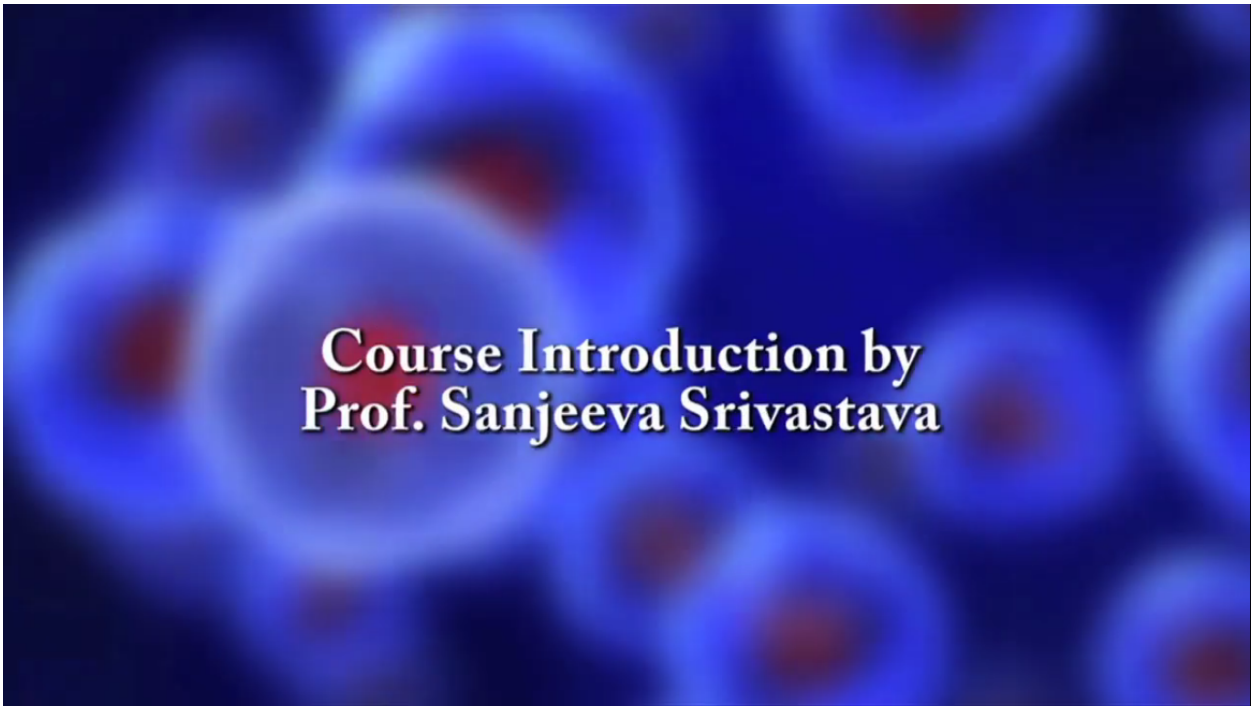
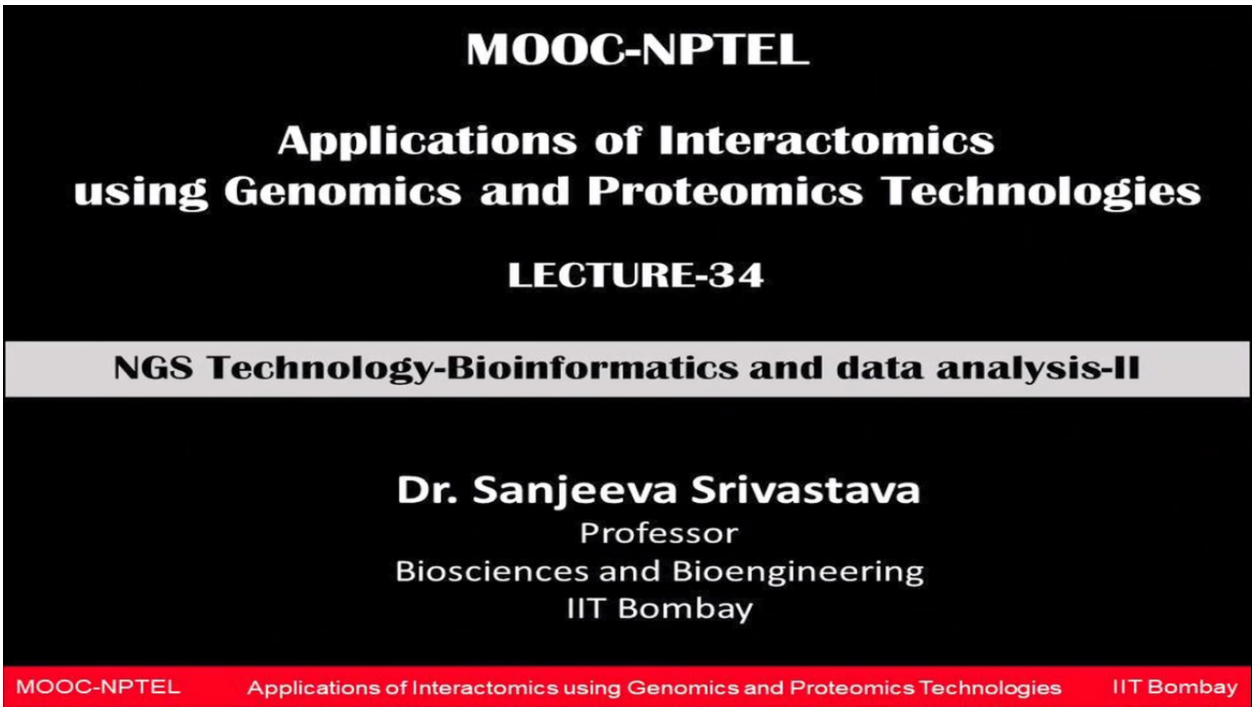
Knowledge is supreme.

