Interactomics: Basics and Applications

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

Lecture – 47

Next – Generation Sequencing Technology – Ion Torrent TM

Today we are going to talk about Next Generation Sequencing Technologies. Before we talk about this you know revolutionary technology platform, let me share a story with you. So, whenever the grant student like you, I was finishing my masters in year around 2000-2001. That time the various genome sequence projects were getting accomplished, it was you know big news that time big breakthrough that time that the human genome project the first draft was completed. And many other model organism and their sequences were also getting completed at the same time.

Just imagine that you know first time we were getting information that what you know makes it human, what our genetic composition is about. It was you know really big science breakthrough because first time we now know that how many genes are actually available to us which makes it a human which makes us so unique and different than other model organisms. Before that they were only speculation that you know what makes human and you know how many genes you may have, but the first time the evidence for genes and the our genetic composition the genome was first time revealed right.

It was definitely big revolution in the science field in general and definitely in the biology and medicine area, but its not end right it is not finished that time. So, many times a technology showed as a promise and then you know it fades away, but what is happened in this area of genomics that you know 2002 and 3 first time the draft human genome were released. But since then there has been tremendous advancement which has happened in this field of genomics reaching to where we are which we say as a next generation sequencing technology.

The series of advancement happened for the first generation, second generation, third generation of next generation sequencing technology and I must admit that advancement

made by engineers and technologists, big data scientist and biologists together has really brought where we are currently in our magnitude of doing sequencing based experiments. So, what has changed from 2000 to in 2018? Main things which has changed is our speed, the speed at which we can now sequence the genome imagine the human genome projects took more than 10 years time to accomplish.

Now, within 2 days time or 48 hours time you can do sequencing of a given individual with the next generation sequencing technology that is humongous change. Additionally what 2 billion dollar took accomplish one human genome sequence now it takes maybe only a thousand dollar to accomplish the same sequencing technology based work.

So, both cost wise as well as the speed wise as well as accuracy wise the next generation sequencing moving from the Sanger sequencing based platform have made tremendous revolution. So, I am sure now you are excited to know that what NGS technologies are which are the currently big industry player, big industry technologies which are leading this field.

And in this light we have invited a guest doctor Atima Agarwal from thermo fisher scientist who is manager a commercial service training and tech support at thermo fisher scientific who will be taking this lecture. And she will try to provide you the basics of next generation sequencing first and then give you little more detail and specifics for one one of the technologies of ion torrent. So, let me welcome doctor Atima Agarwal for her lecture on NGS based technologies and its applications.

So, I am Atima Agarwal from thermo fisher I have been with this company since many years now. So, if you remember or if you have heard in pre 1990 era the sequencing used to be done on gels with radioactive labels yeah.

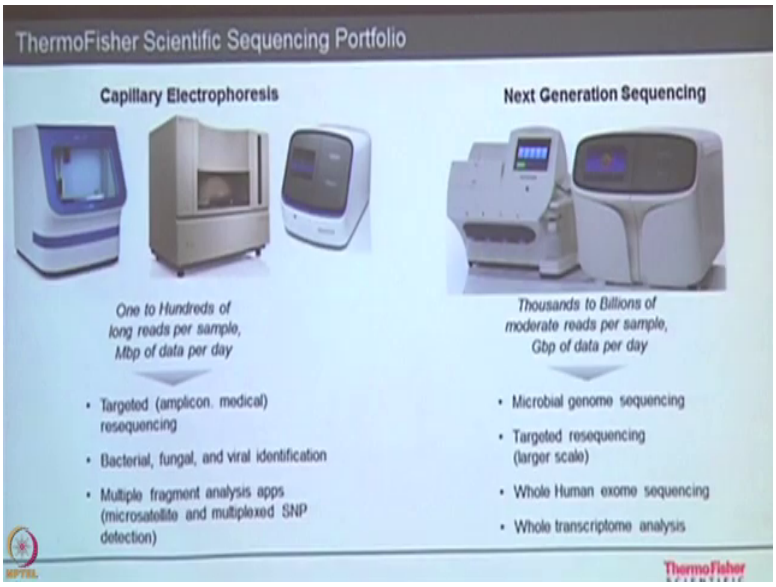
So, and it used to be a very very cumbersome affair, but then it had contributed a lot to the science in those terms that first human genome was being sequenced on these systems and there were many platforms many of these capillary electrophoresis platforms, if you look into the pictures of Sanger institute during that time in lines there were these kind of systems and

they were generating sequences day and night and that is how the first human genome got sequenced and got assembled.

So, those were a lot of like I would say that benchmarks and sequencing suddenly somewhere during 2004 to 2005 we started hearing about next generation sequencing. So, if I were to talk about next generation sequencing it is more aptly called as massively parallel sequencing. So, with Sanger what you are doing is that either you are running it on 96 well plates or you are running it on strip tubes. So, one well is giving you one sequence yeah. So, one sequence which can which is a stretch of ATGC and this can be as long as a 1000 base pair 1100 base pairs and it can be as short as a 100 base pair fragment.

So, one well gives you one sequence now suddenly with massively parallel sequencing you can imagine what is happening is you have millions of wells and you have this sequencing per say happening in all those millions of wells and that is how you are generating a lot of data per run. So, that is massively parallel sequencing.

(Refer Slide Time: 06:13)



So, just to tell you that just because we are now nowadays talking a lot about NGS it is not that capillary sequencing has lost its charm or has lots its lost its utility, it is still a very user friendly technique for a lot of applications.

