Interactomics: Basics & Applications

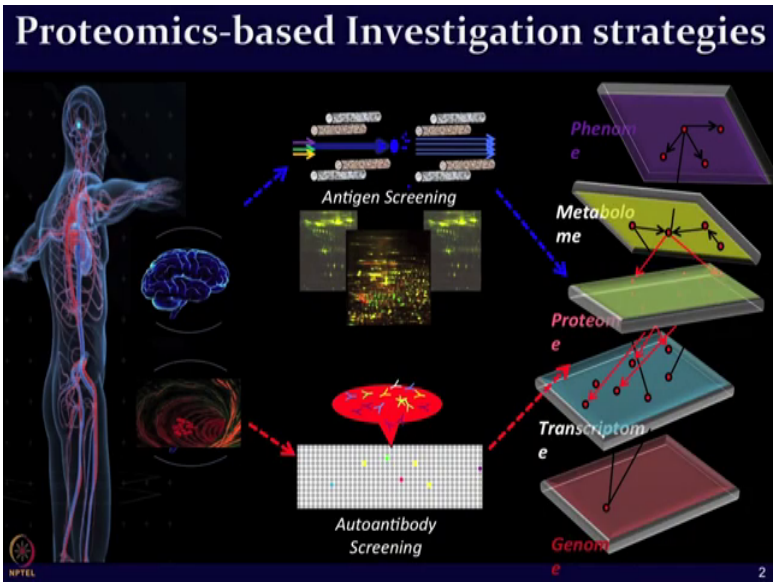
Prof. Sanjeeva Srivastava
Department of Biosciences and Bioengineering
Indian Institute of Technology, Bombay

Lecture - 22
Applications of Protein Arrays in identification of Autoantibody Signatures-I

Hello students. I hope you are enjoying various applications which can be then using different type of protein microarrays. You have got exposure of variety of platforms and different type of applications which are possible by using protein microarrays. I am going to give you a glimpse of another type of clinical application using human proteome arrays, I will focus mainly on the brain tumors.

So, let us talk today about autoantibody screening in brain tumors. But before I come to the workflows I would like to tell you that when you have variety of samples from which you can do proteomics investigation especially in the clinical context, then you can choose different path, different technologies to address different type of biological questions.

(Refer Slide Time: 01:11)

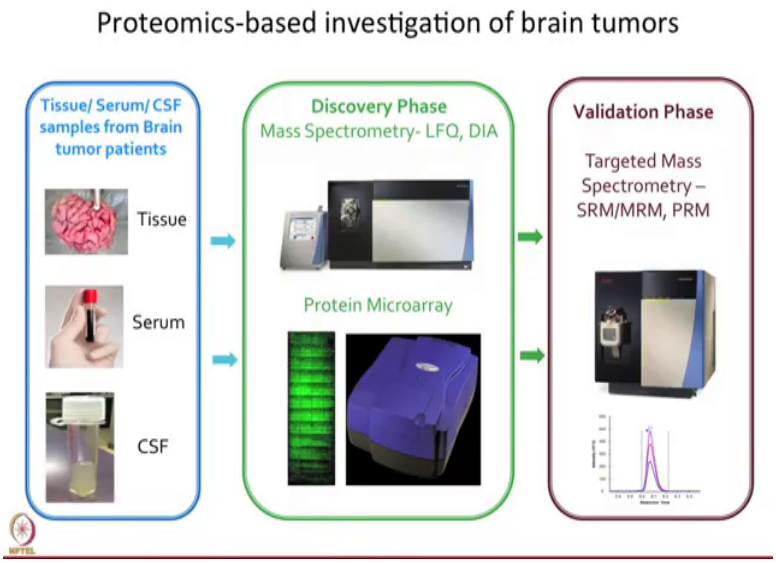


Here I have shown you two different paths – one essentially to look for the most abundant antigen screening which one could do either using the gel based platform or using mass spectrometry based proteomics. Alternatively, especially for various types of bio fluids one could use protein array platform and look for the autoantibody detection. Offers in the same samples one could obtain more complementary information and all of that could contribute towards, the systems level information.

If you recall from my very beginning lecture, when I talked to you about different technology platforms in the field of proteomics and how field is going to progress more towards system biology, I think we have to become very unbiased and appreciate variety of technologies and what kind of new insight they can offer from the same type of clinical samples. In this manner,

it is important that we should understand that which technology, which platform can provide what kind of unique insight.

(Refer Slide Time: 02:29)



So, this works what I am going to describe today essentially proteomics-based investigation of brain tumor focus on the autoantibody screening. This work was done in collaboration with Tata Memorial Hospital in Mumbai, where we obtain variety of clinical samples from the patient, a tissue, serum, plasma or cerebrospinal fluid; all of these kinds of samples provide some unique and interesting information.

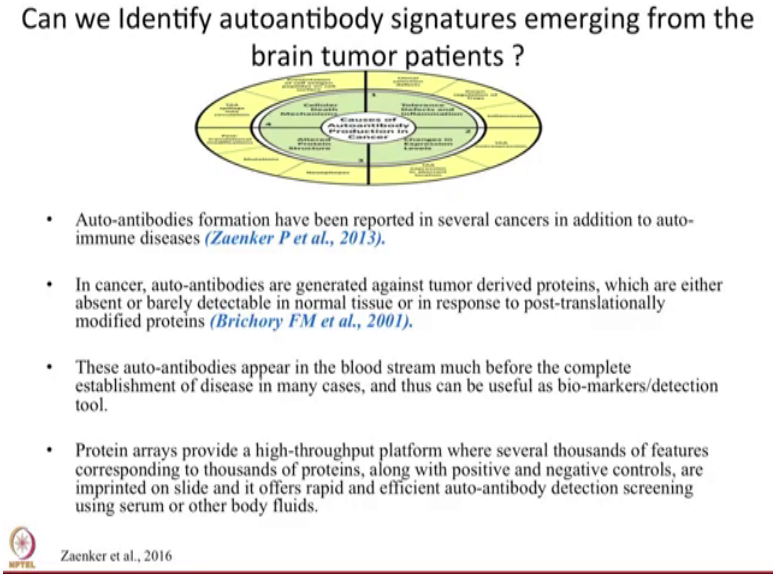
And as I mentioned you cannot decide what is the biological question you want to address, and accordingly you can choose different type of platform either mass spectrometry based or gel based proteomics investigations or using protein microarrays which I have shown in the

middle panel here, that the discovery phase could be done using variety of platform from the proteomic based technologies.

But eventually it comes to validation and validation one could do either using antibody based approaches or one could also use the mass spectrometry based new assays which are coming forward especially based on the selected reaction monitoring, where you are not involving antibodies and directly measuring the concentration of a given peptide or protein using mass spectrometers.

Again the entire field of mass spectrometry based proteomics and target proteomics it itself is a full content for the course, but here I am kind of trying to give you the balanced view that how you should be applying the same samples to understand different type of questions by applying on different type of technology platforms.

