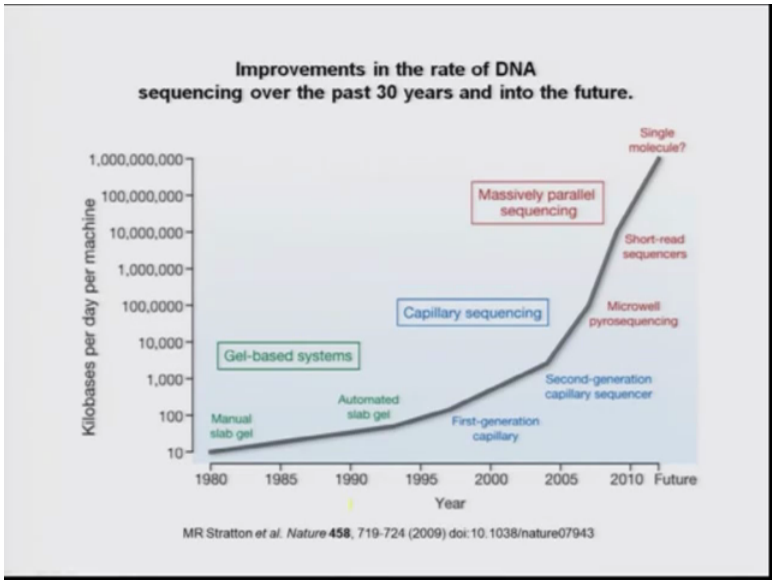
Functional Genomics
Professor S Ganesh
Department of Biological Sciences & Bioengineering
Indian Institute of Technology Kanpur
Lecture No 11
DNA Sequencing Methods Part 2

So welcome back to this course on functional genomics.

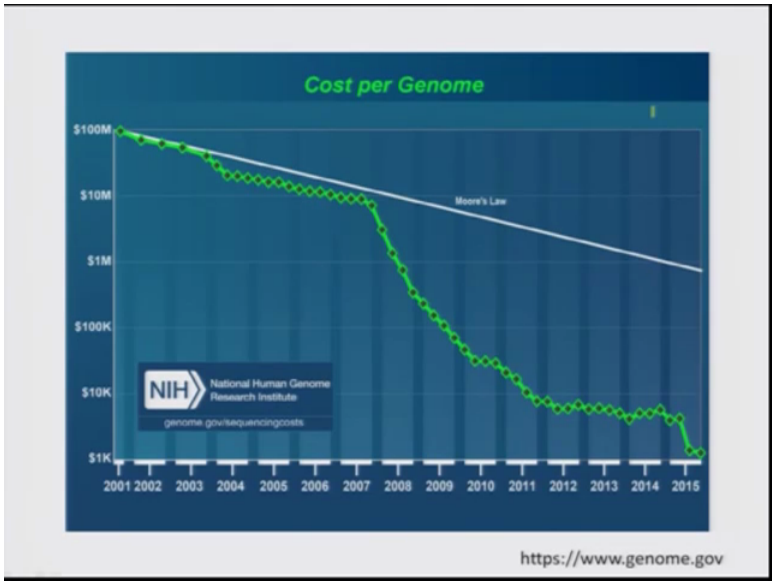
(Refer Slide Time: 0:21)



If you recall that we discussed human genome project when it was initiated in 1990, it was majorly funded by the US government not many other countries could join because of the cost involving sequencing. And also the infrastructure required and how the technology required, so you can see that if if you look into the you know the capability kilobases per day, per machine kind of calculation if you do.

You know you can see there in 1990 when people were using manual slab gel to automated slab gel to what you have capillary sequencing it has you can see that it has really increased the speed at which we are able to sequence. But with what is called as massively parallel sequencing or next generation sequencing the speed at which you are able to sequence the DNA has tremendously increased. Now we are looking at almost probably in the future is single molecule just (())(1:22) one DNA and then look at what are this you know transcript that are made so one can do and go on looking to that.

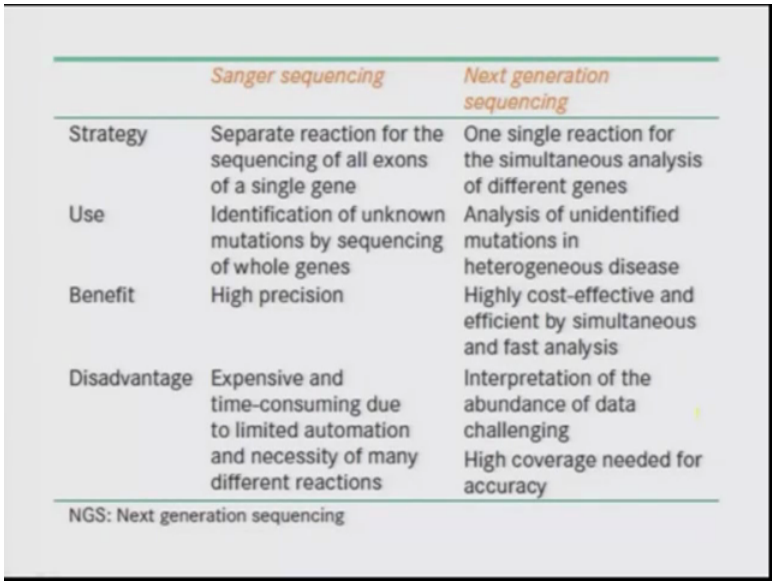
(Refer Slide Time: 1:36)



At the same time there is another important element that really helped us is the cost per genome has dramatically fallen. Now we can see that when the genome sequence sequence was at peak when the draft sequence was about to be published you can see that cost per genome for example human genome is about and you can see that is 100 million dollars to sequence one human genome.

But now it has come almost like you know 1000 dollars you give you will be able to sequence the human genome. So that is one of the greatest benefits of the projects because now you know one has to sequence the genome to understand what changes you have that may give you risk or of developing some disease then it is doable because it is 1000 dollars may be with time it come down to few thousand rupees, so that is the tremendous achievement. So that is that is something which is unexpected and which did not even follow what is Mores law which is the Intel co founder you know said that this kind of technology is exponential growth but this is beyond that exponential growth is really given beaten that Law.

(Refer Slide Time: 2:44)



	Sanger sequencing	Next generation sequencing
Strategy	Separate reaction for the sequencing of all exons of a single gene	One single reaction for the simultaneous analysis of different genes
Use	Identification of unknown mutations by sequencing of whole genes	Analysis of unidentified mutations in heterogeneous disease
Benefit	High precision	Highly cost-effective and efficient by simultaneous and fast analysis
Disadvantage	Expensive and time-consuming due to limited automation and necessity of many different reactions	Interpretation of the abundance of data challenging High coverage needed for accuracy

NGS: Next generation sequencing

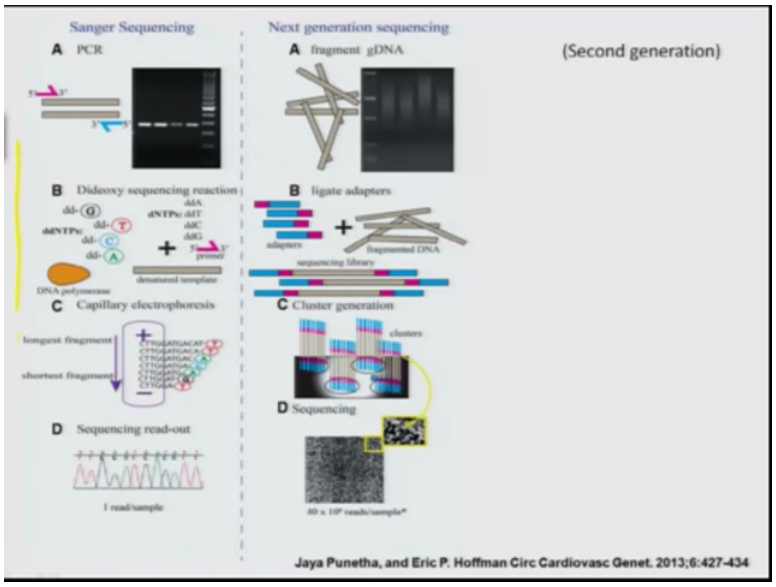
So that is the power of the what is called as next generation sequencing. Let us see what is the difference between the Sanger sequencing that we discussed and the next generation sequencing that we are talking about, so first thing is that strategy. So if you see that Sanger sequencing you have to separate reaction for sequencing of all exons of a single gene. So you have to for example individual exon should be taken out and put in a tube to sequence the small segment of the genome.

So here in the next generation sequencing one reaction for the entire genome as many you know segments can be sequence that is the huge difference. So you do not really do independent reactions. So what is the use major use of Sanger sequencing, so if you are looking at the particular segment of the gene, then identification of the unknown mutation by sequencing you know that is the use.

Then this next generation gives you much better way that is for example analysis of unidentified mutations in heterogeneous disease, right you do not know what it is but you really want to look at everything so you will be able to get some changes that possibly lead to the disorder. The benefit is it is the Sanger sequencing is still better in terms of precession because the accuracy which liable to tell you change the DNA is very high, whereas in next generation sequencing it is highly cost effective you do not spend that much money.

But it is efficient because is very fast you can finish it very quickly that is the huge benefit. What is the disadvantage of Sanger it is expensive and takes lot of time and then limited automation many are still requires manual in approach, whereas in next generation sequencing the interpretation of the data that becomes challenging because it really requires (())(4:35) because it produces huge amount of data and at times the accuracy may not be that that high, so high coverage needed for accuracy. So when I say accuracy high coverage a small segment if you want to understand what changes are there that needs to be sequenced at least 50 times. So that is that is where you come in accuracy.

(Refer Slide Time: 5:03)



Let us see how it is done, so this is the Sanger method so basically we describe already that you know these people use PCR fragments, you have primers with all the you know four bases with one of the dideoxy base and then you separate in a capillary and then you read out the sample depending on which florescent moiety is being released, right at the end of the capillary what comes out what florescent it gives that is how you read the sequence.

Whereas in next generation sequencing, so what you do is you take the old DNA and then you fragment them. So that is what shown here, so here you have a single fragment so that is multiple copies of that particular region of the DNA. Here you have you know fragments of you know certain range as a desired for a given chemistry but these are heterogeneous they them the

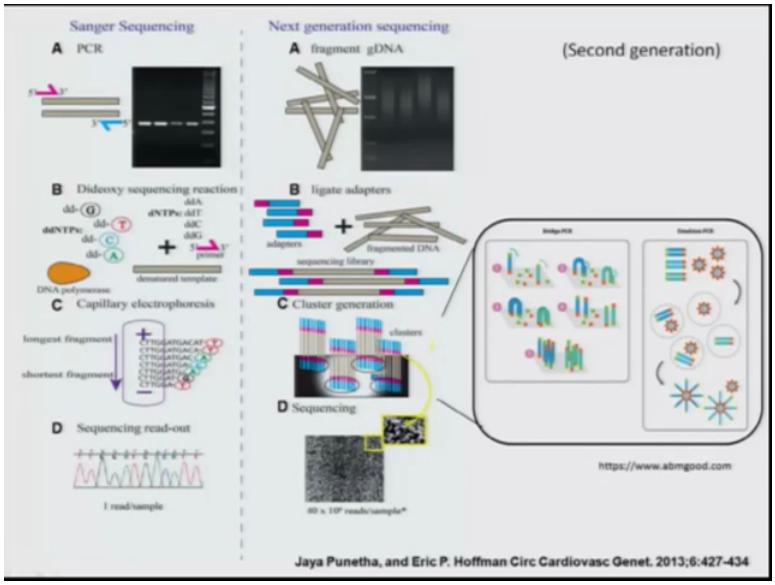
percent they particular you know genome it represent the entire genome if you are looking at the entire genome.

So what you do is that then you have what is called as adaptors, so which are added at either end of each one of the fragments, right and then to create what is called as a library. So here every fragment that is generated from a given species in terms of DNA is legated, joined with the adaptors that are shown here. So these are identical sequence added another side of the DNA to make what is called as a library.

Now these libraries are used for what is called as cluster generations, so basically what you do is that you have a microfluidics kind of a chamber in which the individual DNA molecule get in and then because it has got the adaptors another side you have regions that are complimentary to that, one or few of the molecules will make a complimentary sequence and then this you do a very similar kind of approach where like in a DNA polymerase is given so it makes multiple copies that we will see later.

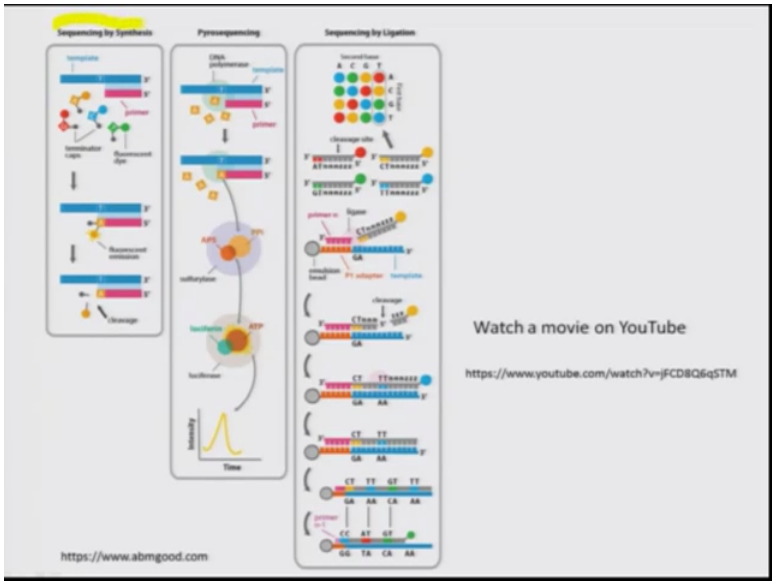
Once you make multiple clusters meaning each cluster now represent several copies of the identical you know sequence, exactly like what is used here but everything is done by automation mode here. So this cluster is having identical copies, right and then you are going to sequence each cluster and then read the sequence, then now that queues you the sequence of all the fragment that you are able to capture in this method. So that you know in one goal we are going to look into large number of samples.

(Refer Slide Time: 7:33)



So the question is how do you really you know develop this clusters, there are different methods we will see one of them so basically what you have is, you have solid devices in which you have these sequences that are complimentary to the fragments that are having these adaptors. So the sequence are complimentary to the adaptors, so when you send these you know fragments that are having the complimentary adaptors into the devices the fluidic devices then they will go and anneal to these sequences because of hydrogen bonding and then people add for example all the four bases and DNA polymerase then it would extend you know it will make copies of it and then we are able to create clusters like that which I will be able to explain you know movie that I am going to show the next.

(Refer Slide Time: 8:52)



So once you make these clusters, right so as I told you these clusters represent DNA copies, this is almost like a cloning, so it is like a PCR, so one copy now you have several copies of the same sequences and they are going to sequence over there. So the way you sequence it differs depending on which platform you are using. So the one is sequencing by synthesis, the other one is pyrosequencing and the third one is sequencing by ligation. So the three major methods people use in next generation sequencing, in these two first two methods what you basically use is the DNA polymerase obviously you have a primer, you have a DNA polymerase that goes and extend and there are bases that are having different modification for scoring whether it is which base it is.

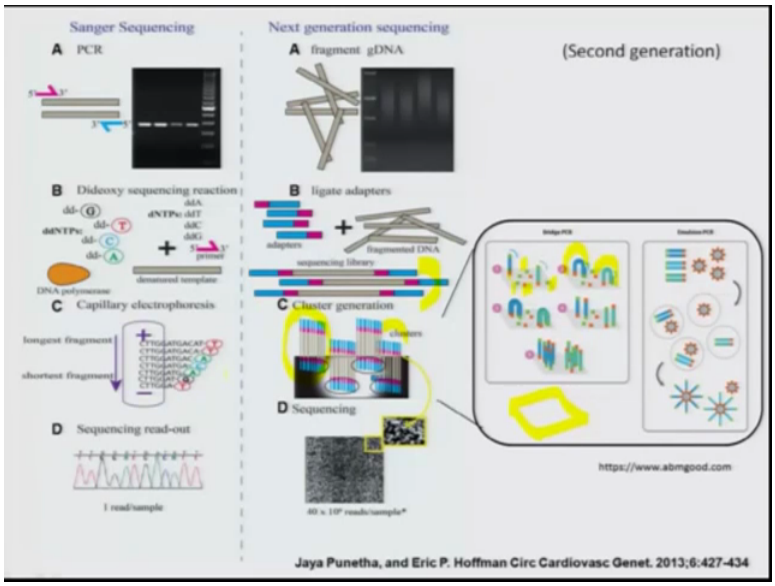
Whereas sequence ligation does not involve any DNA polymerase, let us look into these two. So in the first one the sequence by synthesis what happens is that you have a primer that binds to you know these primers are nothing but exactly the same you know adaptor sequence that is present in all the DNA of the library. So when ones it goes and binds, then what you do is the difference between this and Sanger method is here the bases that are used for calling the base whether it is A, T, G or C, the way it is done is very different in in in this chemistry.

So you have a base which has complimentary say for example this is the first base after the primer the polymerase has to find the base that is complimentary to Ts, so obviously it will go for A. And when it goes this base is all invariably having the terminator, right that is the terminator

cab. So at a time only one base can be added, so once it is added and then you all the reaction in the cluster will add one base and it will stop, ok. And then you wash out all the bases now, so no other base is left and then you excite them with a laser and then see what florescence it comes out, if it is A is going to give a particular fluorochrome and therefore the machine is able to read it as A.

