Meeting Summary

Proteomic Technologies Reagents Resource Workshop

Sofitel Chicago O'Hare
Chicago, Illinois
December 12-13, 2005

Office of the Director
National Cancer Institute
National Institutes of Health
U.S. Department of Health and Human Services

Table of Contents

Proteomic Technologies Reagents Resource Workshop
December 12-13, 2005

Welcome, Introductions, and Meeting Overview

Gregory Downing, D.O., Ph.D., National Cancer Institute
Leland Hartwell, Ph.D., Fred Hutchinson Cancer Research Center
Brian Haab, Ph.D., Van Andel Research Institute
Amanda Paulovich, M.D., Ph.D., Fred Hutchinson Cancer Research Center

Dr. Gregory Downing welcomed attendees and thanked them for their participation on behalf of Dr. Anna Barker and Dr. Andrew von Eschenbach. He reminded participants that this workshop would concentrate on the development of reagent resources that would help in the identification of proteins and peptides of particular interest to the cancer research community. Previous meetings had identified access to reagents and associated characterization data as major stumbling blocks for biomarker discovery, cancer diagnostics development, and therapeutics monitoring. The goal of the present workshop is to examine the technologies and options thoroughly through engaged discussion and debate so that the National Cancer Institute (NCI) can develop a framework for future efforts aimed at developing reagent resources.

After briefly covering the proposed agenda for the workshop (see Attachment 1), Dr. Downing highlighted supplemental materials. He informed participants that the Report to the National Cancer Advisory Board Working Group on Biomedical Technology has been one of the guiding lights for technology development of molecular diagnostics and genomics. For a perspective on the proteomics initiative at the NCI, including a discussion of a reagents resource, Dr. Downing referred participants to the *Journal of Proteome Research* article.[1] Participants also were informed about new funding initiatives under way at the NCI. One recent Request for Applications (RFA), Advanced Proteomics Platforms and Computational Sciences for the NCI Clinical Proteomic Technologies Initiative, was included in the workshop packet. An additional RFA, Clinical Proteomic Technologies Assessment, and the reissue of the Innovative Molecular Analysis Technologies RFAs also will occur in the near future. These RFAs encompass many opportunities for collaborations between corporate and academic investigators to develop new capture technologies and platforms. Dr. Downing also added that the planning committee for the workshop had emphasized a need for participation by resource developers at the meeting. An open process was established in advance of the meeting where any entity developing antibodies or other capture technologies could submit an abstract and participate in a poster and presentation session. Twenty-five companies from around the world participated in the meeting.

Dr. Downing concluded his remarks by thanking the steering committee and other major contributors for their hard work in making the workshop possible. He then introduced Dr. Leland Hartwell, who thanked all of the participants for attending the workshop and stressed the importance of the problem to be solved. Dr. Hartwell stated that he was motivated by the fact that although a tremendous amount of knowledge has been gained during the past 40 years, minimal progress has been made beyond reducing deaths from a few isolated types of cancer.

[1] Aebersold R, Anderson L, Caprioli R, Druker B, Hartwell L, and Smith R. Perspective: A program to improve protein biomarker discovery for cancer. *J Proteome Res* 2005; 4:1104-9.

However, by applying fundamental knowledge to the cancer problem through the improvement of molecular diagnostics, the cancer scientific community was on the verge of a breakthrough.

Participants were reminded that during the past decade, DNA diagnostics have been applied to differentiate cancers and monitor their responses to therapy by using molecular markers in DNA that accompany the development of different types of cancer. These mutations and polymorphisms have been used to make targeted drugs but seldom have been applied to diagnostics. This happens more easily in the protein realm because DNA diagnostics requires sampling the tumor using invasive procedures. In turn, there are thousands of circulating protein markers in the human body that can be sampled in less invasive ways to aid in the early detection of disease.

Researchers and clinicians have been pleading for better biomarkers, not only from the academic community, but from the pharmaceutical community, which is reluctant to take a drug into clinical trials without biomarkers to follow responsiveness. It is currently relatively easy to develop a catalog of potential candidates for biomarkers in any particular disease state. This can be done by comparing diseased and normal tissues at the transcript, protein, and/or bioinformatics levels. What cannot be done easily, and forms a bottleneck, is measuring the concentration of protein candidates at very low levels in serum and other body fluids using simple tests. This challenge will need to be addressed through the development of new protein capture technologies and platforms.

Workshop Co-Chair, Dr. Brian Haab, gave a brief perspective on the problems and challenges that lie ahead. Dr. Haab stressed the need for highly validated affinity reagents with associated background information. One of the chief advantages of using affinity reagents is the ability to perform high-precision, high-throughput measurements. So far, the only way to do this is with specific affinity reagents like antibodies. The true value of what is being considered at the workshop is not only having a centralized repository of a wide variety of antibodies, but also access to the validation information and publications associated with each specific affinity reagent.

Workshop Co-Chair, Dr. Amanda Paulovich, noted that the next revolution in medicine would be in individualizing medicine, both in treatment and in screening for disease detection. Achieving this goal will require better diagnostic tests, especially protein diagnostic tests. Dr. Paulovich further stated that well-validated affinity reagents like antibodies can have a huge impact in biomarker discovery and validation. Many of the difficulties in discovering biomarkers arise because they occur in incredibly complex samples at very low abundances and, thus, require the use of affinity reagents for detection. Biomarker testing and validation also suffer from the time and investment required to develop affinity reagents needed for the current validation standard, enzyme-linked immunosorbent assay (ELISA). Some antibodies are commercially available, but they often are not annotated or validated extensively for the specific application required. Dr. Paulovich concluded by stating that there is a tremendous opportunity at this workshop to provide feedback and influence future technology development. She encouraged each participant to actively challenge the pre-workshop concepts, assumptions, and strawman proposals.

Session 1: Vision of a Shared Proteomics Reagents Resource

N. Leigh Anderson, Ph.D., Plasma Proteome Institute

In this session, Dr. N. Leigh Anderson discussed the details of his strawman proposal (Proposal A, pp. 37–39 in the workshop program) for creating and operating a pipeline to generate affinity capture reagents against candidate cancer biomarkers. He began with the following assumptions, which he encouraged participants to challenge:

- Antibodies represent the only affinity reagent technology mature enough for high-volume production and use today.
- Monoclonal (as opposed to polyclonal) antibodies should be the focus of the present research effort in order to generate a reproducible and renewable resource.
- Multiple monoclonal antibody clones to a given target are likely to be required for optimization of the properties and performance in different applications, and for performing "sandwich" assays.
- High-throughput screening methods will be needed in each application.
- Monoclonal antibodies will be produced and distributed through commercial channels, and clones will be deposited in a common repository.

This conceptual model was organized around three separate sections, each of which contains a series of tasks:

Section A. Target Selection and Reagent Tracking

- Build and maintain a database of candidate cancer biomarkers that is linked with the existing associated affinity reagents for those candidates.
    - Candidate cancer biomarkers can be obtained from the literature, proteomics discovery efforts, microarray studies, pathway analysis, or systems biology.
    - Candidate cancer biomarker affinity reagents can be obtained from catalogs of existing commercial antibodies, as well as web-based resources.
- Provide a database and user interface to support the management of the whole affinity reagent generation process.
- Develop a process for prioritizing candidates and new affinity reagents.
    - Candidates can be prioritized by number and impact factor of literature citations, requests from the user community, and/or strategic input from the steering committee.
    - The particular application for each candidate also can be prioritized.
    - The target application can be prioritized and placed in the queue within 3 months of startup.

Section B. Reagent Pipeline

- Develop a high-throughput mechanism for generating each antibody type.

