LECTURE-26

Interactomics: Yeast two-hybrid immunoprecipitation protein microarrays

TRANSCRIPT

Welcome to the proteomics course. Today we will talk about interactomics, which is studying interactions of proteins and other biomolecules. We will discuss various types of techniques such as Yeast 2-hybrid, immunoprecipitation, and protein microarray.

So as you know, proteins are dynamic molecules, which interact with a wide variety of biomolecules such as lipids, nucleic acids as well as various small drug inhibitors, metabolites and many other biomolecules to provide different type of significant information for physiological actions.

Proteins also interact with one another to form larger complexes. These complexes regulate various fundamental processes such as signal transduction and gene regulation. A detailed understanding of protein interactions provides an opportunity to understand protein functional behavior.

In today's lecture, we will discuss about what is interactomics: different ways of studying interactions such as Yeast-Two Hybrid, Immunoprecipitation and then we will move onto high throughput approaches such as protein microarray.

So first let us start with the interactomics. Interactomics comprises the study of interactions and their consequences between various proteins as well as other cellular components. The network of all such interactions, known as interactome, which aims to provide a better understanding of genome and proteome functions.

Interactions can have different kinds of impacts. It could be positive as well negative. The one which is shown in the slide shows the negative impact. Biology has evolved several mechanisms that regulate interact interactions, such as post-translational modifications and presence or absence of activator or inhibitor molecules. Interactions can also be modulated by altering the expression levels of proteins.

The aberrant interactions, such as the one shown in the slide can lead to disregulation of cellular functions and ultimately, diseases such as cancer may happen.

Proteins interact with a variety of biomolecules. The interaction of proteins with proteins, as well as other biomolecules help to execute their functions. These proteins after interactions, form complexes that modify the proteins, help in protein transportation as well as many more properties in which they are involved.

Interactomics aims to identify the function of uncharacterized proteins so that one can assign the new role of various proteins. The mechanism to regulate protein activity can also be understood by studying the interactions. The network of protein interactions provide very valuable information for processes such as signal transduction and various type of pathways in which these proteins could be involved.

There are different methods of studying interactions because interactions occur in various ways. There are few interactions, which are very transient, that will be for very short duration whereas other interactions could be permanent which may alter the activity. Few interactions are very weak whereas other interactions could be very strong.

Interactions can also be obligate or non-obligate. They can form homo-oligomers as well as different kinds of oligomeric units could combine to form a hetro-oligomer.

This gives you a glimpse of the complexities involved in the interactions because studying the interactions require various types of technologies just because of the diversity of the interactions as well different types of complexities involved.

So what are the physical reasons for these interactions? It could be electrostatic, hydrophobic, steric or hydrogen bonds etc. Earlier in our modules, when we talked about proteins and its different properties we discussed some of these physical reasons.

These proteins could be involved in various biological pathways and form multi-protein complexes such as one shown in the figure here. And by studying the interactions one

can slowly start getting the information for the complex biological pathways in which these proteins can be involved. So these protein interaction studies ultimately help to development the protein interactions network and the wiring diagrams for different type of quantitative information.

Since proteins are involved in mainly binary interactions, an understanding of the underlying biochemistry of proteins and biological interactions is a critical element for the development of novel therapeutics and diagnostics.

Protein-Protein interactions have various potential effects. Few effects are described in the slides such as cataysis, protein inactivation, altered kinetics, alteration of substrate specificity for substrate binding, new binding site information etc. These are just a few examples of potential effects of protein-protein interactions. But they alter multitude of effects, which ultimately lead to either positive effects or dysregulation which may result in various diseases.

So protein interactions are very essential for any cellular mechanisms, whether to talk about signal transduction, DNA replication, translation, cell cycle control, how the metabolic processes are governed, the motility, how the growth and morphology are altered, splicing, transcription etc. I am sure you will be able to add many more examples here, how protein interactions are essential for the activity of cell.

So what is the significance of studying protein interactions? These are very dynamic which interact with a wide various of biomolecules as we discussed previously, such as the biomolecules lipid, nucleic acids, small drug inhibitors and many other type of biomolecules. Proteins also interact with one another to form the large complexes to regulate signal transduction and gene regulation.