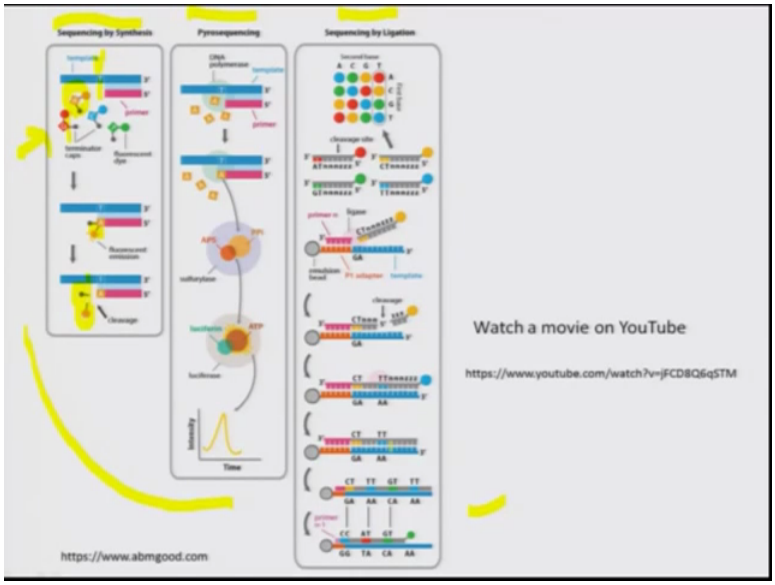
Once the reading is done what you do is you basically remove the terminator like what is florescence is taken after that you remove the terminator, we also remove the florescence moiety from the base then you wash it again it give all the four bases with the terminator and the florescence and then allow the reaction to go and add another base, again read it, again cleave it, again wash it, again go for this third bases. So this is the way you have do it.

(Refer Slide Time: 11:23)



So in a cluster you will find that all of them, all the molecule that are present in a cluster that we have shown here have identical sequence therefore from the base they will add one base. So all of them will be excited for a given fluorochrome, therefore you are again to get the florescence like this. So your machine would call it as A for the given dots, so this would be A, this would be B, C or whatever it is. And then you go on doing the reaction and you are able to complete it.

(Refer Slide Time: 11:47)



In Pyrosequencing it is depending on the reaction kind of a chemical reaction that happens here. So you know you have reaction in which you know whatever mitis that are added the you know the chemical that is released would give different color depending on the base, right. Again it is similar method what florescence comes out for a given base that would be read out. The difference is at a given time you are going to give only one particular base. For example you have the primer that is annealed to the template, so then you flush the system with only the base A and then see whether that got added, if it got added and when you do a reaction you will get a florescence. Suppose that is not A, then this will not get incorporated so you will not get any florescence.

So it could be G, so you will get a florescence, so you for everything you add one base and see whether florescence comes, wash it out give the other base wash it out, give the fourth base. So in this way you can you know identify at the same time you know whether a base is added or not added accordingly you can call the sequence. So this is what you know is done in this case. So this is the way they are able to determine the sequence. The Ligation method is very different you do not use any DNA polymerase, here you have oligonucleotides that have combination of various bases, the only difference is that the first two bases you know are fixed.

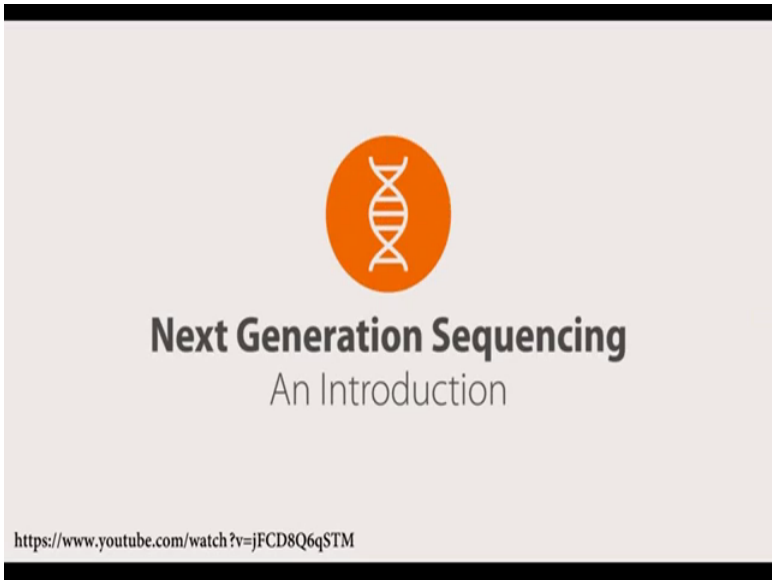
So you have either for example AT, CT, TT, GT, for example these are the bases like what is shown here, this is a first base, this is a second base. The rest of the bases could be random all

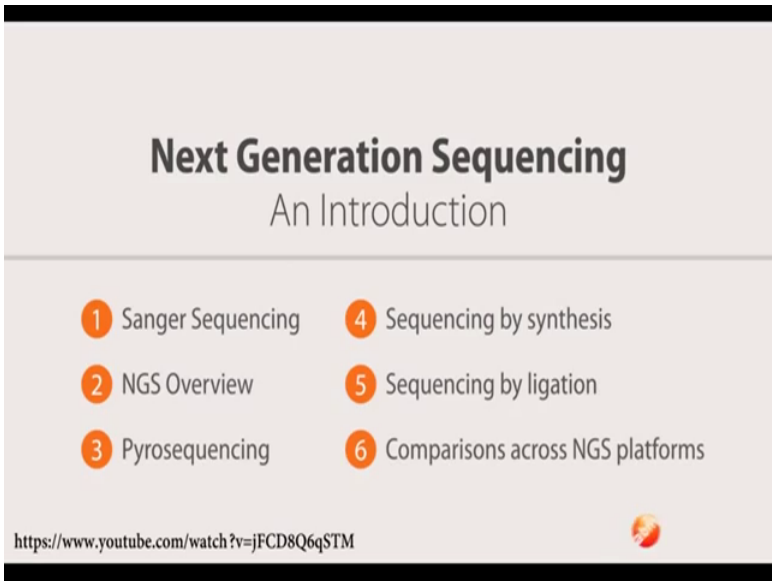
combination exist. So now it all depends on whether for example you have a primer that is aligned to your sequence which is single standard. Now since the primer ends here, you have to have a small oligonucleotide that is complimentary to the first two nucleotides of your template. If it is GA, so you should have this CT containing oligonucleotide that can come and anneal, once it anneals the Ligation reaction happens and this is sealed to make a larger fragment, right.

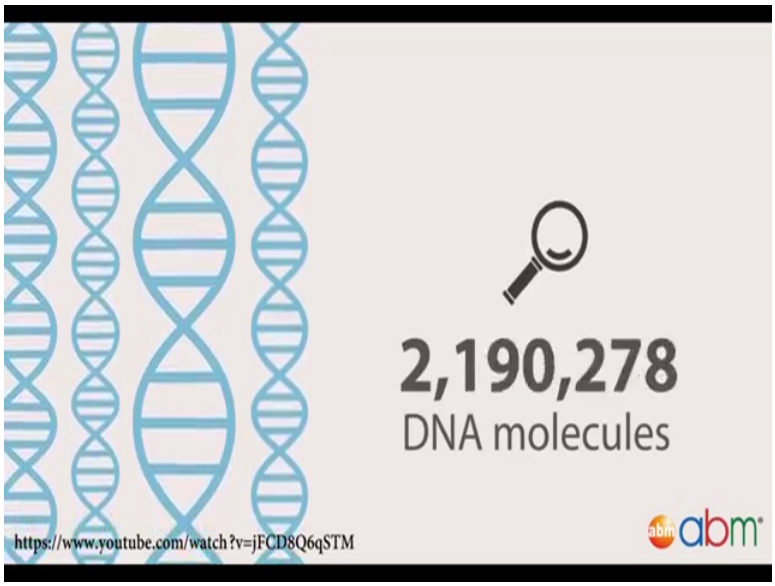
So then the system is able because this is annealed and this gives you different fluorochrome, therefore the system is able to call it as that this next two sequence after the primer is C and T. Then you have another reaction that goes on you know we remove this fluorochrome, then you have another base that comes in and again this would be calling for whatever bases that is present after the first you know oligonucleotide is annealed. So now in between you are not able to read but after three base you are able to again read the two bases of the DNA and go on it goes like this.

And then you have different primers that anneal at different segment, for example you can this is the primer if you consider the ensures it is 0, then it anneals at 1. You could have another sequences annealing at -1, or -2, or +1, or +2, therefore you are able to read the entire you know segments and get the you know sequence. So this is another method which is not very successful but these are the two most common chemistry people use for sequencing using next generations. So we are going to have video that is playing to you, movie that is come from the YouTube that I am going to show you now and this particular website you can go later as well and then watch the movie.

(Refer Slide Time: 15:29)




Next Generation Sequencing
An Introduction
1 Sanger Sequencing
2 NGS Overview
3 Pyrosequencing
4 Sequencing by synthesis
5 Sequencing by ligation
6 Comparisons across NGS platforms
https://www.youtube.com/watch?v=jFCD8Q6qSTM


2,190,278
DNA molecules
https://www.youtube.com/watch?v=jFCD8Q6qSTM
abm

NGS is revolutionizing fields,

https://www.youtube.com/watch?v=jFCD8Q6qSTM

Personalized medicine

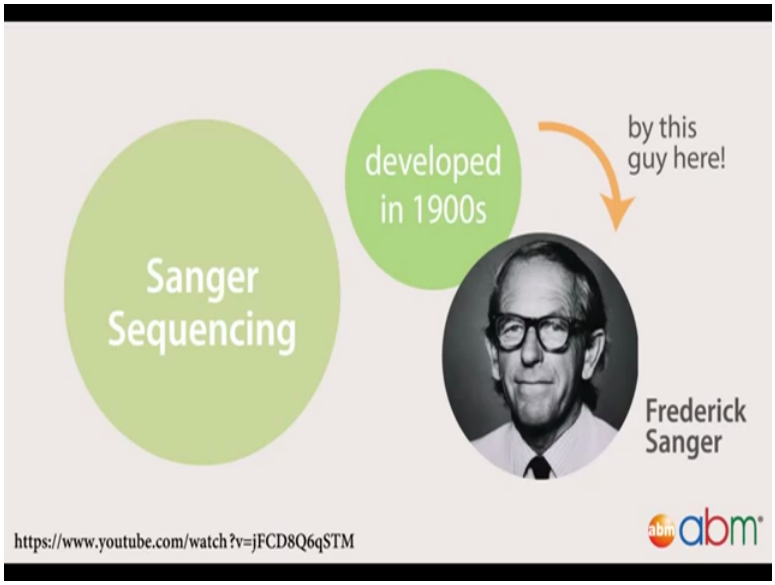
Genetic diseases

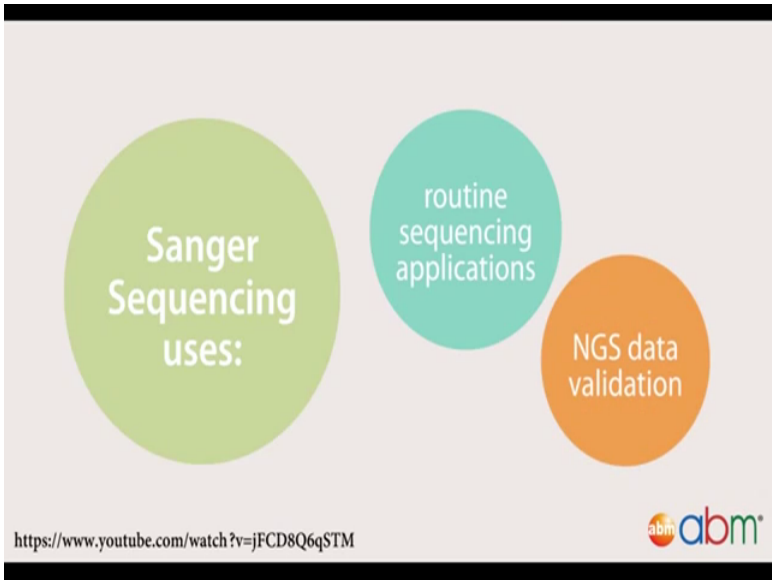
Clinical diagnostics

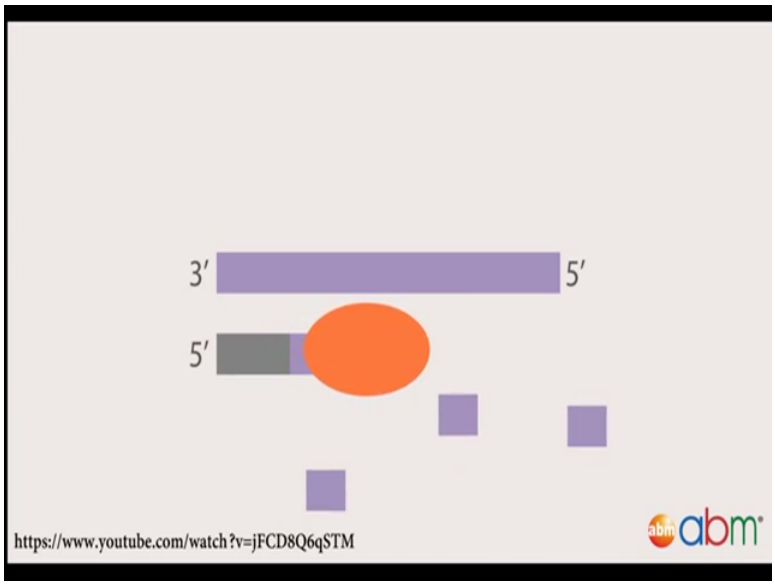
https://www.youtube.com/watch?v=jFCD8Q6qSTM

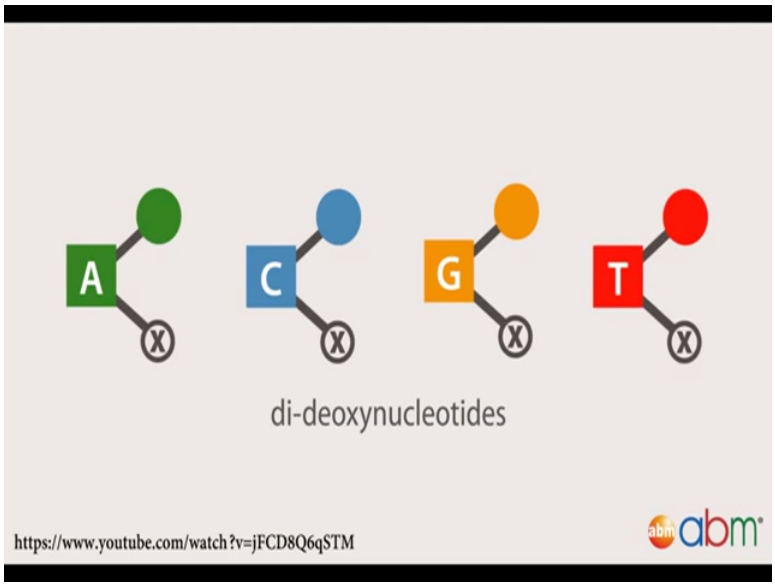
Next Generation Sequencing or NGRs is a powerful platform that is enable to sequencing of thousands to millions of DNA molecules simultaneously. This powerful tool is revolutionizing fields such as personalized medicine, genetic diseases and clinical diagnostics by offering a high (())(15:51) option with the capability to sequence multiple individuals at the same time.

