LECTURE-33

APPLICATIONS OF PROTEIN MICROARRAY

TRANSCRIPT

Welcome to the proteomics course. In today's lecture we will talk about applications of protein microarrays. The protein microarray technology has potential to become one of the very powerful tools in proteomics field. Due to its enormous potential in diagnostics and drug discovery these high density protein arrays which are printed on the glass arrays have been used for proteome wide analysis, whether it's human, yeast or different type of bacterial species these wide variety of microarrays have shown wide variety of applications and shown great potential for high throughput applications.

In today's lecture first I will give you an overview of protein microarray technology which we have already discussed in previous lecture. I will then give you an overview of how to perform protein microarray based experiment, which is general can be applied for any type of application whether you want to study for biomarker screening protein-protein interaction. After discussing the overview and general strategies for protein microarray experiments we will then talk about different applications, I will give you case studies of studying biomarker identification, different type of immuno screening as well as protein interactions as well kinase substrate identification.

So as you know proteome is very dynamic and it represents very high complexity. Whether it's study at genome level, transcriptome level or proteome level as well as metabolome level all of these omics approaches require high throughput techniques and high throughput platforms. Now that's why proteomics field on one hand aiming towards studying the various types of biological problems simultaneously people are trying to integrate various type of technology so that a robust high throughput platform can be generated to study the dynamic proteome. As you can see in the slide these dynamic proteome are quite linked with physiological actions happening in biological systems and that's why we need to integrate many of those to obtain a comprehensive

image of what is happening inside the living organism. That's why when we want to study the protein we cannot study those in isolation we need to study these proteins at the high throughput level so that not only we study a given protein in a context but also we know how that protein interacts in that biological system. So to answer many of these questions protein microarray platform and as well as some other proteomic techniques have been generated and shown potential for high throughput application.

In the previous lecture I have given you and overview of different type of protein microarray technology we had discussed both abundance based and function based protein microarray platforms. We have talked about different type of labeling including direct labeling, sandwich type of assay, we have talked about reverse phase arrays and then we discussed different ways of making function based protein microarrays, which includes immobilizing the proteins which is purified or peptide fusion with the tags as well as different type of cell-free expression based approaches. With this refresher I hope now you remember that there are many kind of protein microarray platforms which can be used for addressing different type of biological questions.

So let me give you an overview of protein microarray experiment. Because regardless of the kind of application which you aim for you have to perform certain steps so that you can use these protein chip for addressing your biological question so as shown in the slide first of all you need to take the open reading frames the cDNA clones for which you want to make the proteins those are nowadays available in many repositories, you have those vectors and then either you need to purify the proteins. Most of the time, you have to do this in 96 well plate format, in the more high throughput protein production system because you cannot rely on individual tubes for purifying the proteins for high throughput applications. Alternatively, one can also use cell-free based expression system to generate the protein content. In either way once you have generated the protein content one can print those proteins of the glass slides which is shown on the right side and then assay performance can be tested on these chips by variety of chemistry. Once you are ensured that a good quality of chip is ready then various types of applications and biological questions can be proved on these chips.

So let me show you one video and give you an overview of this process of showing the protein microarrays how that experiment works. So here I am giving you an overview of protein microarray experiment regardless of what application one want to study the overall work-flow of microarray experiment remain the same. Here I am showing you an experiment of protein interactions performed on the E. coli proteome chip. So first of all these chips are stored at -80°C if they are purified proteins you want to store them at the very freezing condition so that protein can remain functional. Now these chips should be carefully removed from the freezers and then allowed to thaw briefly followed by the washing steps. So first of all transfer these chips from the -80°C to the slide holder. After these chips are thawed briefly at the room temperature then one can either directly block it in the blocking solution or use the PBS tween for brief rinsing. Blocking is usually performed at room temperature for an hour or at 4°C overnight depending on different type of applications. PBS and milk or superblock or BSA are commonly used blocking solutions. One can typically use even a pipette tip box or a small box with approximately 30 ml or 50 ml blocking solution, dip the slide inside the solution and once the blocking is completed you can remove the slide from the blocking solution and tap against a paper towel to remove the excess milk. While performing the blocking experiment or blocking step make sure continuous mixing of the slide even if you are preparing the entire set. If slides are left sitting on their sites without mixing then they will dry and will appear dark background after the screening.

