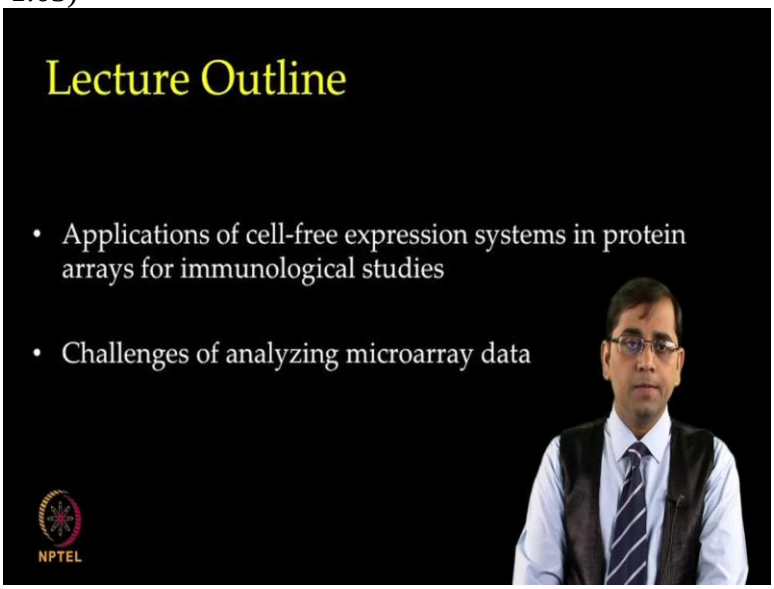
Interactomics: Protein Arrays & Label Free Biosensors
Professor Sanjeeva Srivastava
MOOC NPTEL Course
Indian Institute of Technology Bombay
Module 7
Lecture No 3
Application of cell-free expression protein microarrays in immunological studies

Welcome to the mooc interactomics course. The combination of proteomic technologies, especially the protein microarrays have potential to be applied for wide variety of biological applications. The application of cell free based protein microarrays have seen rampant increase because of ease of synthesizing the proteins by using cell free expression based system as compared to the cell based traditional way of protein purification and then printing on the array surface.

We discussed this aspect in our previous lecture when we talked about use of cell free expression system for biomarker discovery.

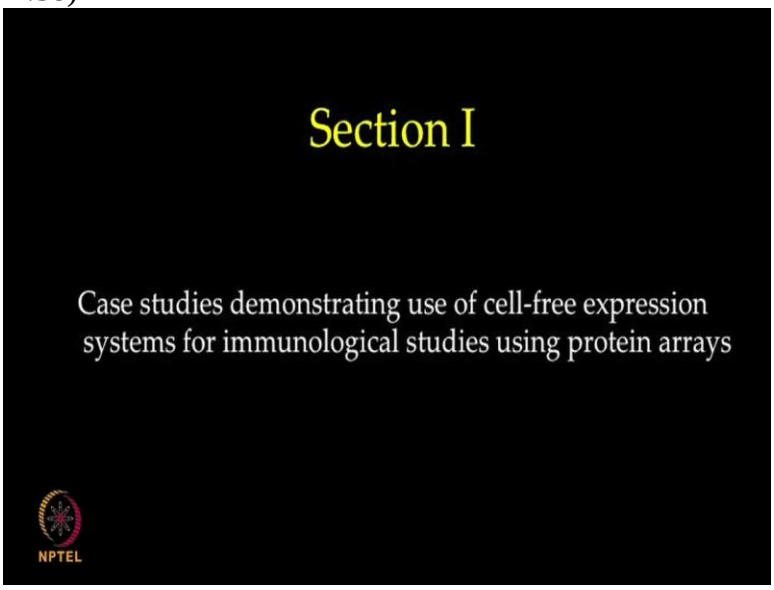
(Refer Slide Time: 1:03)



Today we will continue our discussion on applications of cell free expression system and other biological goals. They help us to achieve using few case studies. Finally we also touch upon the challenges of analyzing the microarray data regardless of which experiments you perform and what biological questions you may want to address.

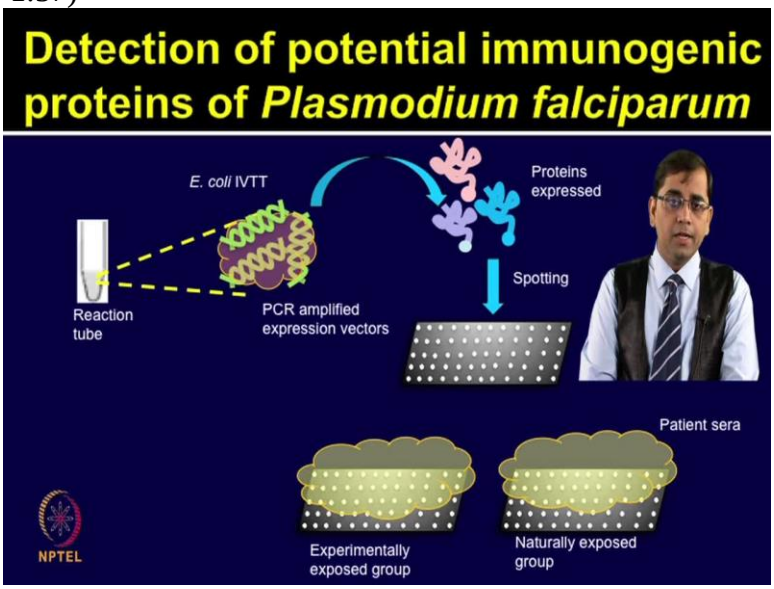
However the data analysis becomes very crucial and very challenging in microarrays experiments.

(Refer Slide Time: 1:38)



The first case study for today's detection of potential immunogenic proteins of plasmodium falciparum. The study performed by Dooley Nichol in 2008.

(Refer Slide Time: 1:57)

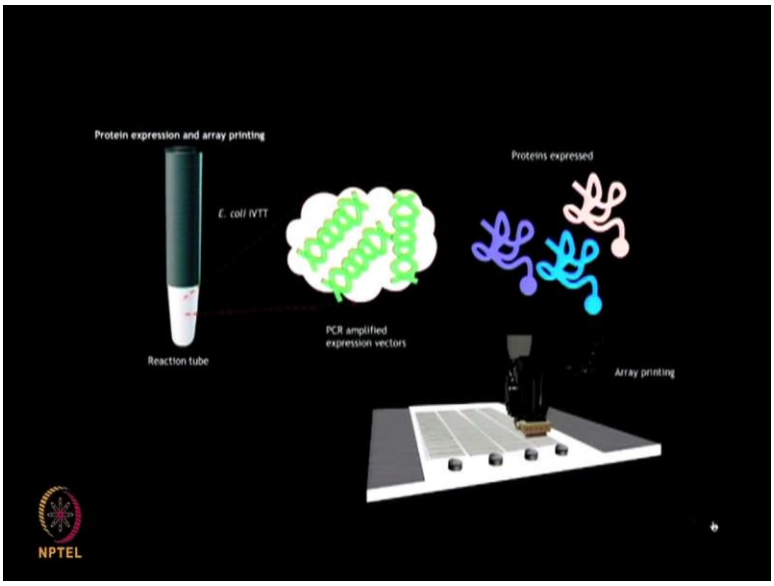


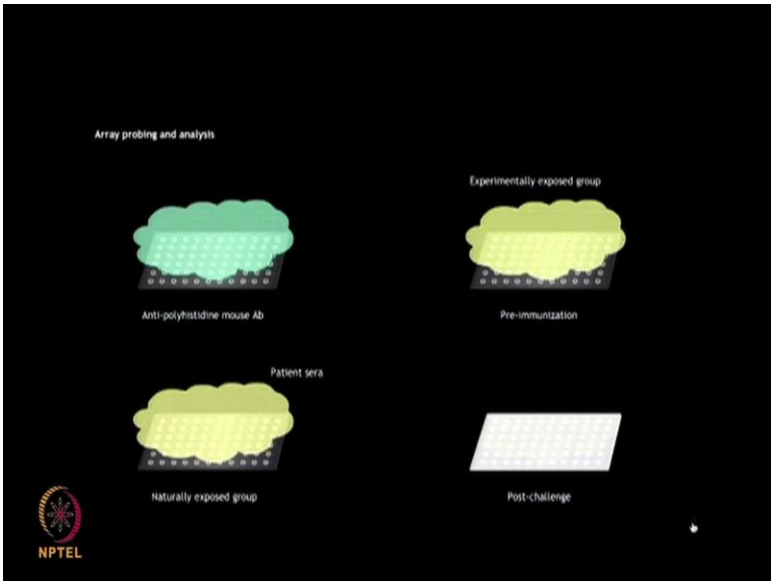
Dooley Nichol used E.coli based cell free in vitro transcription and translation system to produce 250 plasmodium falciparum generated by the polymerase chain reaction and recombinational cloning procedures.