- Acquire existing commercial antibodies against high-priority candidates and submit them for screening and validation.
- Establish the vendor pipeline for the production of antigens.
    - Expressed whole target protein antigen
    - Protein epitope signature tag (PrEST) protein domain antigen
    - Synthetic tryptic peptide antigen coupled to a carrier
- Establish the vendor pipeline(s) for making monoclonal antibodies against the targets.
    - Mouse monoclonal antibodies to expressed protein targets
    - Mouse monoclonal antibodies to PrEST protein domains
    - Mouse monoclonal antibodies to tryptic peptides from targets
    - Rat monoclonal antibodies to failures from above
    - Rabbit monoclonal antibodies to failures from above
- Select two or more scale-up production vendors to produce 5 to 50 mg of each antibody selected by the screening and validation process.
- Establish a process to bank selected clones with the ATCC and implement an open intellectual property (IP) policy so that investigators can obtain clones and produce antibodies independently, if desired.

Section C. Reagent Screening and Validation

- Canvas the user community to determine the basic standards required for antibodies to be useful for each application.
- Establish primary screening and validation centers that focus on distinct applications of antibodies (ELISA in plasma, immunohistochemistry, stable isotope standards and capture by anti-peptide antibodies [SISCAPA]) and screen large numbers of clones.
- Establish secondary screening and validation centers to further characterize (binding affinity, specificity, cross-reactivity) the antibodies selected in primary screening.

Session 2: New Opportunities To Advance Proteomic Technologies for Cancer Research

Gregory Downing, D.O., Ph.D., National Cancer Institute
Adam Clark, Ph.D., National Cancer Institute
Mathias Uhlen, Ph.D., KTH Biotechnology

In the first part of this session, Dr. Adam Clark explained the organizational structure and overall research goals for each of the various core programs within the Clinical Proteomic Technologies Initiative for Cancer. This allowed for the present workshop to be placed in context. Dr. Clark also highlighted programs with upcoming RFAs, Requests for Proposals (RFPs), or Notices of Intent of interest to participants. In the second part of the session, Dr. Mathias Uhlen provided an overview of his work with the Swedish Human Proteome Resource (HPR). This research effort is aimed at exploring the human proteome with antibody proteomics by combining high-throughput generation of affinity-purified (monospecific) antibodies with protein profiling using tissue arrays. Data from this work also have been incorporated into the Human Protein Atlas (www.proteinatlas.org), which Dr. Uhlen also discussed.

Clinical Proteomic Technologies Initiative for Cancer. Dr. Clark provided an overview of the Clinical Proteomic Technologies Initiative for Cancer that has been under way at the NCI for the past 2 years. The 5-year program is designed as an integrative approach to developing and enhancing proteomic technology measurement capabilities for cancer. Researchers from multiple disciplines, including clinicians, laboratory scientists, computer scientists, and statisticians are vital to the success of this initiative. The main objectives are to:

- Develop public resources
    - Reagents, biospecimens, reference sets, protocols, algorithms, and databases
- Accelerate protein-related discovery research and applications
- Enhance the knowledge base to support discovery in translational research
    - Identify and validate 1,500 features of interest in cancer biology
    - Characterize and document candidate-based approaches for peptide and protein identification

Reagents and Resources. The present workshop is supporting this program by providing recommendations for developing and making available affinity reagent resources that can help in the identification of proteins and peptides of particular interest to the cancer research community. It is intended to serve the investigator community as a central public resource by offering access to well-characterized proteomic reagents and resources. The key program objectives are:

- Developing standard reagents (antibodies, proteins, peptides, alternative affinity reagents)
- Performing characterization and quality assurance/quality control (QA/QC) protocols
- Providing an interactive resource "catalog" through the cancer Biomedical Informatics Grid (caBIG)
- Expediting acquisition and distribution of reagents and associated data

Clinical Proteomic Technology Assessment Consortia (CPTACs). These multi-disciplinary consortia will be developed to assess current proteomic technologies and ensure that the methods and protocols employed are standardized, reproducible, and comparable within and among various research institutions.

Advanced Proteomic Platforms and Computational Sciences. This program is organized through program coordinating committees made up of individuals at the NCI and principal investigators who are involved in the various programs. Both the CPTACs and the Advanced Proteomic Platforms and Computational Sciences program will require the development of a wide variety of technologies including:

- Sample preparation and labeling
- Sample fractionation
- Mass spectrometry (new methods and platforms)
- Protein capture and microarray
- Data analysis methods
- Microsimulation models (predict peptides seen in mass spectrometry, likelihood of having different proteins or isoforms binding to microarrays)

- Validation techniques

Knowledge about techniques, methods, standards, protocols, and algorithms generated from this initiative will be exchanged freely between these programs to increase the efficiency of the effort. In addition, all data will be entered into a common bioinformatics platform through caBIG, so that researchers will have open access to an integrated, searchable database of reagents and associated characterization and validation data and optimized platforms.

Dr. Clark discussed other NCI resources that have served as a model for this resource initiative. The NCI's Cancer Genome Anatomy Project (CGAP) (http://cgap.nci.nih.gov/) allows the research community to access the gene expression profiles of normal, precancer, and cancer cells. Researchers can access all CGAP data, bioinformatics analysis tools, and biological resources via a web site to find "*in silico*" answers to biological questions. The Mammalian Gene Collection (MGC) web site (http://mgc.nci.nih.gov/) provides a resource for cDNA clones for human, mouse, and rat genes, and links directly with vendors for ordering clones. It is hoped that the MGC site can serve as a model for web-based resources that come out of the proteomics initiative. Finally, a bioinformatics resource is provided by the SAGE[2] Anatomic Viewer (http://cgap.nci.nih.gov/SAGE), which displays gene expression in human normal and malignant tissues by shading each organ in 1 of 10 colors, each representing a different level of gene expression.

The new web site for Clinical Proteomic Technologies for Cancer at http://proteomics.cancer.gov also was highlighted. By the end of December 2005, the web site will host a 45-minute feature presentation outlining the goals of the NCI and the different components of this initiative. RFAs, RFPs, and Notices of Intent also will be listed on the site.

Clinically relevant, cancer-specific peptides and proteins can assist in the early detection and treatment of patients with cancer. The following multidisciplinary programs have been put in place by the NCI to assist in this effort:

- Clinical Proteomic Technologies Initiative (CPTI)
- Mouse Proteomic Technology Consortia
- caBIG
- Biorepository Coordinating Committee (BCC)
- Interagency Oncology Task Force

By the end of the 5-year CPTI, the NCI hopes to have technologies and well-characterized procedures to identify and validate proteins and peptides within the dynamic ranges of putative cancer proteins in human plasma or serum. Moreover, each of these protocols, reagents, resources, and analytical platforms will be made publicly available.

Human Proteome Resource. Dr. Uhlen began his presentation by describing the Swedish HPR, which operates as a nonprofit entity and is hosted by AlbaNova University Center in Stockholm and Uppsala University in Uppsala. The Swedish HPR functions much like a factory in an academic setting. In the production process, upstream bioinformatics information is used to clone

[2] SAGE: serial analysis of gene expression

about 100 new genes each day, which are verified by sequencing. Approximately 100 new human proteins are produced each week and QA checked by incorporating them into tissue microarrays. Finally, the data obtained are deposited in a web-based image database.
Dr. Uhlen's research group has decided to pursue the development of protein fragments as antigens, as opposed to synthetic peptides or full-length proteins. These fragments, called PrESTs, are recombinant proteins that are cloned and expressed from *Escherichia coli*. An effort is made to avoid homology with other human proteins and transmembrane spanning regions, but nothing is known about how they are folded. The advantage with this method is that it allows for the generation of "epitope"-specific antibodies.

The HPR produces what are referred to as monospecific antibodies (msAbs). These are affinity-purified fractions of polyclonal antibodies made by passing the polyclonal antisera over an antigen-labeled column. The disadvantage with this approach is that only small quantities of msAbs are produced and collected. If large quantities of antibodies are required, as in the present strawman proposal, then the use of monoclonal antibodies may be advantageous, because they can be produced in larger quantities, constitute a renewable resource, and possess a defined binding parameter.

