perform quality assurance
- Hold regular data monitoring committee meetings
- Revise protocol as needed
- Update Manual of Operations as needed
- Distribute revisions and updates to performance sites

# Flow Chart of DMCC Study Coordination



# Validation Study Information Management System (VSIMS)

EDRN's web-based data management system to promote consistent study execution and high-quality study data collected using Common Data Elements (CDEs)

Data Entry  
QA  
Specimen Tracking  
Administrative  
Flexible  
Consistent  
Easy to Implement  
Easy to Maintain  
Traceable



**VSIMS** An infrastructure for supporting validation studies in the EDRN  
Validation Study Information Management System Sponsored by National Cancer Institute

Early Detection Research Network

Protocol: DCP (ID: 111) User: Stephanie Page-Lester (Site ID: 5) Logout Home Back Print VSIMS

- Submit Data
- Confirm Eligibility
- Specimens
- Reports
- Issue Tracking
- Data Transfer
- Study Info
- Administration

**Study Update - DCP:**  
The EDRN Validation Study entitled "Validation of Serum Markers for the Early Detection of Hepatocellular Carcinoma" will investigate a new diagnostic test called des-gamma carboxyprothrombin (DCP) for liver cancer [Study Overview](#)

**Documents:**  
[DCP Newsletter - March/April 2007](#)  
[DCP Newsletter - January/February 2007](#)  
[DCP Newsletter - November/December 2006](#)

**Recruitment Plots:**  
[New recruitment/month](#)  
[Cumulative recruitment](#)

**Recruit a Matched Control:**  
[Click here](#) to determine if a potential Control has a matching Case.

**Message Board**

As of July 15, 2007, enrollment for the DCP Study has ended. Cases and Controls pending enrollment prior to July 15th can still be confirmed to the study. Please remember to continue the 6-month follow-up for all confirmed Controls.

Secure site maintained by COMPASS, Fred Hutchinson Cancer Research Center®. Last updated on 8/8/2007. Contact: [edrdmcc@fhcrc.org](mailto:edrdmcc@fhcrc.org)

# VSIMS Functionality

## Data Entry

- Audit trail
- Built-in edit checks
- Easy access to instructions
- Print CRFs for site's records
- Participant privacy

## Confirm Eligibility

- Consent form status tracking
- Required/optional data elements
- Assigned study group
- Visit schedules

## Issue Tracking

- Data Clarification Form
- Inquiry to Coordinating Center
- Request from Coordinating Center
- Status
- Full History

## Specimen Tracking System

- Tracks Original and Child Specimens
- 2-D barcodes & heat thermal printed labels
- Tracks specimen-related CDEs
- Assay laboratories are blinded to participant information
- All specimens collected in standardized fashion
- Cases and Controls are paired according to matching criteria defined by the protocol
- Pull Lists are created and Algorithm Batch Lists are designed for processing lab

## Study Reports

- Participant Reports
  - Enrollment
  - Adherence
  - Completeness
- Specimen Reports
  - Defective List
  - Specimens Collected
  - Status of Assay Results
  - Pull Lists

## Study Information

- Protocol
- Manual of Operations
- Bulletins
- Mailing List Subscriptions
- Directory
- Study-specific newsletters
- IRB

Satisfies 21 CFR Part 11 requirements for data systems and so can be used to support studies intended for submission to FDA

# EDRN'S CONTRIBUTIONS TO BIOMARKER RESEARCH

- **Phase structure for biomarker discovery and validation studies is being used extensively by the extramural community**
- **Recognition of chance, bias, and other confounders in biomarker research through EDRN experience; Corrective measures and methodologies developed by EDRN are acknowledged by the statistical community**
- **Dissemination of methodologies, study design issues and validation concepts to the research and regulatory communities, e.g., EDRN-FDA meeting, statistical workshops**
- **Study designs and methodologies developed for proteomic-based diagnostics are proving useful to other researchers**

