An Introduction to Proteogenomics

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

Supplementary - 2
Advancement in Cancer Genomics

Welcome to MOOC course on Introduction to Proteogenomics. In today's lecture, we have our industry expert Dr. Mukesh Jaiswal who will talk about advancement in cancer genomics. He will also give a brief introduction and overview of cancer genomics followed by recent applications of genomics in the area of cancer research. Dr. Jaiswal will talk about challenges of doing cancer research especially accurate diagnosis of cancer, as a result the cancer is becoming global burden and how by using new technologies like next generation sequencing technologies one could try to provide better diagnosis and treatment strategies.

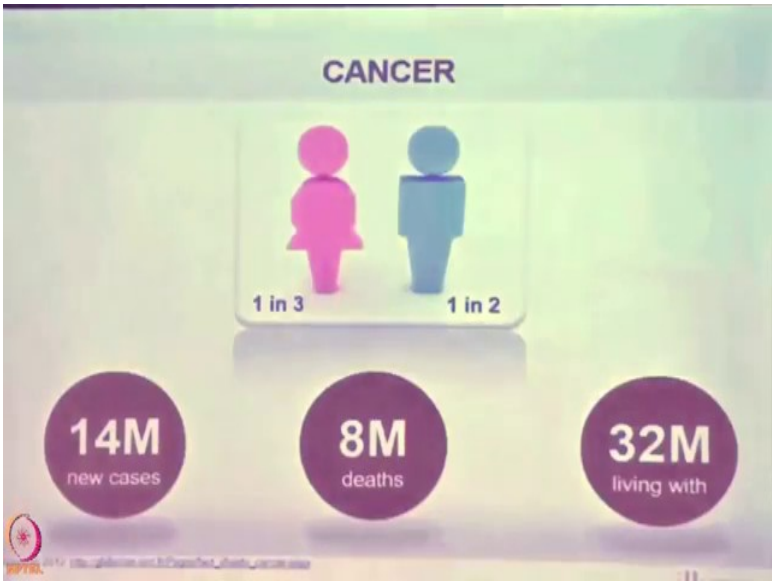
He will also talk about different strategies of treatment for example, the conventional chemotherapy, radiation therapy and various type of drugs which have been used for cancer treatment but more focus will be in which way one could start using new diagnostic tools by utilizing NGS technology. So, let us welcome Dr. Mukesh Jaiswal.

(Refer Slide Time: 01:48)

I am going to give little drive how Illumina do the cancer genomics. Today's agenda I am going to give brief introduction of cancer genomics, right and how this it is utilized for the

companion diagnostic and how it is used for the treatment part. So, application part and then some introduction.

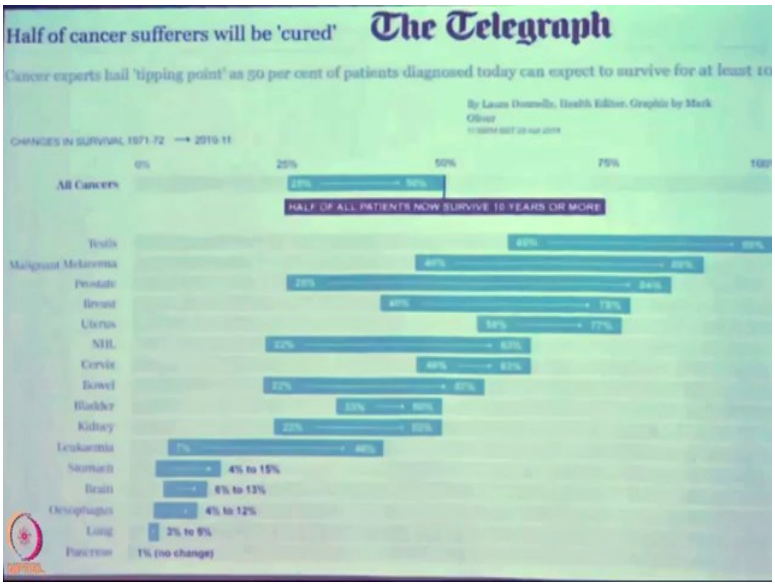
(Refer Slide Time: 02:08)



So, let us start with some introduction. Cancer is basically if you see in the life span of any women out of 3, 1 going to be have a cancer; a man is 1. So in your lifestyle lifetime every 3 women going to, everyone 1 out of 3 going to be a cancer and every 1 man out of 2 going to be a cancer, right. 14 and 14 million new cases coming every year, right and 8 million death every year and 32 million basically is living with cancer, right. So, it is a big number. It is a global data, it is not Indian data. It is a global data but every year we getting number of people adding there, right.

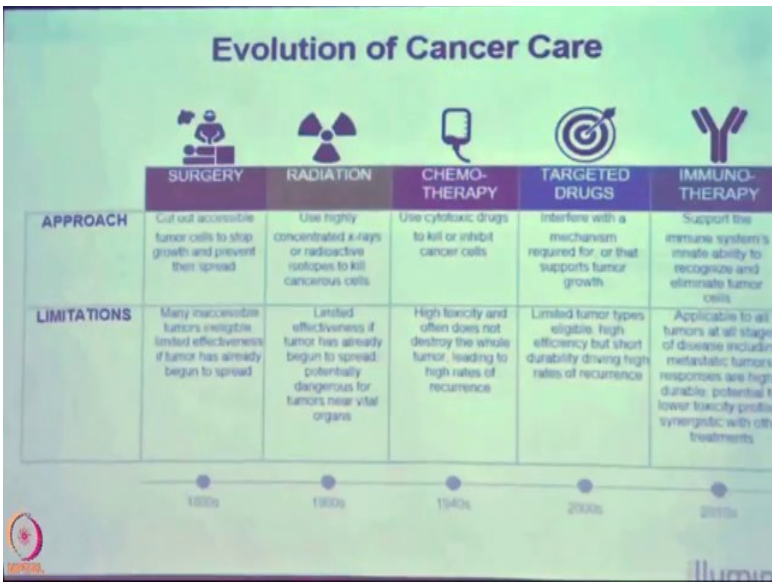
So, it is very important how to diagnose the cancer and then the treatment, right; but good thing is that also of this figure actually, nowadays technologies advances in such a way, right we increase the life span of the cancer patient 10 years, right.

(Refer Slide Time: 03:23)



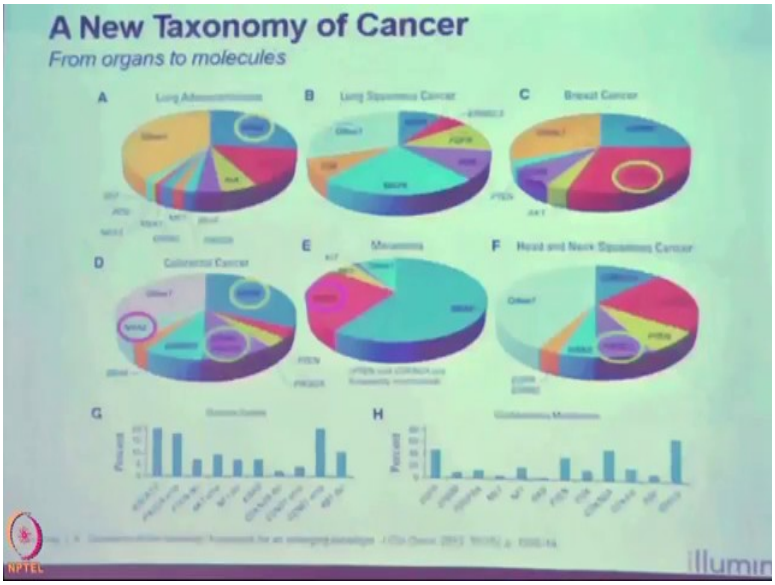
So, out of this is the data, like out of like 50 percent of the who are cancer, 10, half of the patient basically alive till 10 year of the cancer treatment. So, the time life the frame is increasing and this is varied for different cancer type, like cancer if it is testis cancer, it is basically 69 to 98 percent vary but see the lung cancer is still is like 3 to 5 percent, right. So, if you see this is the more it is better for treatment, this is has the less treatment, right. So, we need to really work out which is the best biomarker for the diagnostic and got with the better treatment possibility, right.

(Refer Slide Time: 04:14)



So, like if you start with like very early stage how we basically do the cancer care? Initially, like very early stage, surgery and radiation, chemotherapy, right. 1940s now still we will be doing chemotherapy, right but now after the discovery of genomics, right 2000 we start targeted sequencing, targeted drugs, right and now is the novel drug therapy called as immunotherapy, right. So, Illumina focus is basically this part, really help the cancer patient for the targeted therapy and followed by the novel immunotherapy. So, I am going to cover these.

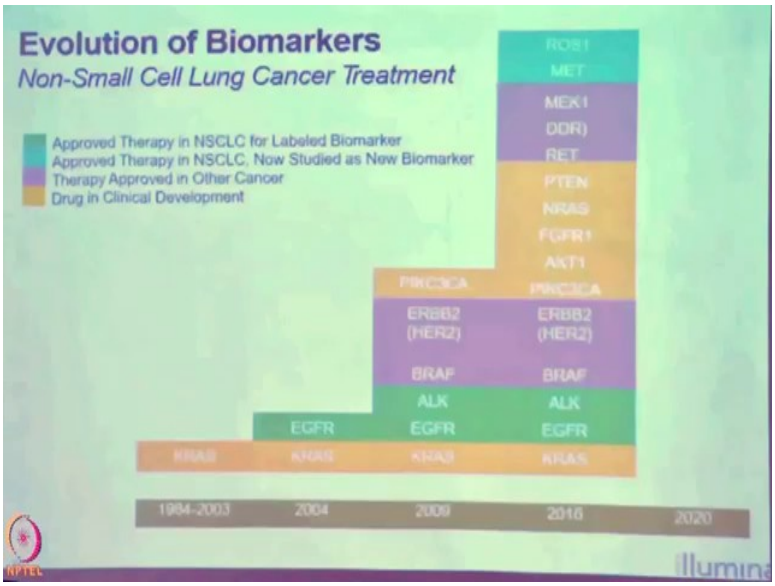
(Refer Slide Time: 05:01)



Initially like, when there is a cancer people say this is a lung cancer, right, breast cancer, but now because of these technologies these organs goes to molecular laboratory, right. When there is a lung cancer these subsets of the genes base; subset of the genes which get mutated and the; these are differential mutation. Some organ has different type of mutations some has different, right.

