Clinical Proteomic Technologies Initiative for Cancer
Animated Video Tutorial

Transcript

Proteomic Technologies and Cancer

In this feature presentation, we will take a closer look at the process for discovering clinical biomarkers, from patients to samples and back to patients, that may provide a new approach to early detection and treatment.

Introduction

Despite recent progress, cancer takes more than a half million American lives each year. In many of these cases, we lost the battle because we detected the enemy far too late.

In this feature presentation on Proteomic Technologies and Cancer, we will take a close look at how scientists and physicians are working to identify clinical biomarkers – the molecular signatures that indicate the presence of cancer.

These signatures, made up of proteins and other molecules found in patient samples such as blood, will usher in a new approach to medicine based on early detection, and rapid response.

The identification of proteins associated with cancer in the patient will enable us to diagnose and provide treatment even before clinical symptoms appear. The sooner we find the disease, and the sooner we treat it, the greater our chances will be in providing a cure, or enabling the patient to live a longer, productive life.

Reading the molecular signatures of cancer could also guide us toward the right combination of therapies that are appropriate to each individual, opening the door to personalized cancer treatment.

But the benefits of this approach will not be realized until we improve the set of advanced technologies used to identify proteins found in patient samples. This set of technologies is called Proteomics.

Proteomics is the identification and characterization of the many hundreds of thousands of proteins expressed in an organism or cell type at a given time.

Proteins are made up of long chains of amino acids, folded into a unique structure. They're the catalysts for most biological activities within the body, including the process of cell division and growth.

When these functions go awry, such as when too many or too few proteins are produced, or when a critical network is disrupted by a mutant protein, it can result in the uncontrolled division of cells … a tumor forms … and the very earliest stages of cancer begin.

But when a molecular network is disrupted in a cancer cell, it also leaves a characteristic signature measured in the differences in proteins that it contains.

In recent years, proteomic technologies have been developed to read these signatures and identify a specific set of proteins associated with cancer.

Proteomics is a technology that is more familiar to research laboratories than doctors' offices. As a science, it's still young. Much work remains to be done to improve its reliability before it can provide us with the biomarkers we can use to diagnose patients. In this presentation, we will describe how proteomics contributes to clinical biomarker discovery – and the challenges we face in moving this powerful technology from the laboratory bench to the patient bedside.

Proteomic Technologies and Cancer Diagram

During the presentation, we will use this diagram to follow the process of candidate biomarker discovery.

The first step is **Sample Preparation**, in which we obtain a pure mixture of proteins from patient tissue or body fluid. During **Separation and Capture** we will isolate single proteins from this mixture. The next step involves **Analysis and Identification** of the proteins by their chemical composition. Scientists can then use this information to develop a diagnostic test.

The **Clinical Validation** stage is needed to establish the proteins as true biomarkers in a clinical trial, allowing the diagnostic test to be used on real patients.

In the final section on **Improving Proteomic Technologies**, we will take a close look at the steps being taken to sharpen the tools of protein discovery and biomarker validation.

You will hear about many new and established proteomic technologies. Some we will describe in detail.

You can learn about the challenges faced in each technology by clicking on this icon [Additional Information Icon].

We will often make use of symbols to represent certain kinds of molecules. For a legend describing what each symbol represents, click on this icon here [Legend Icon].

Sample Preparation

Sample Acquisition

We will now continue to the first stage of candidate biomarker discovery, Sample Preparation, in which we obtain a pure mixture of proteins from patient tissue or body fluid.

The process that ends with the discovery of clinical biomarkers and a diagnostic test, begins with the patient. The patient's role is critical in providing the quantity and quality of samples needed to conduct large-scale proteomic experiments on a wide range of cancer types and stages of disease.

We will also need samples from both healthy and diseased individuals, so we can compare their proteins and find out how they are different.

Samples of tissue or body fluid are collected with patient consent, usually as part of routine medical care, including surgery, diagnostic procedures, and medical examinations. They may also be obtained from patients enrolled in a clinical trial.