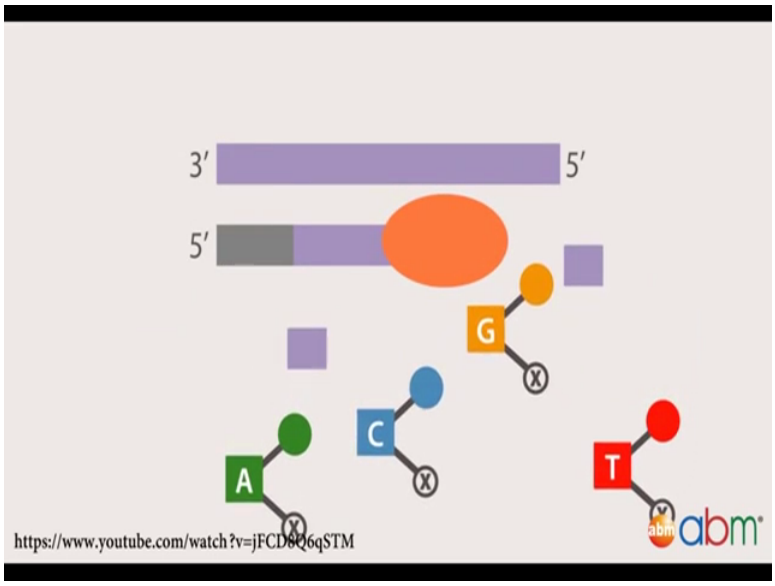
(Refer Slide Time: 15:57)

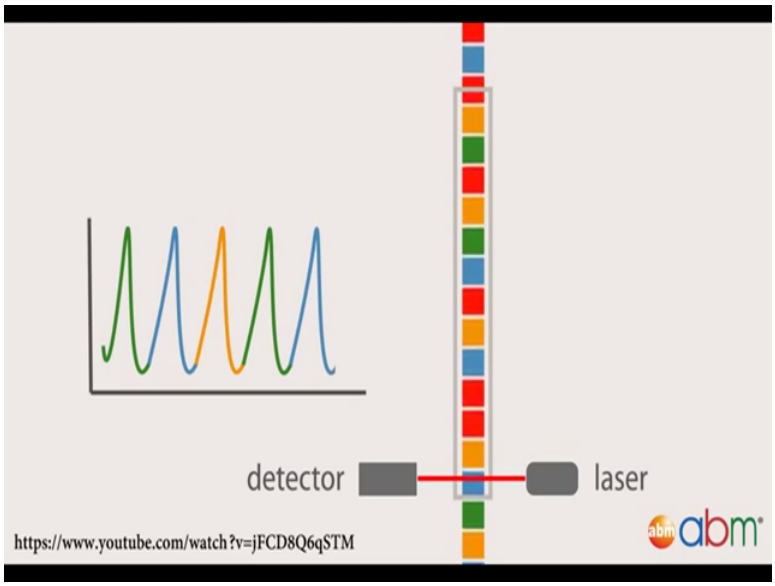


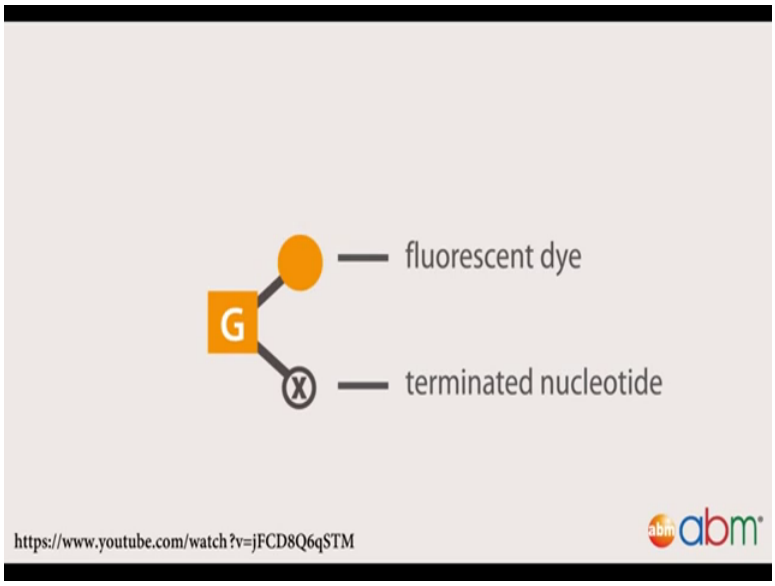


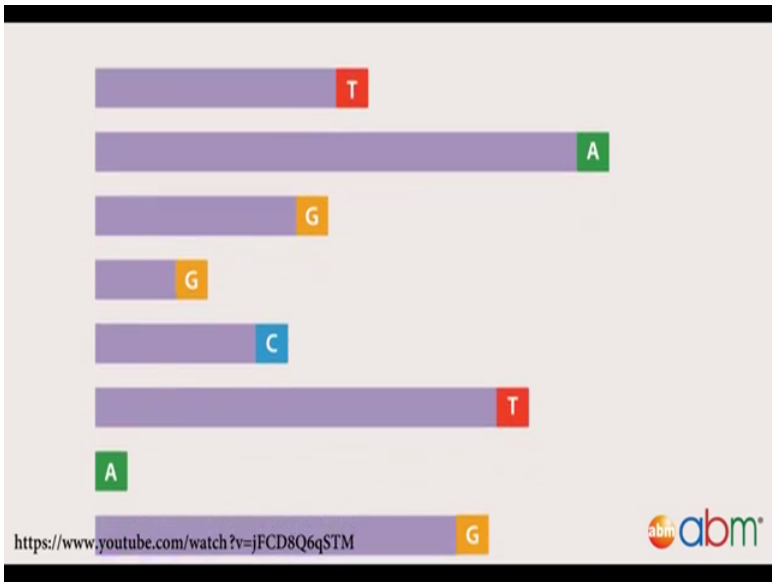

3′
5′
5′
https://www.youtube.com/watch?v=jFCD8Q6qSTM
abm

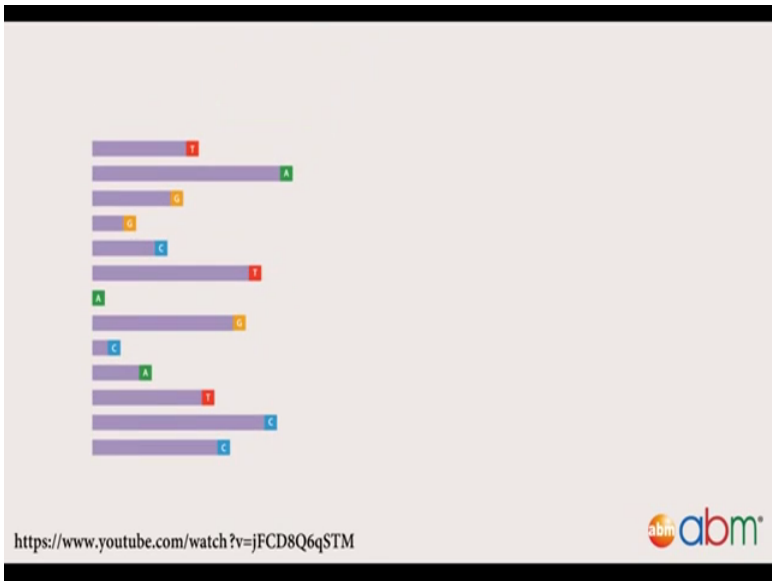

A
C
G
T
X
X
X
X
di-deoxynucleotides
https://www.youtube.com/watch?v=jFCD8Q6qSTM
abm

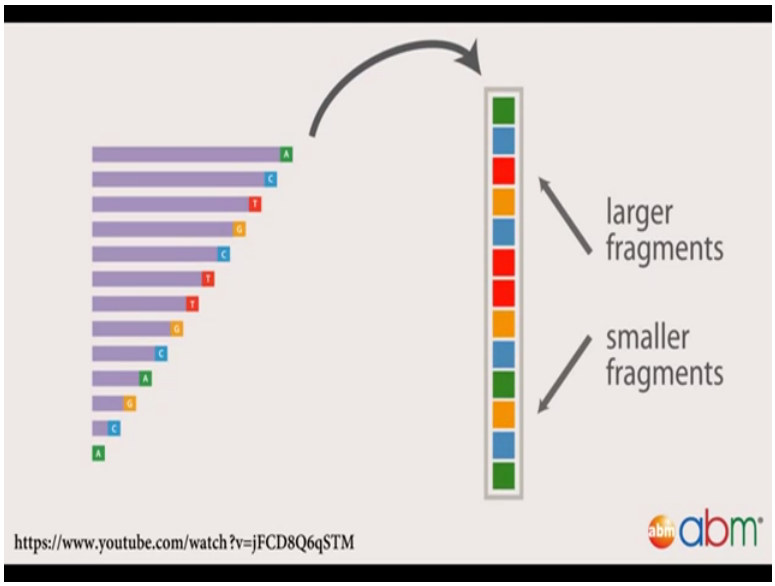

3′
5′
5′
G
C
A
T
https://www.youtube.com/watch?v=jFCD8Q6qSTM
abm


detector
laser
https://www.youtube.com/watch?v=jFCD8Q6qSTM
abm


fluorescent dye
G
X
terminated nucleotide
https://www.youtube.com/watch?v=jFCD8Q6qSTM
abm


T
A
G
G
C
T
A
G
https://www.youtube.com/watch?v=jFCD8Q6qSTM
abm





Sanger sequencing was developed in 1900s is a growth standard for DNA sequencing and it is still used today extensively for routine sequencing applications and to validate NGS data. Digitalize the high fidelity DNA to polymerase to generate a complimentary copy to a single standard DNA template. In 8 reaction a single primer complimentary to the template initiates a DNA synthesis reaction from (())(16:21) end.

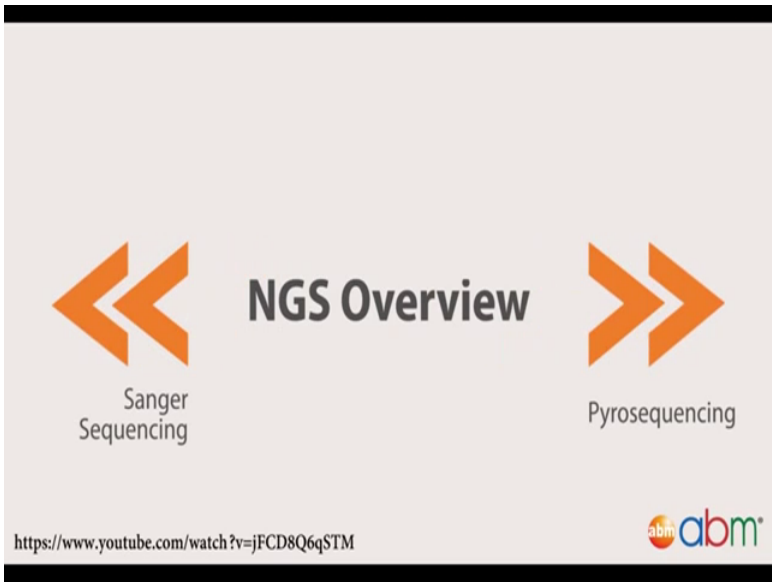
Dideoxy nucleotides are simply nucleotides are arranged one after the other in a template dependent manner. Each reaction also contains a mixture of four dideoxy nucleotides, one for each DNA base, these dideoxy nucleotides resemble the DNA Mormons enough to along

corporation to the growing strength however it differ from natural dideoxy nucleotides into race. One they like a (())(16:47) which is required for further DNA extension resulting in chain termination once incorporated in the DNA molecule.

And two each dideoxy nucleotide has a unique florescent dye attached to it, along for automatic detection of the DNA sequence. As a result many copies of different (())(17:03) DNA fragments are generate in each reaction terminate at all of the (())(17:08) of the template molecule by one of the dideoxy of nucleotides. The reaction mixtures are loaded on a sequencing machine is manually going to slab gels or automatically was capillaries and our lecture for is to separate the DNA molecules by size.

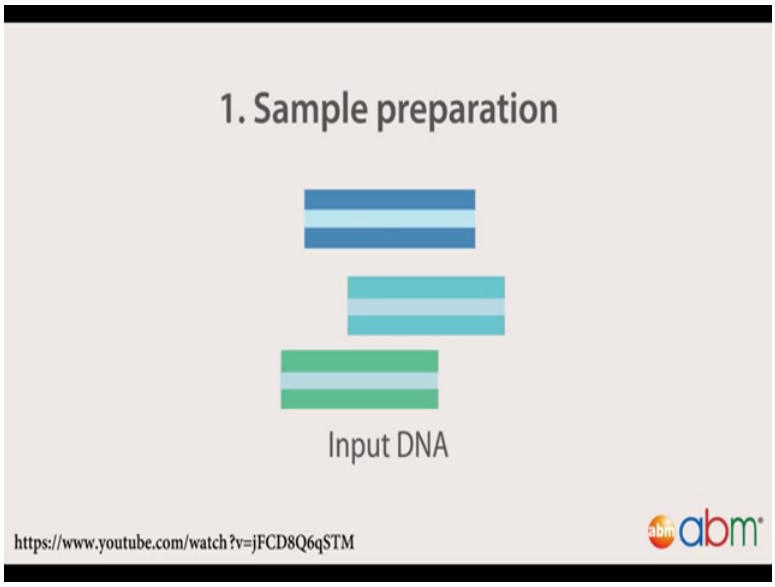
The DNA sequence is (())(17:24) the florescence emission of the dideoxy nucleotide as it flows through the gel. Modern day Sanger sequencing instruments is capillary based automated (()) (17:33) which typically analysis 8 to 96 sequencing reactions simultaneously.

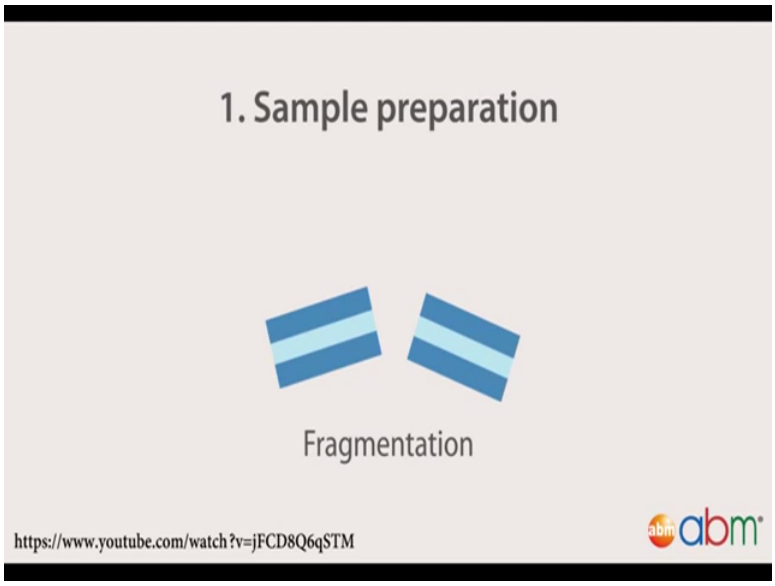
(Refer Slide Time: 17:47)

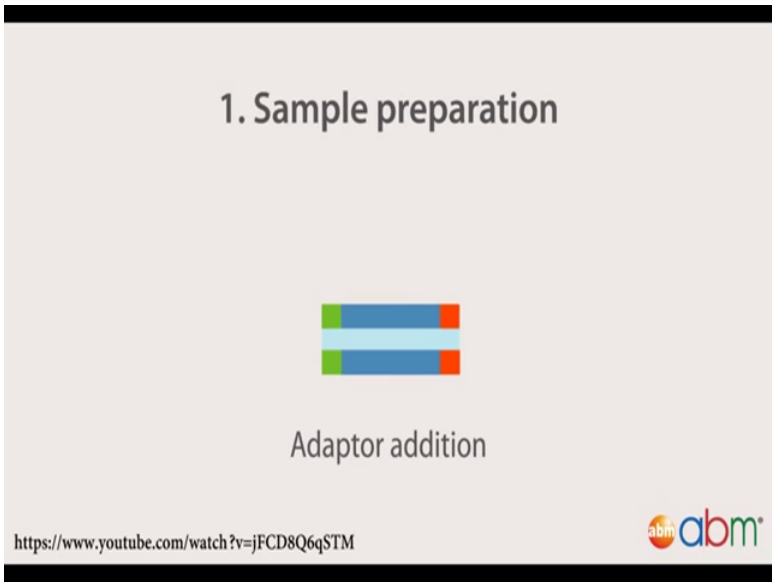


Next generation sequencing systems have been introduced in the past decade that allow for massively parallel sequencing reactions. These systems are capable of analyzing millions, even billions of sequencing reactions of the same time about different machines have been developed with various different technical details, they all shares in common features.

