Interactomics: Basics and Applications

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

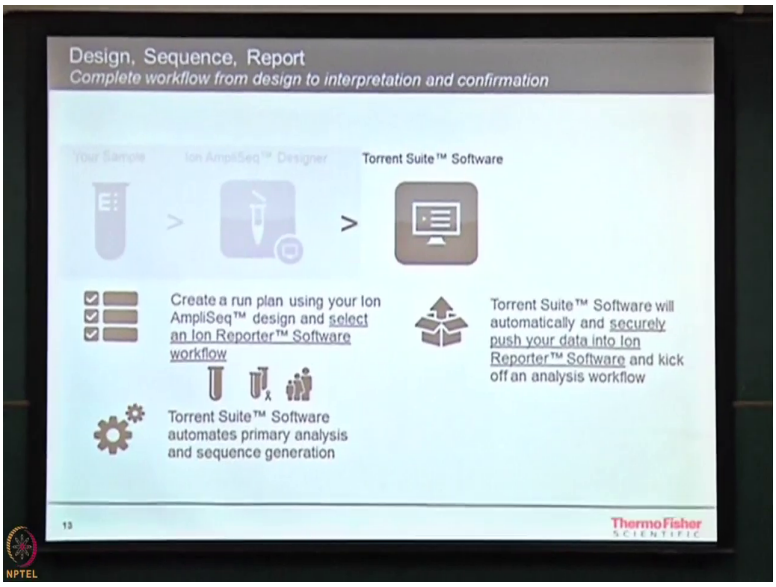
Lecture - 49
NGS Technology - Bioinformatics and data analysis-II

In the last lectures we started discussing about the revolution of next generation sequencing technologies and 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 a specific software called Oncomine knowledgebased reporter used for the oncology studies. Many times when you are able to generate large data sets or big dataset from the you know these kind of big technology platforms like NGS or mass spectrometry.

It is really crucial to bought in your views and think about how to now use that data to think about at a systems level information. For example, you have obtained now genomic information can you try to integrate that from the other data sets available you know from the other community whether it databases can we try to integrate the you know DNA at the with the RNA level as well as protein level. So, now when you have the information from genomics transcriptomics and proteomics you can try to integrate that and obtain some of the systems level information.

And some of these softwares 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 todays lecture is going to be a useful for you to get understanding about one of the software and resource available for you to do a deep beta analysis. So, let us continue with his lecture.

(Refer Slide Time: 02:17)

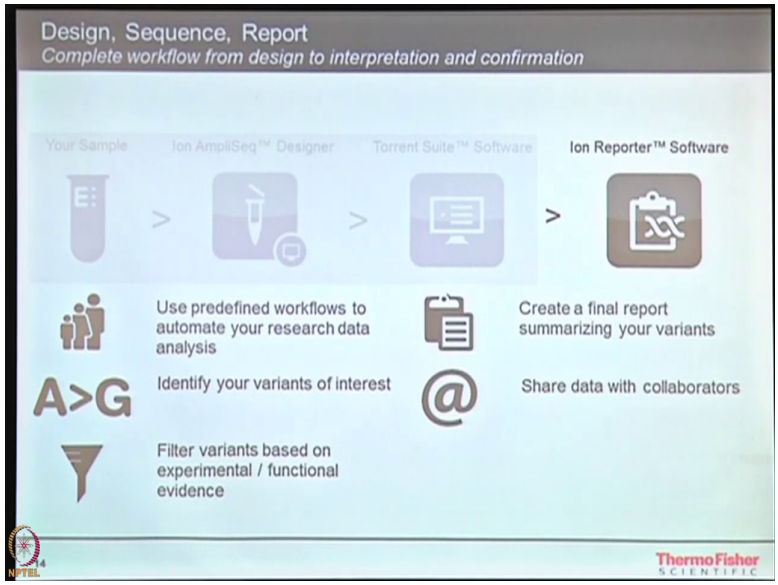


So, I will take a step by step mode; I first did designing on ampliseq alright, then I did the torrent suite software where decodes all the bases and provides you the sequences for it, it provides you alignment with the reference genome. It does the coverage analysis for you which looks for the regions which have been designed your interest gene of interest and then take it further and do variant calling.

So, once we have done with the variant calling you just have variants with you right you have just the SNPS 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 right? So, you need to know

something more about that. So, for doing that there is one more tool called as ion reporter software.

(Refer Slide Time: 03:07)



So, it has a lots of information in built into it. So, this helps you to correlate your variant 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 us VCF which is generated over here ok.

