DRAFT

Office of the Director
National Cancer Institute (NCI)

National Biospecimen Network (NBN) Pilot Use Case and Common Data Element (CDE) Review Meeting

National Institutes of Health (NIH) Campus,
Building 31, Room 8A28, Bethesda, MD

April 15,, 2005
10:00 a.m. – 4:00 p.m. EST

Meeting Summary

List of Participants:

Mark Adams	Booz Allen Hamilton (BAH)
Harsh Bal	BAH
Angelo DeMarzo	Johns Hopkins University (JHU)
Paul Fearn	Memorial Sloan-Kettering Cancer Center (MSKCC)
Ian Fore	National Center for Bioinformatics (NCICB) Contractor
Andrew Hruszkewycz	Organ Systems Branch, NCI
George Komatsoulis	NCICB
Steve O'Krepky	Rose Li and Associates
Mark Rubin	Dana-Farber Cancer Institute (DFCI) (via telephone)
Julie Schneider	Office of Technology and Industrial Relations (OTIR), NCI
Sharon Settnek	Science Applications International Corporation (SAIC)
Bruce Trock	JHU

Purpose:

The purpose of the NBN Pilot Use Case and CDE Review Meeting was four-fold: (1) To draft a first iteration process flow diagram; (2) to draft first iteration high-level use cases; (3) to understand use case priorities to assist in defining project scope and timelines; and (4) to discuss CDEs associated with prioritized use cases. Fore more information, please refer to the meeting agenda (Attachment 1) and roster (Attachment 2).

Presentations and Discussions:

Introduction. Sharon Settnek invited meeting participants to introduce themselves. She then reviewed the meeting agenda (Attachment 1) and the meeting goals and objectives.

Sharon Settnek explained that the group will be exploring the workflow of both the prospective and retrospective studies and that the workflow diagrams will assist in defining the limits of the system. She further explained that the NBN Pilot differs from caTISSUE in that the NBN Pilot requires the capture of

detailed protocol CDEs associated with clinical, specimen, and experiment protocols as well as the capture of biomarker study results. However, the NBN Pilot can benefit from caTISSUE's specimen/segment/sample management and tracking facilities and high-level capture of clinical protocol CDEs (including URIs to existing protocols). Additionally, due to the requirement for capturing biomarker study results, the scope of the Prostate SPORE NBN pilot extends beyond that of caTISSUE. For example, the Prostate SPORE NBN Pilot will capture experiment results and end results for comparative genomic hybridization (CGH) experiments, gene expression studies, immunohistochemistry (IHC), fluorescent in situ hybridization (FISH), and enzyme-linked immunosorbent assay (ELISA). The goal of the group discussion is to understand the workflow of the prospective, retrospective, and biomarker studies and the data required to develop an NBN Pilot that facilitates translational research. Working towards this goal, the purpose of the meeting was to discuss and develop first iteration workflow diagrams and high-level use cases as follows:

- First iteration workflow diagram – Questions that will be addressed in the workflow diagrams include sample distribution to Prostate Cancer Specialized Programs of Research Excellence (SPORE) sites, and biomarker study data collection by specific sites. Sharon Settnek stated that the group will first review the submission use cases. Next, the group will review example searches that will be executed for the different data types that are identified.

- First iteration high-level use cases – After reviewing the workflow, Sharon Settnek stated that the group will determine the types of data that need to be submitted, who will submit the data, and the level of access (i.e., identified, de-identified, anonymous, aggregate) needed for different users. The group will review the definition of a use case and will identify successful processing and error conditions and actions. Next, the group will outline the CDEs associated with specific use cases. Both the caTISSUE and preliminary Inter-Prostate SPORE Biomarker Study (IPBS) CDEs are available as a guide.

Review of Example caTISSUE Process Flow Diagram and Use Case Document. Sharon Settnek reviewed the high-level caTISSUE workflow diagram (page 8 of caTISSUE Core Use Case document) as a model that would be altered for an NBN Pilot workflow. She also reviewed the different user roles and data that would be submitted and processed throughout the workflow. Sharon Settnek explained that the caTISSUE system currently addresses inventory tracking and sample distribution. Julie Schneider asked Sharon Settnek to review definitions such as "site protocol," "segments" and "accession" for the group's benefit. In response to this question, Sharon Settnek reviewed the definition of terms in the starting on page 39 of the caTISSUE Core Use Case document.

