

## Proteomics and Cancer

### Key Points

- Proteomics is the study of all proteins in a cell, tissue, or organism—including their identity, their biochemical properties and functional roles, and how their quantities, modifications, and structures change in response to the needs of the body or in disease (see Question 1).
- The human proteome is much larger and more complex than the human genome. It has been estimated that the human body may produce more than 1,000,000 different protein species from its 20,000 to 25,000 protein-coding genes (see Question 2).
- Researchers hope that proteomics research will help find new ways to diagnose cancer early, identify the best treatments for individual patients with specific types of cancer, and determine whether cancer has come back (recurred) after treatment. However, many technical challenges must be overcome before proteomics techniques can be used in the clinic (see Questions 4 and 5).
- The National Cancer Institute (NCI) currently supports programs both intramurally (within NCI) and extramurally (within the larger cancer research community) to advance the field of proteomics (see Question 6).

### 1. What is proteomics?

The term proteome refers to the totality of the proteins in a cell, tissue, or organism. Proteomics is the study of these proteins—their identity, their biochemical properties and functional roles, and how their quantities, modifications, and structures change during development and in response to internal and external stimuli. The field of proteomics has been propelled by advances in mass spectrometry and other techniques that have made it possible to analyze proteins in large numbers of biological samples rapidly and at low cost (1, 2).

Proteins are products of the genetic code (DNA), and they drive the workings of cells, tissues, and organs. The proteins produced in a specific cell determine that cell's function in the body.

The human proteome—like the proteomes of all organisms—is dynamic, changing constantly in response to the needs of the body. It differs widely between people depending on factors such as age, sex, diet, level of exercise, and sleep cycle. The proteome also changes in response to cancer and other diseases, making the proteome of great interest to medical researchers.

For example, cancer cells often secrete specific proteins or fragments of proteins into the bloodstream and other bodily fluids, such as urine and saliva (3). Researchers hope to discover groups or patterns of proteins—called protein signatures—in these easily accessible fluids that provide information about the risk, presence, and progression of disease (4–6). This knowledge could ultimately help doctors better detect cancer before symptoms are present and customize treatment to the individual patient.

### 2. How is proteomics different from studying genes (genomics)?

The proteome of an organism is much larger and more complex than its genome. Many genes have the potential to produce more than one version of the protein they encode. In addition, proteins are frequently modified by cells after they are made.



Protein modifications include the addition of various chemical groups, such as phosphate, acetate, and methyl groups, or the addition of carbohydrate (sugar) and lipid (fat) molecules. These modifications help regulate the function of proteins, as well as their location inside or outside of cells. Taking protein modifications into account, it has been estimated that the human body may produce more than 1,000,000 different protein species from its 20,000 to 25,000 protein-coding genes (7).

This higher level of complexity is compounded by the fact that the protein composition of an organism or a tissue changes constantly as new proteins are made, existing proteins are eliminated, and proteins become modified or demodified in response to internal and external stimuli. In contrast, a person's genome remains relatively unchanged over the course of his or her lifetime.

### **3. What techniques and technologies are used in the study of proteomics?**

Two main approaches are currently used in cancer proteomics research: Protein identification and pattern recognition (8).

For protein identification, researchers can use several different techniques. For example, based on what is known about the biology of a certain type of cancer, researchers can select antibodies that bind to proteins thought to be overexpressed in that cancer type. These antibodies are then placed into wells on a protein microarray (similar to a DNA gene-expression microarray, sometimes called a 'gene chip'), and a sample of the fluid being tested (such as blood or urine) is washed over the microarray (9, 10). If present, the proteins will bind to the antibodies and can be detected by a technique known as fluorescence microscopy. The amounts of these proteins found in samples taken from patients with cancer can be compared with those from patients without cancer.

Gel-based electrophoresis techniques can also be used to isolate proteins of interest. Gel electrophoresis is a technique that separates proteins based on their mass and electrical charge. Once proteins that are more abundant in cancer patients have been isolated using this method, they can be identified by an enzyme-linked immunosorbent assay (ELISA). ELISA uses antibodies to identify proteins in a manner similar to microarrays.

For pattern recognition, a technique called mass spectrometry (MS) can be used to measure the masses and relative quantity of all proteins in a particular biological sample. MS machines can produce protein profiles (signatures) that can be compared between samples taken from patients with and without cancer, but they cannot identify the individual proteins that produce the signatures (11, 12).

Other technologies include chromatography and a different type of MS, called tandem MS, which can be used to identify individual proteins in an MS signature (6). Identification of the actual proteins that create a protein signature identified by mass spectrometry is important because it reduces the likelihood that differences in protein signatures observed between biological samples taken from patients with and without cancer are actually due to bias. Differences in the way samples from patients with cancer (cases) and from those without (controls) are handled during collection, storage, or processing can all introduce bias—the appearance of a difference in proteins between cases and controls due to cancer when no difference actually exists (13, 14).