So, now we can say if you have a lung cancer these are the subsets of genes which get mutated, right and these are multiple, it is not one now, it is a multiple. It start from the one gene now it is going to be a several gene for one; one cancer site, cancer type and some of those are common also, some are common also, right.

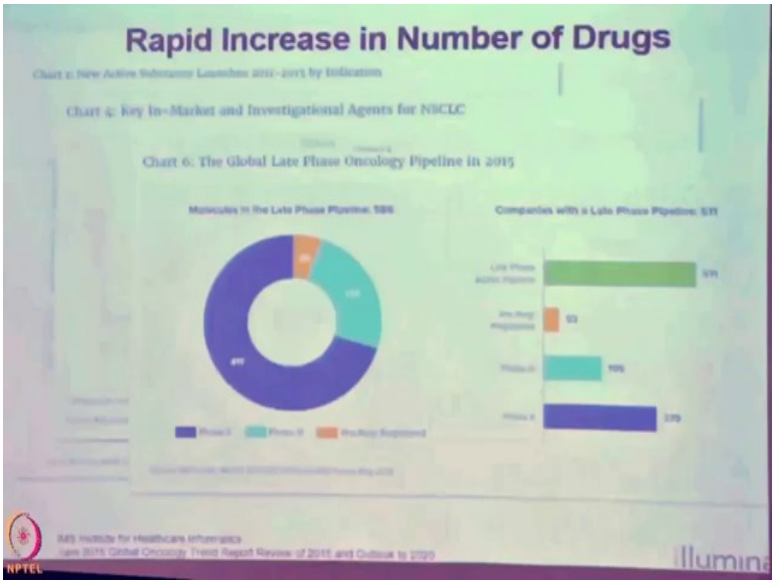
(Refer Slide Time: 05:48)



So, coming to the one example see the lung cancer, right. See 2003, near KRAS only now 2016 we have like 17 different gene added, right and these are basically all with the discovery phase and some are basically in the clinical trials also. There are drugs available for that because of the discovery of new genes which are mutated we find to the targeted therapy for that and we have several drugs available for a treatment. Only important thing you have to do proper diagnostic at right time.

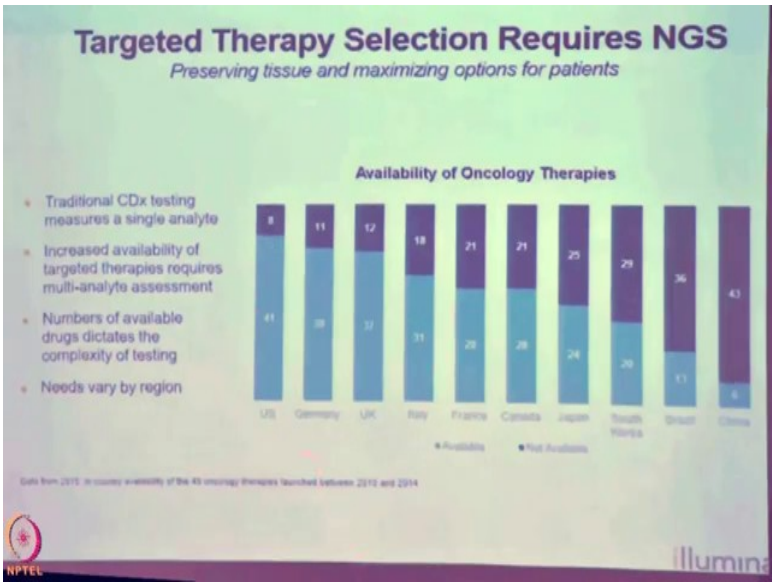
Student: So, for treatment purpose we should have target only one or multiple targeting.

(Refer Slide Time: 06:41)



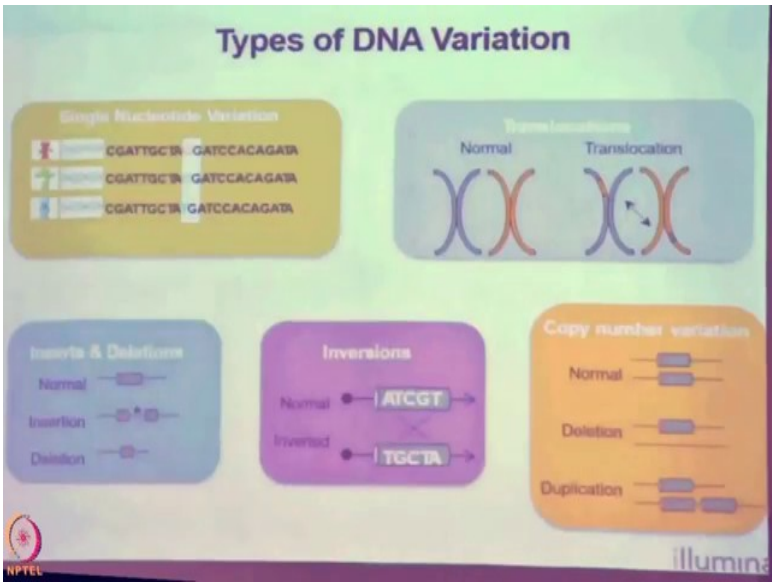
Multiple drug targeting sometimes. So, I will place some example. So, it may be multiple also. So, if you see the rapid increase of the drugs also, see here 511 drugs are in the late clinical trials. So, drugs are coming for the cancer with NGS, it significantly increased, right.

(Refer Slide Time: 07:54)



If you see the country wise, if you see the US number; see here this light blue is available drugs for the treatment of different cancer type. So, US have 14 different targeted therapy for the for a treatment of these cancer. Now see India, India right now it is I think it is not in number but see the China it is the they are increasing well also and 6 target therapy is there. So, I think we are still in discover phase but I think coming future we going to be come somewhere in this number, right.

(Refer Slide Time: 07:32)

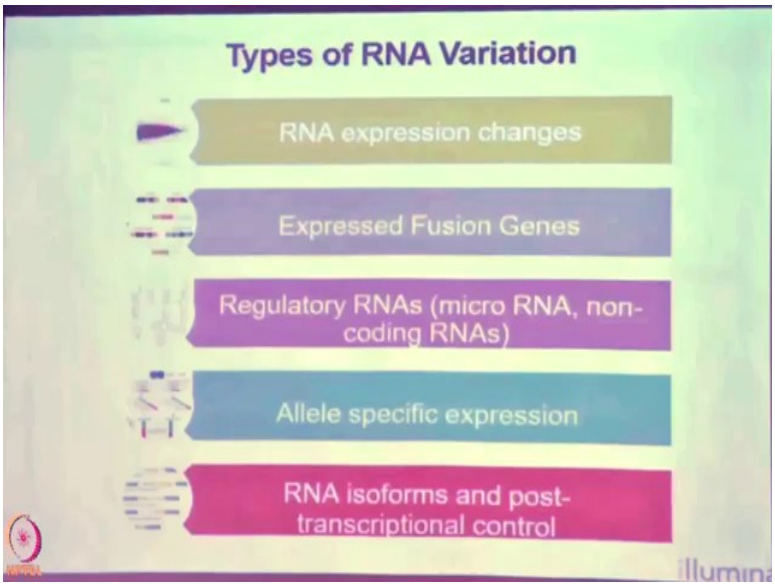


So, what is how we detect the mutations? That is really important. I just taking the very basic things what is the mutation in the gene, right. It can be multiple format. So, if we go gene; DNA variation, it is a copy number it is a single nucleotide variation that is called as snip; that means, one base change.

If you see this is DNA; see this is the one base change this is called as snip, right. Another form of DNA variation is translocation, right. This is normal chromosomes and some part of the chromosome breakout and it translocate to each other, right. It is called as translocation. Then, there is some part of a DNA inserted or deleted in the part that called insertion deletion, inversion and sometime the some part of the chromosome that is it duplicated or deleted; thus has copy number changes.

So, there are multiple variants occur in the DNA, it may be formed of single copy, SNP format, CNV, insertion, deletion. So, we need to detect this type of mutation very in very accurate way, right. That is how Illumina helps to the cancer patient for a detection of these variants plus also through the companion diagnostic we also prescribe the treatments, right.

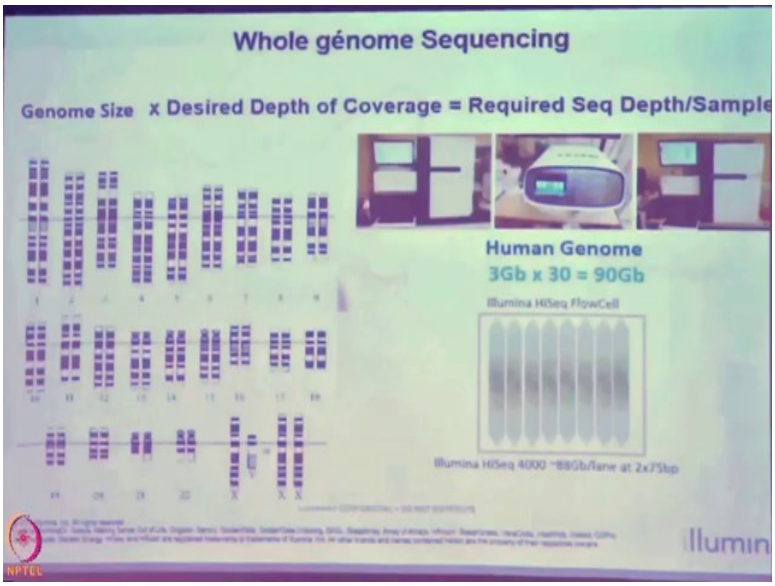
(Refer Slide Time: 08:56)



If you go for the RNA variants, so DNA is DNA we talked about, RNA we use. Sometimes the genes are fused, means RNA fusion, right. Sometime the expression of genes is changed, right, so we need to change look also for the treatment of the cancer. So, these are multi form of variation in the RNA format.