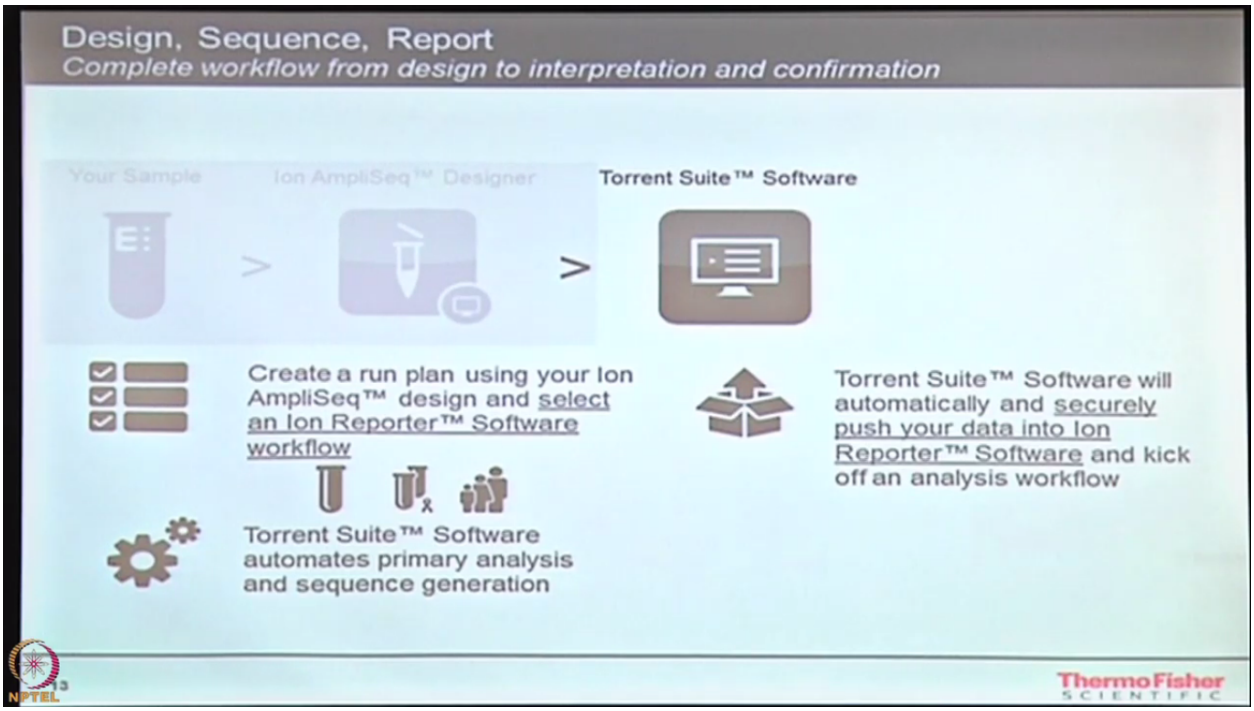

Applications of Interactomics using
Genomics and Proteomics technologies


Course Introduction by
Prof. Sanjeeva Srivastava



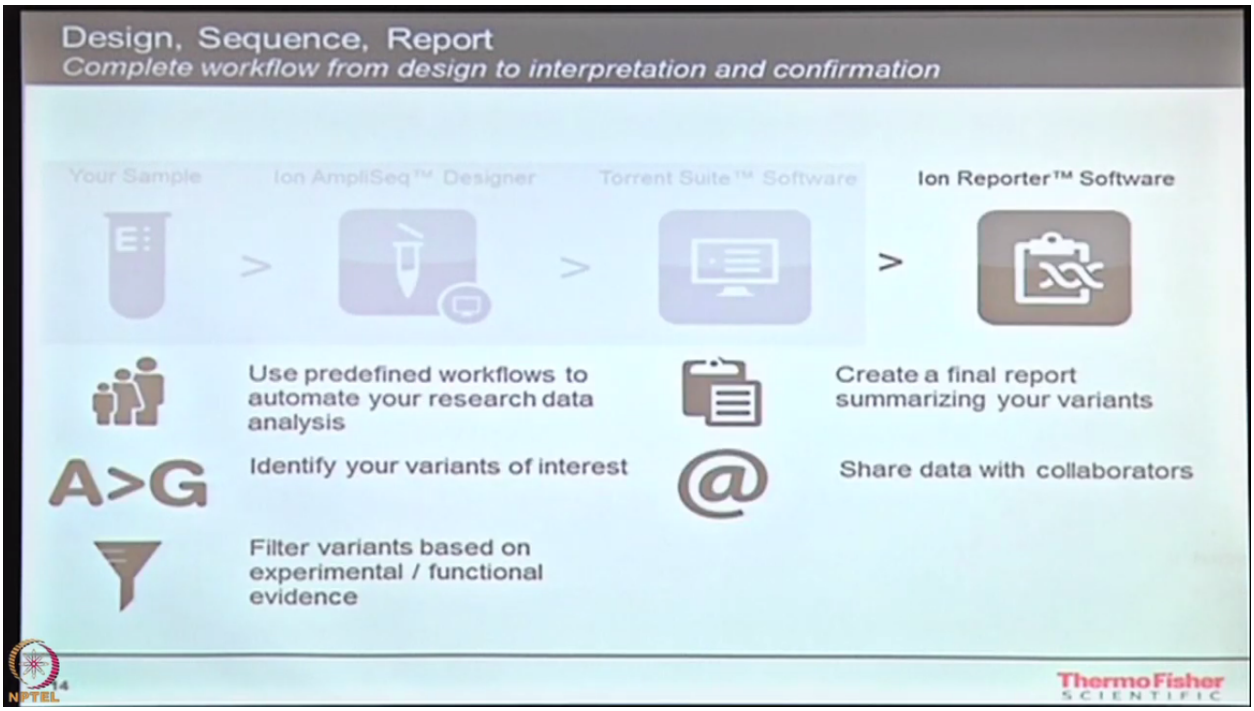
Sanjeeva Srivastava: Welcome to MOOC course on application of interactomics using genomics and proteomics technologies. In the last lecture we started discussing about the revolution of next generation sequencing technologies. An intention is to provide you the details of current leading technology platforms. In this slide we discussed about Ion Torrent technology, and today Mr. Pravin Nilawe will talk to you about specific software called Oncomine knowledge based reporter use for the oncology studies. Many times when you are able to generate large data set or big data set from the – these kind of big technology platform like NGS or master [00:01:18] it's really crucial to brought in your views and think about how to now use that data to think about at the system level information. For example you have obtained now genomic information, can you try to integrate that from the other database available from the other community that have databases, can we try to integrate the DNA at the – with the RNA level as well as protein level? So now that you have information from genomics transcript genomics and proteomics you can try to integrate that and obtain some of the systems level information. And some of these software which are available even commercially or even open access softwares they do provide you ability to integrate information from variety of data sets available from different databases. So I hope today's lecture is going to be useful for you to get understanding about one of the software and resource available for you to do a deep beta analysis.

So let's continue with this lecture.



Pravin Nilawe: So I will take a step by step mode. I first did designing on Ampliseq then I did the Torrent Suit software where it decodes all the basis and provide you the sequences for it. It provides you alignment with the reference genome. It does the coverage analysis for you which looks for the regions which have been designed, interest, gene of interest. And then take it further and do varient calling.

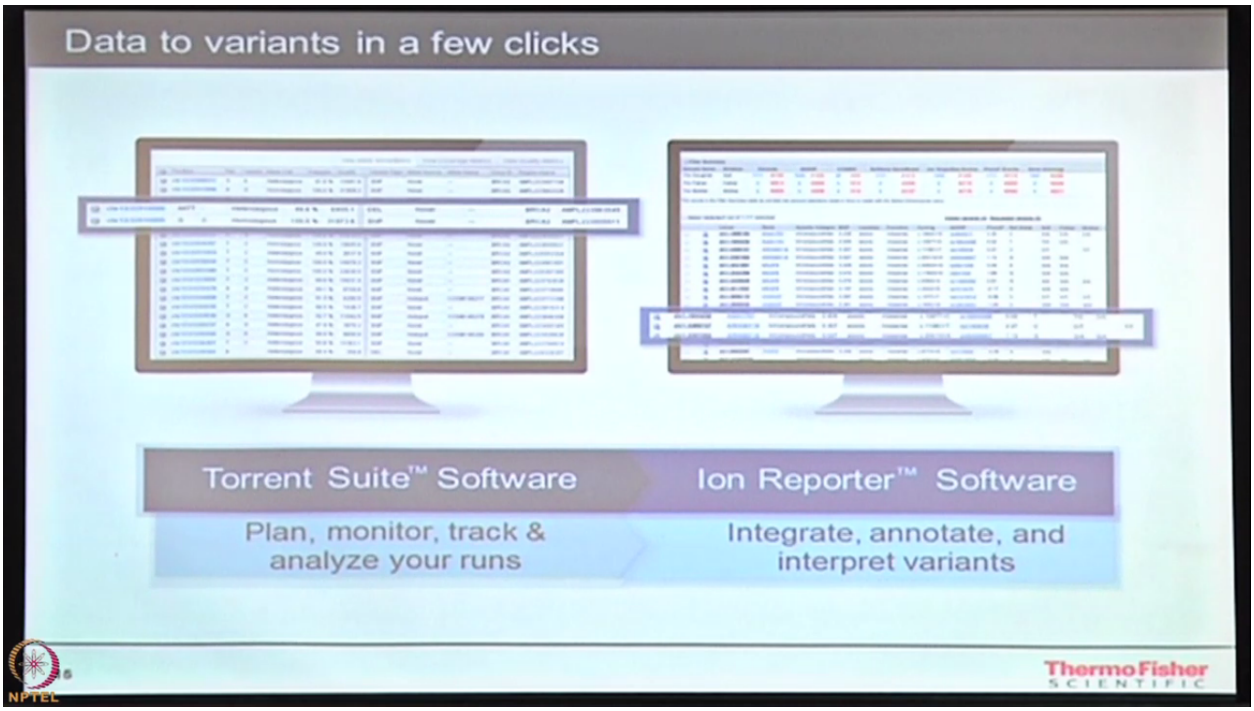
So once you have done with the varient calling you just have varience with you. You have just the SMPs giving you the change from A to T, C to T or just the deletions, A is deleted or T is deleted. But you still need to know something more about it where exactly this is happening, which gene it is happening, whether it is actually having deleterious effects or not. Whether it is having any effects with the patients or not. So you need to know something more about that. so for doing that there is one more tool called as Ion Reporter software.



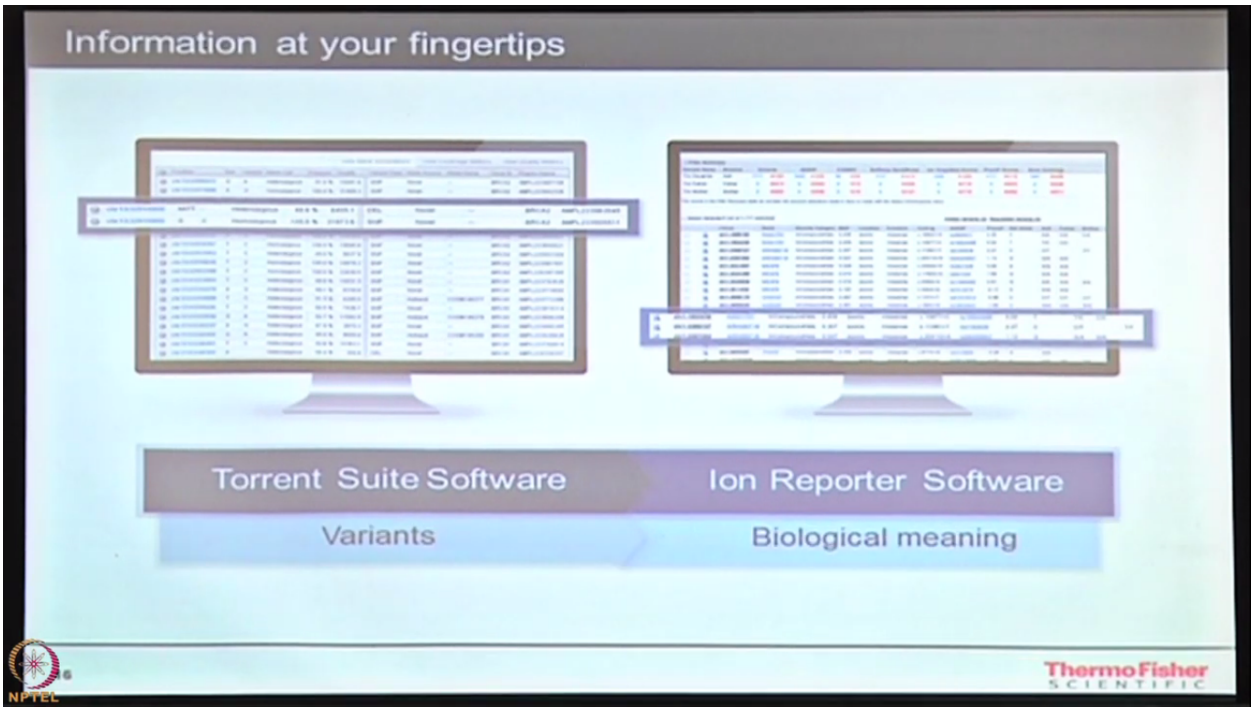
So it has lots of information in built into it. So this helps you to coorelate your varient with the information that is already stored into databases.