Next, Sharon Settnek reviewed the caTISSUE Core Process Model on page 8 of the caTISSUE Core Use Case document. She explained the definitions and user roles in relation to the caTISSUE Core Process Model. Sharon Settnek also mentioned that using URIs in the system to reference clinical protocols may allow the system to access protocol documents/data from external protocol systems.

The group discussed the need to capture information from conventional pathology reports (currently one of the main functions of the cancer Biomedical Informatics Grid's (caBIG's) cancer Text Information Extraction System (caTIES) module, which codes text-based pathology reports). Andrew Hruszkewycz raised a concern that the conventional pathology report may not be sophisticated enough to address the annotation needed for quality interpretation of the data. Angelo DeMarzo added that one of the goals is to develop a system that addresses the annotation required. Bruce Trock mentioned that the clinical annotation used in standard pathology reports is typically quite general. The system developers usually determine what specific data elements are required for the biomarker study. Mark Rubin remarked that in his experience, it is usually fairly straight-forward to develop pathology reports with required data elements prospectively. However, it can be extremely difficult to collect consistent pathology information

across sites retrospectively. Sharon Settnek recommended that, due to NBN Pilot requirements to capture detailed information from the pathology reports for quality assurance/quality control purposes, the data elements required for a study should be defined and submitted via the system. Obtaining this information through the coding of text in pathology reports, although very useful, is usually only 90% accurate. She also mentioned that using structured pathology reports in the system could be detailed as part of the requirements to help improve accuracy on prospective reports. Sharon Settnek then gave a brief overview of related caBIG work already in process as follows.

- A prototype of caTISSUE will be released sometime in June 2005.
- The caTISSUE use cases are currently being reviewed and finalized by the caBIG committee.
- Washington University has created wire frames for the system that help illustrate the system interface.

Sharon Settnek recommended that the group review the caTISSUE wire frame interface to assist in the development of CDEs and a general layout for the Prostate SPORE NBN Pilot system. She would provide the Web address to the group at a later date. Sharon Settnek next presented an early version of the caTISSUE wire frames to the group. Paul Fearn asked how a blood sample would be handled within the caTISSUE structure. Sharon Settnek stated that blood samples would be handled in a manner similar to solid tissue specimens; however, additional data would identify the sample as a blood sample in the system. Sharon Settnek then explained that the workflow and use cases will determine the structure of the system, which will require extensions to caTISSUE. However, if the systems can be designed leveraging the base caTISSUE standard, then data can be shared across the NBN and caBIG. For example, if different institutions capture gene expression data in different formats, but all the institutions standardize format based on the Micro Array Gene Expression (MAGE) standard, data can be shared across the different systems.

Mark Rubin asked whether the Prostate SPORE NBN Pilot system, involving multiple institutions, would allow each institution to determine whether the data would be open for the other Prostate Cancer SPOREs to access, or if it would only be available for that institution. Sharon Settnek responded that such questions will be resolved by defining the workflows and actor and basic flow use cases for the system. She added that the system could utilize the common NCICB security module that is capable of protecting data down to the object and element level.

Data Workflow. Sharon Settnek again reviewed the caTISSUE Core Process Model. She mentioned that, based on the user role in the system and the use cases to be developed, different users will have different levels of access to information and data included in the system. She stated that data tracking is a built-in feature that audits all data entry and changes to data in the system. Sharon Settnek next asked the group to review page 6 of the caTISSUE Core Use Case document which outlines the actors and goals, and defines the roles and data access privileges of specific users. Mark Rubin asked if new user roles could continue to be created after the initial system prototype was developed. Sharon Settnek stated that continual addition of new user roles was possible as long as (1) the user role and the corresponding data access privileges were thoroughly defined and (2) the necessary flexibility was initially built into the system requirements.

Sharon Settnek then presented a draft workflow to the group that she had developed based on preliminary system documentation. She used this document to stimulate discussion within the group.