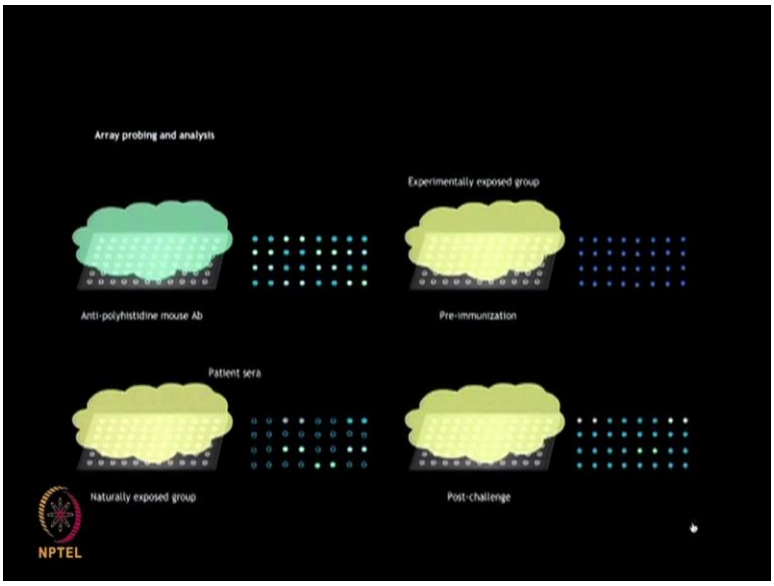
After synthesizing the proteins from these two hundred fifty open reading frames authors profiled antibodies that developed after natural or experimental infection or after the vaccination with the attenuated organism. These are exposed to the plasmodium falciparum either naturally or experimentally and were screened by using protein microarrays.

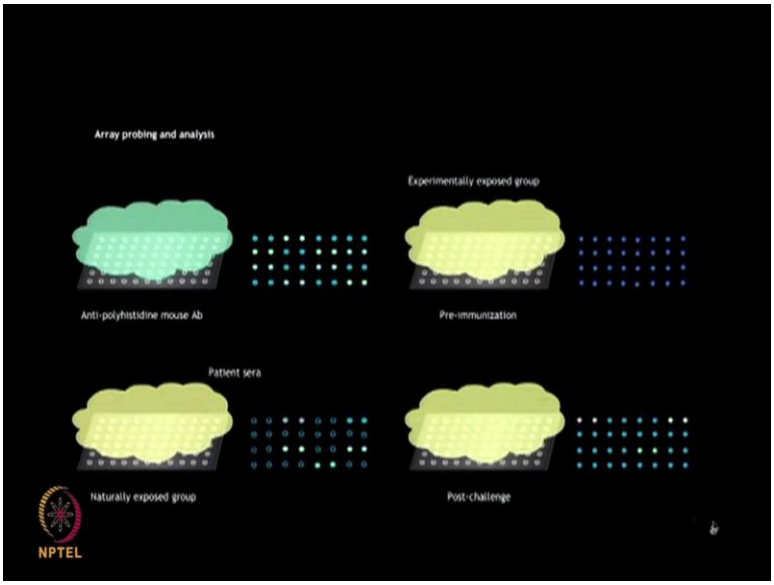
In this study they identified 72 highly reactive plasmodium falciparum antigens. The proteins express specifically in pre (())(2:56) stage of plasmodium which was CSP as well as some liver stage specific antigens such as LSA1. They also identified successfully several proteins by applying cell free expression based protein micro arrays. Let us now discuss this experiment by looking at this animation.

(Refer Slide Time: 3:23)


Protein expression and array printing
E. coli IVTT
Reaction tube
PCR amplified
expression vectors
Proteins expressed
Array printing
NPTEL


Array probing and analysis
Anti-polyhistidine mouse Ab
Experimentally exposed group
Pre-immunization
Patient sera
Naturally exposed group
Post-challenge
NPTEL


Array probing and analysis
Experimentally exposed group
Anti-polyhistidine mouse Ab
Pre-immunization
Patient sera
Naturally exposed group
Post-challenge
NPTEL


Array probing and analysis
Experimentally exposed group
Anti-polyhistidine mouse Ab
Pre-immunization
Patient sera
Naturally exposed group
Post-challenge
NPTEL

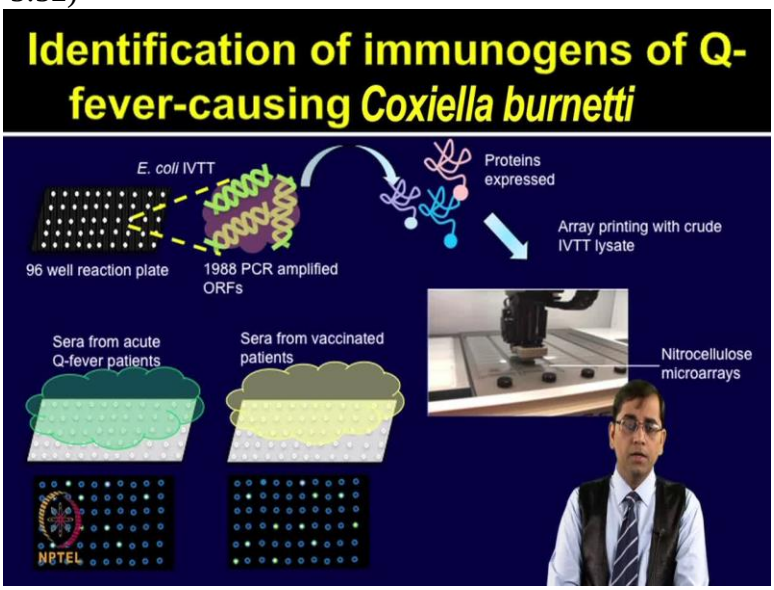
Let us now discuss the immunological studies in this animation. The use of cell free expression based protein microarrays for detection of potential immunogenic proteins of plasmodium falciparum was studied by Dooley Nichol 2008. In this study authors carried out cell free expression of PCR amplified vectors using (())(4:00) coli in vitro transcription and translation system. They expressed 250 putative proteins that were printed directly on to the microscopic array slides without any need for protein purification.

These arrays were probed with serum samples from patients who had been naturally exposed to plasmodium falciparum and who were experimentally exposed by means of radiation attenuated plasmodium falciparum. Authors successfully identified 72 highly immunoreactive protein antigens as well 56 previously uncharacterized antigens that were pseudo-dominant.

The study has shown some of the newly identify targets can serve as potential vaccine targets. Let us now talk about the next case study, identification of immunogenes of Q fever causing coxiella burnetii study performed by (())(5:26) 2008. Q fever is widely spread (())(5:31)disease caused by coxiella species. So, identification of immunogens of Q fever causing this disease were identified by using protein microarrays.

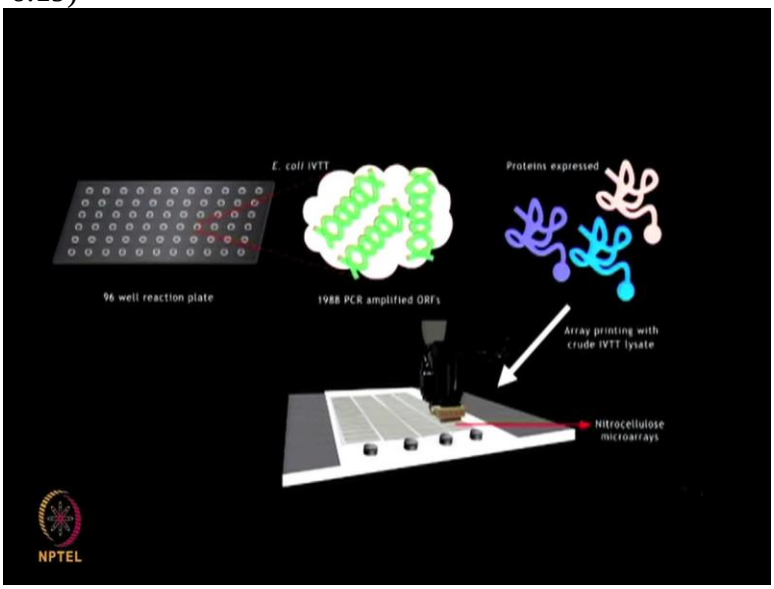
In this study authors used coxiella burnetii protein microarrays to identify immunodominant antigens.

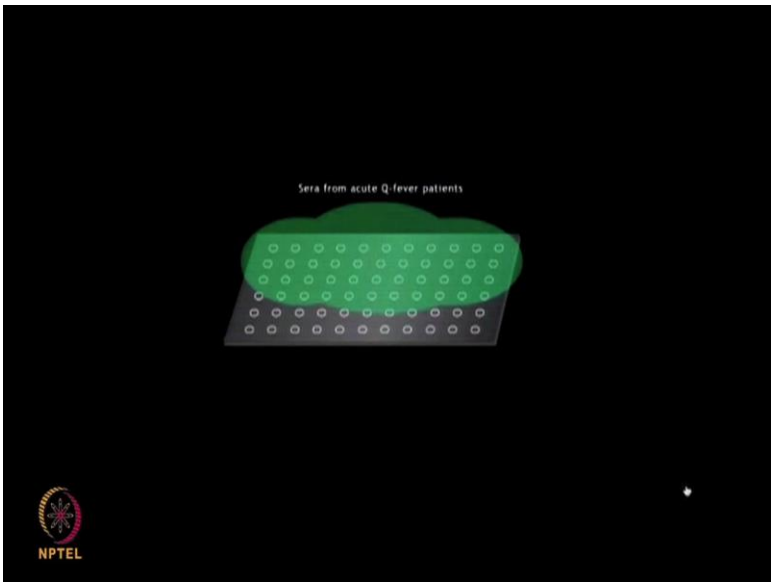
(Refer Slide Time: 5:52)