Fluid samples include blood, serum, plasma, saliva, or urine.

If we can establish the presence of clinical biomarkers in these fluids, they could be used as part of a routine screening for cancers that would otherwise be hard to detect, such as ovarian or pancreatic cancer.

Tissue samples are obtained from a biopsy, or during surgical removal of a tumor. Clinical biomarkers measured in tissue may be used to definitively classify the type of tumor found in the patient.

Often, samples are retrieved from biorepositories, which function like libraries that collect, store, and distribute the samples for biomedical research.

Let's choose a blood sample from the biorepository.

Blood Selected / Tissue Selected

Blood is an excellent source of candidate biomarkers because it comes in contact with tissues and cells throughout the body. Because blood contains both cells and serum, we need to separate those components. In most cases, we’ll search for cancer-related proteins in the serum. After a simple step of centrifugation, the serum is easily drawn from the tube as a clear liquid.

For comparison, let's now choose a tissue sample from the biorepository. With a tissue sample, we can actually purify proteins that come from only a single type of cell. Here's how it's done.

When we get a tissue sample from the biorepository, we cut frozen sections of the sample, and treat it with a chemical stain that helps us to visualize different cell populations.

Laser Capture Microdissection

A significant challenge in identifying the molecular changes that occur in cancer is that the tissue samples often contain a mixture of cells. Laser capture microdissection has elevated molecular analysis to a new level by enabling researchers to harvest *pure* cancer or healthy cell populations.

Laser capture microdissection is an automated system that uses microscopy, cell recognition algorithms, and drawing software to definitively isolate the cells of interest.

The system uses images from a live video camera to view and identify different aspects of the tissue, such as the presence of normal, pre-malignant, or malignant cancer cells.

For example, here is a mixture of different cell types. In this case, we're interested in the areas shown here, because they have features that may represent abnormal or cancerous cells.

With our software-drawing tool, we highlight the sets of cells we're interested in.

Then, we create a field of targets shown here as a series of tiny red circles. Each red circle will become a target for a laser pulse.

The purpose of the laser is to heat a section of the tissue so we can extract it for further study.

Here's how it works:

The key to this process is this tiny polymer cap.

On the tip of the cap is a thin layer of optical quality plastic material that melts from the heat of the laser. This is placed in contact with the tissue sample.

The laser is fired through the polymer cap that is in direct contact with the tissue and adheres to the tissue causing the cells under the target to embed in the material. The polymer cap is then lifted from the tissue. The cells remain embedded in the polymer for further study.

How small are the cells extracted? Can we actually extract a single cell? The answer is yes. The extraction process can remove a sample as small as 4 microns.

This level of precision allows the researcher to selectively retrieve individual cells or regions of tissue to obtain a homogeneous sample, representing a single cell type or disease state.

In the final step, the polymer cap is placed over a tube containing a lysis solution, which breaks the cells apart in order to release their proteins. We are now ready to continue with sample purification.

Sample Purification

Samples derived from serum or tissue can contain thousands of different proteins. The samples may also contain other components such as nucleic acids, lipids, carbohydrates, and salt, which can all cause serious problems in later steps of analysis.

There is no single protocol for cleaning up the protein sample. But researchers may combine a number of different procedures, some of which are shown here [spin concentration; desalting column; antibody affinity column], to eliminate unwanted components and reduce the number of proteins examined in an experiment.

Proteins can be separated from smaller molecules such as salts and lipids by spin concentration, in which a sample is centrifuged through a semi-permeable membrane in a test tube. Smaller components pass through the membrane while the larger proteins stay behind.

In a desalting column, porous micron-sized glass beads can separate proteins, which move around the beads, from smaller molecules which must travel a longer distance through the bead pores.

An antibody affinity column uses beads coated with antibodies to remove and capture highly abundant proteins, making it easier to find other proteins that are expressed in very low quantities.

The end result is a relatively pure mixture of proteins.