- Angelo DeMarzo clarified that a major objective of the prospective biomarker study is to correlate IHC results with clinical outcome data.
- Bruce Trock added that several biomarkers included in the IPBS will be measured in serum and plasma. Therefore, the workflow diagram must address both serum and plasma.
- Mark Rubin asked whether detailed protocol information for each biomarker would be available to the system user. Angelo DeMarzo agreed that access to this detailed protocol information is important and should be available in the system.
- Mark Rubin added that the system should capture the protocol data in the database so that if the protocol is varied over the course of the biomarker study, this variation can be tracked. The system should also require adequate flexibility to allow the addition of new biomarkers as well as new institution sites.
- Angelo DeMarzo stated that the group needs to identify CDEs for the protocols (i.e., dilution, clone type, pretreatment, etc.). Mark Rubin added that a "laboratory book module" would be helpful after studies were published. This module would detail various protocols employed. Bruce Trock summarized the protocol discussion by stating that there would be different protocols employed for each biomarker.
- Angelo DeMarzo suggested that pilot studies would be helpful to determine how the system would handle variations in protocols. Bruce Trock mentioned that the IPBS protocol includes a provision that the Central Pathology Core would periodically query individual sites performing the biomarker assays to identify variations in protocols to ensure specimen integrity.

Sharon Settnek led a whiteboard session to diagram the main prospective study workflow. The first iteration of the high-level, prospective study workflow is shown in Attachment 3. Sharon Settnek explained that the best process to determine the system details is as follows:

- First, complete a high-level system workflow.
- Second, determine the key actors (people) involved in each stage of the workflow.
- Third, define the roles of each system actor.
- Fourth, develop use cases for each stage of the workflow.

Sharon Settnek started with the prospective study workflow (Attachment 3), then proceeded with the retrospective study workflow (Attachment 4) and the Biomarker Workflow (for both prospective and retrospective studies; Attachment 5), and finished by drafting prospective and retrospective actor/action diagrams (Attachments 6 and 7, respectively).

Key Research Questions that the Prostate SPORE NBN Pilot Should Support. Time did not allow for a full discussion of the research questions that the Prostate SPORE NBN Pilot should support. Such questions will need to be developed at a later date.

Actors and Data Access Privileges. Sharon Settnek reviewed each of the workflows and helped identify actors at each step in the workflows. The following actors were derived from the group discussion and are also included in each of the workflow diagrams.

Actors for Workflows

Main Workflow – Prospective	Main Workflow – Retrospective	Biomarker Experiments Workflow – Pro- and Retrospective	Prospective Actor/Action	Retrospective Actor/Action
• Protocol Manager Research Coordinator • Site Investigator • Site Pathology Technician • Pathologist	• Study Lead • Resource Director • Principal Investigator (PI) • Research Coordinator • Pathologist • Pathologist Technician • Tissue Microarray (TMA) Technician	• PI • DFCI Lab Technician • Lab Technician • Research Coordinator • Pathologist	• PI • Research Coordinator • Pathologist • Pathology Technician	• Study Leads • Biomarker Requestor Research Coordinator • TMA Technician • Provider Site Research Coordinator

High-Level Use Case Analysis. Sharon Settnek briefly reviewed a sample use case (page 27 of the caTISSUE Core Use Case document). This use case details submitting/editing segment data. The actors, triggers, data elements, pre- and post-conditions, error conditions, related use cases, flow events, etc. are defined for the process. Bruce Trock asked whether the data changes are tracked when a user utilizes this system to edit existing data. Sharon Settnek responded that it would and that there is an auditing use case that illustrates this on page 41 of the caTISSUE Core Use Case document. Julie Schneider added that the auditing trail is also important to comply with Food and Drug Administration (FDA) regulations requiring a full audit trail of data. Sharon Settnek recommended that the group review and use the caTISSUE Core Use Case document as a template for the Prostate SPORE NBN Pilot.

Prioritizing Use Cases. Time did not allow for use case development beyond the preliminary actor/action diagram included in Attachment 6. Use cases will need to be developed in subsequent discussions.

Detailing Prioritized Use Cases. Time did not allow for discussion on detailing prioritized use cases. This topic will need to be addressed at a later date.

Next Steps. Sharon Settnek concluded the meeting by working with the group on determining appropriate next steps. It was suggested that the group complete steps 1-5 listed below via teleconferences. Step 6 would likely require a face-to-face meeting.