So, for tech for things wherein you do not really require a lot of data things can be done simpler for a lot of microbial research or a lot of single gene based tests and all those things they can be still very well and very efficiently being done on capillary sequences till they this has been the gold standard technology for doing a lot of things like targeted.

So, you have certain amplicons to be sequenced or you have a de novo organism to be sequenced to fill up those kind of assemblies then you have multiple there are multiple fragment analysis applications where in you are not actually sequencing the fragment you are

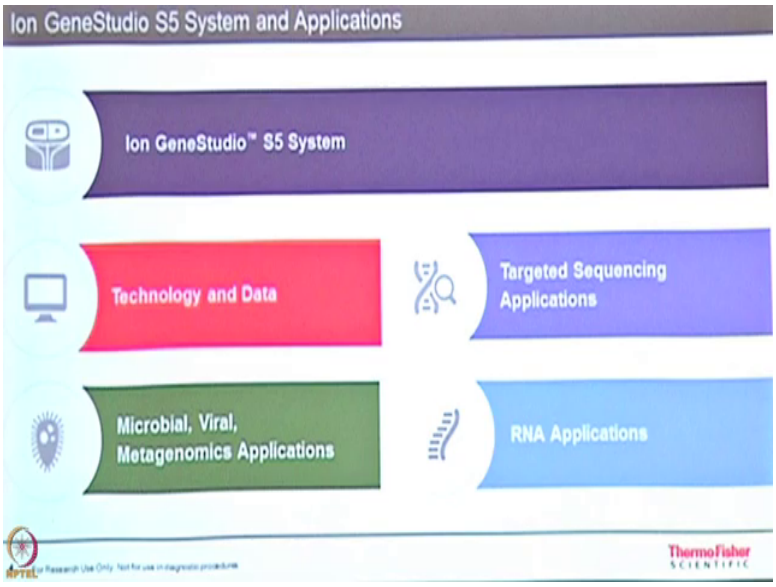
just looking into this size and based on the size polymorphisms you are telling you are giving certain scientific answers.

So, these this is about capillary sequencer coming to next generation sequencing. So, now, as I mentioned that this is a technology where in you have millions of wells.

(Refer Slide Time: 07:34)

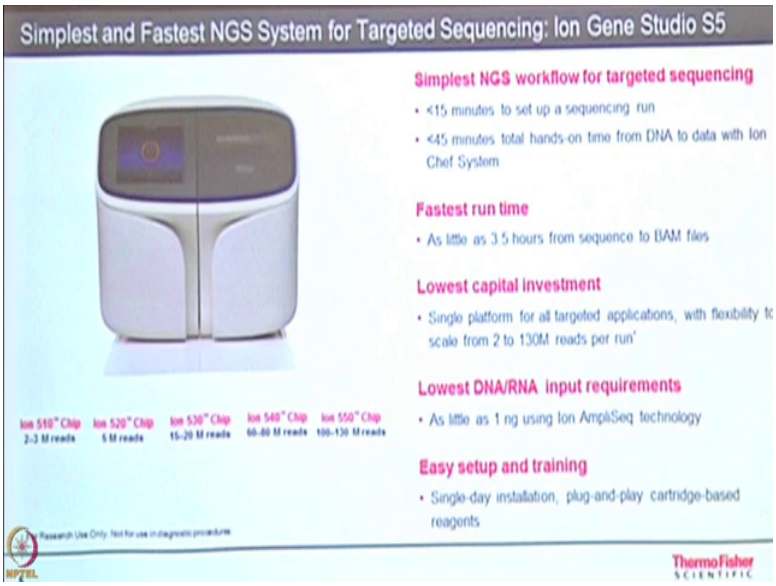
So, if you see that these are the chips which are used with ion technology and you have millions of wells on these chips. So, millions of wells. So, you can imagine that each well has a sequence and that is how you are generating millions of sequences in a run yeah.

(Refer Slide Time: 07:51)



So, the agenda how we have planned out the session for you is that we are going to talk about our new series of NGS platforms which is ion gene studio S5 systems. The technology behind the gene studio S5s and some of the applications. So, this is the fastest sequencer which is available in terms of an NGS which can produce starting from once you have started your sequencing run it can produce data in as little as 3.5 hours from a sequencing run.

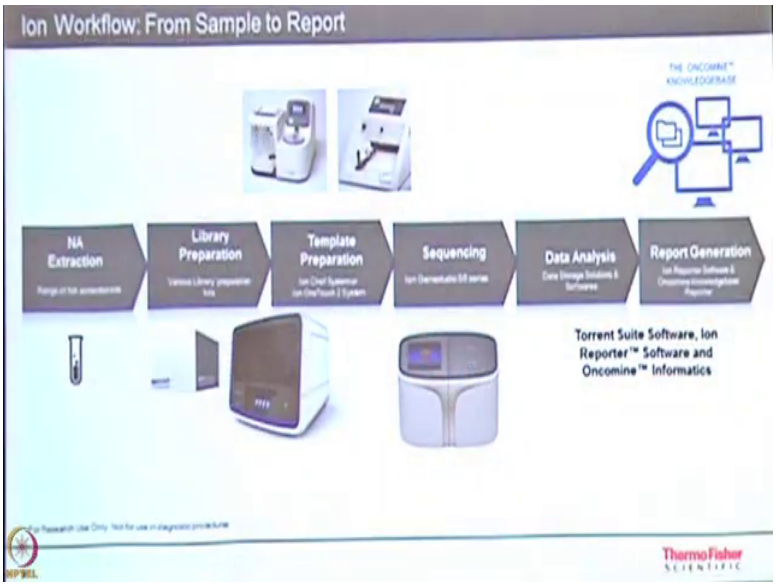
(Refer Slide Time: 08:28)



There are lots of different kinds of applications people want to work on NGS and lot of these samples are very challenging samples like when you extract DNA or RNA from FFT blocks.