(Refer Slide Time: 18:06)

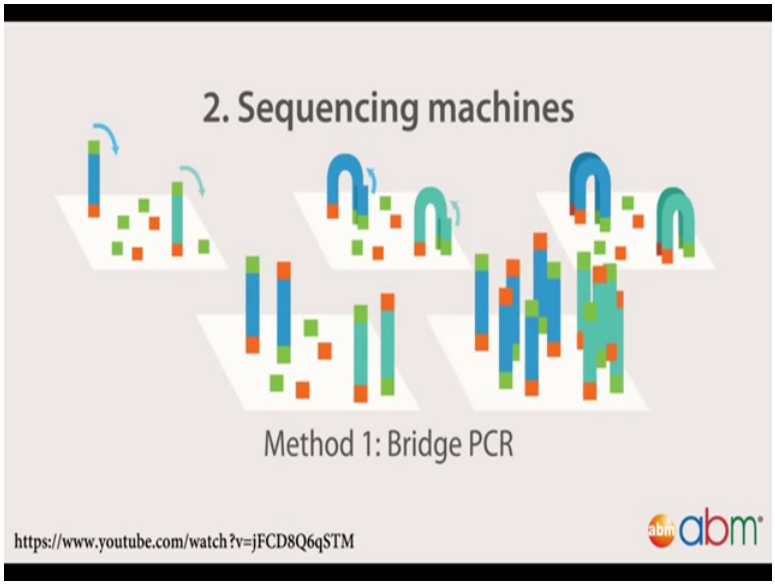


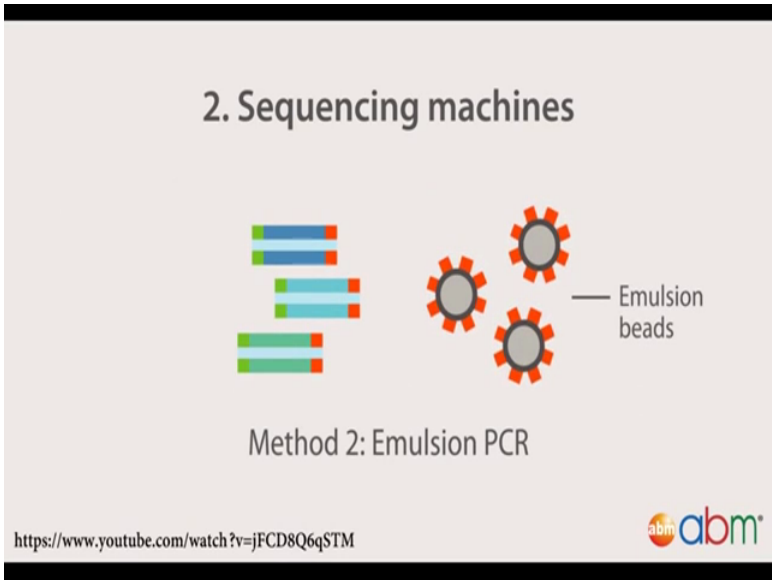


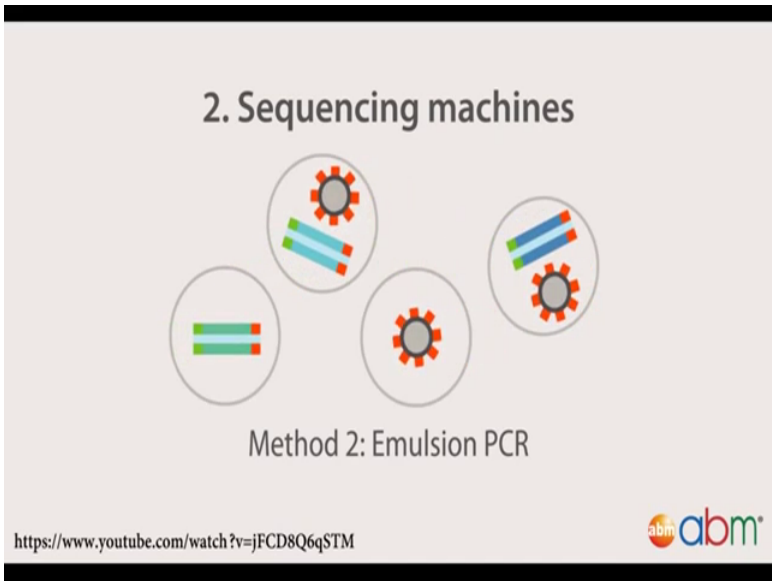


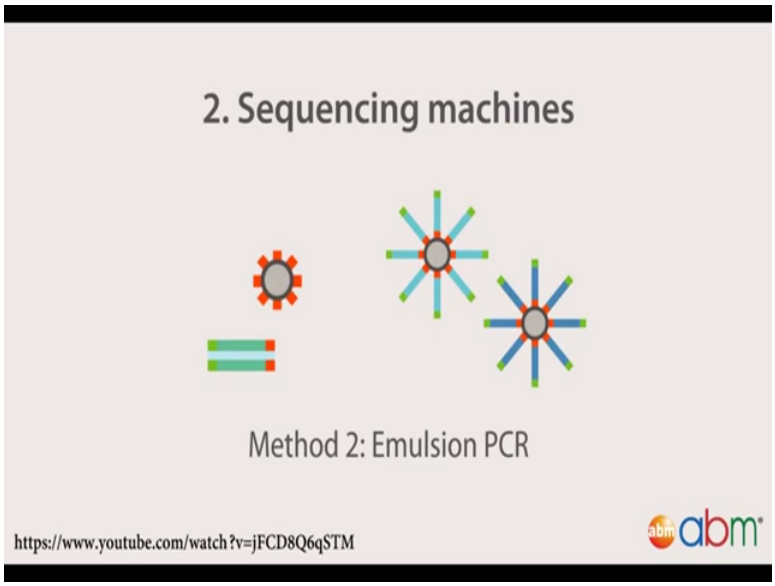
1, Simple preparation all next generation sequencing problems require a library of 10 either by amplification (())(18:11) custom adaptor sequences.

(Refer Slide Time: 18:16)



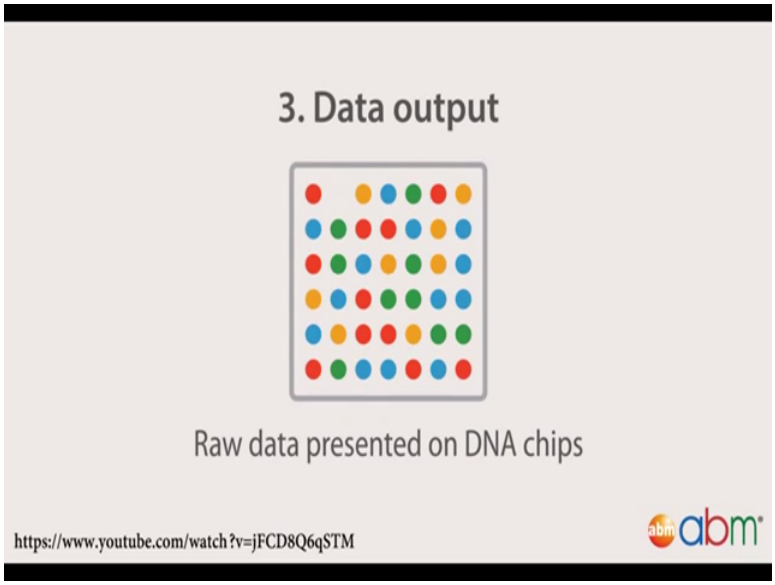


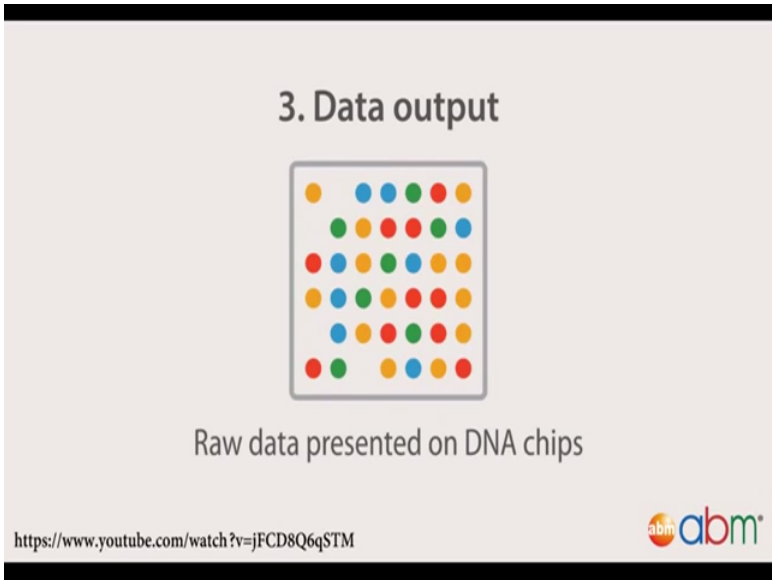




2. Sequencing machines, each library fragment is amplified on a south surface with (())(18:20) attached DNA linkers that (())(18:22) the library adaptors is amplification creates clusters of DNA, each originating from a single library fragment. Each cluster will act as a individual sequencing reaction.

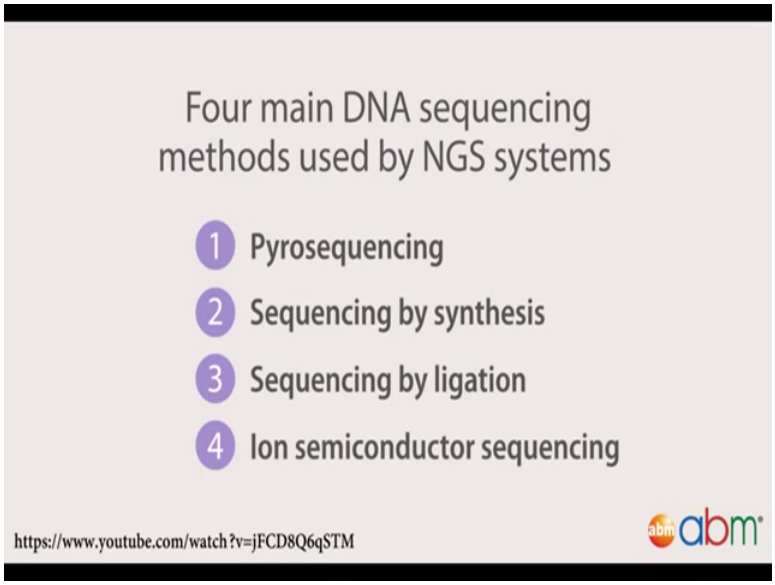
(Refer Slide Time: 18:36)





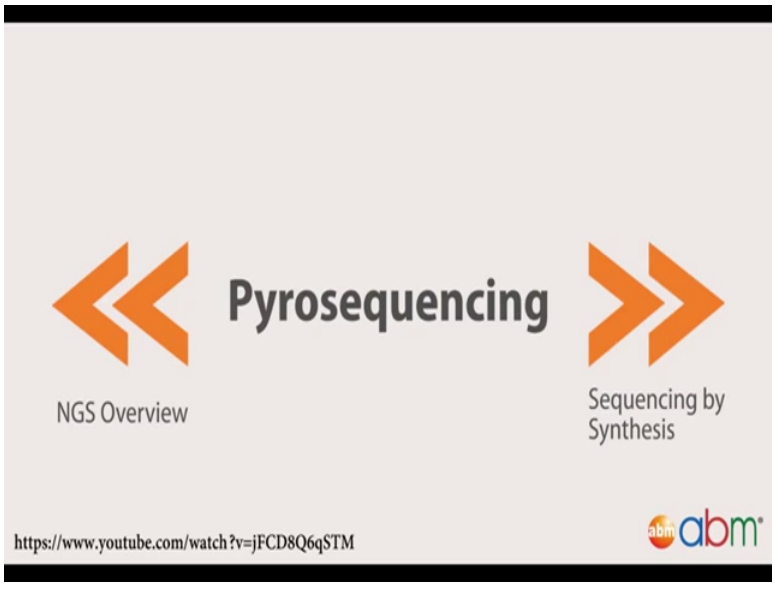
And 3, Data output, each machine provides a raw data at the end of the sequencing run, this raw data is a collection of DNA sequences that was generated at each cluster.

(Refer Slide Time: 18:51)

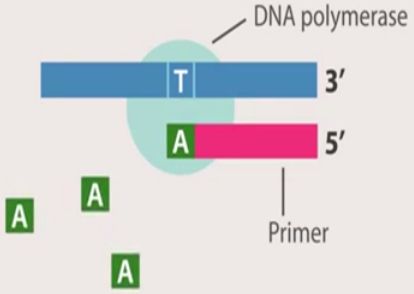


The differences between the different next generation sequencing problems by mainly in the technical details of the sequencing reaction and can be categorized in 4 groups Pyrosequencing, Sequencing by synthesis, Sequencing by Ligation and Ion semiconductor sequencing.

(Refer Slide Time: 19:10)



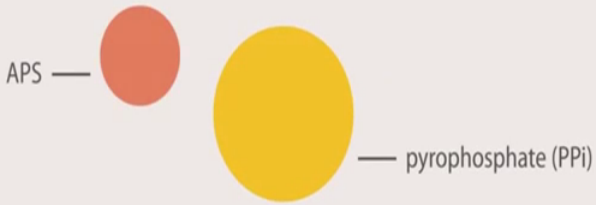
Pyrosequencing



https://www.youtube.com/watch?v=jFCD8Q6qSTM

abm

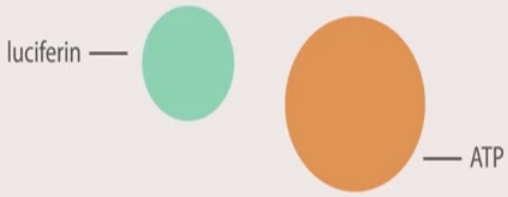
Pyrosequencing



https://www.youtube.com/watch?v=jFCD8Q6qSTM

abm

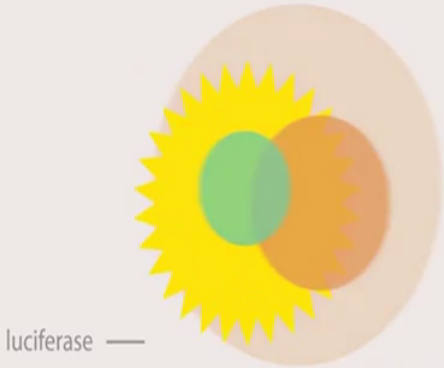
Pyrosequencing



https://www.youtube.com/watch?v=jFCD8Q6qSTM

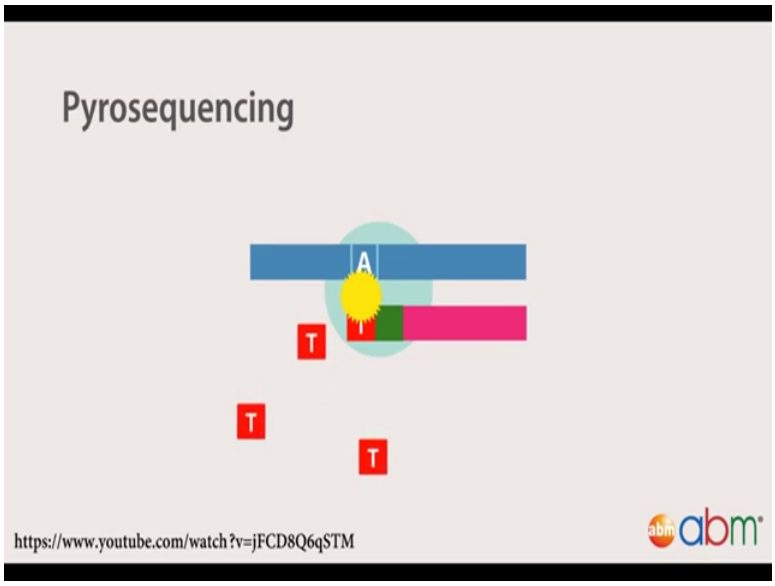
abm

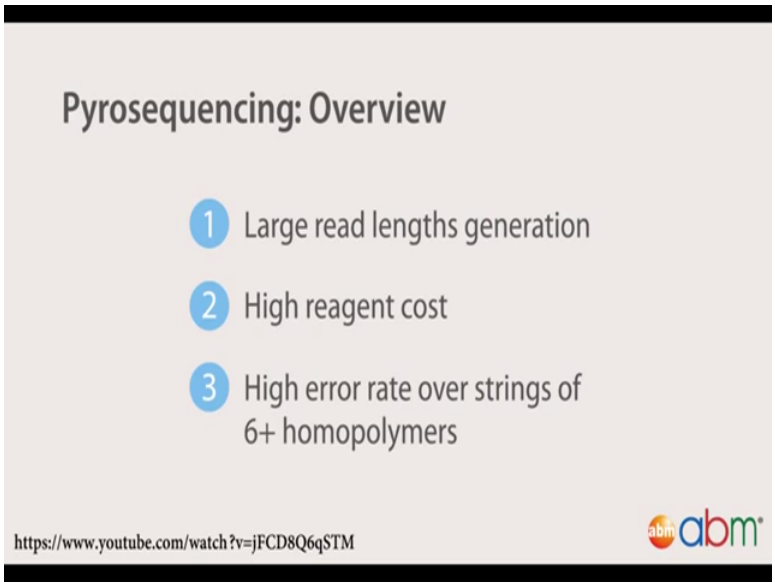
Pyrosequencing



https://www.youtube.com/watch?v=jFCD8Q6qSTM

abm



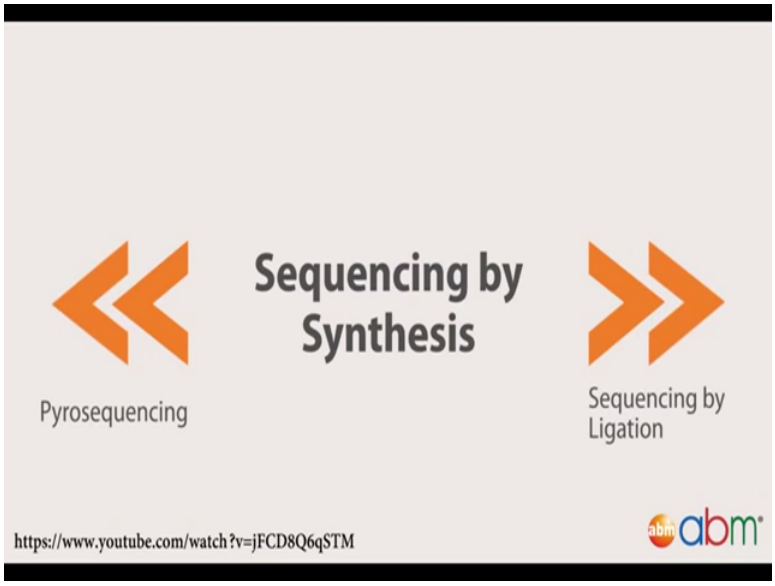


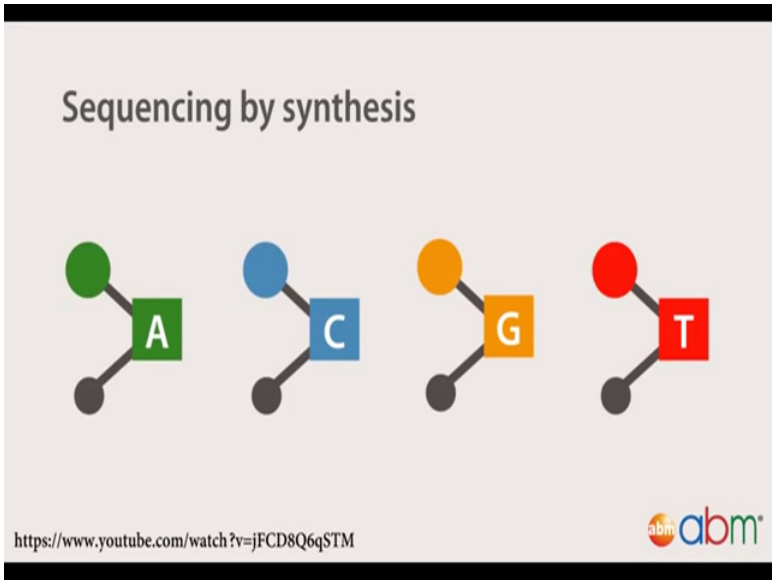
In Pyrosequencing, the sequencing reaction (())(19:09) released at the pyrophosphate doing each nucleotide incorporation, the release pyrophosphate is used in a series of chemical reactions resulting in the generation of light. Light emission is detected by a camera which requires the appropriate sequences of the cluster.