1. Review prospective and retrospective high-level workflow and use case definitions
2. Complete high-level use cases for retrospective biomarkers, administrative use cases, and queries
3. Determine CDEs for each use case
4. Prioritize high-level use cases
5. Begin object modeling
6. Conduct caTISSUE and Prostate SPORE NBN Pilot Tissue Resource Exchange (T-REx) gap analysis

Meeting Adjournment. Dr. Schneider thanked participants for their interest and time and officially adjourned the meeting at 4:00 pm.

Attachment 1

NBN Pilot
Use Case and CDE Review Meeting
AGENDA

Friday • April 15, 2005
10:00 AM – 4:00 PM
NIH Campus, Building 31, Room 8A28

Attendees: Mark Adams, Angelo De Marzo, Sue Dubman, Greg Eley, Paul Fearn, Andrew Hruszkewycz, Steve O'Krepky, Mark Rubin, Julie Schneider, Sharon Settnek, John Speakman, Bruce Trock

Agenda

10:00 AM – 10:10 AM **Introductions**

- Introduce Team Members
- Discuss Meeting Goals/Objectives:
  - Draft 1st Iteration Process Flow Diagram
  - Draft 1st Iteration High-Level Use Cases
  - Understand Use Case Priorities to Assist in Defining Project Scope and Timelines
  - Discuss CDEs associated with Prioritized Use Cases

10:10 AM – 10:30 AM **Briefly Review Example caTISSUE Process Flow Diagram and Use Case Document**

10:30 AM – 11:30 AM **Discuss Data Workflow**

- Discuss Types of Data Collected (Protocols, Clinical, Tissue, Pathology, Biomarkers - Genomic/Proteomic, Site/PI, etc.)
- Whiteboard Process Flow Diagram for Data Collection

11:30 AM – 12:00 PM **Discuss Key Questions that the NBN Pilot Should Support**

12:00 PM – 1:00 PM **Lunch**

1:00 PM – 1:30 PM **Discuss Actors and Data Access Privileges**

- Example Actors: Clinicians, Researchers, etc.
- Example Data Access Privileges: Identifiable, De-identifiable, Private, Aggregate Data Views

1:30 PM – 2:30 PM **Begin High-Level Use Case Analysis**

- Example User Cases: Data Submission (Tissue, Clinical, Pathology, Biomarkers – Genomic/Proteomic, Site/PI, etc.), Data Retrieval, Data Analysis

2:30 PM – 2:45 PM **Break**

2:45 PM – 3:15 PM **Prioritize Use Cases**

3:15 PM – 3:45 PM **Briefly Detail Prioritized Use Cases –** *Time Permitting*

- Detail Basic Course of Action
- Discuss CDEs

3:45 PM – 4:00 PM **Discuss Next Steps**

Additional Information

Example Process Flow Diagram and High-Level Use Case: caTISSUE Process Flow Diagram and High-Level Use Case Slide

Example Detailed Use Case Specification: caBIG Use Case Specification

Prostate SPORE NBN pilot Website: prostatenbnpilot.nci.nih.gov

Attachment 2

Office of the Director
National Cancer Institute (NCI)
National Institutes of Health

National Biospecimen Network (NBN) Pilot:
Use Case and CDE Review Meeting

NIH Campus, Building 31, Room 8A28
Bethesda, Maryland

April 15, 2005
10:00 a.m. - 4:00 p.m.

ROSTER

R. Mark Adams, Ph.D.
Project Manager
Cancer Biomedical Informatics Grid
Senior Associate
Booz Allen Hamilton
Eighth Floor
1101 Wootton Parkway
Rockville, MD 20852
(240) 314-5500
adams_mark@bah.com

Harsh Bal
Cancer Biomedical Informatics Grid
Booz Allen Hamilton
6010 Executive Boulevard
Rockville, MD 20852
(301) 998-6929
bal_harshawardhan@bah.com

Angelo DeMarzo, M.D.
Associate Professor of Pathology, Oncology,
and Urology
Department of Pathology
Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University
CRB, Room 153
1650 Orleans Street
Baltimore, MD 21231-1000
(410) 614-3252
(410) 502-9817 Fax
ademarz@jhmi.edu

