INDIAN INSTITUTE OF TECHNOLOGY

NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING

(NPTEL)

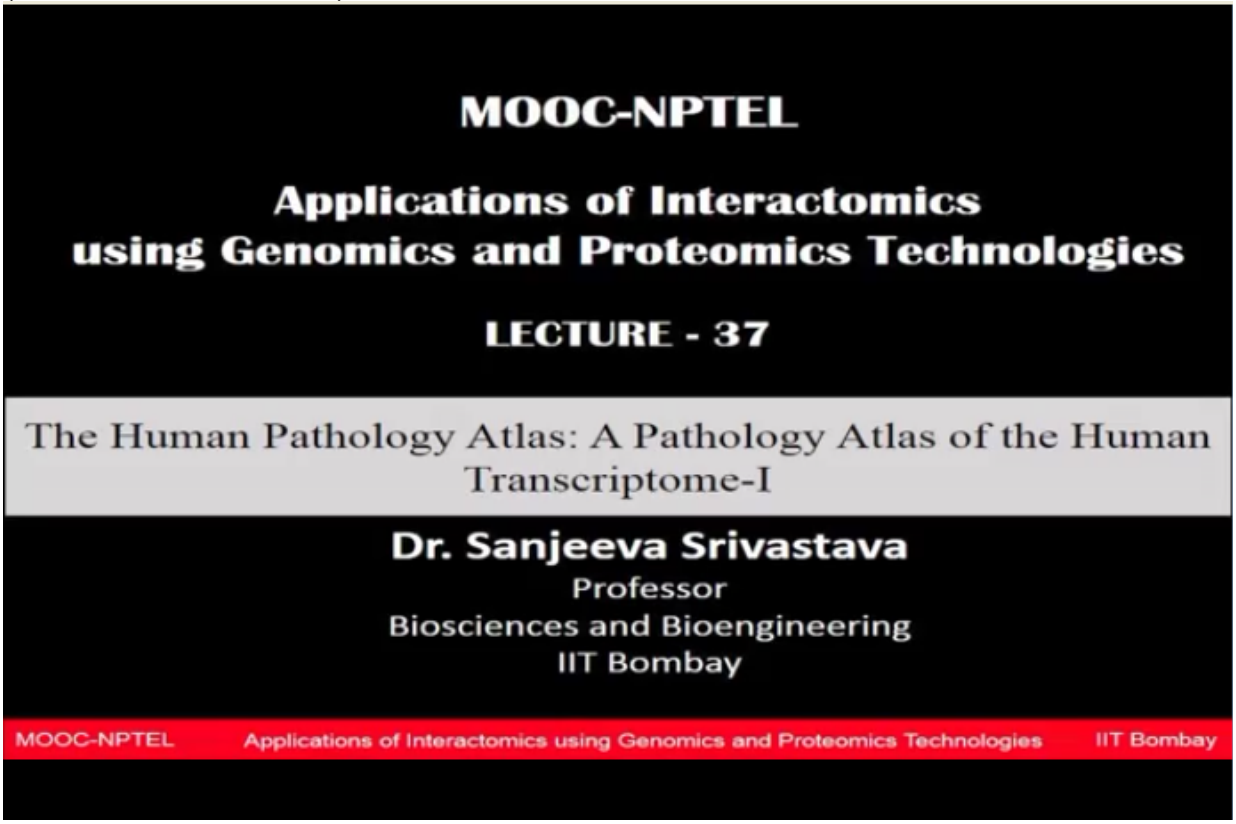
C-DEEP
IIT BOMBAY

Applications of Interactomics using Genomics and Proteomics technologies

Course Introduction by
Prof. Sanjeeva Srivastava
Biosciences and Bioengineering
IIT Bombay

Lecture – 37
The Human Pathology Atlas: A Pathology Atlas of the Human Transcriptome-I

(Refer Slide Time: 00:26)



Welcome to MOOC course on Applications of Interactomics using Genomics and Proteomics Technologies. Today we have distinguished scientist who is going to talk to us about the latest advancement in the areas of human protein atlas.

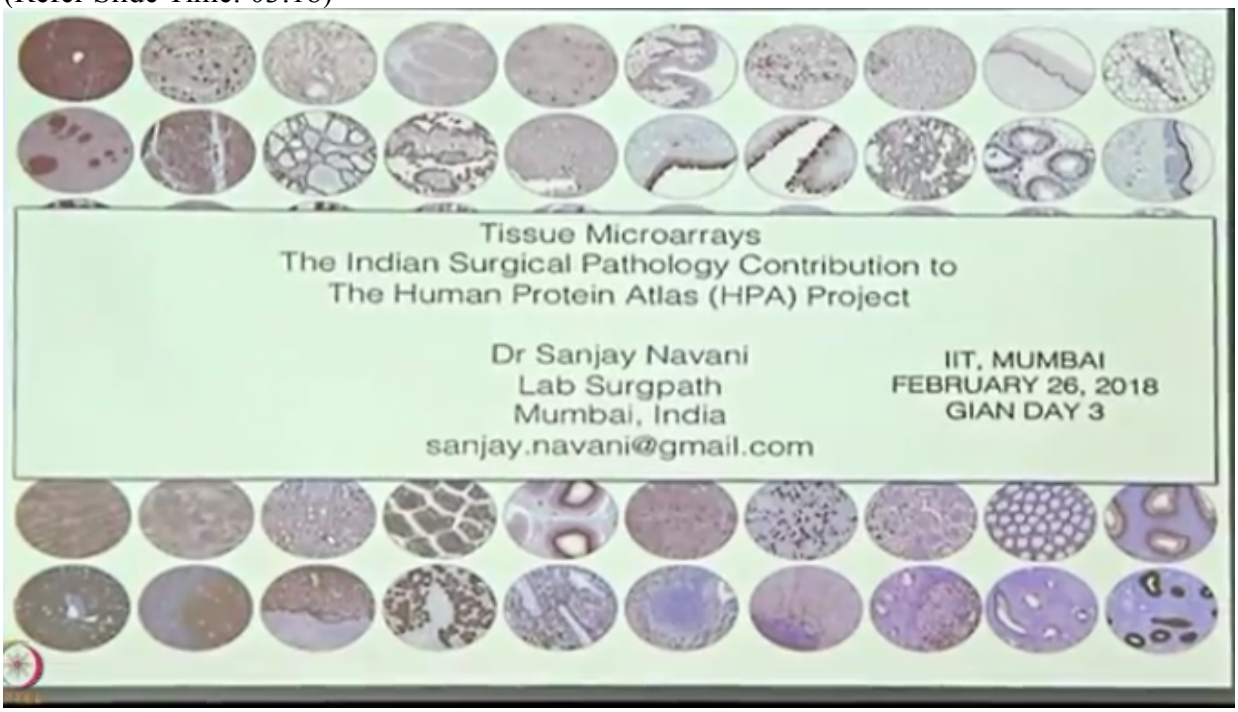
This lecture will be delivered by Dr. Sanjay Navani. Dr. Navani is a consultant surgical pathologist and immunohistochemist who provides surgical pathology services through his laboratory Lab Search Path. Lab Search Path is currently the one of the laboratories in the world which offers immunohistochemistry stained its light services for more than 200 diagnostics and 20,000 research laboratories. Lab Search Path pathologist have contributed immensely to the human protein atlas HV project by manually annotating approximately 60 millions immunohistochemistry images over a period of 8 years.

In today's lecture Dr. Navani is going to provide more detail about his contribution on pathology atlas. Just to give you some understanding that there is a quest to know all the proteins present in human, both using mass spectrometry base approach and the protein array base platforms. Dr. Mathias Uhlen and his group has really contributed immensely to the antibody waste array based platforms where they have tried to look for each protein specific antibodies and then trying to see where they are localized, this is one of the major task mammoth effort conducted by his team where Dr. Navani and his team has also contributed immensely for the pathology project, so I am sure we are all very excited to have Dr. Navani's talk today and let me welcome him for today's lecture.

(Refer Slide Time: 02:40)

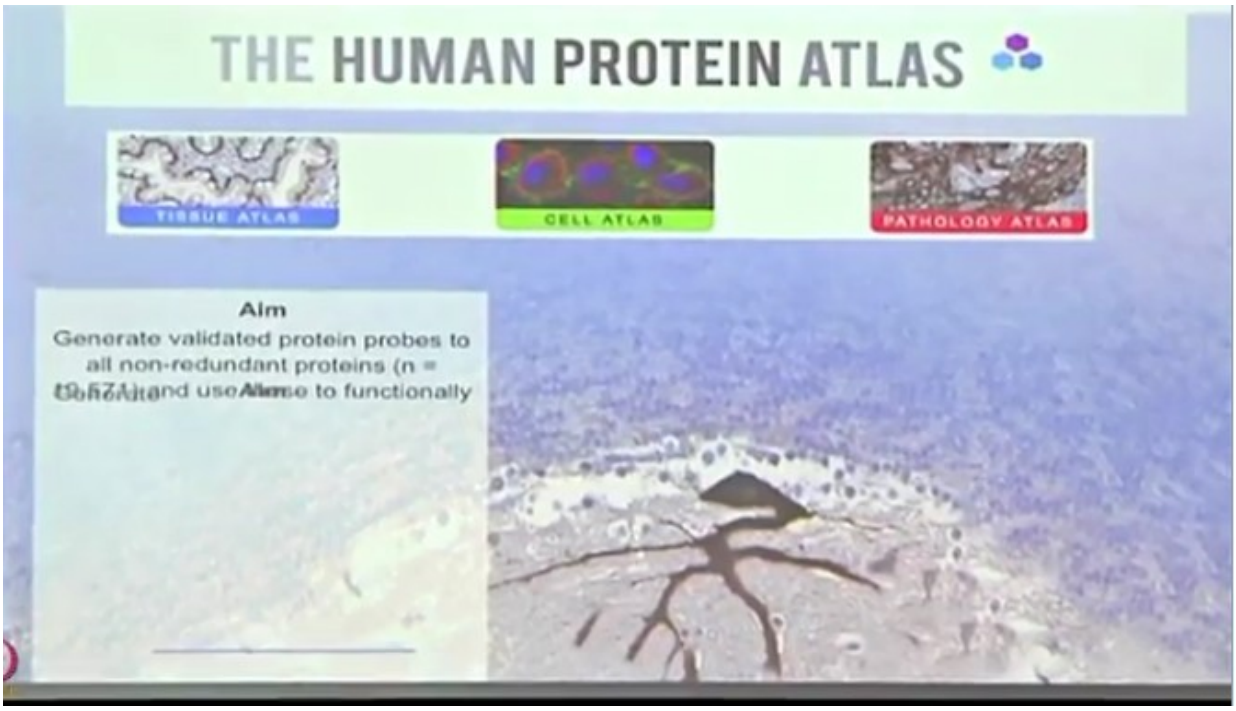
Dr. Sanjay Navani: So I'm the one is called the surgical pathologist, formerly trained as a physician with my MBBS, and then MD in surgical pathology which means that I'll look at biopsies and I do a lot of immunohistochemistry.

We are also the pathologist for the protein atlas, as you know all the images that you see on the protein atlas website has been annotated in India,
(Refer Slide Time: 03:18)



and I hope to tell you a little bit today about what, not only what we've done but what the protein atlas has actually currently doing.

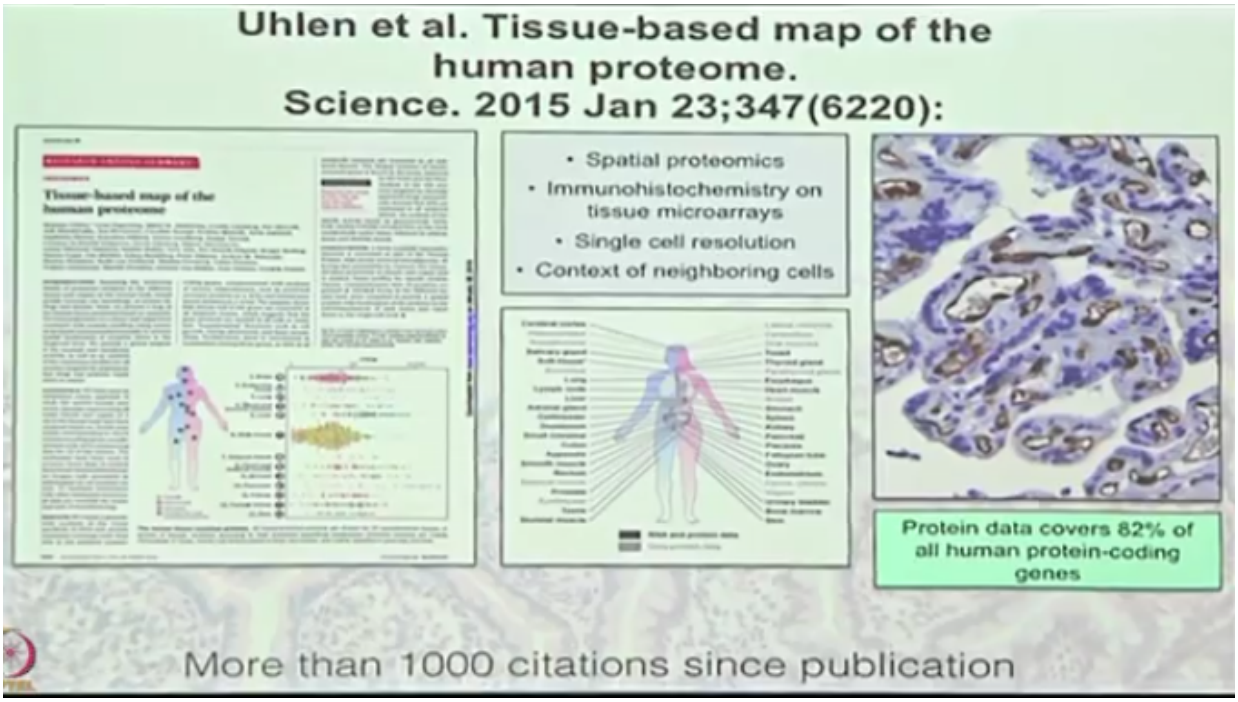
To give you a brief history about the protein atlas, it's started off in 2003,
(Refer Slide Time: 03:38)



the idea was to generate antibodies and study their presence in normal and cancer tissues, and to produce one antibody for every non-redundant gene.