Both protein identification and pattern recognition research also require high-powered computing and bioinformatics systems to process the enormous amount of data that are produced by these studies.

Confidence in the results of proteomics studies and identified biomarker signatures will require promising results to be reproduced in different populations and by different laboratories, a process called validation.

### **4. What technical challenges must be overcome to advance proteomics research?**

Proteomics research is currently limited by the technologies that are available for analyzing proteins. For example, mass spectrometers can accurately measure very small amounts of protein—proteins that are 1 in 100 to 1 in 10,000 times less common than other proteins in a sample of tissue or bodily fluid; however, proteins produced by cancer cells are often present in even smaller quantities, making their detection difficult (15). Researchers are working to improve the sensitivity of mass spectrometry to allow scientists to detect these rare cancer proteins.

In addition, research programs, including the National Cancer Institute's (NCI) Antibody Characterization Program (<http://antibodies.cancer.gov/about/>), are developing antibodies and other molecules called affinity reagents to accurately identify proteins in blood and tissue samples taken in the clinic; however, the

production of these molecules still lags behind the needs of researchers (16, 17). This lack of antibodies slows both protein identification research (which requires antibodies to indicate the presence of known proteins in biological samples) and pattern recognition research (researchers can use antibodies to identify the proteins that make up a protein signature detected by MS).

Another challenge is that proteins are more likely to be degraded or otherwise altered during isolation, storage, and handling than molecules such as DNA. Therefore, stringent precautions are needed to maintain the integrity of proteins to ensure proteomics research results are accurate. To address this issue, researchers are developing best practices and standardizing procedures for collecting, processing, storing, and sharing tissue and bodily fluid specimens, which are collectively known as biospecimens.

Ultimate confirmation of the existence of individual protein biomarkers or protein signatures and their association with specific types of cancer—a process known as biomarker validation—will require the reproduction of findings in different laboratories and additional populations of patients (18).

## **5. How do researchers hope to use proteomics to improve early detection, diagnosis, and treatment?**

Few screening techniques exist to find cancer early, before it causes symptoms. Cancer researchers hope that proteomics will enable them to identify secreted proteins that can serve as biomarkers of disease. These proteins may be secreted into blood or other bodily fluids, including urine, saliva, or sweat, which can be obtained noninvasively. The proteins could then be used to detect specific types of cancer before the disease has become advanced or symptoms appear (4–6).

To date, most secreted proteins studied as cancer screening biomarkers have yielded too many false-negative results (failure to detect cancer in those who have it, a phenomenon known as low sensitivity) and/or too many false-positive results (indicate the presence of cancer in someone who does not have it, a phenomenon known as low specificity). For example, the protein CA-125, which has been used to screen for ovarian cancer, has low sensitivity: Blood levels of CA-125 are elevated in only 50 to 60 percent of women who have early-stage ovarian cancer (8). CA-125 also has low specificity: Benign conditions, such as endometriosis and pregnancy, can elevate CA-125 levels. In addition, the protein prostate-specific antigen (PSA), which is commonly used to screen for prostate cancer, has low specificity: Only 25 to 30 percent of men referred for prostate biopsy on the basis of increased blood levels of PSA actually have prostate cancer (19). PSA also has low sensitivity: In one large study, more than 15 percent of men who had PSA blood levels in the normal range were found to have prostate cancer (20).

Furthermore, preliminary studies have suggested that groups of proteins (protein signatures) may be much more accurate tools for detecting cancer than individual proteins (5, 21, 22). Most proteins produced by cancer cells are not unique to cancer, meaning that noncancerous cells can produce them too (10, 12). It is also unlikely that all cancer cells of a given type will produce the same amounts of a single protein.

Researchers also hope to use proteomics to predict the likelihood that cancers will respond to specific treatments; to show that cancers are, in fact, responding to treatments; to predict the likelihood that specific treatments will cause unacceptable side effects in individual patients; and to monitor patients for signs of cancer recurrence.

Several studies have already shown promising results in defining protein signatures for ovarian, breast, prostate, bladder, pancreatic, lung, and head and neck cancers (23). However, techniques used to collect biospecimens and analyze proteins often differ from laboratory to laboratory, and these results have proven difficult to verify using current technologies. As noted previously (see Question 4), standardized practices and procedures for collecting and analyzing biospecimens will be required to have confidence in the results of proteomics research and to establish their ultimate clinical utility.

## **6. How does NCI support the field of proteomics and its potential to advance cancer research?**

NCI currently funds proteomics programs both intramurally (within NCI) and extramurally (within the larger research community).