Paul Fearn
Clinical Database Specialist
Sidney Kimmel Comprehensive Cancer Center
Memorial Sloan-Kettering Cancer Center
1275 York Avenue
New York, NY 10021
(646) 422-4382
fearnp@mskcc.org

Ian Fore, D.Phil.
NCICB Contractor
National Center for Bioinformatics
6116 Exec Blvd, Room 403
Rockville, MD 20852
(301) 496-3355
(301) 480-4222 Fax
forei@mail.nih.gov

Andrew Hruszkewycz, M.D., Ph.D.
Program Director
Organ Systems Branch
Office of Centers, Training, and Resources
National Cancer Institute
National Institutes of Health
Room 7017
6116 Executive Boulevard
Rockville, MD 20852
(301) 496-8528
(301) 402-5319 Fax
hruszkea@mail.nih.gov

George A. Komatsoulis, Ph.D.
Associate Director, Biomedical Informatics Core Infrastructure Group
NCI Center for Bioinformatics (NCICB)
6116 Executive Blvd., Suite 403
Rockville, MD 20852
(301) 451-2881
(301) 480-4222 Fax
komatsog@mail.nih.gov

Steve O'Krepky
Information Technology Consultant
Rose Li & Associates
P.O. Box 34242
Bethesda, MD 20827
(410) 903-3200
(501) 641-9856 Fax
steveok@roseliassociates.com

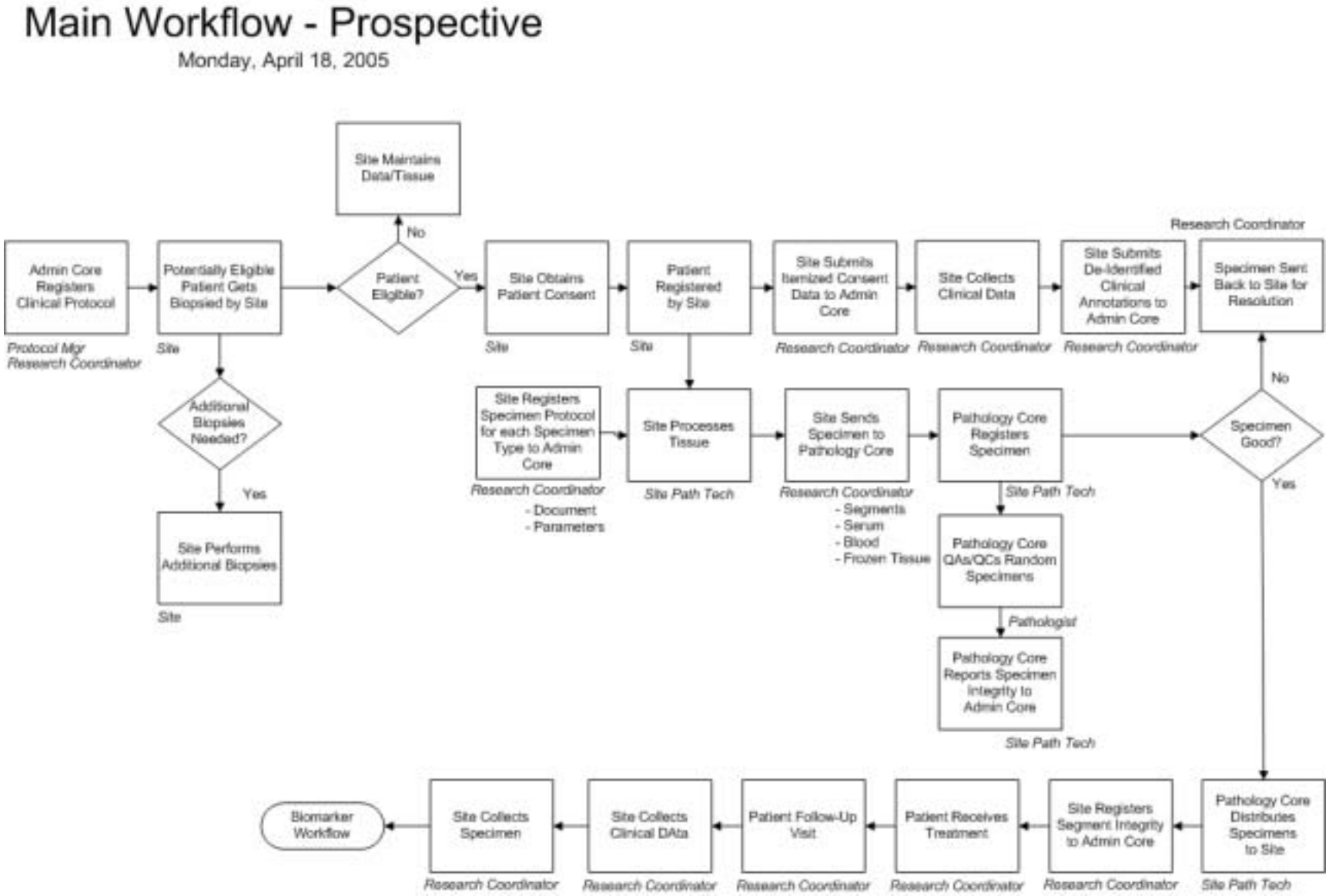
Mark A. Rubin, M.D. (by phone AM only)
Associate Professor of Pathology
Harvard University Medical School
Chief, Urological Pathology
Department of Pathology
Brigham and Women's Hospital
75 Francis Street
Boston, MA 02115
(617) 525-6747
(617) 264-5169 Fax
marubin@partners.org