At the HPR, 80 percent of the time is spent on assays for the QA of antibodies. Assays employed include:

- Protein assays (ELISA, arrays)
- Adsorption assays
- Western blots
- Immunohistochemistry
- RNA/protein comparisons

The first two methods require both the antigen (in substantial amounts) and the antibody, which could be difficult for a centralized facility as recommended in the strawman proposal. The third method, Western blot, is a more powerful assay, because it gives both the antibody size and specificity. Because one of the ultimate end-user applications is immunohistochemistry, the HPR tends to use Western blot as a QA assay. Another way that antibodies are validated is through the use of RNA/protein comparisons, which for some antibodies gives a good correlation between RNA expression and protein expression. Dr. Uhlen stressed that one needs to have two different antibodies to the same target protein to perform what he believes is the ultimate QA assay. In this approach, one compares the results of one antibody with a second antibody that binds to a different epitope on the same target protein.

Due to the myriad possible applications for antibodies, the HPR has limited the application-specific validations of antibodies to the following four applications:

- Western blot (2 cell lines, 2 tissues, plasma)
- PrEST (protein) array (10 tissues, 10 cell lines)
- Immunohistochemistry
- Confocal (fluorescent) microscopy (3 cell lines)

The Human Protein Atlas. The Human Protein Atlas is an open-access, web-based database created to show the expression and localization of proteins in a large variety of normal human tissues and cancer cells. Data are presented as high-resolution images representing immunohistochemically stained tissue sections.

Dr. Uhlen concluded by informing participants that the HPR is interested in working with academic and commercial providers to validate monoclonal antibodies in a consistent, conventional, or standard way. In the model he envisions, the validation data will be returned to the provider. Approved antibodies will be entered into the Human Protein Atlas and web-linked back to the provider. The HPR currently has the capacity to add approximately 1,000 antibodies to the database each year. The web site address is www.proteinatlas.org, and the database contains approximately 400,000 images (20 terabytes). It receives approximately 30,000 hits per day and is an open-access resource that benefits the scientific research community.

Version 1.1 of the Atlas contains 716 antibodies, approximately half of which are from commercial sources. The goal is to add an additional 1,000 antibodies each year. The data, while not completely quantitative, are provided by pathologists who annotate each image. An antibody information page also is provided for each antibody using data from the Universal Protein Resource (UniProt). The web site also contains a simple application form (with guidelines) to allow users to submit new antibodies for inclusion into the Atlas. After approval by the HPR, collaborators are asked to send in an aliquot of the antibody.

The NCI's First Generation Best Practice Guidelines for Biorepositories. Dr. Downing provided an overview of the Biospecimens and Biorepositories Initiative (http://biospecimens.cancer.gov/) at the NCI on behalf of Dr. Carolyn Compton, who was not able to attend the workshop. There are currently no national biorepository standards, and the NCI's current systems do not enable accurate accounting or analysis of funded biospecimen resources. There are also few guidelines governing biospecimen QC; ethical, legal, and policy issues; and access and retention of biospecimens. Given these unresolved issues, data supporting many areas of science are in doubt.

To address these important issues, the NCI formed the Biorepository Coordinating Committee and the Board of Scientific Advisors Tissue Subcommittee, and launched the National Biospecimen Network Prostate SPORE Pilot. Furthermore, two multisector workshops were convened last summer to finalize white papers from background documents and prior meetings. This has resulted in the first generation of NCI guidelines for biorepositories based on multiyear due diligence, white papers, workshops, and input from additional experts.

The first generation guidelines include recommendations for:

- Common best practices for research biorepositories
- QA and QC programs
- Implementation of enabling informatics systems
- Addressing ethical, legal, and policy issues
- Establishing reporting mechanisms
- Providing administration and management structure

Although these guidelines are voluntary, in the next several years, there should be a self-credentialing or self-assurance practice in place. The full and detailed set of guidelines can be viewed on the NCI Office of Biorepository and Biospecimen Research web site at http://biospecimens.cancer.gov/biorepositories/NCI_First_Generation_Biorepository_Full_Guidelines.pdf.

One participant inquired about future NCI efforts to centralize the informed consent and Health Insurance Portability and Accountability Act approval process across laboratories. As it stands now, when specimens are acquired from different laboratories, the entire approval process and necessary paperwork must be repeated. Dr. Downing stated that one of the harmonization efforts under the National Institutes of Health (NIH) Roadmap deals with this very issue. There are also misperceptions and disparities among institutions as to the strictness of the requirements by different agencies. For example, the U.S. Food and Drug Administration (FDA) requirements for de-identification are very different from those of the NIH, and this has been a major obstacle in utilizing clinical trials specimens and data. The NCI is actively engaged with the FDA in resolving these types of issues.

The issue of cost was raised by another participant, who inquired about the total cost of producing 50 mg of a validated monospecific antibody using the current resources and technologies available at the HPR. Dr. Uhlen responded that the total cost (including overhead) is approximately $5,000 per validated antibody, with half of this cost going to QC. However, it would be more expensive to produce monoclonal antibodies than polyclonal antibodies, because they would require selections and screening steps prior to validation. Another difference between polyclonal and monoclonal antibodies is that one probably would want to validate between 20 and 100 different reagents for the latter, as opposed to the single reagent that is validated when polyclonal antisera is produced.

Another participant stressed that the rate-limiting step in the development of antibodies is producing and having access to high-quality protein antigens. He emphasized the importance of establishing a centralized repository for protein antigens, not only for antibody development, but also to have targets for antibody validation. Monoclonal and polyclonal antibodies are relatively complementary, so to focus on one to the exclusion of the other probably would be a mistake for national and international consortia. Polyclonal antibodies are easier to produce, may be useful for screening, and can be optimal in some applications such as immunohistochemistry. There is also the chance that the clone for a monoclonal antibody can be lost. However, because polyclonal antibodies can be made and stored in such large quantities (10 g or more), they may never need to be renewed.

The assumption that antibodies should be derived from animals was challenged by one of the participants, who suggested that in the 21st century one should be discussing *in vitro* methods. This project will not end with antibodies against 30,000 individual epitopes from individual gene products. It will progress to spliced variants and post-translational modifications, which ultimately may require millions of antibodies. This can be addressed only through the use of *in vitro* methods. Dr. Uhlen agreed that *in vitro* selection methods and combinatorial methods are improving, and that the associated costs are about the same as those for producing a monoclonal antibody. He suggested the possibility of workshop participants recommending that the NCI support and fund more development of *in vitro* methods.

Session 3: Proposed Models for an Antibody Reagent Resource

N. Leigh Anderson, Ph.D., Plasma Proteome Institute
Brian Haab, Ph.D., Van Andel Research Institute
Ed Harlow, Ph.D., Harvard Medical School
Joshua LaBaer, M.D., Ph.D., Harvard Institute of Proteomics

This session included a discussion of many of the major points and emerging questions from earlier sessions. Dr. Joshua LaBaer provided a summary outline of discussion topics in the form of proposed questions, which were first addressed by panel members and then by participants.

Should affinity reagents be made against proteins of interest to those in the clinical research community, the basic science community, or both?

One panel member commented on the tensions that exist between those that favor the clinically oriented approach, in which potential cancer diagnostic targets are identified, and those that see the need for the basic research community to ramp up basic technologies to quickly and inexpensively generate a much larger set of affinity reagents. Even if the smaller clinically oriented set is emphasized, the opportunities and problems that will arise concerning the generation of a much larger collection of affinity reagents should not be forgotten. The focus should not be too narrow or deep on the clinical applications because one could end up with many antibodies that are unsuccessful.