After blocking steps are completed then depending upon your application you can either apply primary antibody if you want to do quality control check for example or one can use a query protein for example if you want to study the protein-protein interactions. So this study we are talking about protein interactions so let's say a query protein of interest for which you want to study the protein interactions, you can take that query protein and then apply that on the microarray chip slide, after addition of this query protein you need to cover it with the cover slip and incubate it at the room temperature for an hour or one can optimize this condition depending upon their experimental requirement. Once this step is done then you need to wash the slide with PBS tween for

three times, usually 5 minutes wash for the 3 times at the room temperature with a gentle shaking on a rocker shaker is most commonly used. In the microarray experiment one need to ensure the proper washing and gentle shaking throughout various steps to ensure that slide are washed very neatly. Otherwise you will see very high background on the slides. After this step one can add secondary antibody labeled with the either cy3 or cy5 conjugate labels or one can use the HRP based detection system for detecting the signal.

But prior to this step one need to dry the chip by centrifugation. One can also use compressed air for drying these slides. Once the slides are dried then they can be scanned at appropriate wavelength. This just gives a glimpse of an overview of microarray experiment which is shown in the 3D animation here but depending upon your biological question various type of modifications can be made.

After giving you an overview of protein microarray experiment, let's now talk about protein microarray based applications.

So protein microarray chips have been applied to assay for a wide range of applications such as protein-protein interactions, protein-lipid, and protein-nucleic acid including protein-DNA and then protein-small molecules or protein substrate interactions. Protein chip has also been used for drug and drug target identification as well as kinase substrate identification. I will try to give you few studies in more detail as well as few studies to just make you familiar with what are different types of applications which one can use for protein microarrays.

So in general protein microarrays have been widely used for biomarker identification and protein-protein interaction as well as various types of protein modifications. As you can see in the slide the protein interaction networks transcription regulation, DNA damage repair, antiviral response, drug target identification etc. are various type of applications which people have tried on the protein chips similarly studying about the kinase networks, the dynamic functional regulation as well as studying protein turn over and protein modification have been used on protein chip.

So let's first move onto the biomarker detection. This is one of the most commonly used applications which people have tried using protein chips. Because the clinical samples are very limiting you usually do not have access of large amount of clinical samples from the patients. On one hand one would like to study as many proteins as many possible by using these protein chips or any high throughput proteomic technology, on the other hand you have limitation of total amount of sample so in that regard protein microarray offer a very appreciable platform because you can use a few micro liters of the sample and that can be used to probe thousands of proteins. So on one hand you are using very small amount of sample then you can actually probe that for large number of proteins. Often some of these things are very challenging for other proteomic technology that's why protein microarrays have offered many applications for biomarker discovery.

So let's start with the case study 1 "the identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays"- study performed by Hudson et al.

In this study authors have taken serum from 30 individuals suffering from the ovarian cancer and 30 healthy individuals. For each of these types of samples they used 300 µl of the serum sample, diluted that in correct buffer and then applied that on human proteome arrays. The proto-arrays are obtained commercially so they wanted to compare the antibody response of various types of proteins in ovarian cancer serum as compared to the healthy controls. The identification of tumor associated antibodies and targeted protein antigens were performed on protein microarrays. These protein microarrays contain more than 5000 GST fusion proteins, which were first probed with anti-GST antibody. As you can see in the slide, first of all quality control experiment was performed by using anti-GST antibody because all the clones contain GST tag so first of all there was a need to ensure that all the clones express protein properly so uniform amount of proteins is printed. Now once the quality control experiment is done then one can actually screen the different types of biomarkers by using these chips. So authors used the sample from ovarian cancer and the healthy controls and then applied it on the

human proteome chip. Now the different boxes are shown on the bottom panel on the right hand side which indicates there are several positive and negative controls which were printed on the chips to ensure that assay is working fine and then there are certain proteins shown in the centre which are potential markers.