When you are working with liquid biopsies like you do not you really are looking at cell free DNA yeah or you have very little RNA from a cell and still you want to study the whole transcriptome of this cell. So, our technology is such that it is compatible with as low as 1 nano gram of DNA or RNA and the backbone of this technology is ampliseq technology, we will come to those details later.

(Refer Slide Time: 09:09)



So, coming to the workflow now. So, how we are generating these millions of reads? So, you started nucleic acid now this nucleic acid can be DNA can be RNA. So, where do you think you would be starting with the DNA? What kind of experiments do you think you would be starting with the DNA on NGS DNA as a sample?

Student: Whole genome sequencing.

Whole genome sequencing, yes.

Student: Plasmids.

Plasmids yes.

Student: Targeted read sequencing.

Targeted read sequencing. So, when you are not. So, targeted read sequencing is basically when you do not want to sequence the whole genomes you just want to sequence some targeted regions of the genome that is targeted re sequencing yeah. So, where do you think you would be using RNA as your starting material?

Student: Transcriptomes

Transcriptomes, yeah again like.

Student: Targeted.

Targeted targeted you are looking at some targeted transcripts or you are looking at some targeted fusion events yeah anything else.

Student: RNA seq with protein atmost.

Yeah RNA seq and RNA seq typically why would you do an RNA seq? You want to sequence.

Student: Because DNA will make function that DNA will be negative that assumption curve provides Rna.

Yeah. So, this DNA is transcribed into Rna.

Student: Yeah.

And then.

Student: (Refer Time: 10:42).

Hence there are more probabilities that this particular part of the genome is actually getting transcribed or further translated into a protein, but that is just a possibility there are many long there many non coding RNAs also yeah its like small RNAs, yeah.

Student: Small RNAs (Refer Time: 10:57).

Yeah.

Student: (Refer Time: 10:58).

Which are?

Student: (Refer Time: 10:59).

Yeah. So, which are they are involved in lot of regulation activities of how the transcriptome is shaping within a cell yeah. So, people do a lot of different kinds of experiments now for your what you would look out for in your NGS is that all these experiments because the data requirement for all these experiments is very different yeah. So, when you are doing a seq genome its a very straightforward calculation that you generally tend to generate a 100 x data.

So, whatever is the genome size you multiply by hundred there are definitely ifs and buts around it we will not go into those details, but then the data requirements for these different kinds of experiments are very different yeah. So, that is where you need that flexibility that single NGS system can cater to all those requirements.

So, you started with your DNA or RNA yeah to begin with. So, you isolated DNA RNA from cells or from some culture isolates or whatever. So, now, once you have isolated this DNA and RNA, the next step is that you prepare a library. Now, library is basically a collection of

fragments whether it is coming from DNA or coming from RNA which you finally, want to sequence yeah. So, this can be an enriched portion yeah.

So, for targeted sequencing you would need to enrich that those targets first and then only you can go for the sequencing for whole genome sequencing since your aim is to sequence whole genomes. So, you do not need any enrichment you will just extract DNA and you will go forward for library preparation.

So, library preparation is basically collection of fragments whether DNA or RNA which you finally, want to sequence yeah. So, what is the basic aim in library preparation is that you are now these next generation sequences you we talked about that you are generating 200 base pair it reads you are generating 400 base pair reads you are generating 600 base pair reads.

So, supposingly we take an example of whole genome sequencing now what you gonna do is with whole genome sequencing you want your aim is to sequence that genome as fast as possible yeah and with as bigger reads is possible. So, probably you will pick up a 400 or a 600 base pair chemistry and so what you need to do with the genome is you share this genome yeah you break this genome into smaller pieces. So, that can be done by enzymatic treatment or that can be done by sonication.

So, once you have broken down this genome now these are the fragments which you finally, want to sequence, but you do not know what is there. So, you need to have some adapters yeah or linkers which will help you facilitate sequencing these regions yeah. So, in library preparation what you are doing is you are enriching the region of interest and second is you are ligating it at both the ends you are ligating it with linkers, these linkers are double stranded DNA fragments which are basically exploited.

So, that the primer can then sit on these kind of fragments and then. So, you have your fragment of interest in between you link the adapters at both the ends and so why? So, that the now the these adapter sequences you know because these are coming from a chemistry and

we tell you that these are the adapter sequences and then the primers comes in and sits on these adapters, so that you can sequence this region in between yeah.

So, that is basically. So, with library preparation what you are doing is you are enriching your region of interest and you are ligating these fragments to adapters yeah, then comes template preparation. So, basically now these are the things which comes which are specific to a technology. So, in template preparation what you are doing is that you are taking these library molecules and you are amplifying onto ion spheres.