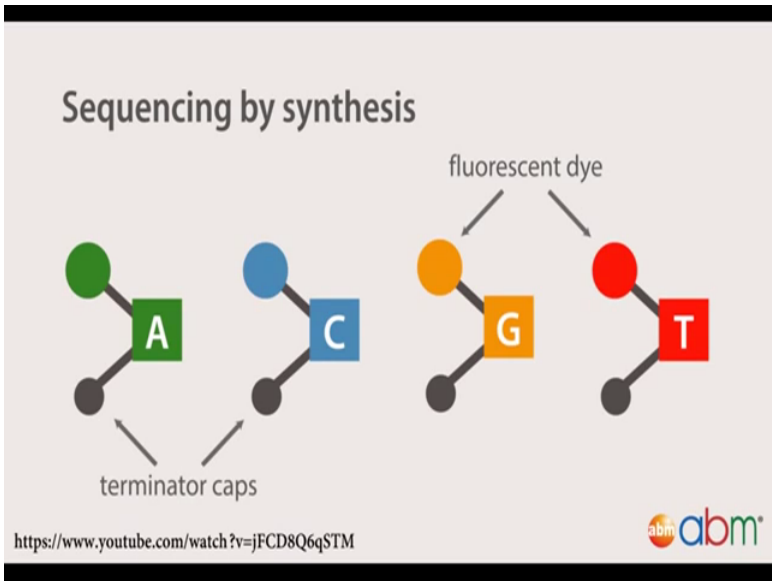
The sequencing proceeds by in keeping one base at a time, measuring the light emission (()) (19:29) degrading the unincorporated bases and then the addition of another base. This technology is capable of generating large read lengths, much comparable to the read length or Sanger sequencing, however high reagent cost and high error rate over strings of 6 or more

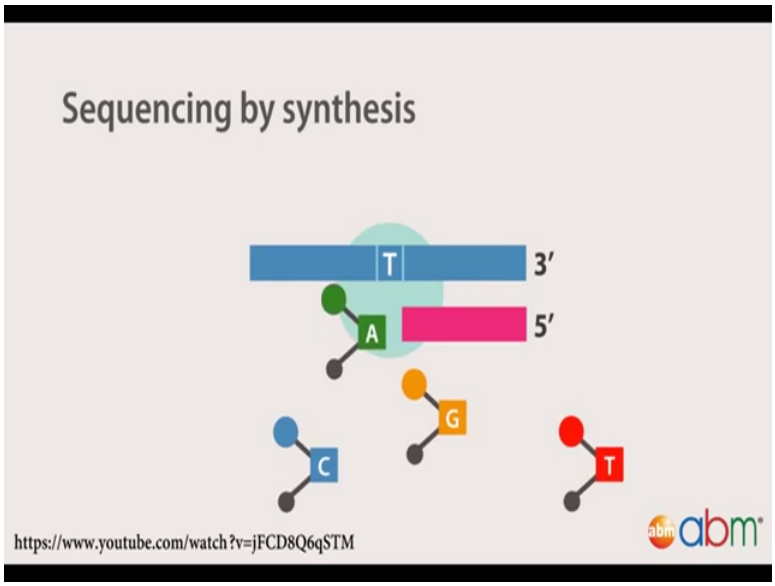
homopolymers have reduced this applications. For more details on the technical aspect of this technology please visit knowledge base at the link provided in the description below.

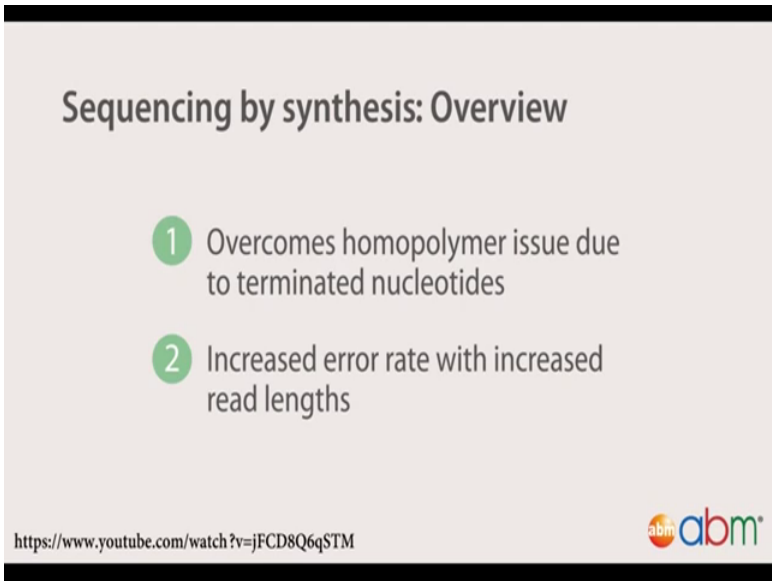
(Refer Slide Time: 19:59)






Sequencing by synthesis
fluorescent dye
A
C
G
T
terminator caps
https://www.youtube.com/watch?v=jFCD8Q6qSTM
abm

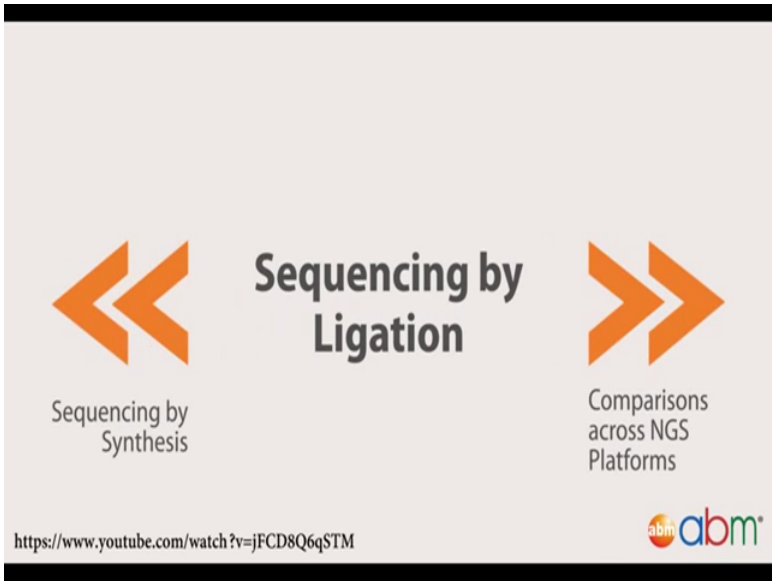

Sequencing by synthesis
T
3′
A
5′
G
C
T
https://www.youtube.com/watch?v=jFCD8Q6qSTM
abm

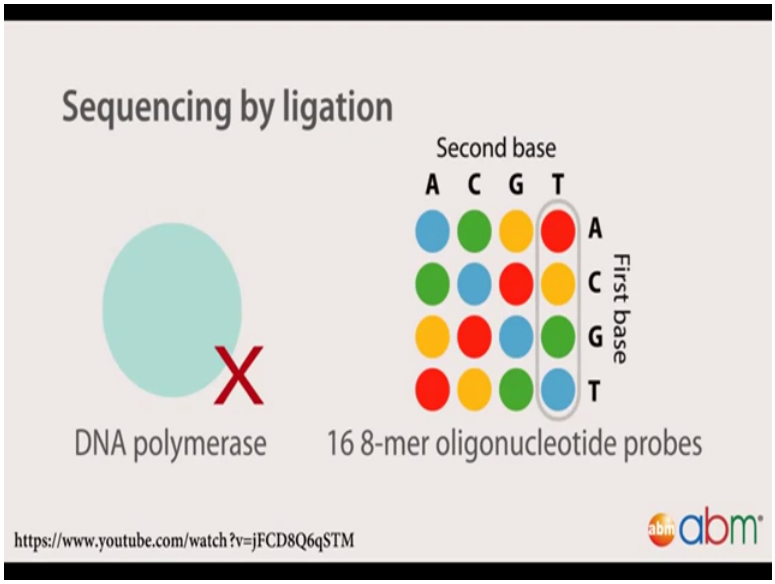


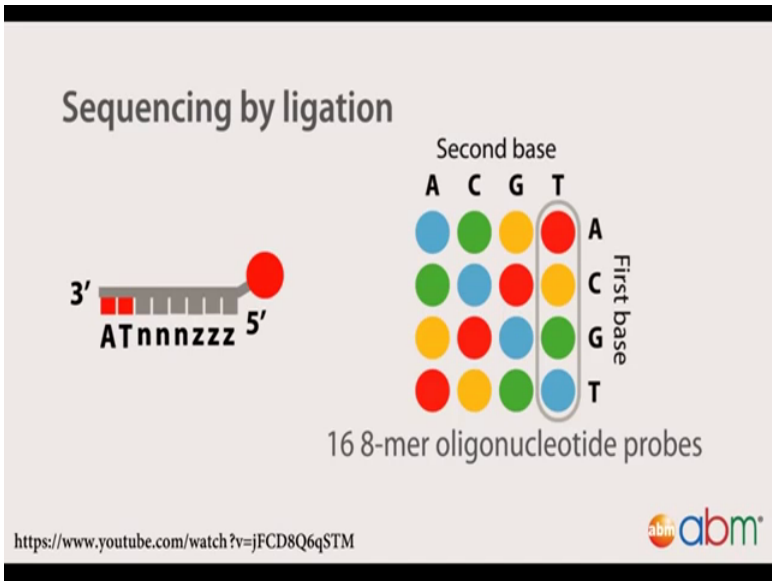
Sequencing by synthesis utilize the step by step of reversely florescent and terminator nucleotide. For DNA sequencing and is used by alumina and GS type. All four nucleotides are added to the sequencing chip at the same time and after nucleotide incorporation the remaining DNA bases are washed away. The florescent signal is read at each cluster and recorded. Both the florescent molecule and the (())(20:23) group are then cleaved and washed away. This process is repeated until the sequencing reaction is complete. This system is able to overcome the disadvantage of the pyrosequencing system by (())(20:35) and single nucleotide at a time.

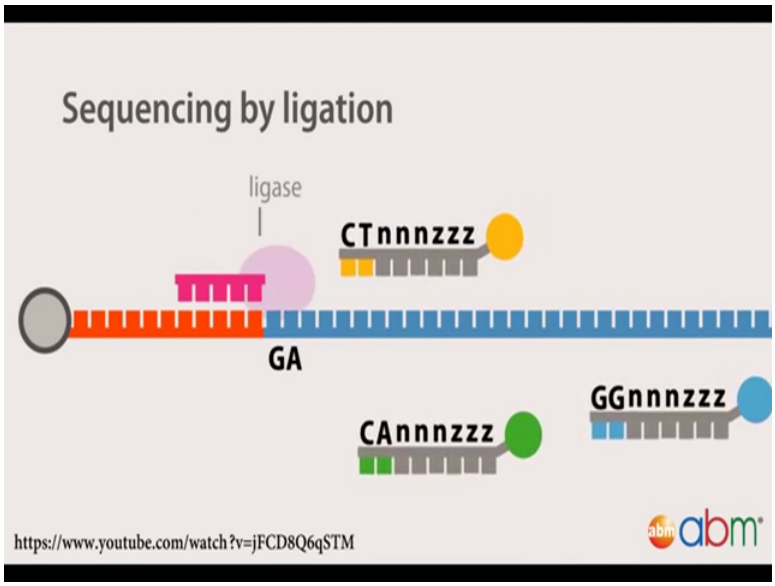
However as a sequencing reaction proceeds the error rate of the machine also increases. This is due to incomplete removal of the florescence signal, which leads the higher background noise bubbles. NGRs and introduction knowledge based provides more technical details about this technology.

(Refer Slide Time: 20:58)

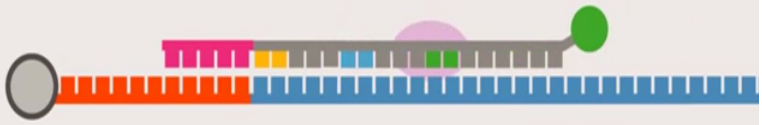





Sequencing by ligation
Second base
A C G T
A
C
G
T
First base
3′
ATnnnzzz 5′
16 8-mer oligonucleotide probes
https://www.youtube.com/watch?v=jFCD8Q6qSTM
abm


Sequencing by ligation
ligase
CTnnnzzz
GA
GGnnnzzz
CAnnnzzz
https://www.youtube.com/watch?v=jFCD8Q6qSTM
abm

Sequencing by ligation



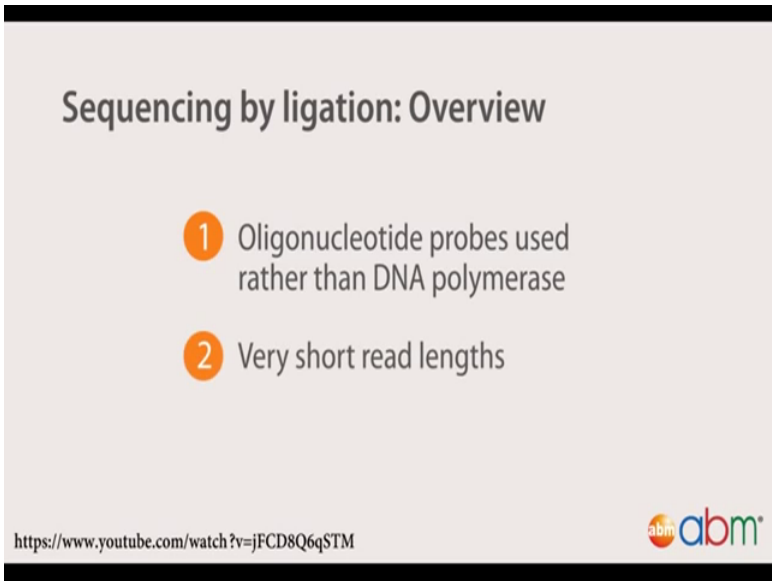
https://www.youtube.com/watch?v=jFCD8Q6qSTM

abm

Sequencing by ligation

https://www.youtube.com/watch?v=jFCD8Q6qSTM

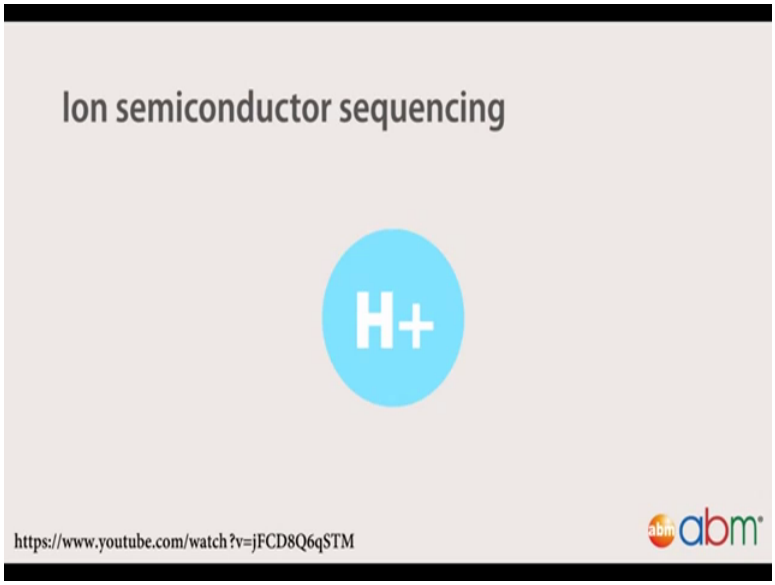
abm

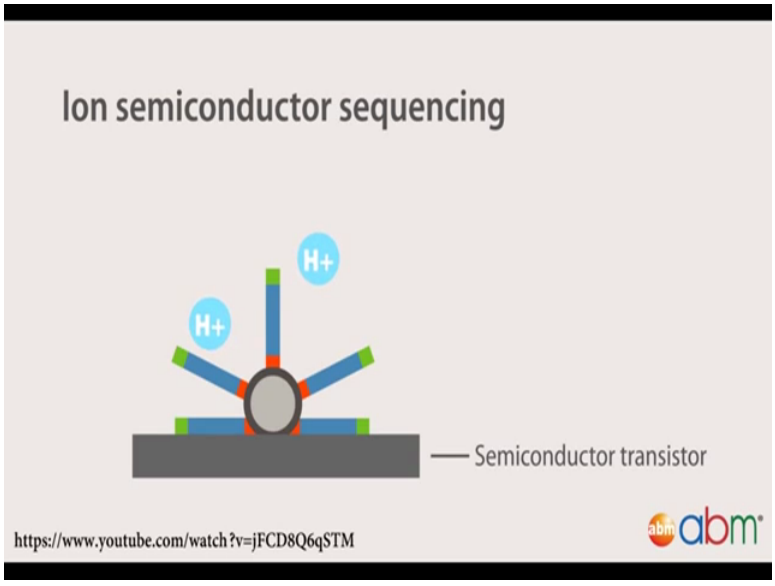


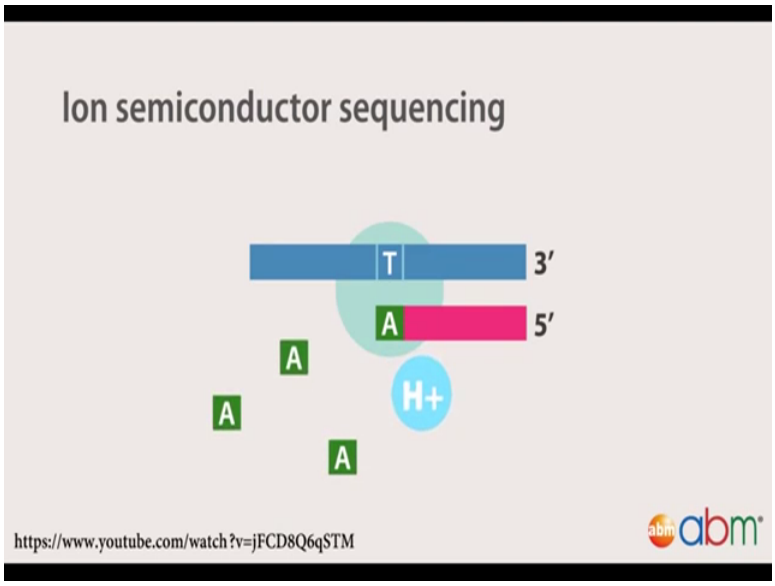
Sequencing by Ligation is different from the other two methods, since it is not utilized at DNA polymerase incorporate nucleotides instead it relies on 16 (())(21:06) oligonucleotide probes. Each with one of four florescent dyes attached to its five prime end but at ligated to one another. Each (())(21:14) consist of two probes specific bases and six to general bases. The sequencing reaction commences by binning at the primer to the adaptor sequence and then hepatisation of the appropriate probe.