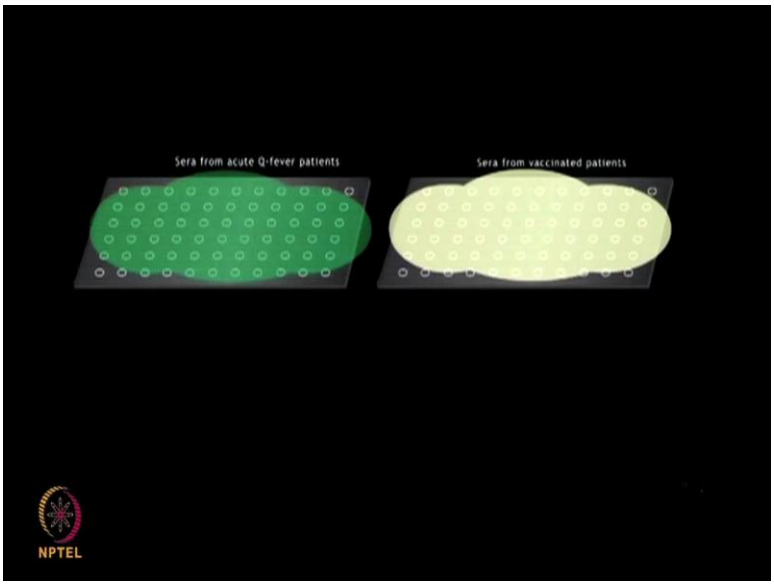


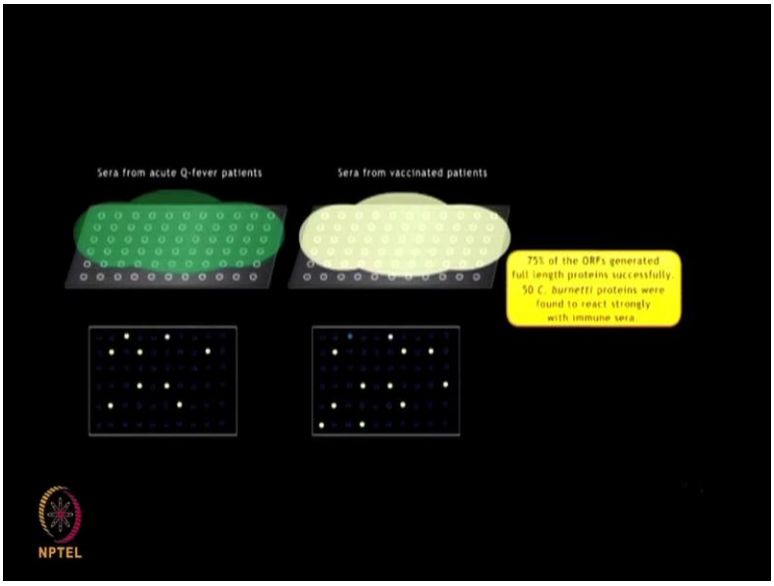
Almost 2000 open reading frame ORFs were generated by using the cell free expression based approach. E. coli in vitro transcription translation system and then employed this protein microarray platform for identifying the immunodominant antigens. Some of the steps involved in this experiment will be discussed in the following animation.

(Refer Slide Time: 6:19)







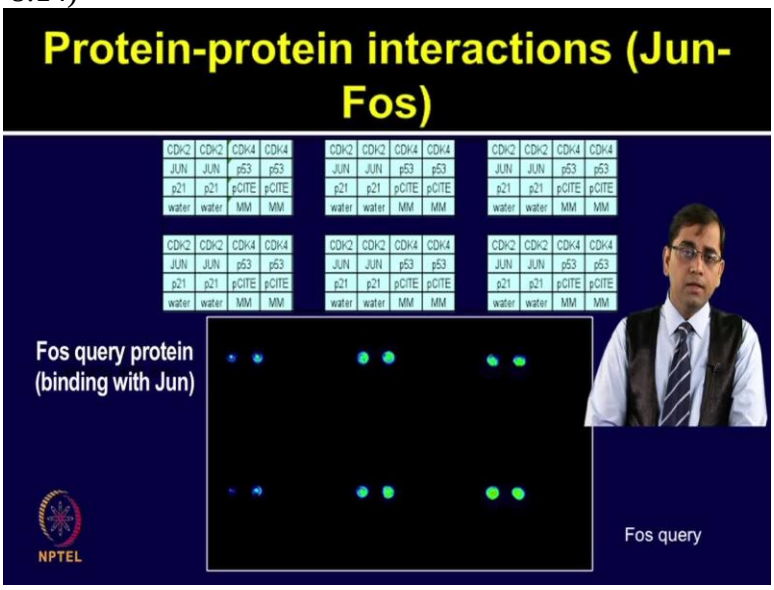


Case study 4-identification of immunogenes of Q fever caused by coxiella burnetii study by Berito 2008. (())(6:36)carried out in vitro transcription and translation of nineteen hundred eighty eight. Open reading frame of C burnetii by using E. coli based cell free systems. 75 percent of the open reading frames were successfully generated as full length proteins by using cell free expression system and then spotted on to the nitrocellulose arrays.

These cell free expression based microarrays were probed with sera from the patients who had been vaccinated as well as acute Q fever patients. 50 proteins were identified that were found to react strongly with the immune sera. From the previous lecture and this lecture you got a glimpse of application of protein microarrays for biomarker discovery and several immunological studies. It is time now to look another widely used application which is protein-protein interaction by using cell free expression based protein microarrays.

In this part of lecture I will mainly focus on nucleic acid programmable protein array or NAPPA and how they are been applied to study protein-protein interactions.

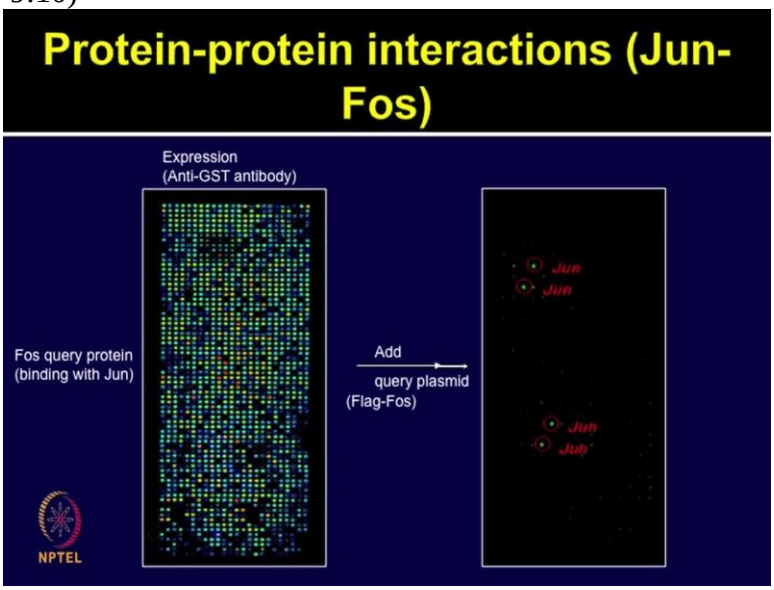
(Refer Slide Time: 8:14)



In this slide as you can see there is a small test array which we have used to teach proteomics course in coldstream Harvard laboratory in New York. The students made these array themselves and as you can see the array layout there only five handful genes were printed in duplicate on the chip along with vector control master mix and water.

Now if you want to study the Jun and Fos protein interaction and you can use Fos as a query protein, it will bind to Jun protein spot and therefore two spots of Jun will light up as you can see in the slide. So, all the six blocks they are duplicate of Jun proteins which are interacting with Fos protein and so Jun Fos protein interactions can be established by using the protein microarray system.

(Refer Slide Time: 9:10)

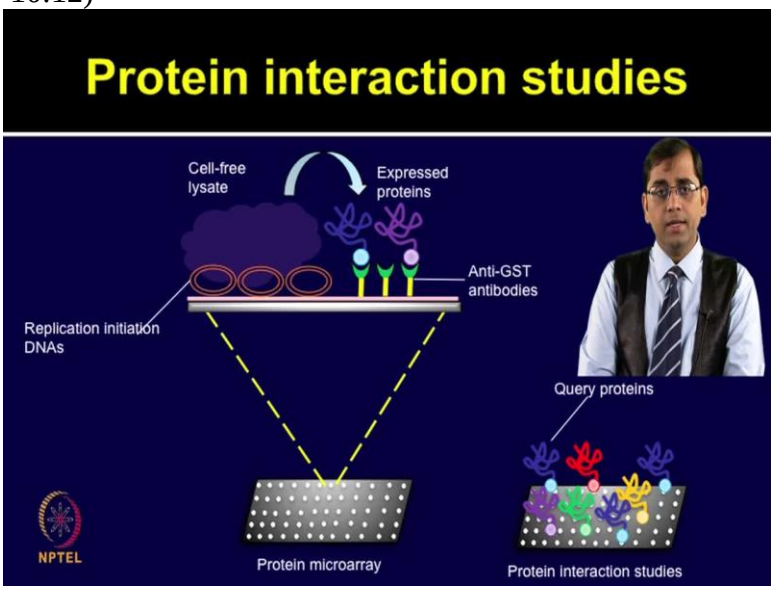


The previous array which we talked was spot testarray but if you really want to perform the protein wide screen to test protein-protein interaction so, you have to use high density protein arrays. Now in this slide I am showing you high density protein arrays to test the same protein-protein interaction of Jun and Fos protein pairs.

