Clinical Proteomics and Biomarker Discovery in Cancer Research

National Cancer Institute Symposium

National Cancer Institute
National Institutes of Health Campus
Bethesda, Maryland

September 24, 2004

Background

An effective platform for clinical biomarker discovery would revolutionize clinical oncology. Such a platform would enable detection of disease early, when it is curable. It would also provide a means of distinguishing quickly those patients who are responding to a given therapy from those who are not, as well as yielding perhaps the ultimate preventive tool capable of identifying those at risk for cancer and even predicting their response to prevention interventions.

To catalyze the development of an effective platform for clinical biomarker discovery, the National Cancer Institute (NCI) is proposing to develop a strategic scientific initiative that would bring the power of proteomic technologies to bear on the problem of discovering biomarkers for cancer. The premises behind this technology-based initiative are the following: Such biomarkers exist in readily accessible body fluids, panels of such markers will be required to achieve high specificity and sensitivity, current technology is capable of discovering these panels, and current application of this technology can be improved. Since no current technology interrogates more than 1 percent of the proteome at a time, a systematic approach to biomarker discovery requires teams of investigators sharing and aggregating data. Achieving this will involve setting standards, ensuring quality control, and developing an informatics platform capable of aggregating and comparing data across laboratories.

Because of his extensive experience in proteomics, the NCI has enlisted Dr. Lee Hartwell, President and Director of the Fred Hutchinson Cancer Research Center, to assist the NCI in its planning activities in clinical proteomics. To catalyze discussion among researchers, Dr. Hartwell prepared a white paper for a focused clinical biomarker discovery initiative. This white paper includes six components organized as shown in Figure 1 and discussed below:

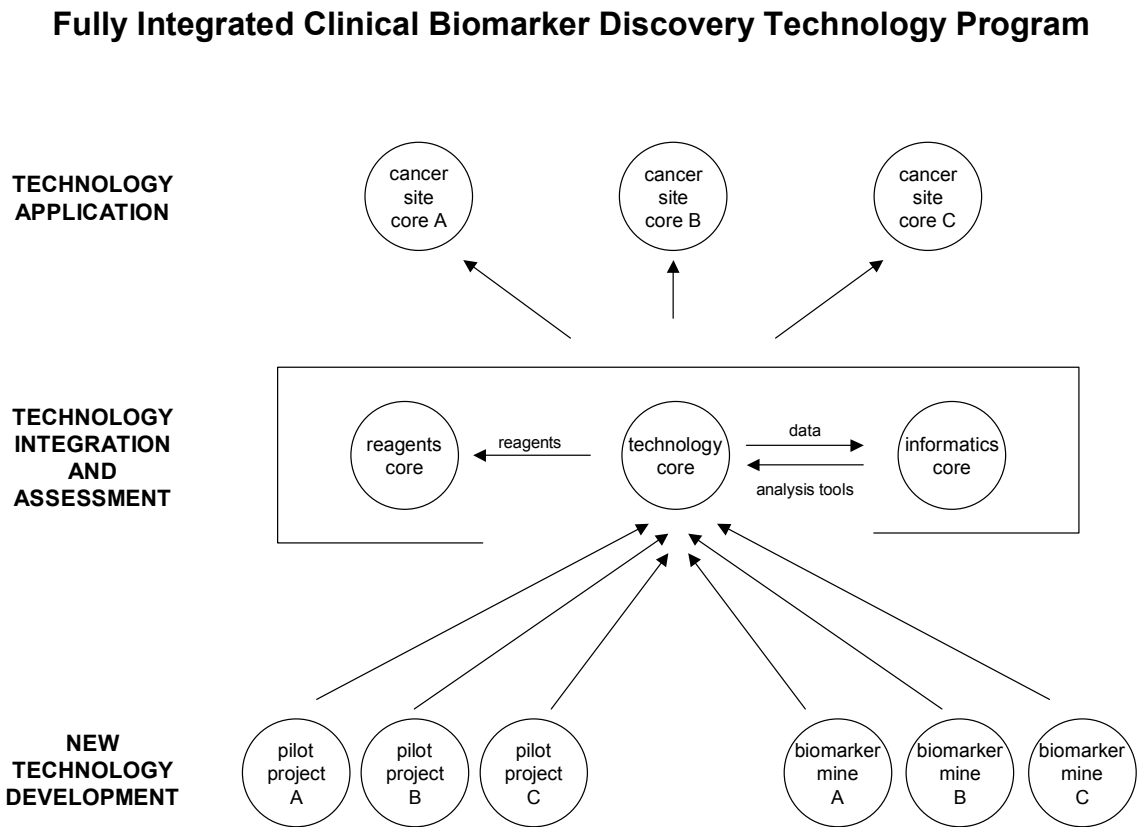

Figure 1. Organization of an effective clinical biomarker discovery program

1. The **Informatics Core** will develop tools to allow laboratories to communicate efficiently and to compare data. This Core will develop a standardized data format to facilitate cross-platform comparisons, and it will provide an open-source suite of analytical tools compatible with this standard data format to facilitate standardization of data review and analysis across laboratories and allow meaningful comparisons of results. Additionally, a central database for storing the curated data of the programs will be housed in the Informatics Core and made accessible to the public.

2. The **Reagents Core** will organize tools for biomarker discovery. This Core will provide a central virtual source for reagents, including human and mouse tissue samples, mouse models, antibodies, and other reagents as needed. Reagents, along with data on reagent performance and quality, will be acquired and dispersed quickly to other core facilities and satellites.

3. The **Technology Assessment Core** will assess technologies central to biomarker discovery in order to provide laboratories with the best possible techniques and

protocols. Initially, this Core will systematically compare existing technologies in each component of the biomarker discovery platform using standard reference plasma. The best-performing technologies from each component will then be integrated into an optimized platform against which new technologies (discovered via pilot projects and biomarker mines; see Figure 1 and below) can be tested. The ability of the integrated and optimized platform to identify biomarkers will be assessed using mouse models. This Core will also provide data to the Informatics Core for algorithm development and will also deposit useful reagents (including reference plasma and mouse tissues) into the Reagents Core for dissemination. Finally, this core will collaborate with Cancer Site Components to implement optimized technologies to find biomarkers in human samples.

4. **Cancer site components** are a team of investigators dedicated to biomarker discovery at a particular cancer site, such as breast, lung, prostate, colon, and others.

5. **Biomarker mine components** are investigators or small groups dedicated to optimizing the methods for discovery in a particular class of biomarkers, such as cell surface or secreted proteins.

6. **Pilot projects** are single investigator projects designed to test a new technology for biomarker discovery, e.g., protein chips, antibody production, etc. Where appropriate, promising new technologies will be reproduced and tested against current standards by the Technology Assessment Core.

One of the major points of the white paper is that a large, concerted effort is required to advance the field of biomarker discovery. The white paper also argues that the field of proteome-based biomarker discovery must answer the following questions:

- What combination of ionization source and mass analyzers will give the best achievable reproducibility, dynamic range, mass accuracy, and throughput for biomarker discovery?

- What techniques for fractionating plasma simplify the proteome sufficiently to allow significant depth of coverage, given the dynamic range of conventional mass spectrometers?

- What techniques for fractionating plasma are capable of the level of reproducibility (tested on multiple repeats of the same sample) that will be required to allow biomarkers to be detected while analyzing samples from a large number of cases and controls in high-dimensional data?

- Which techniques for quantitative mass spectrometry are robust enough for biomarker discovery?

- Can a standardized data analysis pipeline be established (and available electronically) so that results obtained from different laboratories can be analyzed using the same tools, allowing direct comparison of results?

- Can the best conventional fractionation schemes and mass spectrometry instrumentation interrogate a large enough "space" of the plasma proteome to discover diagnostic biomarkers?

Answering these questions will require a large, concerted effort in which the best existing instrumentation and data analysis tools, being developed in multiple laboratories throughout the scientific community, are integrated into one platform for data collection and analysis. Only then will researchers be able to compare rigorously multiple schemes for processing plasma, head to head on a controlled platform using identical samples, using multiple plasma processing schemes in a controlled manner.