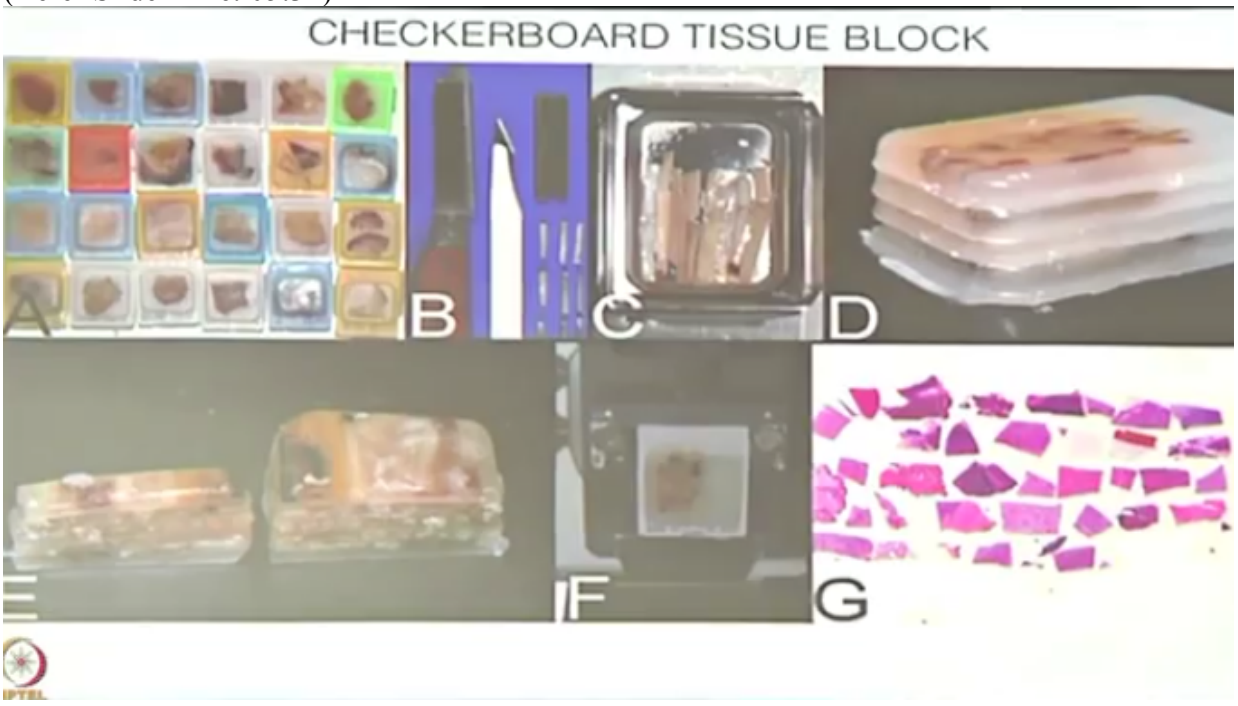
The protein atlas is composed of 3 main parts, the tissue atlas which is normal tissues, which we had the nice discussion about and I will like to expand some more on it, the cell atlas which is now the center of many things happening, and the pathology atlas which consists of the use of antibodies on pathology tissue, but mainly when I say pathology I mean that it's a cancer atlas.

First publication came out in 2015, that when I say first publication of course there were many publications before this,
(Refer Slide Time: 04:48)



but that was when we put out the full draft of the first human proteome as seen in normal tissues, the main platform for the protein atlas is immunohistochemistry and tissue micro arrays, and all of these images which are evaluated are available on the website for any researcher, student to see.

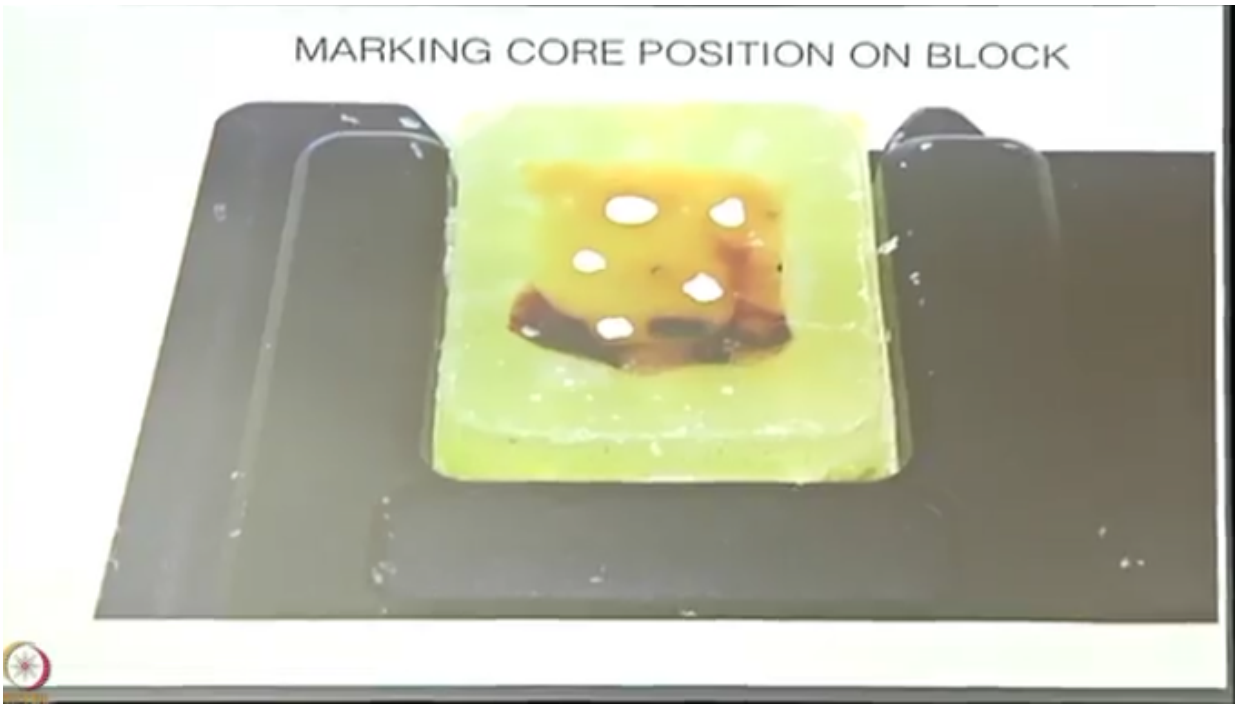
To give you a bit of a background about surgical pathology, and what it is, and how we interacted with this project, I thought I'll just explain what tissue micro arrays meant, and I put this slide in just to show you how tissue micro arrays actually originated.
(Refer Slide Time: 05:34)



With the advent of immunohistochemistry more reagents became available, and these reagents could tell us things about the tissues that we didn't normally see under a light microscope, if we apply them in sequence, so they told us that this looks like this, but it's actually got an estrogen receptor, and that was valuable information because we knew where that protein was being expressed at that particular time.

The problem with immunohistochemistry amongst other things was also that it was in continuous to be a fairly expensive investigation and therefore the expenses directly related to the amount of reagent that you use on a slide, so if your population for a study is 5,000 your entire funding could go down only in staining your slides, and therefore you see now the need why a different technique was required and why pathologies decided to start, to try and cramp all tissues on to one slide, put all the tissues on one slide, put everything on one slide, use the same amount of reagent, save your money, get the answers, now the origin of that idea which finally today is called tissue microarrays and forms the main platform of the HPA program, started with something like this, this was what was called a sausage block, so people took out interesting bits of tissue that they wanted to study from many blocks, from different patients, and they put it all into one block, and then cut a slide that look like this, so one piece of this was lump tissue, one piece of it was colorectal cancer, one piece of it was whatever you like, and with all these tissues on the same slide we put just one antibody and it stains everything, and our cost goes down 80 to 90%, so that was the idea.

The sausage block came about because people wanted controls for immunohistochemistry as you need with all test, you need to know whether it's working properly or whether it's not working, and therefore if you have multiple tissues on a block, some of which you know are positive and some of which you know are negative and then you have some in there which are also your tests, it's looking good, just to give you an idea how the tissue micro arrays are made, this is a usual tissue block, this is the kind of block that I usually use to give out a diagnosis,
(Refer Slide Time: 08:54)



so we've cut out a piece from a tumor, we have taken a slice of it, I've looked at it, I've said this is cancer and the reports gone.

And for diagnostics that is the only way that it must be done, tissue micro array is a research to it, it's not a diagnostic tool, so pathologies can't do that on every case that comes in to the lab, but once this has been established as breast cancer for example, we can use to take cores out of this tissue and put it into another block, where there will be tissues from other blocks, and finally we will create a tissue micro array block, I hope I am quite clear so far, okay.

Now the way that this is done is a slice of this, full piece of tissue is cut and give it under the microscope, the pathologist is asked to mark, characteristic, diagnostic areas in that slide which contains the whole section, so he or she does that, that slide is passed on to the technical person, and he is instructed to take cores from those places that the pathologist has identified as classic, because you must remember that when you see a tissue block of this size, there is tumor, there is trauma, there is normal tissue, so somebody needs to tell the technician where to put the core.

Unidentified Speaker: Is that tissue stabilize **_10:48_**

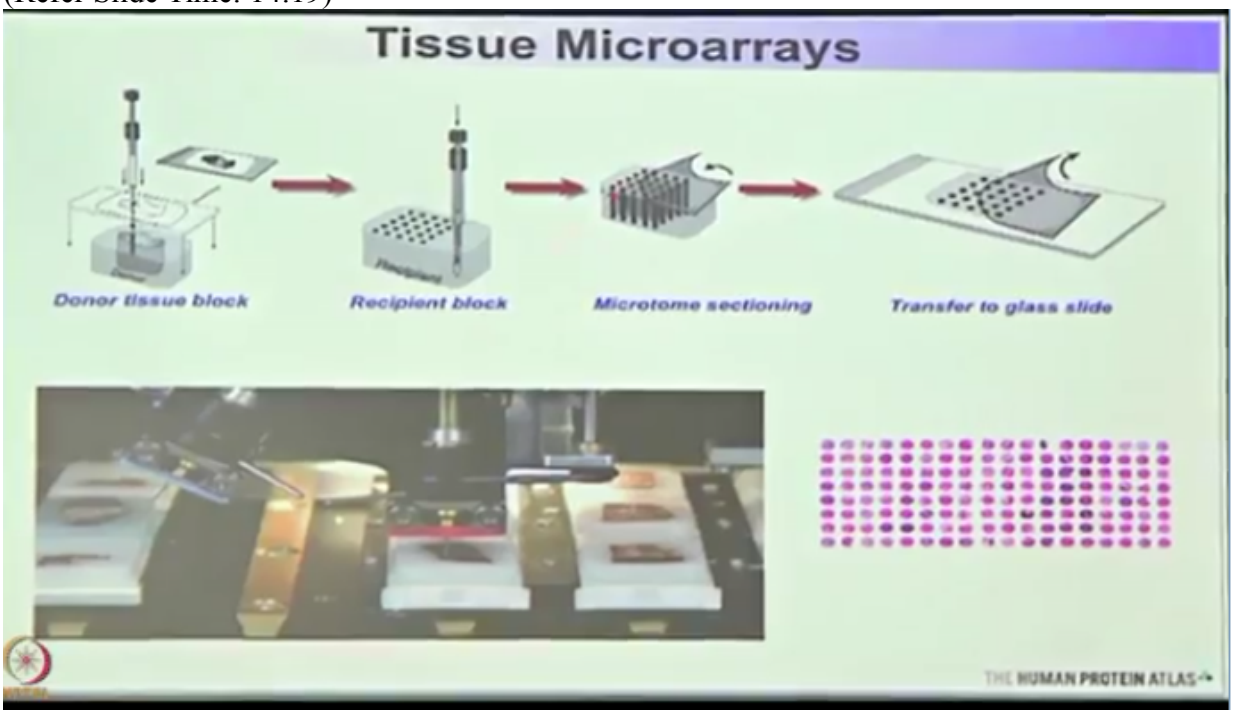
Dr. Sanjay Navani: Yeah, so that's a standard process followed in surgical pathology, as soon as the samples are taken out of the human body the specimens are in sized cut into small thin slices, usually which do not exceed 2 to 3 millimeters in thickness, and not more than one and a half to two centimeters in maximum length, and they are placed in a fixed of call formalin, once it's fixed and formalin the tissue is so to say preserved.