So based on this study, authors were able to identify several differentially expressed proteins which included Lamin A which is one of the nuclear membrane organization protein. Structure specific recognition protein (SSRP1) which is involved in the regulation of transcription, RAL binding protein (RALBP1) which is important in the transportation, IRF6 is crucial for transcription regulation, MEGEB4 is reported as a cancer marker, COIL which is found in the nuclear coiled body although its function is unknown. NOB1P which is adeno-carcinoma antigen the function of this protein is unknown, then CBLB which is involved in signal transduction. So by using high throughput approach high throughput microarrays and screening for more than five thousand proteins, authors were able to identify several differential expressed proteins, these are few of those which they consider quite interesting and selected few of these for further validation. So once the protein targets are identified authors used immunoblot and western blot for doing the validation. The protein sample of cancer and the normal cells were analyzed by using immunoblot assay and they used antibodies specific for Lamin A or C proteins and SSRP1. These proteins were considered interesting based on their proteomics findings. The Lamin A and C were greatly elevated in the cancer sample as compared to healthy control that appeared quite interesting although the western blot analysis was performed by using health and deceased ovarian samples and then they used antibodies specific for Lamin proteins and as well as p53. As you can see on the right hand side of the slide, the Lamin A and C were probed by using western blot similarly p53 protein was probed by using anti-P53 antibody so these immunoblot assays and western blot analysis confirmed and validated that the proteomic finding base on the protein microarrays were quite relevant and the elevation of these proteins were validated by independent techniques.

Now further to the immunoblot assay authors also used tissue microarray based

analysis so the tissue microarray analysis of different stage of ovarian cancer was performed. These microarrays containing the representative tissues from various stages of ovarian cancers such as stage 2, stage 3 and stage 4 tumors were probed for Lamin A or C, SSRP1 and cancer antigen 125. So these results are shown in the slide which is redrawn from the results presented the manuscript. The Lamin A and C shown in the top panel, SSRP1 shown in the middle panel and CA-125 in the bottom panel so overall this was a impressive study which revealed that the three candidate tissue markers which were immune-stained, they can produce a very robust signature of ovarian cancer in the tissue sections. Although, authors have used serum and the motivation was to develop the serum or blood based assays. They thought that at least the robust signature in the tissue indicates that these proteins are overly present in the serum sample and probably this type of markers can be used for blood based or serum based assays.

Let's now move on to case study-2. The identification of novel serological biomarkers for inflammatory bowel disease using E. coli proteome chip, Crohn disease and ulcerative colitis these are chronic, idiopathic and clinically hetropathic intestinal disorders which are collectively known as inflammatory bowel disease also known as IBD. In this study authors have used E. coli proteome chip for screening and identification of novel biomarkers associated with inflammatory bowel disease, study was performed by Chen et al., in 2009. So in this study first of all authors have used E. coli proteome chip, they used more than 4200 proteins obtained from E. coli and then they collected serum from healthy controls and clinically well characterized patients with inflammatory bowel disease, 66 samples from the Crohn disease and 29 from the ulcerative colitis were collected and compared with the 39 healthy controls. The protein spots that were recognized by the serum antibodies were visualized and quantified by using cy3 labeled goat anti-human antibody. After looking at the schematic or the overview of the experiment, now let's look at the some of the results. So these E. coli proteome chips were probed with the sera from CD patients which is shown in the left side which is the Crohn's disease and healthy controls is shown in the right panel in this

slide. The cy3 labeled anti-human immunoglobulin antibodies were probed on the chips which allows the visualization of immunoreactive protein spots. Some of these spots are shown in the center and it shows the comparison of CD versus healthy controls. Now after doing this screening on the chip, author generated a heat map of 273 differentially expressed proteins, which are identified with the healthy control and the Crohn's disease sample. They also performed a comparison of UC, HC and CD which are shown in slide in venn diagram on the right side. It shows that differentially expressed proteins showed very limited overlap among the HC versus CD and CD versus UC. After studying these differentially expressed proteins, authors used these proteins to understand their functional role and how they are distributed in the cellular components. So HC, CD and UC were then further used to define the cellular components and functional role which is shown in this slide based on the membrane, the cell wall, the macromolecular complex, intracellular and periplasmic space and the cell projections they tried to categorize the proteins in all the three groups and looked at their differential response. Further, by using a supervised learning algorithms KTOP scoring pairs, authors identified two sets of serum antibodies that were novel biomarkers for specifically distinguishing Crohn's disease from healthy control.