**Early  
Detection  
Research  
Network**



# Connecting Biomarker Research Across the EDRN

***Dan Crichton***

***Program Manager, Planetary Data System Engineering***

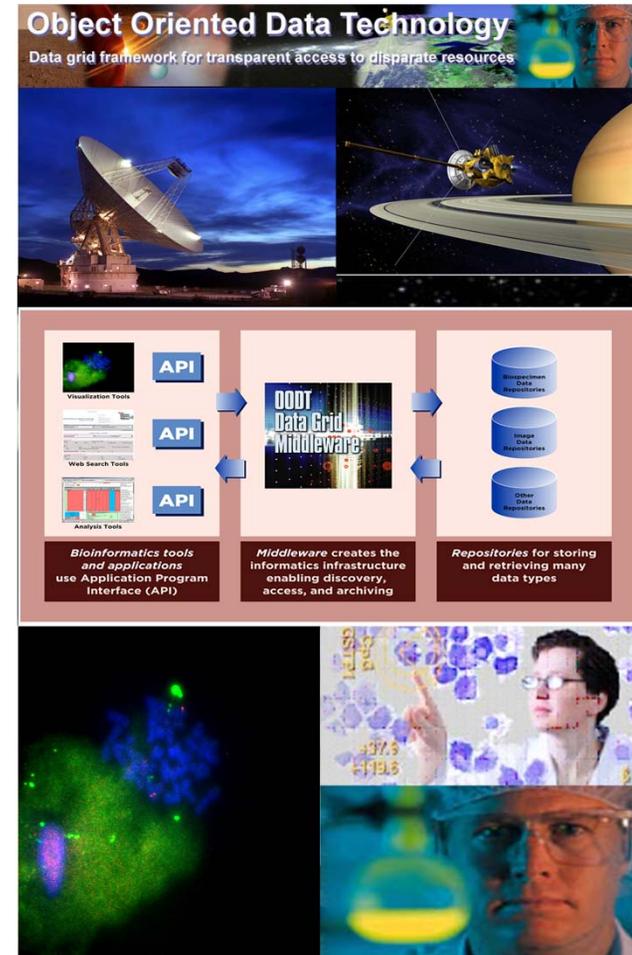
***PI, EDRN Informatics***

***Principal Computer Scientist***

***Jet Propulsion Laboratory, Caltech***

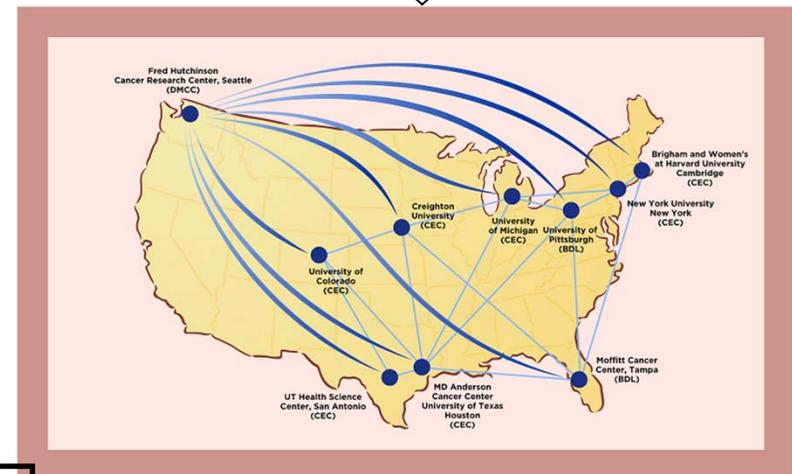
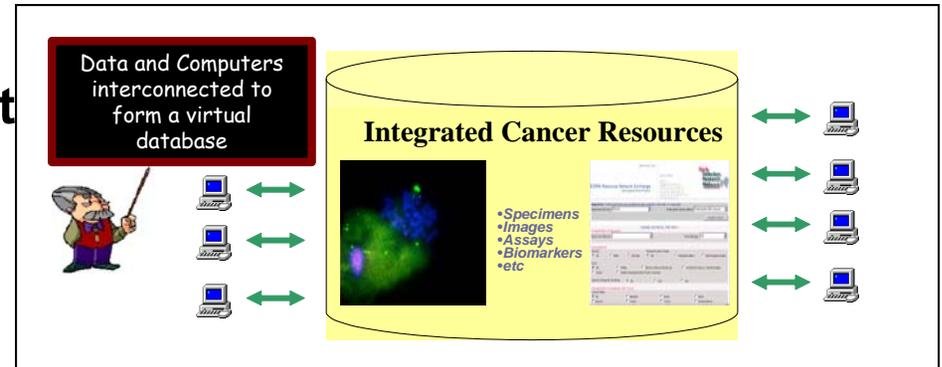
# Motivation and Challenges for Informatics in Science

- **Distributed informatics environments are being developed to support the capture and dissemination of scientific results across physical and life sciences**
  - Good success cases exist (e.g., Planetary Data System)
  - Provides a data management research platform for distributed communities
  - Technology is not the limiting factor; technologies and standards exist to connect enterprises together
- **The challenge is informatics infusion into the scientific discovery process**
  - Requires an “informatics architecture” to support the scientific discovery process
  - Operational model needs to include cross-disciplinary team including working scientists; teams are often organized around sub-disciplines



# Informatics for Biomarker Research

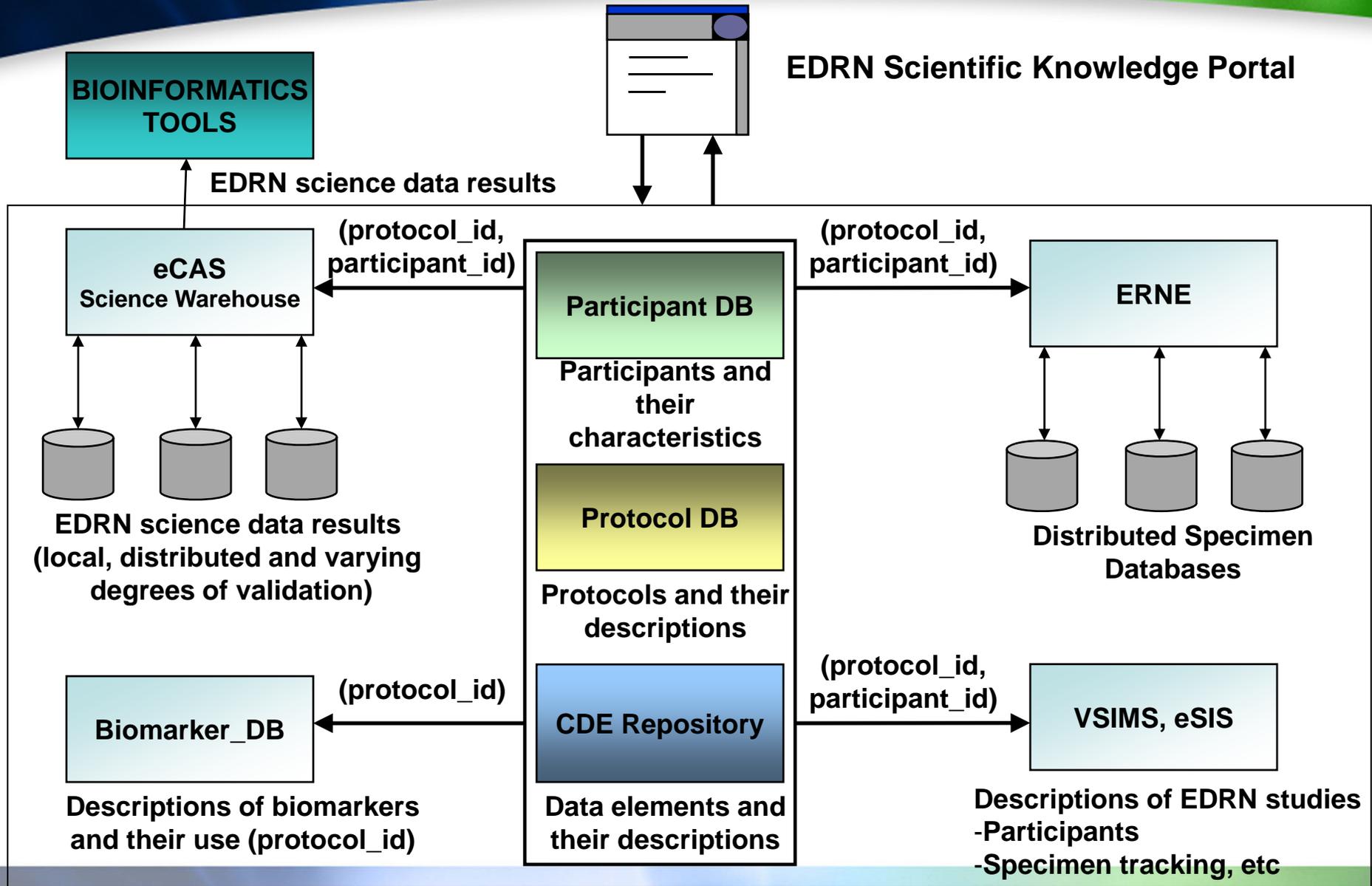
- EDRN has pioneered in the use of informatics technologies to support biomarker research
- EDRN has developed a comprehensive infrastructure to support biomarker data management across EDRN's distributed cancer centers\*
- It supports capture and access to a distributed diverse set of information and results
  - Biomarkers
  - Proteomics
  - Biospecimens
  - Various technologies and data products (image, micro-satellite, ...)
  - Study Management