(Refer Slide Time: 09:22)

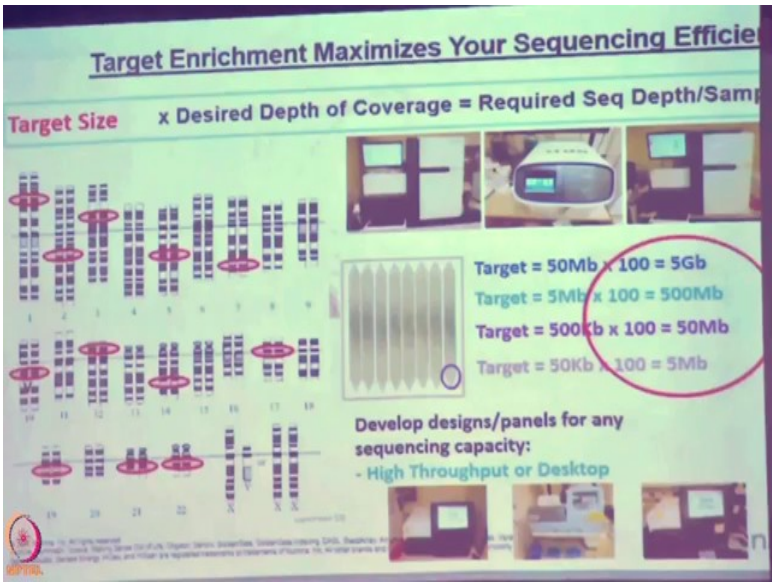
(Refer Slide Time: 09:24)



To detect these variation we do two approaches, one is the whole genome sequencing approach; that means you are sequencing whole genome, right. That means, you are sequencing all 24 chromosome pair, right, chromosome 1 to x and y, right. To do this you need. So, what sequencing does; it reads chromosome number 1 one time; 2, 3, all the time. It generates 3 Gb data when you sequence all the chromosomes, right, all the chromosomes 3 Gb.

But if you read 3 time 30 times; that means, 3 Gb multiply by 30, 90 Gb data is required to sequence genome in 30 times means for the accurate rate, right. What happened when you start reading, you start one time you add some error, right. So, we do multiple reads, right. So, for human genome it is required that you read 30 times. So, for one human genome; that means, you going to generate 90 Gb data and which is can be done in Illumina platform only from HiSeq to NovaSeq, right. I will tell little little all on that.

(Refer Slide Time: 10:40)

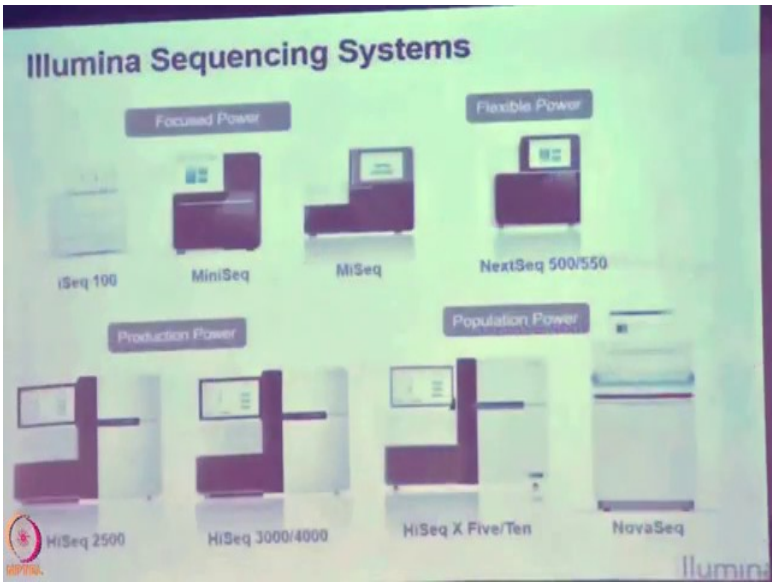


So, another approach is targeted therapy targeted sequencing; that means you are not targeting your genome, you are targeting only the small portion of genome; that is called a targeted therapy.

Suppose, your cancer patient has multiple gene is spread on different chromosomes what you do, during the library preparation you only pool down these parts of the chromosome, rest is removed, right and these red circled one is captured there is a different way of capture, I will take it in detail how we captured that thing. After capturing, what we do; we do the sequencing, it is a same calculation. If you do targeted sequencing and read by 100, 100 x means 100 times; that means, you generate that much data.

Suppose, this these all red part of the DNA its total target size is to 50 Mb, if you read 100 times; that means, 0.5 Gb data that is very simple calculation. It is totally depend how much sequencing data is required basically comes upon what is your target design, what is your target size. If your target size is 50 Kb only you want to do 100 as sequencing, you look at only 5 Mb data.

(Refer Slide Time: 12:08)

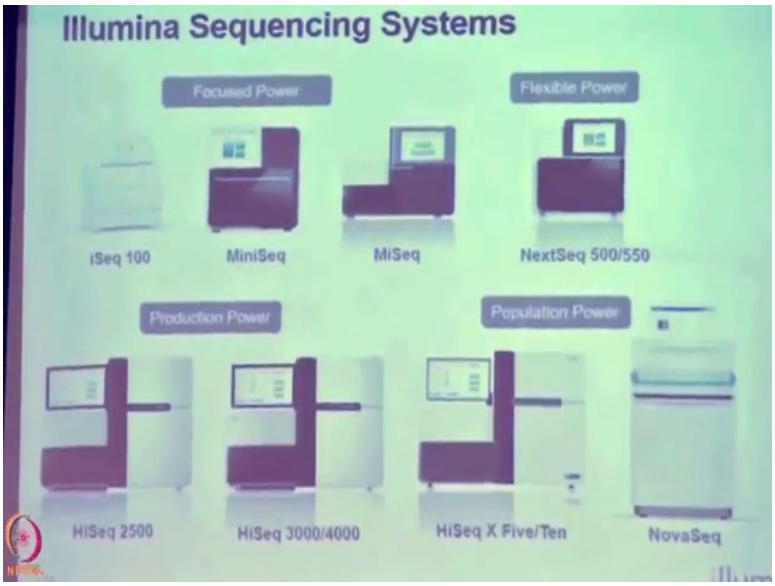


So, depending on your target, depending on your application you might require different type of sequencer. So, these are the Illumina sequencer. It start from the very low data output it is called iSeq; iSeq generate only 1.2 Gb data; that means, it is a; it is good for single gene, like BRCA1, BRCA2, right very small panel, right. MiniSeq it generate around 7.5 Gb data maximum, MiSeq 15 Gb data, NextSeq is 120 Gb data. It is all depends, it is height support, it is totally depend on what you want to do sequencing.

Suppose, if you want to do exome sequencing means the part of a DNA which express their gene, right. So, for one exome data it requires 5 Gb data, right. So, if you choose MiSeq you can do only 3 samples, requires 15 Gb divided by 5, so total 3 samples but if you go NextSeq 200 120 Gb that means divided by 5 you can do around 6 around 20 to 30 samples, right. Then HiSeq, these are the HiSeq series butwe are not basically selling this one.

Now, we have a NovaSeq, NovaSeq generate like 6 Tb data means 6000 Gb data; that means, you can do around 500 exome, 2400 genome in one run. So, capacity of the sequencing is really increased by the Illumina technology, right. I think none of the platform here in a in any player which can do this much data generation right 6 Tb data.

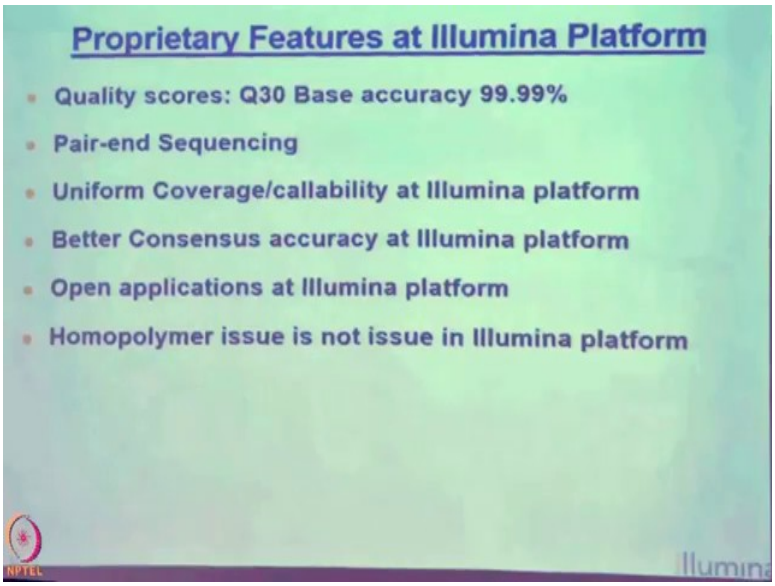
(Refer Slide Time: 13:45)



And these type of the sequence like NovaSeq really help the community for the cancer care because you need to do very deep sequencing to see that variant, right because there is two type of variant; one time germline variant which inherited to the mother and father, right; one is somatic which is generally doing your lifestyle, right and their frequency or frequency of detection is very low, 0.01 percent.

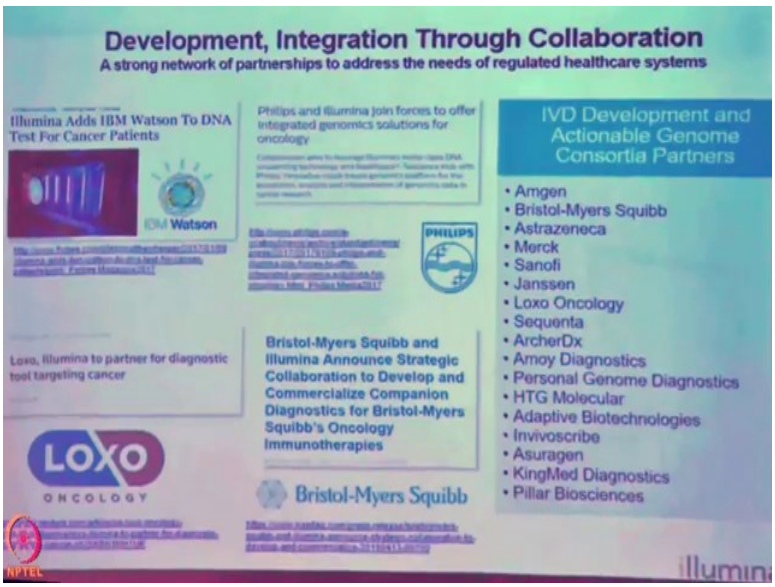
That means to detect that variant you need to go for high x read like 5000 x read, you need to go to 10,000 x read, right. For that you need to give at least have the bigger sequencer to get these data, right. So, Illumina, what Illumina is platform benefit is the quality, the quality of data is basically very high. The Q30 is 99.99 percent; that means Illumina sequencing incorporate 1 error in 1000 base pair, so really really high sequencing data, right.