So as you can see in the slide that healthy controls, Crohn's disease and ulcerative colitis all these samples can be distinguished and especially the healthy control versus Crohn's disease cross by using supervised learning algorithm.

After studying the applications one based on the biomarkers another based on E. coli proteome chip. Now let's look at some of the biomarker discovery in more quick way just so that we have more time for looking at different type of other applications. So I will try to cover some of these applications in brief and just to give you an overview of different type of applications happening on these protein microarrays.

So in this study, we will talk about "identification of potential diagnostic markers for infection from *Neisseria meningitides*. *Neisseria* is most common cause of meningitis disease and also causes epidemic outbreaks. To investigate the immune responses to

the phase variable expressed proteins, Steller et al., applied protein microarrays to screen the meningitis patient serum. So this is actually the first study which aims to investigate the genetic phase variations in the pathogens. So in this study authors first amplified all the 102 known phase variable genes, obtained from the *Neisseria meningitides* and expressed and purified these proteins in the heterologous system in *E. coli.* They were able to purify 67 recombinant proteins because not all the proteins were able to produce the proteins so 67 were successful and these were further used to generate protein microarrays. Now these protein microarray platform was used to screen 20 patient serum as well as healthy controls. After screening authors were able to identify 47 immunogenic proteins out of which 9 proteins were quite reproducible including a phase variable opacity protein OPAV which was very reproducible in many patients.

So let me give you the overview of the study in the following animation. This study was performed by Steller et al., Bacterial protein microarrays for the identification of potential diagnostic markers for infection from *Neisseria meningitides* infections. Authors amplified and subcloned 102 genes from *N. meningitidis* for expression in *E. coli*. These clones were grown for overnight at 37°C in antibiotic containing medium, following which protein expression was induced by addition of isopropyl-beta-D-thiogalactoside (IPTG).

The cells were harvested four hours after induction and then protein was purified. The proteins were purified based on the specific Ni-NTA binding followed by analysis on SDS-PAGE to check the purity. This shows the SDS-PAGE separation of the proteins based on the molecular weight. The 67 purified proteins obtained were then printed on nitrocellulose coated glass slides using a robotic arrayer. The array was probed with sera from 20 convalescent patients by incubating it overnight at 4 degree centigrade. The array was then washed with PBS and detection was carried out by means of Cy5 labeled secondary antibody. The excess detection antibody was washed off and the array then dried and scanned by microarray scanner. Authors detected 47 immunogenic

proteins, one of which showed response in 11 of the patients. These protein microarray platform was successfully used for detection of several other disease biomarkers and this is one of the applications which is shown here.

Let's now move on to case study-4. The human prostate cancer screening for the indentification of potential biomarkers study by Miller et al., so in this study authors used antibody microarrays containing 184 unique antibodies which were printed on the microarray surface and they used two different type of substrates containing polyacrylamide as well as poly-l-lysine coated glass slides and further they used this platform to screen the prostate cancer biomarkers. So in this study authors used 33 cancer patients and 20 controls and obtained serum sample from these subjects and employed these to study the abundance of various proteins present on the microarray surface. First they optimized various type of parameters for microarray measurements and once these conditions were optimized then they used this system for identification of various potential biomarkers. Let me show you this study in brief.

Authors used robotically spotted 184 unique antibodies on polyacrylamide based hydrogels and poly-l-lysine coated glass slides, which they probed with sera from prostate cancer patients as well as healthy controls. Let me use this interactive animation and show you the results obtained for the hydrogel slides. Left slide is probed with the prostate cancer sera and the right one is probed with control group sera. From this study 5 proteins were shown to have significantly different expression levels in prostate cancer patient samples compared to the control group. vWF protein was found to be elevated and the remaining four proteins were found to be down-regulated as compared to the control group.