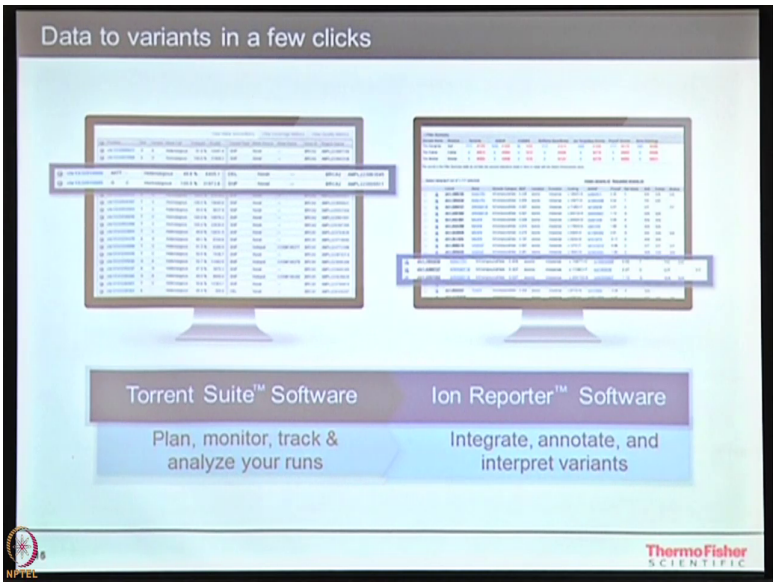
So, whatever variants that you call up these would be provided into a VCF format and these VCF formats could be easily uploaded on ion reporter software ok. Now, this ion report software will help you to understand the variant positions take it further and do certain statistical calculations, have you heard about the protein evolutionary study parameters which are called a sift s i f t, polyphen, grantham ok.

So, these are certain parameters which actually calculated when there is a protein change happening into a particular like you have a variation coming in and a protein change happening into it. So, this values or these parameters also defines whether a particular change is a deleterious to that protein or not a deleterious for this region or not ok. So, these parameters helps you to also filter out whether these variants would be really deleterious or not.

So, this software ion reporter helps you to calculate those protein evolutions over here with that it assigns databases from D B S N P cosmic O M I M D G V where it correlates with all the information that is regard related to your genetic disorders, cancer ok.

And various other genes that are oriented to it genes transcripts that are related to it. So, you have you must have heard about NCBI right? Yeah? NCBI is a database where you have lots of information stored, whether it is gene did a OMIM DBSNP. So, this every information from there is correlated over here in ion reporter. So, what happens over here.

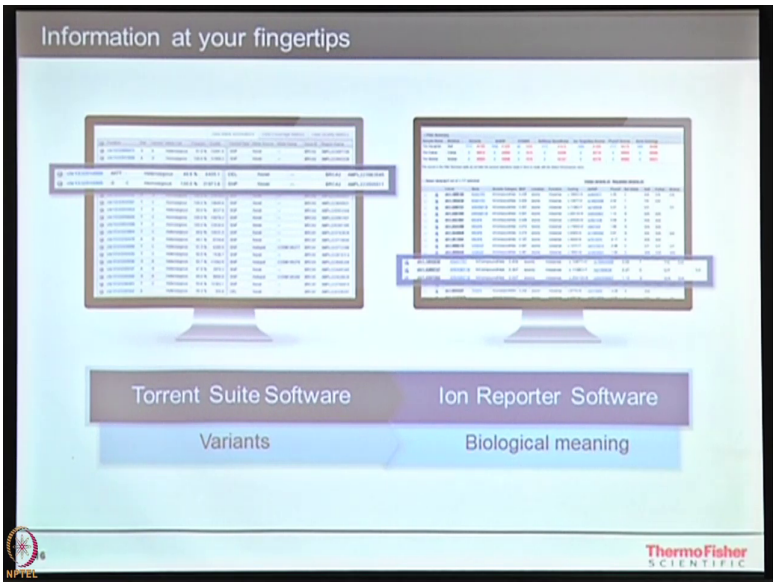
(Refer Slide Time: 05:36)



So, you have torrent suite software which helps you to plan your runs run your data generate your data generate your variants and understand what is the coverage across those regions right. You getting results over there as variance, mutations; you take that further and put it into ion reporter. My ion reporter will help you in integrating annotating and interpreting my results right.

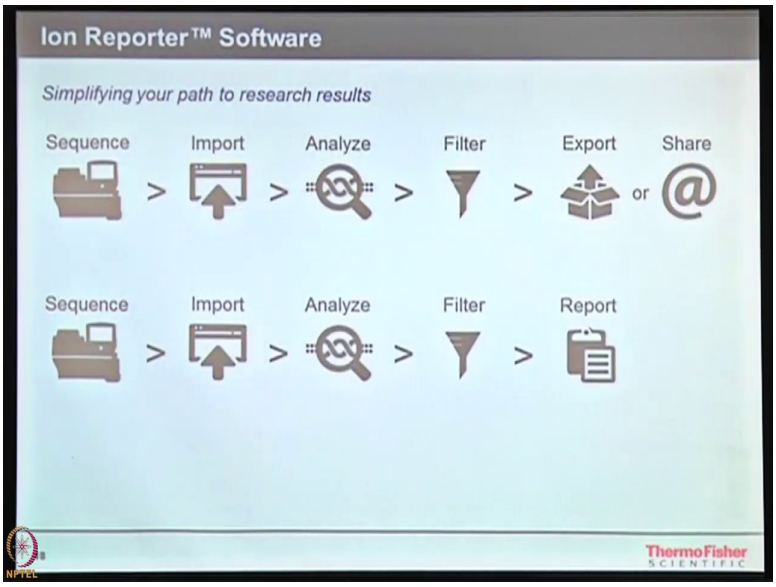
So, you have more information to read about where you will understand your variants are really important or not; whether you want to keep it, whether you want to utilize it or you want to filter them out ok. So, this tool comes in a lots of help over here ok.