Another panel member suggested that the clinical focus is deserving of the bulk of the present research effort because it is underserved and current cancer biomarkers are poor. Therefore, it is worthy of some real emphasis in this initiative. Concerning the problem of which biomarkers should be selected, it is unclear whether 5 additional years of basic research would improve the knowledge of optimal candidate choice. Even if 1,500 suboptimal candidates are chosen and allowed to move forward into population samples, which is where the real sifting will begin, the hypothesis is that some useful biomarkers will be identified. The source of the reagents is basically an issue of risk reduction, which establishes the argument for the use of monoclonal animal antibodies. Regarding the issue of for whom these affinity reagents are being developed, one panel member indicated that he was in favor of developing resources for the broader cancer research community. The best biomarkers are going to be biologically based. Discovery should be linked to biology, and the information should be exchanged freely between the two scientific communities. It is not a good idea to predetermine the types of applications that the affinity reagents will address.

Should the prioritized list of the 1,500 most desirable targets (biomarkers) be based on literature sources, protein expression levels in cancer, mass spectrometry data, or research investigator-generated proposals?

Several panel members expressed skepticism that a rationale currently exists for making an intelligent decision regarding optimal target choice. Many also were concerned about limiting the total number to 1,500. With such a small number, it is very likely that none will prove to be useful. Further comments noted that a pilot project of this size—complete with annotation and characterization data—would be a legitimate starting point for a resource that is likely to be affected by emergence of newer technologies in the near term.

One panel member felt that the choice of target should be driven by research investigators. They should submit proposals delineating which antibodies should be made for this resource. This is the best way to champion these projects and push them forward. Information resulting from funded research projects, as well as information about the thousands of existing antibodies, must be stored in a centralized database. The real value of this resource will be the information in the database, which will contain validation information for the antibodies with links to the associated biological studies. If companies want to submit antibodies to the database and their antibodies can pass rigorous, standardized validation procedures, then the market value of those reagents should increase. In this way, companies can play a part in deciding which antibodies end up in the resource. Incentives are provided for both companies and the user community. One company representative requested that the NCI simply publish a list of the most desirable targets, however they are determined, to aid vendors in deciding which proteins to produce.

Are animal-based antibodies the best affinity reagents available?

One panel member stressed the importance of having affinity reagents be renewable resources. Therefore, monoclonal antibodies (and their hybridomas) represent the best choice. Even though polyclonal antibodies can perform better in many assays, affinity reagents must be widely available to the community; therefore, a requirement is that they be produced in a reproducible and renewable fashion. As long as this requirement is met, it does not matter if the methodology is DNA, synthesis, or antibody based. The reagents produced must serve the existing user community, which currently has a huge reliance on antibody-based methods and platforms. For this reason, affinity reagents directed against the first round of 1,500 targets primarily should be antibodies.

Other participants noted that it may not be so important to impose the choice of biomarkers to be studied or affinity reagent platforms to be employed. Rather, it may be better to establish a common reagent validation process that provides users with information on antibody interactions with biological materials. This effort would provide the user community with a rating and comparison system for existing commercial antibodies and allow for the development and use of any alternative platform that could be validated. This effort should focus primarily on a centralized mechanism for validating affinity reagents and storing this information in a database. After completing a validation process, the NCI reagent "seal of approval," the reagent would be listed in a publicly available database along with additional information about the reagent. Participants, particularly those representing antibody producers, noted that this approach may provide a financial incentive for the development of validated, high-quality antibodies for which a commercial market already exists.

Should antigens be made by individual affinity reagent producers or by a centralized facility?

One argument for centralizing the production of antigens (as important sources for antibodies and as applied technologies for affinity capture methods) is that they can be made available to all of the research investigators who may need them. Additionally, it may be easier to ensure the QC of antigens if they are all produced in a centralized facility.

What are the mechanisms for NCI government contracting for the production of affinity reagents, and what terms and conditions should apply?

The panel expressed some concern that a government contracting process would result in the same (large) companies ending up with most of the antibody production business. This would result in the same vendors producing the same affinity reagents that currently exist. Instead, to ensure that the right antigens are produced in a timely fashion, government supported efforts should pay only for the validation step for antibodies against antigens of high interest. The resulting validation process and "seal of approval" and the subsequent listing in a centralized database may provide commercial incentive for antibody producers. Another panel member suggested that one may still need to have an additional NCI-contracted production of antibodies for targets that are primarily of interest only to the academic community, because companies tend to focus on the production of antibodies that result in higher revenues. Providing funding for the development of new antibodies with unknown commercial potential is especially important to ensure that small, young companies participate in the present initiative so that new antibodies directed against new targets are produced. The NCI also could consider including the cost of validation in research proposals aimed at producing antibodies against biomarkers of high importance. A participant also encouraged the NCI to put more resources into the development of uniform standards and validation methods. There is a need to establish trusted regional centers to perform validations. One participant pointed out that there are already many existing antigens (recombinant proteins, peptides) available from research institutes and commercial sources. This group of antigens could be used to produce new antibodies and validate existing ones. Many participants were of the opinion that a validation system, building on the Protein Atlas concept, should be available for any currently existing antibodies.

How should antigens and antibodies be validated?

One participant suggested that the NCI issue a challenge to companies to provide customers with the sequence of the antigen (epitope map) so that they could determine where in the protein the antibody is binding. This sequence also would allow for the development of different antibodies that bind to different parts of the antigen. There is also the possibility of using NCI resources to validate antibodies in the academic community using RNAi and cDNA technologies. It would be very difficult for commercial antibody producers to use this type of technology; therefore, it would be better to support decentralized production of antibodies followed by centralized validations. The idea of proposal-based validation was brought up by a participant. This method would ensure that the broad scientific community was engaged in the validation process. Researchers who submit proposals would have a vested interest in understanding the targets and making sure that the antibodies worked. On the other hand, relying on the research community for validation may be difficult. Although the community would be incentivized, it does not have the same QA/QC advantages of a centralized facility operating with standardized validation methods.

Should distribution of the affinity reagents be handled by a centralized repository or by individual vendors using the existing commercial distribution channels?

There was agreement among panel members that there should be a centralized database to store characterization and validation information about affinity reagents but that distribution would be

handled best by existing commercial mechanisms due to their higher efficiencies. One participant remarked that it was a mistake to consider physically distributing antibodies. Instead, antibodies should be distributed via the web by providing sequences that can be synthesized to genes relatively inexpensively and then produced in an investigator's own laboratory for subsequent use.

Will IP considerations limit distribution or make it difficult to build multiplexed arrays containing multiple proprietary affinity reagents on a proprietary platform?

One participant stated that it is unrealistic to expect reagents and platforms to be "free" of IP constraints. It is possible to have IP protection as a technology is moved into the clinic while still being able to distribute reagents and platforms to the research community.

How many affinity reagents are needed per target?

A participant commented that there must be multiple capture reagents that (1) are independent agents, (2) recognize distinct epitopes on an antigen and (3) do not cross-react. The avidity of any of the reagents that will be generated is not sufficient such that a single antibody can be relied on for detection. There must be multiple antibodies directed against each target to accomplish the goals needed for research and clinical reagents.

Which assays should be included in the standard validation of antibodies?

One panel member was of the opinion that the choice of assay for validation should be dictated by the ultimate application for which the antibody will be used. For example, if the antibody is going to be used for immunohistochemistry, then it should be screened for that application in the first round, or it may not be useful in the end. One participant noted the importance of ensuring that the concentration of the correct protein is being measured, especially for low-concentration, high-sensitivity applications. RNAi knockdown experiments can be used to ensure that the signal that is being detected actually decreases when the level of protein is decreased. This procedure should be a standard step in the validation process. Antibodies should not be barred from inclusion in the database solely due to poor performance for a given application, as long as they have been characterized fully.

Session 4: Applications of Affinity Capture Methods in Cancer Research

Amanda Paulovich, M.D., Ph.D., Fred Hutchinson Cancer Research Center
Mathias Uhlen, Ph.D., KTH Biotechnology

In this session, Dr. Paulovich led a discussion aimed at highlighting the most useful and important applications that should be supported by the reagent resource. Application-independent reagent validation methods also were proposed and discussed.