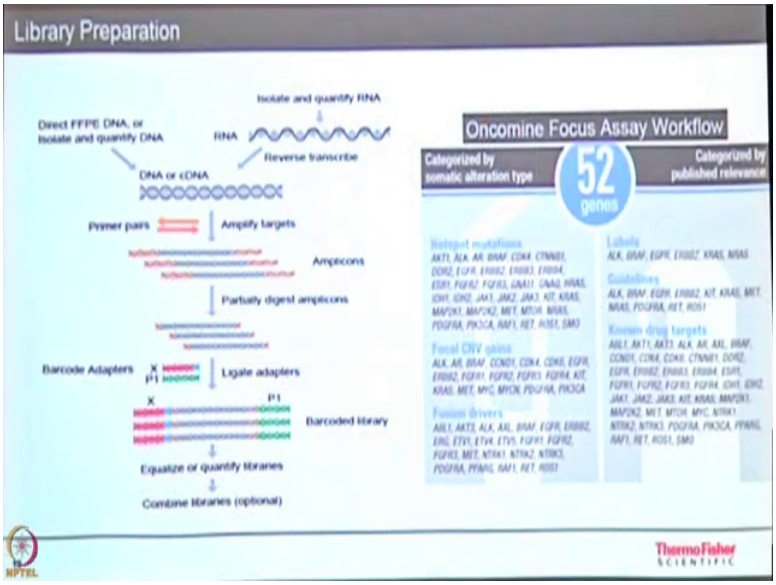
So, we are talking about ion torrent technology. So, we are using ion spheres as the mode these ion spheres get coated onto those or get loaded onto those chips and finally, these ions spheres are carrying those regions of interest or those library molecules from which you are generating all those reads yeah.

So, that process of taking these library molecules and amplifying them onto the ion sphere is called template preparation. Then comes sequencing now you have your ion spheres which are templated which have library molecules amplified onto them and now you want to sequence. So, you load these ion spheres onto the chip and then you generate the data. So, once you have loaded you initialize your sequencer which we discussed this going to take only fifteen minutes and then your sequencing run goes for anywhere from 2.5 to 4.5 hours depending on the kind of the read length you are targeting.

So, talking about the instrumentation which is required in this workflow. So, we have. So, library preparation can be done manually, can be done on ion chef there are some kind some parts of library which can be done on ion chef yeah. So, this is an automated way of preparing libraries otherwise you have manual way of preparing libraries wherein you are using certain magnetic beats, certain library preparation kits from us. And the process generally for targeted sequencing or for whole genome sequencing the process is around 4 to 4 and a half hours for manual and probably for RNA seq its a days protocol because you have to do certain levels of enrichment and you and you have to do more QCs in that.

So, if you talk about manual preparing libraries manually. So, that is the kind of time frame otherwise you can use ion chef for preparing your targeted sequencing libraries automatedly, then comes the template preparation. Now, as we discussed during template preparation you are loading these library molecules onto the ion spheres and then finally, these ion spheres are loaded onto the chip. Now these loaded chips you take from here and you just load it in on this system and they are good to go for sequencing.

(Refer Slide Time: 17:37)



So, just coming into few details on library preparation. Now, how we go about library preparation is since you are targeting variations at the DNA level as well as at the RNA level you isolate DNA and RNA yeah then you reverse transcribe this RNA. So, now, you have DNA or CDNA. So, you take this primer pool and you use this DNA and CDNA as your target all these genes yeah get amplified at one go and then since the primers are coming from

conserved regions wherein you do not have any hotspots or any kind of variations to be looked into we partially digest these primer sequences and then we ligate adapters.

Now, when we say that you can do smaller experiments also. So, you have a 52 gene panel and you have some 12 samples to be done. So, what you are going to do is you index these 12 samples by 12 different barcodes. Now, these barcodes, so when we said that library preparation is all about enrichment of the targets and ligating adapters. So, when we are ligating these adapters you are using bar coded adapters to differentiate one sample from another yeah.

Now, this bar code is a 10 base pair unique sequence which is present in the adapter yeah which is which will finally, help you to differentiate between different kinds of samples which you are loading on 1 chip. So, once you have; once you have ligated these adapters you pool all these libraries now you are you are working with probably 6 samples on a chip. So, with based on the number of samples you plan to multiplex on a chip you are going to use the number of barcodes yeah.

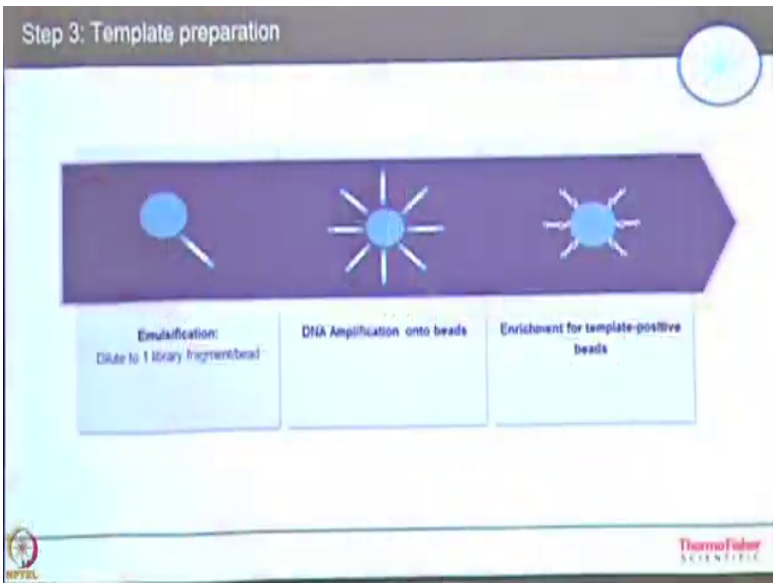
Once you have done that now you want equal amount of data for all those libraries what you are gonna to do is in terms of their molarity concentration you will use equal molar concentration for all those libraries and make one pool of that. But the aim is that when you are doing small RNA sequencing for all these samples you are aiming that you generate equal amount of data for all these samples, but because otherwise probably some of the samples have worked out pretty well and some will not. So, to ensure that you need to pool all these libraries in equimolar proportion.