(Refer Slide Time: 06:19)



The same thing that I was talking about. So, you have got variants from torrent suite software, from torrent variant column and then you are bringing it to a biological meaning where you can understand what the variant does into you are samples.

(Refer Slide Time: 06:31)



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 it onto the ion reporter software, analyze them for whatever you want and take it for the filtering based on your parameters available. So, what happens filters maybe of different types; it may be your coverage filters you want to take variants which are at a very high coverage see you want 1000 rids should be covering a particular region then only you want to take those as a particular variant called variants right.

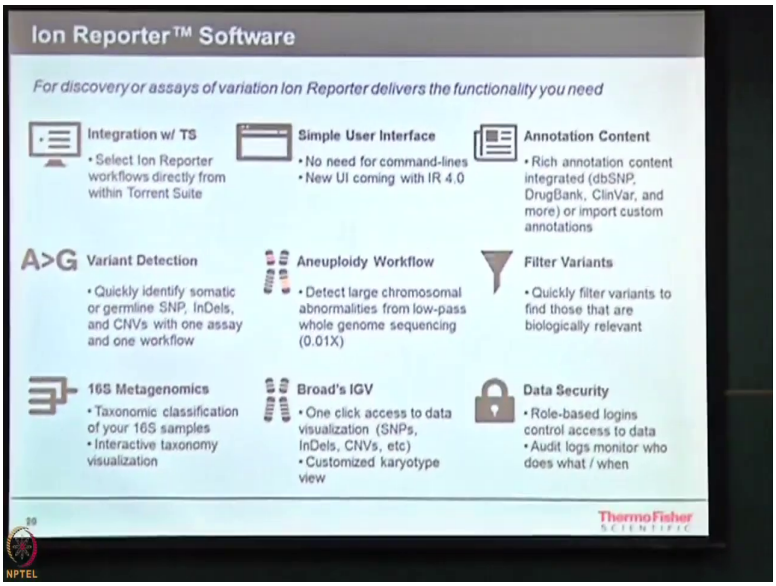
You have filters for sift polyphen grantham the protein evolution parameters that play a important role over here. You can also look into databases like you have ucsc common snp

databases which has information which are the common snps which are not affecting 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 database information and can be downloaded or even shared with different people across ok. 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 ok.

So, what happens over here is I have in 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.

(Refer Slide Time: 08:21)



We have something for the variant detection that is one ok, but we also have something called as 16S Metagenomics with us this is something a way 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 data coming for 16 S sequencing we have a kit which helps you to do that 16 S sequencing onto the ion torrent system and then you can do a straight forward analysis in ion report over here. So, differentiates your microbial content as well as we have something called as aneuploidy.

So, you let to look into large deletions or like you would like to look for the aneuploidies across your embryos right. So, you like to see whether your embryos you must have seen about heard about IVF right. So, IVF they normally go forward with certain embryo screening, where they like to know whether deletions or there are certain duplications happening into your data on not at that movement 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 ok. So, what are visualize the region gene region or that reads that are mapping the same time same thing can be done over here onto the software itself.

So, you can visualize the data over here. So, if you are see you are looking at particular variant you can click at the variant and visualize it into IGU particularly Integrated Genomic Viewer. At the same time databases I have talked about. So, it is dbSNP, DrugBank, ClinVar, 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 their yeah. So, with this this is about the ion reporter part. So, any questions till now? Ok yeah.

Student: Do you have anything novel gene discovery of (Refer Time: 10:51).

Novel gene discovery as in sense in RNA seq or as?

Student: (Refer Time: 10:58).

So, you already have certain free tools right available. So, what happens over here in sequencing you can do the sequencing right; you get generate your data for RNA seq and then take it for the for different tools that are available, like you must be knowing about if you have an idea about it you have idea about cufflinks. So, that is one of the tool which is actually free which is a command line based tool which helps you to define which are the new ISO forms that could be generated or the RNA seq data ok.

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, cufflinks is a tool that helps you to do that we have a workflow which fits into this, but it does not has this particular workflow it has to be done separately we have a RNA seq plug in available which does a human based alignment of the data.

And gives you the human based gene code based annotations are utilized and you get the gene information entered as well as the gene counts and transcript counts for it ok. Then you can take it further and do differential expression, but when you are looking about particularly for those isoforms we are looking more into the command line softwares that are available freely available or the other commercial tools that are available you have to do it separately yeah.

Student: Sir, one question.

Yeah.

Student: When we use say (Refer Time: 12:38) when we analysis the database.

Yeah.

Student: (Refer Time: 12:43) 16 S the database is continuously.

Updating.

Student: Updated. So, is that the preloaded or automatically goes through the internet and do. So, (Refer Time: 12:51).

So, for the software that is online; that is already updated ok. So, whenever there is an update coming in 6 or 3 to 6 months; that gets updated over there on to the cloud system that we have. So, when I was talking about ion reporter it already has certain databases which have been available over there ok.

Student: So, we have to work on the cloud for this (Refer Time: 13:13).