Let me describe you few concepts involved in the interactomics field by showing you this animation.

Animation - Interaction studies of proteins with various biomolecules help in deciphering and understanding the functions of various proteins in the complex network of cellular

pathways. Proteins interact with other biomolecules such as nucleic acids, lipids, hormones etc to execute a multitude of functions in living organisms such as signal transduction, growth and regulation and metabolism. These are few examples. There are many other cellular processes and functions in which these are also involved.

Protein interactions with other biomolecules can be of several different types. They may be weak or strong, obligate or non-obligate, transient or permanent. The physical basis for these interactions include electrostatic, hydrophobic, steric interactions, hydrogen bonds etc.

So now let's talk about different methods to study the protein-protein interactions. Understanding the protein-protein interactions provides important clues to the function of proteins. The identification of interactions with known proteins may suggest the functional role played by a novel protein.

There are 2 different ways of studying these protein-protein interactions- one is the traditional approach and then there are a few high-throughput approaches.

The traditional approach has heavily used yeast two hybrid, affinity chromatography and Immunoprecipitation. The more recent include the high-throughput approaches such as protein microarrays and different types of label free technologies. The label free technologies will be described in greater detail in the following lectures and different types of protein microarray platforms will also be elaborated in the following lectures, but today we will focus on the traditional approaches, YST, IP and conventional protein ma. As I discussed, there are many types of microarray platforms which will be covered in the following lectures.

The traditional ways of studying protein-protein interactions involve YST and IP. These are 2 widely recognized technologies which have been used to map the protein-protein interactions on a large scale. These are yeast two hybrid also known as YTH and IP with MS identification also known as IP-MS. Both of these approaches, YST and IP

have been used to identify thousands of novel interactions in different organisms in humans, *C. elegans*, drosophila etc.

Let us talk about IP technique. IP or tandem affinity purification also known as TAP is a technique used to purify protein complexes and study their protein-protein interactions. Depending on the protein that needs to be purified, different types of tags can be attached to the bait protein.

Through the IP method, the purification of protein complexes by IP or tandem affinity purification methods is performed. The target proteins and its interacting partners are isolated from a given complex sample, and then by using by using different kinds of tags these proteins can be isolated from a mixture and then further processed, which I will describe in the next slide.

In the IP method, the antibody specific to the bait is attached to the whole cell extract which forms the complex. Remember, this step is performed in the native, or non-denaturing conditions. The protein-protein complex is immobilized on Protein-A or Protein-G sepharose beads. Protein complex is eluted and further analyzed on SDS-PAGE gel. As you can see on the slide it shows that the antibody is used which is binding with the antigen and the unknown ligands and this complex is separated on a SDS PAGE gel. This condition is done under denaturing condition, whereas the first step was under non-denaturing conditions.

What are the merits of using the IP method? In this method, proteins are isolated in the native state. Why the native state is important? The native state will allow native complexes to be formed. It also allows the formation of post-translational modifications. It is essential to perform these steps in the native or non-denaturing conditions. Interactions by using the IP method are natural and by using this method, large order complexes can be observed because the native state will promote the native complex formation.

What are different demerits of using IP method? It has been reported that frequently, sticky proteins may be picked up-proteins that are not specific could also be eluted, because they are sticky on the other protein surface. It means few proteins that are specifically interacting with target of interest will be isolated along with sticky proteins. It is unclear whether this interaction is direct or indirect because the proteins which are directly interacting will be bind, along with those indirectly bound to the interacting proteins, those will also be eluted. This method is also quite expensive because it needs a good equipment setup and different type of analytical and computational analysis. Applications in organisms in other than yeast have been challenging, but now people have applied this to different types of organisms.

So let's discuss the immunoprecipitation method, how this mechanism occurs in more detail by showing you this animation.