Ion Reporter software is available as cloud based. So you can register yourself and sign in and utilize it for your analysis. During variant calling we have a format called as VCF which is generated over here. So whatever variants that you call up these would be provided into VCF format. And these VCF formats could be easily uploaded on Ion Reporter software. Now this Ion Reporter software will help you to understand the variant positions, take it further and do certain stastistical calculations. Have you heard about the protein evolutionary study parameters which are called shifts, polyfem, granthum. So these are certain parameters which are actually calculated when there is a protein change happening into a particular like you have a variation coming in and protein change happening into it. So these values or this parameters also defines whether a particular change is deleterious to that protein or not or deleterious for this region or not.

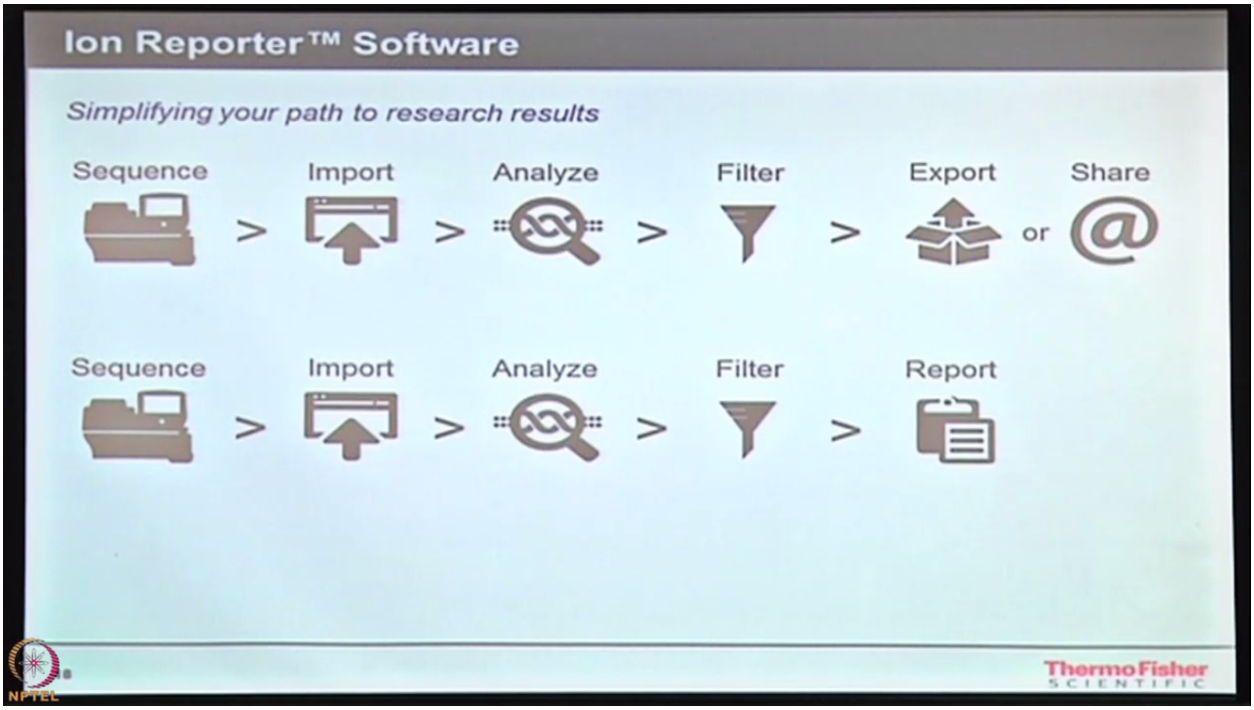
So these parameters helps you to also filter out whether these variants would be really deleterious or not. So the software Ion Reporter helps you to calculate those protein evolution over here. With that it assigns databases from DBSMP, Cosmic, OMIM, DGB where it corelates with all the information that is related to your genetic disorders, cancer and various other genes that are oriented to it. Genes transcripts that are related to it. So you must have heard about NCBI. NCBI is a databases where you have lots of information stored whether it is gene, OMIM, DBSMP. Just every information from there is correlated over here in Ion Reporter.



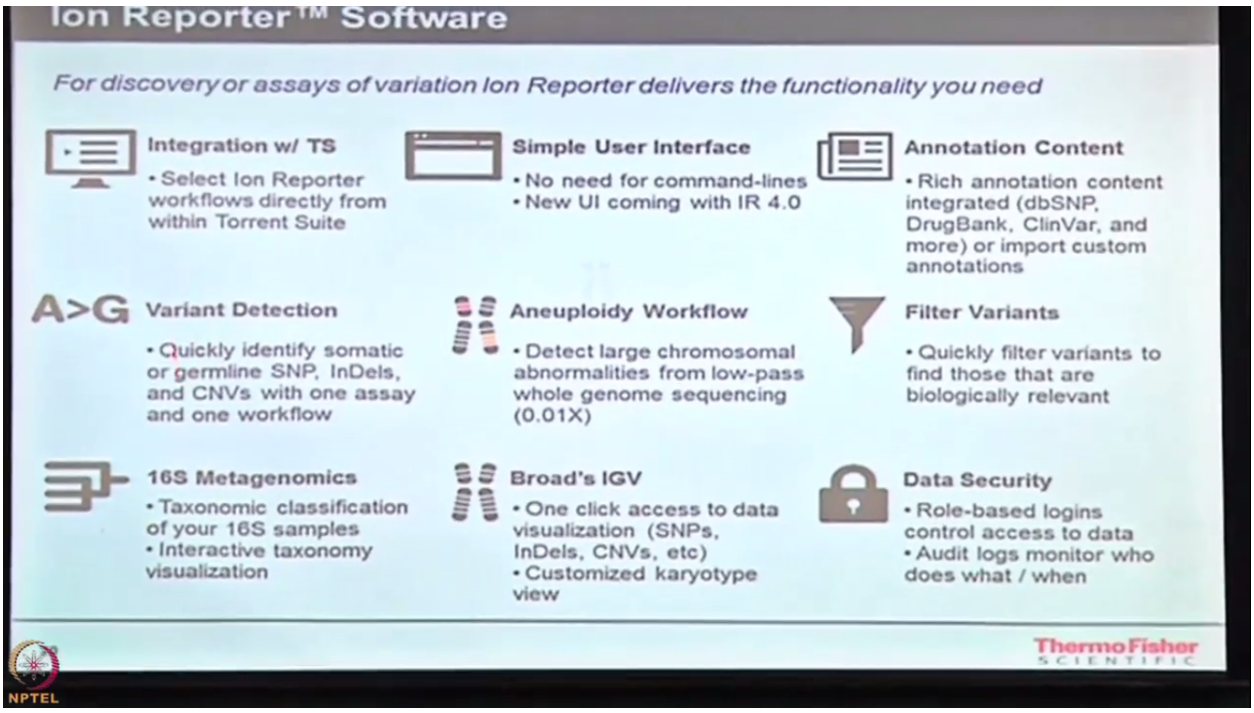
So what happens over here. So you have Torrent Suit software which helps you to plan your run, run your data, generate your data, generate your variants and understand what is the coverage across those regions. You are getting results over there as variants, mutations. You take that further and put it into Ion Reporter. Ion Reporter will help you in integrating annotating and interpreting my results. So you have more information to read about where you will understand your variants are really important or not, whether you want to keep it, whether you want to utlize it or you want to filter them out.



So this tools comes in a lots of help over here. Okay, the same thing that I was talking about so you have got variants from Torrent Suit software, from Torrent Variant color and then you are bringing it to a biological meaning where you can understand what the variant does into your samples.