Based on experience with planning other technology development programs, the NCI has committed to holding a series of symposia designed to introduce the proposed program to the research community and to solicit comments from a broad range of experts in proteomics, clinical oncology, biomarker discovery and technology, bioinformatics, and drug and diagnostic technology development. The first of the series of meetings was held on September 24, 2004.

Introduction

Dr. Anna Barker, NCI Deputy Director for Strategic Scientific Initiatives, began the meeting by noting that the NCI is committed to fostering the development of technology capable of detecting cancer at its earliest stage, treating it most effectively, and preventing it from developing. In that vein, the Institute is planning a technology development effort in proteomics, with a focus on identifying clinically relevant biomarkers. Dr. Barker commented that the logic behind crafting such a biomarker discovery initiative is that proteomics is at a stage where a concerted technology development effort should enable researchers to provide the signatures necessary for diagnosing cancer earlier, stratifying patients for clinical trials, and matching patients with the proper therapy. What is still not settled, however, is how the NCI should empower this effort. Indeed, the purpose of this meeting, she told the assembled experts, is to ask questions about what the research community needs in terms of supporting such an initiative in proteomics as applied to biomarker discovery. Dr. Barker also noted that the NCI is eager to begin this initiative and, by doing so, to serve as the leader for the international research community. Currently, no such leader exists, but if the United States does not take on such a role soon, other countries are set to do so in the near future.

Presentation of the Proposed Plan

Dr. Hartwell began his presentation by first noting that he believes that science can make a significant impact on medicine but wonders how science will have an impact on patients. It is this concern that led him to conclude that scientists can have the biggest impact by developing technology capable of dramatically improving early detection of cancer and, in his mind, proteomics is the technology that can best accelerate the

discovery of the cancer biomarkers that will drive early detection. From this conclusion, he went on to develop a white paper titled *Clinical Biomarkers Discovery Initiative*, which he put forth as **one** possible development model that he hopes will serve as a vehicle for discussion and input from the research and development community.

To begin this discussion, Dr. Hartwell laid out his arguments for NCI to fund an initiative in clinical proteomics, stating that, at present, discovery of biomarkers is the limiting step for improved early detection of cancer. While this will be a challenging endeavor, given the potential of various proteomics technologies, it should be possible to identify a sufficient number of biomarkers, at a relatively low price, to revolutionize early detection. Dr. Hartwell chose to emphasize early detection because it has been proven to be the best way of dramatically improving clinical outcomes for patients. As examples of early detection strategy, he cited the progress made in treating colon cancer and cervical cancer.

As far as cost is concerned, it should be possible to develop a panel of 60 biomarkers for early detection for as little as $6 million (or $100,000 each) if the validation trials are conducted along with existing clinical trials. This compares to development costs of $800 million for a new drug and $100 million to conduct a prevention trial. Dr. Hartwell noted that not many biomarkers are in the testing and validation pipeline, so reagents are a primary expense at this point.

Indeed, he added, this has not moved forward rapidly despite technological advances because the individual investigator does not have access to reagents; these need to be supplied to the research community globally. He added that validation of biomarkers should not be done in individual trials, but rather in multiplexed diagnostic validation trials run in tandem with therapeutic effectiveness trials. In contrast to the drug development pipeline, which begins by screening huge numbers of molecules and then testing only one at a time, the biomarker pipeline will identify candidate biomarkers and then test each one in tissue and blood of 100 cancer patients and healthy patients to ensure that the marker is capable of distinguishing cancer from non-cancer (Figure 2).

Blood protein biomarkers for cancer are likely to exist, Dr. Hartwell noted, since tumors and the vessels around them are leaky. It is easy to estimate that 1 million different types of proteins are produced by the body, but our current tools can measure only about 100 types. However, even with current technologies, a more systematic approach should allow us to go deeper into the proteome. In fact, he added, it should be relatively straightforward to screen 1,000 biomarker candidates at a time and 10,000 random proteins.