No, no, no not nothing like that. So, what happens the data gets generated onto the S 5 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 got get all the variants into it ok.

So, we have come got the variance through the system itself variance would be all the mutations 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, you are taking this and then we have a tool which is actually cloud based which is actually on website.

Student: So.

So, we take this data particularly we upload it onto this cloud and we can try to do the analysis.

Student: So, is it like a blast (Refer Time: 14:15) where you run it.

It is.

Student: On data (Refer Time: 14:16).

It is not like a blast it is not a like a blast, but there is something similar to it.

Student: (Refer Time: 14:22).

Yeah. So, what happens over there we have a tool called as sky map which does the alignment of the data to the or a mapping of the reads to the genome over here ok. So, when you are mapping that particular to genome it is already done onto the system itself or the server that we have over here with this system itself. So, it is 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 ok. So, the 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: (Refer Time: 15:12) time the every few month, every time.

So, there is an update happening on cloud every few months and though.

Student: So, the license goes every year?

Ah there is nothing like license over there actually what happens we are given around a hundred GB of space of space onto the cloud ok. So, if you are taking the same system at your as a local server there is a two version out of it. One is a ion reporter cloud version which is online which gives a hundred GB free space once you have spaces full you need to buy some more space it is 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.

Student: (Refer Time: 15:58).

And in that you can do the analysis further and do the same analysis that you are doing on to the cloud just that you need to update the server every 3 or 6 months that is it ok. So, you have the database in both of them is available just that you need to apply upgrade them and utilize them with the data that you have.

Student: Sir.

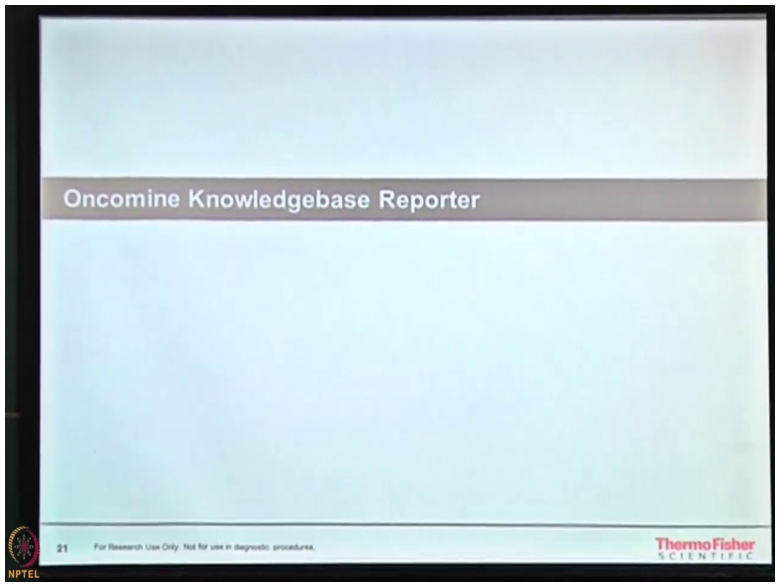
Yeah.

Student: (Refer Time: 16:17) because, but sometimes the software goes corrupted (Refer Time: 16:20).

Right. Over here we have everything want to the system itself and what we have done is it has like an appliance; it is like your toaster and kind of thing where you are just putting your data or doing sequencing and generating your data you cannot 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 to take it.

And do the analysis on cloud ok. So, at this stage if you have like you have got the idea about the variants right now what are the variants that are coming in.

(Refer Slide Time: 17:05)



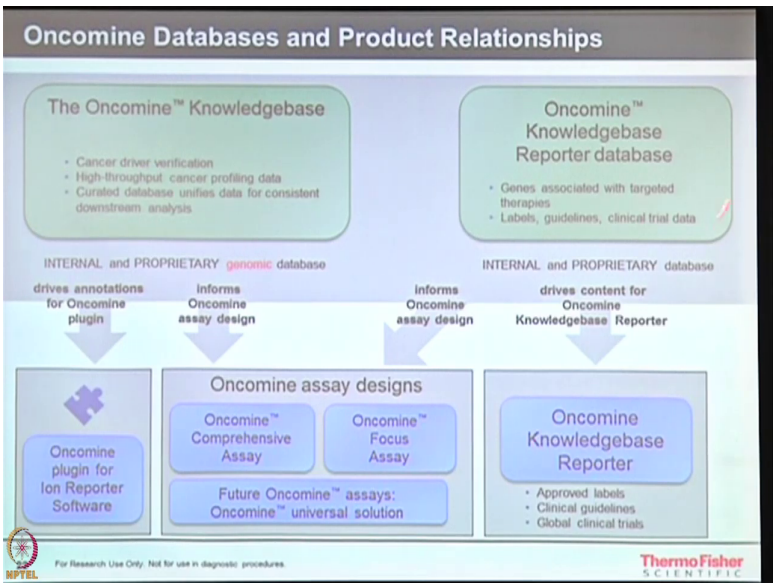
But even after doing so much of study what happens normally. Nowadays, we require something where we the results would be very quick; quick as in sense even if you are got the variants we need to get to know what is it affects and what is it is counter effects happening across. And then what could be the drugs which could be available for such type of diseases or variants that you are getting right.