With the same I will just take it forward with Ion Reporter software which helps you to understand what can be done over here. So I have a way for going forward with that. You sequence your data. You take your data, you can import in onto the Ion Reporter software, analyze them for whatever you want and take it further filtering based on your parameters available. So what happens filters maybe of different types. It maybe your coverage filters, you want to take variants which are at very high coverage. Say you want 1000 read should be covering a particular region, then only you want to take those as a particular variant called variants. You have filters for shift polygrame grantham, the protein evolution parameters that play an important role over here. You can also look into databases like you have [00:07:44] databases which has information which are the common SMPs which are not effecting at all in any conditions to any of the proteins. So you want to filter them out and you have to take them out and keep it separately. So whatever remains back you take the databases, you get the databases information and can be downloaded or even shared with different people across.



So once that is there, once you have done that, the same thing that I was talking about, importing your data onto the system, analyzing them by various filters that are available and taking it further, reporting it further. So what happens over here is I have an Ion Reporter software I will be detecting different things. Not just this Ion Reporter software is not just for your variant analysis actually. There are lots of other tools that are also available in Ion Reporter. We have something for the variant detection that is one. But we also have something called a 16S Metagenomics with us. This is something a bit different from what you do with the genomic level. This is something with the microbial content or microbial identifications that you do. So if you have bacterial samples, or if you have certain samples that are coming for 16S sequencing we have a kit which helps you to do that 16S sequencing onto the Ion Torrent system and then you can do a straightforward analysis in Ion Reporter over here. So differentiates your microbial

content. As well as you have something call as Aneuploidy. So you like to look into large deletions or like to look for the Aneuploidies across your embryos. So you like to see whether your embryos you must have seen – heard about IVF. So IVF they normally go forward with certain embryo screening where they like to know whether they have deletions or certain duplications happening into your data or not. At that moment this particular workflow would be very much helpful for you.

So not touching much about this rather than the other part of us so you have done the variant analysis. After that the same thing that I have shown you as IGV visualizing your data. So what I visualize the region, gene region, or that reads data mapping at the same time same thing can be done over here onto this software itself. So you can visualize the data over here. So if you see you are looking at particular variant you can click at the variant and visualize it into IGV in particular, Integrated Genomic Viewer. At the same time databases I have talked about so it's DBSMP, drug bank, cleanware, OMIM, so lots of things are available over here. So if you can play with them anytime. Then the filtering is happening and you have a data security. Security as in sense you have your own logins that are given to you or you can register your own logins which are encrypted. So there is nobody who is going to take away your data over there.

So with this this is about the Ion Reporter part, so any questions still now? Okay. yeah.

Student: Do you have any tool that can help in discovery of novel genes or isoforms?

Pravin Nilawe: Novel gene, gene discovery as in sense in RNA or as in –

Student: [00:11:18]

Pravin Nilawe: So you already have certain free tools available. So what happens over here in sequencing you can do the sequencing. You can generate your data for RNAC and then take it further for different tools that are available like you must be knowing about if you have an idea about it you have idea about cufflings. So that's one of the tools which is actually free which is a command line based tool and which helps you to define which are the new isoforms that could be generated for your RNAC data. So it helps you to know which are the regions which are getting aligned by your sequences and then defines whether it is an actually exone or not which is actually fitting into your particular data or not or particular gene or not. And then comes out into a particular transcripts for it and gives the different isoforms. So cufflings is a tool that helps you to do that. we have a workflow which fits into this but it doesn't has this particular workflow. It has to be done separately. We have RNAC plug in available which does a human based alignment of the data and gives you the human base gene code based annotations are utlized and you get the gene information into it as well as gene counts, and transcript counts for it.

Then you can take it further into differential expression. But when you are looking about particularly for those isoforms we are looking more into the command line software that are available, freely available or the other commercial tools that are available. You have to do it separately.

Student: [00:12:54] the database is continously updating. So does it get updated automatically or we have to go to the internet and do it on our own?

Pravin Nilawe: So for the software that is online it is already updated. So whenever there is an update coming in six or – three to six months that gets updated over there onto the cloud system that we have. So when I was talking about Ion Reporter it already had certain databases which have been available over there.

Student: So do we have to work on the cloud for that?

Pravin Nilawe: No, nothing like that. so what happens the data gets generated onto the S5 system or the instrument. Once the data is generated you try to do the analysis basic analysis at the system level itself. So we have a server with this which helps you to do the basic analysis like aligning the data to the reference genome. Then looking at the coverage analysis and then coming to a variant level where you get all the variants into it. So you have got the variants through the system itself. Variants would be all the mutation that you are looking for. Now what happens this mutation or variants that you are getting we need to annotate them or we need to know more about them. So we are taking this and then we have a tool which is actually cloud based which is actually on website. So we take this data particularly, we upload it onto this cloud and we can try to do the analysis.

Student: So it's like a BLAST?

Pravin Nilawe: It's not like a BLAST. But there is something similar to it. So what happens over there we have a tool called a steam map which does the alignment of the data to the – or mapping of the reads to the genome over here. So when you are mapping that particular to genome it is already done onto this system itself or the server that we have over here, with this system itself. So it's locally done over there. Once you align it whatever you get is in a alignment format. And then we try to call the variants across it or the variations into it, onto the server. So whatever variations you pick up you can take it further and then put it further into the cloud system. So that cloud will help you to correlate the positions that you have got and the database information that we have for those position and can be correlated together for all the databases.

Student: So the database get updated automatically.

Pravin Nilawe: So there is an update happening on cloud every few months.

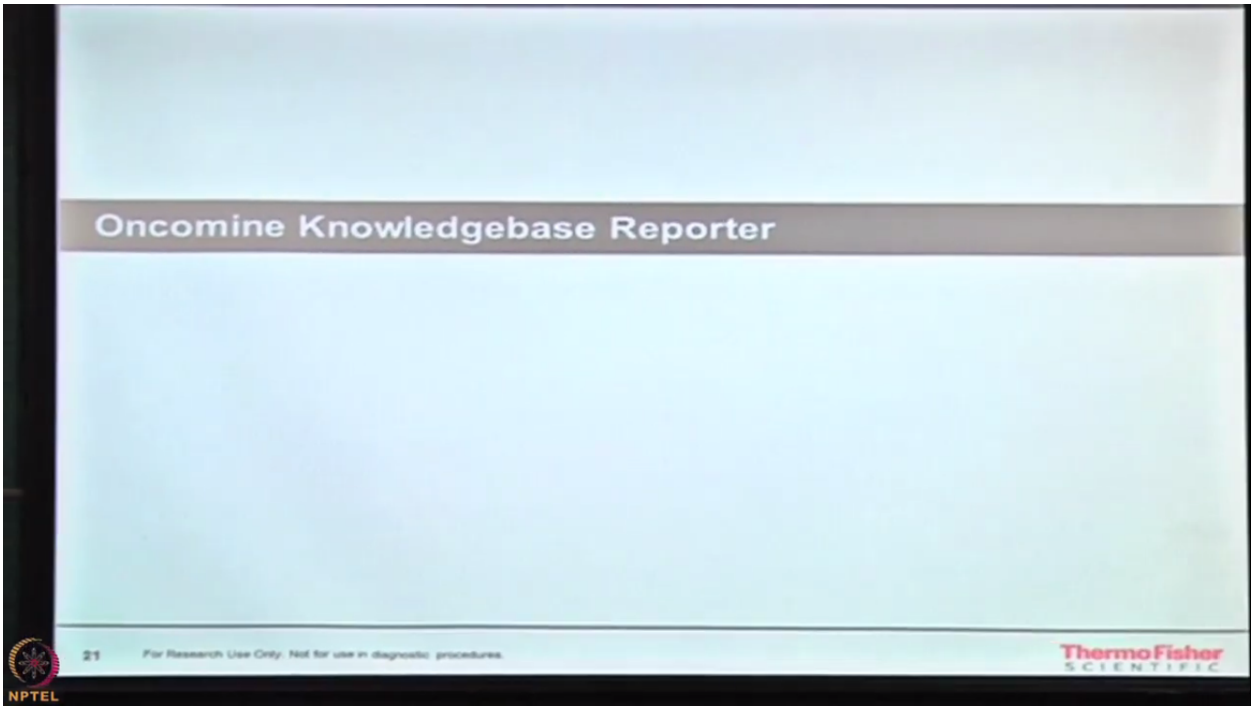
Student: Do we need to get the license every year?

Pravin Nilawe: There is nothing like license over there actually. What happens you are given around 100 GB of space onto the cloud. So if you are taking the same system at your – as a local server there is a two version out of it. One is Ion Reporter cloud version which is online, which gives 100 GB free space. Once your space is full you need to buy some more space. It's like a cloud system that you have. Otherwise it is like you have a local server also where if you already have the system you can push the data to that local server something like a Linux server and in that you can do the analysis further and do the same analysis that you are doing onto the cloud. Just that you need to update the server every three or six months. That's it. So you have the database in both of them is available just that you need to upgrade them and utilize them with the data that you have.