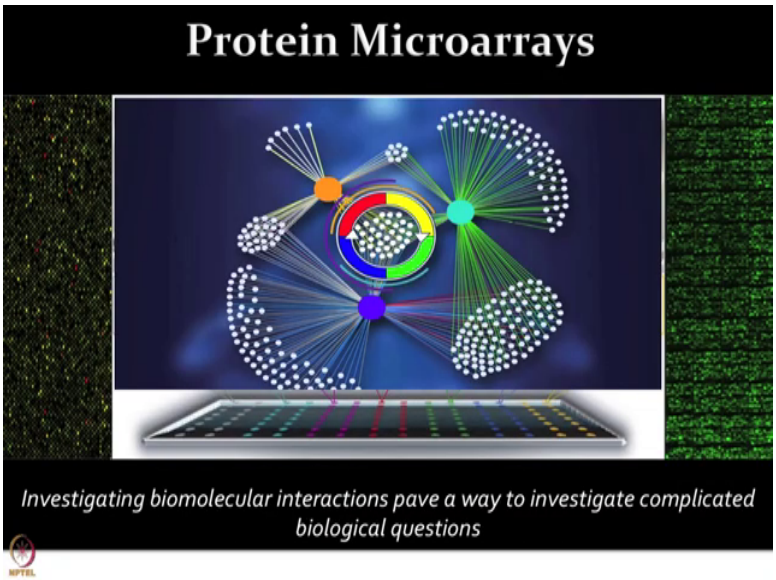
(Refer Slide Time: 04:03)



So, question is can we identify the autoantibody signatures emerging from the brain tumor patients. This is one of the very interesting hypothesis where many reports and many publications they have reported that in addition to the auto immune diseases even a different type of cancer, auto-antibodies are emerging from the patient samples, and especially in the cancer the auto antibodies are generated against the tumor derived proteins which are either absent or barely detectable in the normal tissue or in response to the post translationally modified proteins.

These auto-antibodies appear in the blood serum or plasma much before the complete establishment of disease in many cases. Someone could actually use them as early detection biomarkers, because one could do the testing from the blood with a routine kind of blood based tests which people usually do for the health check up. So, protein microarrays provide one of the strong platform to look into these autoantibody screening from the patients serum or plasma samples.

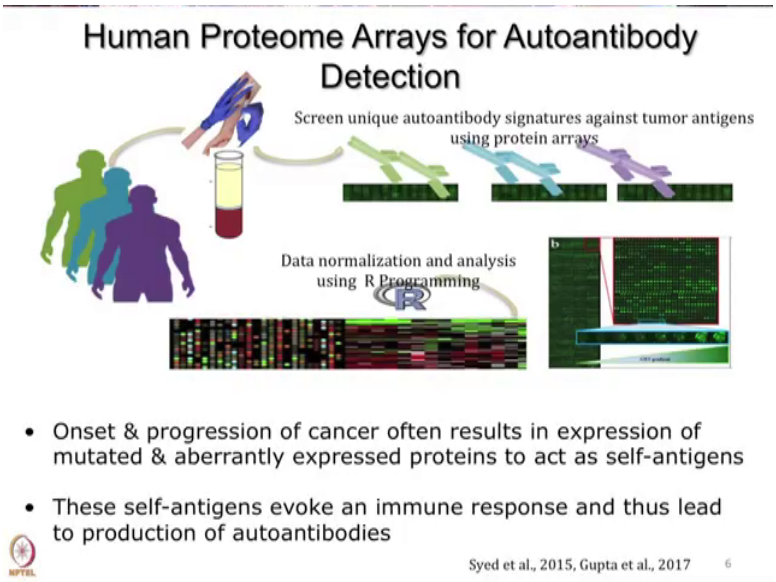
(Refer Slide Time: 05:18)



So, as we have seen that protein microarrays, they offer a variety of applications and different ways one could make the arrays, and finally, use those protein array platform which is essentially the miniaturized arrays when you can print your proteins or your CD array, and you want to make the proteins from that which could be utilize for many applications. But biomarker screening using auto-antibodies is one of the very important application which is we are going to focus today.

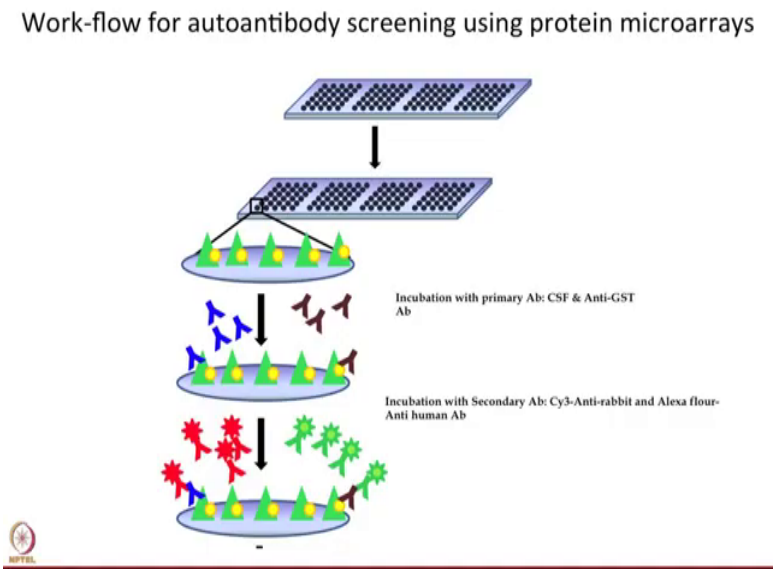
So, as I mentioned the onset and progression of cancer often results in expression of mutated, and apparently expressed proteins which could act as the self-antigens. Now, these self-antigens they evoke an immune response which leads to the production of auto-antibodies.

(Refer Slide Time: 06:06)



Here I have shown you a workflow how to use human proteome arrays for detection of auto-antibodies. Just keep in mind that when we are talking about human proteome arrays, one could either use the purified proteins or use a CDNA-NAPPA based approaches. In this case, we are using purified protein based arrays which consists of almost 18,000 human proteins printed in duplicate. Idea is if we add the patient's serum sample, we can detect unique autoantibody signatures again the tumor antigens using protein array platform. Of course, while assay looks very straightforward and simple, but data analysis and further data interpretation becomes very challenging.

(Refer Slide Time: 06:55)



Here I have shown you the work-flow for autoantibody screening using protein microarrays. You can think about a similar experiment which you have done using western blots, exactly the same concept is here except that you do not have a membrane talking about the protein arrays on the different substrate especially on the glass or nitrocellulose membrane. All the protein separated on the chip, first you want to block your slides, so that all the other region, whereas proteins are not printed they should not give you the background, the blocking can be done either milk dissolved in the PBS or you can use the super block.