In this study students used Fos as query protein and then identified Jun printed four times on chip as a target. Jun Fos was used as model system to demonstrate how protein-protein interactions can be studied by using cell free expression based NAPPA microarray system. Let us look at some more case studies where protein microarrays have been used for study the protein interactions.

The next case study is performed by Ramachandra Nitol 2004 to identify novel protein-protein interactions using NAPPA microarrays.

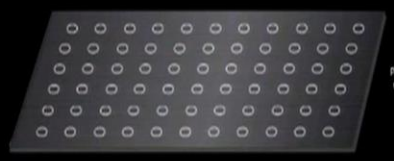
(Refer Slide Time: 10:12)



Authors reported generation of self-assembling microarrays which was one of the novel technology reported in science in 2004 in this study Ramachandran Nitol used a pair wise interaction among 29 human DNA replication initiating proteins which recapitulated the regulation of CDT1 binding to the selected replication proteins and map its geminin binding domain by using NAPPA approach. Let me describe some of the steps involved in this experiment by showing you this animation.

(Refer Slide Time: 10:54)

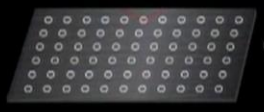
Protein expression using NAPPA



Protein microarray

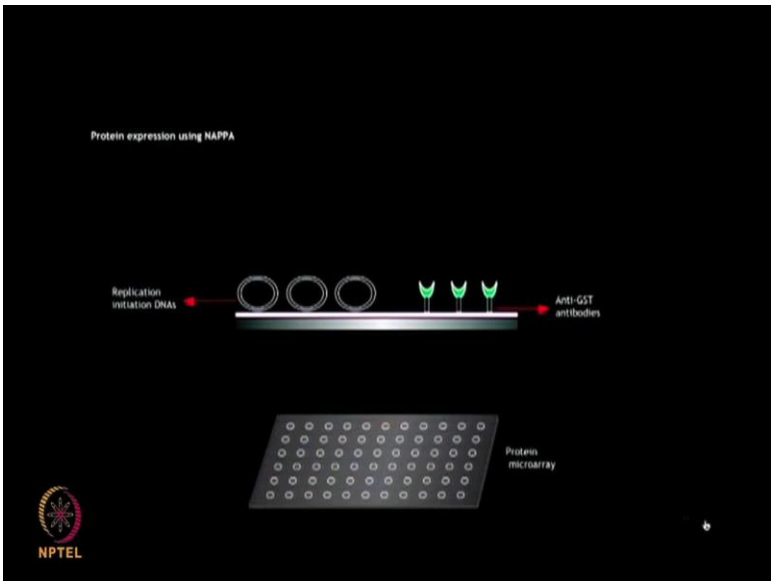
NPTEL

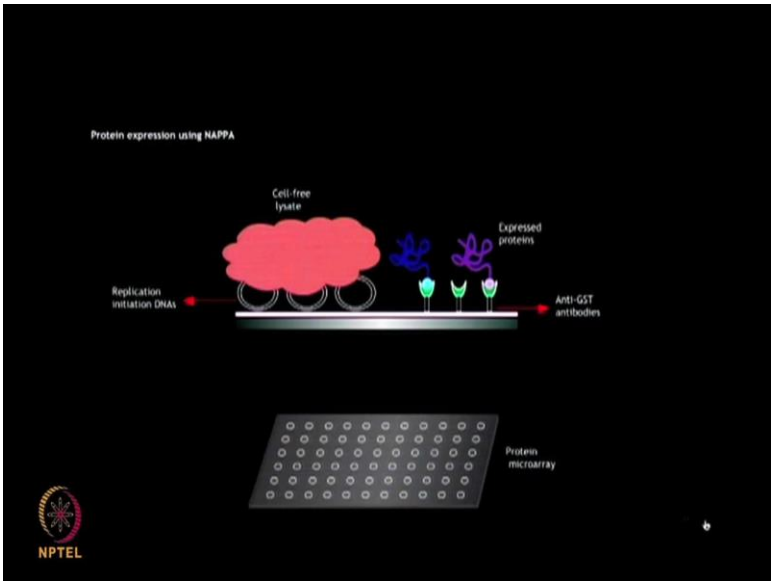
Protein expression using NAPPA



Protein microarray

NPTEL

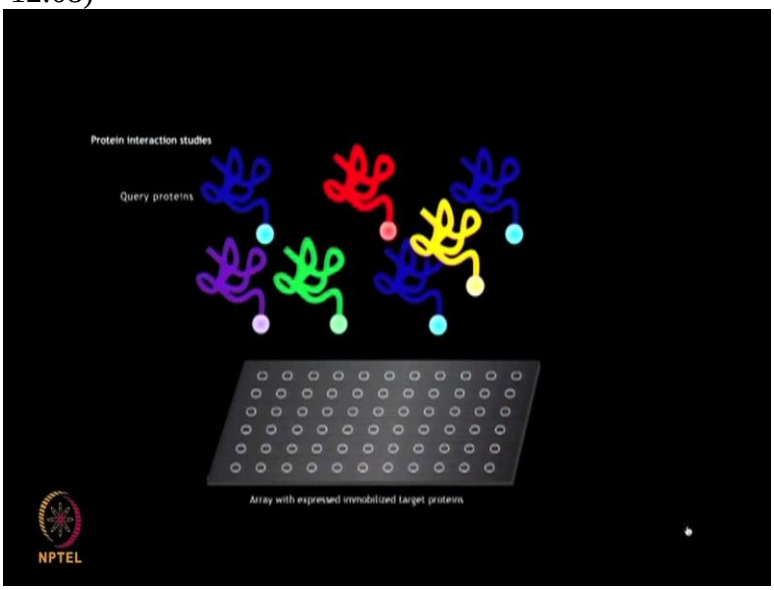

Protein expression using NAPPA
Replication initiation DNAs
Anti-GST antibodies
Protein microarray
NPTEL

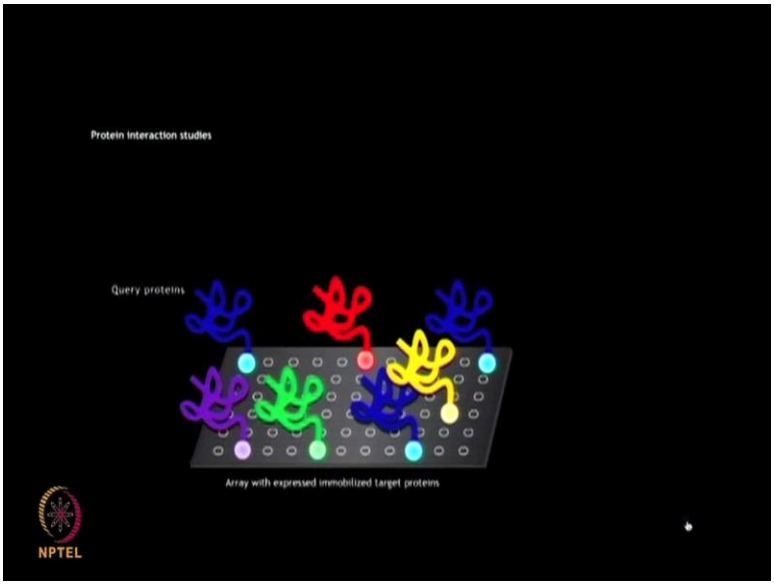

Protein expression using NAPPA
Cell-free lysate
Expressed proteins
Replication initiation DNAs
Anti-GST antibodies
Protein microarray
NPTEL

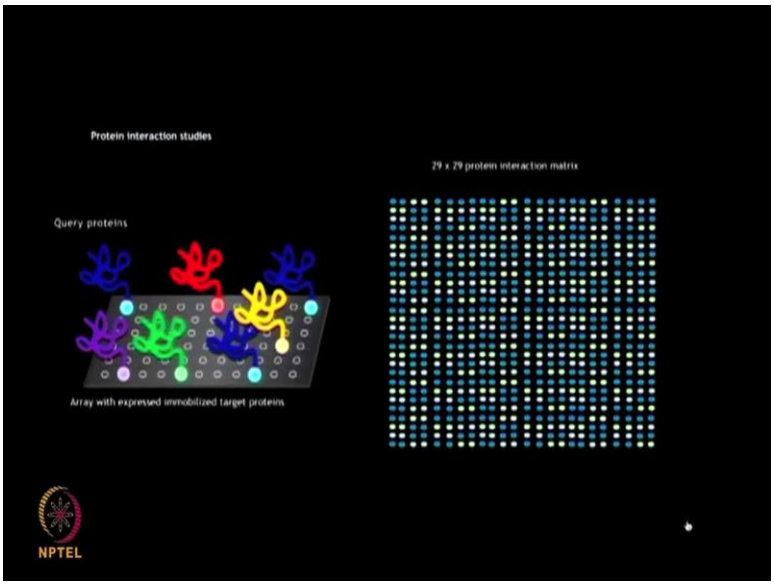
Protein interaction studies. Case study five-Identification of novel protein-protein interactions using nucleic acid programmable protein microarrays, study by Ramachandran Nitol 2004. Ramachandran Nitol tested the use of NAPPA microarrays by immobilizing twenty nine sequence verified human genes involved in the replication initiation on the array surface and then expressing them in duplicate with rapid ratico slide lysate.