So, once you have pooled that after your library preparation step now you started with 12 samples or you started with 20 samples now that becomes 1 sample yeah. And now how are you going to differentiate? Now, because it is now one sample, but you started with twenty samples. So, when you generate the data how are you going to differentiate?

Student: Using the bar codes.

Yes, we use bar coded adapters. So, the system is first thing its going to use when its sequences it is sequences the bar coded barcode sequences and then it sequences your region of interest and then it makes the bins that this read belongs to this sample and it is in this bin.

(Refer Slide Time: 20:57)



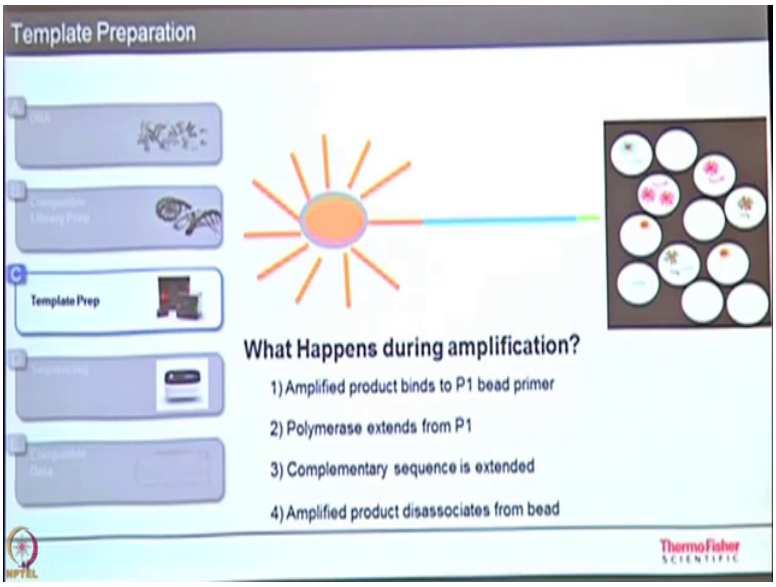
Once we are through with our library preparation, then the next step is template preparation.

So, in template preparation the aim is that you clonally amplify this. So, you have prepared you started with targeted sequencing or you started with whole genome sequencing or you started with an RNA seq experiment you have tons of target millions of target molecules which are to be sequenced. So, what we are aiming during template preparation is that one library molecule gets clonally amplified onto one ion sphere.

So, this is the basic aim of this particular step and so this library molecule gets amplified onto this ion sphere and this is one clone which is getting amplified all over on the ion sphere then there could be some ion spheres which have not got amplified yeah. So, you do not want to load them on the chip. So, you negate all those ions spheres at the enrichment step yeah. So, what is happening is that we are using ion spheres which have an oligo which is linked to these.

So, this oligo is complementary to one of the adaptors which you used during library preparation. So, if we go back a few slides, so we used 2 adapters P 1 and one was bar coded adapter. So, and we mentioned earlier that we are basically exploiting these barcode these adapter sequences for us to enable the sequence to sequence these millions of fragments later on one chip. So, we are using this particular P 1 adapter sequence yeah for the library molecules to complimentary go and bind onto the ion sphere.

(Refer Slide Time: 22:52)



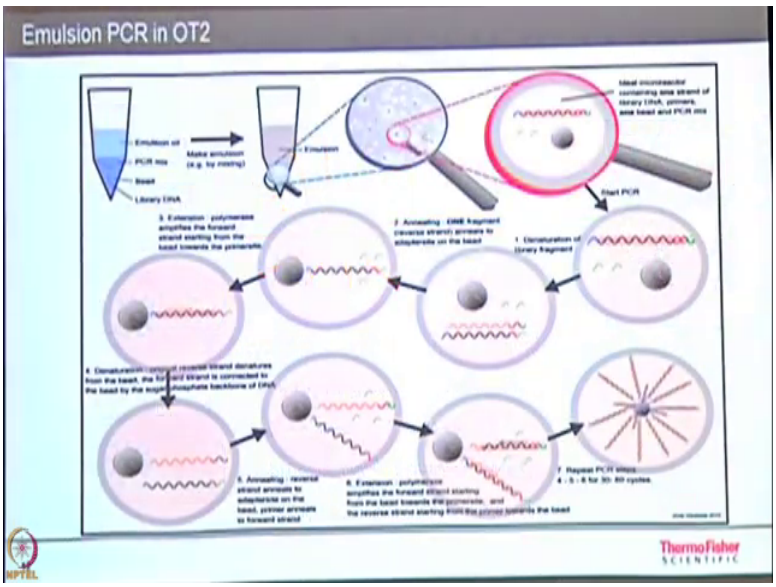
So, now if we go back to that slide. So, these ion spheres are loaded or are coated with oligos which are complimentary to the P 1 adapter. So, when these emulsions are being created you are basically making these water molecules which is surrounded by a oil and you are trying to amplify these library molecule. So, when the this emulsion is being created this is a random distribution of these ion spheres with the library molecule and enough PCR components for it to amplify this library molecule onto this ion sphere

So, what is happening is a library molecule is coming it is getting denatured and it is sitting on the complimentary sequence which is complementary to the P 1 adapter and then it goes ahead and it sequences this part and that is how you sequence. Now this goes through a cycle of denaturation annealing and extension like your normal PCR, but when while we are talking

all this has been. So, you are not manually involved with all this process all this is being done by these systems either semi automatedly or fully automatedly.