Once blocking this step has been done then now you can use the arrays where you can add the patient serum sample. And think about we are talking about very thin glass slides where you can add even simply 10 to 15 or 20 microliter of the patient serum sample which is diluted in

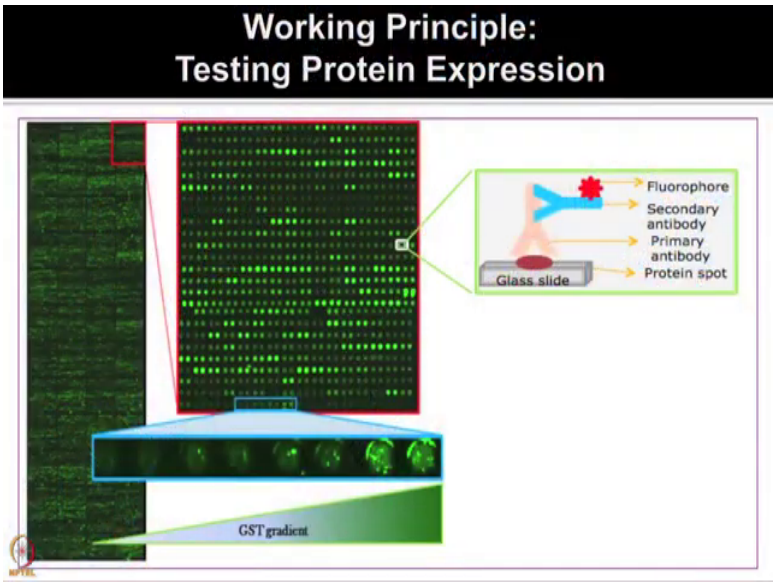
the buffer. And therefore, even hardly with 200 microliter, you can cover the entire slide. Idea is that these patients serum sample contain antibody against tumor.

And if these antibodies are produced, then can they bind to the antigens which are printed on the chip. So, we have 18,000 human protein antigens printed on the chip. And these antibodies which are part of the serum sample, once they come on the chip, can they come and bind. And if they are binding, then one could try to retake those antibodies using secondary antibodies which is antihuman iggs linked with the psi 3, or psi 5, or various type of (Refer Time: 08:34) fluorophores.

So, the idea is a primary antibody is coming directly from the serum sample, and secondary antibody you are adding either antihuman iggs or igms, and then further you can detect the signal with any different type of detection strategies. So, so far I think you got a glimpse at doing the protein array based experiment is not very difficult, and you do not need a lot of you know big setup and big instrumentation in any standard molecular biology lab, these experiments can be done.

But when we talk about high support experiments, you have to keep in mind that these experiments has to be done very meticulously, and lot of attention has to be paid that what kind of data you going to get. High throughput does not mean that you know you can get data in a very high quality in a very very short time. High throughput requires a lot of attention lot of you know careful thoughts that how well your experiments are proceeding is the data what you are going to generate is of high quality and can rely on that data.

(Refer Slide Time: 09:39)



So, in this slide, first of all if you have the printed slide, you would like to make sure that protein is actually printed properly or if you are using NAPPA array, so proteins are expressing properly. So, if all the clones which you have printed eventually which from which you made the protein if they contain some tag for instance the GST tag or make tag any other tag. Now, if you can use the antibody against that, so all the features where the protein has been printed or protein is expressed, then one could test out from the antibody.

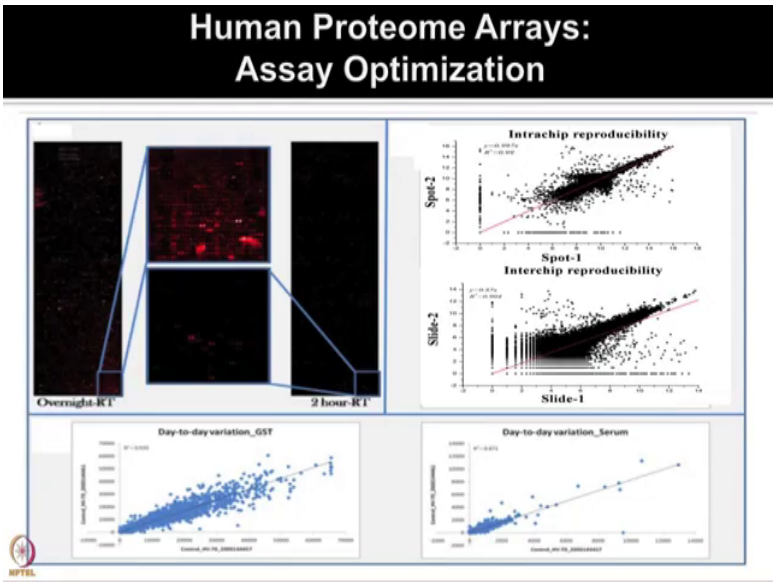
In this case, first we want to test on the human proteome arrays which we collaborated from the Johns Hopkins Laboratory in US. So, these arrays contain the purified protein, and each one of these clones contains the GST tag. So, as you can see on the slide that first you want to make sure that wherever proteins are printed, so they are all showing is the signal which is coming from anti GST antibody detection.

Also we have the you know series of purified proteins printed for the GST protein which shows the gradient and different concentration of the GST spots from the lower to higher concentration which also helps us to decide that you know how much protein is expressing on any given feature.

The next thing comes that you know you have now the patient serum sample, you want to apply that on the human proteome arrays. Idea is what should be my best experiment which I should try to apply on the protein arrays. If we are let us say having the serum with very high concentration that might get very sticky and you see lot of background a nonspecific signal.

So, you have to make the first appropriate dilution of the serum sample, then further how long I should keep the serum on my you know the slide, should I keep it just for an hour, 2 hours, overnight, very long incubation, just on the room temperature or doing the 4 degree or 37 degrees variety of you know optimizations are required right. So, before you actually start applying your hundred patient sample you know it is good idea for you to optimize the signal first and that is what we try to optimize the assays.

(Refer Slide Time: 11:50)



If you look at the left side panel on the slide, the top, the center image show that you know when we are adding for the overnight incubation of the serum, we see lot of non specific signals coming. So, now the first image I showed you earlier was the GST with the you know green signal, and now we are showing you the entire human igg signal coming as a red signal here.

So, now, we are looking at autoantibody responses, and they should be response which we want to detect, but it should not be everywhere which is nonspecific. So, the middle top panel shows that there is lot of red signals which is now coming on a specific throughout the slide.

The bottom panel what you see like that looks pretty much clear that denotes everything is not lighting up, only few spots show red signal, and that is with the less incubation when we have at the room temperature. So, these kinds of optimizations are very much required. Also what

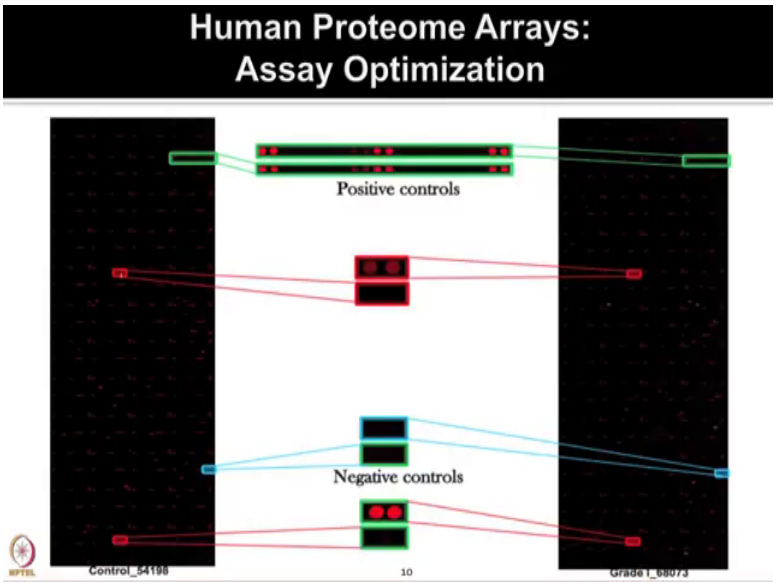
you want to test out when you are using the same patient sample, apply on two different chip how much reproducibility you can see slide one to slide two, also the spot to spot reproducibility across different chips. So, I hope you recall from my previous lecture that how much you know the QC checks are important right.