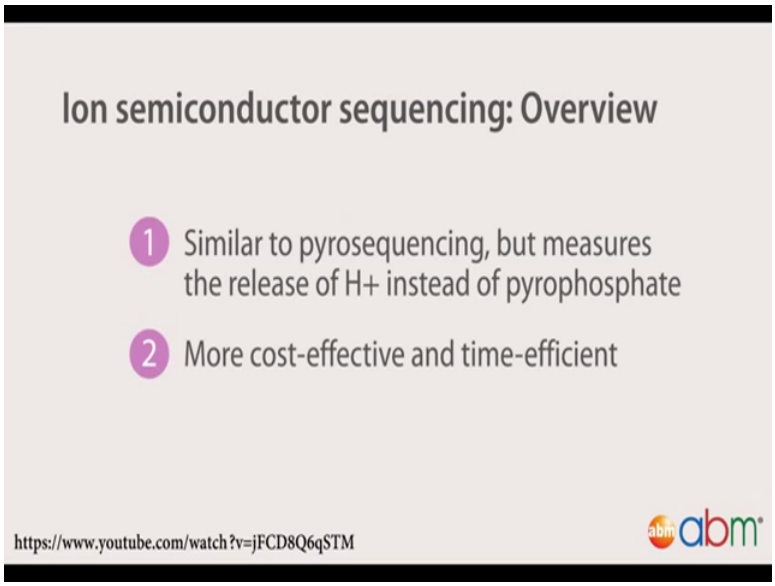
This hepatisation of the probe is (())(21:28) by the two probe specific bases and upon annealing is ligated to the primer sequence through a DNA ligase. Unbound oligonucleotides are washed away, then the signals detected and recorded. After that the florescent signal along with the last three bases the optimal probe are cleaved and then the next cycle commences after approximately seven cycles of ligation the DNA strand is denature and another sequencing primer are set by one base from the previous primer is used over (())(21:56) steps in total five sequencing primers are used. The major disadvantage of this technology is the very short sequencing reads generated.

(Refer Slide Time: 22:10)



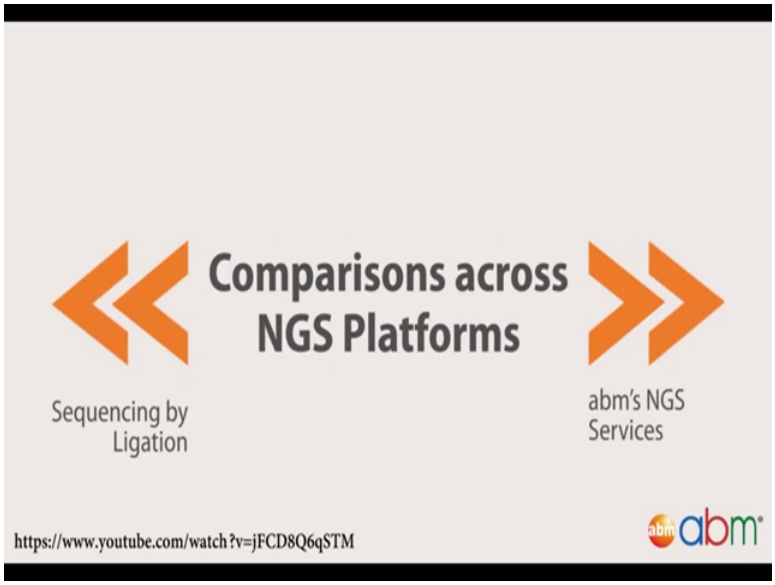


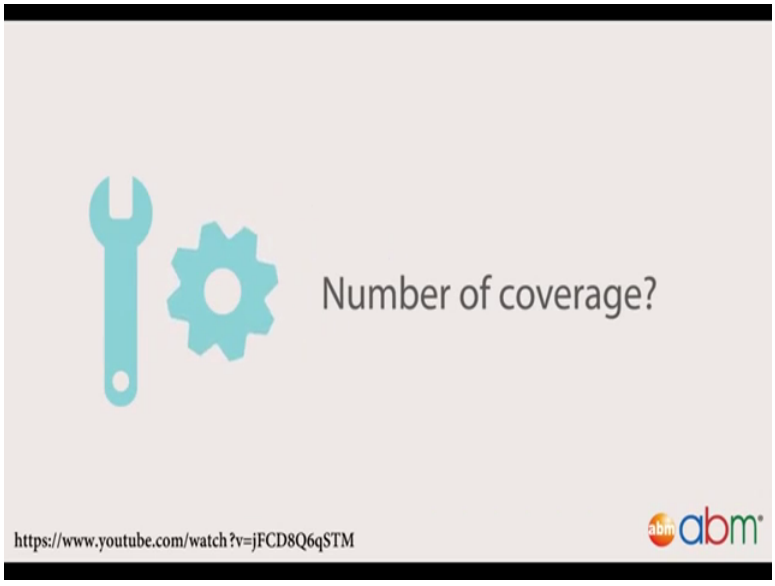


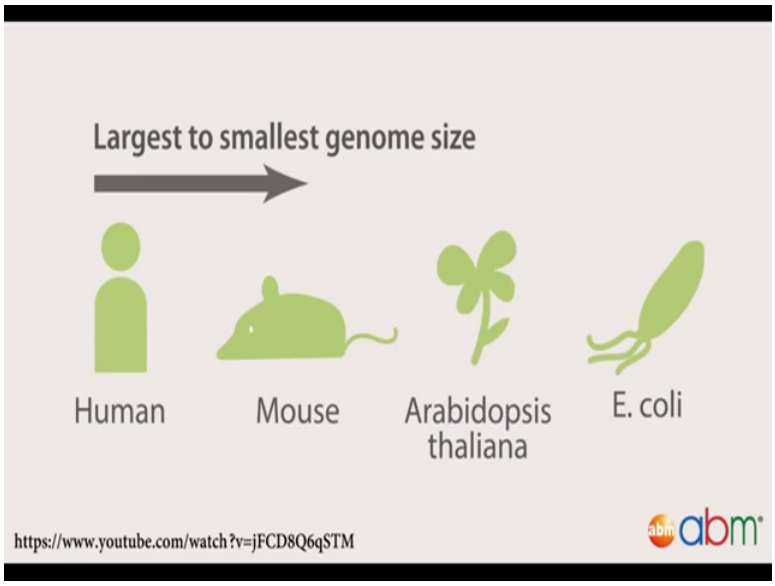


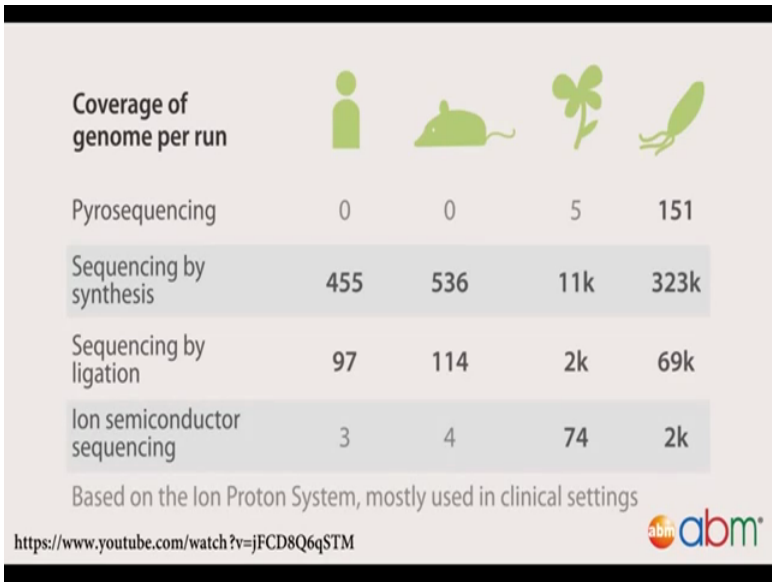
Ion semiconductor sequencing utilizes the release of hydrogen ions during the sequencing reaction to detect the sequence of a cluster. Each cluster is located directly above a semiconductor transistor which is capable of detecting changes in the PH of the solution. During nucleotide incorporation a single hydrogen ion is released into the solution and it is detected by the semiconductor. The sequencing reaction itself proceed similar to pyrosequencing but at fraction of the cost. Please use our knowledge base for further details on ion semiconductor sequencing and the sequencing by ligation techniques.

(Refer Slide Time: 22:45)





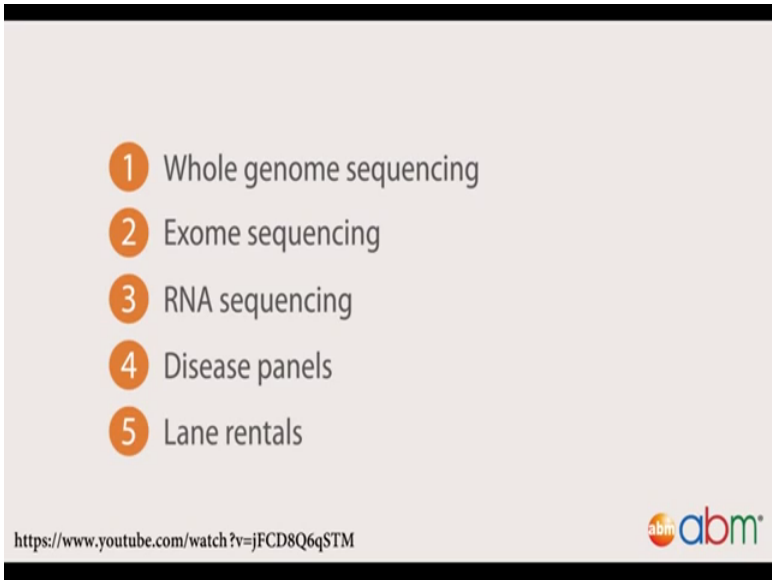




In order to be able to showcase and compare the different technical aspects of each of the above technologies the number of coverage the each one generates when sequencing the whole human, mouse, Arabidopsis thaliana, and E.coli are calculated and presented here. The presented (()) (23:00) is based on a most powerful machines of which technology further detail can be find on the knowledge base. For whole genome sequencing (())(23:08) to be useful a minimum of 30 times coverage is required, as it can be seen the pyrosequencing method is only able to sequence the E.coli genome and enough coverage to resolve entire data.

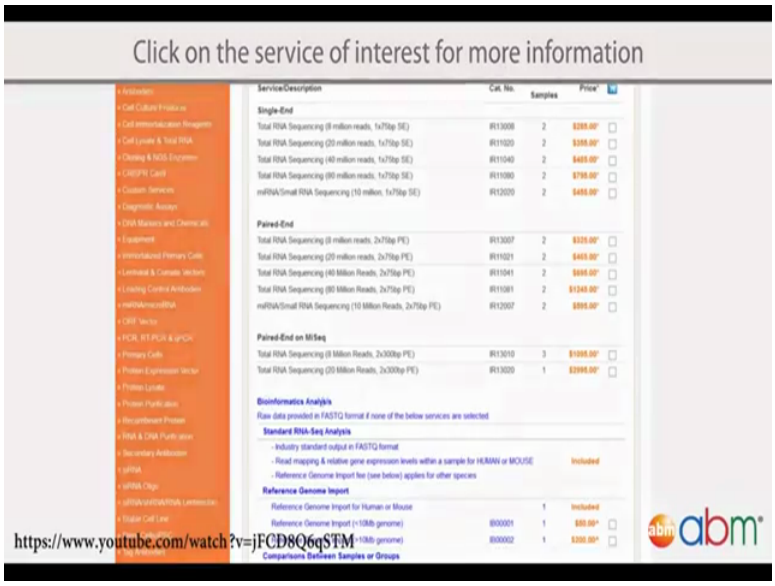
The sequencing by synthesis method which is the most popular method (())(23:22) market is able to generate hundreds of coverage per run. In fact with this machine it is possible to sequence 15 individuals within three and half days. The sequencing by ligation method also generates enough coverage for all genomes to be used, however it is incapable of generating millions much output illumina hiseq machines. The ion proton machine is used mostly in clinical settings because it is able to provide sufficient size output within two hours.

(Refer Slide Time: 23:53)



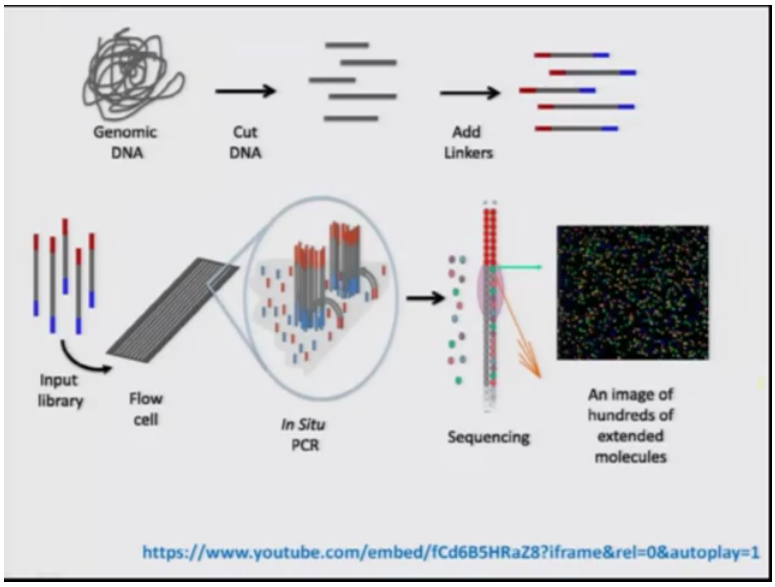
ABM offers a wide range of next generation sequencing services, these include whole genome sequencing, exome sequencing, RNA sequencing, disease panels, lane rentals and much more.

(Refer Slide Time: 24:05)



To be able to access our services please visit our website at www.abmgood.com and from there click on the NG sequencing services link. This will load our NG service webpage (())(24:14) available services, clicking on a service of interest will showcase the technical of details pressing at bioinformatics solutions that are related to that particular status.