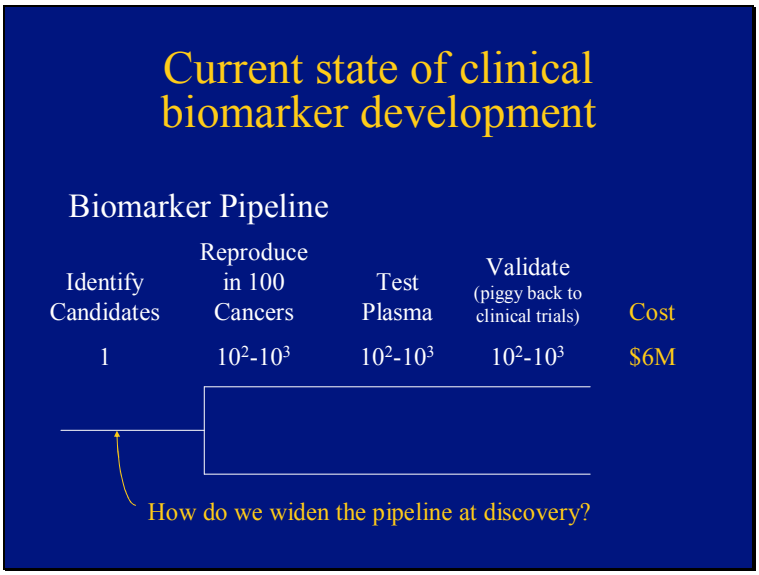

Figure 2. A model for a biomarker development pipeline

Next, Dr. Hartwell discussed two possible methods for identifying biomarker candidates from tissue and fluid samples. One method involves a search for candidates based on likely biological properties; the other is based on the function of likely candidates (Figure 3). He also noted a third method—looking randomly at all proteins—but did not discuss this approach in detail. Searching for biomarkers by properties would entail looking for proteins that are differentially expressed, are found on the cell surface or are secreted, are found either with or without albumin, are of low molecular weight, or are glycosylated, phosphorylated, ubiquinated, or methylated. Searching by function would involve looking for biomarkers associated with angiogenesis, lymphogenesis, metastasis, DNA repair, proteolysis, mitosis, stress response, growth signaling, wound healing, and other processes that may be associated with cancer. Dr. Hartwell noted that the reagents needed for this approach have yet to be developed to any great extent.

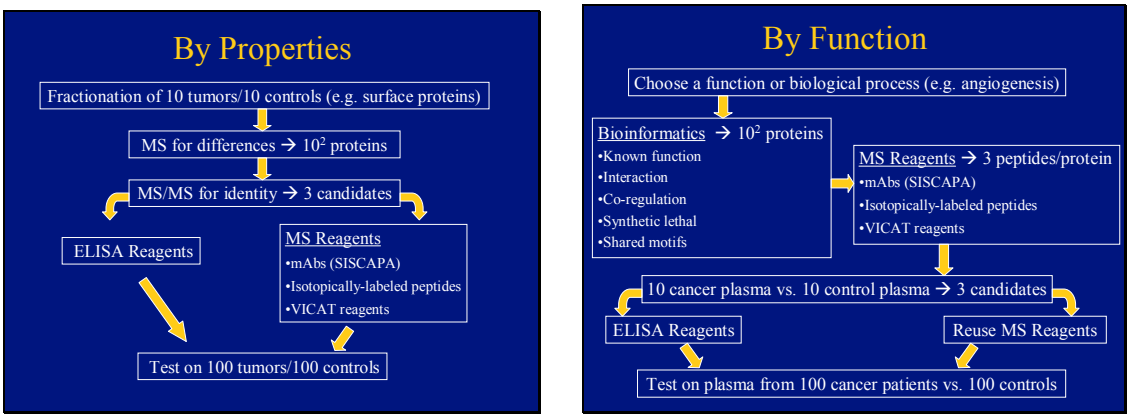

Figure 3. Schemes for biomarker discovery based on properties and function

Outlining how development efforts would proceed for the two approaches, Dr. Hartwell laid out two different development pathways, both of which would lead to testing a panel of biomarkers. Both approaches would also entail development of a host of new reagents and refinement of mass spectroscopic methods. The goal of both pathways would be to reach what Dr. Hartwell called validation stage 1: showing that a biomarker is capable of distinguishing those individuals with cancer from those without it. Once a biomarker-based diagnostic test completed this stage of validation, the private sector would assume responsibilities for the final stages of development and validation. Dr. Hartwell estimated that this approach would yield 3 validated biomarkers from every 100 identified as possible candidates. For the "properties" approach, reaching validation stage 1 for a single cancer site would cost approximately $2 million, while the "function" approach would be more costly for the first biomarker identified, perhaps as much as $18 million to $20 million (Figure 4). Since many of the reagents developed and used in the functional approach can be reused for other cancer sites, the cost for subsequent sites would drop to approximately $4 million. Dr. Hartwell also noted that having a supply of reagents available to the research community is critical for either of these two approaches to succeed at the lowest cost and in the shortest time possible.