\* EDRN was highlighted for its informatics in a recent NCI-FDA-AACR report on biomarker research

# Biomarker Informatics Architecture

EDRN Scientific Knowledge Portal

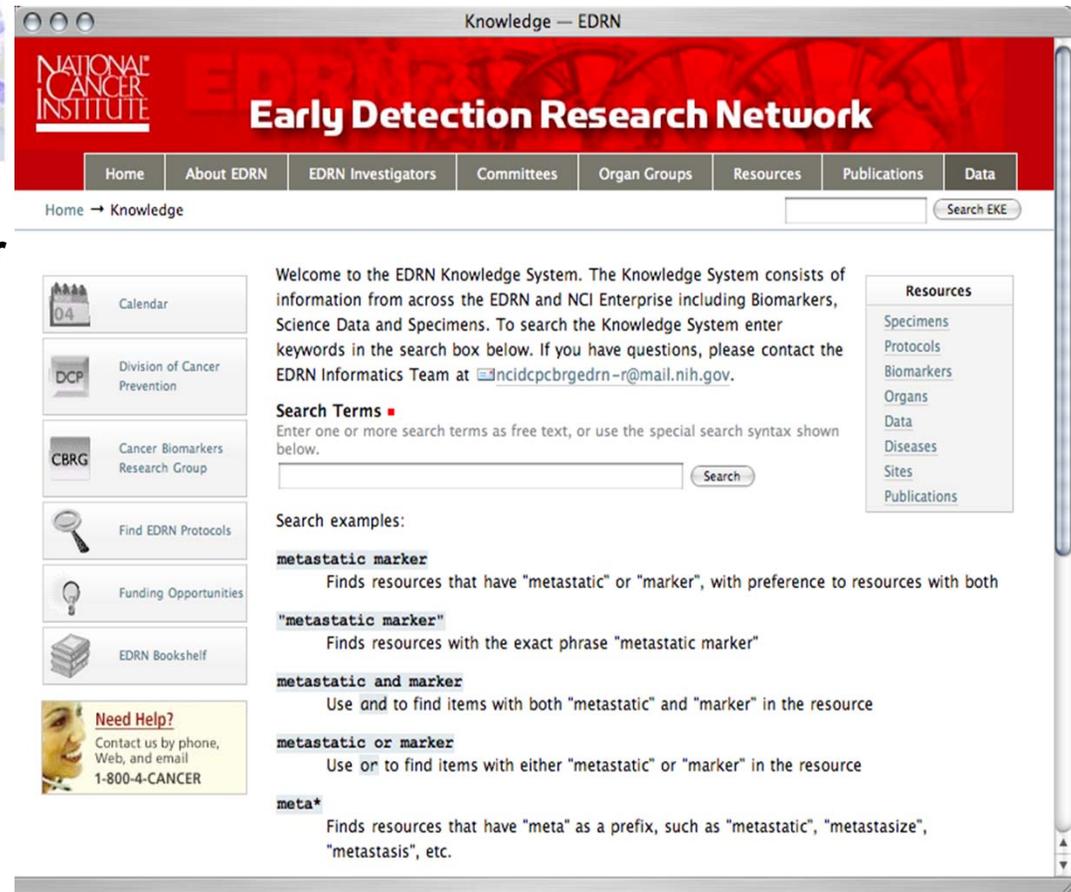


# Integrated EDRN Knowledge Portal

## EDRN Knowledge Environment



- Highly scalable scientific information portal for biomarker research
- Adapts to a dynamic set of “object types” (mass spec, immunohistochemistry, etc) based on the biomarker information model
- Provides access to distributed repositories of information
- Provides “google-like” search of the object types



Knowledge — EDRN

NATIONAL CANCER INSTITUTE

## Early Detection Research Network

Home About EDRN EDRN Investigators Committees Organ Groups Resources Publications Data

Home → Knowledge  Search EKE

Welcome to the EDRN Knowledge System. The Knowledge System consists of information from across the EDRN and NCI Enterprise including Biomarkers, Science Data and Specimens. To search the Knowledge System enter keywords in the search box below. If you have questions, please contact the EDRN Informatics Team at [ncidcpcbrgedrn-r@mail.nih.gov](mailto:ncidcpcbrgedrn-r@mail.nih.gov).