Separation and Capture

High Pressure Liquid Chromatography

We successfully purified the proteins from other components in the sample. Now it’s time to separate the proteins from each other.

In the separation stage, we will isolate individual proteins that are expressed differently between healthy and diseased samples.

There are a number of technologies used to separate proteins. One of the more commonly used technologies, high pressure liquid chromatography, separates proteins according to unique characteristics, such as their size, their affinity for water, or the number of charges on their surface.

All of these methods are used often, but let's take a closer look at one method called ion exchange chromatography.

Ion Exchange Chromatography

Ion exchange chromatography allows proteins to be separated according to their charge. This is done using a cylindrical column packed with porous micron-sized glass beads. The beads are coated with a layer of molecules that have either a net negative charge, or a positive charge. Here we show the beads coated with a negative charge.

If a protein has a net charge, it will bind to the beads that carry the opposite charge. For example, if we load a mixture of proteins onto a column full of negatively charged beads, the proteins that have a greater net positive charge will bind more tightly to the beads, while proteins with a neutral or net negative charge will pass through the column more rapidly.

Now, we want to separate the bound proteins based on the strength of their net charge. To do this, we steadily raise the concentration of salts and other charged ions passing through the column, which displace the proteins from the beads. As we increase the ion concentration, the more weakly bound proteins will be displaced first, while the more tightly bound, highly charged proteins will fall off the beads last.

If the proteins fall off the beads and flow through the column, the liquid containing increasingly charged proteins is collected into tubes.

Summary

Usually, a single step of liquid chromatography is not enough to resolve the many thousands of proteins found in a sample. So we may choose to run the samples through another round of liquid chromatography, or proceed to other separation methods.

Two Dimensional (2D) Gel Electrophoresis

A commonly used and powerful technology to separate proteins is two dimensional gel electrophoresis. This method is capable of resolving several thousand proteins on a single rectangular sheet of polymer gel.

Our goal is to sort and resolve the proteins on the gel by their unique physical properties, so we can easily pick out the individual proteins for analysis and identification.

The first dimension of separation is called isoelectric focusing. Here, special molecules called ampholytes are immobilized along the length of a strip of gel and create an environment that varies in pH. At one end, the pH is acidic, which means it has a high concentration of hydrogen ions. As you move along the gel, the pH becomes more basic, where there is a lower concentration of hydrogen ions.

When we apply the sample, proteins will soak into the gel and take on an electric charge depending on their particular chemical composition and structure, and the pH of their environment.

For example, here are two proteins. Both are absorbed into the gel at a location where the pH value is seven. Because of its unique structure, one protein takes on a highly negative charge. The other, a more positive charge.

When an electric field is applied, the protein with a positive charge will migrate toward the negative electrode giving up hydrogen ions to the surroundings until its net charge is zero – where it stops moving. This is called its isoelectric point. Meanwhile, the protein with the negative charge will migrate toward the positive electrode and *pick up* hydrogen ions from its environment until *its* net charge is zero.

Each protein in the sample will move under the electric field until it reaches its isoelectric point, where it stops moving. Because every protein will have an isoelectric point that depends on its chemical and physical structure, we'll end up with a pattern of protein bands along the length of the gel.

Now we are ready to separate the proteins in the second dimension, this time according to size. The isoelectric focusing strip is placed at one end of another, rectangular gel. The gel contains SDS, a detergent that unfolds the proteins into extended amino acid chains, and coats them with a uniform distribution of charge.

When we apply an electric field, the amino acid chains snake through tiny pores in the gel toward the positive electrode. Larger proteins, which have longer chains, become more "entangled" in the pores and migrate more slowly. Smaller proteins with shorter chains can move more rapidly through the pores.

At a precise time, the electric field is removed and the proteins remain in place, now sorted according to their size. The end result is a unique pattern of protein spots.

By comparing two-dimensional gels of healthy and diseased samples, we can identify differences in the spot pattern, such as the one shown here, that may contain proteins associated with cancer.