Dr. Paulovich began her presentation by attempting to place the discussion of applications in the context of three guiding principles, which were taken from previous workshop discussions. These principles are listed below:

1. The reagent core should serve the entire scientific community.

2. Antibody validation and annotation and database coordination and curation are the most valuable services that the reagents core can provide.
3. New, inexpensive, and scalable technologies with higher throughput should be developed for the future expansion of this project to target the entire human proteome.

As to the first principle, there is no inherent conflict in serving the research community and facilitating biomarker discovery. The need to understand the underlying biology or the pathophysiological process is now well recognized by the biomarker community. The second principle addresses the components that are missing. There are actually many existing antibodies, and commercial pipelines are in place for making many more. The most valuable service that this core can provide currently is the coordination and maintenance of an open-access database containing characterization and validation information for new and existing affinity reagents. The third principle expresses the idea that the existing expensive and laborious methods and technologies for producing affinity reagents ultimately will fail if the future goal is targeting the entire human proteome. Therefore, while the first round of the initiative may employ monoclonal antibodies, there should be a simultaneous research effort aimed at generating novel, high-throughput technologies that will lower the cost of future affinity reagents.

Recommended Target Applications. In light of these principles, what applications should the reagent resource support? Dr. Paulovich referred to sales data for commercial antibodies taken from four companies from which the following list of the most popular applications results:

- Western blot
- ELISA
- Immunofluorescence/fluorescence activated cell sorting
- Immunoprecipitation
- Immunohistochemistry

She noted that an additional technology, SISCAPA, also should be included in the list. Although this method is not yet in widespread use, it is an important emerging technology. In this technique, an isotopically labeled version of a peptide to be quantified is spiked in at known concentration into a complex biological sample containing an unknown concentration of an endogenous peptide. Antibodies immobilized on beads then are used for enrichment by binding the labeled and unlabeled peptides. A mass spectrometer is used, in effect, as the secondary antibody to provide the relative concentration of the endogenous to the isotopically labeled peptide. This method offers two advantages over the more common ELISA technique. The first advantage is the relaxed stringency requirements for the antibodies due to the specificity provided by the mass spectrometer. The second advantage is that the method is amenable to the rapid generation of assays because a large number of reactions can be performed in multiplex with a very small sample volume.

Several participants suggested that microarray technology be added to the list. One of the reasons that antibody microarrays are not very popular in the user community is that many antibodies do not work well for this application. The cost of preparing these multiplexed arrays also can be prohibitive. Another participant suggested that the application set be modified to focus more specifically on biomarker discovery. One alternative approach would be to develop sets of antibodies that could be used to screen as many antigens as possible in truly high-throughput

proteomic scale. If any hits result, deeper antibodies sets can be developed for the bulleted assays listed above. Missing from the list are applications that involve highly multiplexed use of the antibodies. Another participant suggested the addition of immunoprecipitation and mass spectrometry to the list of validation methods.

Antibody Validation Methods. Dr. Paulovich outlined a set of priority questions covering important characterization parameters for antibodies used in specific applications. These are listed below:

- How sensitive is the antibody (affinity/avidity)?
- Does the antibody bind the target (high accuracy)?
- Does the antibody bind the target specifically (with a good signal-to-noise ratio)?

One method for validating antibodies is to decide for which applications they will be used and then to apply that same application as the validation. A different approach is to validate antibodies in an application-independent manner. The method is not completely independent because some application or an assay still will be used. Some examples of this latter approach are listed in the following table provided by Dr. Uhlen:

Table 1. Application-independent Validation Methods for Antibodies

Method	Description	Examples	Advantages	Disadvantages
Antigen-based	Assays based on the antigen used for immunization	ELISA, protein arrays, Biacore surface plasmon resonance (SPR), antigen adsorption		Need to have antigen used for immunization; prone to artifacts without knowledge of antigen sequence (epitope map)
Target-based	Analysis of native or partially denatured protein from natural sources (such as cell lysates)	Western blot, immunohistochemistry, immunocapture	Does not require the antigen used for immunization	In the absence of the purified target, it is difficult to determine if the antibody is binding to the target; usually relies on denatured targets
RNA-based	Comparison of expression levels at the protein and RNA levels	Transcript profiling, *in situ* hybridizations		Difficult to know if RNA levels correlate with protein levels
Genetics-based	The use of genetic mutants or recombinant constructions to validate the target	Transgenetics, RNAi, green fluorescent protein (GFP)-fusions (subcellular localization)	If protein levels are observed to be increasing with an antibody or decreasing with an RNAi, then one can be relatively certain that the antibody is binding to the target.	GFP-fusions may be subject to artifacts.
DNA-based	Bioinformatics analysis using predictive algorithms (as compared to experimental data)	Signal peptide, transmembrane regions, localization signals		

Table 1. Application-independent Validation Methods for Antibodies (Continued)

Method	Description	Examples	Advantages	Disadvantages
Affinity-based	Determination of the kinetic parameters for the antibody		Complementary to Biacore SPR measurements	
Epitope-based*	Comparison of two or more antibodies directed to different parts of the same target	Antibodies to PrESTs or synthetic peptides		

*Optimal method for validating antibodies based on Dr. Uhlen's work at the HPR.

The problem with the methods listed in Table 1 is that they are not very quantitative. Many of the assays are open to different interpretations, depending on how experiments are performed. Therefore, all of the primary data from each of the assays should be posted to the distributed database web site for comments and review by the research community.

Session 5: Affinity Capture Reagents and QC and Validation

Brian Haab, Ph.D., Van Andel Research Institute

In this session, Dr. Haab led a discussion aimed at determining the best operational model(s) for performing validations of new and existing antibodies. Issues related to providing a cost-effective means of standardized QC measures and standard operating procedures (SOPs) for antibody validations were highlighted.

Antibody QC and Validation Operational Models. Dr. Haab noted from the discussions that there are two evolving operational models for the antibody validation procedure. The first involves a type of "user-validation" whereby an antibody is validated by the end user, employing a particular set of rules, before it is included in the web-accessible database. As use of the antibody continues, more application-specific data can be added to the web site by other users. The second model would employ a centralized facility or group of certified laboratories to perform standardized application-independent validations for new and existing antibodies. A combination of the two models is also a viable option, depending on the application.

For clinical biomarkers, for example, an independent facility that can validate the expression in tumors may be a cost-effective solution. On the other hand, using this type of facility to validate large numbers of antibodies (both good and bad) involving many assays under different conditions may be prohibitively expensive. A virtual model, however, in which the user community provides additional validation data, feedback, and comments via the web-based, distributed database may be very cost effective. Some participants agreed that the cost would be lower for distributed laboratories entering data into the database voluntarily but worried that the overall effectiveness also would be lower. There was agreement among many participants that entering data into a database after validation experiments would not be a high priority for most research laboratories. Additionally, when many different laboratories perform validations without adherence to strict SOPs, there may be less rigor and reproducibility.

Another group of participants stated that although the biological information provided by individual laboratories would no doubt prove invaluable, the QC process should be performed by a centralized facility operating with SOPs. However, there was agreement that those in

individual laboratories with vested interests in the science involving the antibody tend to do a better job finding the appropriate conditions for optimal antibody performance and, hence, a more rigorous validation. In the end, there was widespread support among panel members and participants for the use of a virtual validation model to augment application-independent QC validations that may need to be performed in a centralized facility.