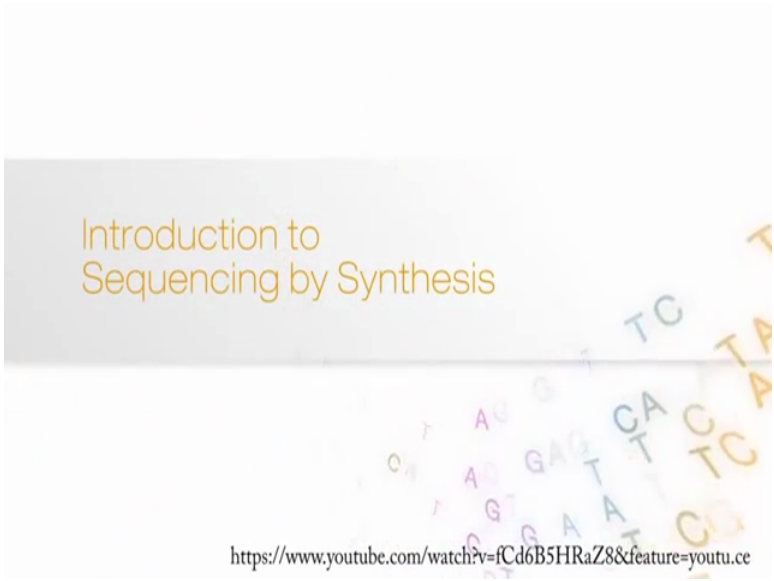
(Refer Slide Time: 24:31)

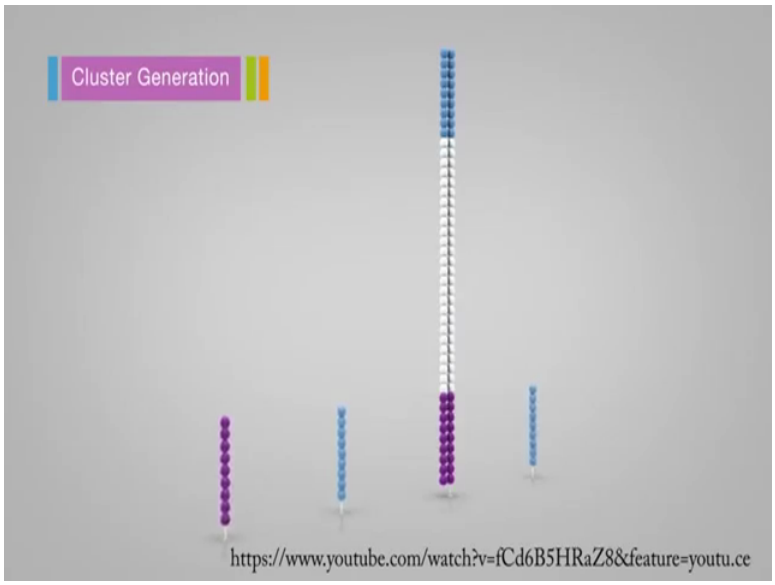


So this is just to give you overview that we have the DNA and that is cut and made into fragments and you have the adaptors that are added another side and then flow them through the grids or the chambers that are present in this cell and they all go and bind depending on what you

know the adaptors complimentary sequence present over there and then you amplify make clusters of identical copies and then you have the DNA polymerase adding this bases and the image comes and you are reading that is what going to be show in this video.

(Refer Slide Time: 25:11)

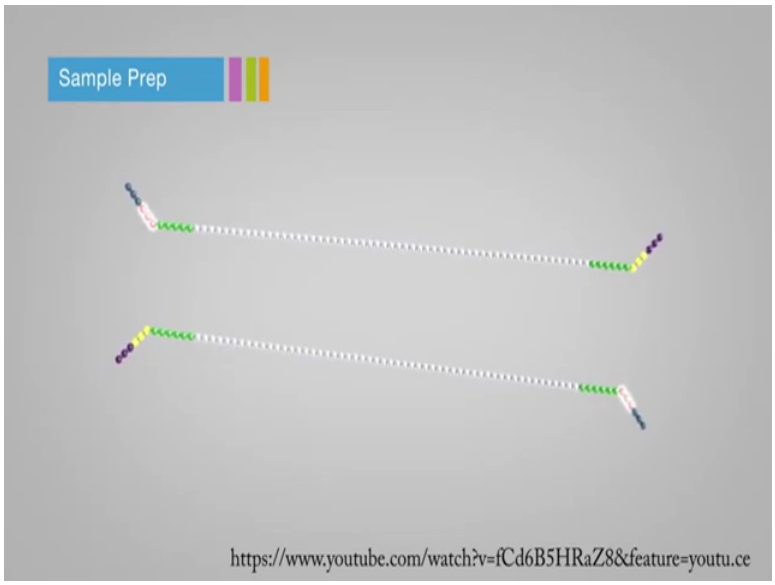



Cluster Generation
https://www.youtube.com/watch?v=fCd6B5HRaZ8&feature=youtu.ce


Sequencing
https://www.youtube.com/watch?v=fCd6B5HRaZ8&feature=youtu.ce

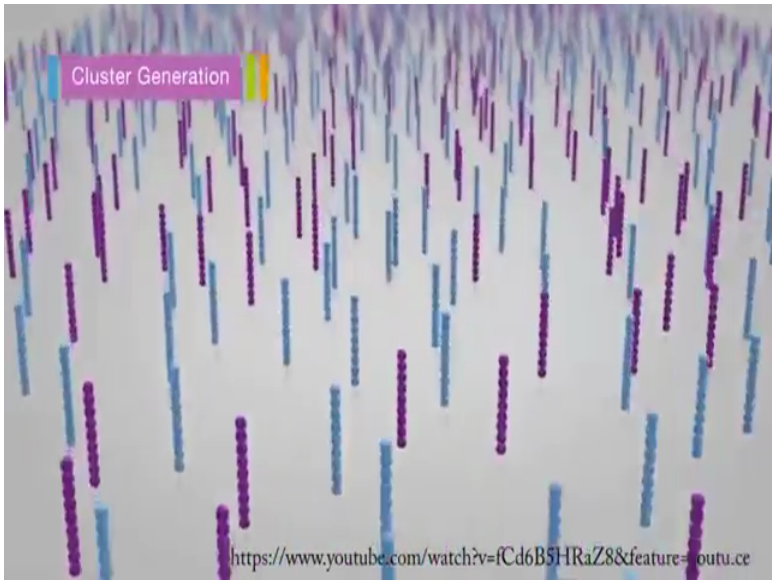
The illumina sequencing workflow is composed of four basic steps Sample Prep, Cluster Generation, Sequencing and Data Analysis.

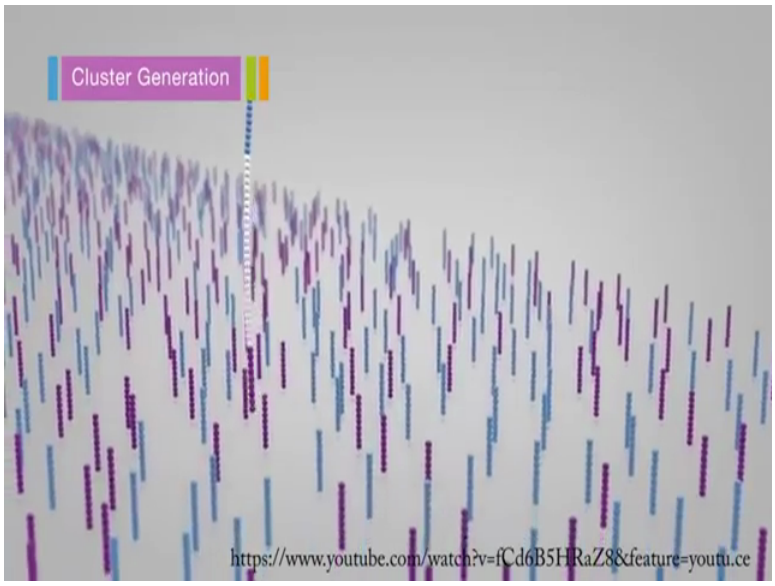
(Refer Slide Time: 25:32)



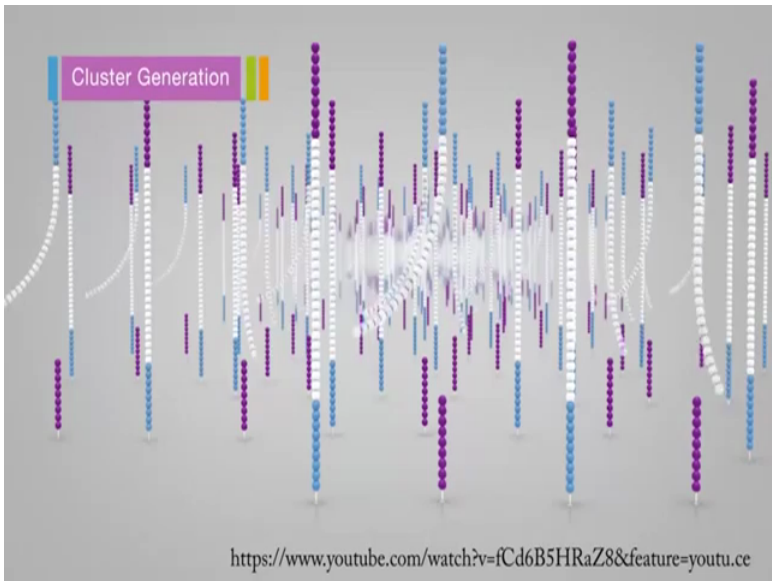
There are number of different ways to prepare samples all preparation methods add adaptors to the ends of the DNA fragments through reduce cycle amplification, additional motives were introduced, such is the sequencing binding side in disease and returns complimentary to the fluocells oligos.

(Refer Slide Time: 25:53)




Cluster Generation
https://www.youtube.com/watch?v=fCd6B5HRaZ8&feature=youtu.ce


Cluster Generation
https://www.youtube.com/watch?v=fCd6B5HRaZ8&feature=youtu.ce

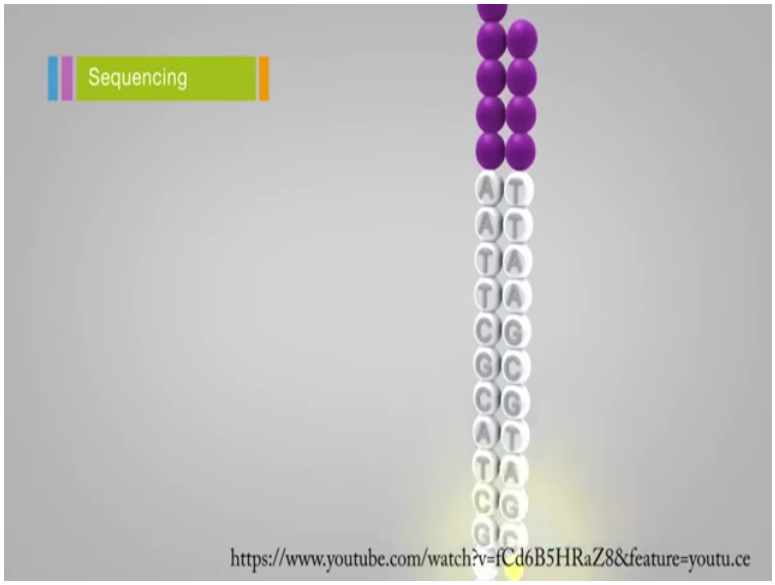


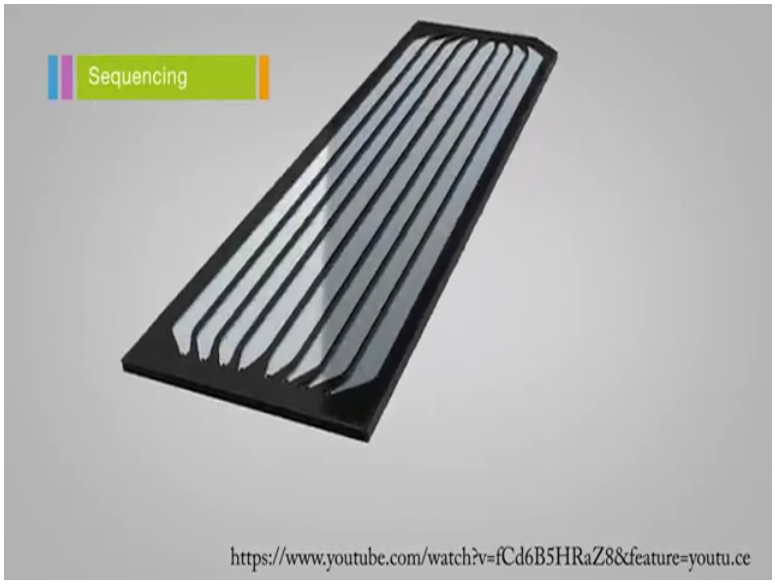
Clustering is a process for each fragment molecule is isothermally amplified. The fluorocell is a glass side with lames, each lame is a channel coded with (())(26:04) composed of two types of oligos. Hybridization is enabled by the first of the two types of the oligos on the surface. This oligo is complimentary to the adaptive region on one of the fragment strands. A polymerase creates a compliment of the hybridized fragment.

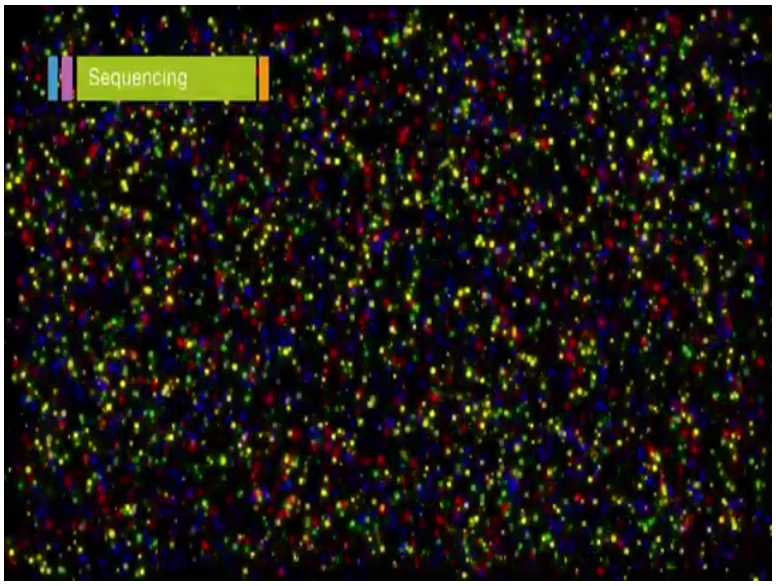
The double stranded molecule is denatured and the original template is washed away. The strains are clonally amplified through bridge amplification. In this process the strain falls over and the adaptive region hepatizes to the second type of oligo on the fluorocell. Polymerize is generally the complimentary strain forming at double stranded bridge, this bridge is denature resulting in two single stranded copies of the molecule that are gathered to the fluorocell.

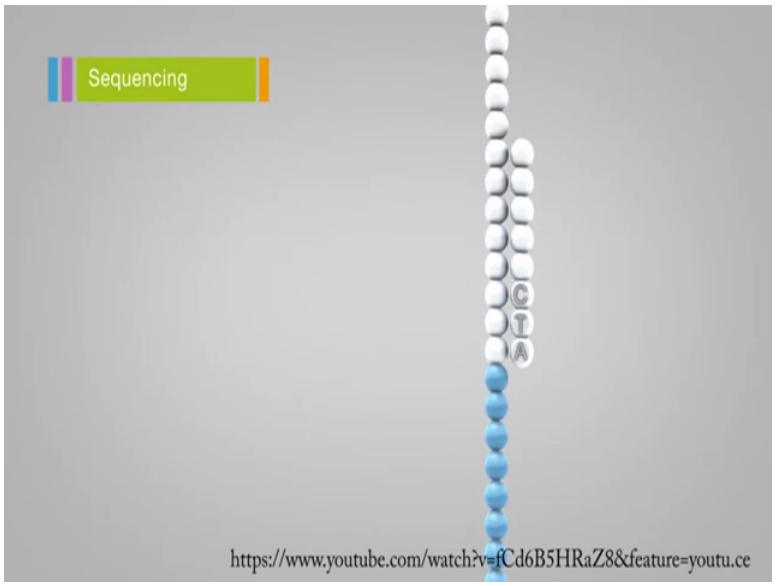
The process is then repeated over and over and they cruise simultaneously for millions of clusters resulting in clonal amplification of all the fragments. After bridge amplification the reverse strands are cleaved and washed off, leaving only the forward strands. The three prime ends are blocked to prevent unwanted priming.

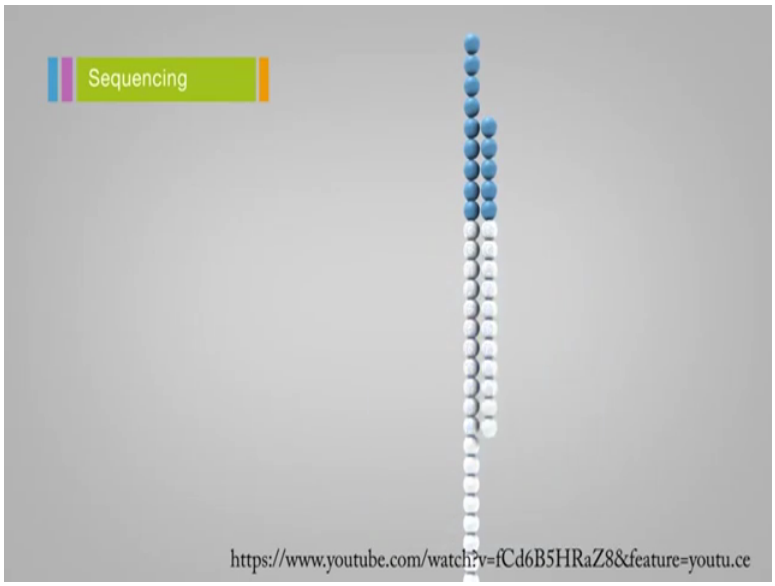
(Refer Slide Time: 27:19)






Sequencing


Sequencing
C
T
A
https://www.youtube.com/watch?v=fCd6B5HRaZ8&feature=youtu.ce



Sequencing begins with the extension of the first sequencing primer to produce the first read, with each cycle for (())(27:26) nucleotides compete for addition to the growing chain only one is incorporated based on the sequence of the template after the addition of each nucleotide the clusters are excited by a light source in the characteristic fluorescence signal is emitted, this proprietary process is called sequencing by synthesis.