We then need to get it ready so that sections can be cut in viewed under the microscope, a process that refer to as tissue processing, and after the tissue is processed the tissue is put into this thing which is actually made of paraffin, so when you hear the term in the papers you read

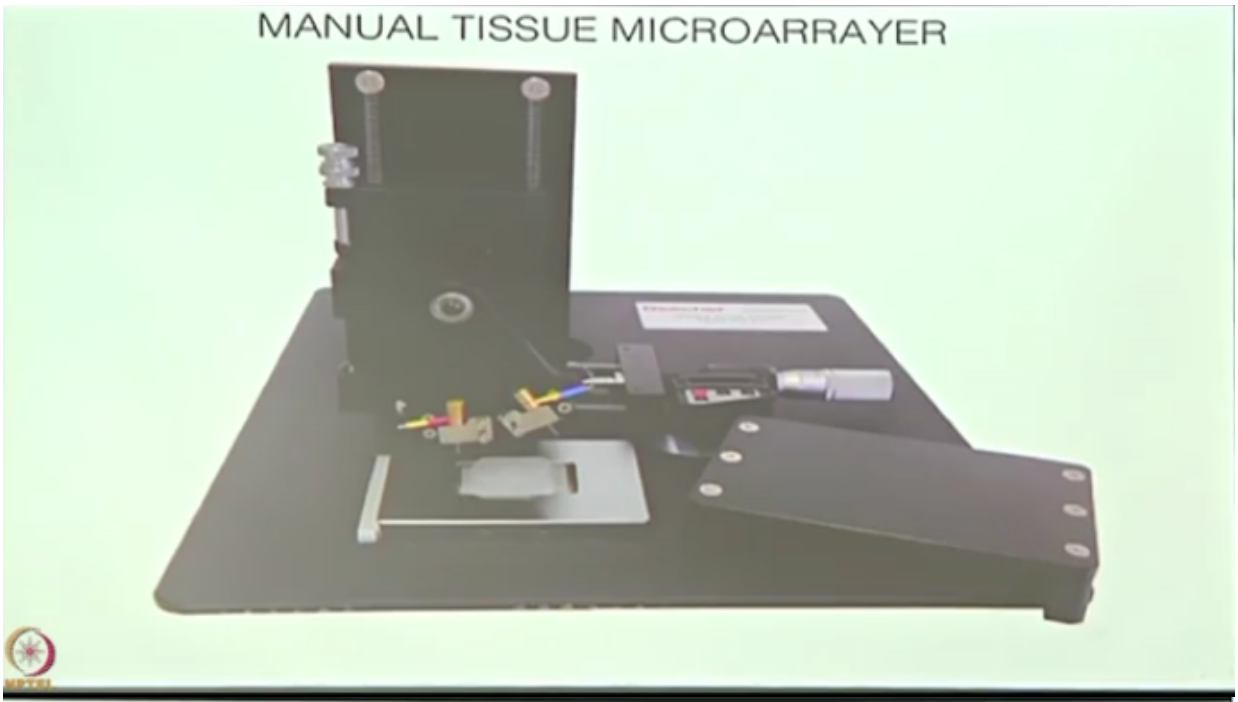
or the people you talked to, and people say FFPE that's Formalin Fixed Paraffin Embedded, amazingly immunohistochemistry has been able to do that, so when formalin fixes the tissue it denatures the protein or if I can quote an example that I have used, if this is your antigen after formalin fixation it becomes this, if there is no formalin fixation it's this, so if it is this there are epitopes regions of the proteins specific enough that can be recognized, to give you a clear staining pattern, but the point that you've raised is very relevant because it brings into question, the question of the antibody, so the antibody has to be good, has to be validated, we had a nice lecture from Dr. Joshua about validation, you can produce something from a rabbit, but does it actually work? That's the big question.

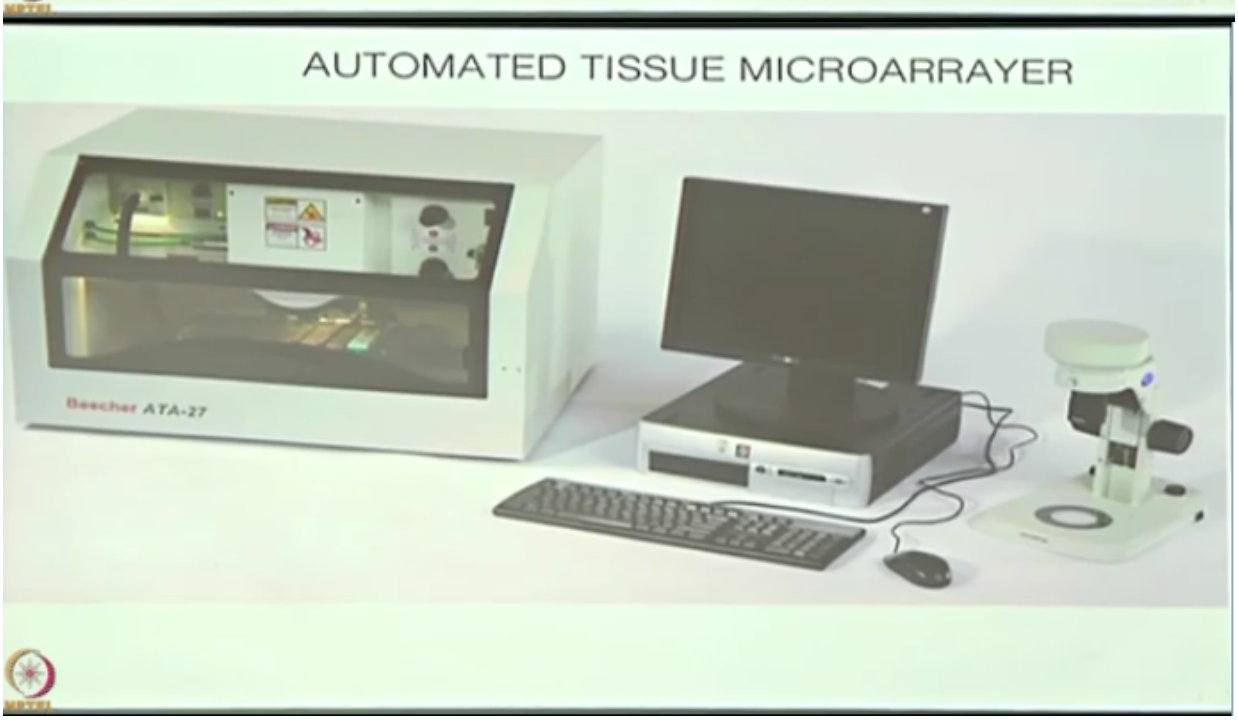
So we have about a 150 to 200 antibodies that we use in the clinic, so the antibody that I use for estrogen and progesterone receptors on the basis of which a women is going to get some drugs for breast cancer, they have to be very good, but most of the antibodies that I am talking about today are research antibodies, whose aim it is to get validated, so just that you know in the spectrum where the antibodies lie, so they are mature antibodies, they are new antibodies, they are antibodies with partial data on them, how good is your antibody, that's the big question.

Okay, let me move a little quickly otherwise we'll never get to the end of the lecture. This just shows you briefly how tissue micro arrays are made, so you take a core from that block that I showed you and you put it into a different block
(Refer Slide Time: 14:19)



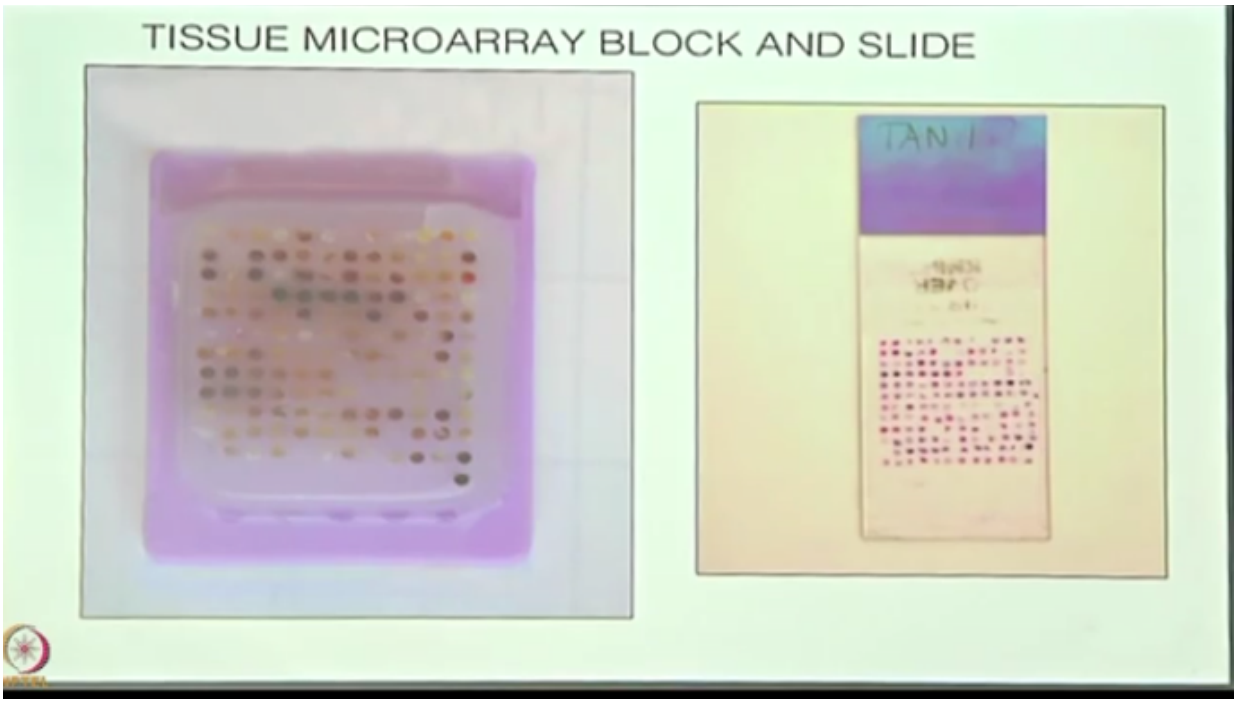
in which there is space for more cores, this is then cut as a section, put on the slide and then when you stain it this is an ordinary H&E stain that's how it looks. That's the machine that does, it's called the tissue microarrayer and there are many types of those, that's a manual tissue microarrayer, the one that we use in our lab, it's a fairly easy to use, good device, but as sweetest colleagues
(Refer Slide Time: 14:48)





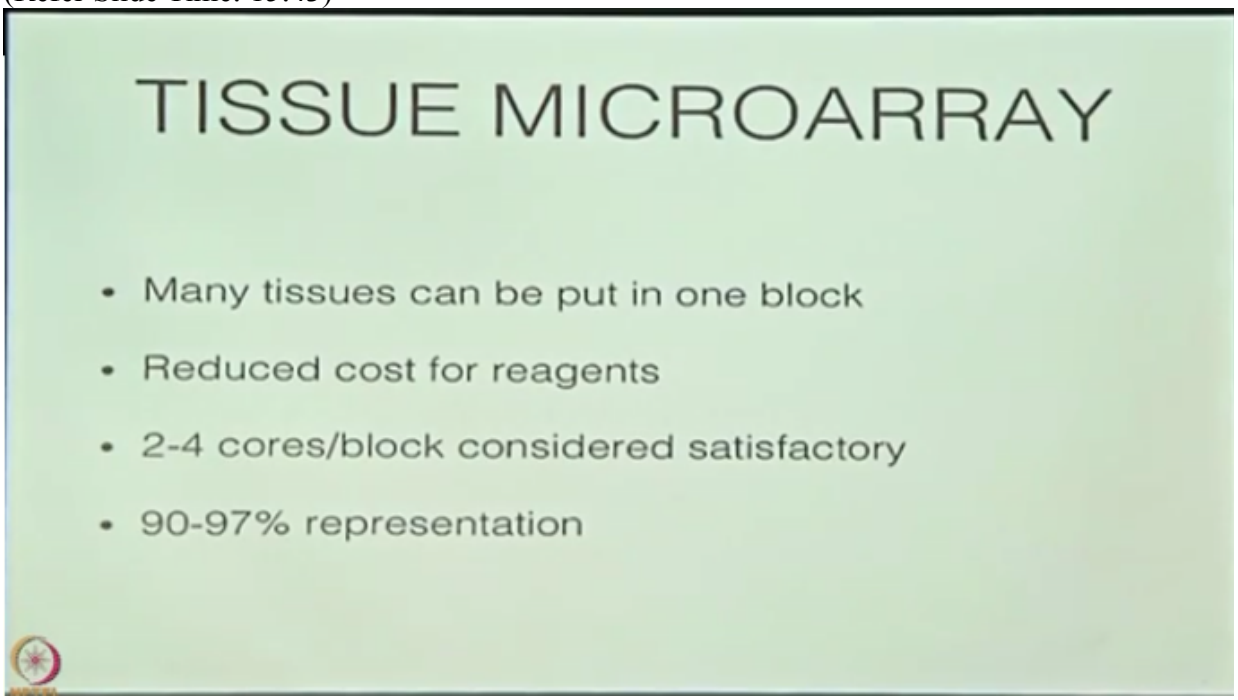
use more sophisticated tissue microarrayers that are automated, automated microarrays can make up to 20, 25, 30 microarray blocks a day, but it's an instrumentation which has its own problems, so if you don't have that workload it's not nothing to get into, you're much better off with the manual microarrayer.