Cost Analysis. Many participants stated that the ultimate choice of an operational validation model, or combination of models, may come down to the issue of cost. One participant remarked that if the amount of antibody that needs to be supplied to the user community is small enough, then one could consider using polyclonal antibodies. While it is true that monoclonal antibodies, once validated, constitute a renewable resource, the use of polyclonal antibodies may not be precluded at low production levels. A careful cost analysis should be performed and tied to the ultimate use of the antibodies. Polyclonal antibodies, if shown to have a clear cost advantage, could be very useful in the target selection or credentialing phase, because they can be made rapidly. Since the ability to detect a differential in a diseased versus a normal state with the polyclonal antibody has been established, the next step would be to make 5 to 10 monoclonal antibodies against different epitopes in the identified target. This may be an effective method for triaging the tremendous amount of effort and cost associated with making affinity reagents of interest.

Session 6: Current Technologies – Overview of Poster Presentations

Mathias Uhlen, Ph.D., KTH Biotechnology

During this session, participants from companies and the research community were given the opportunity to make short presentations (3 to 5 minutes) related to their posters. These presentations were designed to highlight their particular affinity capture technologies and educate the research community on the state of the art. These short presentations are not summarized in this document, but they will be made available on the NCI proteomic technologies web site at: http://proteomics.cancer.gov/meetings_events/symposia_workshops.asp

What Does the Future Hold for Affinity Capture Technologies in Cancer Research?

Patrick Brown, M.D., Ph.D., Stanford University School of Medicine

In this session, Dr. Patrick Brown presented his perspective on what the future could and should hold for affinity capture technologies. He began by stressing the desirability and feasibility of developing inexpensive, fast, microscale, quantitative systems for profiling the entire proteome in tiny biological samples. This topic often has been neglected in cancer research and clinical diagnostics because the focus has been on one-off reagents, offering little economy of scale. The key is to develop a much more valuable and powerful ensemble of thousands of affinity reagents, each with different molecular specificities.

Barriers and Suggested Alternatives for Developing Highly Multiplexed Protein Arrays. Dr. Brown envisions a proteomic parallel of what is possible today for the genome, but so far, several barriers have made this goal difficult to achieve. These barriers include the lack of a comprehensive set of affinity reagents, the prohibitive price of existing commercial antibodies, and the difficultly with unrestricted use of reagents or platforms imposed by IP rights. He cautioned against waiting for the market to remove the obstacles to the development of

multiplexed arrays, citing the lack of customers and the inability of researchers to purchase entire collections of antibodies in tiny aliquots at affordable prices. The NCI and workshop participants were challenged to provide the strategies and impetus to achieve the goal of enabling highly multiplexed, "genome-scale" systematic profiling of complex protein samples.

Two possible alternatives were offered as methods for advancing the goal of large-scale, parallel profiling. The first approach involves engaging and leveraging the labor and expertise of research investigator laboratories with interest in a particular protein or system in concerted effort to develop small-scale protein microarrays. Although this approach may constitute only a modest first step, the data obtained could provide the impetus to drive the technology forward toward large-scale efforts. Incentives provided by the NCI for participant investigators could include: (1) Offering no-cost development and production of a specific affinity reagent for each purified human protein or validated expression/purification system of interest submitted, with a short period of exclusivity of use, (2) requiring grantees to disclose and provide all renewable human protein expression systems, hybridomas, single chain variable antibody fragment clones, etc., to a proteomic reagent clearinghouse, (3) paying investigators for existing human proteins, expression systems, hybridomas, and affinity reagents that meet certain specifications, and (4) providing grant supplements for timely completion and open access to key characterization results.

The second approach would move directly to large-scale microarrays for profiling the entire proteome by developing new strategies and methods to overcome the present scalability problem that limits development. The first challenge would be to find a mechanism for lowering the cost of diverse sets of commercially available affinity reagents such that large-scale multiplexing of assays can begin to occur. A cooperative, bulk-purchasing arrangement with distribution coordinated by the NCI was offered as one suggestion for solving this problem. The second challenge would be to minimize and manage IP rights and issues associated with individual reagents and platforms efficiently such that they do not limit the scale of multiplexing. For incentives for the second approach, it was suggested that the NCI issue challenges or sponsor competitions involving both the academic and corporate research communities so that they would work together to solve the scalability problem.

Another alternative for making large numbers of human proteins available to researchers in small quantities at low cost would be the establishment of a centralized repository. This type of NCI-funded facility could store and distribute affinity reagents, validation data, and published scientific research findings to the academic and corporate research communities.

Attributes of Ideal Affinity Reagents for Multiplexed Microarrays. A list of key attributes for affinity reagents to be used in highly multiplexed microarrays was offered by Dr. Brown. The ideal affinity reagent should have:

- High and well-defined specificity
- High affinity
- High biocompatibility
- The ability to serve as a renewable resource
- The ability to be engineered readily for diverse platforms and applications
- No IP barriers to prevent multiplexing

- An economical price, especially when purchased in large sets

One participant challenged the notion that a large antibody array would function properly. He cited dynamic range differences between low- and high-abundance proteins, and the difficulty in achieving well-defined specificity (each antibody binds only one protein) as the major problems to overcome. Another participant expressed skepticism that it was reasonable to expect IP barriers to disappear as a result of the present NCI-funded initiative given the tremendous investment that has been made in protecting the IP around affinity reagents and platforms. Dr. Brown clarified this point by suggesting that parties with IP rights should work together to ensure that royalties do not scale linearly with the number of proteins being measured on a single microarray. The use of NCI-organized consortia, clearinghouses, and cross-licensing opportunities were given as possible means of overcoming the potential royalty-stacking problem.

Alternatives to Antibodies—Small-Molecule Affinity Reagents. Dr. Brown provided a brief overview of a technology that can be used to produce small-molecule (< 500 daltons) affinity reagents as alternatives to more traditional antibody-based capture technologies. This high-throughput method, DNA display, is basically a method for encoding organic synthesis schemes in DNA and using translation of sequences to guide molecules through synthetic schemes. The technique is a variant of split and pool synthesis chemistry followed by a selection to produce a highly enriched population of relatively high-affinity (nanomolar or greater) small molecules. The technique is very flexible (works with a variety of different chemistries), scalable, inexpensive, and portable.

The difficulty in having truly monospecific small molecules with a high affinity for a given protein was raised by a participant. Dr. Brown admitted that this could be a potential disadvantage, but he also pointed out that this technology could be used to select for things that are more interesting than just binding. Small molecules with undefined, broader specificity also could be very useful for sample prefractionation or as secondary affinity reagents. Another participant echoed concerns about producing individual small molecules with high specificities and affinities but suggested that two of these small molecules could be linked together to yield additive binding energies and the product of their affinities. The current challenge is the development of a truly high-throughput method for this linked small-molecule approach.

Session 7: Target Selection

N. Leigh Anderson, Ph.D., Plasma Proteome Institute

In this session, Dr. Anderson led a discussion of the appropriate mechanisms for selecting an initial set of optimal targets to which affinity reagents would be produced. There was also some discussion about how candidates on the resultant list should be prioritized.

Collection of Candidates. With respect to the design of a systematic process for selecting candidates, one also should consider the null hypothesis (the worst possible case). For instance, what would happen if none of the optimal targets were chosen intelligently and, instead, only random selection was used? By randomly sampling even 5 percent of the total set of human genes (1,100 candidates), much could be learned if broad biomarker patterns are present. Presumably, one can do better using a scientific approach rather than random selection.

Several rational mechanisms for selecting candidates are presented below:

- Literature survey (Polanski and Anderson, 1,200 candidates)
- LaBaer web site (http://hipseq.med.harvard.edu/MEDGENE)
- Swedish HPR data (as it becomes available)
- Extensive analysis of microarray results

Citation frequency analysis of candidate cancer biomarker proteins can be used to rank the order of proteins. These data can be normalized or tracked as a function of time to yield a target selection method based on the publication history of a particular putative cancer biomarker. A second method for selecting targets involves the use of the MedGene database, which is available on Dr. LaBaer's web site. This resource allows users to enter a disease target and retrieve a rank-ordered list of genes associated with that disease based on Medline co-citation analysis. A third method could rely on the existing and emerging data provided by the HPR for candidate selection. All three of these sources potentially can be searched for concordance and pooled to provide a list of at least 1,000 candidates. A fourth method for selecting candidates, extensive analysis of microarray results, was not discussed.