The expressed proteins bound to the anti GST antibodies which are the captured antibodies present on the array surface.

(Refer Slide Time: 12:08)



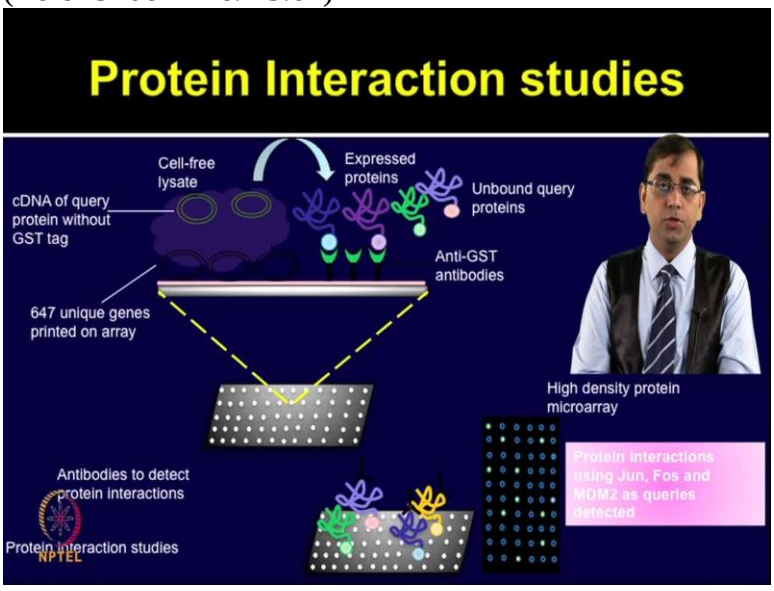




Authors made use of each of this expressed protein to probe another duplicate of array of the same twenty nine proteins thereby generating a 29 by 29 protein interaction matrix. 110 interactions were detected between proteins of the replication initiation complex of which 63 were previously undetected.

Let us now discuss the next case study - high density NAPPA array approach for studying well characterized gene pairs, a study by Ramachandran Nitol 2008.

(Refer Slide Time: 13:02)



The previous study which we discussed was more proof of concept where handful proteins were taken to study the protein -protein interactions whereas this time authors used high density arrays with thousands of protein features will printed.

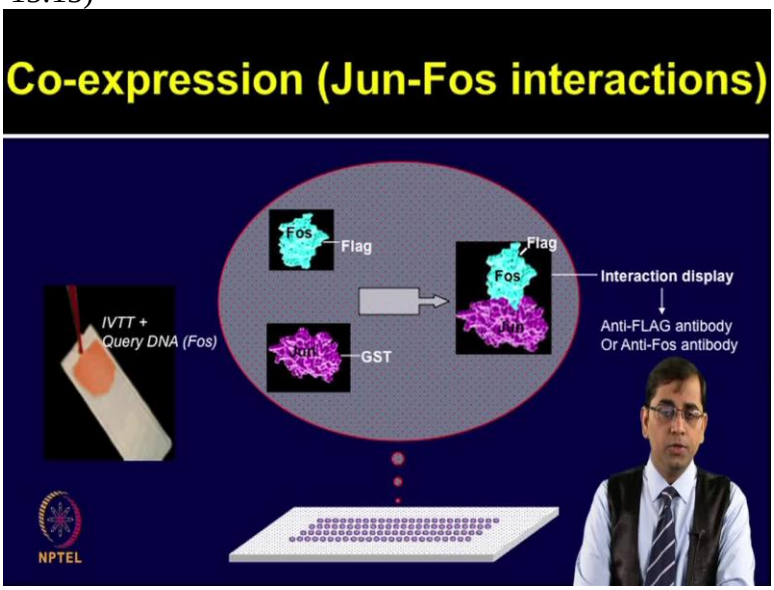
In this study authors used high density NAPPA approach to study the binary interactions between several well characterized interacting protein pairs such as Jun and Fos, p53 and MDM2. Now selective binding to these interactions were identify by using specific antibodies. In protein interaction studies it becomes very tedious to test our protein interactions in whole directions. For example if one testing the Jun and Fos interaction it should work in either way Jun as a query or Fos as query a protein.

If Fos is printed on the array Jun should be able to bind if used as query protein or similarly if is Jun printed on the array then Fos protein can be used as query molecule to test the protein interactions. Many times these interaction becomes unidirectional. It becomes very tedious to show that interaction is working in either of these directions but in this study author showed that protein interaction of Jun Fos protein pair can be shown in both the directions.

In addition to showing that protein expression and protein interaction works it is also interesting to perform the co expression where author showed that even the query protein need not to purify and one could expressed that DNA along with the in vitro transcription translation mix printed on the protein chip surface.

So if you have protein microarrays features are printed on the chip and then you have generated the contents by using cell free expression based system. Now you want to study the interaction and for that you have to purify the protein and used it as interactor.

(Refer Slide Time: 15:15)



By using co expression now you can use protein specific antibody to identify the interaction or you can use tag specific antibody to detect interactions. However in this study authors used co expression. It means the query protein along with the arrays protein were expressed by using cell free expression system so, that there was no need to purify the (crude) query protein as well.

Let us talk about co expression. So, if you have protein microarrays features are printed on the chip and then you have generated the contents by using cell free expression system and now you want to study the interaction and for that you need to purify a protein and use it as interactors. You can use protein specific antibody to identify the interaction or you can use tag specific anti body for detecting the interactions but in this study authors used co expression.

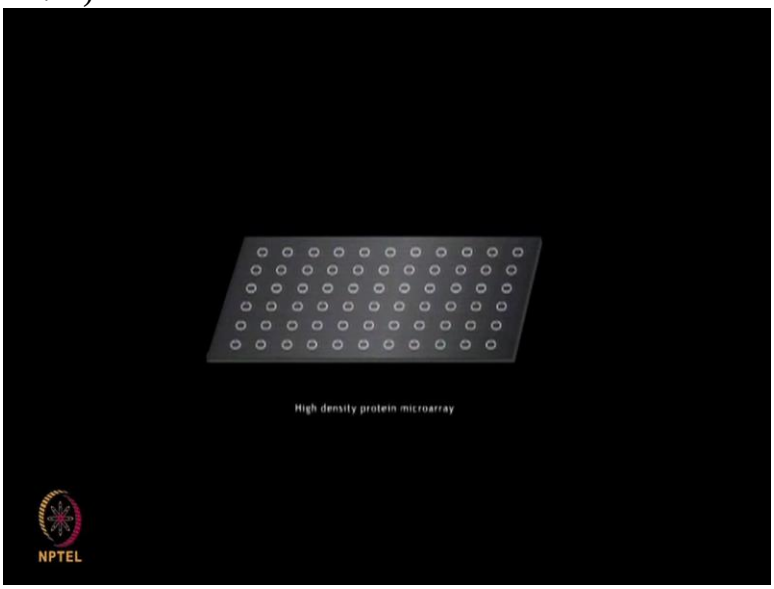
Let us talk about co expression experiment. It means the query protein along with the arrayed protein were expressed by using cell free expression system so that there was no need to purify the query protein also. When the protein interaction has to be performed you can take the cDNA of Fos protein, for example mix it in rapid reticosyte lysate along with other in vitro transcription translation machinery.

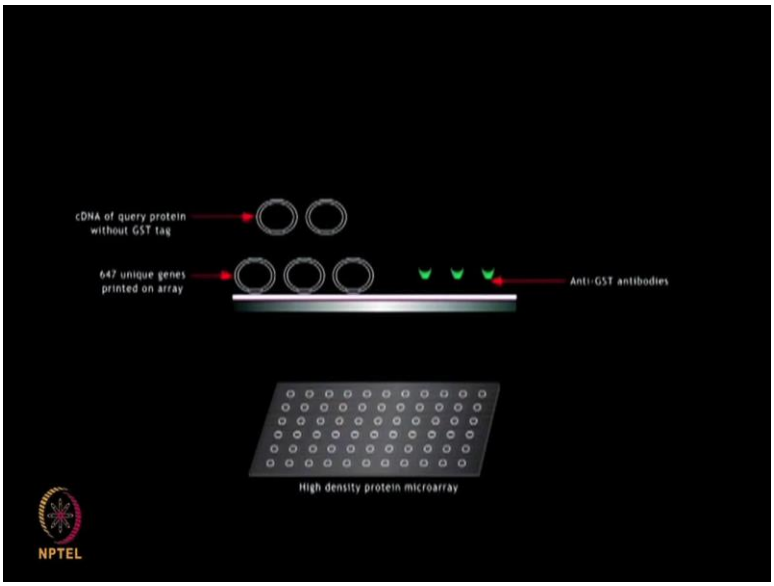
As you can see the slide the mix the whole cell lysate on chip surface and then after incubation when proteins are expressed at the same time query cDNA will also express the proteins and then

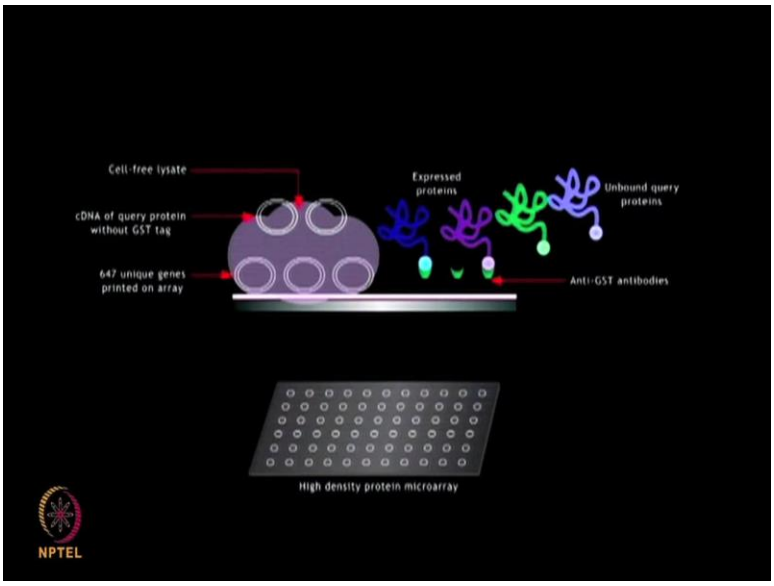
if it finds the binding partner it is going to bind to those features which can be detected by using protein specific or tag specific antibody.