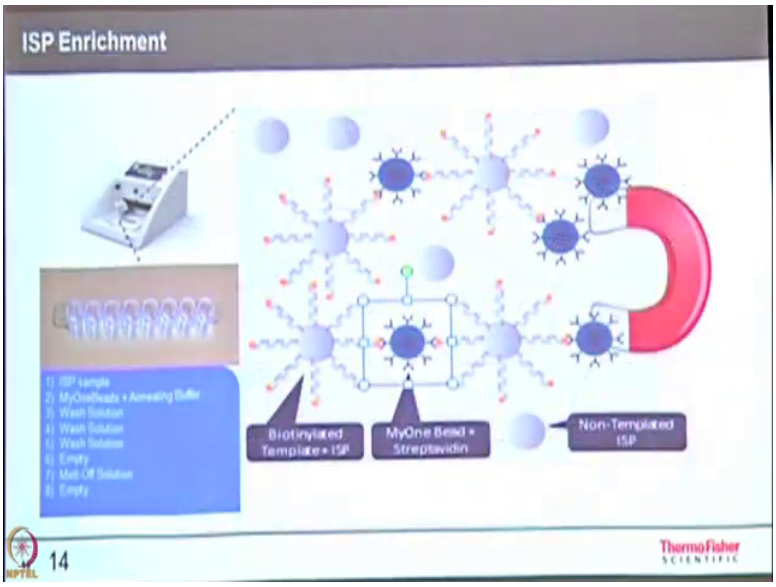
So, then when it gets denatured it then again binds to another oligo on that same bead and you then that is how you are generating you are clonally amplifying this particular molecule onto the ion sphere.

(Refer Slide Time: 24:17)



So, what is if we go into the details what is happening is you have an emulsion oil then you have a library molecule, you have some primers this is your ion sphere this library molecule gets denatured it gets bound to the ion sphere it gets extended again gets denatured. Now this template goes and binds to another oligo and finally, you have this one library molecule clonally amplified onto this ion sphere yeah.

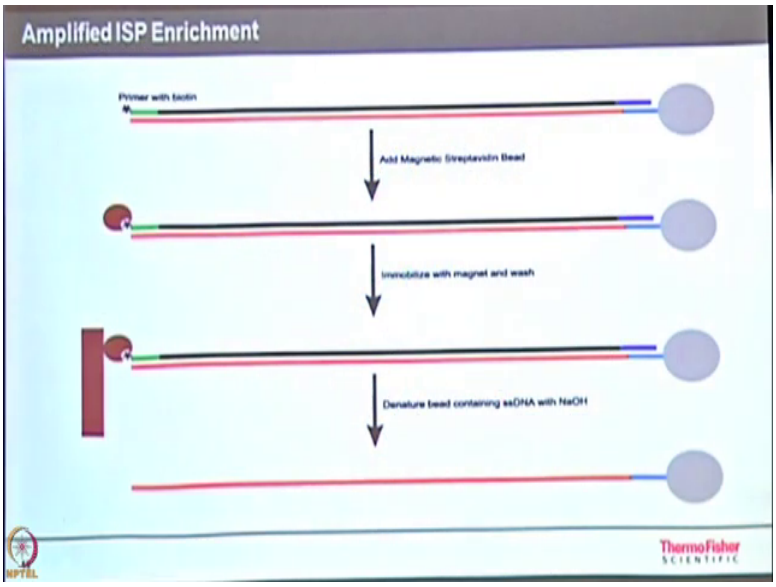
(Refer Slide Time: 24:48)



So, now we said that there could be some molecules or some of those water droplets when the emulsion was being created they did not get a library molecule. So, they will not get amplified also, but we do not want to take them on the chip because they are still going to occupy some space on the chip, so why to lose on that data. So, what we are doing at the enrichment step is when we were doing amplification on the ion spheres one of the primer was biotinylated.

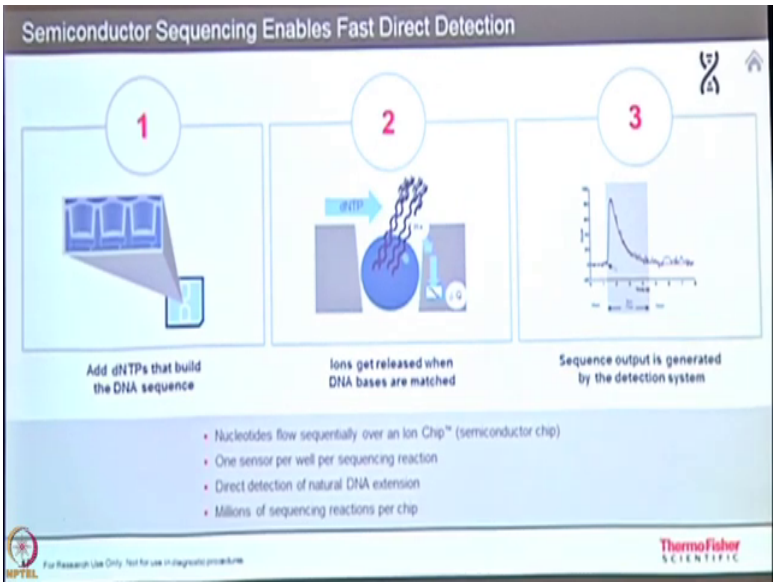
So, now all the amplified ion spheres are biotinylated, we are using streptavidin coated beads to fish all the amplified molecules and rest all molecules remain in the solution yeah.

(Refer Slide Time: 25:32)



And then we just go through. So, this is primarily your ion sphere this will have these molecules all over because we clonally amplified. So, you use streptavidin coated beads and this streptavidin streptavidin coated beads are magnetic. So, with the help of a magnet all these are fished out and then you have only the beads which are amp which have amplified which are finally, to be loaded on the chip.