By removing the spot, we have succeeded in isolating a single protein from a mixture. After extracting the protein from the gel, we have completed the stage of separation.

Analysis and Identification

Overview

We must now analyze and identify the protein by its amino acid sequence.

At the stage of Analysis and Identification, our goal is to determine the amino acid sequence of the isolated proteins. Why is this important? If we can attach an amino acid sequence and a name to the protein, we will enable researchers to share results, prevent duplication of work, and conduct further experiments, including the production of large quantities of the protein and antibodies, which are critical to the development of a diagnostic test.

The most powerful technology used today to identify proteins is mass spectrometry.

Mass Spectrometry

A mass spectrometer is an instrument that can measure the mass-to-charge ratio of individual charged molecules. It's capable of discerning differences of mass as small as a single hydrogen ion.

Can we identify an intact protein by measuring its mass? Unfortunately, no. In the cell, proteins are often modified by the attachment of other molecules. We simply have no way of associating one value of mass with one protein.

So, to identify a protein using mass spectrometry, we often break it into fragments, or peptides, using precision enzymes that cut the amino acid chain at specific locations. If we can identify at least some of the peptides by their mass, we can identify the protein they came from.

Mass spectrometry is made up of three main processes – ionization of the sample, analysis of the charged molecules in an electric field, and detection of the molecules to produce a mass spectrum.

Ionization Overview

The first step in mass spectrometry is to add an electric charge, or to ionize the molecules in the sample, enabling them to be manipulated in an electric field. It's at this stage where the most important technological advances have occurred for modern proteomics research.

In fact, new methods of ionization enabling analysis of large biomolecules, earned the inventors the Nobel Prize for Chemistry in 2002. Earlier, traditional methods of ionization developed for small molecules were much too harsh for proteins and peptides.

The newer "soft ionization" methods are called Electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI).

Electrospray Ionization

Electrospray ionization generates ions by creating a fine liquid aerosol of highly charged droplets.

Evaporation causes the droplets to decrease in size, resulting in even greater concentration of charge on their surface area. Eventually, the repulsion between charges causes the droplets to break apart, and the cycle repeats until only the charged peptides remain and are ejected into the mass analyzer.

Matrix Assisted Laser Desorption Ionization (MALDI)

In Matrix Assisted Laser Desorption Ionization, the peptide sample is first mixed with an excess of a UV light-absorbing matrix compound. The mixture is then spotted onto a metal plate, where the peptide and matrix compound co-crystallize. The metal plate is placed in the mass spectrometer ionizing chamber.

The matrix absorbs energy from a pulsed UV laser, causing both the matrix and the peptide sample to vaporize. The matrix also acts as a proton donor, ionizing peptides so they can be manipulated in the electric field of the mass analyzer.

Mass Analysis Overview

After ionization, the sample is passed through analysis chambers that separate the molecules according to their mass-to-charge ratio. There are several types of chambers available, including quadrupole, time-of-flight, quadrupole ion trap, and fourier transform ion cyclotron resonance.

Let's look more closely at two of these, quadropole and time-of-flight, to see how they work.

Quadrupole Analyzer

Quadrupole analyzers contain four (or sometimes eight) parallel rods, across which an oscillating electric field is applied. Ionized peptides travel through the empty space between the rods, but only peptides with a specific mass-to-charge ratio can reach the detector, based upon the field generated by the rods. All other peptides are thrown out of the original path.

By varying the electric field in the quadrupole, different peptides can be analyzed. Multiple ionizations are required to scan over the entire mass-to-charge range of a sample.

Time-of-Flight (TOF) Analyzer

Time-of-Flight analyzers accelerate all ionized peptides in a sample toward a detector with the same amount of kinetic energy.

The smaller peptides will travel faster, while the larger peptides, with greater mass, will travel more slowly. One can then determine the mass of a peptide by the amount of time it takes for it to reach the detector.

Protein Identification Strategies

We can see that mass spectrometry consists of a number of different choices of components.