(Refer Slide Time: 14:29)



Then we do pair- end sequencing, I think this is something very important when you go for the somatic mutation detection. Coverage is very high, so there is a multiple feature; this Illumina basically has a very good sequencing platform for the data generation and actually total whole human genome 90 percent data is done by the Illumina, right.

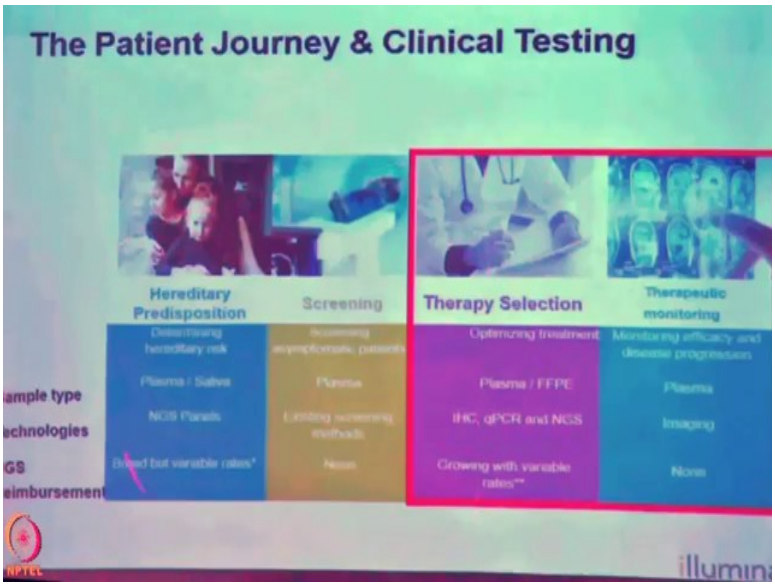
(Refer Slide Time: 15:24)



So, because of this technology, it is the for the cancer patient the different collaboration with the Illumina these are the collaboration people use these technology for the cancer cure, right. So, one is LOXO Bristol Myers, right IBM Watson, these are the collaboration which

basically use Illumina technology for the companion diagnostic. So, I am giving from this training, so I am going to give some of the example how basically these are used for the application for the cancer care.

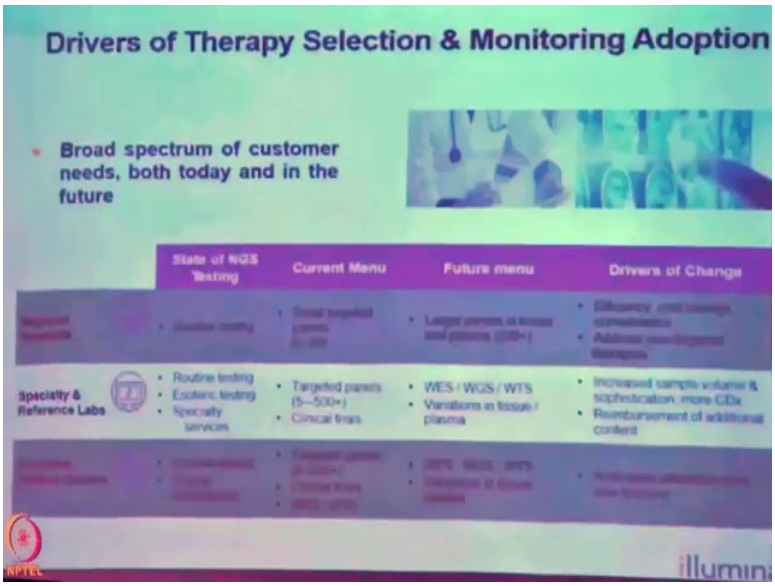
(Refer Slide Time: 15:56)



So, this is the how the patient generally start, right if you see the very important like what type of therapy you are going to give the patient this is really important thing and for this diagnostics is very important. So, nowadays there is a multiple sample size may be plasma may be affected tissues, right.

So, lot of people use IHC, qPCR and NGS also but there is some drawbacks when you go for the real time PCR or IHC because at a time you can do only one gene, right but for the cancer it is a multiple gene require you need to go for very high throughput sequencer to do the NGS, right. So, this is where the Illumina work start, what type of therapy you need to give to the patient and then monitoring.

(Refer Slide Time: 16:46)



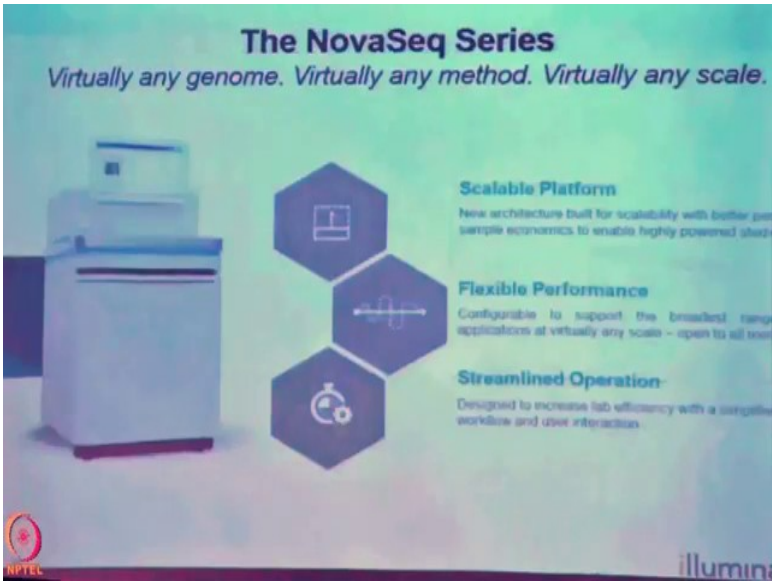
So, for this one depending upon the customer depending upon the if there is a regional hospital they do very small panels, right 5 to 15 gene, some are bigger clinical trials, right, they do 500 gene panels. Some are go doing whole exome or whole genome sequencing and they do different approaches to identify the cancer, what is a, what was the cause of that, right.

(Refer Slide Time: 17:19)

So, we have a complete solution with the library prep, instrumentation, informatics part which tells you like what to be the targeted therapy, what to be and what to be the how you

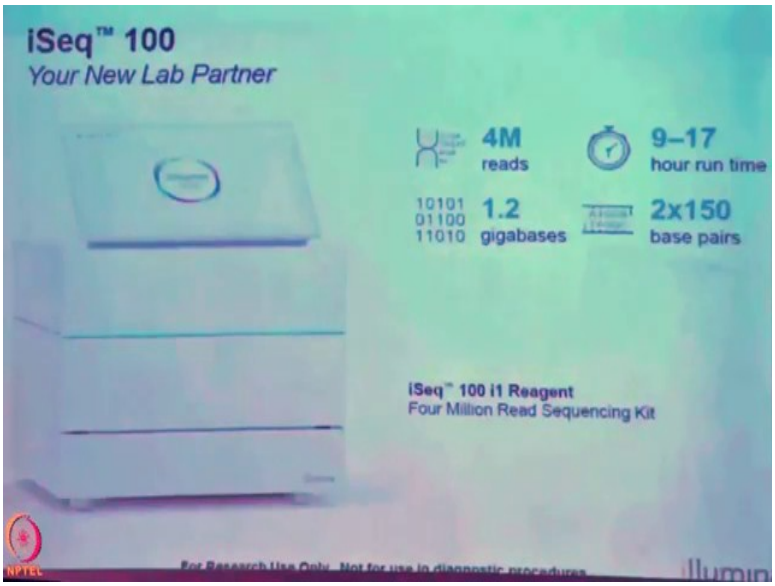
and then lots of clinical trial enrollment is going on and these instrument is also IVD approved, so that means you can directly imply into the IVD mode with research.

(Refer Slide Time: 17:44)

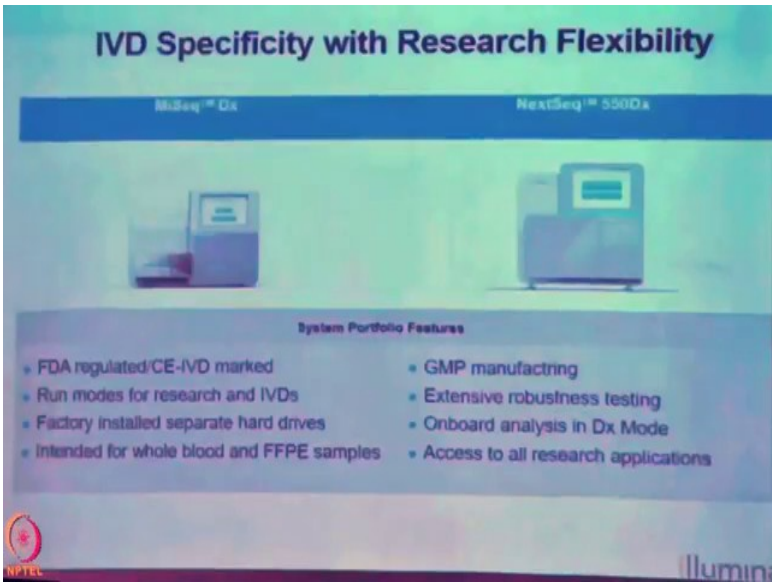


So, I am telling this one is really because this is the NovaSeq series. It genrates 6 Gb data but it can virtually requires any genome, any targeted panel, any method and any scale, right. So, this technology basically people using to use for a targeted therapy.

(Refer Slide Time: 18:06)

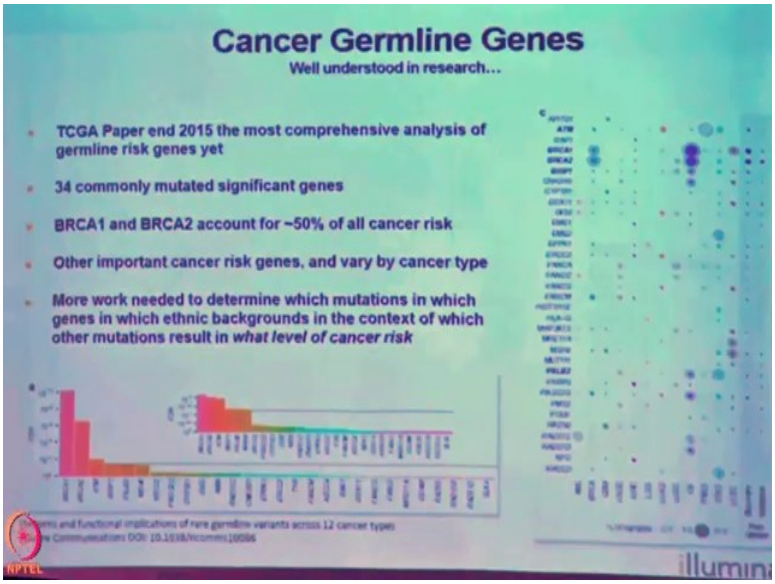


(Refer Slide Time: 18:09)



Let me show some example. This is iSeq. These are two Dx model of the NextSeq and MiSeq, some applications, right.