**Search Terms** ■  
Enter one or more search terms as free text, or use the special search syntax shown below.

Search

**Search examples:**

**metastatic marker**  
Finds resources that have "metastatic" or "marker", with preference to resources with both

**"metastatic marker"**  
Finds resources with the exact phrase "metastatic marker"

**metastatic and marker**  
Use **and** to find items with both "metastatic" and "marker" in the resource

**metastatic or marker**  
Use **or** to find items with either "metastatic" or "marker" in the resource

**meta\***  
Finds resources that have "meta" as a prefix, such as "metastatic", "metastasize", "metastasis", etc.

**Resources**

- [Specimens](#)
- [Protocols](#)
- [Biomarkers](#)
- [Organs](#)
- [Data](#)
- [Diseases](#)
- [Sites](#)
- [Publications](#)

Calendar

DCP Division of Cancer Prevention

CBRG Cancer Biomarkers Research Group

Find EDRN Protocols

Funding Opportunities

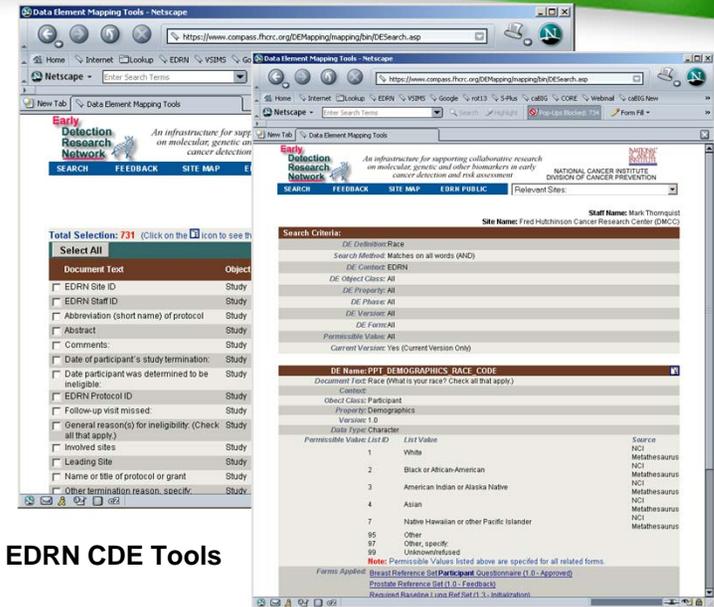
EDRN Bookshelf

**Need Help?**  
Contact us by phone, Web, and email  
1-800-4-CANCER

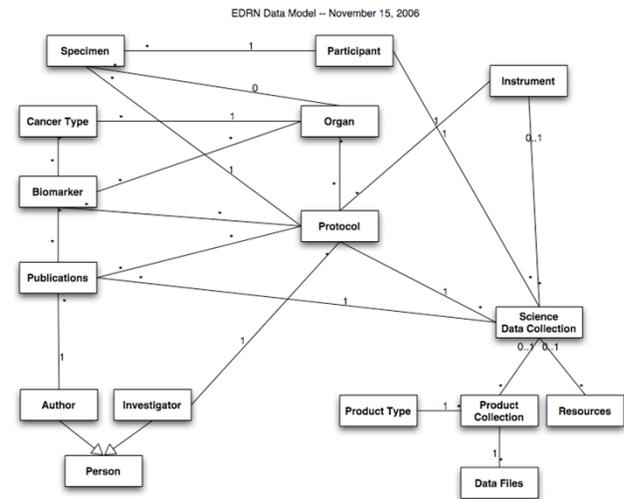
# Informatics Standards for Data Capture and Curation

- EDRN has developed a High level *ontology* model for biomarker research which provides standards for the capture of biomarker information across the enterprise
- Specific models are derived from this high level model
  - Model of biospecimens
  - Model for each class of science data
- EDRN is specifically focusing on a granular model for annotating biomarkers and their studies
- EDRN has a set of EDRN CDEs which is used to provide standard data elements and values for the capture and exchange of data

EDRN's underlying biomarker information model needs to be flexible in order to manage a variety of different types of science data captured by different experiments at different phases in studying biomarkers...



EDRN CDE Tools

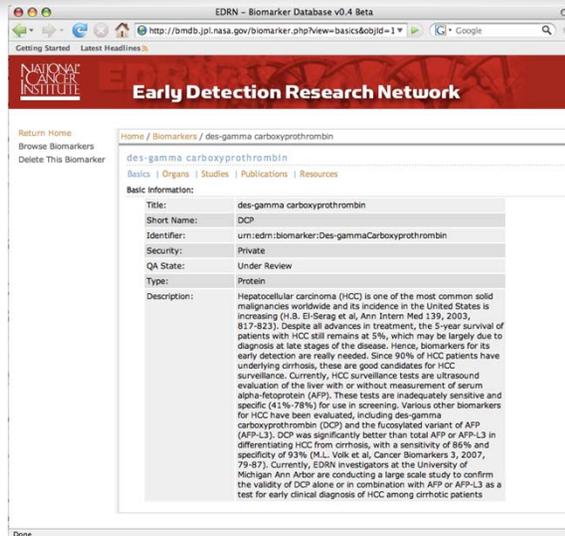


EDRN Biomarker Ontology Model

# Biomarker Data Management

- EDRN provides an infrastructure to capture, curate and manage biomarker information including:
  - Biomarker Annotations (Biomarker Database)
  - Study Design (VSIMS)
  - Protocols (eSIS)
  - Study Results (Science Data Warehouse)

- All databases are designed using EDRN Common Data Elements in order to link EDRN into an integrated enterprise