So, now we are doing the actual project, actual experiment here, aim is to look at the brain tumor patients whether the autoantibody response and can be identify some biomarkers. But before we start actual patient serum sample screening, we would like to make sure our assays are working fine, and the reproducibility of assay has been determined.

So, on the right side, you can see we have measured intrachip reproducibility, interchip reproducibility. And on the bottom panel, you can see we have looked at day-to-day variations across different types of spots and different type of chips. Once we have seen that our CV that tight assays looks fine then only we are ready to proceed for the actual experiment.

Additionally we have to also keep an eye that when we are doing these assays, our positive controls and negative controls they are behaving the way they showed. So, for instance, in this case we have if you are applying the antihuman iggs, so iggs spots which are purified immunoglobin proteins printed on the chip, they should definitely show as the positive controls.

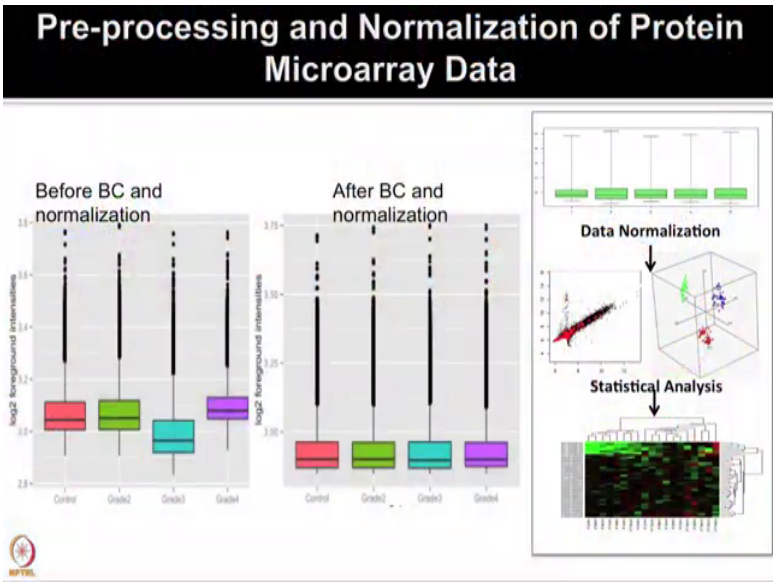
(Refer Slide Time: 14:06)



And that is what you see in the in the middle here. The positive control with the red spots are seen, whereas when we have empty spots or the spots you know when we have only vector we do not have the genes printed on that. So, then those should be negative controls when we are not expect to see protein expression happening. And those are the ones which are shown as the totally black spots which are the negative controls.

So, again now you see that how much we have already looked carefully variety of features before we actually start applying our patient samples and start investigating questions of biological interest. And further once we have done all the screening, then one need to look at different ways of statistical analysis to look for that how my signals were appearing before normalization.

(Refer Slide Time: 14:55)



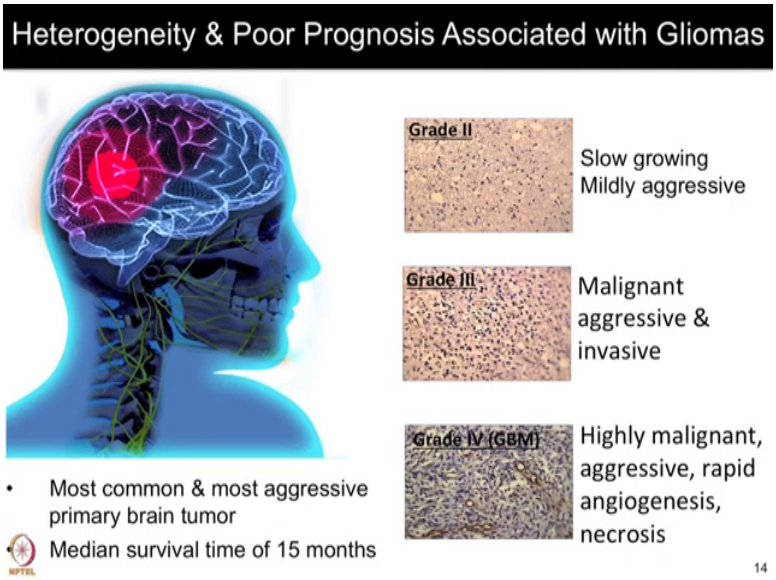
So, if you look at the left hand side, variety of you know different grade of you know patient samples, they are showing different level of signals. But we need to normalize all of them, so that every signal now start from the same baseline, and then only we are able to calculate what should be the protein expression change for the up regulation or down regulation, and then right side shows you the schematics. That after doing those kind of normalization, then one could proceed for looking at the significant proteins the best classifier using multidimensional scaling analysis, and also look at various type of heat maps and find out the best classifiers.

I hope now you are clear that you know what one should do before starting a project of you know the clinical interest or biological interest. So, now, let me kind of walk you through little fast, but the idea is to give you the glimpse how one could now apply this information this knowledge of a doing an essay on the protein arrays for the actual patient related problem. So, I am going to talk to you about some deadly brain tumors, the gliomas, meningiomas, and how

from the serum and CSF cerebrospinal fluid, one could use the same protein array platform and look for the auto antibody biomarkers.

So, I do not want to talk too much in depth about biology of these diseases, but just to reiterate that these are one of the most challenging tumors which you know are very aggressive in nature especially the gliomas tumors, and these are the most commonly found brain tumors especially meningioma shares the largest fraction of all the brain tumors.

(Refer Slide Time: 16:28)

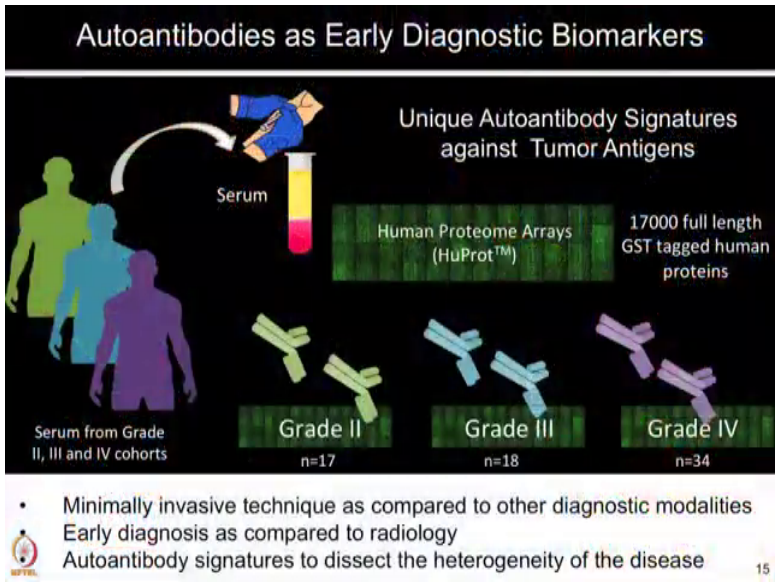


So, just to start we I want to talk to you about the gliomas first, which I have the origin from the glial cells. And ideally you know we have the patient which are having the low grade tumors or the high grade tumors. The low grade gliomas grade I and 2, they are slow growing, and you know they are definitely mild in nature not as aggressive as you know you will see in the higher grades. What are the disease progresses to the grade III and grade 4, we start

seeing that the they are very malignant aggressive and invasive nature. And then eventually the gliomas tumor multi forming or the grade 4 form which are highly malignant most aggressive and most deadly brain tumors which also shows the rapid angiogenesis, necrosis etcetera

The median survival of these patients are very less, sometimes a year or 15 months. Therefore, you know detecting these kinds of diseases at early stage probably can help better diagnosis and better therapeutic strategies for these patients. So, for doing this project, we took the patients from different grades of gliomas, who are suffering from different grades of the brain tumor the grade II, grade III and grade 4.