(Refer Slide Time: 18:15)

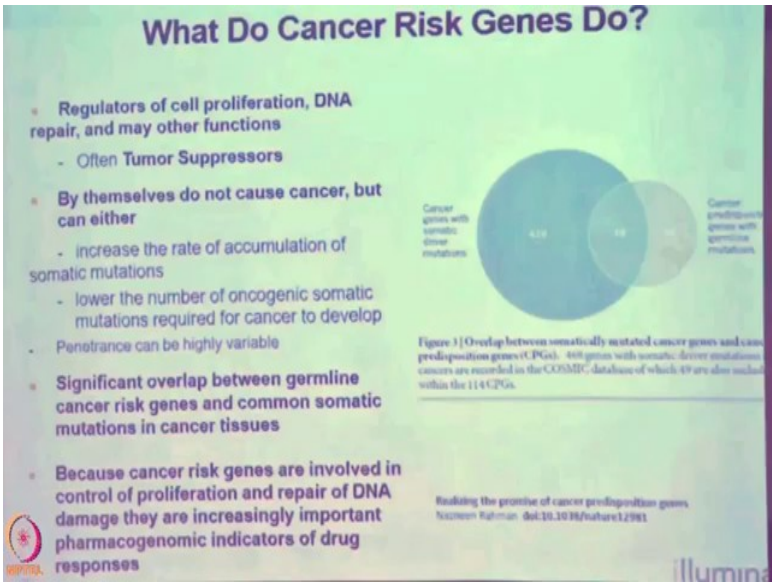


So, this is the; this is the genes basically. This is around 34 common gene, 34 common gene which is expressed in most of the cancer type, different cancer type and these are common, right.

And these are basically germline, germline mutations which are in 34 gene. Then if you see the BRCA here is basically common in 50, 50 type of the cancer. That means, you can there

is question, right, if somebody doing single gene panel and somebody doing the multiple gene panel the accuracy for the multiple gene is much better because it doing the multiple gene and these are common markers for all. So, it make sense you do the multiple gene.

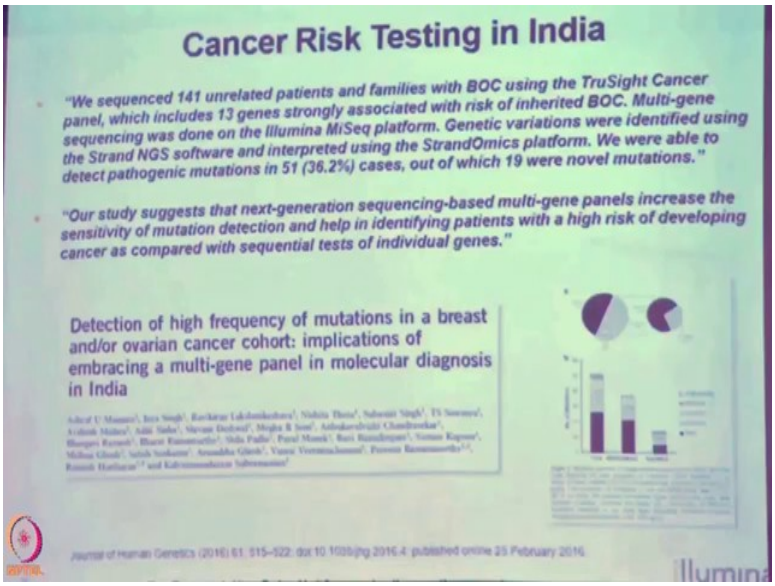
(Refer Slide Time: 19:15)



So, what cancer, cancer gene does; it basically regulate the cell proliferation, DNA repair and other function. So, if you see this is the hereditary gene this is like 114 gene which are overlapping with the somatic cancer, right and overall is around 500 gene basically which is involved in the cancer means involve in the cancer. So, it is already well studied, right. Some are hereditary, some are cancer somatic and both are overlapping in some extent, like 45 genes are overlapping.

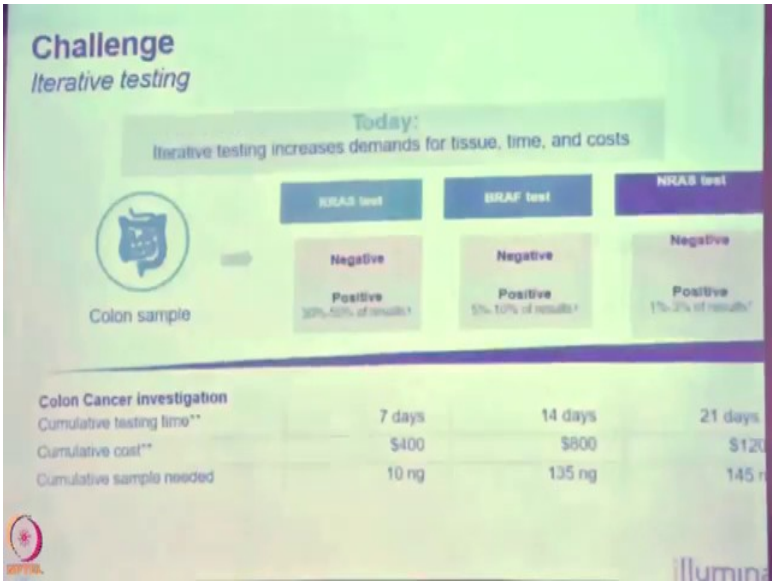
So, we try to design a one comprehensive panel, right to and it can be come from different organ also to detect the these mutation, one comprehensive, not one; one comprehensive panel.

(Refer Slide Time: 20:00)



So, this is one example this is done by the strands in India, strands in Bangalore. So, they use a multiple gene panel which is TST 1, TST 15 as a 15 gene panel and they studied this thing in the breast cancer and ovarian cancer, right and they identified that the there is a 51 pathogenic mutation which are common in the breast cancer and ovarian cancer, right. So, it is a value, if a value is there if you do the multiple gene, if you do only may be you miss it. Let us see the example.

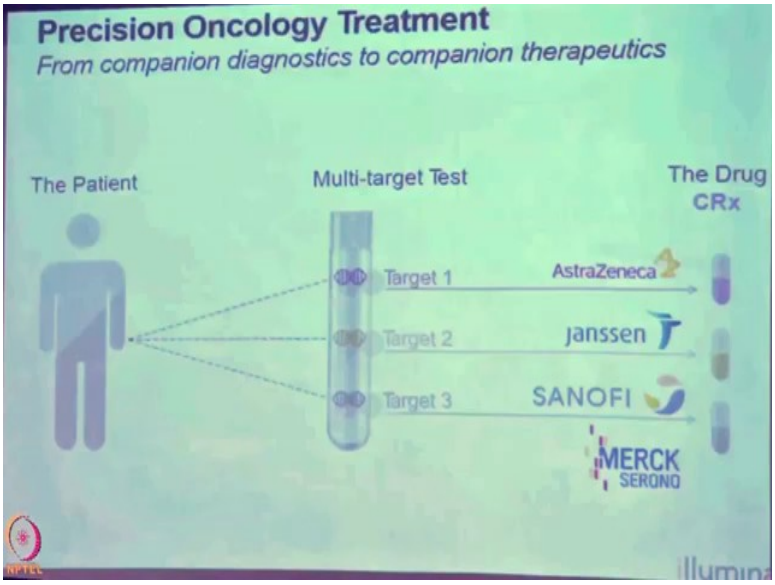
(Refer Slide Time: 20:35)



Suppose, there is a colon cancer and you get a tissue for the for that diagnostic, right and you do only one test KRAS, right and if it is negative, if it is positive that is ok, it is a therapy available, right. If negative, you have to do other another test; it is BRAF, right. If it is negative, it is another test which NRAS, right.

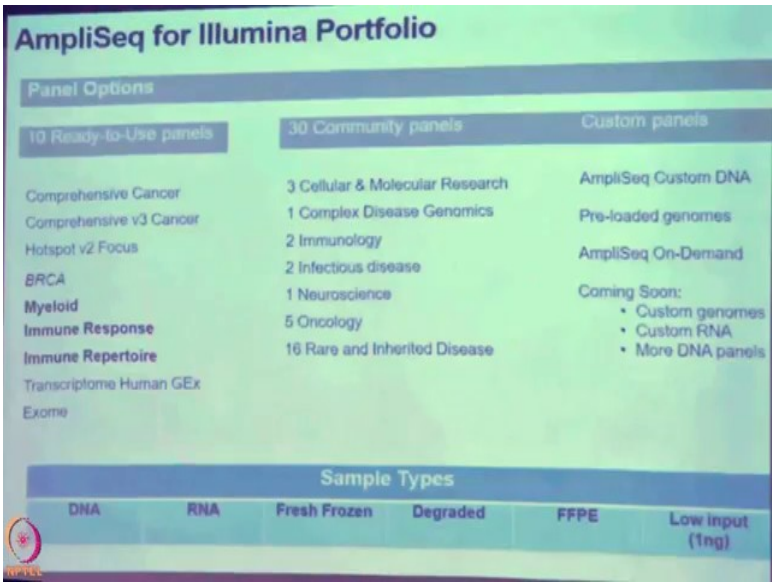
But if you do these tests it takes 21 days and amount of tissue basically gets more you need, every time you need fresh tissue and the cost of the per test is also add up. So, that means, if you have one common panel which can do multiple thing and all the mutation your therapy must be better, right.