**EDRN - Biomarker Database v0.4 Beta**

Getting Started Latest Headlines

**Early Detection Research Network**

Return Home  
Browse Biomarkers  
Delete This Biomarker

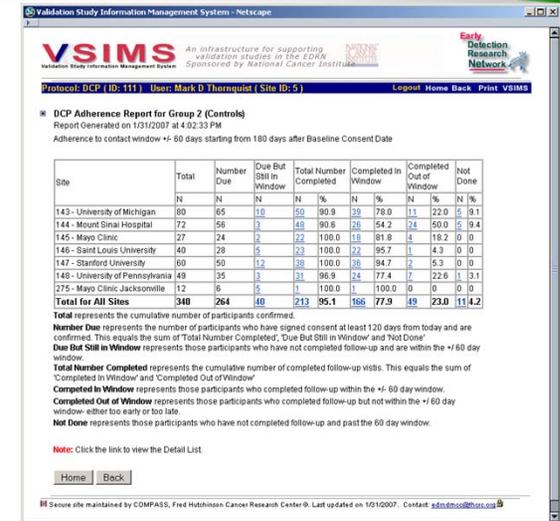
Home / Biomarkers / des-gamma carboxyprothrombin

Basics | Organs | Studies | Publications | Resources

**Basic Information:**

Title: des-gamma carboxyprothrombin  
 Short Name: DCP  
 Identifier: urn:edrn:biomarker:Des-gammaCarboxyprothrombin  
 Security: Private  
 QA State: Under Review  
 Type: Protein  
 Description: Hepatocellular carcinoma (HCC) is one of the most common solid malignancies worldwide and its incidence in the United States is increasing (H.B. El-Serag et al, Ann Intern Med 139, 2003, 817-823). Despite all advances in treatment, the 5-year survival of patients with HCC still remains at 5%, which may be largely due to diagnosis at late stages of the disease. Hence, biomarkers for its early detection are really needed. Since 90% of HCC patients have underlying cirrhosis, these are good candidates for HCC surveillance. Currently, HCC surveillance tests are ultrasound evaluation of the liver with or without measurement of serum alpha-fetoprotein (AFP). These tests are inadequately sensitive and specific (41-78%) for use in screening. Various other biomarkers for HCC have been evaluated, including des-gamma carboxyprothrombin (DCP) and the fucosylated variant of AFP (AFP-L3). DCP was significantly better than total AFP or AFP-L3 in differentiating HCC from cirrhosis, with a sensitivity of 86% and specificity of 93% (M.L. Volk et al, Cancer Biomarkers 3, 2007, 79-87). Currently, EDRN investigators at the University of Michigan Ann Arbor are conducting a large scale study to confirm the validity of DCP alone or in combination with AFP or AFP-L3 as a test for early clinical diagnosis of HCC among cirrhotic patients

Biomarker Database



**Validation Study Information Management System - Netscape**

**VSIMS** An infrastructure for supporting validation studies in the EDRN  
Sponsored by National Cancer Institute

Protocol: DCP (ID: 111) User: Mark D Thorquist (Site ID: 5) Logout Home Back Print VSIMS

**DCP Adherence Report for Group 2 (Controls)**  
 Report Generated on 1/31/2007 at 4:02:33 PM  
 Adherence to contact window +/- 60 days starting from 180 days after Baseline Consent Date

Site	Total		Number Due		Due But Still in Window		Total Number Completed		Completed in Window		Completed Out of Window		Not Done	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
143 - University of Michigan	60	95	18	30	50	90.0	38	78.0	11	22.0	5	10.0	0	0.0
144 - Mount Sinai Hospital	72	56	3	4.2	49	68.0	26	54.2	24	60.0	5	12.5	0	0.0
145 - Mayo Clinic	27	24	2	7.4	22	100.0	18	81.8	4	18.2	0	0.0	0	0.0
146 - Saint Louis University	40	28	5	12.5	10.0	22	95.7	1	4.3	0	0.0	0	0.0	
147 - Stanford University	60	50	12	38	100.0	36	94.7	2	5.3	0	0.0	0	0.0	
148 - University of Pennsylvania	49	35	3	31	96.9	24	77.4	7	22.6	1	3.1	0	0.0	
175 - Mayo Clinic Jacksonville	12	8	5	11	100.0	1	100.0	0	0	0	0.0	0	0.0	
<b>Total for All Sites</b>	<b>340</b>	<b>264</b>	<b>40</b>	<b>213</b>	<b>95.1</b>	<b>166</b>	<b>77.9</b>	<b>49</b>	<b>23.0</b>	<b>11</b>	<b>4.2</b>	<b>0</b>	<b>0.0</b>	

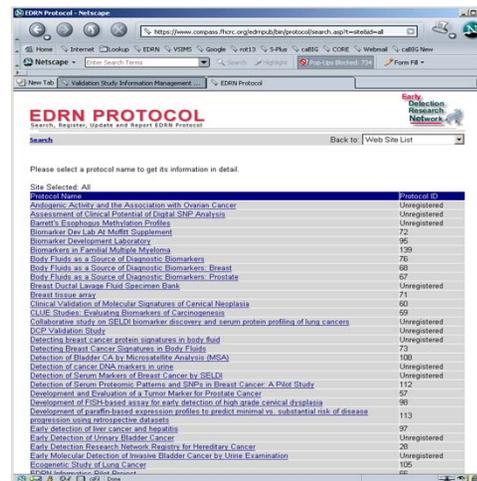
Total represents the cumulative number of participants confirmed.  
**Number Due** represents the number of participants who have signed consent at least 120 days from today and are confirmed. This equals the sum of "Total Number Completed", "Due But Still in Window" and "Not Done".  
**Due But Still in Window** represents those participants who have not completed follow-up and are within the +/- 60 day window.  
**Total Number Completed** represents the cumulative number of completed follow-up visits. This equals the sum of "Completed in Window" and "Completed Out of Window".  
**Completed in Window** represents those participants who completed follow-up within the +/- 60 day window.  
**Completed Out of Window** represents those participants who completed follow-up but not within the +/- 60 day window - either too early or too late.  
**Not Done** represents those participants who have not completed follow-up and past the 60 day window.

Note: Click the link to view the Detail List.