(Refer Slide Time: 26:05)



So, this is the enrichment part of the template preparation then what is happening at the sequencing level now I showed you that chip that chip has millions of wells and each well is working is generating a sequence. So, your; so this is your chip it has millions of wells and if we are now to consider that this is one well this one well has one ion sphere which is loaded with one kind of template. So, what is happening during sequencing is that when we are when ion chef is preparing the sample for the sequencing it is already adding the primer and polymerase through the chip itself. Now, what else it needs for extension?

Student: dNTPs.

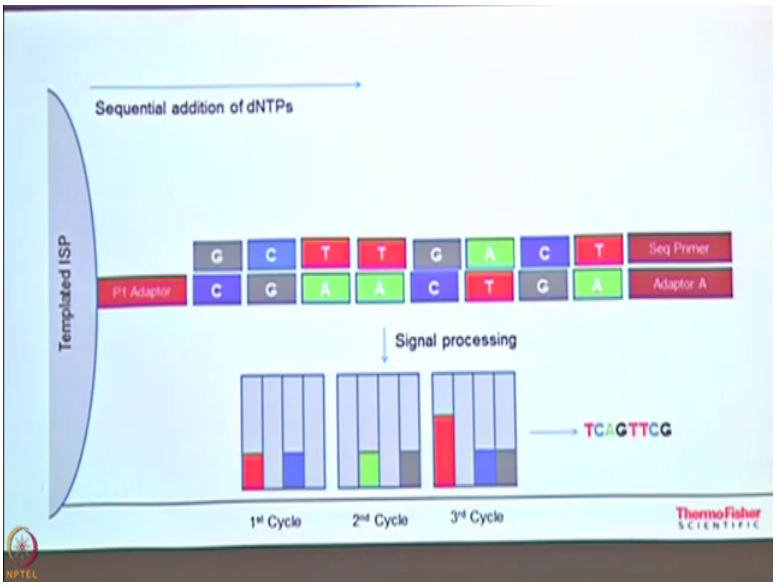
dNTPs yeah. So, what the system does is it is adding dNTPs sequentially onto the chip and then naturally we know that whenever dNTP gets incorporated or gets polymerized into a growing chain of DNA there is a; there is a bond formed between the phosphate group and

the hydroxyl group yeah and this phosphodiester bond when formed this releases a hydrogen ion yeah.

So, now we coated these library molecules onto this ion sphere. So, this same event. So, supposingly the first base here is an a yeah. So, you had flown dTTP, so it will have a complement it is a complementary nucleotide, it will bind and it will release hydrogen ions and these hydrogen ions get converted into a voltage signal which is finally, being captured. So, this ion basically is bring these release of, so many hydrogen ions in one well are bringing up a change in ph for that particular well yeah and when the ph is being changed for that particular well that change in ph is being recorded as a change in voltage.

So, basically you are converting a chemical information into a digital information and that is how you are saying that a base a signal has been received and a base has been called. Now supposingly now we said that T is your or A is your first base and the system had flown a dGTP. So, now, that is G is not going to bind to a yeah. So, what will happen is there will be no binding no hydrogen ion release, so no change in ph no signal detected by the system.

(Refer Slide Time: 28:39)



So, we will just see this by a cartoon. So, considering that this is your one ion sphere these are your adapters yeah this is the barcoded adapter and this is your adapter which you used for ligating these sequences on to the ion sphere. Then what is happening is your primer is coming in binding here you adds the dNTP sequentially. Now, what will happen is this will bind you will have a change in ph which will be recorded by the system and the system knows that it is flowing a dTTP at this particular time and that is why it can associate this signal within A yeah.

So, we are not using any fluorescently label dNTPS, so this is a simple chemistry. Now, what will happen in the next, so we had flown in an A and the next sequence next target base is G they are not complementary. So, it will not bind it will get washed off then again. So, likewise

you will build up the sequence. Now, what will happen at this step? Supposingly we are flowing a dTTP, what will happen?

Student: (Refer Time: 29:58).

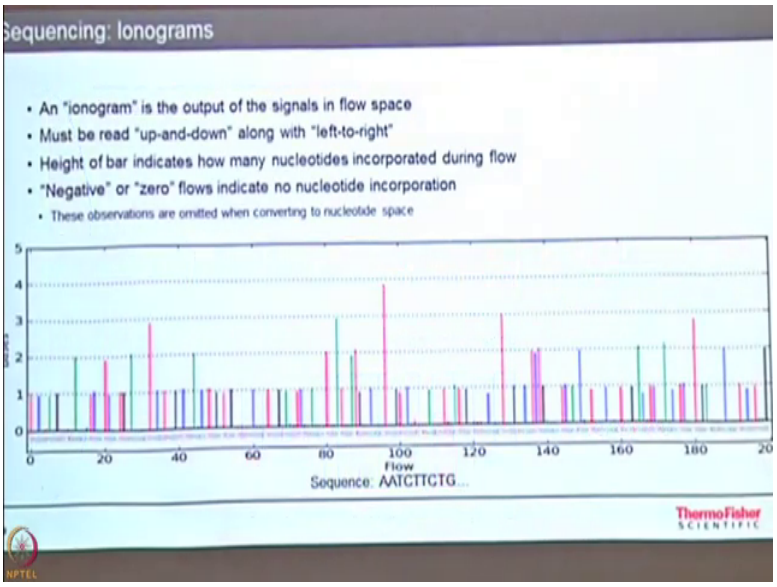
Sorry.

Student: (Refer Time: 30:02).