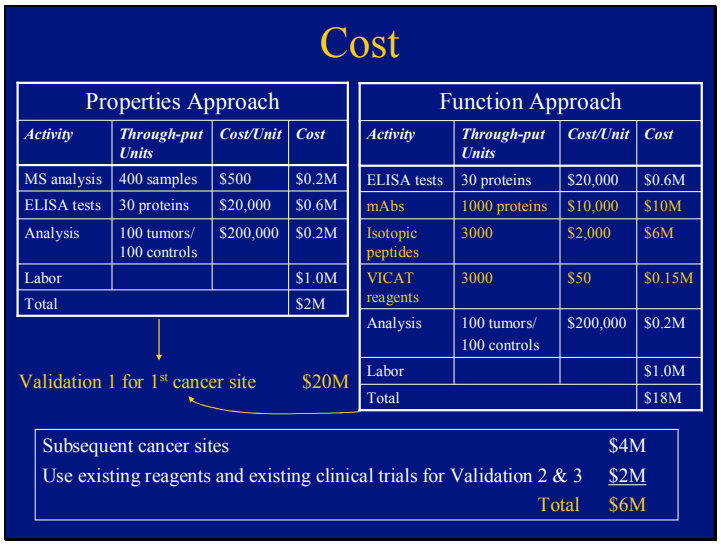

Figure 4. Projected costs of biomarker discovery and validation through stage 1

Dr. Hartwell then closed his presentation with a list of goals for a coordinated clinical proteomics and biomarker discovery initiative:

- Establish criteria and centers for testing biomarker discovery technologies in order to define an effective pipeline for discovery:
  - Develop a technology assessment site to compare competing technologies head to head on identical samples from mouse models of cancer.
  - Identify technology innovations:
    - Pilot grant funding

        - NCI's UIP and Innovative Molecular Analysis Technologies (IMAT) programs
    - Encourage academic/industry collaborations.
    - Promote engineering to improve reproducibility, throughput, and automation.
- Develop a publicly available informatics platform that permits data storage, analysis, searching, and comparison.
- Establish a consortium of collaborating laboratories to discover biomarkers in particular cancer sites and for particular classes of biological molecules.
- Establish repositories of reagents for clinical biomarker discovery, available to the community.
- Promote the translation of new imaging agents to clinical trials.

General Discussion Session

The remainder of the meeting was devoted to discussion of what is needed to drive a successful effort using proteomics for biomarker discovery. One participant noted that the field needs to challenge the paradigm of biomarker discovery so that it is a more cooperative rather than competitive effort, and that NCI can use its clout to accomplish this paradigm change. He also stated his opinion that the problem of biomarker discovery and validation is more complex than acknowledged and noted that it will be necessary to follow the "normal" patients used to validate a biomarker for an extended period of time to ensure that they remain cancer-free for some time after the validation experiment.

Another participant said that the effort to develop markers for cancer so far has been a failure, largely because it is an amazingly difficult task, and added that it is on the technology development front, rather than the reagent side, that an all-out effort is needed to make significant progress. Others argued that reagent development is indeed important for progress—there is a need for monoclonal antibodies against every known human protein, for example—but most important is that both technology and reagent development occur as part of an organized plan. Such a top-down effort would attract talent from a variety of areas, driving the effort more rapidly. From further discussion emerged the question of whether proteomics technology is sufficiently developed to organize a concerted biomarker discovery effort. There was general agreement that technology improvement is needed but that any technology development effort must have a clear goal, with one participant noting that the Human Genome Program did not begin as a technology development program; technological improvements were driven by the overall sequencing effort.