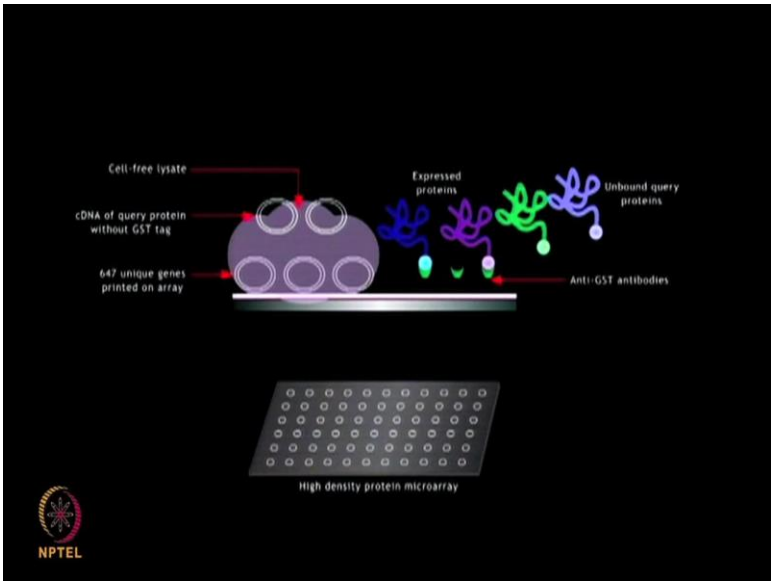
By performing this type of experiment authors allowed co expression, it means involvement from both query and target proteins expressed in the same environment and allowed very natural protein interactions to happen. There is good likelyhood that they are going to identify the right interactors. So, let me show you the steps involved in this study by showing you the following animation.

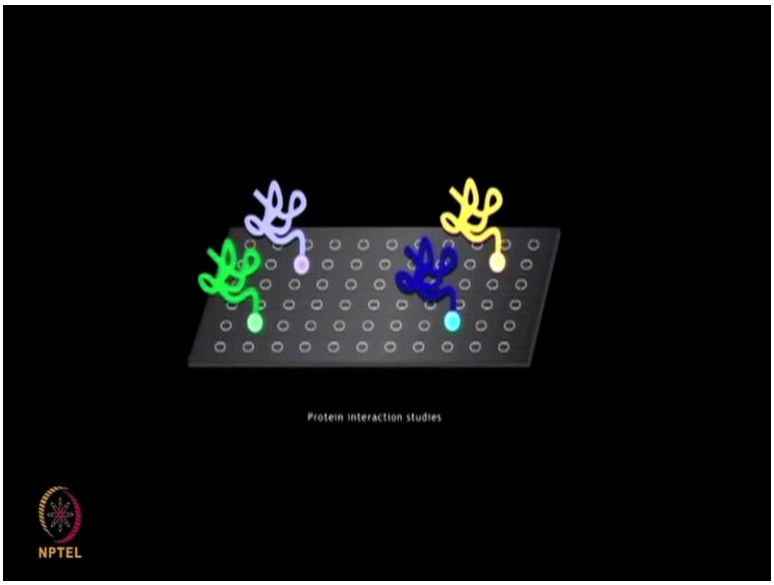
(Refer Slide Time: 17:47)

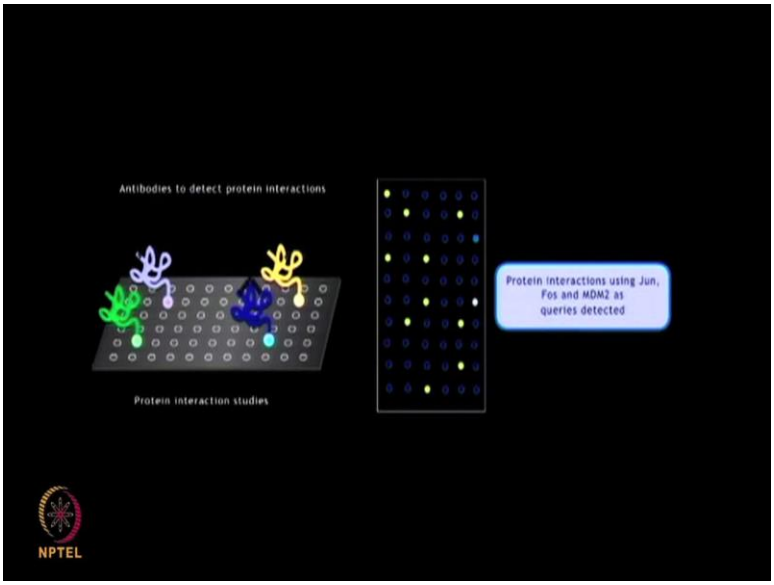


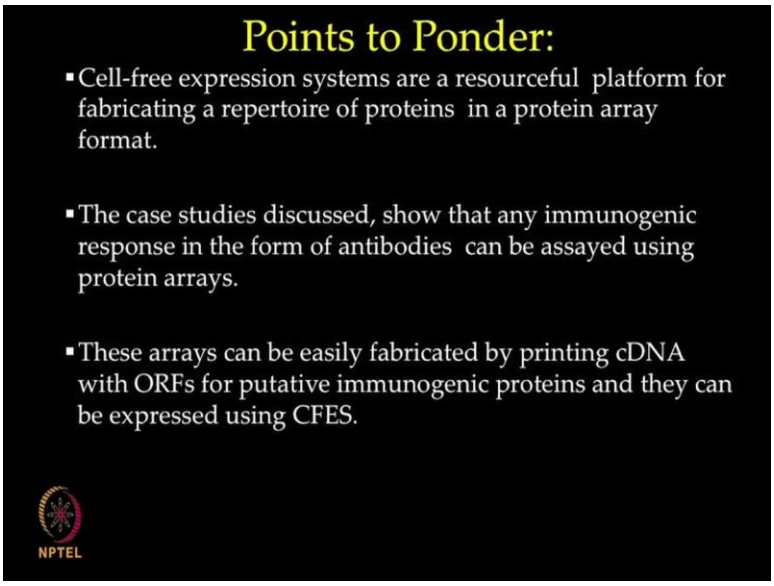



Cell-free lysate
cDNA of query protein without GST tag
647 unique genes printed on array
Expressed proteins
Unbound query proteins
Anti-GST antibodies
High density protein microarray
NPTEL


Cell-free lysate
cDNA of query protein without GST tag
647 unique genes printed on array
Expressed proteins
Unbound query proteins
Anti-GST antibodies
High density protein microarray
NPTEL


Protein interaction studies
NPTEL

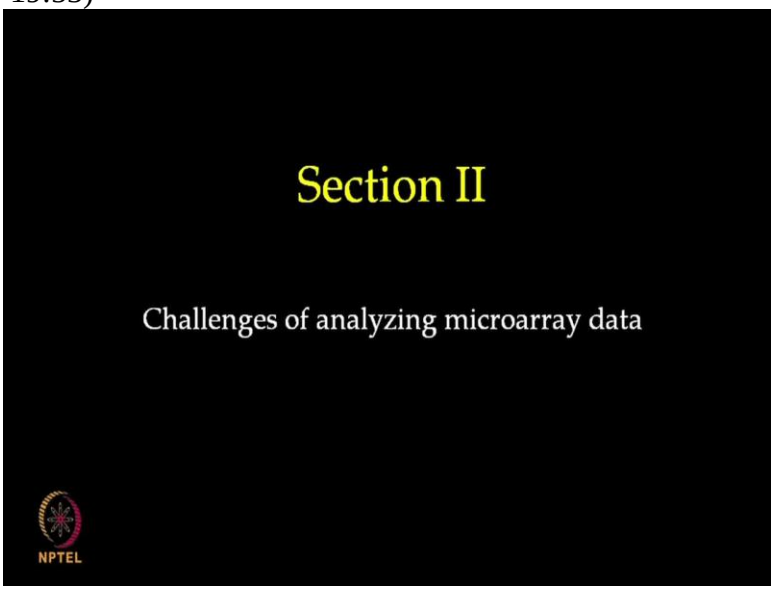

Antibodies to detect protein interactions
Protein interaction studies
Protein interactions using Jun, Fos and MDM2 as queries detected
NPTEL



High density NAPPA approach for studying well characterized genes pits, study by Ramachandran Nitol 2008. In this study authors made use of high density nucleic acid programmable protein arrays to study protein-protein interactions. 647 unique genes were printed on to the array surface and expressed by adding the cell free expression based system. After addition of cell free expression based system proteins containing GST tag were synthesized and bound on to the captured antibody.

cDNA of query protein was also added to the same mixture such that the query was co expressed but remain unbound due to the lack of the tag capturing agent. These protein microarrays were then probed with anti-bodies specific to the query proteins. Authors detected various protein interactions using well known query proteins such as Jun, Fos and MDM2.