Animation - In immunoprecipitation, the protein of interest is fused with a TAP tag which contains a calmodulin binding peptide, a TEV cleavage site and protein A. Depending upon the proteins to be studied, this tag can be modified. The tag is then bound to a column through affinity interactions between the protein A and IgG. The protein mixture whose interactions with the bait protein are to be studied, is then added. Some of the proteins form a complex with the bait protein through specific binding interactions.

The remaining unbound proteins are then washed away. This is followed by cleavage at the TEV site by the TEV protease to release only the protein complex bound to CBP. These reactions constitute the first affinity step.

In the second affinity purification step, the bait-prey complex is bound via the CBP domain to a calmodulin functionalized column in the presence of calcium ions. The column is washed to remove any other unwanted contaminants after which a chelating agent is added which complexes the calcium ions. Once these are removed, the CBP-calmodulin interaction is weakened and leads to release of the purified protein complex.

Once the protein complex has been purified, the components of the complex are separated by electrophoresis under reducing conditions. The SDS gel is then analyzed and the protein components evaluated thereby providing an understanding about the interactions with the bait protein of interest.

So we have discussed the traditional ways of studying protein-protein interaction. First one was immunoprecipitation. Now let's move on to the next approach, yeast two hybrid. In YST, two types of proteins are involved- a bait protein and a prey protein. Bait protein is protein of interest whose interactions is to be studied. The bait protein is fused with the binding domain (BD) of the transcriptional activator by inserting it and expressing along with the binding domain in a suitable manner. The prey protein is a protein whose interactions with the bait needs to be determined and that is also known as prey protein. The prey protein is fused with the activation domain (AD) of the transcriptional activator. So, the successful interactions of the bait and the prey protein activates the transcription of reporter genes.

The YST system was first demonstrated by scientists Fields and Song in 1989 for studying the protein interactions. Since then, this approach has been widely used in different organisms for various types of biological questions. In general the YST system uses the bait binding domain and the prey activation domain. These hybrid proteins are jointly expressed in the yeast nucleus. If the protein interactions established between the bait and the prey, which are coupled with the BD and AD, then, transcription will occur.

As you can see in the slide, the transcriptional activation for the two-hybrid system consists of 2 protein domains, one is DNA BD, which remains bound to the promoter region and is fused with a suitable bait protein whose interactions with another protein is required for study. The AD is the other domain of the transcriptional activator which is fused with the prey protein. This activation domain when bound to DNA binding domain forms the functional transcription activator and brings the expression of reporter genes.

As shown in the slide the bait protein binds with DNA binding domain and the partner protein or the prey with transcribe activator domain or AD.

Once the BD binds with DNA while the AD activates the transcription when both of these are associated. As you can see in the slide the bait with the BD and prey with AD when they bind together the transcription event occurs.

What are the different advantages of using yeast two hybrid system? As I mentioned, since 1989 when it was shown first time for studying the protein interactions it has been used for various types of interactions including protein-RNA, protein-DNA, analyis of particular complexes as well as studying the large protein interaction networks. The protocol for yeast two hybrid is quiet simple. Unlike the other approaches there is no requirement for heavy instrumentation here. This method also allows to screen large libraries.

What are different disadvantages of using yeast two hybrid system? It provides very high positive and negative rates. So high false positive rate, that has been one of the major limitations of this approach. The proteins must localize and interact in the nucleus. So that is the limitation of the approach because it is restricted to the binary interactions or if you compare with immunoprecipitation the complexes can not be formed. So application in a non-yeast environment is also questionable because the system is mostly used in yeast. The protein folding, if you are aiming for studying the mammalian system, is not guaranteed. And it is also quiet sensitive to the toxic genes. Finally it is limited to pair-wise interactions. If you have fair idea about the molecules which are potentially interacting then you can use the pair-wise study by using yeast two hybrid. But also if you just want to generate a list of potential interactors which may or may not be true then those can be quickly screened by using different libraries by using yeast two hybrid method. Once you have established the list of potential interactors then you can use different types of validation approaches for really establishing how many of those proteins are interacting well and how many of those are false positives.