(Refer Slide Time: 15:22)



And that's how the block finally looks, nice, neat, equal sized cores, and when you cut the section that's the slide, okay, now this is already been stained with an H&E which is a regular stain, but any kind of stain can be done on it, any immunostain can be done on it, so the advantage is which I've already outlined many tissues can be put in the block,

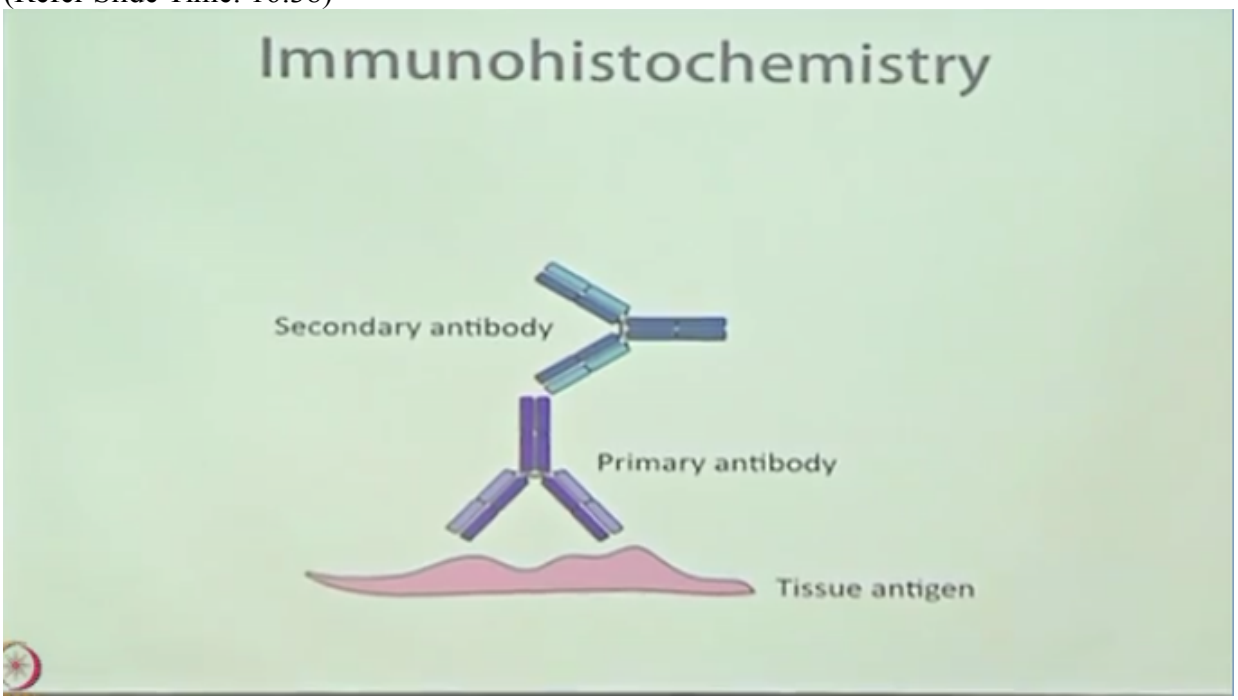
(Refer Slide Time: 15:43)



as a reduce cost for reagents, how many core should you take to obtain, correct representation? 2 to 4 cores is normally okay, most people internationally use 2 cores for a cancer, the protein atlas uses 2 cores for a cancer, but for normal tissues we use 3, and if you do that then we get a

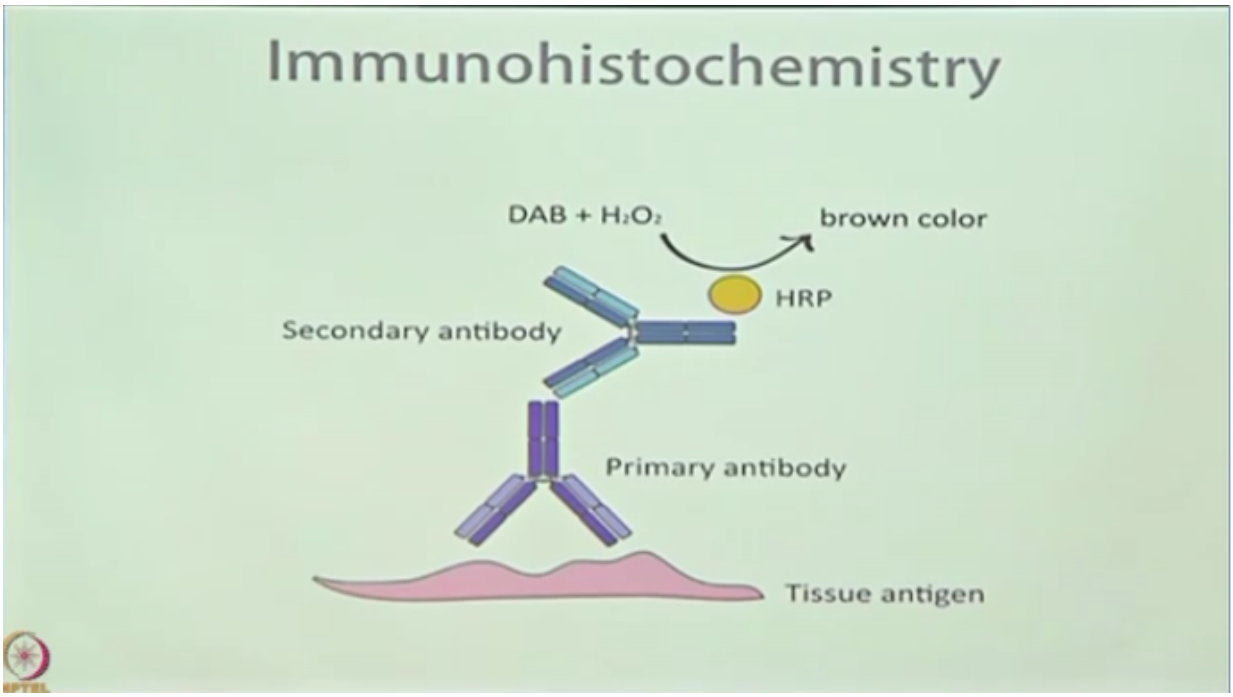
90 to 97% representation of the original block, provided the correct areas I have been marked by the pathologist.

So let me tell you a little bit about immunohistochemistry, you have this antigen which is in tissue which we can't see normally, you apply a primary antibody which goes and binds to the tissue,
(Refer Slide Time: 16:38)



then you apply a secondary antibody which detects actually the primary antibody.

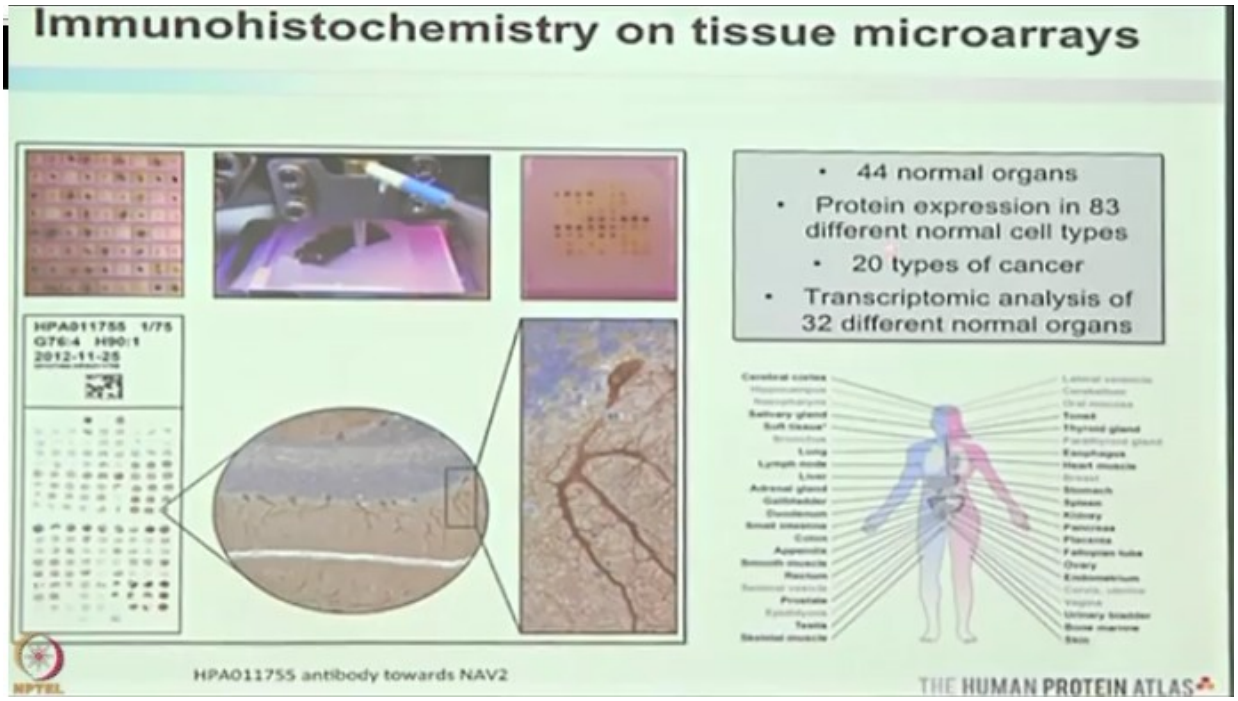
And finally you add a coloring agent which usually gives you a brown colour,
(Refer Slide Time: 16:48)



and when you see that you infer that the antigen is actually present, so everything depends on your skyscraper of immunohistochemistry standing up correctly, if any of those products are faulty including the tissue antigen which may not have been preserved properly, if it's not properly fixed in formalin, you will get a false negative, if you don't follow the other procedures use very highly sensitive reagents you will get a false positive, so it's the only thing which we have, in which you can see the tissue morphology and also see where that protein is localized.

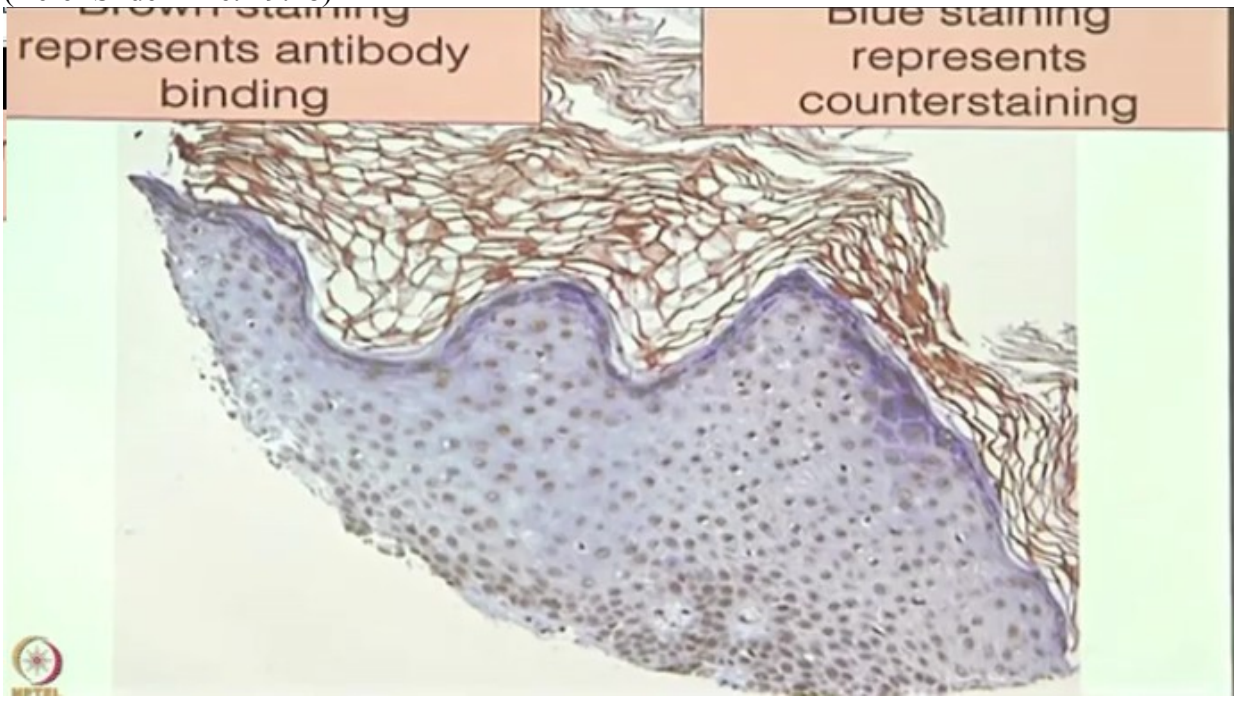
So now you understand what is the importance of using tissues to test your antibodies because then you know where those reagents are, where those proteins actually are, so that's a brief view of exactly what happens, a lot of blocks go and make a TMA, you cut a TMA, you stain it with immunohistochemistry, you look at that stain and that's what it looks like.