(Refer Slide Time: 17:36)

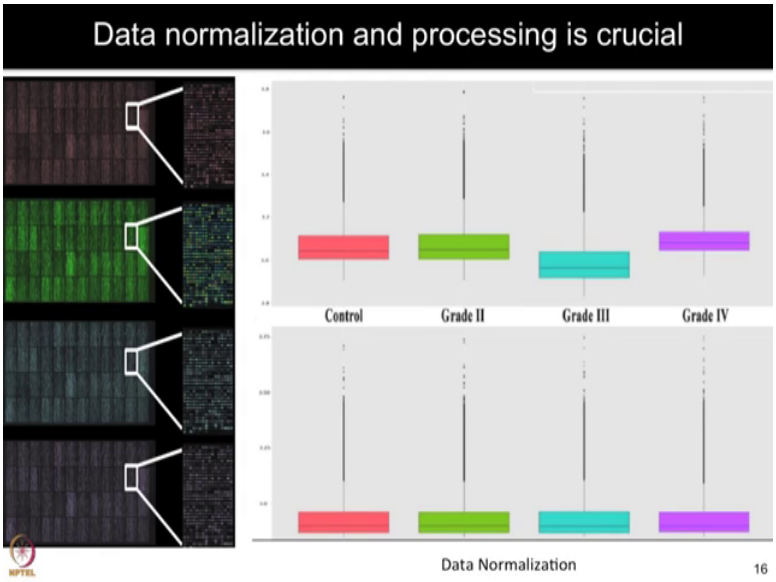


As I said grade 4 at the most deadly gliomas tumor multiforme patients. And we had this human proteome arrays, the first version which we use for this project was having 17000 full length protein all of having GST tag. And idea was to apply the patient's serum sample from

different grades of these patients one at a time on each slide. And then you scan the images and then eventually use the image analysis for looking at any pattern emerging from these images for the different grades of tumors.

So, as I mentioned that you know while a screening is not very difficult to experiment, one could do you know 10-20 per day; but eventually when you want to put together the hundreds of you know patient data images, then the data analysis data normalization becomes very challenging.

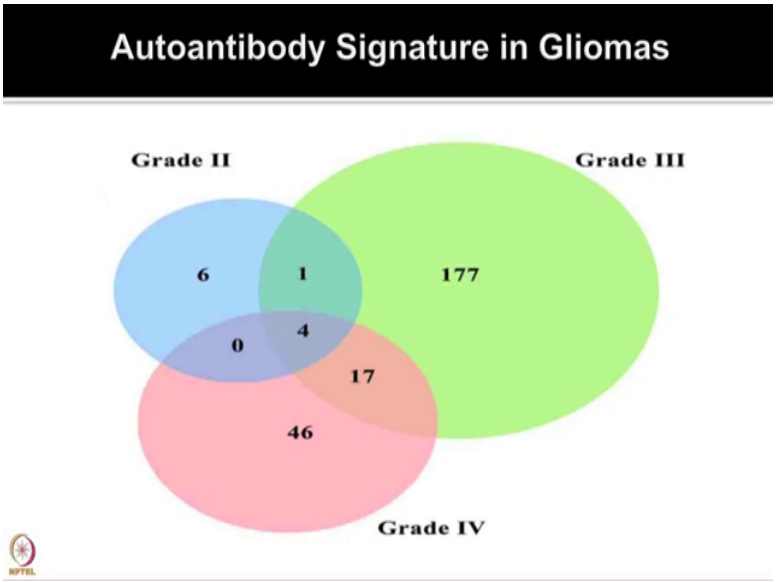
(Refer Slide Time: 18:26)



And this is what we encountered here. The different you know grades of patients and the control. We were doing in batches and then before doing the analysis, we could see that you know their signals are not same. So, we did the data normalization. And further after doing all

the statistical analysis, we found that the grade II patients having the less number of auto-antibodies.

(Refer Slide Time: 18:49)



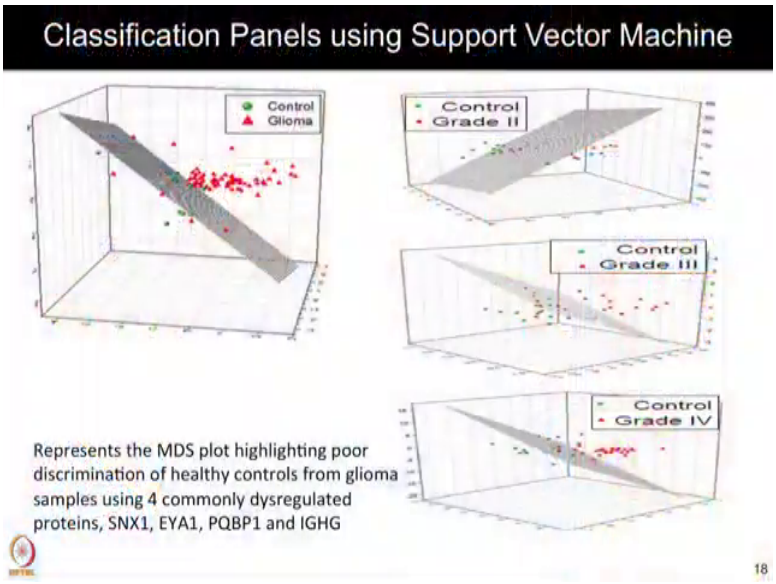
Only 6 we could detect, whereas grade-III showed the highest number of autoantibody response which is 177, and grade-IV showed the moderate numbers. So, while you know we are very surprised that why we could see very few number of autoantibody proteins in grade-II, but when we spoke to our clinicians and our collaborators they are not much surprised. Reason being that if you think about the brain tumors especially the gliomas, they are protected with the blood brain barrier. And once the tumors start going much further and starts going to the aggressive stage of grade-III and grade-IV, then only blood brain barrier is going to breach, and then many of these proteins are going to infiltrate and reach to the bloodstream.

So, therefore, very few proteins are ideally expected to see in the grade-IIs in the blood and grade-III should have large number of proteins. And because of the grade-IV everything you know body's immune system is shutting down everything is almost deregulated, so very few changes we are seeing at the grade-IV level. So, this was actually clinically quite interesting observation. And now the idea was which are these proteins which we can use for our further analysis.

Additionally if you see these Venn diagrams, there are four proteins which are showing they are common in grade-II, III and IV. So, can we find some protein which are showing common you know active these are commonly present in all the tumor, but if there is a trend can we detect them at the very early stage? Because one of the idea for this project was to look at early detection biomarkers.