There are two major types of ionization, and at least four kinds of mass analyzers, two of which we described in detail. These components can be mixed and matched to suit the proteomic application. So, how does one choose?

The decision is driven by strategy. We can identify a protein in one of two ways:

- We can read its peptide mass “fingerprint” using a MALDI Time-of-Flight mass spectrometer; or

- We can directly determine the amino acid sequence from its peptides using a tandem mass spectrometer.

Let’s take a look at these two approaches in more detail.

Peptide Mass Fingerprinting: MALDI-TOF

First, what is peptide mass fingerprinting?

Each type of protein has a unique amino acid sequence. When we cut the sequence, we used an enzyme that targeted one particular amino acid. This way, a protein will have a unique pattern, or a fingerprint, of peptide fragments depending on the original sequence.

This “peptide fingerprint” can be detected using the MALDI Time-of-Flight mass spectrometer.

MALDI Time-of-Flight combines MALDI ionization with time-of-flight mass analysis. If we start out with a mixture of peptides from an unknown protein, a MALDI Time-of-

Flight spectrometer will provide us with a graph revealing the mass of each peptide in a pattern that is unique to the original protein.

Then, a software program performs a virtual "enzymatic digestion" for all proteins in a database and calculates what a mass spectrum might look like for each protein in the database. This is compared against the experimental result, and a "best fit" between theoretical and experimental results reveals the identity of the protein we are measuring.

Amino Acid Sequencing: Tandem Mass Spectrometry

An increasingly popular method for the identification of proteins involves the direct determination of their amino acid sequence using tandem mass spectrometry.

Tandem mass spectrometry places two analyzers in series separated by a gas-filled collision chamber. For example, the analyzers can be a quadrupole/quadrupole, a quadrupole/ion trap, or, as shown here, a quadrupole/time-of-flight combination.

Our goal will be to determine the amino acid sequence of the protein – one peptide fragment at a time. To help separate the peptides, we will first subject them to reverse phase chromatography, which separates molecules according to their affinity for water. The peptides flow directly from a narrow reverse phase column into the mass spectrometer ionization chamber.

Electrospray ionization is ideal for use with a continuous flow of sample.

The quadrupole analyzer channels the peptide into the gas-filled collision chamber.

The peptide is further broken down in the collision chamber into a series of sequentially smaller amino acid chains, from the intact peptide down to a single amino acid.

The time-of-flight analyzer will generate a graph showing a series of peaks, each one smaller than the next by one amino acid. We can directly determine the amino acid sequence of the peptide by reading the difference in mass between the peaks.

We now have the sequence of one peptide. By repeating this procedure for multiple peptides, we can match it to a protein available in a database.

Summary

We have purified, separated, and now positively identified a set of proteins as candidate biomarkers. With a name and amino acid sequence in hand, scientists can share this information and use it to create other reagents necessary to develop a diagnostic or predictive test.

Clinical Validation

Overview

We showed through laboratory analysis that certain proteins may be associated with cancer. Now, we have to determine whether they can reliably predict or diagnose disease in patients. This stage is called Clinical Validation.

We now want to answer the question: Can the set of cancer-related proteins we have identified diagnose disease accurately, or predict a patient's response to treatment? Before we can actually use these proteins as clinical biomarkers of disease or treatment response, we need to confirm, or validate, that they'll work in a real-world setting.

Clinical Validation has two important steps:

- Development of a test, or assay, to measure the presence and quantity of a candidate biomarker; and

- A clinical trial to determine the assay's predictive value on real patients.

Let's look first at how the assay is developed.

Assay Development

Traditional biomarker assays typically measure only one protein at a time. For example, prostate specific antigen (PSA) is used to detect the presence of prostate cancer. CA125 is used for ovarian cancer. But single biomarker tests might miss the earliest signals of cancer, and might sometimes provide false-positive or false-negative results.

If we wish to detect cancer, a particularly complex disease, from the patient's molecular profile, then we might need to measure a larger set of proteins.