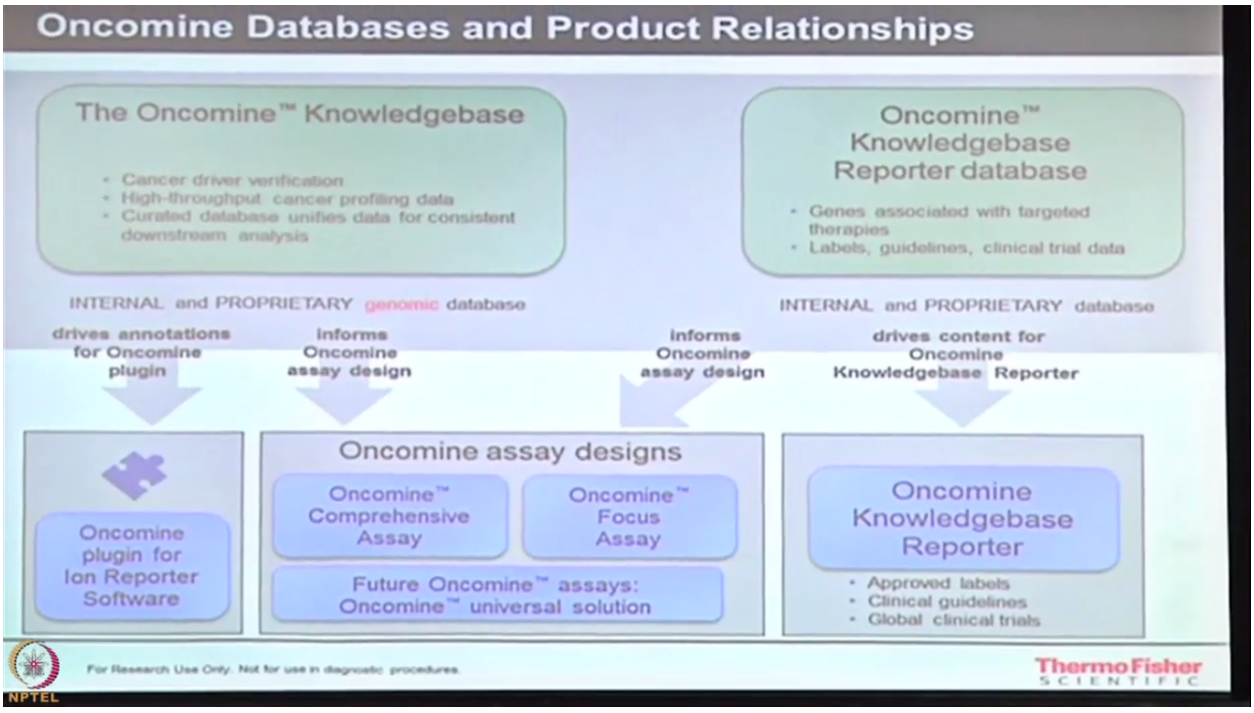
Student: But sometime the software get corrupted.

Pravin Nilawe: Right. Over here we have everything onto the system itself. And what we have done is it is like an appliance. It's like your toaster and kind of thing where you are just putting your data or doing sequencing and generating your data. You can't touch that particular system anywhere. You have to just access it through a web browser only. So nobody can access it externally unless there is some service person coming in and doing something out of it. So we have some data coming out. Take it and do the analysis on cloud.

So at this stage if you have like you have got the idea about the variants now; what are the variants that are coming in but even after doing so much of study what happens normally nowadays we request something where we – the results would be very quick, quick as in sense even if you have got the variants we need to get to know what is its effects and what is its counter-effects happening across. And then what could be the drugs which could be available for such type of diseases or variants that you are getting.

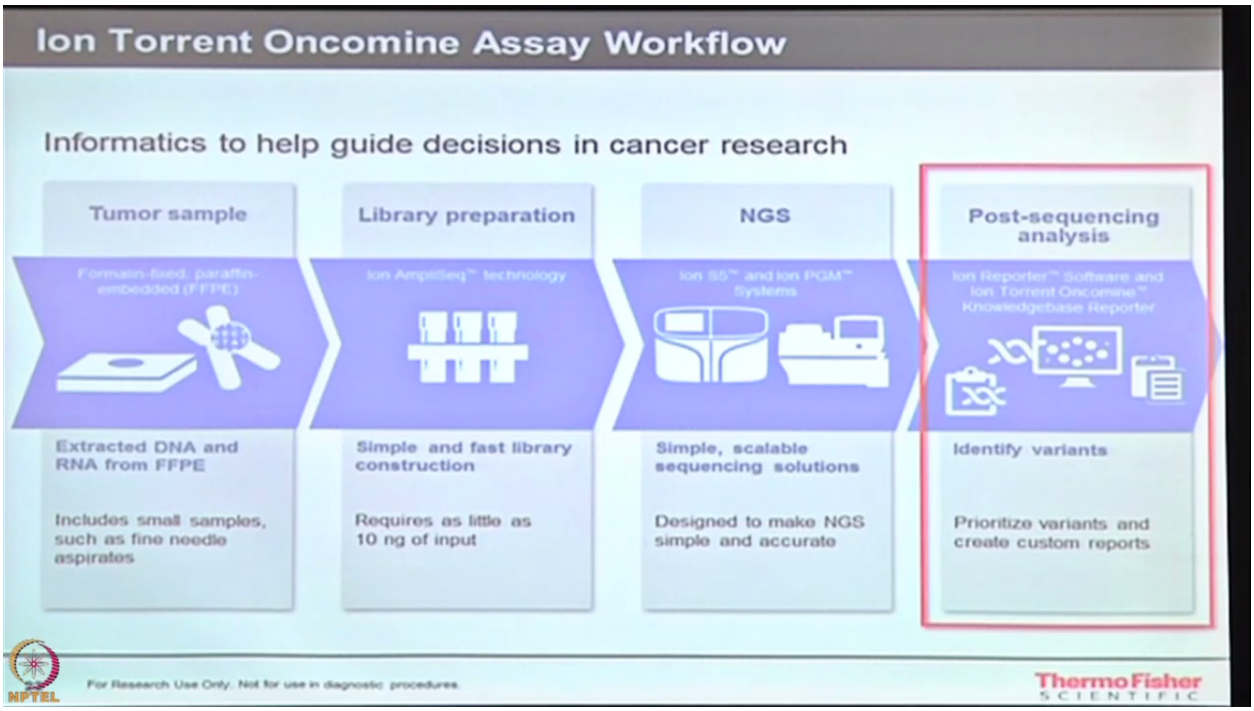


What you are doing over here is you are somewhere correlating your patients samples information to the variants, getting meaningful information and then you are trying to generate the information what types of drugs could be utilized. You have a drug bank database that is correlated to this. But doing this would take some time. Doing all the correlations would take some time and getting you the results out of it. At the same time we have developed something called as Oncomine knowledge base reporter. Oncomine knowledge base reporter has a database called as the Oncomine knowledge base. This Oncomine knowledge base again if you Google it you will get an idea it has lots of information regarding to cancer.

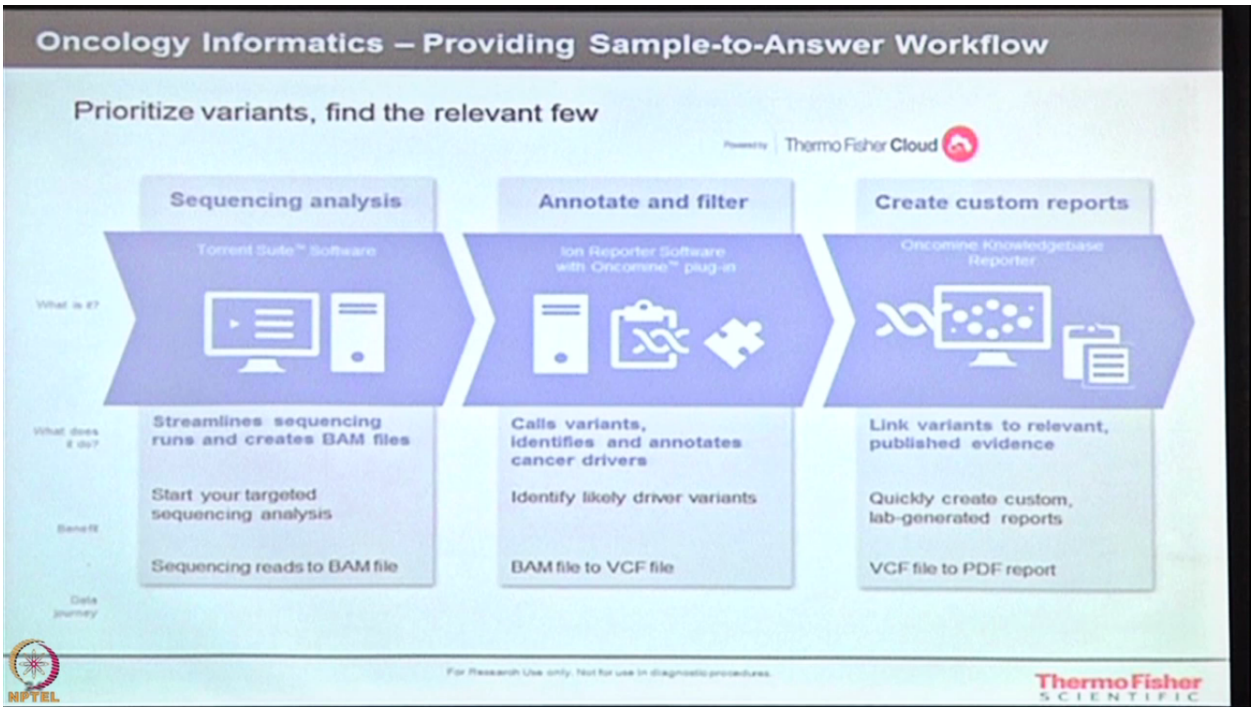


This is specifically for the caner related studies or else this is something like which will help you to correlate your cancer. Cancer is variants to the drugs that are available in the market or into the clinical trials. So we have certain databases. You have certain cancer driving drivers that are available which are verified and studied by the researchers over here. So same database we have generated over here called an Oncomine knowledge base reporter database. So this has all the curated studied information about genes, the targeted therapies, or the drugs that are utilized for treating cancer. And with that you have an interface called as Oncomine knowledge base reporter interface.