(Refer Slide Time: 21:33)



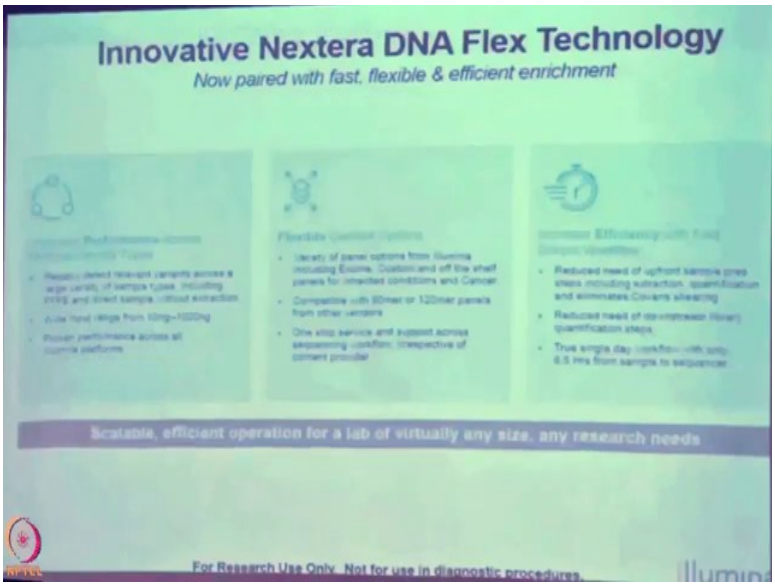
That is how the companion diagnostic or precise oncology treatment done. There is a patient, multiple target, means all type of cancer, one test. If you get something mutated, there is a different different informatics part and you your companion diagnostic treatments available, right. So, it is very important when what panel you are basically using.

(Refer Slide Time: 22:55)



So, Illumina has a multiple cancer panel also. So, if you we have the comprehensive cancer panel, we have a hotspot cancer panel, smaller panel also, the multiple cancer panel. I will go one detail one by one.

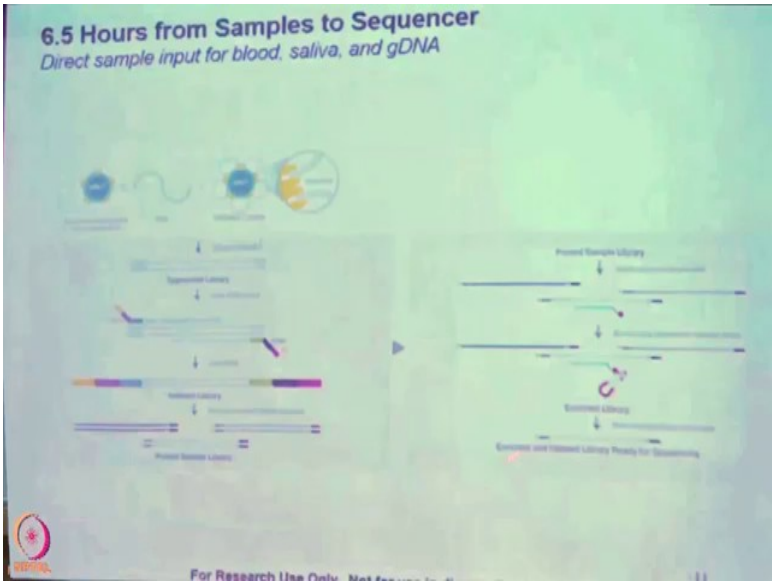
(Refer Slide Time: 22:12)



Let us, let us let me tell you how we make the library, like up to this point I told you we have the different caner panels, right, so, how we make the libraries, right. That is a very important part.

So, we use Nextera DNA flex library preparation, right. In this, this is the very very easy process it takes only 10 nanogram of DNA any type of tissues affected tissues, usually in the cancer type is a lung type, whenever you get you are going to have affected tissues and to make this library is only 6.5 hour. So, it is very easy, right and it is cost effective also. So, how we make it?

(Refer Slide Time: 22:49)



So, just to see this is your targeted DNA, right which you want to target, right. So, this is a transposase based enzyme, it fragmented DNA in the smaller fragment, right. See, so now, this adding the adapters, at this this is the target you we want to sequence and this is the adapters which is going to bind in the Illumina flow cells and indexes.

If you see because we do the targeted sequencing, so that means, you need have to have a probe which captures that target, right. So, this is the blue color is the probe. Probe it is like 80 to 90 mer big long and is biotinylated, right. So, if it is biotinylated is bind to this target and you have streptavidin beads, it pooled out only that part, right.

So, it gets only that part of DNA and that is the enriched part of your DNA; that means you have targeted you have enriched only that smaller part of your DNA, right. So, that means with this technology you can restrict target smaller amount of DNA for the sequence.

Student: How long it will be target that is enough?

So, it is totally depends upon your design, how it, it can be 1 gene only.

Student: I mean with one particular probe how long we can…...

It is a 18 mer long probe but if it is bind to DNA it may be bigger also.

Yeah. So, but we cut the DNA in such a format if you see here, here it is one 150 BSPR. So, it cannot be more than 150 BSPR for sequencing.

Student: So, if that many probes you design.

Yes.

Yes.

Student: I am interested in 100 Kb design.

Yes.

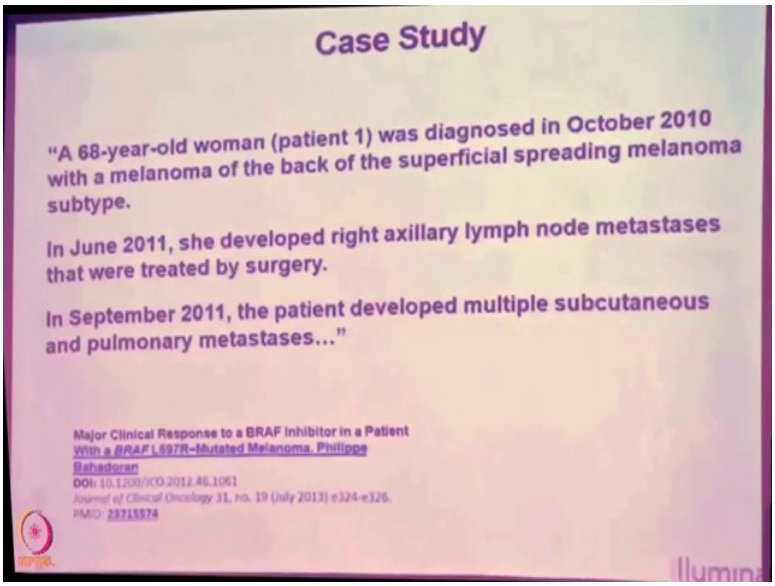
Student: So, can we design for the entire 100 Kb.

Entire 100 Kb, and we try to make sure you all the part of that gene is covered. It might be overlapping probes also.

Student: Ok.

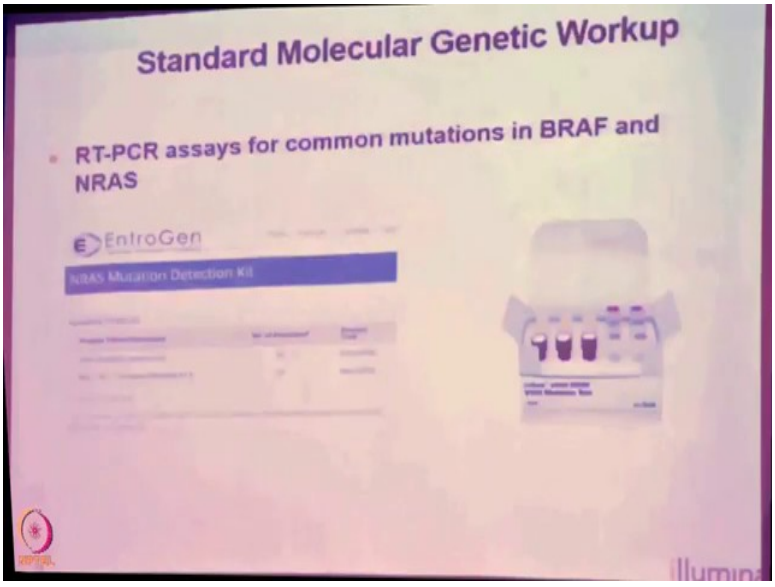
Right. Suppose, it if we do not cover we might lost some part, right. So, that is how we do the library prep, right, ok.

(Refer Slide Time: 25:03)



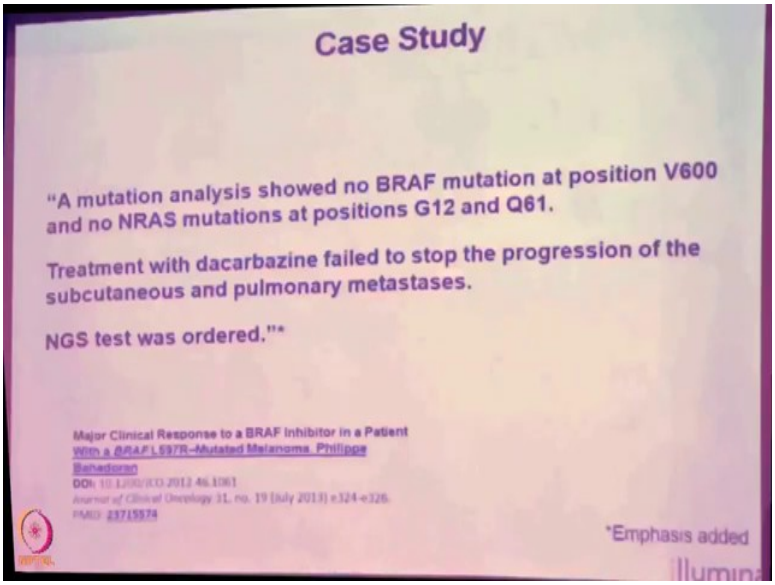
So, after library prep, I want to show one case study here. In this case study, a woman of 68 year old is diagnosed with a melanoma, right. June 2011, she, so it is surgically removed, right. So, after few months it is the that cancer is metastasized and it become, it goes to the lymph node and then after a 2-3 months they developed a subcutaneous pulmonary metastases, right. So, after doing surgery also it basically grow, it is increasing, not basically cleared, is not cured, right.