Protein Microarrays

Protein microarrays offer a new way to measure a large set of proteins in a single assay. A one centimeter squared microarray grid may have as many as 1,000 to 10,000 spots, each capable of measuring a single type of protein.

There are three common methods used to capture and measure proteins on the spots of the microarray. These are the Sandwich Immunoassay, the Antigen Capture Assay, and the Direct Binding Assay.

In the sandwich immunoassay, capture antibodies that specifically bind to the proteins of interest are immobilized in each of the spots on the surface of the microarray. A patient sample, such as serum, is applied. The proteins of interest will bind to the antibodies,

while unbound proteins are washed away. The bound proteins are detected using a second, fluorescently labeled detection antibody.

In an antigen capture assay, the capture antibodies are again immobilized in each of the spots on the surface of the microarray. But this time, the proteins from the patient sample are themselves fluorescently labeled.

The third method, direct binding, is conducted by directly binding the mixture of proteins from the patient sample to the spots on the microarray. The presence of specific proteins is detected by labeled detection antibodies.

The use of microarrays is a powerful tool, potentially providing information not only on the quantity of protein in the sample, but also the degree to which it has been modified by the addition of other biomolecules.

If we can look at hundreds of proteins at this level of detail, on a single microarray, we can use this as an assay in a clinical trial to correlate unique protein patterns to evidence of disease.

Clinical Trial

With a protein assay in hand, we can now conduct a clinical trial to confirm whether the assay is statistically accurate as a diagnostic and predictive tool. The design of the trial will hinge closely on the purpose the candidate biomarkers will be used for.

- Will they be used as an early diagnostic tool to determine the particular class of cancer and the stage of disease?

- Are they intended to predict how well a patient might respond to treatment, such as a drug, or a protocol of chemotherapy?

- Will the candidate biomarkers be used as an alternative for other measures of clinical outcome, such as X-rays, magnetic resonance imaging, or biopsies?

In each case, we must compare against conventional methods of diagnosis, treatment selection, and outcome measurement.

Two statistical parameters of the candidate biomarker assay will be critical – its sensitivity and its specificity.

The clinical trial will determine whether the assay is sensitive enough. That is, whether it can identify everyone who actually has the disease. If it correctly identifies, say, only 50 percent of the patients that have the disease, and other methods of diagnosis are more accurate, then the set of candidate biomarkers cannot be considered validated.

The clinical trial will also determine whether the assay is specific enough. For example, if it falsely identifies a large number of patients as having cancer, who do not actually have the disease, then it cannot be used in routine medical practice.

When we have demonstrated that the assay is statistically robust in its ability to diagnose disease or predict treatment outcomes, and that it performs better than conventional alternatives, then we have completed the validation.

A Faster Path to Patients

Improving Proteomic Technologies

The pathway from candidate biomarker discovery to the patient is long, and involves many technologies and procedures.

Today, only a small percentage of cancer-related proteins identified in the laboratory will overcome the hurdles of clinical validation.

If our goal is to change the way we detect and treat cancer through the widespread use of biomarkers, we need to better those odds.

With conventional medicine, by the time patients are diagnosed with lung or ovarian cancer, more than 60 percent already have hidden or known metastases. For these patients, the diagnosis comes too late, and the available treatments are often unable to reverse the course of the disease.

The ultimate goal in the discovery of clinical biomarkers is to bring about the transformation to molecular medicine. And with that transformation, turn cancer into a treatable, reversible condition.

But we are still in the early stages of the effort to discover proteomic biomarkers of cancer. Mass spectrometry, protein microarrays, laboratory automation, and other advances have set the stage. By focusing on improving and adding to the set of technology tools behind biomarker discovery and validation, we can put ourselves on a faster track to the era of 21st century molecular medicine.

* * * *

A production of the National Cancer Institute, Office of Technology and Industrial Relations with Adam Michael Clark, Ph.D., as narrator.

Additional footage and visuals used with permission from Fred Hutchinson Cancer Research Center.

For more information visit proteomics.cancer.gov