So, after these analysis then one would like to find out can my proteins which we have identified segregate these you know the brain tumor patients with the healthy controls, so that was the idea. And we used you know the multi-dimensional scaling support vector machine based analysis, and try to segregate the control with the gliomas.

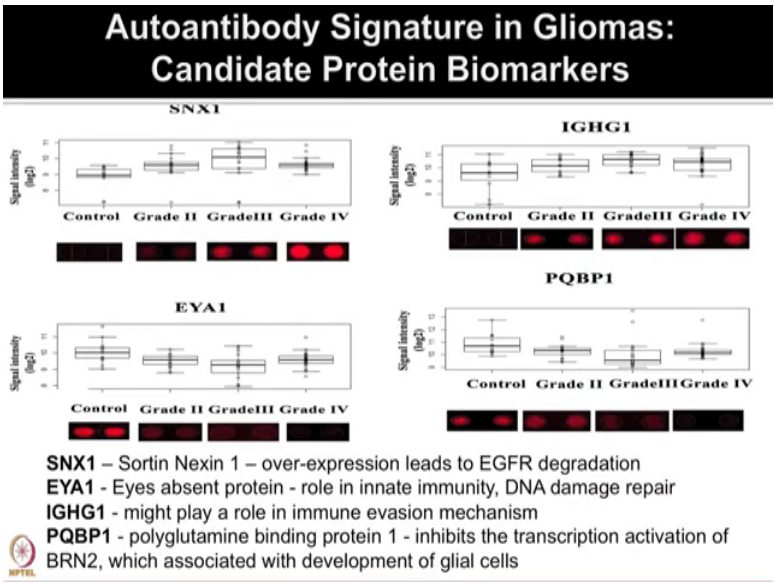
(Refer Slide Time: 20:42)



But of course, the success was not much. As you could see that the pattern was not very clear that you know they can the healthy individual what says gliomas we could not segregate very well based on our protein signatures. Then we started looking at can we look at control versus grade IIs, control versus grade IIIs, control versus grade IV, and as I started analyzing you know variety of these patterns we could see some train in grade III versus control and some train in grade IV versus control. But of course, given the nature of the disease which is very complex it was really not expected that we can segregate it you know magically from the healthy individual to the disease individual. And within disease so easily that which grade the these patients come from.

Nevertheless as I mentioned earlier, we found those four protein which were common in grade II, III and IV, and our idea was can we start investigating these proteins further.

(Refer Slide Time: 21:34)



And what are these proteins. So, these proteins as you can see on the screen one is Sortin Nexin 1 protein, which is usually shows over expression which leads to the EGFR degradation, lot of interesting biology is already understood for this protein. And this protein if you can you know look at from the control to grade II, you can start seeing some signal is coming from the grade II patients, and as it is progressing to grade III and IV, we can now see more intense signal in grade IIIs and IV.

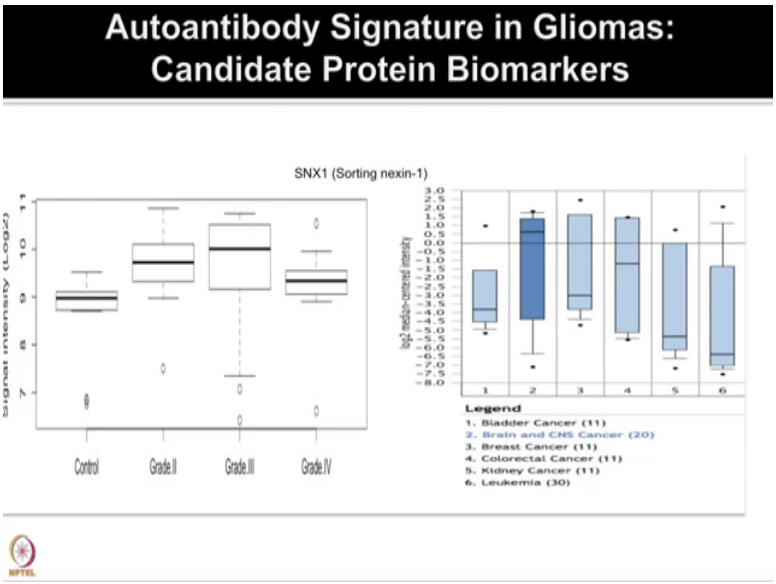
So, what we are looking at we are looking at can we find some protein which can be detected at the early stage of the tumor, and that is why I think that is the next one I look very interesting that you know even in the in the grade II patient we can start seeing those signal. If you look at IGHG 1, this is again one of the immunoglobulin family protein which might play

some important role in the immune evasion mechanism. And this protein again showed a very strong signal even in the grade II, and that remain consistent in grade III and grade IV.

Now, other two proteins, which we identified is EYA 1 protein and PQBP 1 protein. So, EYA 1 protein which is eyes absent protein that plays important role in the innate immune response, and also in the DNA damage repair. So, it is already known to play an important role in cancer. But what we are finding interesting that the level of this protein high in the control and as a disease is progressing from grade II, III, and IV, the signal for these a spots are going down, another protein which was polyglutamine binding protein one or PQBP 1 protein.

Again if you look at their signal from the control to the grade 4, the signal is going down, and this protein is again involved in inhibiting the transcriptional activation of another protein BRN 2 which is associated with development of the glial cells. So, ideally can we start looking at some of these biomarkers. And look at what are their trends across other diseases as well because many proteins might be showing response gen generic response for any tumor type. So, we started looking at the level of SNX1 across other tumor type as well by looking at uncommon database.

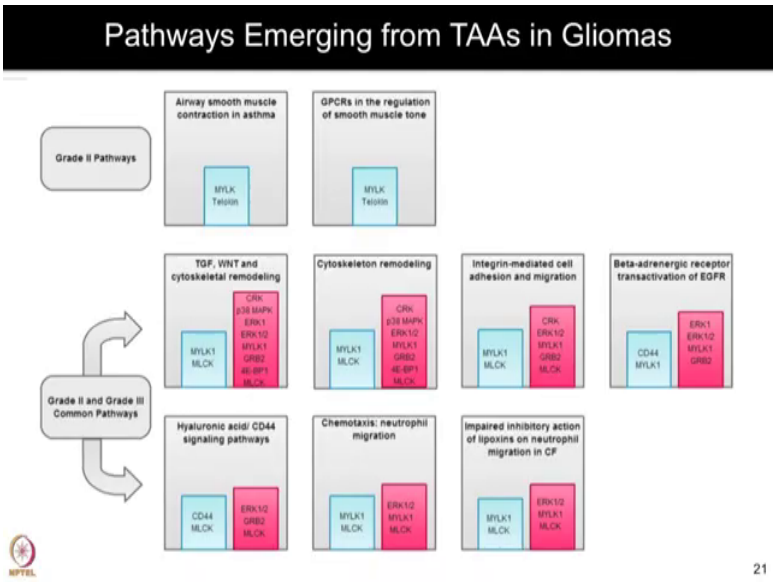
(Refer Slide Time: 23:50)



And what we can see here that if you on the right side, if you look at the image that specially in the brain tumors and CNS tumor the level of SNX 1 is very high, whereas, other tumor type it is showing the low expression. So, definitely this could be good marker good candidate to take it for further investigation for the clinical utility.