What you are doing over here is your somewhere correlating your patients samples information to the variants getting meaningful information and then you are trying to generate the information; what type of drugs could be utilized right. You have a drug bank database

that is correlated to this, but doing this would take some time right 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 us Oncomine knowledgebase reporter ok. Oncomine knowledgebase reporter has a database called as the Oncomine knowledgebase ok.

(Refer Slide Time: 18:18)



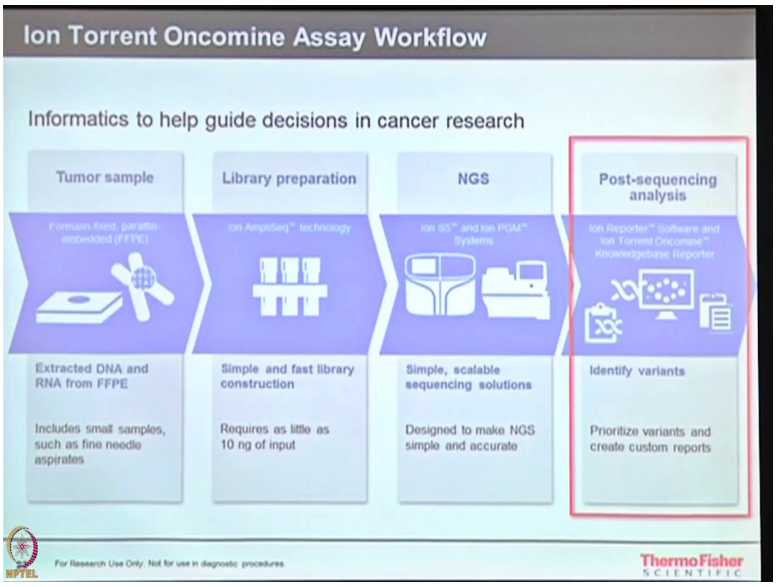
This Oncomine knowledgebase even if you Google it you will get an idea it has lots of information regarding to cancer ok. This is specifically for the cancer related studies or else, this is something like which will help you to correlate your cancer, cancerous variance to the drugs that are available in the market or into the clinical trials ok.

So, you have certain databases you have certain cancer driving; drivers data available which are verified and studied by the researchers over here. So, same database we have generated over here called as Oncomine knowledgebase 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 knowledgebase reporter interface.

So, now what happens over here it has lots of information of your variants called as cancer or a driver information ok. This driver information is correlated with different drugs information and the same thing is available as my design assays. So, what I was talking earlier into ampliseq that I like to design my gene assays.

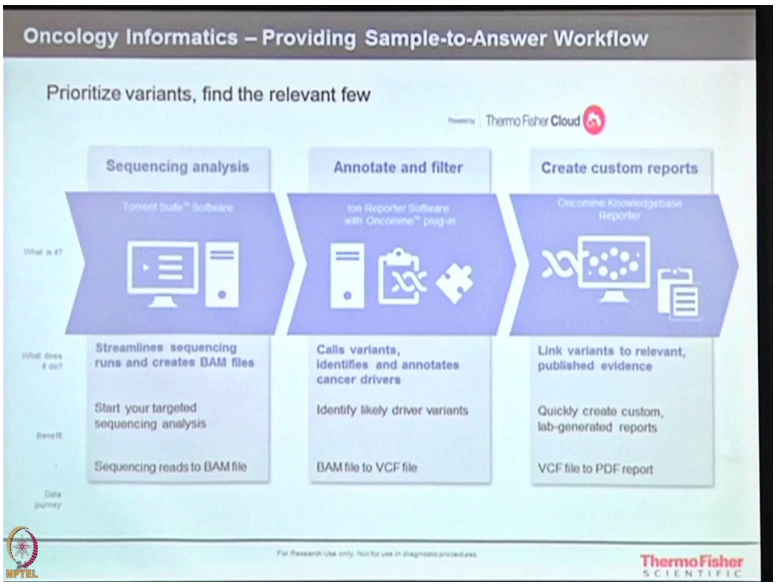
Or gene regions target of interest and then sequence it to know which are the diagnostics variance that you are getting. The same thing is coming over here you have certain assays which are available which can be utilized to detect which are the variants data getting coming into your samples. And, the same samples are utilized to denote whether that variant that has been detected is having any clinical significance or not and with thus clinical significance whether it is correlated to any drugs or not ok.

(Refer Slide Time: 20:17)



So, this is something that you have as a summary you have something called as tumor sample you take it further, repair your library you do your sequencing over here.