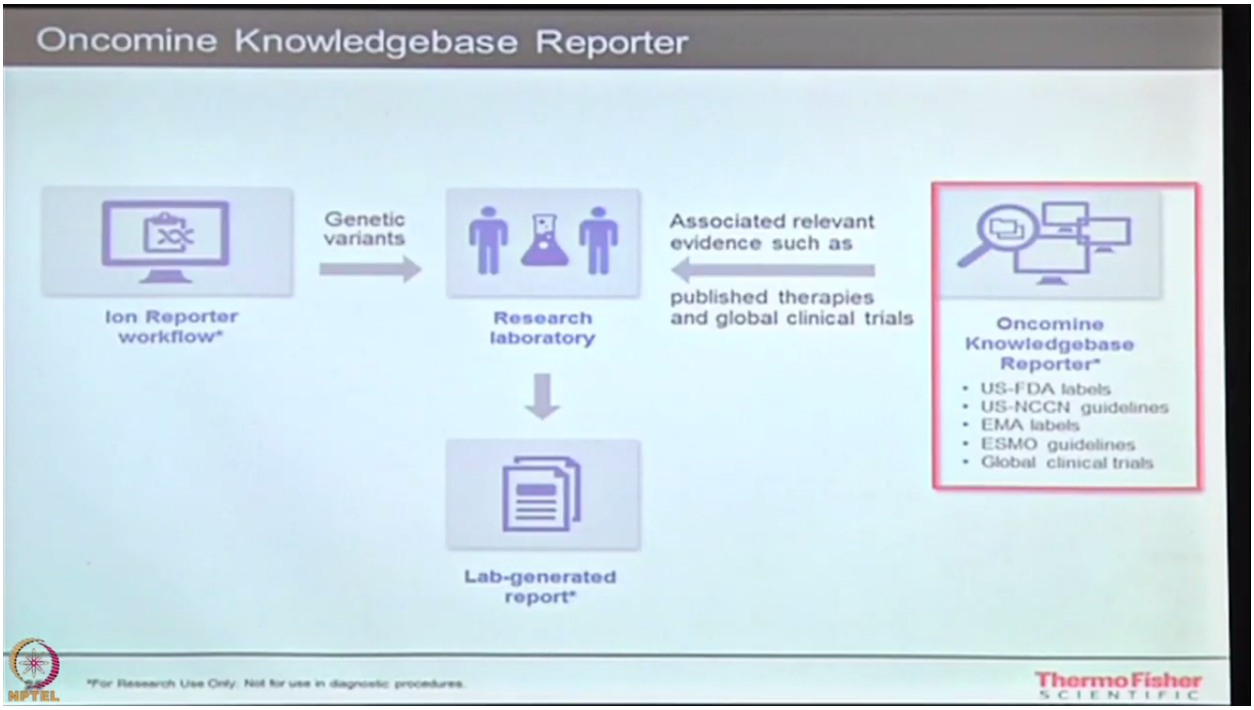
So now what happens over here it has lots of information of your variants called as cancer or driver information. This driver information is correlated with different drug information. And the same thing is available as my design assays. So what I was talking earlier into Ampliseq that I would like to design my gene assays or gene region target of interest and then sequencing to know which are the diagnostics variants that you are getting. The same thing is coming over here. You have certain assays which are available, which can be utilized to detect which are the variants that are getting coming in into your samples and the same samples are utilized to denote whether that variant that has been detected is having any clinical significance or not and with the clinical significance whether it is correlated to any drugs or not. So this is something that you have as a summary you have something called a tumor sample.



You take it further, prepare your library, do your sequencing over here and then you do certain analysis. The certain analysis would be like you do the Ion Reporter analysis, understand the variant information, get the results from there. This Ion Reporter also has information regarding the driver variant. So this driver variants are recorded over here in Ion Reporter as well as these could be pushed up into the tool called as Oncomine knowledge reporter. So once you push up this data to the Oncomine knowledge reporter it will try to look for these driver information, correlated with the different drugs available and then look into the clinical trials that have happened already over here.

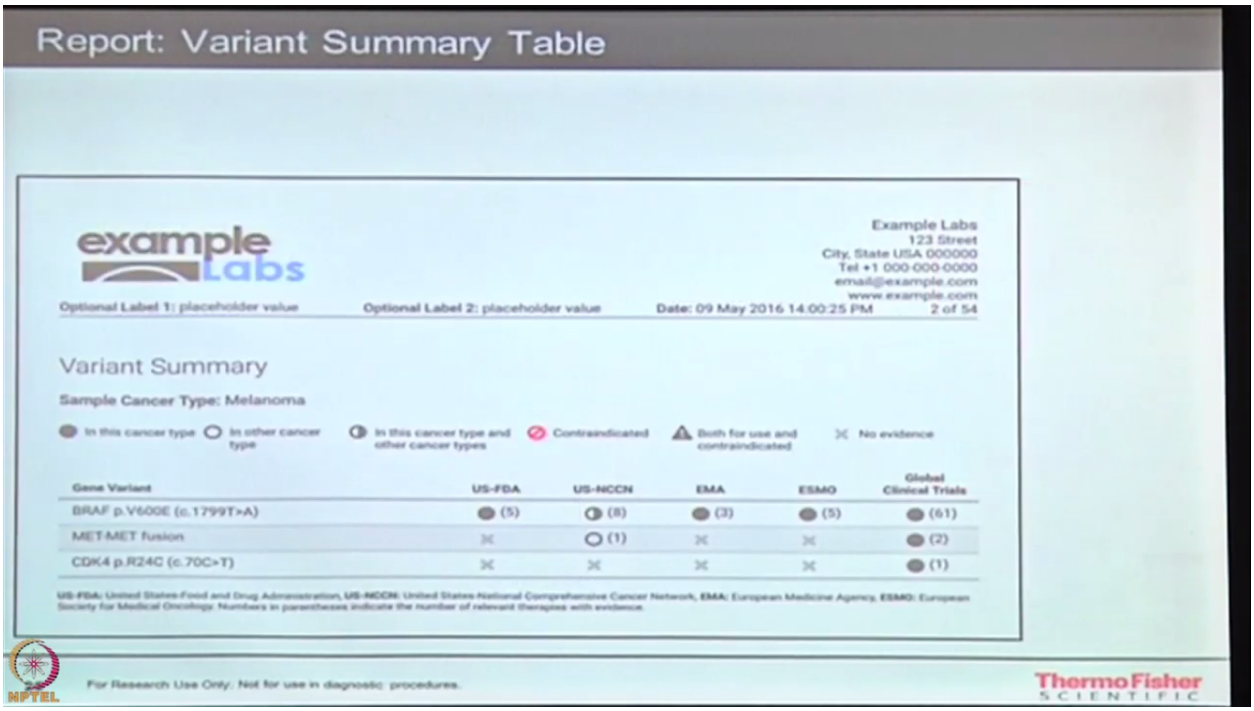


So at this point if I take further so you have something like this.