So the yeast two hybrid screening remain as a novel biological technique which has been used for screening and the discovery of protein-protein and protein-other biomolecular interactions. So let me show you how yeast two hybrid works in the following animation.

Animation - We will now talk about yeast two hybrid. In yeast two hybrid screening, the binding of the transcriptional activator protein composed of binding domain and activation domain to the promoter region is essential for expression of the corresponding reporter gene located downstream of the promoter. The binding domain is fused with the bait protein while the activation domain is fused with the prey protein. Binding of either one of the fusion proteins to the promoter is not sufficient to bring about transcription of the gene.

The two-hybrid screening protocol uses this interaction as the basis for screening for protein interactions. When the bait protein bound with the binding domain interacts with the prey protein fused with the activation domain, there will be expression of the reporter gene which can easily be detected. lacZ is a commonly used reporter gene whose protein product beta-galactosidase cleaves the substrate X-gal resulting in a blue color colonies.

Let me show you the yeast two hybrid concept applied in one of the studies by showing this interaction. The yeast two hybrid screening approach was used by Zhang et al., (2009) to study the protein-protein interactions between the vaccinia virus and human proteins. The virus bait proteins were fused with the C-terminus of the Gal4 binding domain (BD) while the prey human proteins were fused with the Gal4 activation domain (AD) of the transcriptional activator. These interactions were further validated by other techniques which gave a successful validation rate of 63%. So let me shown you by this interaction how this yeast two hybrid screening can be used. Drag and drop the component such that transciption will be carried out. As you can see the first Gal4 domain adhered to the promoter region and now as soon as the activation domain binds there transcription event happens.

After discussing about different types of traditional methods which have been used for studying protein-protein interactions, now let's talk about some of the high throughput approaches which have also been used for studying protein interactions. Though many interactions have been discovered by using yeast to hybrid and immunoprecipitation study, the reports for high false positive rates as well as the poor reproducibility of some of the earlier discussed methods have been a limitation. Aside from these technical issues both the methods, immunoprecipitation and yeast two hybrids are primarily end-point assays that occur in a closed system inside the cells. So modulating the experimental conditions and different types of parameters becomes very challenging. Protein microarrays address some of these limitations such as providing the open system that enables the monitoring the effect of various types of modifications.

So let's discuss some of the high throughput approaches, different types of protein microarray platforms. So as we have been discussing the high throughput genomic and proteomic projects, they capture data at the scale of entire organism and incorporate data into relational databases from which insight into various biological systems, organization of physiological networks can be derived. Different types of hypothesis can be made based on these large data sets. The genomic era has fostered the development of many new methods such as sequencing, SNPs as well as generation of DNA microarrays. The success of DNA microarrays at the time when most of the genes were sequenced it was almost year 2000 and till 2003 when we had availability of all the gene sequences, at that time DNA microarray technology reached to its maximum potential because it was very easy to screen thousands of genes and full genome of any given organism such as human for which almost 30,000 genes were already available. So by using the DNA microarrays scientists have shown the potential of the high throughput genomic technologies. The success of genomic technologies such DNA microarrays have motivated the development of protein microarrays.

Protein microarrays are microscopic arrays which comprise thousands of discrete proteins. Now the concept of microarrays has stirred a great deal of excitement in the proteomics community because it can be applied for several applications such as high

throughput platform, biomarker discovery, protein-protein interaction, functional characterization of different proteins, identification of different substrates, drug inhibitor studies, etc. So once the protein microarray technology is fully realized it promises to enable the study of broad variety of protein features at an unprecedented pace and scale.

The protein microarrays fall into two general broad classes antibody arrays as well as test protein arrays. As shown in the slide the antibody array is an abundance based method which is intended to inform the users or the investigators how much of each protein is present in each sample or to identify the proteins whose abundance is differentially expressed in one sample as compared to the other sample. So antibody arrays, they print thousands of antibodies on the chip surface and it has been used for the measure of proteins or the other biomolecules in different samples to compare the control versus experimental conditions for protein expression. In the test protein arrays the protein are spotted as opposed to the antibodies in the earlier case. This is done by using the procedure that uses the activity of proteins on the surface. The goal of these test protein arrays is to perform functional studies so that different types of functions can be assigned, different types of biological questions related to the protein activity and its function can be studied by using these protein arrays. So these have been used for assaying the protein function, protein interactions studying about small molecule interactions as well as identification of substrates.