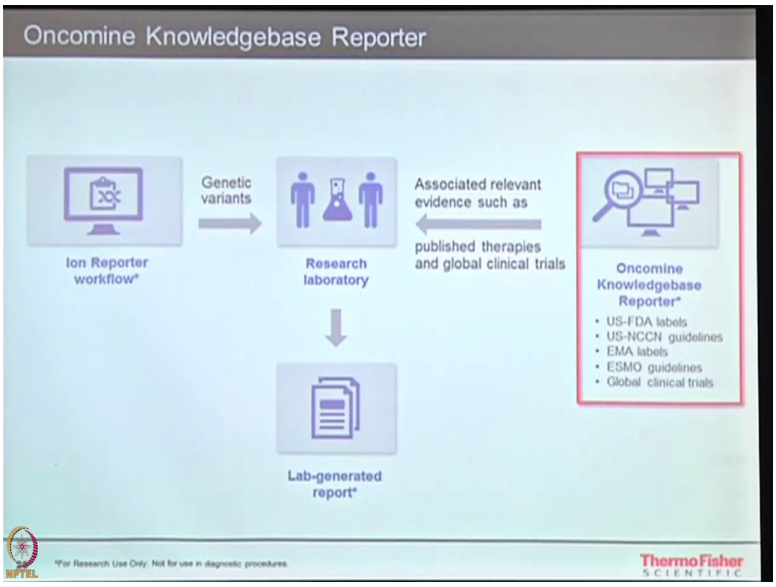
(Refer Slide Time: 20:30)



And, then you do certain analysis your 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 ok. 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 ok. So, at this point if I take further.

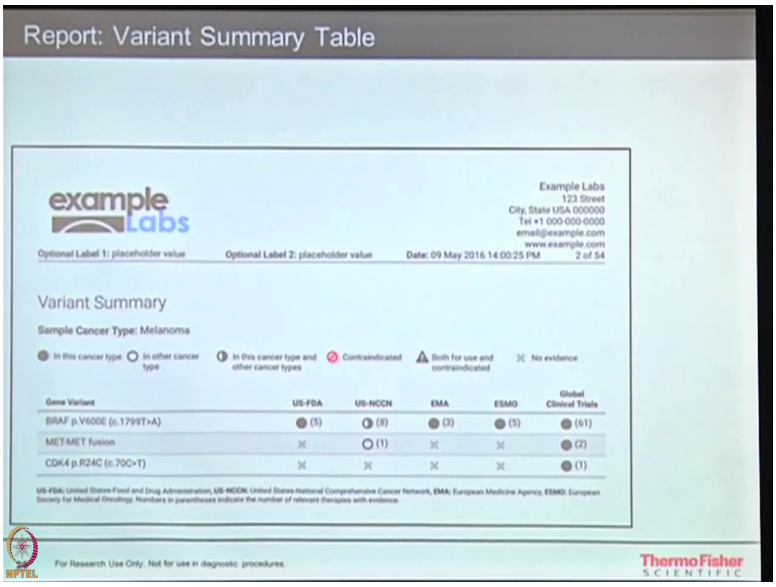
(Refer Slide Time: 21:13)



So, you have something like this you have ion reporter coming into play you have genetic variance. You have a knowledge base reporter which has different guidelines information regarding to your drugs, information regarding to a clinical trials everything over here you correlate them and finally, generate a report for it ok.

And this would be very helpful when you are trying to give a drug to a particular patient right it would be very quick for you. So, a patient comes you do this sequencing on 1 or 2 days. Look for the drug that is actually having any effects give the report to a doctor he could let you know what could be the drugs utilized further for a particular treatment ok. So, this is how we can come to a place which would be very quick in responding to cancers different types of cancers.

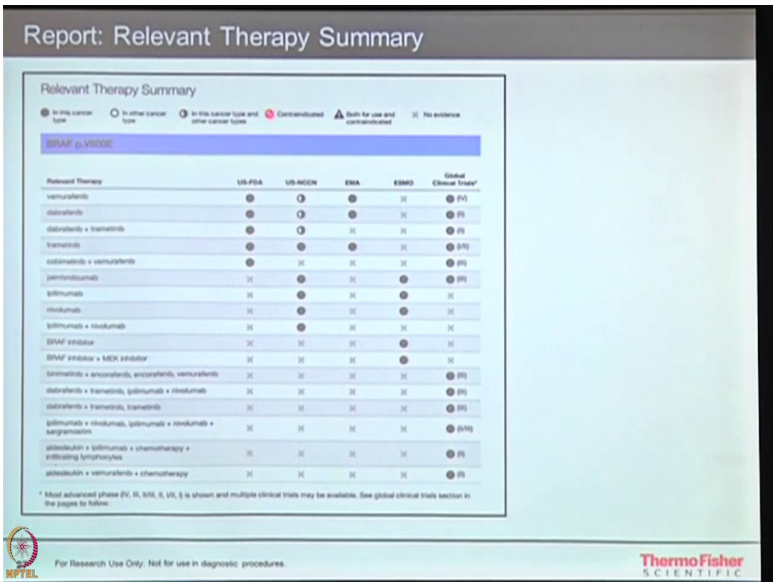
(Refer Slide Time: 22:06)



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 C to T change that is available and there is a T to A change happening over here also there is a protein change happening.