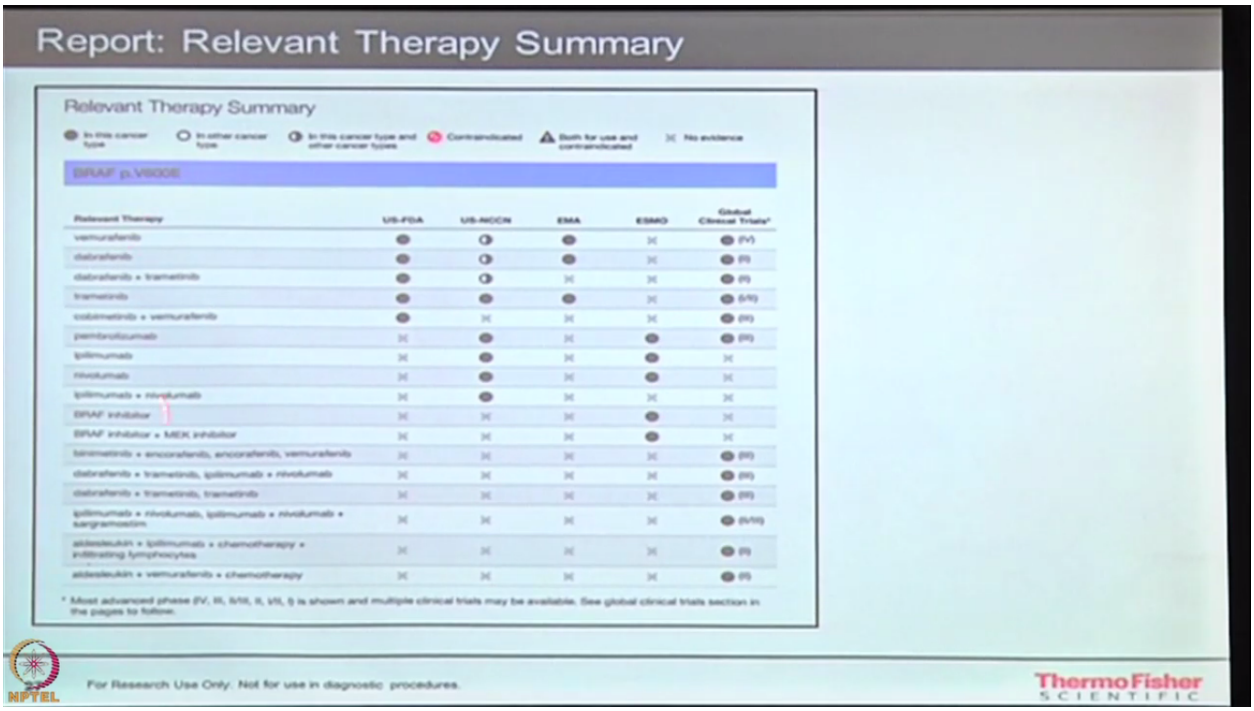


You have Ion Reporter coming into play, you have genetic variants. You have a knowledge base reporter which has different guidelines information regarding to your drugs, information

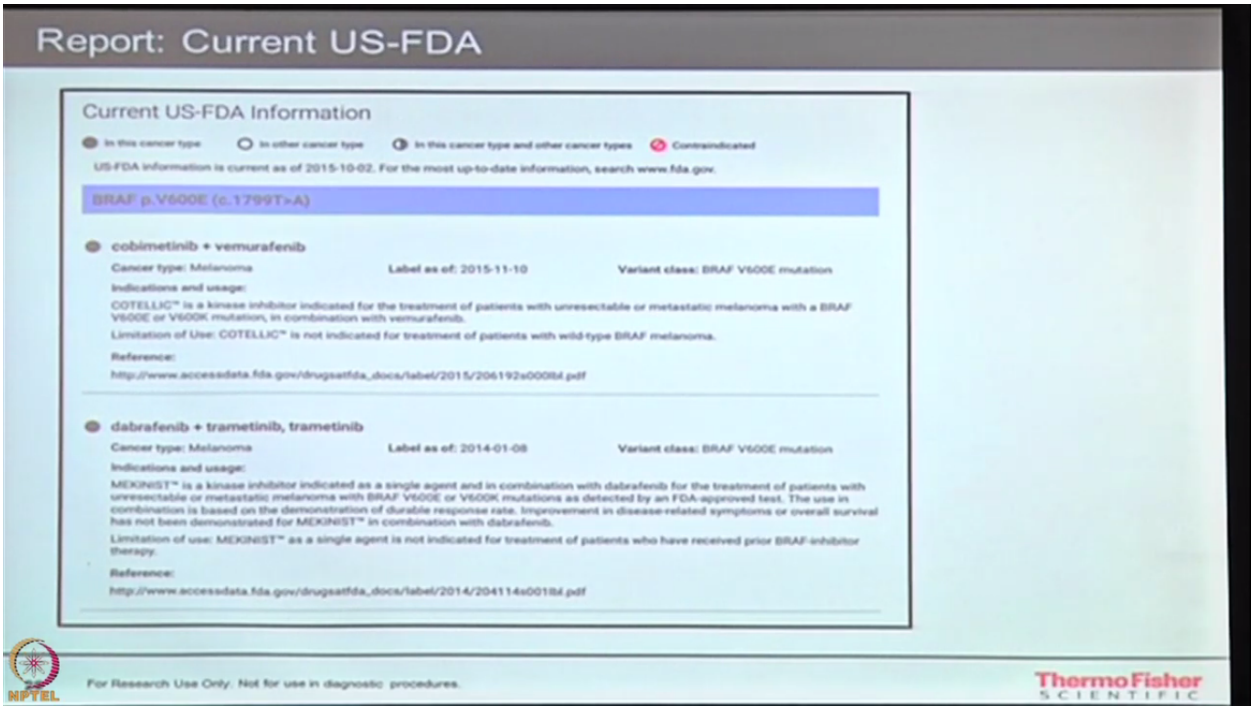
regarding to clinical trials, everything over here. You correlate them and finally generate a report for it. And this would be very helpful when you are trying to give a drug to a particular patient. It would be very quick for you. So patient comes, you do the sequencing one or two days, look for the drug that is actually having any effect, give the report to a doctor, he could let you know what could be the drugs utilized further for a particular treatment.



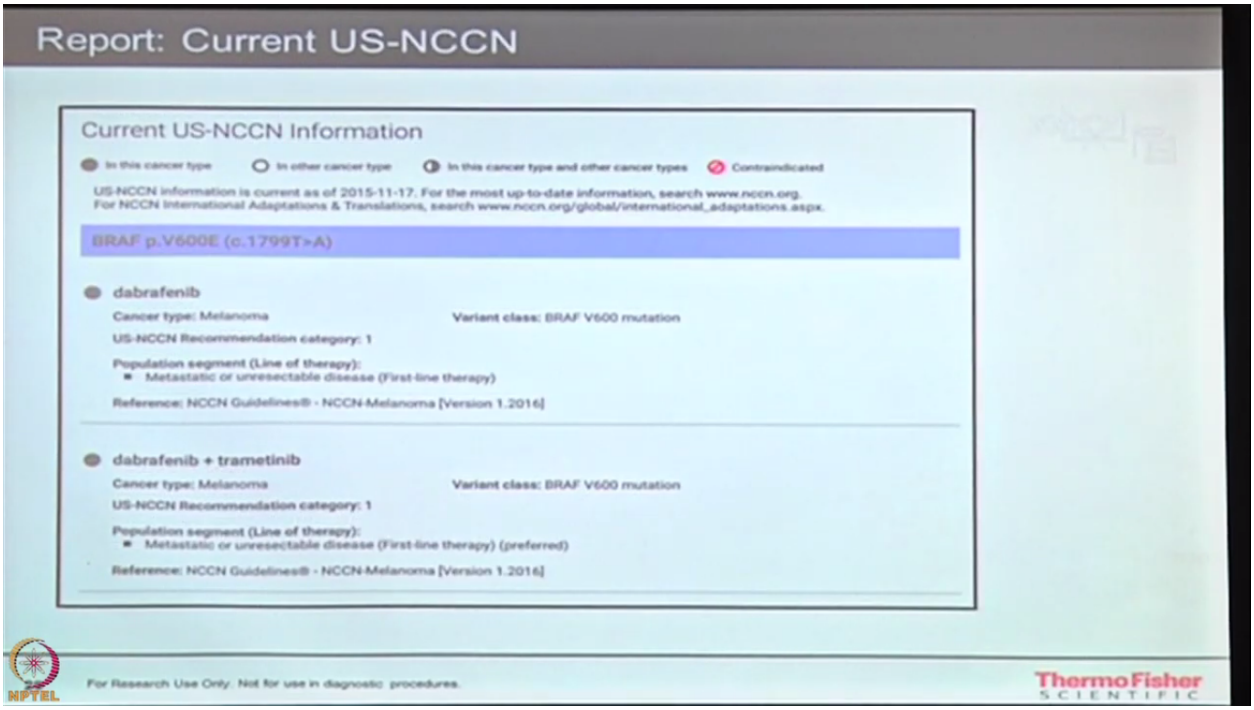
This is how we can come to a place which would be very quick in responding to cancers. Different types of cancers. So an example report for us where you have the variants you can see here are the variants that we are looking at. There is a fusion. There is a C2T change that is available. And there is T2H change happening over here also. There is a protein change happening. V600E that's a protein change happening over here amino acid. And at the same time there are different clinical trials and drugs affiliations available. So if the drug has been studied some place and has been given a go for the utilization so everything is given over here. So this is something like a summary which gives you idea whether it is going into a clinical trial or not, whether it is U.S. FDA approved or not.