(Refer Slide Time: 18:19)



The protein atlas covers 44 normal organs, its studies the protein expression in 83 different normal cell types, 20 types of cancer and it has transcriptomic analysis of 32 different normal organs, and it was a very pertinent question raised about what is considered normal? Normal means in this case not cancer, most of these tissues have been abstained from sweetest subjects, who either underwent an autopsy or who underwent surgery for a non-cancerous condition, that's why in some of these tissues there is a mild inflammation present, but it's as close to normal as we can get, and that's how it appears finally, that's something what I see, so you see this brown stuff,

(Refer Slide Time: 19:18)



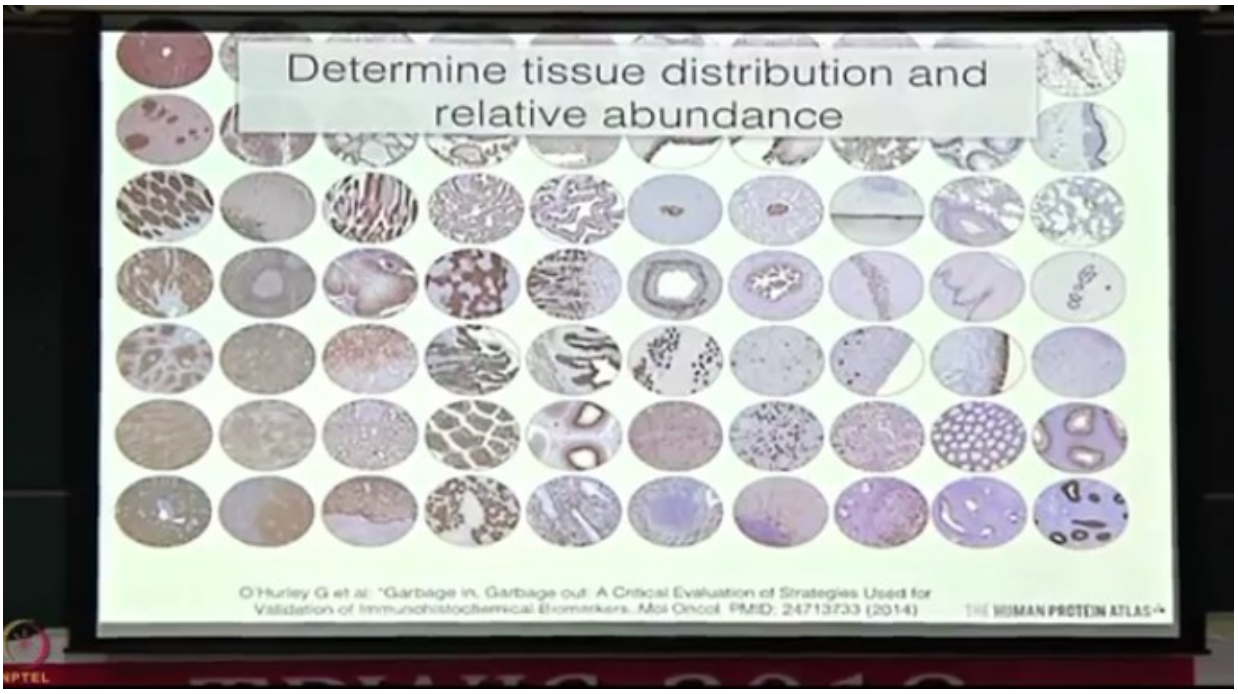
the brown stuff in the nuclei and the blue stuff, the brown stuff is the stain, that means the antibody is positive, the blue is the counter skin and its put only because it helps us to appreciate the brown better, if you are not careful with the blue and you leave the blue on for a little bit longer, it will mask the brown, so immunohistochemistry although if you speak to diagnostic pathologist or if this was a diagnostic pathology platform, you would hear me bossing about how we can do great things with it, but trust me when I tell you it's one of the most fragile platforms that you can ever encounter, today you will get it positive, tomorrow you will get it negative, your quality control, reagents, storage, it's got to be up there.

Unidentified Speaker: So related to that then are there internal controls that you may have, person that tell you that the staining is just right, is that how you do it or?

Dr. Sanjay Navani: So that's a valid point, and usually many tissues that we study at least on the diagnostic platform will have an internal control, and one internal control per batch at least and if we move to external then at least one external control on the same slide in the batch, so that's a very valid point Dr. Joshua has raised, how do you know that everything is working? So to give you the most common example of breast cancer and which we take up fragment of breast cancer tissue, the periphery still contains normal breast tissue, so when you put the stain there and you say it's negative you look up there, and it should be shinning, so you know it's working.

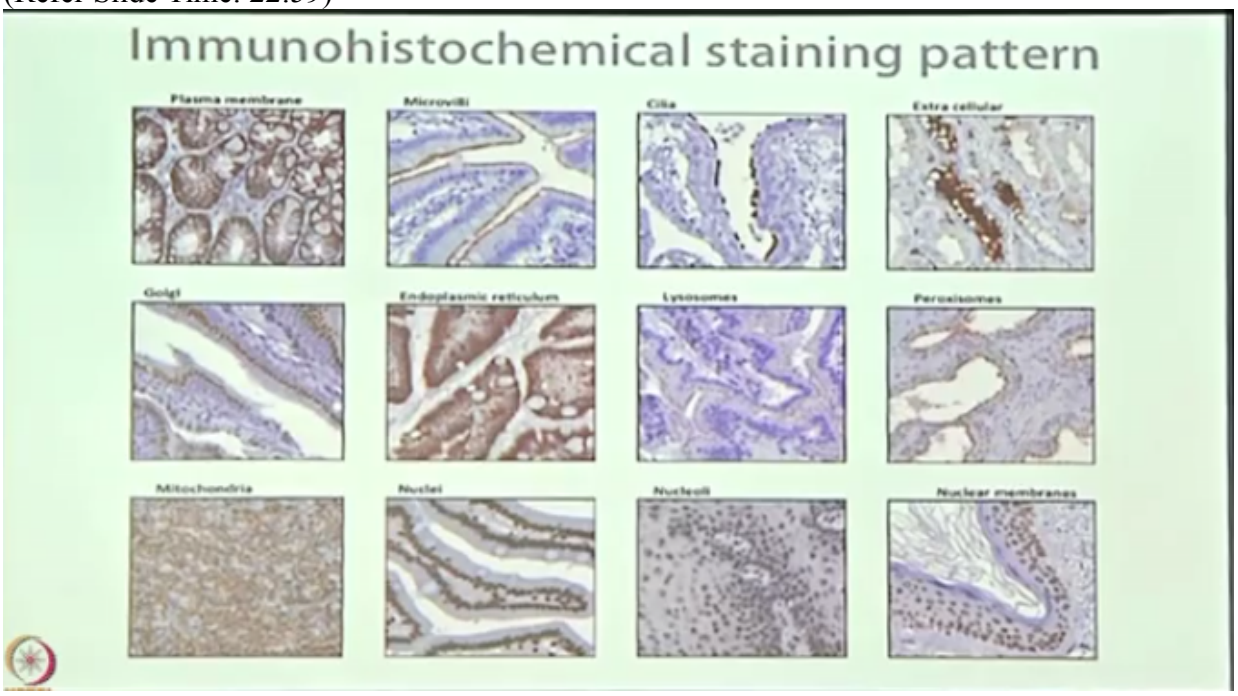
Now in this situation you may well ask the other question, yeah, but what if it's the false positive? It may not be, it's likely not to be because you should have then see in that in the tissue as well, not only in that place, the reason I'm going into this into a bit of detail is I'd really want to get across the point about how fragile a platform this is, so whenever you are getting any data on immunohistochemistry, please read it and thoroughly and look to poke holes in it, and please do not accept it at face value, it's most important, right.

So the idea after carrying out immunohistochemistry on these tissues with all the antibodies is to find out the relatives, distribution and the abundance,
(Refer Slide Time: 22:23)



where in the normal tissues is it present? And how much of it is there? Because you must remember that at the start of this exercise approximately 40 to 45% of the proteins were unknown, we knew only what 50 to 60% of the proteins do where they are, even today we do not know 100%, but certainly we have covered a lot of ground since then.

(Refer Slide Time: 22:59)

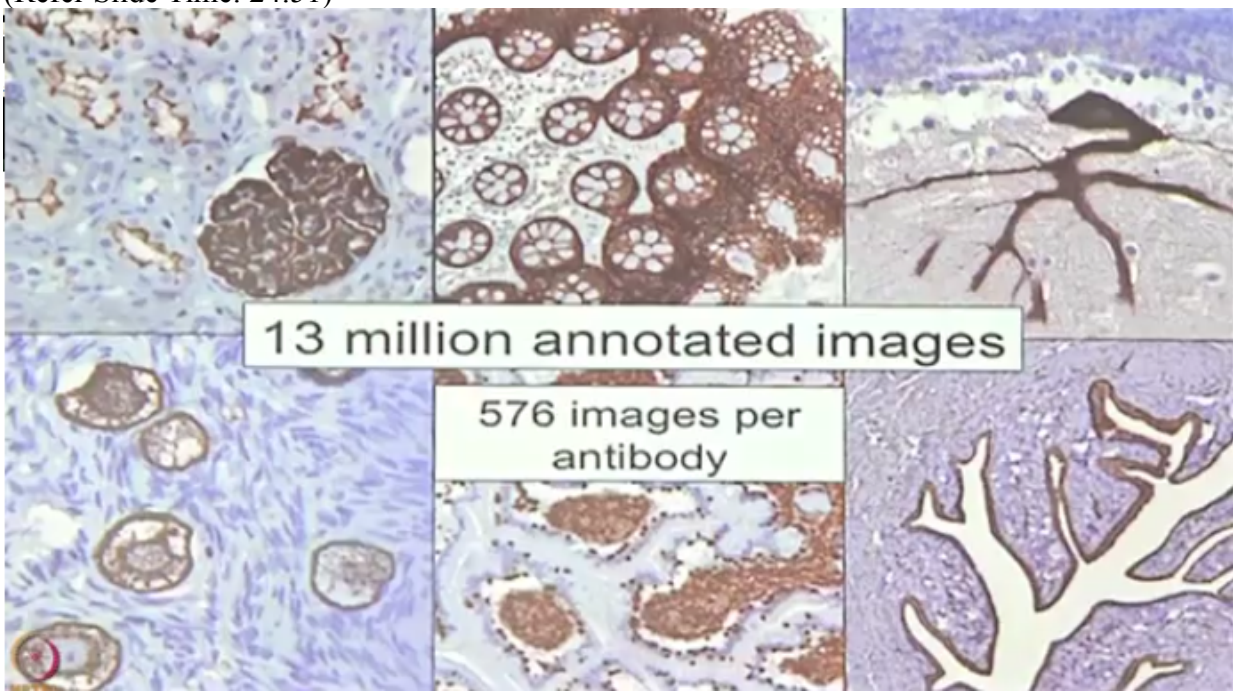


Unidentified Speaker: First of all that _22:58_

Dr. Sanjay Navani: How would that affect? Once the tissue is formalin fixed, the moment that tissue comes out from the body it loses its normal nutritional support in the form of blood vessels, and there are several studies which have documented that best results are obtained if that tissue is obtained within an hour of surgery and placed in a fixed, there are many ways to look at proteins, we are looking at immunohistochemistry in tissues because it gives us the localization, so what she says is correct, it's giving me the localization, but it's not actually correct, how would I check that? You use a different platform, and look whether you are getting the same protein in that platform also, I'll just come to that in a minute.