Julie Schneider, D. Phil.
Technology Program Manager
Office of Technology and Industrial Relations
National Cancer Institute
National Institutes of Health
Building 31, Room 11A-03
31 Center Drive
Bethesda, MD 20892
(301) 496-1550
schneidj@mail.nih.gov

Sharon Settnek, M.B.A.
Program Manager
Cancer Biomedical Informatics Grid
Science Applications International Corporation
Center for Bioinformatics
National Cancer Institute
National Institutes of Health
Suite 403
MSC 8335
6116 Executive Boulevard
Rockville, MD 20852-8335
(301) 451-6345
settneks@mail.nih.gov

Bruce Trock, Ph.D.
Associate Professor
Departments of Urology, Oncology, and Epidemiology
School of Medicine
Johns Hopkins University
Jefferson Building, Room 149
600 North Wolfe Street
Baltimore, MD 21287
(410) 502-2773
(410) 614-8096 Fax
btrock@jhmi.edu

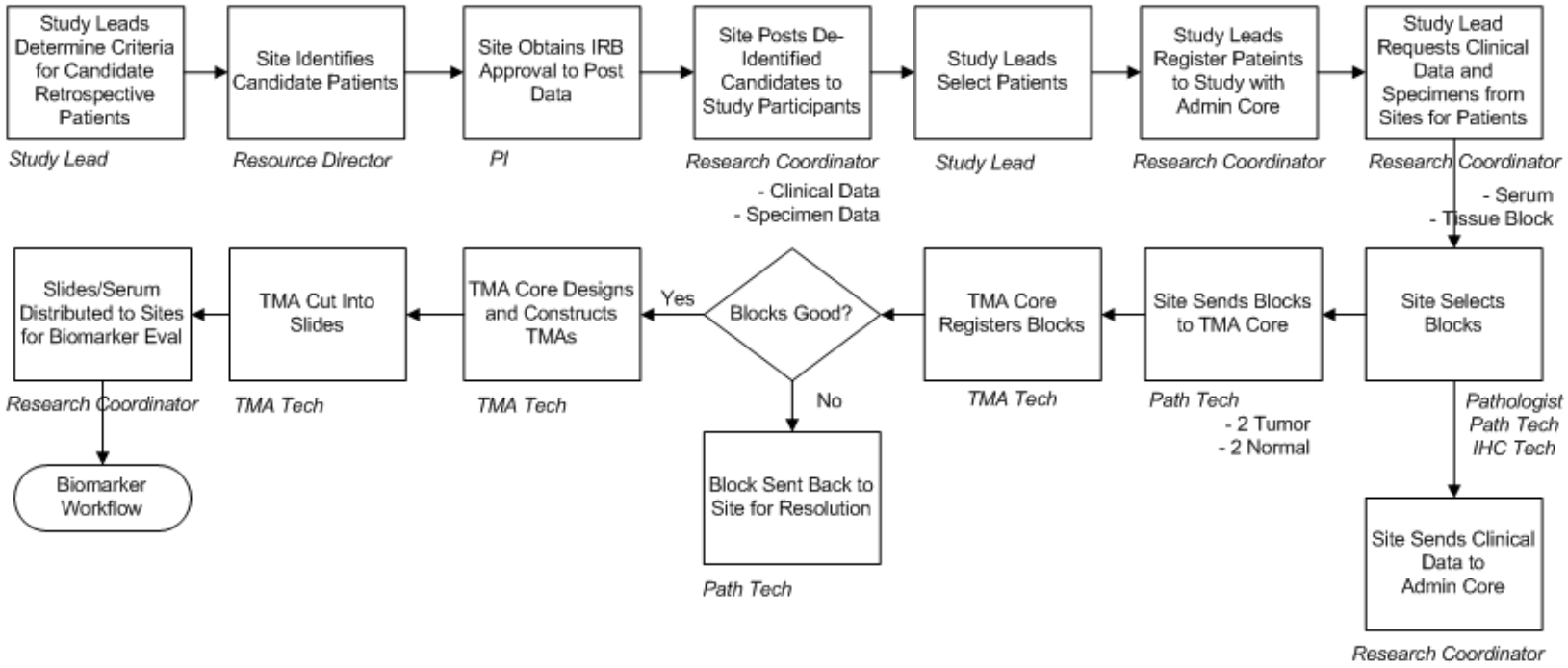
Attachment 3



Attachment 4

Main Workflow - Retrospective

Monday, April 18, 2005

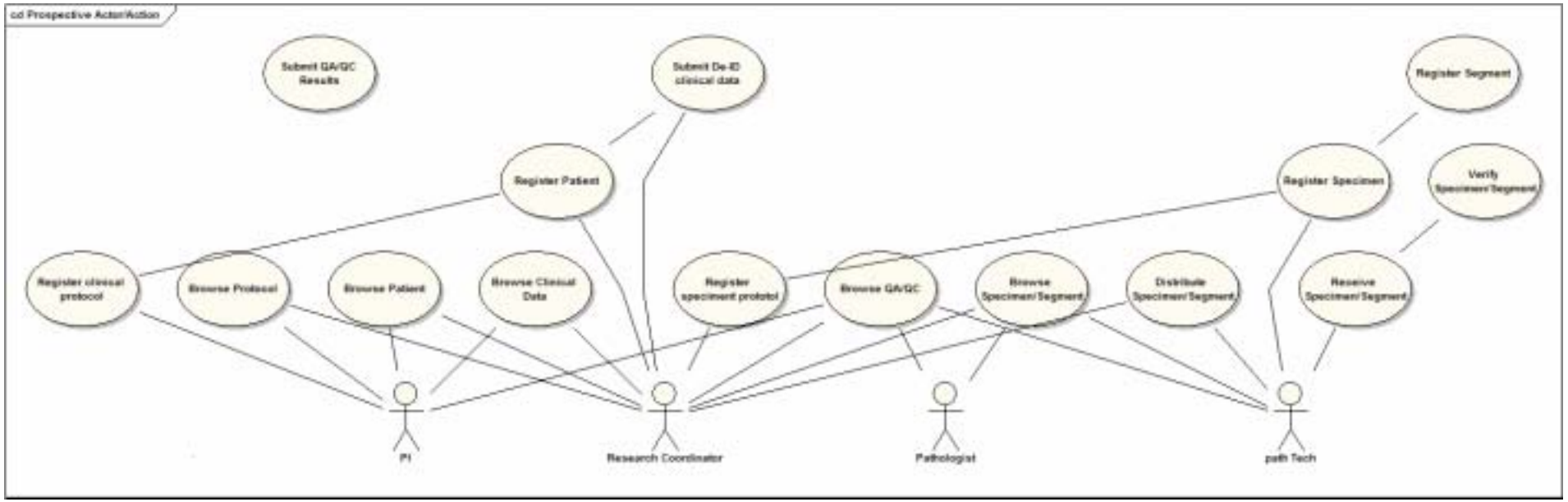


Attachment 5

Biomarker Experiments Workflow - Pro- and Retro- spective

Monday, April 18, 2005

Attachment 6



Attachment 7