Additionally by looking at these tumor associated antigens, and we try to map them that which pathways they belong to. So, we looked at the grade II proteins, grade III proteins and grade 4 proteins that these proteins are mapped in which pathways. As you can understand that you know grade II, we had very large number of proteins, so we could not get some interesting pathways for that. But as a disease started progressing from grade II and III we see that you know many interesting pathways.

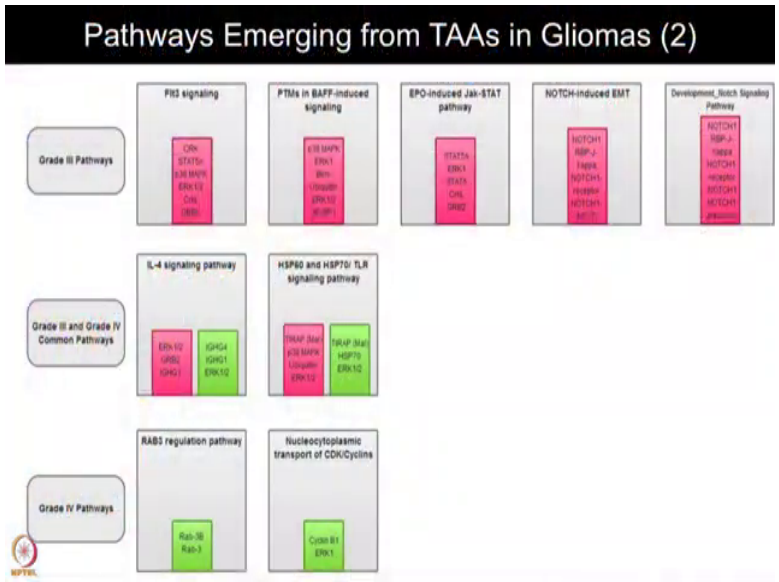
(Refer Slide Time: 24:37)



Especially the you know the TGF pathway, the wind signaling, cystoskeletal remodelling, integrin mediated cell activation signals, and variety of you know the androgenic receptors all of them started appearing on the interface of grade II and III. And as disease started progressing more, then we see there are more new pathways emerging.

So, again if you look at the bottom panel here we have hyaluronic acid or CD 44 based pathways. We have chemotaxis, neutrophil migrations; and we have impaired inhibitory action of the lipoxins on neutrophil migration in CF patients. So, some of these are showing how the disease is progressing towards the higher grades.

(Refer Slide Time: 25:22)

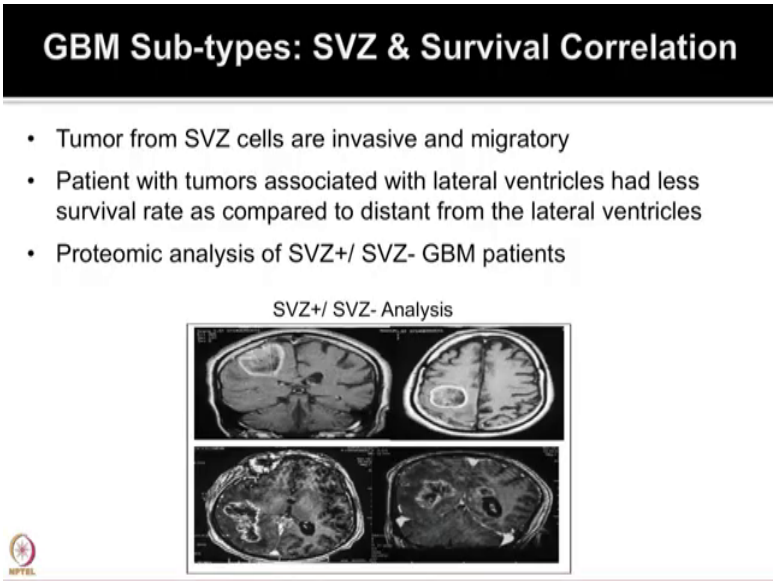


And again if you are now you know look at the disease progression moving from the grade III and IV, now we see some new pathways emerging like IL-4 signaling pathways, HSP 60 and HSP 70 pathways. So, in turn you know while doing this analysis what we realize that some new biology is appearing which we can now see that you know as it is progressing, there are new and new pathways which are also emerging based on the different candidates which we see the progressing from the low grade to high grade of gliomas. Of course, once you have the platform and each patient you are screening individually, then you can do many different ways of investigation.

Now, our tradition collaborators, they had interest that you know how these you know gliomas tumour multiforme patients which are the most aggressive type, some of them survive better, some of them do not survival; in some patients the tumor is very aggressive, in some

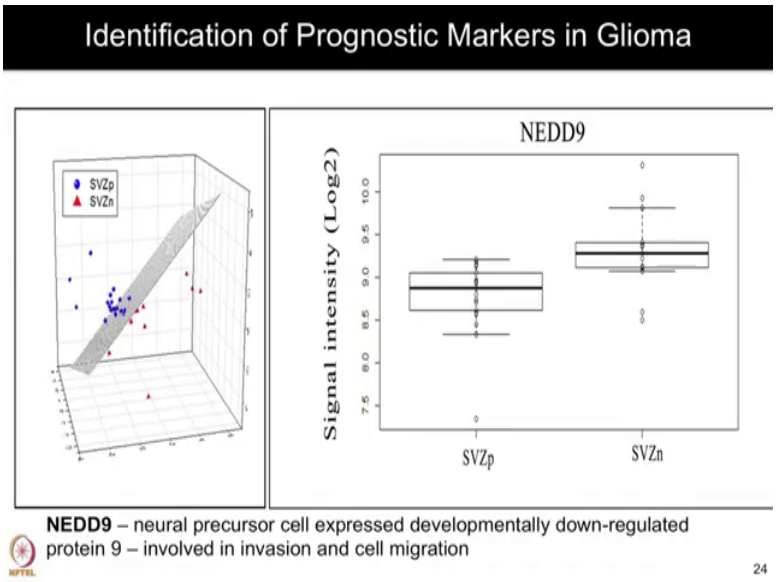
tumor is not aggressive. So, how best we could look at this kind of profile and try to look at the impact of the location of tumor on the patient survival.

(Refer Slide Time: 26:25)



So, they had some interesting observation based on the radiological images. And in this case as you see in the on the side, we have the patients where the tumor is very close to the (Refer Time: 26:32) region, and those are known as a SVZ positive or tumor is moving for a part from the SVZ region, and those are known as SVZ negative. SVZ positive are you know they are very, very aggressive tumors; SVZ negative are less aggressive tumors. So, of course, the patient survival and prognosis will be based on some of these you know features, and can we find out some proteins which might be interesting when we are looking at the impact of the location of the tumor based on radiology.