(Refer Slide Time: 25:48)



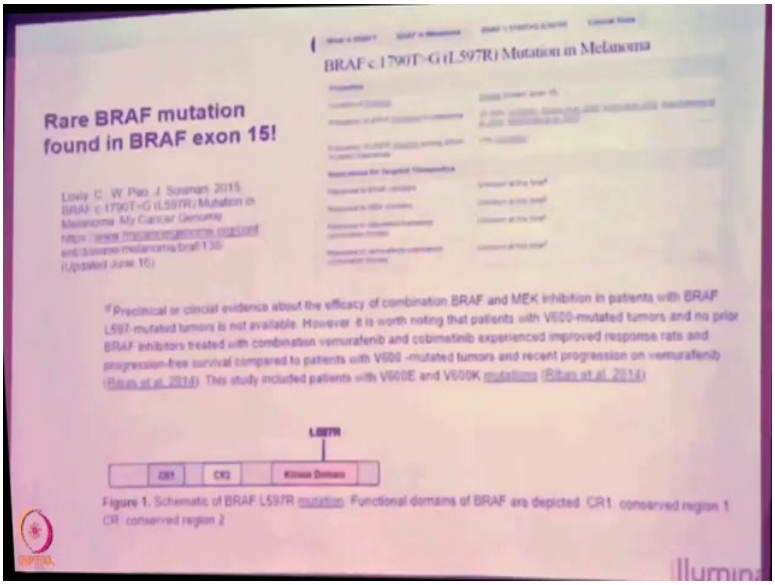
So, what doctor did; what doctor did; doctor ordered a test which covers only two genes NRAS and KRAS, right and there detain only 3 mutations.

(Refer Slide Time: 26:01)



So, when they find that there is a no mutation on BRAF and no mutation on NRAS, right. So, that is why therapy does not work because it does not cover all the things.

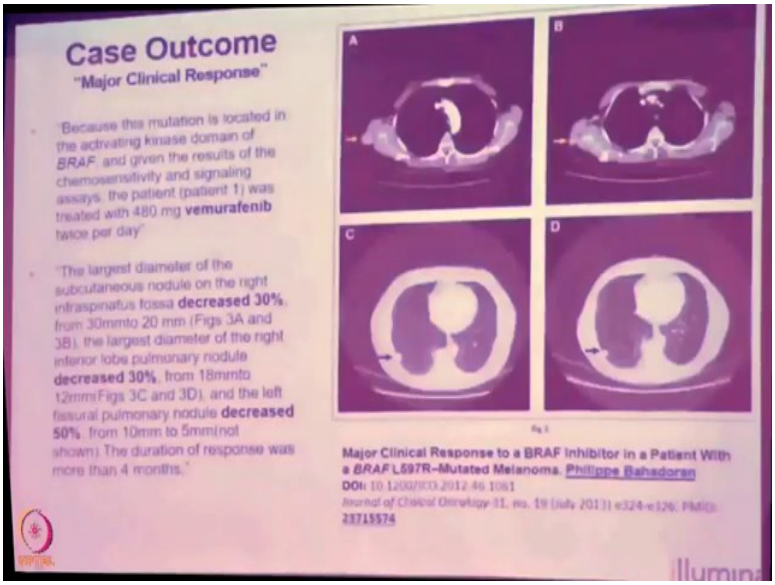
(Refer Slide Time: 26:21)



Then at the end NGS test was ordered, right, and then find very novel mutation, it is a different mutation on the BRAF gene. This is the kinase domain and this is the mutation type. It is totally different mutation, right. Now, when they find this mutation with the kinase

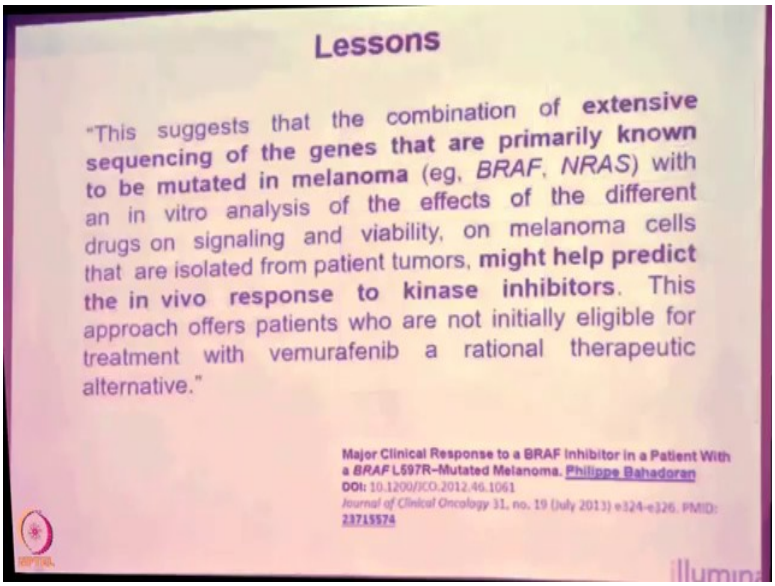
domain what they did, with the help of that chemotherapy they gave another targeted therapy to block this one.

(Refer Slide Time: 26:45)



They gave the kinase inhibitors and this is the name of the kinase inhibitor, right. With the giving of this kinase inhibitor the in few months 30 percent drop and another few months it is 30 percent drop.

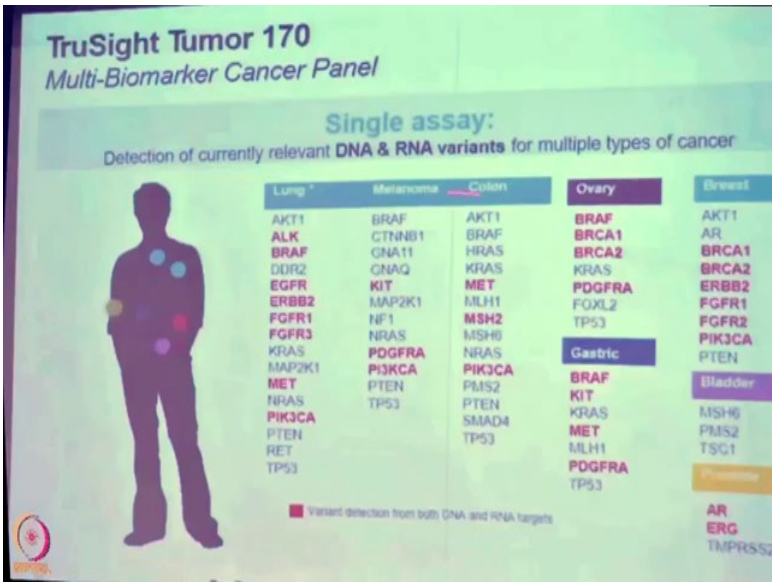
(Refer Slide Time: 26:58)



Finally, basically the case study is saying that study only of by the real time PCR, a few part of the gene is may not enough because you need to go cover all the part of gene, right and

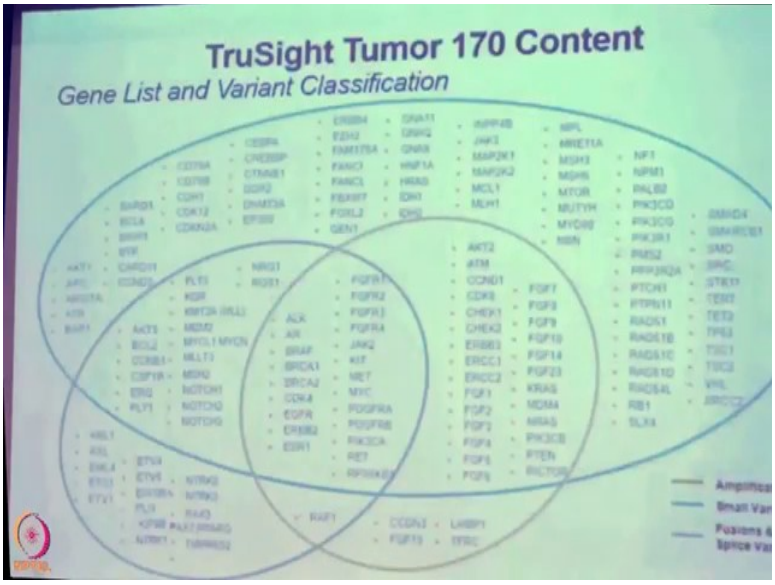
after studying getting this kind of mutation they find that no, you can do one chemotherapy per targeted therapy, kinase inhibitor, mixed together and is better for the treatment. So, this is one case study how people use the targeted therapy, right.

(Refer Slide Time: 27:26)



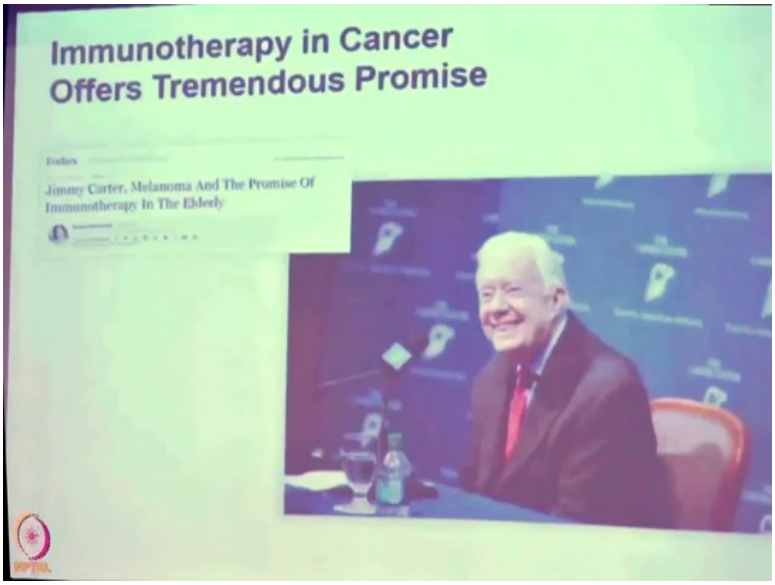
So, coming to this the important point, right doing one gene is not enough, may be 10 genes not enough because cancer is a bigger is big pathway, multiple genes are involved, it is not one gene. So, we recently we launched TST 170 panel, so we launched from this panel and it covers most of the cancer type.

(Refer Slide Time: 27:57)



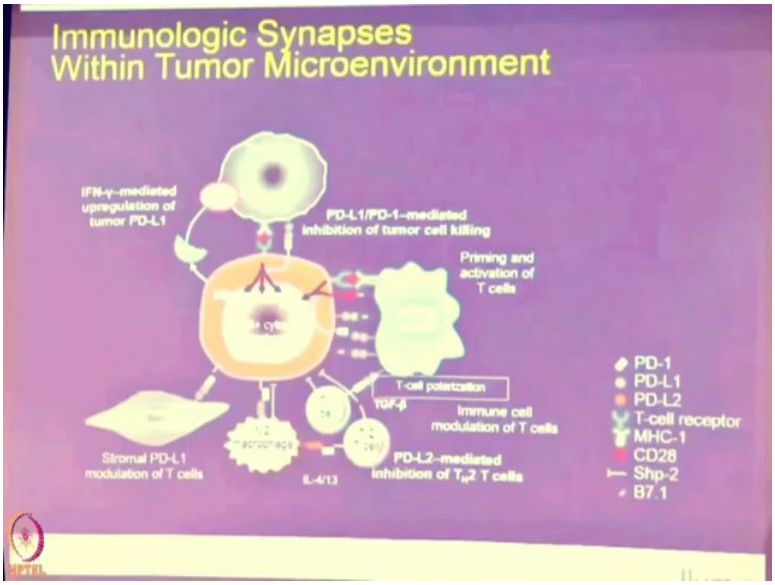
And it is involved to determine the different type of cancer, it also detect the copy number variation, RNA fusion and all. So, one gene panel, ok.