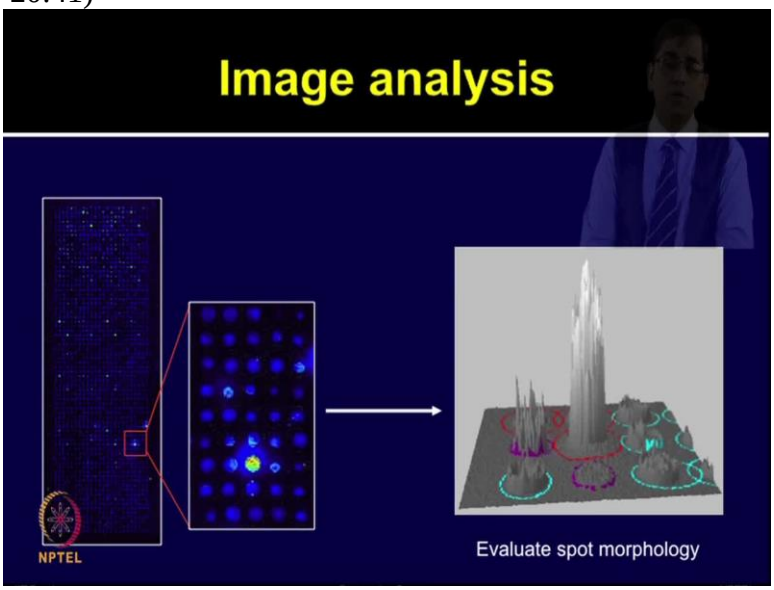
(Refer Slide Time: 19:53)

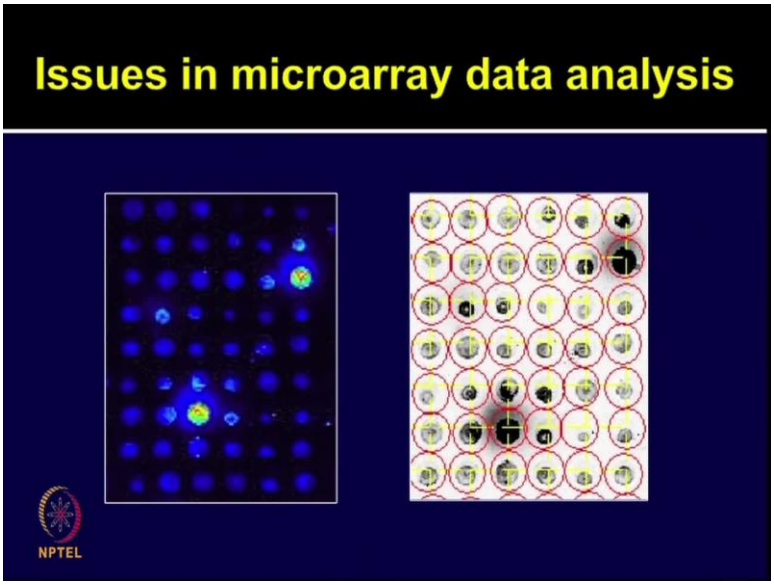


We have talked about various applications by employing cell free expression based protein microarrays and discussed biomarkers screening, immunological studies and protein-protein interactions over the last two lectures. Now regardless of what applications you want to perform on these arrays you are going to generate large amount of data. So, the volume of data generated from microarray experiments are prodigious.

It becomes important to develop the appropriate informatics system so that one can analyze the data uniformly and make some very good output from this data analysis.

(Refer Slide Time: 20:41)





So the image analysis when you are talking about high density approaches data analysis, the image analysis becomes very challenging. For example you can see an image here for the protein microarrays and I have shown a spot. The expression of this particular immunogenic protein is so high that it is spilled over to the neighbouring proteins.

Now one need to correct this type of error and remove the spots which are in the periphery of these proteins. Scaling up is good approach because one wants to perform high throughput experiment so that thousands of features can be studied simultaneously. However while scaling

up especially when you are using cell free expression based approaches one need to be cautious that what need to be the optimum intensity for the arrays because if there is of spillover of expressed protein on the neighbouring protein spot that is going to affect values of the neighbouring spots as well.

The slide shows how protein is diffused to the neighbouring spots and the neighbouring spot need to be corrected for data analysis. Similarly one need to perform the background correction, normalization and use various parameters to perform good microarray data analysis. Let us now discuss with the Dr Sudesh Srivastavaa the micro data analysis. What are the challenges involved and in the subsequent lecture we will talk about in much more detail the different steps and detail of data analysis for microarray based system.

(Refer Slide Time: 22:25)

Sanjeeva Srivastava: What are the major challenges of the microarray data analysis? Also I would like to get your comments that what should be the good statical design when biologists are starting some experiment for the microarray because most of the time there are chemical samples they would like to get some very useful biological information from this.

Dr Sudesh Srivastava: Statistical is very important aspects of this biological experiments and I think the statistician get involved from the beginning of the experiments and the the important of statistician as you mentioned about the sample not only the sample size but also to understand

the experiment and control the variation in any biological experiment and so it is a very good idea to have a statistician on the beginning of the experiment and where they can contribute not only on the data analysis point but also to conducting and performing a optimal design experiment way so you can have as precise and as useful information out of the data of the experiment you are trying to achieve.

Sanjeeva Srivastava: practice very important but there are many ways of analyzing microarray data and what are the different approaches which are available and which one would you consider as good approach?

Dr Sudesh Srivastava: Yes. There are I see a statistical is the methods are a tools and its depends on your objectives. So, when whenever you are trying to perform a experiment I think one has to be very clear with the objectives and which when you talk to the statistician they will let you know what methods are appropriate to your experiment to your hypothesis and in that case statistician will let you know how you should go and perform the experiment and not only experiment but also to control the biological or technical error which are involved in when you are performing experiment.

So, but the methods as I said there are these new methods are involving each and every day in this field and is because the biological problems are very complex. It is not like you can answer by telling one method. So but of course there are some standard methods been used in other fields and people are trying to use the soon like same methods in biological point of view and I would like to say the all the methodology developed in the statistics been based on small sample size and now in the recent years.

So, I would say and nowaadays we are dealing with huge data size particularly in biological field. So, in that case we have to come up with new methods or new methodology and statistic which can handle the more appropriately and more objective oriented statistic statistical method.

Sanjeeva Srivastava: So, I am coming that its not possible to really list out one best method.

Dr Sudesh Srivastava: Yes

Sanjeeva Srivastava: But at least can you provide few possible solutions.

Dr Sudesh Srivastava: Absolutely I will truly agree with you on that and there is no unique method and is not only in biological point of but is in general as well because statistics should be statistical is always dependent more dependent all the experiment dependent techniques so, it only develops type of data or the experiment objectives you are trying to achieve.

Sanjeeva Srivastava: Sudesh in the microarray field biologist apply that to identify differential expression of genes and what type of issues do you see like in terms of analyzing this data sets from the microarrays and what are different ways of analyzing the data and any comments on that?

Dr Sudesh Srivastava: I would like to say the statistical method used in identifying differential expression gene analysis starts from experimented design till the end of the analysis for the till the final conclusion of the analysis.

So, there are a number of methods available and particularly at design stage because design stage there is different kind of a design like you must have heard about the loop design, reference designs and also the factorial design it is depending on your situation. You try to perform your experiment and all this experimental design are basically depend on the statistical tool you wanted use for your data analysis and like when you are using reference design or you using loop design. Sso basically you trying to compare the two treatment of the by the control or sorry yes or the tumor.

So, in that case you used (())(27:19) like when you are comparing just two conditions and in case if you dealing more more than two condition then you go little bit analysis of various kinds of analysis and then you use booststrap method and you use sams method and also there is another method called wrest method. So, these are the all methods appropriate in your your situation when you are dealing with identifying differential expressions.

Sanjeeva Srivastava: So, there are a lot of options available but it always becomes challenging to apply which one is a real good method.and that is why we need

Dr Sudesh Srivastava : Ohh ya, absolutely absolutely

Sanjeeva Srivastava: and that is why we need some good statistical (())(27:55) gene for.

Dr Sudesh Srivastava : I am I am glad to hear that word because ahh most of the times what happen they scientist or basic scientist research they go to the statistician when they are done with their experiment and they go to detect the data to statistician and they ask to you could you please analyze the data and so, in that case I would like to say there is a very famous saying from(())(28:19) ahh when you to the statistician with your data so, on the statistician can do postpartum of your data and can tell you how the data dies.