Home Back

Source site maintained by COMPASS, Fred Hutchinson Cancer Research Center. Last updated on 1/31/2007. Contact: edrn@comp001.husup.edu

Validation Study Database



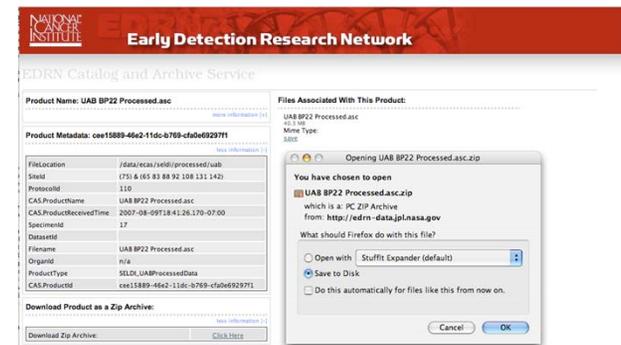
**EDRN Protocol - Netscape**

Search, Register, Update, and Report EDRN Protocol

Please select a protocol name to get its information in detail.

Protocol Name	Protocol ID
Androgenic Activity and the Association with Ovarian Cancer	Unregistered
Assessment of Clinical Utility of Digital SNP Analysis	Unregistered
Breast's Epigenome Methylation Profiles	Unregistered
Biomarker Dev Lab At Mount Sion	72
Biomarker Development Laboratory	96
Biomarkers in Familial Multiple Myeloma	129
Body Fluids as a Source of Diagnostic Biomarkers	76
Body Fluids as a Source of Diagnostic Biomarkers - Breast	68
Body Fluids as a Source of Diagnostic Biomarkers - Prostate	67
Breast Digital Language Fluid Specimen Bank	Unregistered
Breast Tissue Array	71
Clinical Validation of Molecular Signatures of Cervical Neoplasia	60
CLUE Studies - Evaluating Biomarkers of Carcinogenesis	Unregistered
Collaborative study on SELDI biomarker discovery and serum protein profiling of lung cancers	59
DCP Validation Study	Unregistered
Detecting breast cancer protein signatures in body fluid	Unregistered
Detecting Breast Cancer Signatures in Body Fluids	73
Detection of Cancer DNA Methylation in urine	103
Detection of Genes Markers of Breast Cancer by SELDI	Unregistered
Detection of Serum Proteomic Patterns and EDRN in Breast Cancer - A Pilot Study	112
Development and Evaluation of a Tumor Marker for Prostate Cancer	57
Development of CNA based assay for early detection of high grade cervical dysplasia	98
Development of paraffin-based expression profiles to predict minimal vs. substantial risk of disease progression using microarray datasets	113
Early detection of liver cancer and hepatitis	87
Early Detection of Urinary Bladder Cancer	Unregistered
Early Detection Research Network Registry for Hereditary Cancer	26
Early Molecular Detection of Invasive Bladder Cancer by Urine Examination	100
Economic Study of Lung Cancer	102
EDRN Information Data Element	01

eSIS Protocol Database



**Early Detection Research Network**

EDRN Catalog and Archive Service

Product Name: UAB BP22 Processed.asc

Files Associated With This Product:  
 UAB BP22 Processed.asc  
 File Type: ASCII

Product Metadata: ceel5889-46e2-11d6-b769-cfa069297f1

File Location: /data/ncsa/vtd01/processed/uab/...  
 Size: (751 645 83 84 92 108 131 142)  
 Protocol ID: 110  
 CAS Product Name: UAB BP22 Processed.asc  
 CAS Product Creation Time: 2007-08-09T18:41:26.170-07:00  
 Specimen ID: 17  
 Dataset:  
 Filename: UAB BP22 Processed.asc  
 Original: n/a  
 Product Type: SELDI\_UABProcessedData  
 CAS Product ID: ceel5889-46e2-11d6-b769-cfa069297f1

Download Product as a Zip Archive:  
 Download Zip Archive [Click Here]

Opening UAB BP22 Processed.asc.zip  
 You have chosen to open UAB BP22 Processed.asc.zip which is a PC ZIP Archive from http://edrn-data.jpl.nasa.gov  
 What should Firefox do with this file?  
 Open with Stuffed Expander (default)  
 Save to Disk  
 Do this automatically for files like this from now on.