(Refer Slide Time: 27:09)

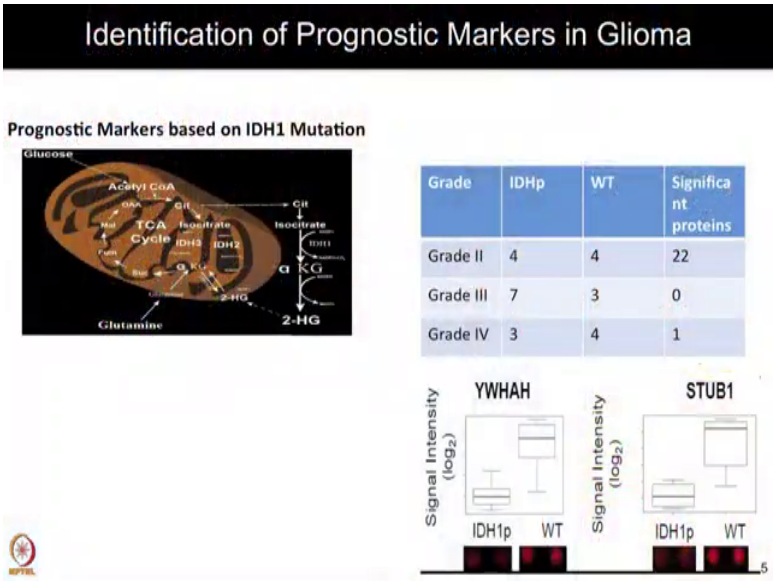


So, in this case, we identified an interesting prognostic biomarker which was NEDD9. And NEDD9 is one of the neural precursor cell expressed developmentally down regulated protein-9, which is involved in invasion and the cell migration. So, what are you talking right now is the same grade of the patient which are the grade IV gbm and within the same grade of patients some patients are having the tumor in a different location SVZ positive, and some in the away from the SVZ region which are SVZ negative.

So, it is very challenging you know problem for us to address because patients have the same grade, but just one minor change with the tumor location. Based on this we found very few protein which were actually differentially expressed from the SVZ positive versus the negative, but this protein NEDD9 look interesting and we still see some differences coming when we

compared for sizable number of patients within the SVZ positive and SVZ negative population..

(Refer Slide Time: 28:15)



Additionally we also looked at artists and prognostic markets of glioma based on the mutation effects. So, IDH 1 or as a trade the hydrogenise is one of the one of the very interesting gene which you have shown a lot of mutation in the glioma patients. So, not the next investigation we want to do we already have the data from individual patients from each slide. Can you now look at just the effect of these mutations on these patients and look at you know are there some interesting differences we see from the wild type and the mutant of the IDH population?

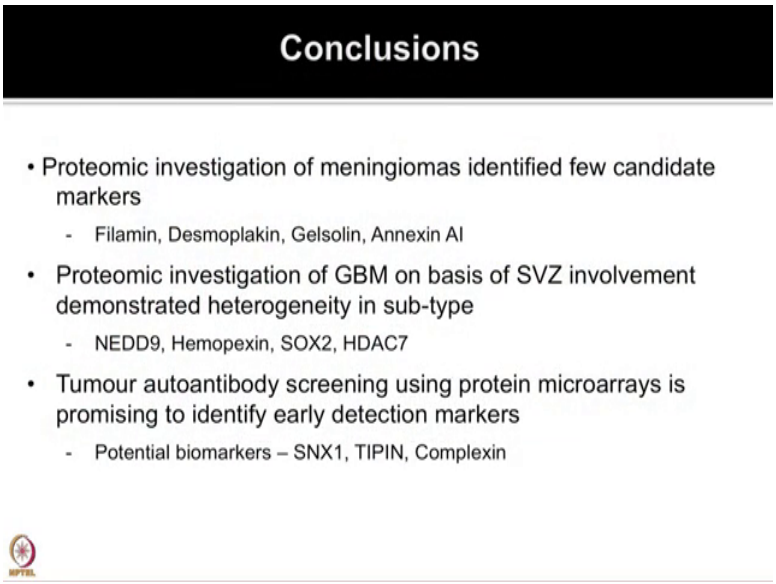
So, all these patients were also sequenced for the IDHG and based on that then only few patients where we could get the confirm sequence data for IDH mutation only those we took forward for analysis based on the protein array autoantibody screening. And in this case we

could identify almost 22 proteins, but two proteins which looks very interesting where YWHAH protein or an STUB1 protein. So, YWHAH protein that is actually known to be involved in the proliferation of glioma cells, and it was found to be upregulated in the wild type cohort as compared to the IDH positive cohort.

Now, it is possible that this kind of you know deregulation might be attributing towards the poor prognosis of wild type patients. So, we can conclude this part that based on protein arrays based screening in a very less effort, one could actually do the screening of thousands of protein from the patient serum sample. Of course, the most challenging part is the data analysis part, and doing the further data investigation using different types of pathway analysis and making biological sense of the data. So, doing experiments are much simpler as compared to doing the further data processing and data analysis aspect. Based on these analysis we found the set of proteins especially four proteins which looks like promising candidate for the early detection of the tumor especially from the serum sample.

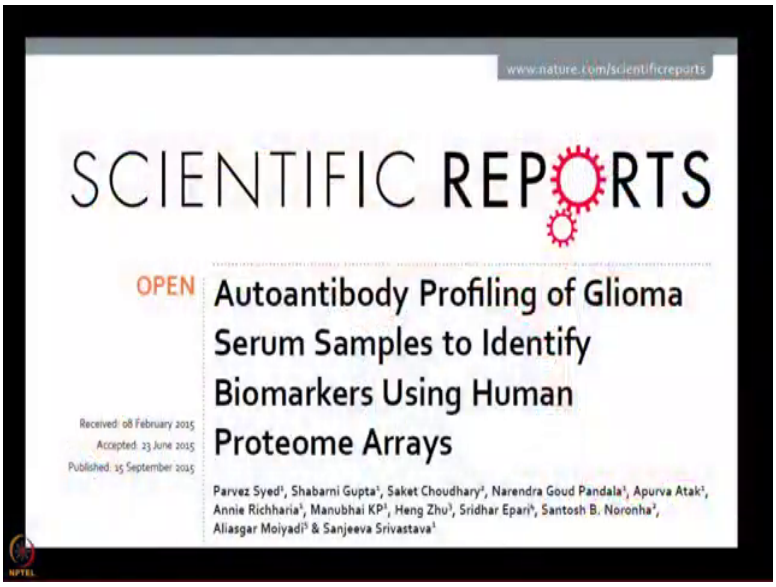
We also investigated the effect of the tumor location based on radiology with that could also reflect in the serum sample without antibody level. And one protein NEDD9 differently should promise. Additionally we could also see some other protein like hemopexin and (Refer Time: 30:33), those are also interesting protein in to differentiate these type of SVZ behaviour. We also looked at are there effect of the IDH mutation on these patients, and we could see around 22 proteins were different in IDH positive and the wild type population.

(Refer Slide Time: 30:50)



So, in general I hope you got a glimpse that how to perform the autoantibody screening using a patient serum sample, and what is the workflow involve, how to do the data processing, and finally, how to make some meaningful insight from the dataset.

(Refer Slide Time: 31:07)



So, this work was published in Scientific Reports, which was a you know big collaborative work from the clinicians from TATA Memorial Hospital from some of the technologists on the protein array platform from the Johns Hopkins University, and of course, my team and you know variety of Ph.D, students, and postdocs from IIT, Bombay.

So, again you have to also appreciate that in this kind of research, you need to make good teams and you need to bring the interdisciplinary strength to really try to achieve some very you know interesting information which is otherwise not possible to opt in all kind of you know specialisation just from your own way. So, good to build the team which are all having the different type of strengths, and then try to work on a given problem, and try to see what kind of meaningful insight we can obtain. I hope you got some idea that one could use protein arrays for the biomarker discovery program and more applications are going to follow soon.

Thank you.

(Refer Slide Time: 32:07)