One breast cancer researcher noted that a recent review of breast cancer biomarkers by the Institute of Medicine found that, while there have been dramatic technologies and technological innovations, researchers in the field are operating in a data analysis and applications development vacuum. Whether a new technology can be implemented in the medical care arena, how that would happen, and at what expense should be considered early in the process. These considerations cannot come at the end of the process. Several participants agreed strongly with this idea that biomarker development must occur in the context of how it would eventually be deployed in clinical practice. More than one

participant commented that any new test faces considerable resistance to adoption because of reimbursement issues.

Discussion continued on the topic of commercial development, with the opinion voiced that systematizing biomarker development, as called for in the white paper, may not be ambitious enough because translational capability does not yet exist to persuade a diagnostic company that it can earn a reasonable rate of return on investment in a stage 1 validated marker. On the basis of this discussion, it may be worthwhile to examine how to bring these developments to the clinic and the market, since the field needs to define the development pipeline and how it would work. Any discussion must include participants from the Centers for Medicare and Medicaid Services (CMS) to address reimbursement issues and from the medical community to determine how a development would be used. Several participants raised the idea of reducing development costs by tapping into drug development funds, since useful biomarkers should have a dramatic effect on reducing drug development costs. Dr. Barker noted that the NCI is working on these issues though two interagency task forces: One task force with the U.S. Food and Drug Administration is examining the role that biomarkers can play and how they will be approved; a second task force with CMS is looking at costs and reimbursement issues. She also commented that NCI is working to lower the risk of development by using the NCI-funded comprehensive cancer centers to advance potential biomarkers through the validation and development process.

After several participants voiced support for a highly organized effort to discover and validate biomarkers, one attendee wondered how the Human Genome Project (HGP) mechanism could be applied to biomarker discovery. One participant noted a strong similarity between the two projects. Before the HGP was established, numerous laboratories were working on specific sequences but there was no uniform effort in bioinformatics, statistical validation, or technology development and standardization. Forcing the field to collaborate on these issues played a critical role in streamlining and accelerating the sequencing effort. Several participants then commented that leveraging existing resources, as was done with the HGP, will be key, with the SPOREs and cancer centers playing a critical role. The ongoing effort at identifying biomarkers for biotoxicity, organized and supported by the National Institute of Environmental Health Sciences (NIEHS), could also serve as a model for NCI's much larger initiative. Some opposition was voiced, however, to a top-down, organized approach, with the arguments made that such an approach would stifle creativity and progress. Therefore, since this effort to pick technologies will continue for years, no one would be happy.

The discussion then turned to what the NCI can do to catalyze biomarker discovery efforts. Participants echoed an earlier comment that NCI should encourage multidisciplinary team-building activities that will bring a wide variety of talent to bear on the problem. There was wide support for the idea that the NCI should not establish centers on the basis of a single expertise; that is, there should not be, for example, one center for proteomic bioinformatics and another for technology development. One participant added that such multidisciplinary centers would help train the next generation of cross-disciplinary researchers, such as pathologists with biomarker experience and

biocomputational experts skilled in the clinical sciences. There was also a brief discussion on how to utilize existing resources, including the SPOREs, Early Detection Research Network (EDRN), comprehensive cancer centers, and NCI intramural program. There was a consensus among the participants that any NCI effort in biomarker discovery must include mechanisms that will put these existing resources into a central position in the research agenda. One participant added that any NCI program must use funding mechanisms that go beyond R01 support, since technology development fares poorly historically in the R01 review process.

The National Institute of General Medical Sciences GLUE grant program was mentioned as a useful model for NCI to study. There was also a suggestion that NCI could provide help in making clinical samples available and promulgating standards for sample preparation. NCI also should establish a central resource for distributing standardized materials, reagents, and even mouse models to help technology development efforts. Participants also voiced support for Dr. Hartwell's suggestion that the new genetic models of human cancer can serve as a test bed for potential biomarkers, although this was not unanimous. Several participants did suggest, however, that while mouse models may not be the best tool for biomarker discovery, they could serve as the mechanism for testing technologies thoroughly before moving on to humans.