Science Data Warehouse of EDRN raw and processed science data

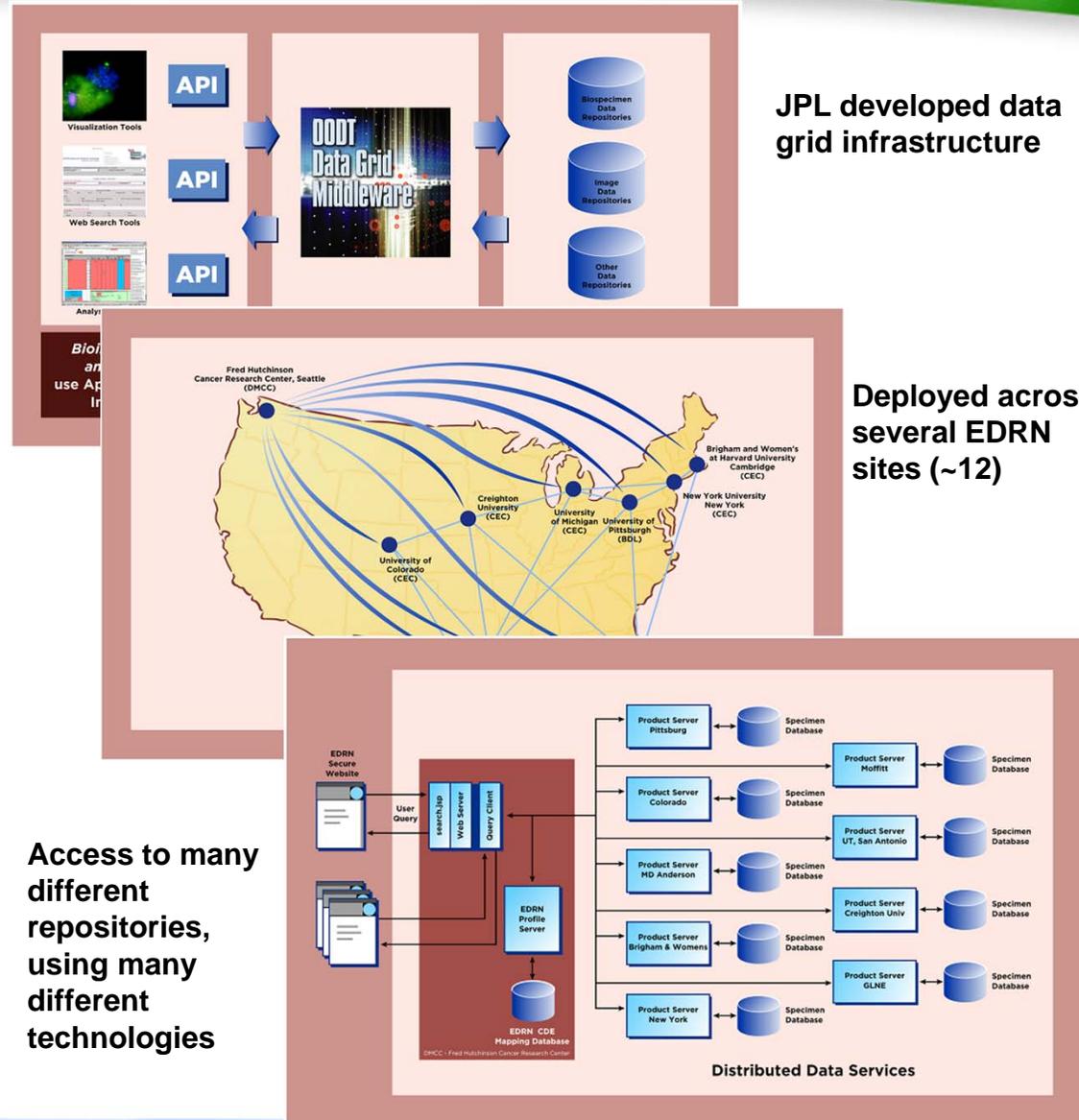
# Access to Specimen Information

- **EDRN has architected, ERNE, the EDRN Resource Network Exchange which provide access to specimens**

- Information about available specimens remains managed at cancer centers
- Access is built using JPL open source software for access to distributed repositories, the same software used to manage access to planetary, earth and astrophysics repositories

- **EDRN has been successful at integrating a diverse community of institutions and technologies**

- Many, many different technologies have been integrated
- EDRN has succeeded in working within different local institutional policies (IRB, Security, etc)
- EDRN and caBIG have worked together to demonstrate how caBIG tools can plug into EDRN (e.g., caTissue)



# Searching for Biomarker Information

Focus on access to science data using general and organ-specific tools

- Google-like Searching
- “Map-based” Searching for organ-centric information

SELDI Validation Study — EDNR Demonstration Site  
<http://edrndemo.org/bmdb/protocols/seldi-validation-study/>

**NATIONAL CANCER INSTITUTE**  
**Early Detection Research Network**

Home | About EDNR | EDNR Investigators | Committees | Organ Groups | Resources | Publications

Home → Biomarker Database → Protocols → SELDI Validation Study

**Biomarker Study Detail**  
**SELDI Validation Study** (see related science) → BioDB: home | **Related Resources**  
 data or specimens | Prostate Cancer Foundation

A Comprehensive Program for the Validation of Prostate Cancer Early Detection with Novel Protein Identification Techniques -- is divided into three phases. The goal of Phase I was to assess the reproducibility and portability of Surface-Enhanced Laser Desorption and Ionization time-of-flight mass spectrometry (SELDI-TOF-MS) using protein profiles generated from serum. Phase I was recently successfully completed at six institutions using a single source of pooled sera. 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.

Prostate Urethra

From biomarker research SELDI Protein Profiles.

Phase	2
Sensitivity Range	36.9 → 46.9 No comment
Specificity Range	46.9 → 46.9 Effective
Positive Predictive Value Range	12.9 → 23.4 Expected range
Negative Predictive Value Range	-12.9 → -9.2 Narrow range
Assays Used	serum protein profile
Technologies Used	SELDI-TOF-MS, IMAC Proteinchip

**Need Help?**  
 Contact us by phone, Web, and email  
 1-800-4-CANCER

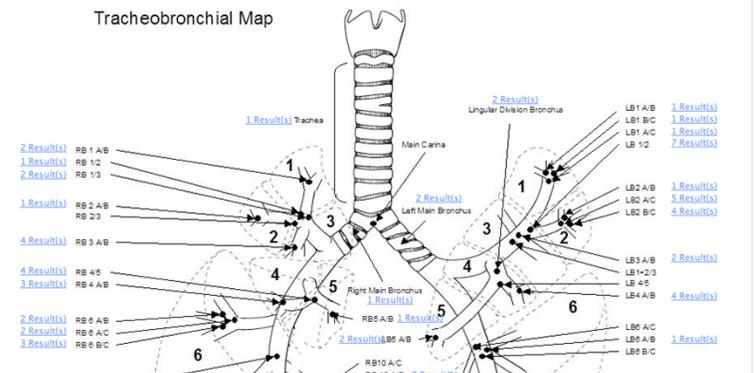
**NATIONAL CANCER INSTITUTE**  
**Early Detection Research Network**

Home | About EDNR | EDNR Investigators | Committees | Organ Groups | Resources | Publications

Home

Biomarker Search: Lung Status: Processed 140 / 426 Profiles [Cancel](#)

**Tracheobronchial Map**



1 Result(s) Trachea  
 2 Result(s) Lingular Division Bronchus  
 1 Result(s) LB1 A/B  
 1 Result(s) LB1 B/C  
 1 Result(s) LB1 A/C  
 7 Result(s) LB1 C  
 1 Result(s) LB2 A/B  
 1 Result(s) LB2 A/C  
 2 Result(s) LB2 B/C  
 2 Result(s) LB3 A/B  
 2 Result(s) LB1+2/3  
 1 Result(s) LB4 A/B  
 4 Result(s) LB4 A/C  
 1 Result(s) LB4 B/C  
 1 Result(s) LB5 A/B  
 1 Result(s) LB5 A/C  
 1 Result(s) LB5 B/C  
 2 Result(s) RB10 A/C  
 2 Result(s) RB10 B/C  
 2 Result(s) RB10 A/B