As compared to the DNA microarrays which have shown its promises and potential in various biological applications there are relatively few studies published in the protein microarrays. The protein microarrays still remain very challenging just because the generation of the content, the protein, is very challenging. MacBeath at Harvard, first demonstrated the feasibility of printing the protein on the chip surface in a high density array similar to DNA microarrays. In 2000 at the time when DNA microarrays have reached to its maximum potential the proteomics community was still wondering whether similar type of platform can be used for studying the proteins, can we study proteins at a scale by using the protein microarrays? So MacBeath first showed the

concept that proteins can be printed in high throughput platform on the chip surface but he used very limited number of proteins. The limitation of the work was that it was performed on relatively small set of proteins which reflects the difficulty of generating the proteins. we do not have an analogue of PCR which just could amplify the proteins. so protein content generation was one of the major challenges. But thereotically this was concept was shown in the year 2000. The *E. coli* purified proteins were spotted on the chip surface but very few proteins were printed.

Success of this study motivated other scientists to start doing the protein microarrays. Other group, in Mike Synder's lab started doing the yeast proteome chips and they used 5800 yeast clones which were Histidine tagged to screen for the known and novel calmodulin and lipid binding proteins. Now this was full-scale yeast protein array and it showed in 2001 the potential of these chip technologies for protein interactions and different types of other functional applications.

So there are different types of platforms which are available for studying the proteins by using protein microarrays. Let's have a quick look on some of these available platforms.

Antibodies have been used to print on the chip surface for various proteomics application in various types of orientation. The one shown here shows the direct labeling where the target proteins are labeled with fluorescence or other tag which allows detection after it is captured by the antibody immobilized on the array surface.

The approach is the sandwich immunoassay in which the target protein is captured by an antibody followed by detection with labeled secondary antibody as shown in the slide.

In the reverse phase protein blot method the complex mixture such as cell lysates are printed and are probed with the specific detection labels. All these 3 methods, direct labeling, sandwich immunoassay and reverse phase protein blot, rely on antibodies. Obtaining good quality of antibodies and obtaining antibodies at the human proteome scale or any proteome scale remains very challenging. So people have started

exploring different methods of printing proteins on the chip surface for different types of applications.

So very conventional or most widely used method for printing the proteins involve the chemical linkage. The purified proteins are immobilized on the functionalized glass slide and it can be used for various applications. If one can purify large number of proteins then this could be ideal approach for printing the proteins on the chip surface and studying the different types of interactions.

Peptide fusion tags is another approach. Peptides can be synthesized artificially. So the proteins fused to GST, six histidine tag are spotted on. Nickel coated slides have also been used for doing the protein microarrays.

Due to the challenges involved in purifying the proteins or synthesizing the peptides, scientists have also explored the ways to eliminate the protein purification process. Josh LaBaer's group at Harvard developed nucleic acid programmable protein arrays method in which the cDNA containing the GST tag are printed on the array surface along with capture and anti-GST antibody. Protein is expressed by using cell-free expression system and captured by the antibody.

Another cell-free based expression method tried to overcome some of the previously used methods' limitations and it tried to print the cell-free expression system as well as cDNA on the chip surface by using multiple spotting technique (MIST). MIST involves cell-free expression *in situ* expression of the unpurified PCR products and the cell-free lysates are printed on top of this spot so that both *in vitro* transcription and translation can be performed on the chip surface.

So we have seen different types of microarray platforms available. As you can see in this overview slide there are different type of abundance based as well as function based protein microarrays available. We discussed about direct labeling, sandwich immunoassay, reverse phase protein blots as shown in the abundance based methods.

and then we talked about chemically linked, peptide fusion, nucleic acid programmable protein array and multiple spotting techniques in the function based methods.