Prioritization of Candidates. One participant suggested that in addition to culling the literature for a list of potential candidates, selection should be driven by organized groups of investigators that want to pursue particular pathways or particular aspects of cancer. These researchers should be able to submit their own proposals, each containing a prioritized list of potential candidates, to the NCI. This would ensure that any new antibody generated to a candidate has an associated research "champion" with a vested interest in seeing the technology advanced. Several conditions and restrictions could be placed on these types of proposals. For example, submissions could be limited to only organized groups of investigators that are studying a particular pathway that already has been implicated in cancer. The group also may have to demonstrate the importance of the resultant antibody to basic science (discovery element). In addition, there could be a requirement for validation in cancer by specifying that one of the collaborators must have access to specimens.

Another approach would be to let users actively submit requests and payments for antibodies via a web site whenever a particular research interest emerges. This would ensure that antibodies for which there are no actively engaged investigators who will be using them in biological studies, are not being produced.

One participant remarked that perhaps it is premature to go through the process of selecting new targets. Instead, all of the existing antibodies should be validated to determine their usefulness by a centralized or "virtual" facility. In parallel, new and existing technologies that are robust enough to increase the throughput should be further developed and proven. At only 1,000 antibodies per year, the pace of affinity reagent development and validation will be exceedingly slow. Using *in vitro* technologies, it is not inconceivable to raise the throughput to 10,000 antibodies per year. The development of new, multiplexed assay platforms also may be important in this effort. The NCI could consider funding investigator laboratories to develop and validate a given number of antibodies in a predetermined period of time in order to determine the most promising and robust technologies.

Session 8: Building a Database

Henning Hermjakob, EMBL-European Bioinformatics Institute

In this session, Mr. Henning Hermjakob led a "requirements gathering" session to determine the immediate and long-range approaches to designing and implementing an antibody database.

Use Cases. Mr. Hermjakob began the discussion by defining 10 use cases, which capture who (actor) does what (interaction) with the system and for what purpose (goal). Mr. Hermjakob asked participants to consider the following use cases:

Use Case 1. Simple lookup

Actor:	Scientist
Interaction:	Lookup
Goal:	Needs to find available antibodies for a favorite protein
Requirements:	Anonymous access, HTML web interface, not interactive

Use Case 2. Feedback

Actor:	Scientist
Interaction:	Data input
Goal:	Needs to enter the antibody validation data into the database, supported by a journal citation
Requirements:	Login, data input, PubMed ID, classification of the experiment that has been done using a controlled vocabulary, ability to make free text comments

Use Case 3. Target request

Actor:	Scientist
Interaction:	Request
Goal:	Scientist did not find a validated antibody for the protein and needs to request an addition to the target list
Requirements:	Login, data input, anonymity of request (may not want others to know of the interest in a certain protein), prioritization algorithm for target requests from many sources (establishes the right weighting for requests from external individuals, NCI grantees, or groups of scientists)

Use Case 4. Curator approval

Actor:	Web site operator/curator
Interaction:	Request
Goal:	Curator needs to be able to approve all submissions and remove obviously wrong or misleading comments and reviews
Requirements:	*Requirements for this use case were not discussed.*

Use Case 5. Curator contribution

Actor:	Curator
Interaction:	Data input
Goal:	Curator needs the ability to enter additional data into the database
Requirements:	Requires curator or curator team, and curation guidelines.

Use Case 6. Large-scale lookup I

Actor:	Research group
Interaction:	Lookup (monthly, with automated ordering whenever the price falls below a certain level)
Goal:	Research group wants a list of all available 4-star antibodies for a list of 500 favorite proteins, also wants the total price for the antibodies and the list updated every month
Requirements:	Automated large-scale read access, anonymous access, secure access

Use Case 7. Large-scale lookup II

Actor:	NCI contract validation laboratory
Interaction:	Lookup
Goal:	NCI contract validation laboratory needs to look up the next batch of antibodies to validate and needs to ensure that it is not duplicating another contract laboratory's efforts
Requirements:	Potential antigen sources, the ability to relay status and to "flag" antibodies that are in the validation pipeline (Laboratory Information Management Systems aspect)

Use Case 8. Large-scale target list

Actor:	Research group
Interaction:	Data input
Goal:	Research group needs to enter 200 new promising targets from a microarray study with associated publication; list will be updated every quarter for the next 2 years
Requirements:	Automated data input from already existing data.

Use Case 9. Large-scale validation

Actor:	Researcher/NCI contract laboratory
Interaction:	Data submission in a defined format
Goal:	Researcher needs to report the results of an antibody array monthly; data includes a technology description, a publication, and 500 data points
Requirements:	Data input in a defined format, same action using different data on a monthly basis, ability to prevent update problems as the data changes

Use Case 10. International collaboration

Actor:	The HPR
Interaction:	Data exchange in a defined format
Goal:	The HPR provides validation information for a few thousand antibodies, updated monthly, with extensive documentation in the Human Protein Atlas. Although stored locally in Sweden, these data should be exchangeable with other databases in other countries, and remotely searchable.
Requirements:	The ability to exchange data with other external databases in an automated manner, synchronization, updates

Data Elements To Be Stored. Targets that are to be stored should be defined with respect to a well-known, publicly accessible, standardized resource, such as the UniProt. Problems will arise

if a researcher submits a gene name that is duplicated or nonstandard, or a clone number from a private laboratory as a potential target. The sequence, accession number, and version of the target are also essential because sequence databases are not static. The species attribution of the protein also should be stored. As already mentioned, the target ranking, associated evidence, and antigen source also should be contained within the database. For antibodies, common attributes (including name, source, price, quantity, mono/polyclonal status, and system) should be stored. Technologies used for validation should be defined clearly within the database. This can be accomplished through technology descriptions with controlled vocabularies and a hierarchical structure. Antibody validations may require the capture of the following information:

- Relevant antigen (sequence) features, e.g., PrESTs
- Technology
- Laboratory
- Score
- Supporting evidence, potentially external, e.g., Western images
- Quantitation
- Technology-dependent parameters
- Many validations for one antibody/antigen pair
- Relationship between antibody, antigen, and technology
- More than one antibody in a validation assay, e.g., sandwich assays

Access Control. It will be necessary to provide read-only access to certain individuals. Write access will be needed by different entities in order to capture anonymous opinion statements and target requests. Internal curators and international collaborators will need write access to the database. Large-scale lookups and complex queries will require automated repetitive read and write access to large data sections. In the current state of technology, automated access to complex data would be provided best through an XML interface and web services (Simple Object Access Protocol style). Whatever the interface, there always will be users who want to download the full database to use and manipulate the data locally.

Data Exchange Format. One possibility would be to incorporate data elements from the antibody database into the XML schema of the Human Proteome Organisation, Proteomic Standards Initiative-Molecular Interaction format (http://psidev.sourceforge.net/mi/rel25/). Additional elements of the controlled vocabulary would need to be added, but much of what is needed for data exchange could be provided by this existing format.

Session 9: Building a Shared Resource and Identifying Action Items

Joshua LaBaer, M.D., Ph.D., Harvard Institute of Proteomics
N. Leigh Anderson, Ph.D., Plasma Proteome Institute
Brian Haab, Ph.D., Van Andel Research Institute

In this session, Dr. LaBaer led the discussion through his summary of preliminary conclusions drawn from previous sessions at the workshop. Areas of agreement and unresolved issues were highlighted with input provided by both workshop chairs and participants. The presentation was organized around individual issues with rationales, advantages, and disadvantages often offered for each conclusion.