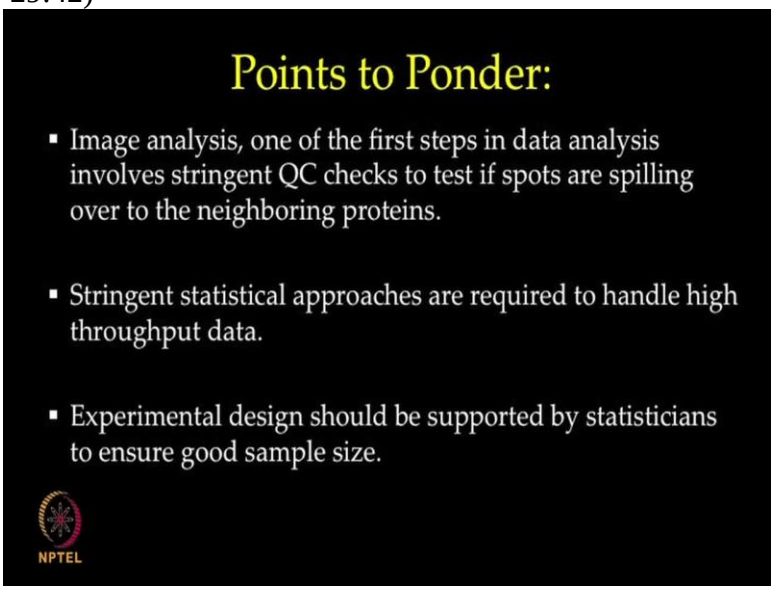
And so with that remark I would like to say if you are trying to conduct any research process or trying to perform any research oriented biological question so, I would say go to the statistician from the beginning and they can help you to at least get the optimal way of experiment itself so that that you can do the best analysis or best mathematics tool to appropriate your situation. So, I would say yes statistician should be or must be involve from the beginning of the experiment.

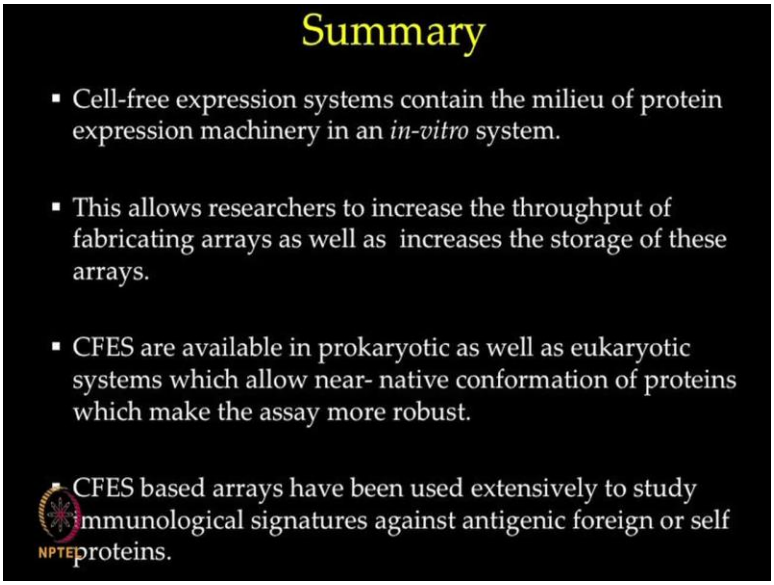
Sanjeeva Srivastava: So, I must say that this is the take home from this interview that its not at the end but actually its on the beginning when we need to involve a statistician for the large set of analysis if we want to perform some different high throughput experiments and especially genomics and proteomics mix is very important because we invest lot of technology, lot of samples and if our central design is not pretty well then later on the thing will fail.

So, with that thought I will conclude this interview and I would like to thank Dr Sudesh for being with us and sharing some experience on microarrays data analysis and challenges. Thank you very much.

Dr Sudesh Srivastava: Thank you.

(Refer Slide Time: 29:42)





Summary

- The main challenge in data analysis is to ensure that the image quality is optimum and "bad" spots are flagged and not considered for data analysis.
- Further the study must be designed always be in consult with the advice from a statistician prior to performing the experiments.
- Stringent data analysis with normalization would ensure robust analysis of the high-throughput data.

References

- Chandra, H., and Srivastava, S. (2010). Cell-free synthesis-based protein microarrays and their applications. Proteomics 10, 717–730.
- Chen, C.-S., Korobkova, E., Chen, H., Zhu, J., Jian, X., Tao, S.-C., He, C., and Zhu, H. (2008). A proteome chip approach reveals new DNA damage recognition activities in Escherichia coli. Nat. Methods 5, 69–74.
- Chen, Y.-W., Teng, C.-H., Ho, Y.-H., Jessica Ho, T.Y., Huang, W.-C., Hashimoto, M., Chiang, I.-Y., and Chen, C.-S. (2014). Identification of bacterial factors involved in type 1 fimbria expression using an Escherichia coli K12 proteome chip. Mol. Cell. Proteomics MCP 13, 1485–1494.

References

- Gong, W., He, K., Covington, M., Dinesh-Kumar, S.P., Snyder, M., Harmer, S.L., Zhu, Y.-X., and Deng, X.W. (2008). The development of protein microarrays and their applications in DNA-protein and protein-protein interaction analyses of Arabidopsis transcription factors. Mol. Plant 1, 27–41.

- Gupta, S., Manubhai, K.P., Kulkarni, V., and Srivastava, S. (2016). An overview of innovations and industrial solutions in Protein Microarray Technology. Proteomics 16, 1297–1308.

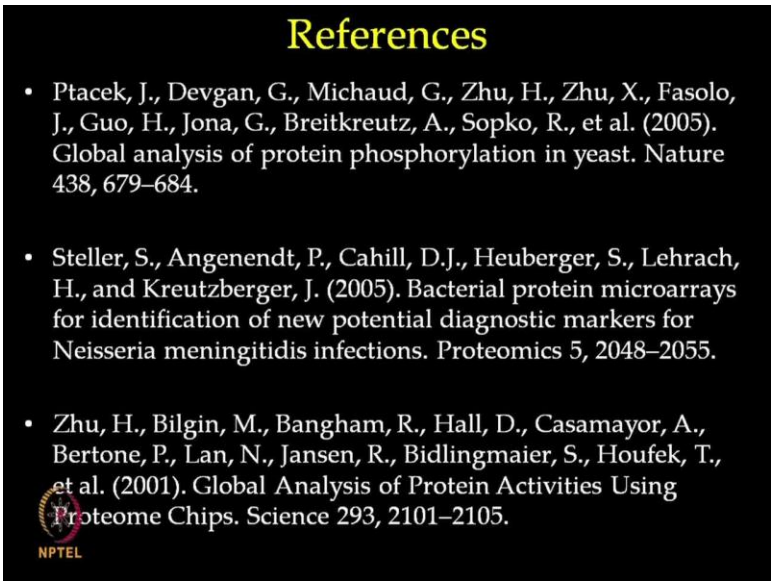
- Haab, B.B., Dunham, M.J., and Brown, P.O. (2001). Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol. 2, RESEARCH0004.

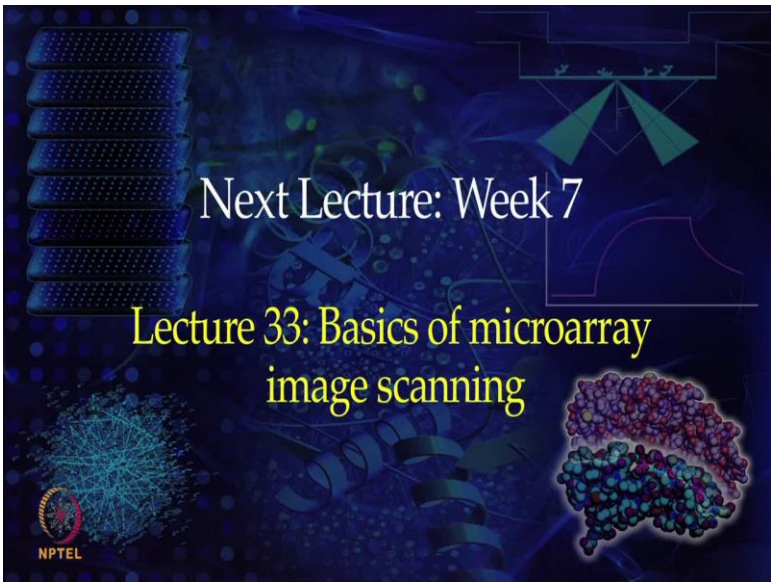
References

- Hudson, M.E., Pozdnyakova, I., Haines, K., Mor, G., and Snyder, M. (2007). Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays. Proc. Natl. Acad. Sci. U. S. A. 104, 17494–17499.

- Liotta, L.A., Espina, V., Mehta, A.I., Calvert, V., Rosenblatt, K., Geho, D., Munson, P.J., Young, L., Wulfkuhle, J., and Petricoin, E.F. (2003). Protein microarrays: meeting analytical challenges for clinical applications. Cancer Cell 3, 317–325.

- Miller, J.C., Zhou, H., Kwekel, J., Cavallo, R., Burke, J., Butler, E.B., Teh, B.S., and Haab, B.B. (2003). Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. Proteomics 3, 56–63.





Sanjeeva Srivastava: Thus in today's lecture we looked at different types of application of cell free expression system through various case studies. We also discussed the role of stringent statistics for robust data analysis with a leading expert, Dr Sudesh Shrivastav.

Once data is obtained from such high throughput experiments one gets a holistic view of protein regulation and cell functioning. That brings in (())(30:32) biology to test biological hypothesis developed from such studies and also understand functional biology at another level. We will look into these aspects in the forthcoming lecture. Thank you.