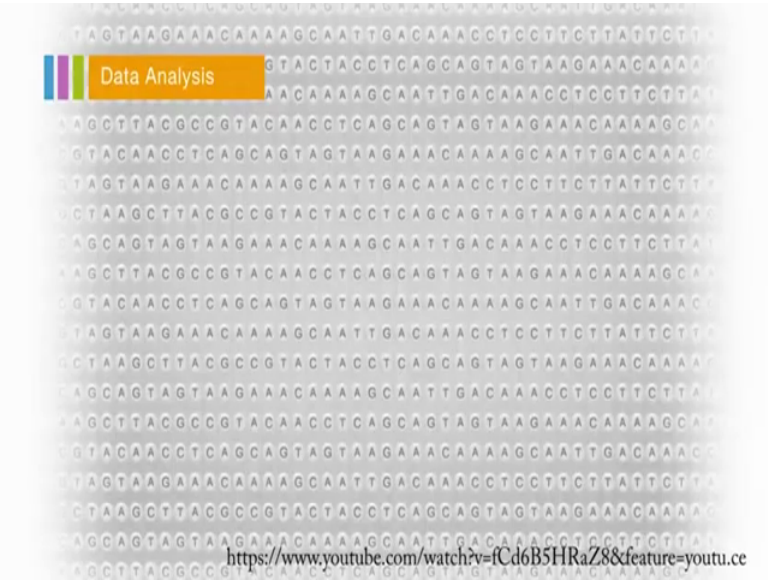
The number of cycles determines the length of the read, the emission wavelength along with the signal intensity determines the base call. For a given cluster all identical strands are read simultaneously, hundreds of millions of clusters are sequenced in a massively parallel process.

This image represents the small fraction of the fluorocell. After the completion of the first read the read product is washed away.

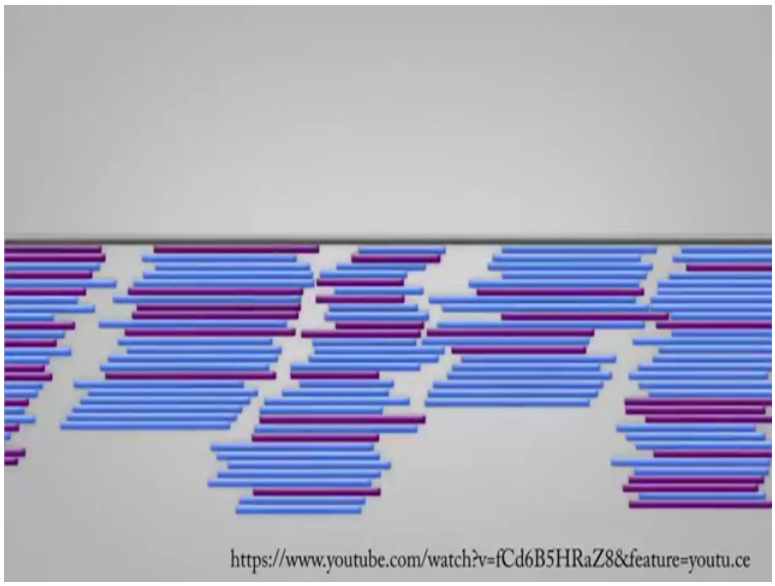
In next step the index one read primers introduced and hepatized to the template, the readers generated simulate to the first read. After completion of the index read the read product is washed off and the three prime ends of the template are deprotected. The template now falls over and binds the second oligo on the fluorocell. Index two is read in the same manner as index one.

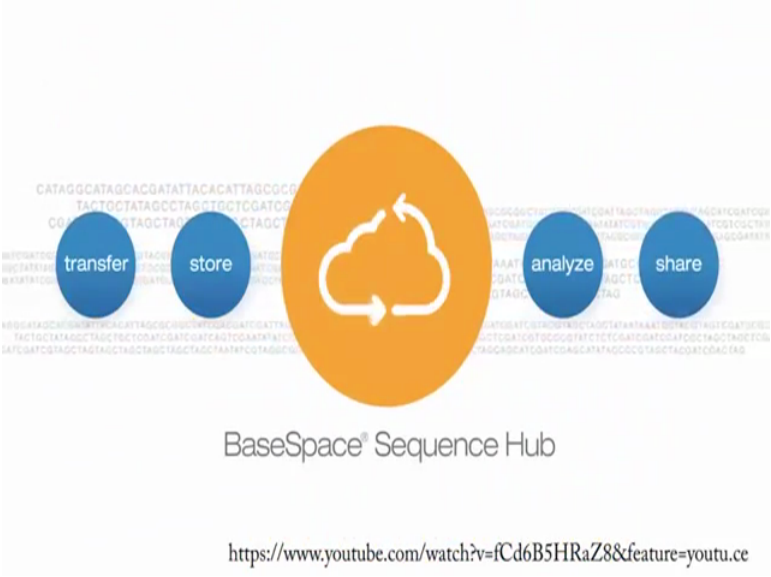
Polymerises extend the second fluorocell oligo forming a double stranded bridge. This double stranded DNA is then mineralized in the three prime ends are blocked. The original forward strand is cleaved off and washed away, leaving only the reverse strand. Read two begins with the introduction of the read two sequencing primer. As with read one the sequencing steps are repeated until the desired read length is achieved. The read two product is then washed away. This entire process generates millions of reads, representing all the fragments.

(Refer Slide Time: 29:33)



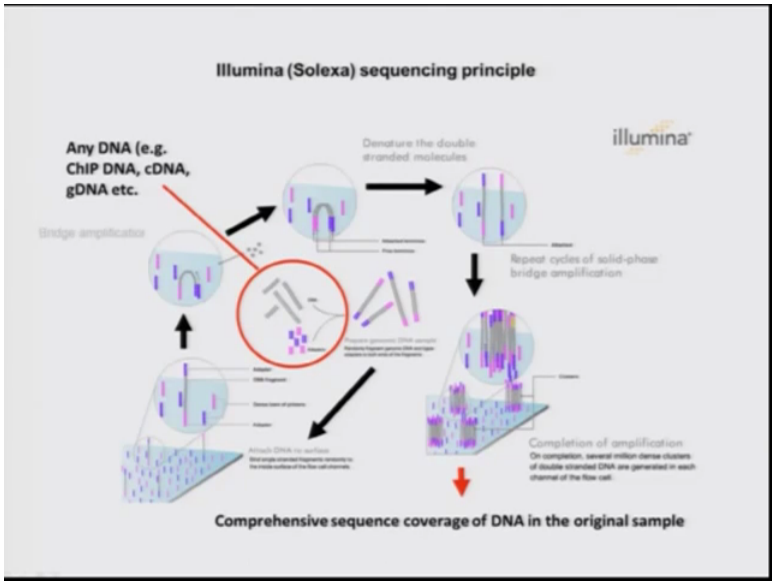

Data Analysis
Forward read
Reverse read
https://www.youtube.com/watch?v=fCd6B5HRaZ8&feature=youtu.ce


https://www.youtube.com/watch?v=fCd6B5HRaZ8&feature=youtu.ce



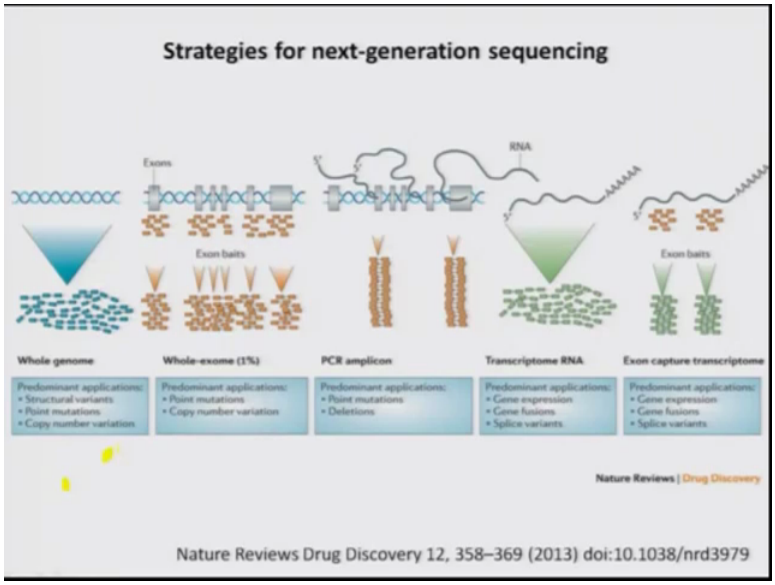
Sequences from pooled sample libraries are separated based on the unique indices introduced during the sample preparation. For each sample we expect similar stretches of base calls are locally clustered. Forward and reverse reads are paired, creating contiguous sequences. These contiguous sequences are aligned back to the (())(29:39) for variant identification. The paired end information is used to resolve ambiguous alignments, genomic data can be securely transferred, stored, analyzed and shared in BaseSpace Sequence Hub. Discover the possibilities of next generation sequencing.

(Refer Slide Time: 30:08)



So what has been shown is the illumina platform, what is called as Solexa sequencing principle and this pretty much gives you comprehensive sequencing coverage of the DNA. So this DNA what you are using could be any DNA could be the genomic DNA, it could be the cDNA meaning you have made the copied DNA of the MRNA now you are sequencing the cDNA therefore it represent the MRNA or it could be the chip determining the DNA that are bound to second transmission factor or histone and so on. So any DNA can be used here to sequence and that would tell you what is that sequence that particular DNA represents.

(Refer Slide Time: 30:52)



So that you know this is a versatile platform you can use it for any and some of these usage is shown here one of course is the version on sequencing, so you can sequence the entire genome using this method or it could be the what is called as exome, whole exome sequencing meaning you do not sequence the entire genome. We are looking at any disease you are likely that the disease is cost by defect in the coding sequence.

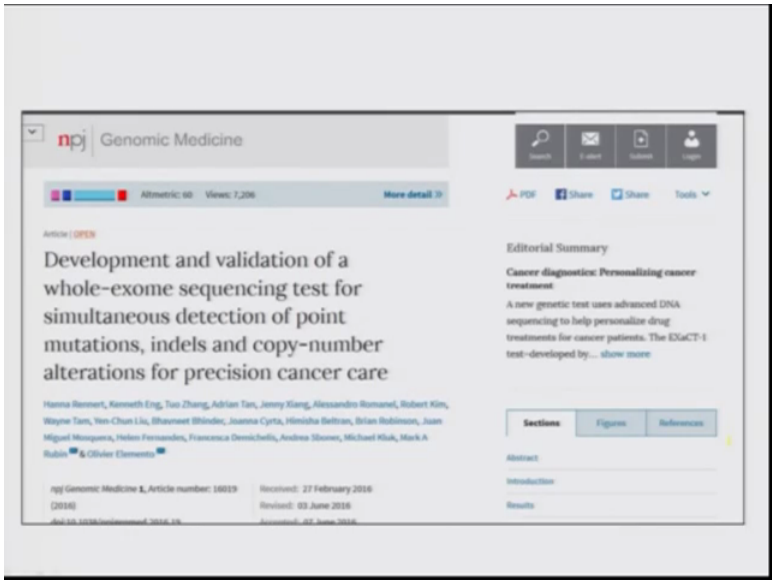
So what you do is you pretty much select the regions that code for the genes, right the protein coding sequences. Therefore you selectively you know enrich these regions and then sequence (())(31:26) genome. So this depends on how do you capture this DNA or you can restrict your analysis to a few set of genes not necessary the entire gene but the region that are most likely mutated in many of the family for example clinical applications.

So you can go on look at point mutations, deletions and so on. So you amplify small segments and go and look at large number of genes but restricted to small regions or you can start with the RNA and convert them into cDNA and use cDNA for understanding gene expression because this also not only sequences the RNA but it gives the (())(32:04) meaning if gene is expressed highly in a given tissue so you have likely to have more (())(32:10) derived from that RNA because you have more copies of the RNA.

And if you have read many of such sequences then the reads number of reads for a given gene would tell you the expression value itself, because it is not only qualitative but it also give the quantity because number of reads represent number of copies which represent the expression level of the gene. And you can also understand you know what is called as you know gene expression with regard to for example the splice variant. Remember we are discussing how the gene can undergo (())(32:43).

So most of the micro ray you have a limitation with regard to what you know genes which splice variant you are looking at. But since you are sequencing here and if you restrict your sequence analysis to be exome intron or exome boundaries you will be able to tell which spliced worm is more abounded which is less, right. That really helps us so that is another big advantage of next generation sequencing.

(Refer Slide Time: 33:13)



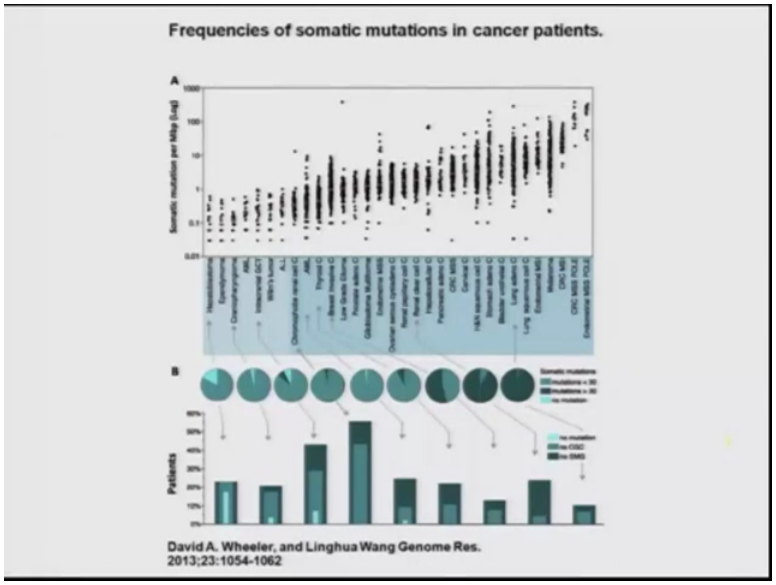
So what you are going to look at is some of the examples of how the sequenced the next generation sequencing has helped. Here is a paper which talks about development and validation of a whole exome, meaning we have the sequencing restricted to the exome. Sequencing test for simultaneous detection of point mutations, insertions, deletions and copy number alterations in precision cancer care.

So really want to type the cancer and (())(33:35) the therapeutics based on the cancer type and you have to do it quick because the patient is already in the hospital so you remove a small cancerous region and then you extracted the RNA and the DNA and go on sequence it within a day or two we are able to get the analysis which tells you what kind of changes happened, which is not restricted to point mutation where one base changed.

But it is also involving what is called as insertion deletions, there are regions in the genome specially in the intronic region or in the promoter region where there are repeat element that are present either in two copies or one copy it is called as insertion and deletion polymorphism that modulate the level of the expression or the splicing or the functioning of the gene. So we can correlate that and then it is also called as a copy number alternations, there are segments in the genome which undergo you know multiple copy number variation they amplify get amplified and cancer the number of copies go up.

So this is here unlike Sanger sequencing where you are cloning and sequencing which do not tell about you know the quantity or you know the qualitative measurement. This next generation sequencing it talks about how many reads, how many molecules we are able to read. It also tell you the copy number variation, that is a huge advantage of this you know next generation sequencing.

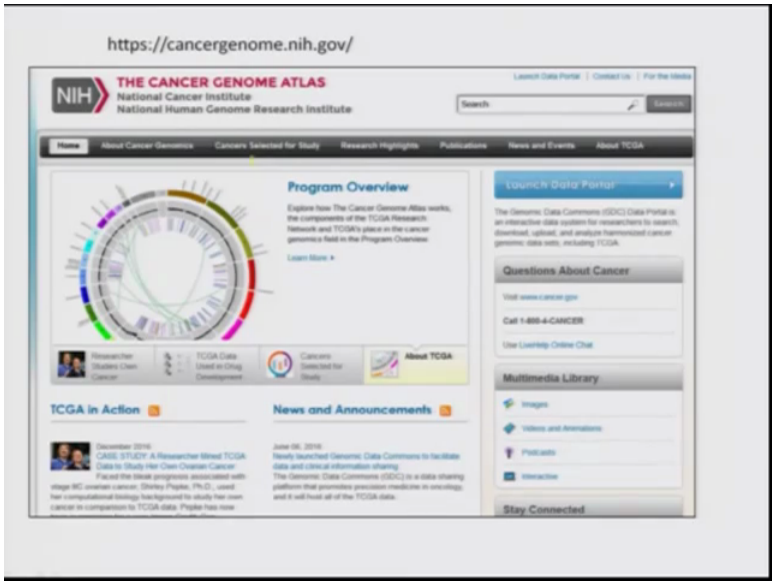
(Refer Slide Time: 35:05)



We are talking about such kind of panel one that came published recently it talks about the frequencies of somatic mutations cancer patients. Because it is huge challenge as you know the cancer is just one gene that is mutated, it is a large number of genes that are mutated as a result you have cancer the number of genes that are mutated could vary from one region of the body to the other even if it is cancer is undergoing metastasis or depending on the cell types.

So this is one such analysis where they have used the next generation was (exome sequen) genome sequencing to understand what are the changes that are present here, right on different cancers that is say these are somatic mutations per mega base pairs, you can see that it varies it varies from patient to patient, right depending on what kind of cancer you are seeing that really talks about the heterogeneity and the need for such kind of approach because you know information indicate from one patient it is one patient is not really help the other patients. So you need to understand what is so unique about that particular you know cancerous population found in a particular patient.

(Refer Slide Time: 36:20)



That is the tremendous advantage and understanding such kind of you know complexity involved in cancer is also now you have what is called as Cancer Genome Atlas which is you know is kind of a repository which adds all those data into the database. So one can go and look at and then come up with a better way to analyze the cancer. With that we will end this lecture, now next

lecture we will see a few more applications of the RNA seq or high throughput DNA sequencing technologies.