## ***Clinical Proteomic Technologies for Cancer Initiative***

In 2006, NCI launched the Clinical Proteomic Technologies for Cancer (CPTC) initiative (<http://proteomics.cancer.gov/>), a 5-year technology development program that harnesses the expertise of investigators in academic institutions around the country. The CPTC was founded to accelerate the development of better technology for proteomics research in response to scientists' needs for more accurate, more standardized, and more reproducible protein measurement techniques.

The CPTC is composed of three major, integrated programs:

- The Clinical Proteomic Technology Assessment for Cancer (CPTAC) network, a collaborative effort involving both the public and private sectors, is conducting rigorous assessments of two major technologies currently used to analyze proteins and peptides—mass spectrometry and affinity capture platforms. More information about the CPTAC is available at <http://proteomics.cancer.gov/programs/CPTAC/> on the Internet.
- The Advanced Proteomic Platforms and Computational Sciences initiative is focusing on the development of new proteomics technologies and the development of computational approaches for analyzing, processing, and sharing proteomics data. More information about this initiative can be found at <http://proteomics.cancer.gov/programs/platforms/> on the Internet.
- The Proteomic Reagents and Resources Core serves as a central source for affordable, well-characterized, and validated reagents and supporting resources for the scientific community. More information about this program is available at [http://proteomics.cancer.gov/programs/reagents\\_resources](http://proteomics.cancer.gov/programs/reagents_resources) on the Internet.

## ***Early Detection Research Network***

Since 1998, NCI has also funded the Early Detection Research Network (EDRN), a group of over 50 researchers across the country working collaboratively to develop and test promising biomarkers or technologies for the early detection of cancer. The EDRN brings together dozens of institutions to help accelerate the translation of biomarker information into new ways of testing for cancer in its earliest stages and for cancer risk. Currently, EDRN funds a number of proteomic-based biomarker discovery research projects utilizing state-of-the-art proteomics technologies. Many projects have yielded potential candidate biomarkers for lung, pancreatic, and prostate cancers and are currently being evaluated (24–26). The EDRN's Web page is located at <http://edrn.nci.nih.gov> on the Internet.

## ***Biomedical Proteomics Program***

NCI's Biomedical Proteomics Program (BPP) focuses on the identification and characterization of the protein signatures of human cancer cells and tissues and the application of proteomics technologies directly to the diagnosis of cancer, monitoring of side effects of cancer therapy, and improvements in treatment. The BPP's Web page can be found at [http://web.ncifcrf.gov/rtp/prot/site/default\\_flash.asp](http://web.ncifcrf.gov/rtp/prot/site/default_flash.asp) on the Internet.

### **7. Where can I find a list of current clinical trials that involve proteomics?**

In the HTML version of this fact sheet on NCI's Web site (<http://www.cancer.gov/cancertopics/factsheet/detection/proteomics>), the text below links to a search of clinical trials that are studying proteomics techniques and are now recruiting patients. The trials are included in the clinical trials database that is searchable on NCI's main Web site, located at <http://www.cancer.gov> on the Internet.

[Active cancer clinical trials studying proteomic profiling](#)

Alternatively, the database can be searched directly using the search form located at <http://www.cancer.gov/clinicaltrials/search> on the Internet.

Additional information about clinical trials is available from NCI's Cancer Information Service (1–800–4–CANCER) or on the main clinical trials page of NCI's Web site at <http://www.cancer.gov/clinicaltrials> on the Internet.