1 Result(s) RB1 A/B  
 1 Result(s) RB1 C  
 2 Result(s) RB1/3  
 1 Result(s) RB2 A/B  
 1 Result(s) RB2 C  
 4 Result(s) RB3 A/B  
 4 Result(s) RB4 A/B  
 3 Result(s) RB4 A/C  
 2 Result(s) RB4 B/C  
 2 Result(s) RB5 A/B  
 1 Result(s) RB5 A/C  
 2 Result(s) RB5 B/C  
 2 Result(s) RB10 A/B

Submit

**NATIONAL CANCER INSTITUTE**  
**Early Detection Research Network**

Home | About EDNR | EDNR Investigators | Committees | Organ Groups | Resources | Publications

Home

Biomarker Search: Lung Link: [Return to Map View](#)

**Search Results**

	40529 RB7-3F.jpg Pathology Score: 940 Participant Id: 6227965
	4171 LU1-19.jpg Pathology Score: 832 Participant Id: 6186616
	41432 LB6-1.jpg Pathology Score: 891 Participant Id: 6155419
	15453 RB7-3F18.jpg Pathology Score: 833 Participant Id: 6341689
	4271 RB1-3F.jpg Pathology Score: 821 Participant Id: 6819142
	4171 RB8-9-3S.jpg Pathology Score: 310 Participant Id: 6097731
	LB6-1F 40243.jpg Pathology Score: 410 Participant Id: 6096040
	40222 LU08-1.JPG Pathology Score: 891 Participant Id: 6058037

Submit

Identifier: b9d4b46-1b2-11d0-aa9a-f30dc0903a6  
 Title: 21432 LB6-1.jpg Specimen Collected Date: 06/20/07  
 Study Participant Id: 6155419 Bronchial Site Code: 8 Pathology Score: 891



# Interactions across NCI

- **caBIG Interactions**

- EDRN and caBIG have worked together to integrate ERNE and caTissue
  - A working pilot exists which will plug in any caTissue site into ERNE
- EDRN's underlying Common Data Elements are built using ISO 11179, a standard that caBIG has embraced for data elements
- EDRN has submitted its CDEs to the caDSR
- EDRN has participated and presented in several caBIG meetings

- **Other interactions**

- EDRN has met with and shared its architecture and success with other groups (e.g., Prostate NBN Pilot )
- EDRN participated in the NCI-FDA-AACR meetings on an infrastructure to support biomarker research providing significant input related to the EDRN informatics architecture
- EDRN participated in the studies on the National Biospecimen Network

# Challenges and Next Steps

- **EDRN has developed a comprehensive informatics infrastructure for Biomarkers Research**
  - It's now at an inflection point as it moves towards an operational model to capture and curate the information from EDRN studies
  - Informatics needs to continue to be agile, evolving to support science research
- **New bioinformatics tools can now be developed on top of the informatics infrastructure which has laid a foundation for access, management and sharing of biomarker information**
- **EDRN plans to continue to pioneer its novel biomarker knowledge system to expand the data and information that is available and to ensure that it is leveraging and developing modern technologies to build the “e-science” infrastructure for biomarker research\***

\* D. Crichton, S. Kelly, C. Mattmann, Q. Xiao, J. S. Hughes, J. Oh, M. Thornquist, D. Johnsey, S. Srivastava, L. Esserman, W. Bigbee. A Distributed Information Services Architecture to Support Biomarker Discovery in Early Detection of Cancer. In Proceedings of the 2nd IEEE International Conference on e-Science and Grid Computing, pp. 44, Amsterdam, the Netherlands, December 4th- 6th, 2006.

# Acknowledgements

**Mark Thornquist and members of Data Management and Coordinator Center at the Fred Hutchinson Cancer Research Center**

**Sudhir Srivastava and Don Johnsey, National Cancer Institute**

**The EDRN informatics advisory group Bill Bigbee, Laura Essermann, Wilbur Franklin, Tony Hollingsworth, Jeffrey Marks**

## **Current sites collaborating on informatics:**

- H. Lee Moffitt Cancer Center
- University of Texas, San Antonio
- Creighton University
- University of Colorado
- University of Pittsburgh
- University of Michigan/Dartmouth University (Great Lakes New England Consortium)
- Brigham and Womens
- MD Anderson
- New York University
- Duke
- Johns Hopkins University
- University of San Francisco
- Fox Chase Cancer Center
- Beth Israel
- Center for Disease Control
- Roswell Park

**NASA Jet Propulsion Laboratory**

# Early Detection Research Network



## Backup Slides

# Phases of Informatics Development



Operational deployment of biomarker databases

- EDRN deploys its knowledge portal for science access
- EDRN deploys its curation infrastructure for biomarker data
- EDRN builds new tools for organ-specific access to data
- EDRN establishes a curator network for data

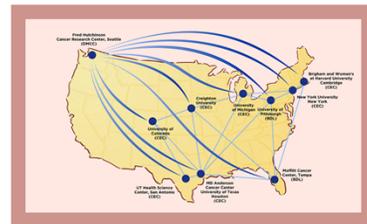
EDRN begins full operational deployment of its tools, focusing on curation of the data



EDRN Portal

- EDRN begins building new tools for study and biomarker management
- ERNE is extended to new sites
- EDRN deploys a public portal

Public Access to EDRN Data



Extend Core Framework

- Extend middleware to link nine EDRN sites together
- Common tools to manage deployed system
- Portal developed to access information
- Validation Study Infrastructure

Major EDRN sites connected

## Initiate System Architecture

- EDRN works with NASA-JPL to Develop Architecture
- Basic Information Model Developed
- Initial set of sites connected

Middleware for linking EDRN distributed data repositories

2002

2004

2006

2008