Yeah. Since there are 2 as in a stretch and there is nothing stopping it its a growing chain yeah. So, you will have double the signal yeah. So, there are double molecules double number of hydrogen ion release which will bring up it double change in the voltage and that is how it is being recorded as signal which is double and that is why you can tell that they were 2 consecutive a s in the sequence.

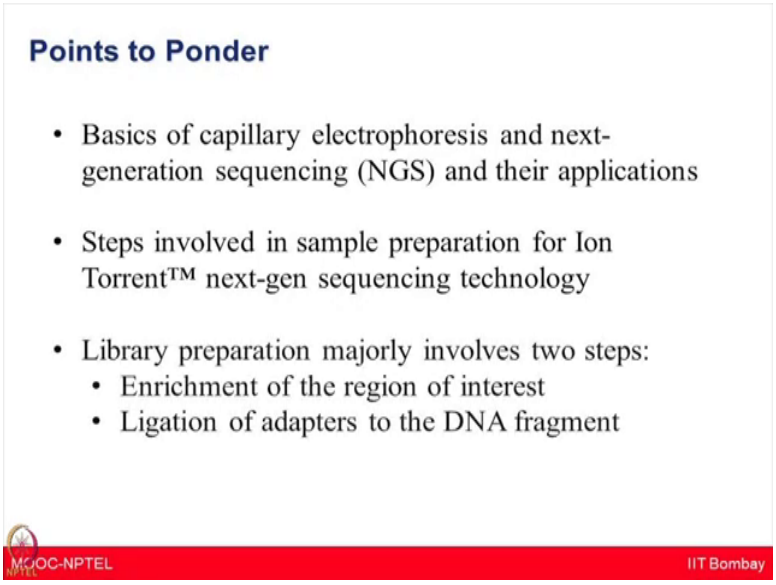
So, then this by adding the these dNTP sequentially finally, the system is being is creating a sequence which is finally, being read is an ionogram.

(Refer Slide Time: 30:41)

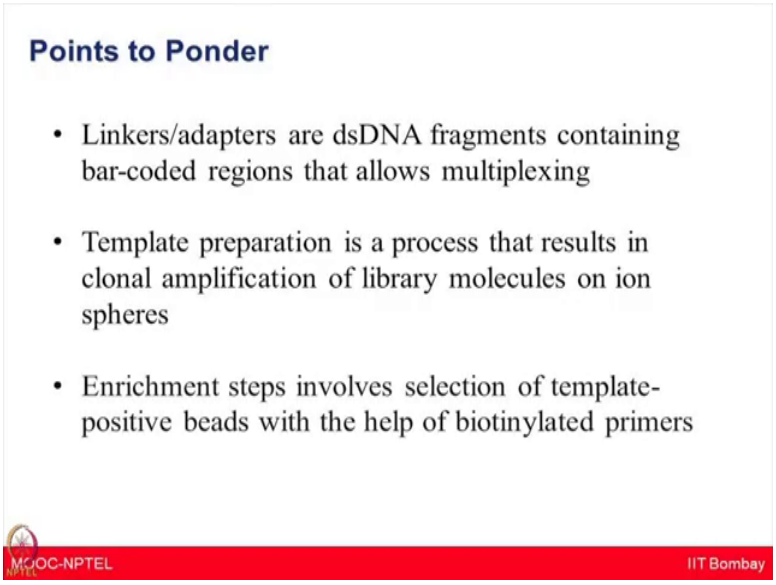


So, this ionogram is again being read by the sequencer, this has to be read from left to right and for the number of bases. So, whether you get a signal or not that will define the sequence of the dNTPs, the intensity of the signal will define how many of those dNTP how many of those As or Ts were in a row.

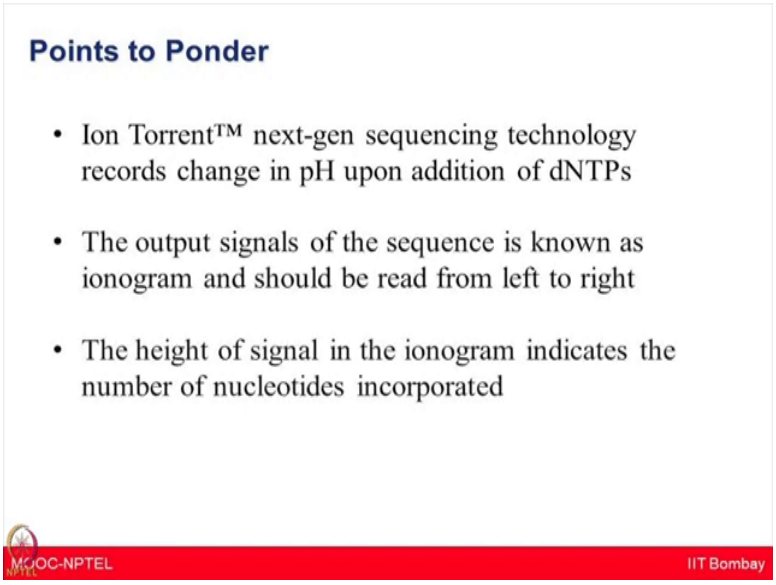
(Refer Slide Time: 31:11)



(Refer Slide Time: 31:21)



(Refer Slide Time: 31:29)



So, hope you got a glimpse of next generation sequencing technologies which has really revolutionized the field of genomics you were given the basics and some possible applications using this technology platform and when a specific technology of ion torrent was also illustrated and discussed. These concepts will be again covered in more detail in the following lectures when we have invited some more industry experts representing different technology platforms who will also provide you the basics and the current status of these technologies. So, see you in the next lecture.

Thank you.