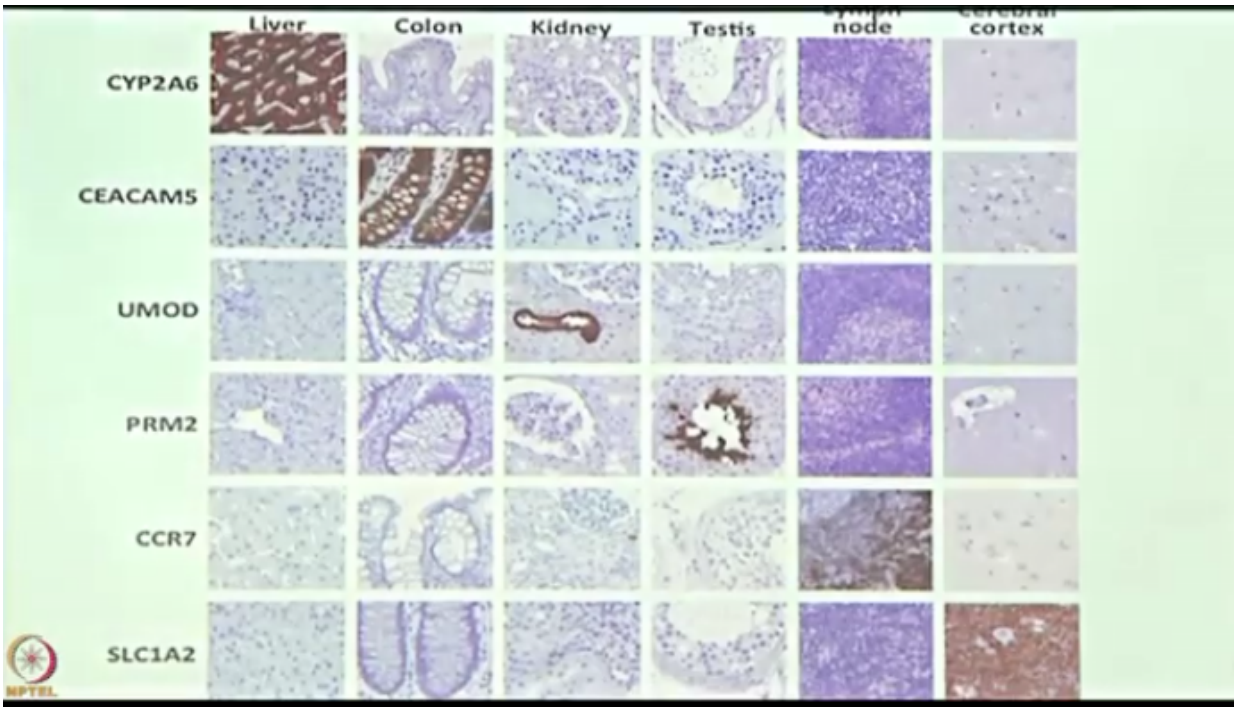
So with the antibodies that were produced, as you can see there was some lovely images that we got to see, and we were very happy about it, we were the first people in the world who saw these images, came to know where those proteins are located, a lot of them didn't have any localization patterns predicted for them, and we don't normally see all these proteins in diagnostic pathology, so it was in many ways a quite a visual treat for us to see and know where proteins were being localized, so that's the job that Indian pathologist did, we looked at 30 million images over a period of 7 to 7.5 years,

(Refer Slide Time: 24:51)



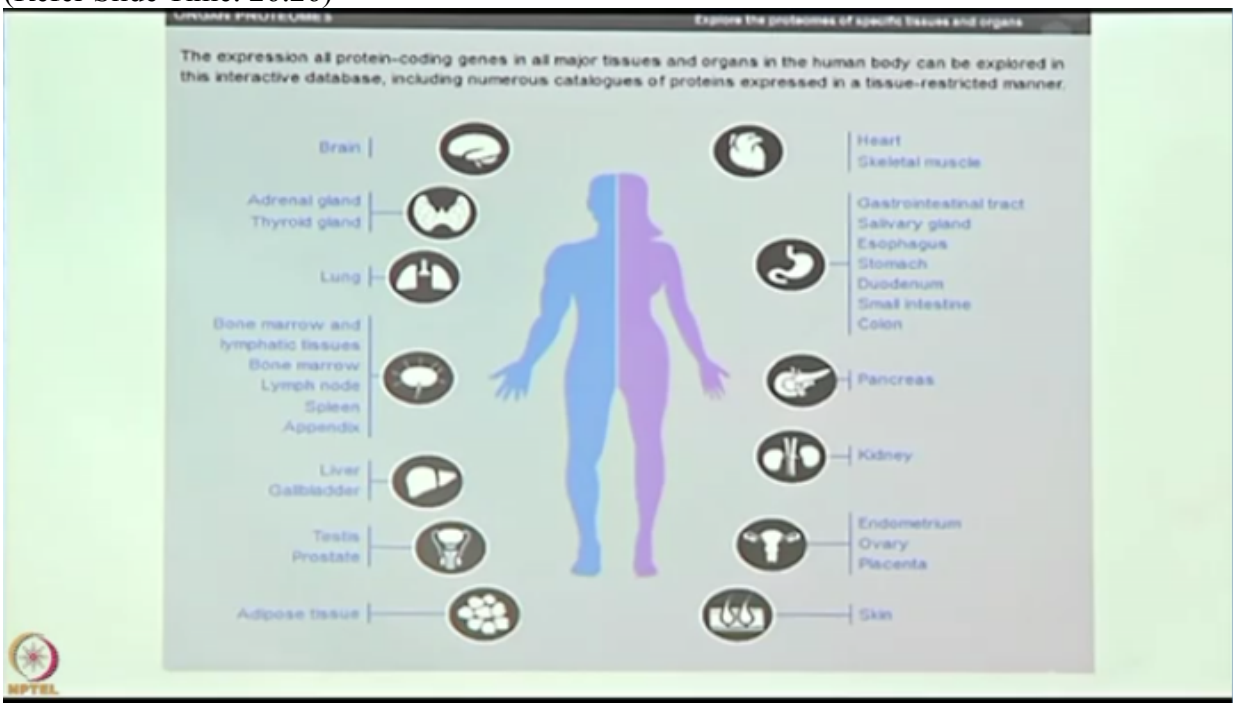
it's the largest effort in human history, and every image on the atlas has been annotated by an Indian surgical pathologist, there are about 20,000 genes as you well know and the aim of the protein atlas project was to produce one antibody for each one of them, so if you produce one antibody we wanted to see the global distribution means all normal tissues, 20 types of cancer, so that gives rise to very large number of images, so there were 576 images for each antibody.

(Refer Slide Time: 25:41)



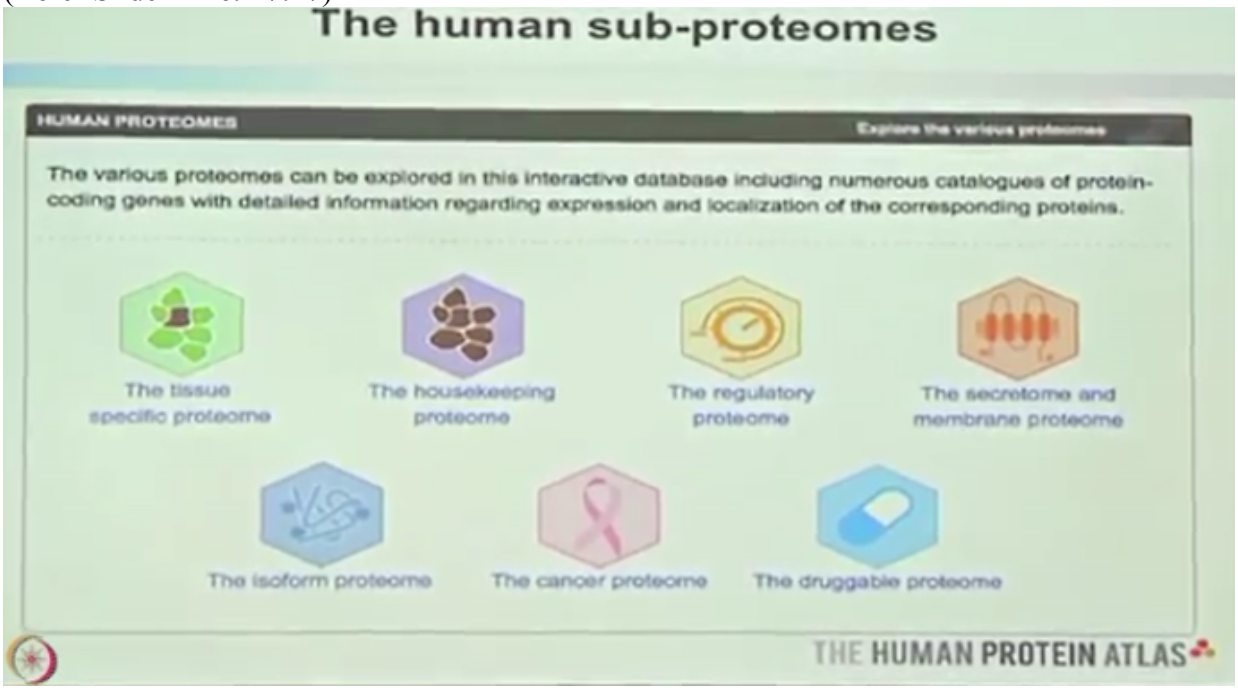
And that's a nice picture that's put together by one of my sweetest colleagues which shows you the organs on the top, the proteins along this side and how each one of them is showing a specific pattern in a different organ, I must point out to you that this is a nice picture to show you in the lecture, but I don't think that you should be going out expecting that's how it's clear is going to be, this is just to get the point across that there are some proteins expressed in some tissues that are vastly different from others.

(Refer Slide Time: 26:26)



If you go to the protein atlas you will find this caricature and if you click any of these links, and you can do that, you don't need to register, it's a freely accessible site, anybody can use it and if you access anything, say you had click pancreas, it will immediately take you into the genes that we found in the pancreas, and you can take it further from there.

You can actually also see the images that we annotated, we get several emails from people all over the world, some of them complementing, some of them not so happy, after looking at the images and giving us their feedback, we basically classified the normal tissues into 7 different types of proteomes,
(Refer Slide Time: 27:17)



and this is also available on the website, everything that I say is available on the website, and in far greater detail, so there was the tissue specific proteome that is specific proteins in specific tissues, there was also the house keeping proteome which I'll tell you a little bit further down the line about how important that it is, the cancer proteome which I'm going to just start talking about.

And very importantly also the druggable proteome, most of the targets in the pharma industry are proteins, we manipulate them and then we get the desired result with some side effects, in the pharmaceutical industry particularly the bio discovery area, nobody speaks of using a drug by itself, they always talk about benefit risk ratio, because everybody knows that a drug cannot be only good, therefore how should one look at these targets and is there something that we can do that will make these drugs do their job and at the same time not give a side effects.

Unidentified Speaker: Sir, I have one question, like in this tissue specific proteome, there are many proteins which expresses as specific proteins, but many a times I have seen that they are dependent upon stratification also.

Dr. Sanjay Navani: It depends on many things going into the tissue specific proteome, every tissue proteome I don't think this is in the scope of this lecture, but I would very much encourage you to visit the website.

Unidentified Speaker: Sir, I have one more question. The clinical samples, they will give you the lock masses of the tissue.

Dr. Sanjay Navani: Yeah, yeah**.**

Unidentified Speaker: And the same kind of disease there can be different different tissues.

Dr. Sanjay Navani: Yeah.

Unidentified Speaker: So, after fixation how you counter or how you reduce that biasness?

Dr. Sanjay Navani: That's the job of the surgical pathologist. The surgical pathologist can look at it and tell you where to classify it, that's what a person like me does. Each tissue looks different, so if you don't tell me what the tissue is and you sleep it under my microscope I can tell you what it is, and I'm glad you ask that question at this time, because you might be wondering whether you are in an abstract art class or in the middle of the science lecture, but I'm going to try and make a point about surgical pathology.

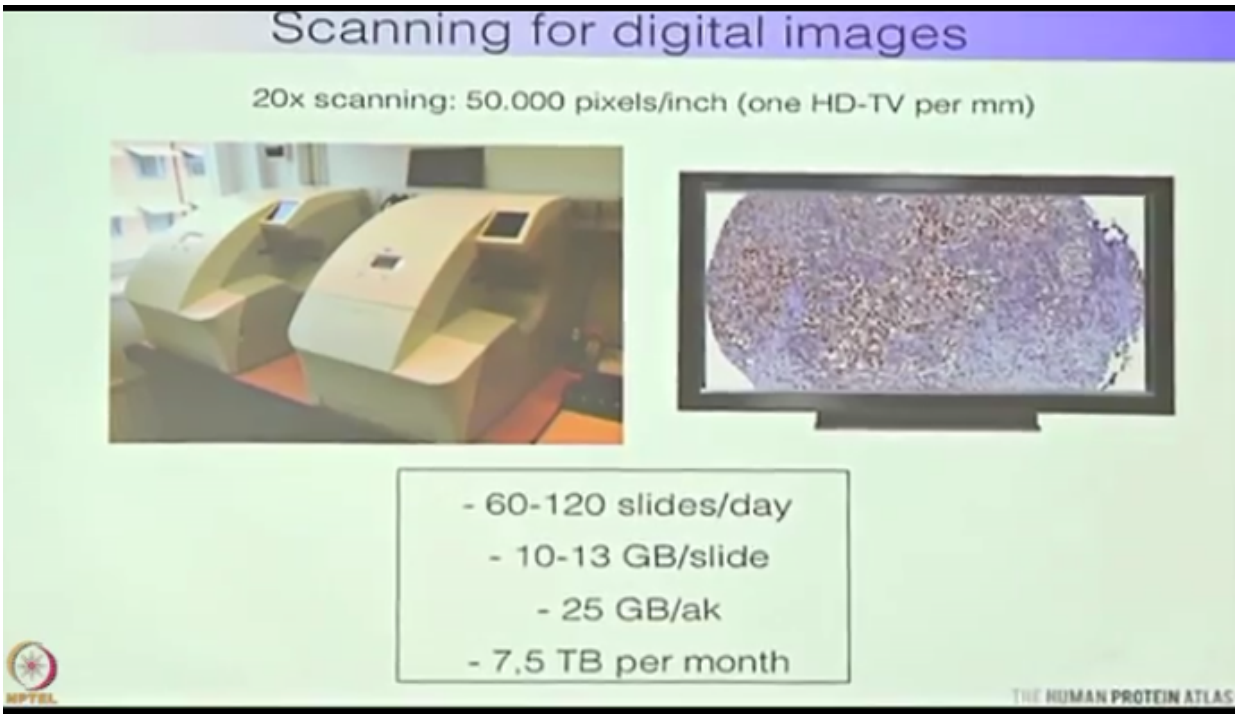
So when surgical pathologist first see H&E stain slide under the microscope, the first thing they see is something like this, there is a lot of things happening, some big, some small, some tiny, there is no central focus, it's one picture that you have taken fully, and then you begin to start thinking what's important.