Target Selection. Two different strategies were identified for choosing optimal targets (biomarkers). The first involves utilizing targets already under active investigation by the scientific community. The rationale for this "user-requested" approach is the abundance of existing data, suggesting that these proteins are either important to the pathogenesis of cancer or may serve as future biomarkers. One recognized advantage is that a user market already exists for antibodies produced to these targets, which could provide commercial incentive for companies developing affinity reagents. On the other hand, there are already mechanisms in place to generate antibodies to these existing targets, so little additional research capability is likely to be introduced by this approach.

The second strategy involves deliberately choosing targets infrequently studied by the scientific community in an effort to find new clinical biomarkers for detecting and categorizing disease. These targets may include low-abundance proteins in serum, proteins that are extracellular domains, or proteins that are predicted to be soluble, secreted, or on the surface of cells. The rationale for this approach is that there is limited data, suggesting that these targets could be good biomarkers in the future, and duplication of effort could be avoided by not developing additional antibodies directed at existing targets. A potential disadvantage is that it may be difficult to incentivize companies to invest in research and development to produce these new antibodies without a well-developed research or commercial market.

Affinity Reagent Platform Selection. There was widespread support for antibodies as the current affinity reagents of choice. The rationale for this choice was based on the assumption that antibodies represent the most mature technology. The fact that existing companies can produce 500 to 1,500 monoclonal antibodies per year, and that tens of thousands are already commercially available, also were seen as major advantages for antibodies. However, many of these antibodies are known to be poorly characterized and suboptimal for many applications. Although alternative platforms were recognized as needing more development work, several participants challenged this assumption by citing the tremendous potential of high-affinity aptamers to substitute for antibodies in today's microarray applications. There was extensive support among participants for the continued development of affinity reagent technologies that do not rely on antibodies.

The use of monoclonal (as opposed to polyclonal) antibodies was suggested as more desirable for most applications. The rationale for this selection was the fact that although more costly, monoclonal antibodies constitute a valuable renewable reagent because they do not have to be rescreened once validated. The use of less expensive polyclonal antibodies, however, was suggested for initial screening applications used to identify meritorious candidates for further reagent development.

Assay Platform Selection. The need to establish a consistent platform (bead based or microarray) was stressed. This recommendation was based on the idea that a standard platform would greatly simplify the development of high-throughput (proteome scale) reagents.

Antigen Production. There was significant agreement that making antigens is a key bottleneck in the development of new affinity reagents. Access to high-quality antigens is especially important, since they are required for both antibody production and subsequent validation. Both company representatives and the user community expressed preference for a central repository

that would store antigens (and associated data) produced by individual laboratories. The idea of having this centralized facility perform protein expression, microarray fabrication, and even distribution also was advanced, but the logistics of organizing and managing such a facility remained unresolved. Participants also discussed the formation of a partnership between the user community and the NCI in which the NCI would pay for the commercial development of a monoclonal antibody after a user had invested time to produce and characterize a high-quality antigen.

Antibody Production Process. Validation of affinity reagents was seen as the most important step in the production process. Although some companies were open to the idea of partnering with academia to make antibodies to "popular" proteins at low or no cost, the question of how to incentivize companies to produce and validate antibodies to infrequently studied targets was not resolved fully. Several alternatives were suggested, ranging from simply having the NCI pay for either production and/or validation, to providing companies with a list of recommended proteins based on their predicted potential to serve as useful biomarkers. One participant remarked that it remained uncertain whether either of these alternatives would be sufficient to mitigate the financial risks assumed by antibody producers.

Distribution of Antibodies. Two different mechanisms were discussed for the distribution of affinity reagents. The first involves the use of a centralized repository and distribution center. With this model, a standard set of QA/QC validation parameters could be used for all antibodies produced. The logistics of developing high-throughput applications also would be simplified by a centralized distribution center containing all of the required reagents. Various IP issues requiring the use of multiple agreements also may be avoided or streamlined under this model. Possible disadvantages that were discussed included the present lack of infrastructure for "privatized" centralized distribution, duplication of fully functioning commercial distribution channels, and distribution limited to only affinity reagents produced and paid for by a centralized effort.

The second mechanism involves use of the existing commercial distribution center. One concern raised by participants from the user community was that the current catalog price of antibodies makes it economically unfeasible to fabricate high-density antibody microarrays for research purposes. Company representatives explained that they currently do not have the capability to package and distribute the extremely small (more economical) quantities of antibodies needed to spot on microarrays. However, they did express an interest in working with the academic community to provide these small aliquots, as well as making surplus antibodies of low commercial value readily available.

Validation of Antibodies. There was strong support for the development of a centralized database containing characterization data for all present and future antibodies. There was strong support for using those in the research community with a vested interest in developing the appropriate conditions for optimal antibody performance to validate existing commercial antibodies initially. This strategy was seen as preferable to having validation performed by a regulated, centralized facility operating under a regulated set of SOPs. There was some objection from commercial vendors who pointed out that there are particular skills required in the validation process and that the usefulness of a given antibody depends on the particular application. One suggestion offered by a participant was to have commercial vendors work in

close partnership with the academic research community to develop an appropriate standardized validation procedure and SOPs.

For reporting mechanisms, participants favored an interactive, centralized database in which users deposit characterization data and comments for sharing across the entire research community. This type of web-based rating and review system was suggested as the most incentivized, cost-effective, and timely means of providing information on existing antibodies. Possible disadvantages of this approach are the difficulty in ensuring data and review consistency across individual user laboratories and the lack of an existing enforcement mechanism to ensure that data from the user community is deposited in the database. These possible drawbacks led several participants to recommend the use of a centralized facility to handle some functions, including resolving discrepancies in user validations and reviews, monitoring QA/QC as a function of lot number or storage time, and providing official independent certification through routine standardized assays.

The suggested choice of applications that should be validated routinely included Western blot, ELISA, and immunoprecipitation. Several commercial vendors indicated that these tests were already part of their standard validation procedures for all reagents sold. There was a suggestion that applications that foster high-throughput uses also should be validated. The desirability of capturing and including in the centralized database quantitative information for affinity reagents (K_d, on-rate, off-rate) also was discussed. Several participants noted the importance of understanding the actual binding epitope (linear or conformationally dependent) in order to interpret the quantitative characterization.

Wrap-up and Adjournment

Gregory Downing, D.O., Ph.D., National Cancer Institute
Joshua LaBaer, M.D., Ph.D., Harvard Institute of Proteomics
Brian Haab, Ph.D., Van Andel Research Institute

The committee chairs provided closing comments and impressions on the workshop. All were quite enthusiastic and positive about the opportunity and progress made. On the corporate side, several chairs remarked that the companies had expressed an interest in receiving help in choosing optimal targets, screening existing hybridomas, having commercially available antibodies validated, and participating in a centralized database. On the research community side, some users appeared to be frustrated by the lack of available antibodies for particular applications, the absence of reagent validations, and high reagent costs. Several of the chairs recommended that the NCI act as a liaison to help spur the development of a centralized database that would serve both the research community and commercial vendors. It also was suggested that the NCI work with both academia and the commercial sector to develop and impose a set of antibody validation standards. One unresolved issue concerns the logistics and scope of a possible NCI program to fund the production of antibodies in the commercial sector to reduce the financial risks to companies and lower prices to end users. Finally, several participants expressed a desire to see more funded development of alternative (non-antibody-based) affinity reagents in the near future. Many echoed concern about the ultimate limitations of antibodies and suggested the development of more *in vitro* platforms.

Dr. Downing concluded the workshop by thanking all of the participants and informing them that the NCI is fully committed to developing a plan for going forward with specific action items. He highlighted the agency's commitment to developing guidelines or recommendations for antibody characterization and validation standards. The commitment for developing a web-based, accessible, centralized database to store data and user comments obtained under a set of standardized parameters also was stressed. Dr. Downing concluded by stating that although it may be a difficult challenge, the development of new resources and *in vitro* technologies will continue to be high priorities at the NCI.