So after looking at various type of applications for biomarker discovery including two studies which we discussed in more detail and two studies which I showed you in the animation. I hope you able to get a feel of various types of assays, which people perform to identify biomarkers on protein microarray platform. Now let's look at other

application protein interactions, again this is very widely applicable and widely used assay on the protein microarray surface. Different type of interactions, including protein-protein, protein-DNA, as well as with other bio-molecules have been probed on the protein microarray surface.

So let's discuss case study-5. The development of protein microarrays and their applications in DNA-protein and protein-protein interactions, analysis of Arabidopsis transcription factors studied by Gong et al. So transcription factors play a very crucial role in cellular and developmental processes in organisms including plants. Authors generated an ORF expression repository and then used this system for generating the protein microarrays based on the transcription factors of Arabidopsis.

So this slide shows you an overview of steps involved in generating TF Arabidopsis protein microarrays. So from Arabidopsis they have generated the ORF repository of transcription factors and these clones are transformed to the various type of recombinant vectors which were further used to purify the protein. Once the proteins are purified they were printed on the chip surface and used for the different applications including protein-protein interaction and protein-DNA interactions.

So this slide shows you the Arabidopsis transcription factor protein microarrays the left side is used as a quality control check to show that all the proteins expressing well on the chip surface so authors used anti-His antibody because all these clones contain the His-tag and the right side shows the western blot image because they wanted to first QC each of the clones expressing the proteins. Authors first defined the studies for the detection of protein-protein and protein-DNA interaction. As you can see in this slide the left panel shows the sandwich assays which they developed for the protein-protein interactions. Where the first immobilized transcription factor proteins which were used to binding proteins or the DNA labeled with the biotin tag and then further streptavidin and Cy dye chemistry was used to detect the signal. Similarly right panel shows assay conditions for DNA-protein interactions. Now based on protein DNA binding studies, In

this slide as you can see the authors examined 82 ERF transcription factors and their cognate cis elements and they reported that protein microarray provide very efficient and high throughput platform for genome wide screening for TF-DNA interactions. Further they used the protein chips for detecting the protein-protein interactions and they discovered four viable protein partners, which interacted with HY5.

Let's now discuss case study-5. Analysis of yeast protein activities using proteome chips. In this study, Zhu et al., cloned 5800 ORFs from yeast and over-expressed and purified these proteins to build the yeast proteome microarrays.

As you can see in this slide, yeast proteome microarrays was built after each of these clone were expressed and purified then further printed on the chip as shown on the left slide and then it was used to screen the interaction of the proteins and phospholipids. This study revealed that there are many new calmodulin and phospholipid interacting proteins. Let's discuss this in the following animation.

The study was performed by Zhu et al., "Global Analysis of yeast protein activities using proteome chips". Authors generated a yeast whole proteome array by expressing 5800 purified proteins on a single nickel coated slide. The chips were probed with anti-GST antibodies to determine the reproducibility of protein expression and immobilization. This shows uniform signal across whole proteome array. The signals were observed for more than 90% of protein samples and over 90% of the features contain 10 to 950 femtogram of proteins. To understand the potential applications of such whole proteome arrays, the authors screened the immobilized proteins for protein-protein and protein-lipid interactions. They used biotinylated calmodulin in presence of calcium and phosphoinositide liposomes, respectively. The detection was carried out by using Cy3 labeled streptavidin. Arrays were washed, dried and then scanned by using microarray scanner. So 6 of the known calmodulin targets and 33 potential partners were identified with 14 of these proteins possessing a consensus sequence. The phosphoinositide liposomes were able to identify 150 protein targets, of which 45 were found to be

membrane associated, predicted to have membrane spanning regions. This study testified the tremendous potential of using whole proteome arrays in identifying new potential targets.

After discussing two broad categories of applications based on the biomarker discovery and protein interactions, let's briefly look at some other applications, which people have used on the protein microarrays.

Case study-6 "detection of antigen-antibody interactions at various concentrations using antigen and antibody microarrays" a study performed by Haab et al. In this study authors used protein microarrays to measure the abundance of many specific proteins in complex solutions, the protein microarrays can provide a practical mean to characterize pattern of variation as you can see in this slides and then hundreds and thousands of proteins can be probed on these type of platforms. Let me show you details of the study in this animation.