Protein microarrays have provided high density, high throughput platform which was one of the major achievements of this technology. Very small volume of clinical or biological samples or pharmaceutical samples can be used on these array surfaces and multifunctional assays can be performed. However there are limitations and challenges if using the protein microarray which includes generating the protein content, purification as well as keeping it functional, stable on the bench or in the shelf surface.

So the development of protein microarrays on which thousands of discreet proteins are printed at high spatial density offers a novel tool to interrogate the protein function in high throughput manner. In this animation I will discuss different types of features, different types of processes in the protein microarrays.

Animation - Before we talk about how protein microarrays are generated let's discuss the need for protein microarrays.

Functional analysis of proteins is a time consuming process that requires many steps. Analysis of a single protein at a time would be a tedious and laborious process.

Analysis of several protein samples will undoubtedly take a long time if they are run one at a time. Protein microarrays successfully overcame this hurdle by allowing analysis of several samples simultaneously.

How to express the proteins and purify? The gene coding for the protein of interest is expressed in a suitable heterologous host system such as *E. coli* by means of expression vectors like plasmids. The host cell machinery is used for transcription and translation which results in a mixture of proteins consisting of the target protein along with other host proteins.

Since the protein of interest is expressed along with other proteins native to the host, it is essential to purify the target protein before it can be used for microarray applications.

This can be done by chromatographic procedures to obtain the pure target protein. Protein purity is tested on SDS-PAGE gels. Tags like His6 are often fused with the protein of interest to facilitate the purification process due to its specific affinity towards nickel.

Array functionalization: The array surface is functionalized with a suitable chemical reagent that will react with groups present on the protein surface. Aldehyde and silane derivatizations are commonly used as they interact well with amino groups present on the protein surface resulting in firm capture of the protein.

Robotic array printing: The protein solution is printed on to the array surface in extremely small volumes by means of a robotic printing device that has small pins attached to it for this purpose. The slides are kept for a suitable duration following the printing step to allow capture of the protein on to the array surface. The unreacted sites are then quenched by a blocking solution such as BSA which also prevents any non-specific protein binding in subsequent steps.

Types of arrays: There are two types of protein arrays that are commonly used. In forward phase arrays, the analyte of interest such as an antibody or aptamer is bound to the array surface and then probed by the test lysate which may contain the antigen of interest. In reverse phase arrays, however, the test cellular lysate is immobilized on the array surface and then probed using detection antibodies specific to the target of interest.

Protein detection and analysis

In the direct labelling detection technique, all the target proteins are labelled with a fluorescent or radioactive tag that facilitates easy detection upon binding to the immobilized capture antibody on the array surface. In the sandwich assay, however, a fluorescent tagged secondary antibody that recognizes a different epitope on the target antigen binds to it and is detected by means of the fluorescence.

The protein microarray is then scanned at in a microarray scanner that allows detection of the fluorescently labeled proteins or antibodies. The output from this scanner is then received by an appropriate software on which the data can be analyzed. Certain well-characterized proteins can be printed on the array as shown in this animation. Now a proof of concept array is shown here where well characterized proteins are printed on the array surface along with their corresponding query molecules shown on the left labeled with different fluorescent dyes. Now by using this interactivity let's match the protein interacting pairs, Jun & Fos, p53 & MDM2, by dragging the query to the correct protein on the array surface in order to see the signal output. So on the array surface there are both p53 and Fos proteins present, now if we drag the Jun protein it should interact with the Fos protein as you can see by the interactions here. MDM2 proteins interact with p53 protein. Once this interaction is established then these signals can be detected by using a scanner.

So in summary today we discussed the significance of studying interactions in the field known as interactomics. We discussed different types of traditional methods which have been used for studying these protein interactions such as yeast two hybrid, immunoprecipitation assay. Then I have shown you a variety of platforms currently being used of studying proteins in the high throughput mannersby using protein microarrays. There's different type of advancement happened in the protein microarrays such as use of cell-free expression based methods as well as coupling the protein microarrays with the label-free detection systems. I will discuss some of these advancements and techniques in the subsequent lectures.

Thank you.