One participant asked whether there is a database of all the biomarkers that have been discovered and where they are in development. Dr. Barker noted that the NCI has efforts under way to assemble this information, and another participant added that the SPOREs are engaged in putting together such a list. Estimates from around the table suggested that hundreds of biomarkers are available for study, but little has been done to determine whether any of these have potential clinical use. Indeed, said one participant, most of the biomarkers discovered to date have had poor prognostic value, at least when used individually.

The issue of whether the NCI could provide standardized samples for testing candidate markers was raised. If a set of standardized samples were available from NCI for use by the field, it would provide a good deal of credibility. One participant questioned the need to identify every protein in the human proteome, given that an NCI pilot project has already identified more than 3,000 proteins; perhaps another 2,000 would provide a good base for biomarker discovery efforts. According to the ensuing discussion, new technology is needed to measure small concentrations of these proteins, as well as standards for how to measure them and for tissue collection and processing methods that will be used to assay them. This was seconded by several participants.

The issue of intellectual property in the biomarker arena was raised by several participants. One of the concerns voiced was that problems may lie ahead when it comes to aggregating markers in a single test. There seemed to be widespread agreement that researchers in the field must agree to grant nonexclusive licenses only to those who wish to explore the usefulness of any marker. Dr. Barker noted that the NCI has commissioned a study by the National Academy of Sciences on intellectual property issues. She added

that the field needs a good cross-licensing mechanism similar to that developed by the computer industry.

Synopsis of Key Points

Following the meeting, Dr. Hartwell provided the following synopsis of the recommendations that emerged from the day's discussions:

- A more systematic effort in biomarker discovery is warranted at this time.
- There is a need to define a molecular diagnostic pipeline and to establish standards for validation at each step. This effort should look beyond the pharmaceutical development pipeline, which may be unsuitable for biomarker development.
- High-quality reagents such as antibodies, aptamers, and isotopically labeled peptides are needed. A catalog or central resource for such reagents should be developed to maximize availability to the field.
- A high-quality database with algorithms for processing and analyzing mass spectrometry data is needed.
- Any new program should integrate with existing programs, such as the EDRN, mouse consortium, SPOREs, and NCI intramural program.
- A new program should encourage innovation, particularly in the area of technology development.
- Any program should follow a design-build model to allow flexibility to incorporate in real time new developments in technology and science.
- Access to high-quality tissues is crucial, and involving pathologists in any proteomics effort could help ensure that such tissues are available.
- It is important to use mouse models as well as human tissues.
- A program should recruit innovations already existing within the community.
- A database of candidate biomarkers already discovered in the community should be constructed.
- Academic investigation of diagnostic markers needs to progress to higher levels of validation in order to attract industry.
- Other diseases and institutes at NIH should be considered for inclusion.
- It is necessary to consider how a marker will be used clinically early in the development process.
- Any effort should start with one tumor site and prove that a systematic approach is effective.
- Any effort should emphasize functional information and therapeutic response for end goals as much as early detection.
- There is a need to recruit the most effective scientists and maximize resources through a GLUE grant mechanism.
- There is a need for a systematic examination of collection and storage issues.
- A new training program for physician-scientists, and particularly pathologists, is needed to recruit more of these professionals to the proteomics and biomarker discovery and development field.

- Dovetailing biomarker validation with drug clinical trials may represent a way of reducing development costs for biomarkers and new pharmaceuticals while simultaneously providing clinicians with experience in utilizing biomarkers in clinical practice.

Concluding Comments

Dr. Andrew von Eschenbach, Director of NCI, closed the meeting by thanking the participants for their comments and suggestions on how to best design a proteomics initiative that would accelerate clinically useful biomarker discovery. The NCI, he noted, is eager to launch this program, which will build on the existing cancer research enterprise and leverage the resources currently available to the NCI through its cancer centers, SPOREs, EDRN, and intramural program. Dr. von Eschenbach noted that the NCI will take a design-build approach to its proteomics initiative, designing the program and launching and modifying it in order to accelerate progress immediately while leaving the program sufficiently flexible to incorporate new technologies and scientific advances. He also noted that the NCI will fill the leadership vacuum that exists in the proteomics arena, in terms of planning and directing the effort and ensuring that the field has the necessary resources needed to achieve success. Dr. von Eschenbach then assured the participants that NCI will continue to solicit input and help in developing this plan but the effort will be going forward in the very near future.