(Refer Slide Time: 28:07)

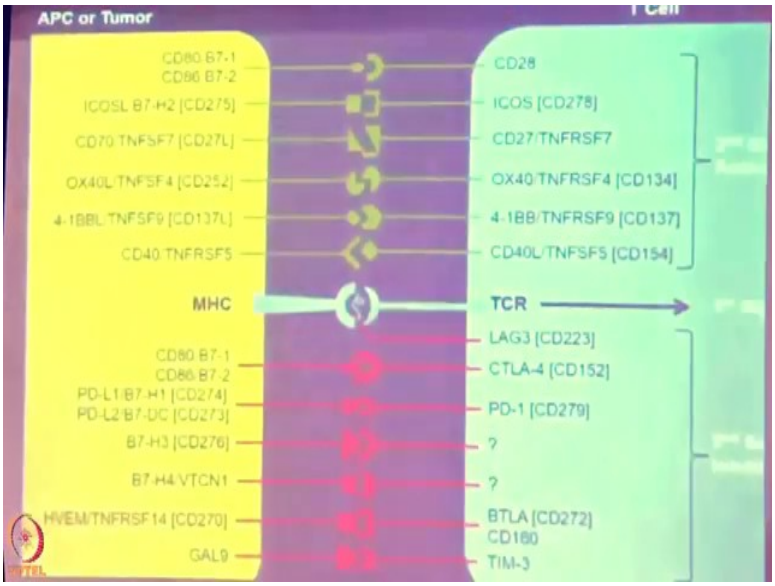


So, one last point like I want to share. This is the novel therapy by Onco, it is immuno onco. So, up to this point I showed like there is comprehensive panel made and different ways of therapy, right. I do not have time to cover all but let us go this one. Immunotherapy is really a very novel way to show like how we pointed to find immunotherapy.

(Refer Slide Time: 28:35)



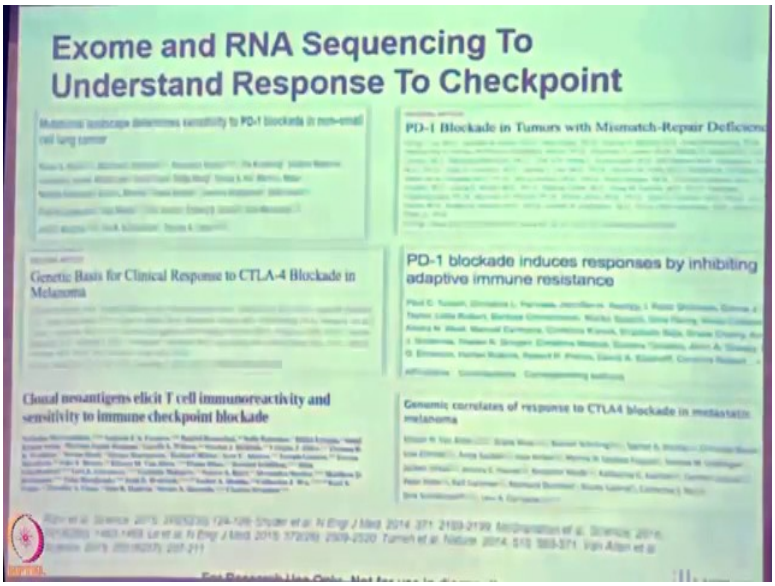
(Refer Slide Time: 28:42)



So, if you see here one simple example PD-L 1. So, PD-L 1 is a PD-1 is basically if you go this slide, this is the tumor and this is the T-cells and the PD-1 is basically suppressed in the T-cell, and PD-L 1 is suppressed in the tumor cell. Their interaction basically blocks the T-cell activation and allow the tumor cell grow, right. So, if the doctor give the chemotherapy, right, it not going to help because T-cell is not responding.

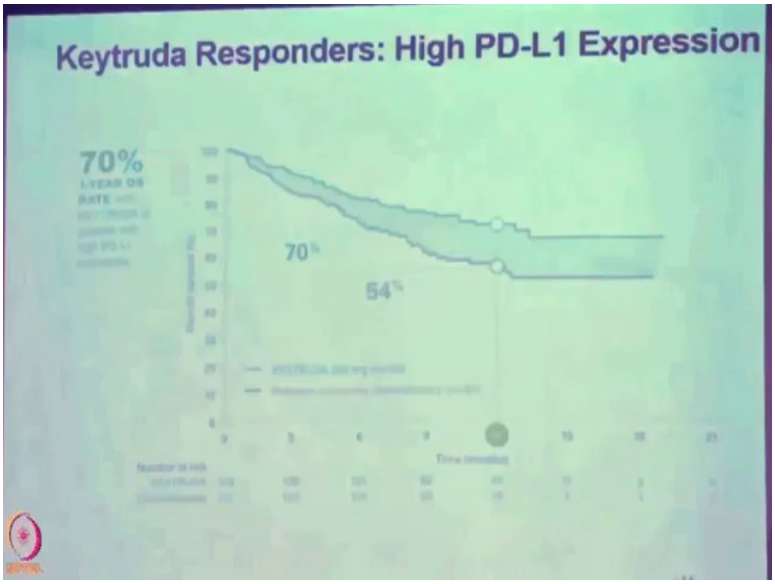
So, what the best way of the treatment? Give the dual therapy targeted therapy to block this PD-L 1, right, to block this PD-L 1 and give the targeted therapy. It will help the patient to recover fast, right because at that time T-cell get activated and kill all the cancer environment.

(Refer Slide Time: 29:18)



So, these are the multiple paper where people use anti PD-L 1 for the treatment of the different cancer type.

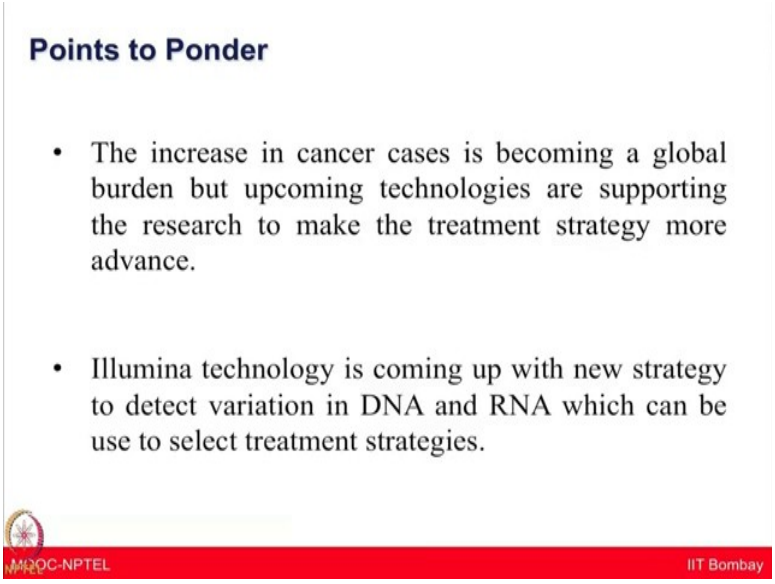
(Refer Slide Time: 29:41)



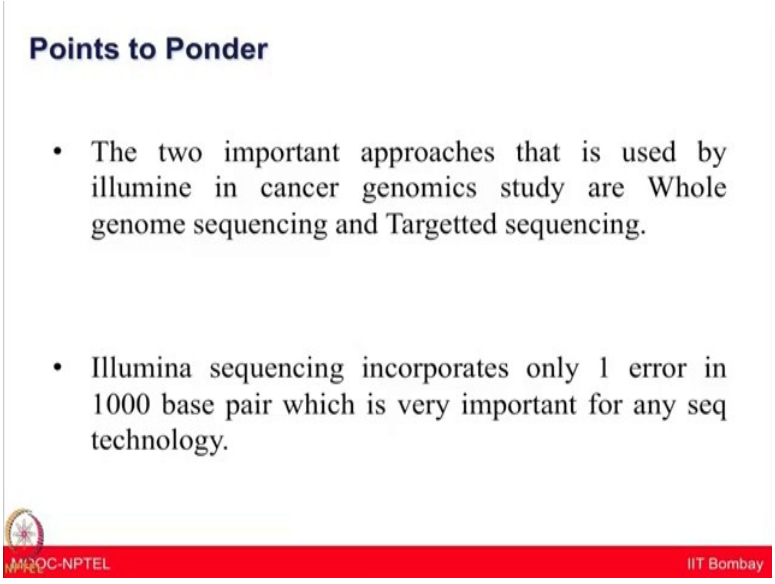
So, if you see the example this is the inhibitor of the PD-L 1. If the tissue have the high PD-L 1 expression, if you give this drug, the response is basically better, it gets better it reached for the treatment purpose it reached to 70, 70 percent. So, that means, you by giving these type of immunotherapy also it is better for the treatment of the patient, right

So, I am stopping here. The key home message like targeted therapy basically helping the cancer patient for the for a treatment and we are moving towards the companion diagnostic thing, right. So, there is a one panel which going to be the say this is kind of mutation you have and this is going to be targeted therapy or targeted drug going to be available and drugs are numbers are increasing, may be after 5 year you are going to have multiplied drugs, right. So, I think these things are changing with the targeted therapy and it is going so fast.

(Refer Slide Time: 30:40)



(Refer Slide Time: 30:53)



In today's lecture, Dr. Mukesh Jaiswal talked about different mutations and variations in the gene which need to be diagnosed primarily and accurately to deal with cancer. We also heard how a specific industry technologies, such as Illumina platform is going to provide new strategy to detect variations in DNA and RNA which could be used to select treatment strategies. He also talked about two important approaches that is used by Illumina in cancer genomic studies, first whole genome sequencing and second targeted sequencing.

Thank you.