In this study, Haab et al. used protein microarrays for high parallel detection and quantitation of specific proteins and antibodies in complex solutions. Authors printed six arrays of 114 different antibodies on to poly-L-lysine coated glass slides by using a robotic arrayer. These were used to analyze interactions in six unique antigen mixtures ranging in concentration from 1.6 micro gram/mL to 1.6 nano gram/mL. The antigens were tagged with Cy3 and Cy5 fluorescent labels. After, the antigen-antibody binding reaction was complete, excess unbound antigens were washed off using phosphate buffered saline and water at room temperature.

Once the excess antigens were washed off, the bound antigen-antibody interactions were detected by means of a microarray scanner at wavelengths of 532 nm and 635 nm. The authors found that such microarrays of antibodies could detect their corresponding antigens at concentrations as low as 1 nano gram/mL. In a complimentary experiment, the authors generated six antigen arrays having 116

different antigens, which they probed with Cy3 or Cy5 labeled antibodies of varying concentrations. The antigen-antibody binding reaction was allowed to go to completion and excess unbound antibody was washed away using PBS and water at room temperature. After washing and drying, these slides were scanned at 532 nm and 635 nm. It was found that these antigen arrays allowed detection and quantitation of antibodies down to absolute concentrations of 100 pg/mL. These detection limits can further be improved by using high affinity and purity antibodies, thereby it demonstrated great promise for high throughput and sensitive clinical applications.

Let's discuss now case study-7. A proteome chip approach reveals new DNA damage recognition activities in E. coli study by Chen et al. In this study first of all, authors generated the E. coli chip and to do that they purified over 4200 proteins from the E. coli K12 and then developed assays for identification of protein interactions which are involved in the recognition of potential based damage in DNA. So this slide shows you the overview that how different type of mutated proteins or protein activity can be identified by using this type of protein microarray platform. So first of all, to give you an overview of the high throughput approaches involved and various steps which people perform for generating the protein content. This slide shows you that how different type of proteins can be made in the high throughput platform, this is the E. coli proteome chip manufactured and to do that authors have used the 96 well format for high throughput E. coli protein expression and purification. But high throughput approaches require lot of quality control and that is why quality and quantity of the purified proteins were determined by the commassie staining and the western blot by using anti-His antibody as shown in this slide and then they used the anti-His monoclonal antibody followed by cy5 labeled secondary antibody to show that quality of the chip is good for performing the further assay.

So in this study by using a group of DNA probes as shown in the slide here, each containing a mis-matched base pair or an abasic site authors found a small number of proteins that could recognize each type of probe with high affinity and specificity. So

they evaluated two type of proteins, one YbcZ which is shown in this slide and YbcN which is shown in the following slide.

After looking at various applications, let's cover the last application. Case study-8 "protein phosphorylation study in yeast" study by Ptacek et al. 2005. So in this study authors cloned more than 5800 ORFs of yeast, over-expressed and purified these contents and then build the protein microarrays for yeast chip. This slide gives you an overview of process involved in the identification of kinase substrates using protein chips. As author defined overall scheme to identify the kinase substrates. First of all, each kinase was over-expressed, purified and assayed on the chip and it contain over 4000 protein spots which were used for kinase substrate identification. These type of high throughput approaches can further be used to build the map and to show the connection between kinases and substrate.

So in summary, microarrays have been used for surveying both antigen and antibodies and have been used for various type of applications including in human, yeast, E. coli and other organisms. I have tried to cover few applications based on protein microarrays however there are numerous applications which I could not cover due to the time limitation. But it just gives you a glimpse of how these type of high throughput approaches could be used to address various type of biological question. So in summary today we talked about and overview of protein microarray experiment, I refreshed you about different protein microarray platforms which we discussed in the previous lecture and then we talked about different type of applications including biomarker discovery, protein-protein interaction, protein-DNA interaction, DNA damage study, kinase substrate identification etc. we will continue our discussion in the following lecture on microarray data analysis and challenges associated with high throughput data. Thank you.