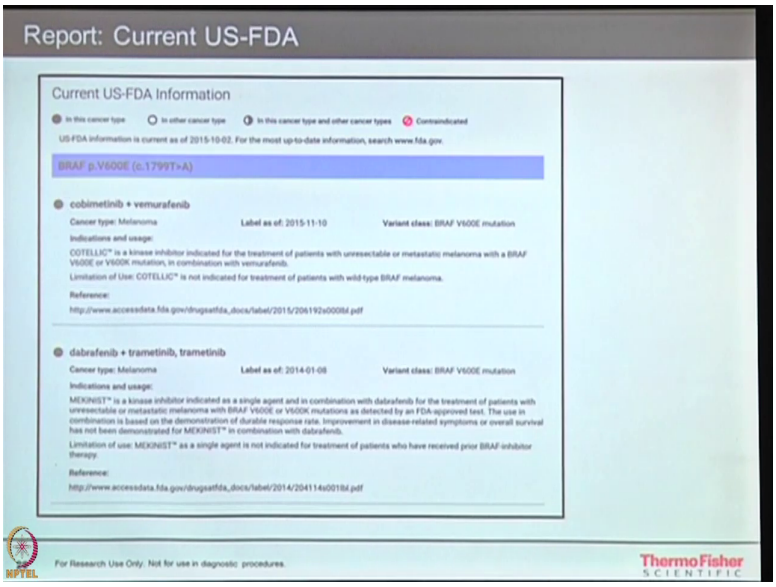
V600E that is a protein change happening over here amino acid. In the same time there are different clinical trials and drugs affiliations available. So, if the drug has been studied someplace and has been giving; given a go for the utilizations. So, everything has given over here. So, at the same this is something like a summary which gives you idea whether it is going into a clinical trial or not whether it is a US-FDA approved or not ok.

(Refer Slide Time: 22:50)



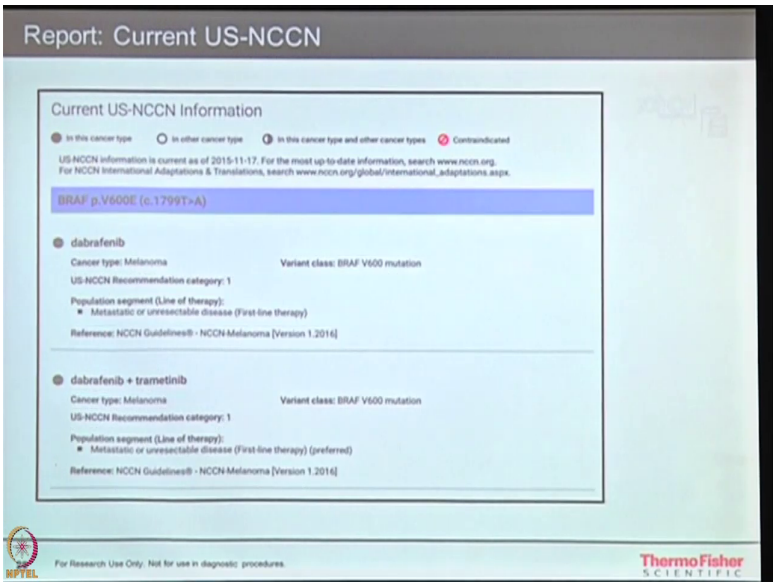
At the same time you can get the more information about more drugs center. or different variant that you are looking at V600E. So, these are the various therapies that are available drug therapies. So, who have given which approval which stage is approved global clinical variants, in which stage it has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

(Refer Slide Time: 23:12)



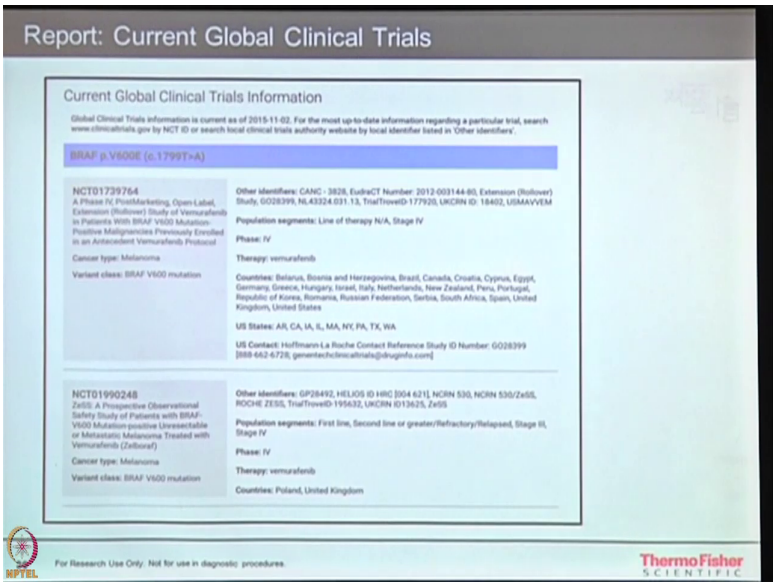
At the same time what I like to know is in detail in depth. If I am looking for a particular drug I am concentrating on that what study it has been done across those particular drug. So, that studies are coming over here ok. So, what type of drug effects are happening over here. So, those information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

(Refer Slide Time: 23:36)



And this finally, comes to a level where you can finalize it and give a report to the doctor ok. So, based on my different information or different approvals that we have got this report comes in.

(Refer Slide Time: 23:51)



So, once this report is here you can utilize 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 and the doctor, who wants to give that drug and take it further so.

Student: (Refer Time: 24:12) these are accepted by the (Refer Time: 24:15) each other.

So.

Student: In general.