## Selected References

1. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: The long and uncertain path to clinical utility. *Nature Biotechnology* 2006; 24(8):971–983.
2. Smith L, Lind MJ, Welham KJ, Cawkwell L. Cancer Biology Proteomics Group. Cancer proteomics and its application to discovery of therapy response markers in human cancer. *Cancer* 2006; 107(2):232–241.
3. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature Clinical Practice Oncology* 2008; 5(10):588–599.
4. Latterich M, Abramovitz M, Leyland-Jones B. Proteomics: New technologies and clinical applications. *European Journal of Cancer* 2008; 44(18):2737–2741.
5. Jurisicova A, Jurisica I, Kislinger T. Advances in ovarian cancer proteomics: The quest for biomarkers and improved therapeutic interventions. *Expert Review of Proteomics* 2008; 5(4):551–560.
6. Paulovich AG, Whiteaker JR, Hoofnagle AN, Wang P. The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. *Proteomics Clinical Applications* 2008; 2(10–11):1386–1402.
7. Jensen ON. Modification-specific proteomics: Characterization of post-translational modifications by mass spectrometry. *Current Opinion in Chemical Biology* 2004; 8(1):33–41.
8. Nossov V, Amneus M, Su F, et al. The early detection of ovarian cancer: From traditional methods to proteomics. Can we really do better than serum CA-125? *American Journal of Obstetrics and Gynecology* 2008; 199(3):215–223.
9. Gulmann C, Sheehan KM, Kay EW, et al. Array-based proteomics: Mapping of protein circuitries for diagnostics, prognostics, and therapy guidance in cancer. *Journal of Pathology* 2006; 208(5):595–606.
10. Whiteley GR. Proteomic patterns for cancer diagnosis—promise and challenges. *Molecular BioSystems* 2006; 2(8):358–363.
11. Simpson RJ, Bernhard OK, Greening DW, Moritz RL. Proteomics-driven cancer biomarker discovery: Looking to the future. *Current Opinion in Chemical Biology* 2008; 12(1):72–77.
12. Solassol J, Jacot W, Lhermitte L, et al. Clinical proteomics and mass spectrometry profiling for cancer detection. *Expert Review of Proteomics* 2006; 3:311–320.
13. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. *Nature Reviews Cancer* 2005; 5(2):142–149.
14. Whiteley G. Bringing diagnostic technologies to the clinical laboratory: Rigor, regulation, and reality. *Proteomics Clinical Applications* 2008; 2(10–11):1378–1385.
15. Engwegen JY, Gast MC, Schellens JH, Beijnen JH. Clinical proteomics: Searching for better tumour markers with SELDI-TOF mass spectrometry. *Trends in Pharmacological Sciences* 2006; 27(5):251–259.
16. Haab BB, Paulovich AG, Anderson NL, et al. A reagent resource to identify proteins and peptides of interest for the cancer community: A workshop report. *Molecular & Cellular Proteomics* 2006; 5(10):1996–2007.
17. Blow N. Antibodies: The generation game. *Nature* 2007; 447(7145):741–744.
18. Addona TA, Abbatiello SE, Schilling B, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring–based measurements of proteins in plasma. *Nature Biotechnology* 2009; 27(7):633–641.
19. Parekh DJ, Ankerst DP, Troyer D, et al. Biomarkers for prostate cancer detection. *Journal of Urology* 2007; 178(6):2252–2259.

20. Thompson IM, Pauler DK, Goodman PJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level  $\leq 4.0$  ng per milliliter. *New England Journal of Medicine* 2004; 350(22):2239–2246.
21. Visintin I, Feng Z, Longton G, et al. Diagnostic markers for early detection of ovarian cancer. *Clinical Cancer Research* 2008; 14(4):1065–1072.
22. Sardana G, Dowell B, Diamandis EP. Emerging biomarkers for the diagnosis and prognosis of prostate cancer. *Clinical Chemistry* 2008; 54(12):1951–1960.
23. Roboz J. Mass spectrometry in diagnostic oncoproteomics. *Clinical Investigation* 2005; 23(5):465–478.
24. Faca VM, Song KS, Wang H, et al. A mouse to human search for plasma proteome changes associated with pancreatic tumor development. *PLoS Medicine* 2008; 5(6):e123.
25. Qiu J, Choi G, Li L, et al. Occurrence of autoantibodies to annexin I, 14-3-3 theta and LAMR1 in prediagnostic lung cancer sera. *Journal of Clinical Oncology* 2008; 26(31):5060–5066.
26. Taylor BS, Pal M, Yu J, et al. Humoral response profiling reveals pathways to prostate cancer progression. *Molecular & Cellular Proteomics* 2008; 7(3):600–611.

# # #

#### Related NCI materials and Web pages:

- NCI's Antibody Characterization Program Web Page (<http://antibodies.cancer.gov/>)
- NCI's Biomedical Proteomics Program Web Page ([http://web.ncifcrf.gov/rtp/prot/site/default\\_flash.asp](http://web.ncifcrf.gov/rtp/prot/site/default_flash.asp))
- NCI's Clinical Proteomic Technologies for Cancer Web Page (<http://proteomics.cancer.gov>)
- NCI's Early Detection Program Web Page (<http://edrn.nci.nih.gov>)
- NCI's Understanding Cancer Series: Molecular Diagnostics (<http://www.cancer.gov/cancertopics/understandingcancer/moleculardiagnosics>)

#### How can we help?

We offer comprehensive research-based information for patients and their families, health professionals, cancer researchers, advocates, and the public.

- **Call** NCI's Cancer Information Service at 1–800–4–CANCER (1–800–422–6237)
- **Visit** us at <http://www.cancer.gov> or <http://www.cancer.gov/espanol>
- **Chat** using LiveHelp, NCI's instant messaging service, at <http://www.cancer.gov/livehelp>
- **E-mail** us at [cancergovstaff@mail.nih.gov](mailto:cancergovstaff@mail.nih.gov)
- **Order** publications at <http://www.cancer.gov/publications> or by calling 1–800–4–CANCER
- **Get help** with quitting smoking at 1–877–44U–QUIT (1–877–448–7848)

**This fact sheet was reviewed on 8/7/09**