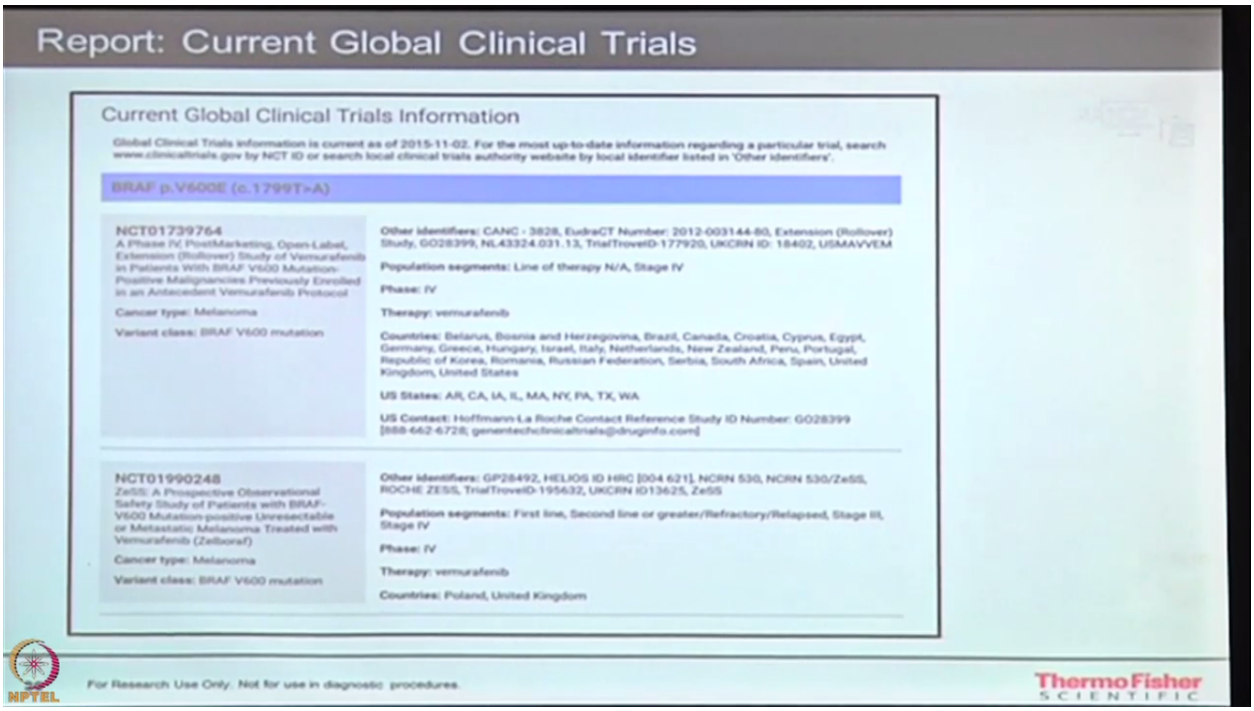
At the same time you can get more information about more drugs into it by different variant you are looking at. V600E so these are the various therapies that are available drug therapies. So who have given which approval, which stage is approved, global clinical variants, in which stage it has been approved. These details are coming into this particular report particularly.



At the same time what I would like to know is in detail, in depth if I am looking for a particular drug I am concentrating on that what studied has been done across those particular drug. So that studies are coming over here. So what type drug effects are happening over here. So those information comes in complete details for each and every therapy that you are looking at.



And this finally comes to a level where you can finalize it and give a report to the doctor. So based on my different information or different approvals that we have got, this report comes in. So once this report is here, you can utilized the same report, goes to the patient and you can directly assign drugs over there particularly or there could be a discussion between a clinician and a person under doctor who wants to give that drug and take it further.



Student: Are these reports accepted by doctors?

Pravin Nilawe: So in general as in sense somebody who is using the Ion Reporter, sorry who is using the Oncomine Reporter are making their own formats of reporting this particular variants. This particular data. So they are doing what is if they are targeting a particular say they are targeting a particular cancer as such, they have studied the cancer properly. They have already done certain diagnosis study. They come to a point where the have all these drug information, they just take out which is the drugs that are really required and provide the details to the clinicians or to the doctor as such.

Student: [00:25:15]

Pravin Nilawe: So that's what, the person who are working on this are giving into a standard format. They are –

Student: [00:25:34]

Pravin Nilawe: Right.

Student: [00:25:39]

Pravin Nilawe: So this has been integrated. This has been done. These are the approvals that they have got through the U.S. FDA, the drug whatever information we are giving is based on the approvals that are there available like U.S. FDA or the European I don't get that name, yeah ESMO. So they have utilized the same information over here to summarize and give it to it. So how to represent it it is all about the person who is going to use the NGS system. They plan it how to give it as a format or as a like what you say approval. So they try to work on that and then give it.

Student: [00:26:31]

Pravin Nilawe: No idea about it. We don't have an idea about it because whatever we are going okay –

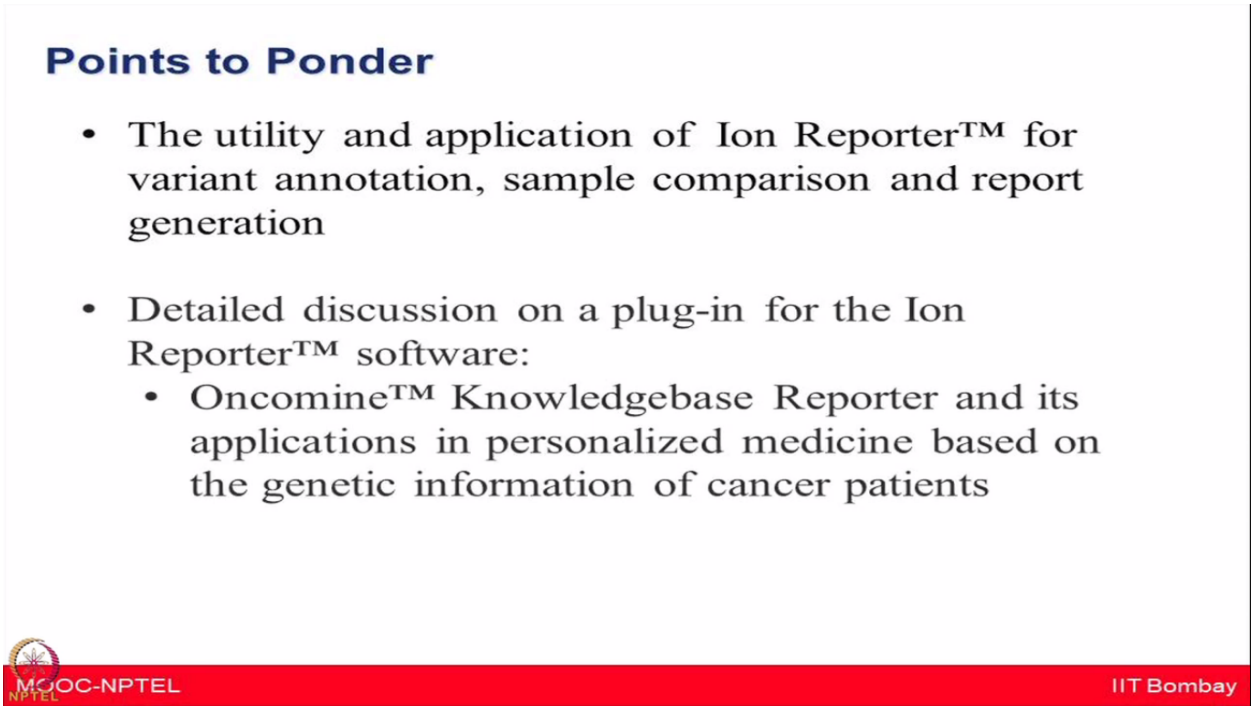
Student: [00:26:51]

Pravin Nilawe: Actually that's why we have brought this particular tool like this particular software which helps you to get the information in such a way that it could be acceptable.

Student: [00:27:04]

Pravin Nilawe: Yeah. So. Okay, yeah. So any questions about this? I hope so you have got an idea or a feel how the data analysis happens in NGS. The only thing over here was to take you through the entire workflow, how we start with the analysis, how we would like to look into the particular gene designs, come to a level where you do these sequencing for the gene designs and then get it annotated through Ion Reporter. It's a tool that is freely available and then looking for something like a quick response there is Oncomine knowledge base reporter which helps you to do that. okay.

So thank you very much for your time.



Sanjeeva Srivastava: So today's lecture you got a good understanding about one of the useful resource which is Oncomine knowledge based reporter from where you can get lot of information for the cancer cells. And again while we are really getting biased towards cancer or clinical applications but there are similar kind of resources available for your model organism of interest as well. So it is important for you to dig deeper and know your available resources which many times now made publicly available and anybody can use those resources and if you have access to them then a wealth of information could be integrated from variety of data sources. So these database also contains various critical information about cancer. It's variant forms, different type of drug and therapies impact and it can be really usefulf for you to now start adding information on on top of a data from the curation and see that whether you can build now some hypothesis which could be executional hypothesis now to take forward for your experiments. So in this lecture you you have learned more about how to do variant data analysis obtained from NGS platform which could be correlated with the patient data to study cancer. We will continue more about NGS platform. I am sure it's one of the revolutionary technology and exciting a reas to discuss and I will also try to bring more application orientation ad application scientist to speaker to you about let the developments happening in the NGS based platform.

Thank you.