In general as in the sense somebody is using the ion reporter; sorry was 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 they have all this drug information they just take out which is the drugs that are really required and provide the details to the clinician or to the doctor as such.

Student: So, buddy information has (Refer Time: 24:55) clinician (Refer Time: 24:57).

Student: but the clinicians do not increment that is our experience (Refer Time: 25:01) incremented, because it is not at the standard medical power regime or it is a not a standard.

So, that is that is what the person who are working on this are giving into a standard format.

Student: It should be in the WHO accepted total values.

Right.

Student: (Refer Time: 25:18) so

Right.

Student: (Refer Time: 25:19) So, clinicians are also responsible.

Right, right your right.

Student: (Refer Time: 25:23) certain treatment back on the (Refer Time: 25:25) contraction so they do not what to change or they (Refer Time: 25:29).

Ok.

Student: So, has this (Refer Time: 25:31).

So, this has been integrated; this has been done like these are the approvals that they have got through the US-FDA is a drug whatever information we are giving is based on the approvals data their available like US-FDA or the European I do not get that name.

Student: ESMO.

Yeah.

Student: ESMO.

ESMO, ESMO. So, they have utilized the same information over here to summarize and give it to it. So, how to represent 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 the scope they try to work on that and then give it.

Student: But, my question is have can be incenses where the move has (Refer Time: 26:16) approached clinicians and we try to make some changes under the w h o food policy (Refer Time: 26:23) approach (Refer Time: 26:24).

No idea about it; no idea about that we do not have an idea about it because whatever we are going.

Student: Research but.

Ok.

student: It is not a (Refer Time: 26:32) because of (Refer Time: 26:34).

Actually, that is why we have brought this particular tool, like a this particular software which helps you to get the information in such a way that it could be acceptable.

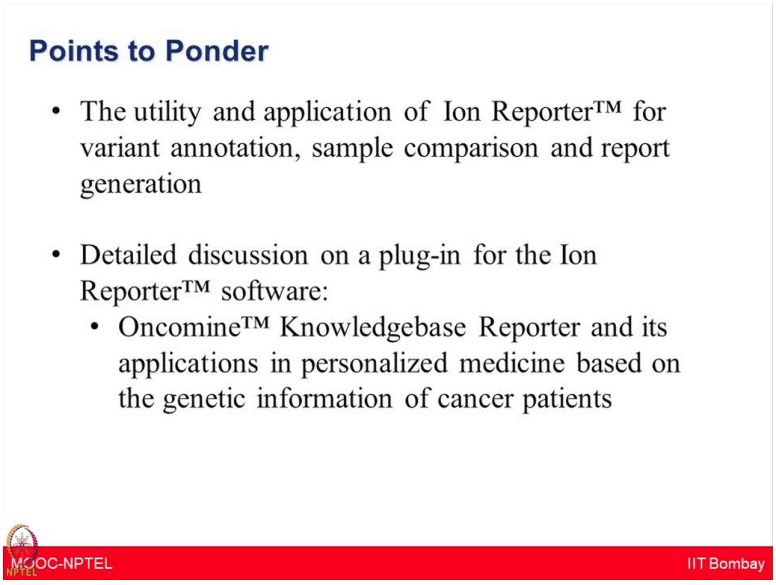
Student: So, we are working with the clinicians (Refer Time: 26:46) at the level where we are doing (Refer Time: 26:48) (Refer Time: 26:49) participating in those (Refer Time: 26:51) available for this tools, but then this is called something to do beyond (Refer Time: 26:58) beyond.

Student: (Refer Time: 26:59).

Student: So, beyond that (Refer Time: 27:01) it also creating the awareness (Refer Time: 27:03) we are doing that were (Refer Time: 27:04) tools (Refer Time: 27:06). So, now, for you to implement these transpose comment on it is after that, but then case in terms of bringing the awareness that we are already working to it.

Yeah. So, yeah. So, any questions about this I hope so you have got an idea or a feel how the data analysis happens in NGS ok. The only thing over here was to take you through the entire workflow how we start with the analysis, how we like to look into the particular gene designs. Come to a level where you do the sequencing for the gene designs and then get it annotated through ion reporters ok. It is a tool that is freely available and then looking for something like a quick response thus Oncomine knowledgebase reporter which helps you to do that ok. So, thank you very much for your time.

(Refer Slide Time: 28:10)



So, today's lecture you got a good understanding about one of the useful resource which is Oncomine knowledgebased reporter from where you can get lot of information for the cancer research. And, again you know while we are really getting biased towards cancer or clinical applications, but there are similar kinds of resources available for your model organism of interest as well. It is really important for you to dig deeper and know your available resources which many a 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, this database also contains a very scripted information about cancer, its variant forms different type of drug and therapy is impact. And you know it can be really useful for you to now start adding information on top of our data from the curation and see

that you know whether you can build now some hypothesis which could be actionable hypothesis now to take forward for your experiments.

So, in this lecture you have learnt more about you know how to do variant data analysis obtained from NGS platform which could be correlated with the patient data to study cancer will continue more about NGS platform I am sure it is one of the revolutionary technology and exciting areas to discuss. I will to also try to bring more application orientation and application scientists to speak to you about latest developments happening in the NGS base platforms.

Thank you.