If you look at it for a little bit of time, it crystalizes a bit, and you begin to think yeah that actually I think this looks like branches of a tree, but it's no, there are no leaves in it and I think its branches of a tree but I'm not sure, and then we look at it some more and then we say yeah now I'm sure, yeah definite, but there are no branches on the tree, so I'd like to equate this image with what a pathology sees on a normal H&E section, when he sees a H&E section and he comes to this point, this is sure, this is a tree, they are no leaves, and this is it, when I say I'm sure I mean as much as can be possible, maybe some of you didn't noticed Joshua laughing about it, because he is probably encountered the problem, people see the tree and they say this is not a tree, but I'm just trying to build a grade to get my whole concept there.

Now that's all that we see, if it's so clear 10 out of 10 pathology should tell you that's a tree, and that's a big job in itself to get those 10 jokers to agree is a big thing, so there will tell you now this is a tree, now starts and that something we've been doing for more than 100 years, now starts what more can you tell me about it,
(Refer Slide Time: 32:20)

because if you go only on this then all 10 trees should die without water, but they don't, only 3 die, 7 do very well, so if you have 10 invasive carcinomas of the breast which are staged alike, which look the same, and then why she is dying and she is having a party, there comes in the question of is there something as, so the first immunos might help you and they are exemplified by the leaves, so that looks different, that's not that tree, that's going to die, this is not going to die, that's what we are trying to do, and I am using breast cancer only as an example, but this is very far from the ultimate holy grail I should say which is we want to say it like this, there is a tree, there are leaves, they have different colours, this is yellow, this is orange, this is red, this means that, that means that, well you are in the middle of it right now, that's what the whole effort is about, which protein goes where, what is its role, is it going to fall off, is it going to continue, how important is it, so now you get the picture I hope from the surgical pathology point of view, right.

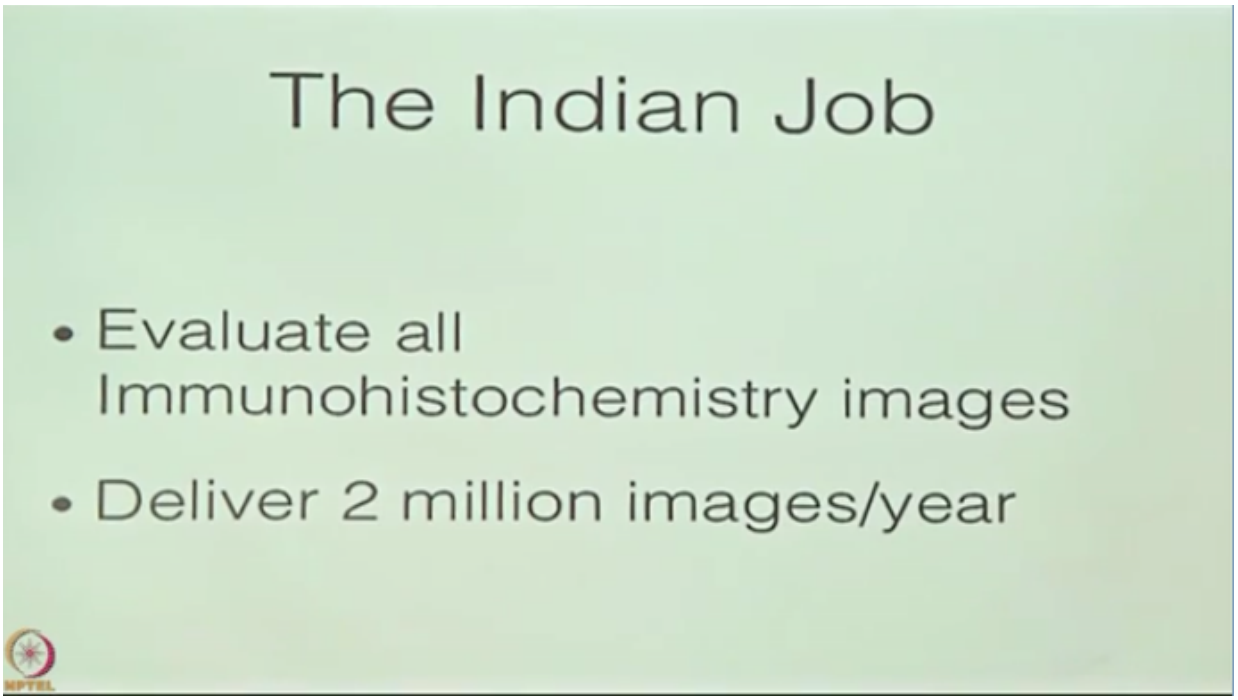
Let me go on to the next slide, when all these immunostains are done on the tissue microarrays, if you give that slide to the pathologies with all those 150, 200 or sometimes even 1000 cores they are never going to talk to you again,
(Refer Slide Time: 34:26)



because it's a physical challenge to look at that each core under the microscope, and then say that, that core shows this, it's one of the most frustrating exercises in the laboratory, so comes in the need for the digitization, just scan the whole slide and present it to the pathologist one core or two cores at a time, and if you add a software to that, so that they can go and click of the boxes then the pathologist is going to invite you for dinner.

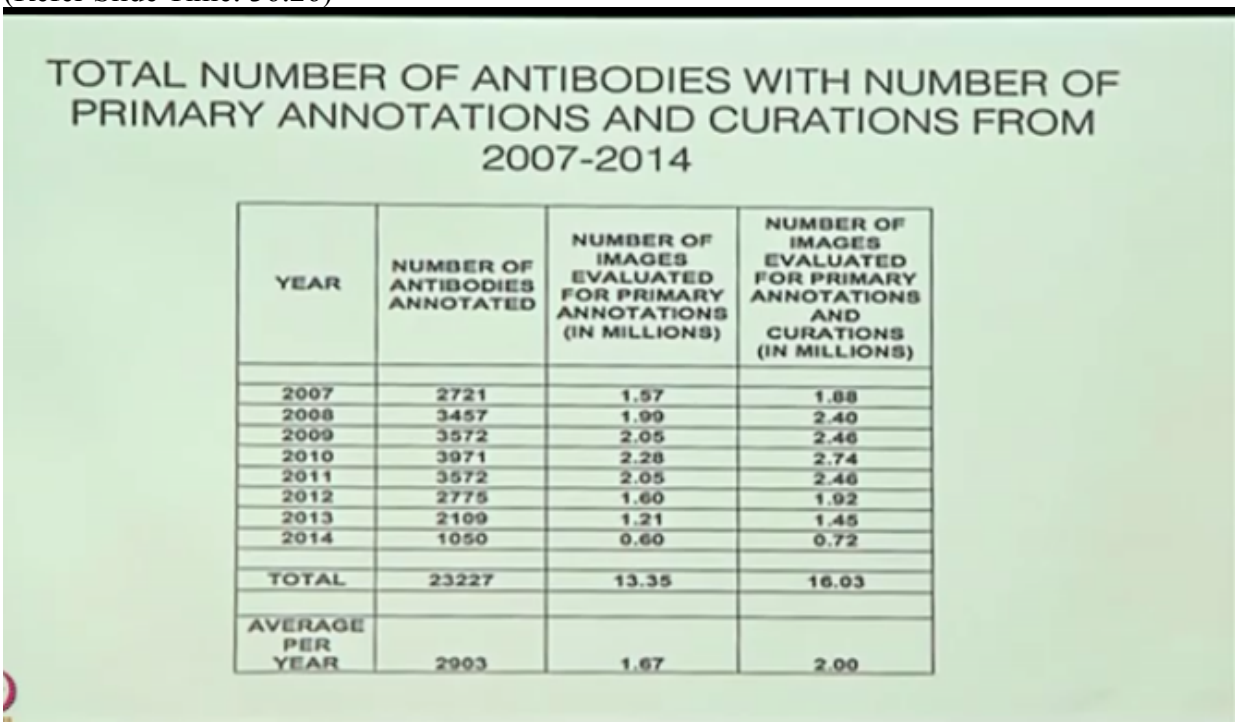
So these are some of the scanners that I used, and I think the ones that we made use off in this project, internet in India was just coming in a big way, the big lines you know we haven't had these high speeds more than, for more than 12 to 13 years, actually less than 10 years, and the big challenge was how would all these images be transferred to India for the pathologies to see it, if you don't have internet that fast enough, so we actually download it in several million images and it was a difficult time for us to work it, because when we started the speeds were not that fast, and people use to be waiting and you would get only half the core, then a glitch in the line, then another half, now luckily over the last 5, 6 years it's much better, so that was our job when we took it on,

(Refer Slide Time: 36:06)



we had to evaluate all the high images and we had to deliver 2 million images a year, 2 million images a year is roughly 7,000 images a day manually.

(Refer Slide Time: 36:26)

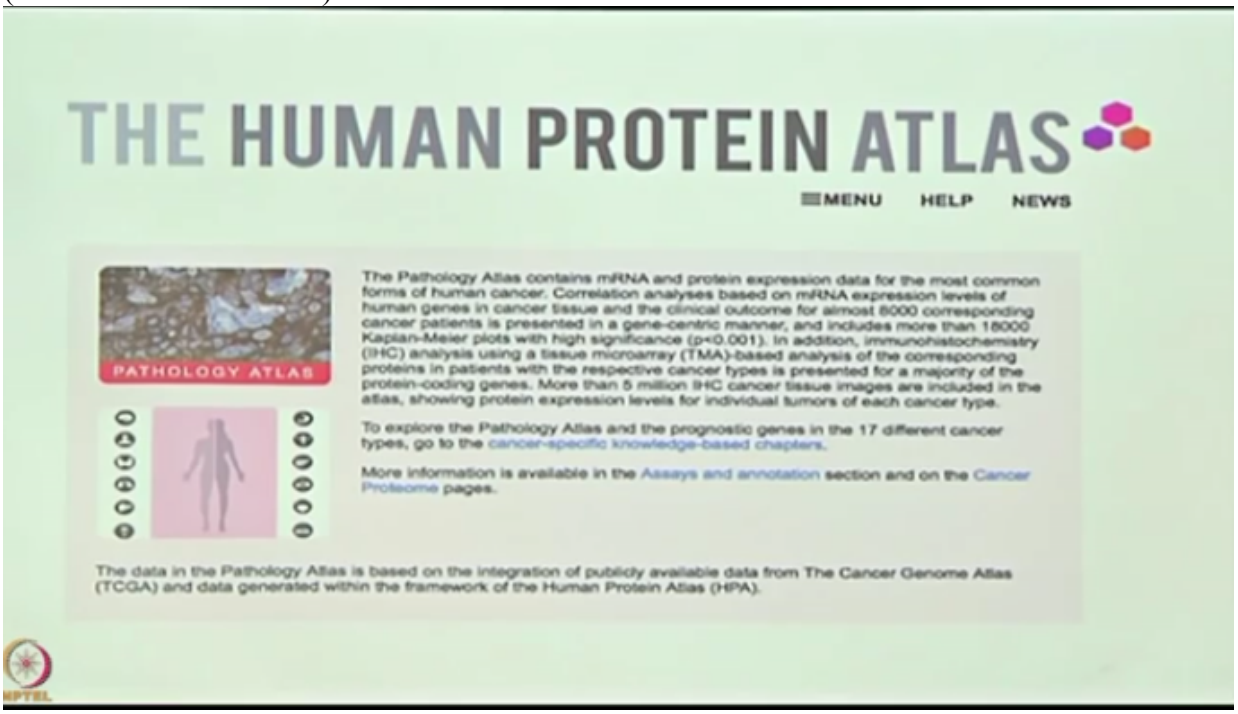


YEAR	NUMBER OF ANTIBODIES ANNOTATED	NUMBER OF IMAGES EVALUATED FOR PRIMARY ANNOTATIONS (IN MILLIONS)	NUMBER OF IMAGES EVALUATED FOR PRIMARY ANNOTATIONS AND CURATIONS (IN MILLIONS)
2007	2721	1.57	1.88
2008	3457	1.99	2.40
2009	3572	2.05	2.46
2010	3971	2.28	2.74
2011	3572	2.05	2.46
2012	2775	1.60	1.92
2013	2109	1.21	1.45
2014	1050	0.60	0.72
TOTAL	23227	13.35	16.03
AVERAGE PER YEAR	2903	1.67	2.00

And those are the statistics that we followed from the year 2007 to 2014 that gives you an average of 2 million a year, it gives you 2 million because they were original images that were annotated and then came the big question what about quality control, this is from India, and we had to work very hard to prove that there was really no difference whether they were annotated

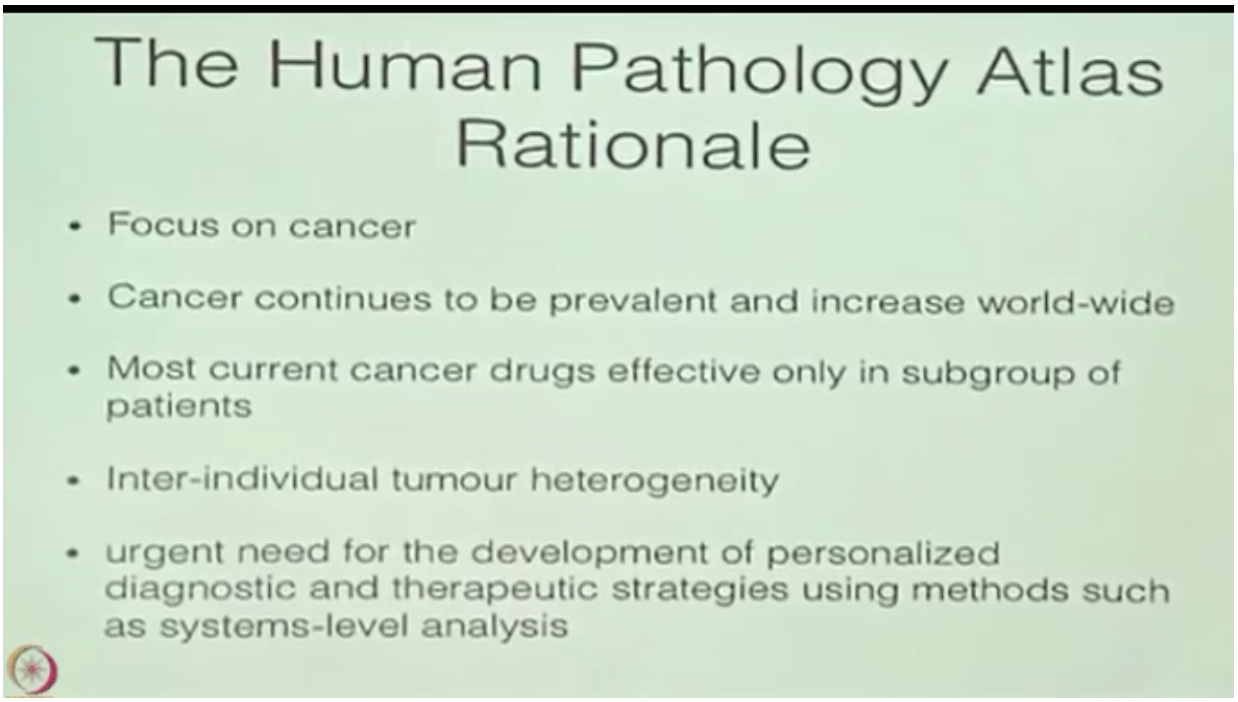
in India or in Sweden or in Norway or anyway else and everybody agreed, the statisticians became our best friends because they prove that.

(Refer Slide Time: 37:15)



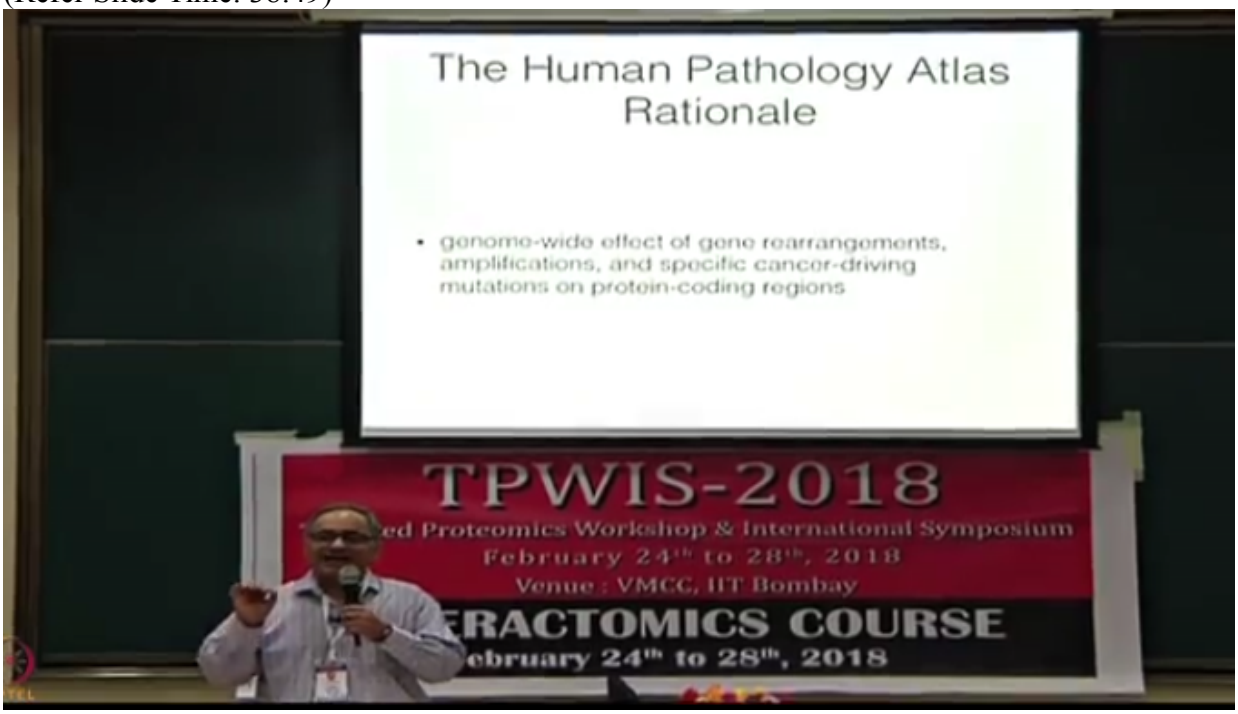
Now let me go on to the second part which is a more recent event in the protein atlas, what I just spoke about was normal tissues, but we also have a pathology atlas section that deals with pathology, mainly cancer, and that was published in August 2017, the big proteome of the pathology actress according to that transcriptome.

The focus for the human pathology atlas was on cancer, for reasons that most of you are already aware about,
(Refer Slide Time: 37:55)



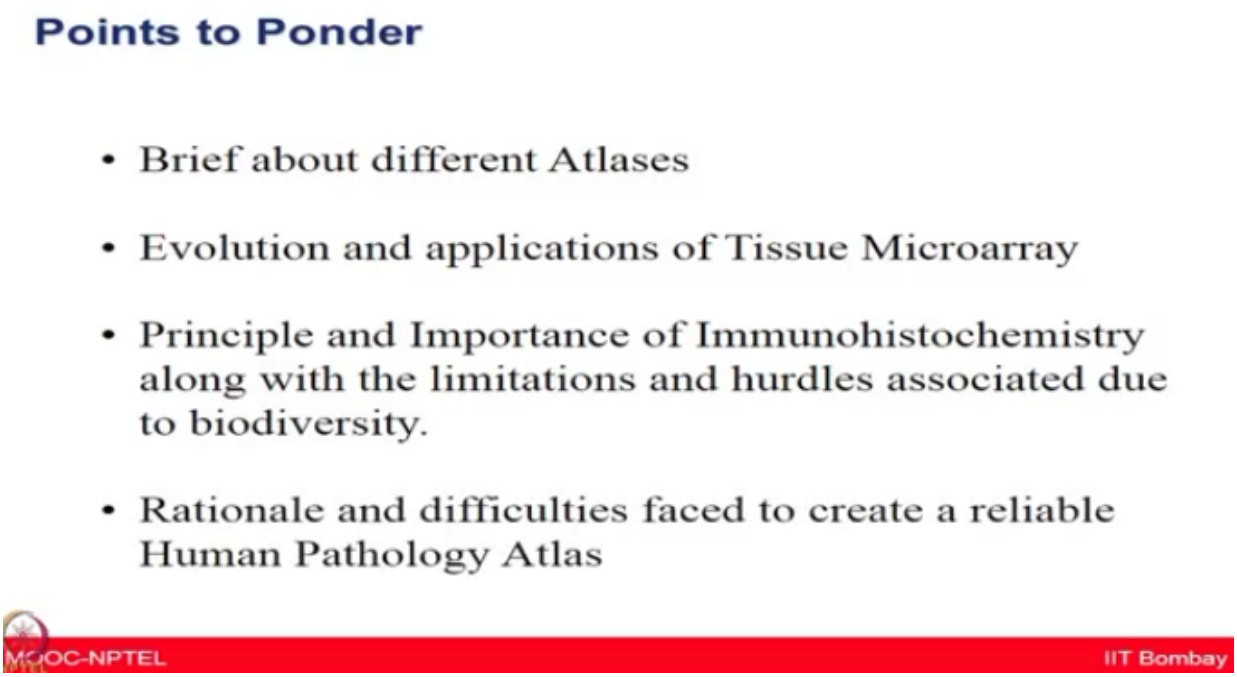
the big challenge was that most drugs in cancer are effective only in a small subset of patients, and therefore was there any difference between 2 tumors from 2 different people, looking the same under the microscope, remember the tree without branches and remember the tree with leaves, and therefore if there is so much difference, should the treatment protocol get personalized according to what that patient has, or should they just be hit with everything, so what was studied in this paper was a huge effort, it was a genome wide effect of gene rearrangements,

(Refer Slide Time: 38:49)



amplifications and specific cancer driving mutations in all the cancers that we could see.

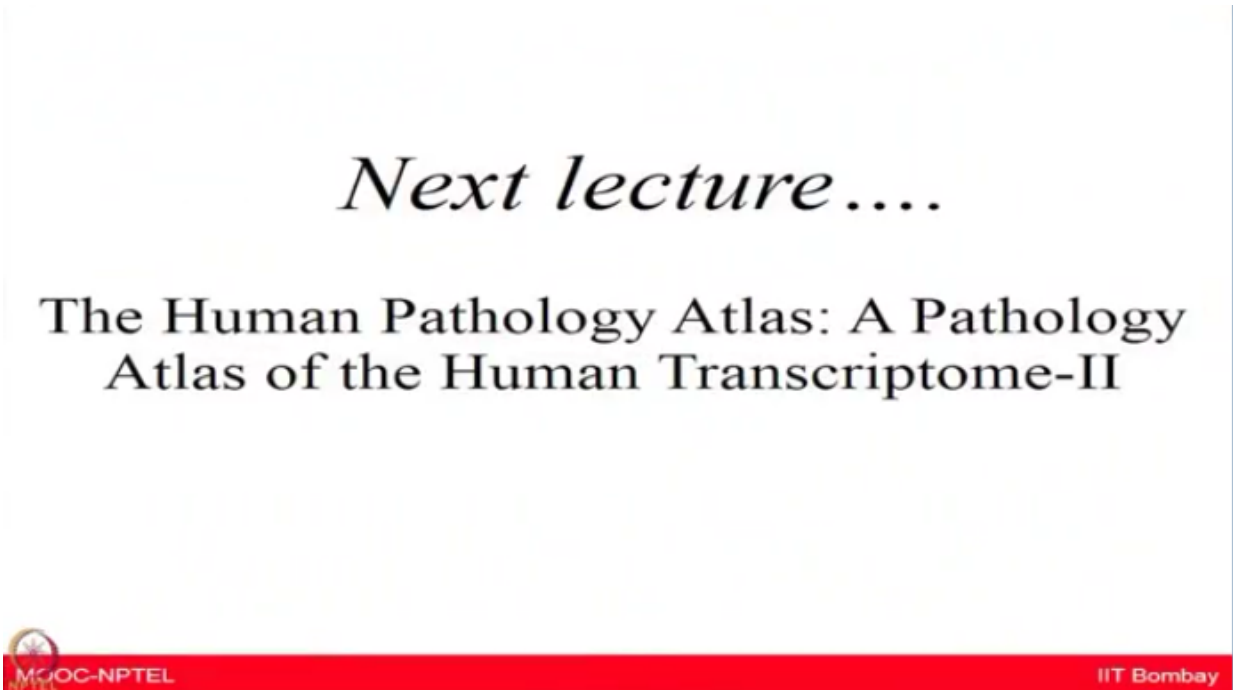
(Refer Slide Time: 39:00)



Sanjeeva Srivastava: In today's lecture you got the understanding of a big project, human protein atlas and also the major contribution by Dr. Navani's group on human pathology atlas.

The pathology atlas contains MRNA and protein expression data from 17 different forms of human cancer, these are the findings obtained from this project have really contributed immensely to the whole biologist and the entire field. Let's continue this lecture and more discussion about the human protein atlas and human pathology atlas in the next lecture. Thank you.

(Refer Slide Time: 40:00)



Prof. Sridhar Iyer

NPTEL Principal Investigator
&
Head CDEEP, IIT Bombay

Tushar R. Deshpande
Sr. Project Technical Assistant

Amin B. Shaikh
Sr. Project Technical Assistant

Vijay A. Kedare
Project Technical Assistant

Ravi. D Paswan
Project Attendant

Teaching Assistants

Apoorva Venkatesh

Nikita Gahoi

Shalini Aggarwal

Bharati Sakpal
Project Manager

Bharati Sarang
Project Research Associate

Nisha Thakur
Sr. Project Technical Assistant

Vinayak Raut
Project Assistant