Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis

Prof. Aritri Bir

Dr. B.C. Roy Multi-Speciality Medical Research Centre

Indian Institute of Technology Kharagpur

Lecture 51 : Serial Analysis of Gene Expression

Namaskar. Welcome back to the NPTEL lecture series Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis. We are at the eleventh week where we are going to discuss molecular diagnostics in medicine. For the first class, I am going to discuss a new technique which is known as Serial Analysis of Gene Expression. In this class, the basic concepts which we are going to cover regarding the method that is Serial Analysis of Gene Expression is the basic principle, the workflow process or step by step how the Serial Analysis of Gene Expression is occurring and then finally, the application. So, now Serial Analysis of Gene Expression is one technology which basically is helping to identify the transcripts not only to identify, but rather quantify the transcripts.

So, what is the purpose that in a cell suppose there are in a tissue there are multiple cells. Now different cell populations are not behaving in a same manner all the time. Suppose in a hypoxic condition, the brain cells are not behaving like the cells of heart, like the cells of liver. Similarly in brain, neurons are not behaving like the astrocytes or the microglia.

So in the cell in a tissue different types of cells can behave differently not only that cells do not a common pool of cells do not express similar type of mRNA in all the condition. In a hypoxic condition, the neuronal cell is behaving and expressing certain particular proteins via mRNA which might not be present during the normoxic condition or during some stress cells might behave differently because they express some different proteins. So in a cell the genome the whole part of genome are not expressed in a specific phase of time. So the gene expression is variable in terms of tissue to tissue, cell to cell even in a single cell different stressful condition, different physiological condition, different pathological conditions they modulate the gene expression in different way. So basically if we want to study that in this specific condition or in this physiological condition specific physiological condition what are the genes which are expressed in a cell for that what we need to do? We need to do that pool of proteins in our cell.

Those proteins are basically coming from the mRNA. So if we can check at a specific point what are the mRNA which are being expressed then we can study the gene expression profile of a cell in a specific phase or a specific time frame that is the purpose of serial analysis of gene expression. Only we are analyzing the gene expression based on certain tags. So there are specific tags which are tagged to each mRNA of that cell transcriptomic pool. So in a transcriptomic pool of a cell from a specific time frame if there are 10 different types of mRNA then those 10 different types of mRNA will be tagged via 10 different types of tag and those tag will give us the idea of gene expression.

So basically based on the tag we can study the transcriptomes and what we do we basically form we basically convert the mRNA to cDNA and after that that cDNA is sequenced. So what we are getting is a snapshot of transcriptome and after that we can check the abundance of each mRNA. So this gives us an idea of global gene expression profile of a particular type of cell or tissue from a specific time frame not only that. It helps to identify a set of specific gene in a specific cellular condition and that can be compared between two sets of cells. A cell a pool of cell which is stressed versus a pool of cell which is not stressed.

A pool of cell which is in a physiological condition suppose express a set of protein and we can compare with to the malignant version of that cell to the diseased version of that cell. So there can be a comparative transcriptomic study. So let us see how this serial analysis of gene expression or SAGE can be done in step by step. So these are the steps we are going to discuss. So here you can see a cDNA cDNA is formed from mRNA after that those cDNA are basically tagged those tags are isolated and they are amplified then those amplified tags are ligated to some bacterial cell then they are sequenced and those sequencer studied.

I think I am a very I have given a very gross idea. So let us go in depth so that we can study the individual steps in details. So the very first step is isolation of mRNA and from that cDNA synthesis. Now isolation of nucleic acid isolation of mRNA or DNA I think at the end of 10th week it is very much clear to all of you because based on that we have discussed a lot in the previous weeks. So this I am not going to discuss how to isolate the mRNA, but from mRNA to cDNA synthesis there is something special.

So if you remember at the end of mRNA there is a poly A tail and for cDNA synthesis in reverse transcription process we use one oligo DT primer which is having multiple T's which basically complementarily bind with this poly A tail well and good, but there is something special in this reverse transcription. Here the primer is basically biotinylated. So this is our biotin molecule which is attached to this oligo DT primer during reverse

transcription. Now I want to highlight a very special character of this biotin. Biotin can avidly bind with streptavidin and based on this biotin streptavidin attachment or binding we can isolate multiple types of molecules like if a base is attached to streptavidin and the molecule is attached to biotin because of these binding between biotin and streptavidin we can isolate the molecules that is our target here.

So here while synthesizing the cDNA we are using one oligo DT primer which has biotin at its end. After that we are removing the RNA by treating it with RNA's H and there is synthesis of double strand cDNA which has in one strand one biotin molecule. After that we are going to isolate this cDNA via capturing them over these streptavidin beads. So this is our streptavidin beads. These blue colors are basically our streptavidin beads and these streptavidin beads can bind the biotin attached to the cDNA.

Now what happens after this capturing of the cDNA these cDNA's are cleaved by anchoring enzyme. Anchoring enzyme is a type of restriction endonuclease. Why we have given this specific name because in sage there are two types of restriction endonucleases which are used. One type is anchoring enzyme the second time I am going to discuss it later. So what happens these streptavidin bound cDNA are now treated with restriction endonuclease giving rise to the restriction sites which is a common one sticky end restriction site that is GTAC.

So there are multiple types of mRNA. So for one mRNA this sequence is different for another type of mRNA this sequence is different, but for all types of mRNA which is common the end the sticky end is common in all the mRNA which are attached to the streptavidin beads. Now the part which is cleaved up by restriction endonuclease which is not attached to the streptavidin bead is basically washed off. So what is the thing that here is our streptavidin bead based on that biotin attached cDNA fragment is there and we have given restriction endonuclease cut over this. So this part which is not attached to the streptavidin bead is basically washed off.

So what we have now a pool of cDNA with a common end common restriction site GTAC treated by the anchoring restriction endonuclease enzyme which is attached over the streptavidin bead. Then we are dividing this pool of mRNA into two halves why we are going to discuss it fine. So finally, what happen those streptavidin beads attached cDNA pool are divided in two halves. Now in these two halves what we attach is one linker or adapter those are oligonucleotide probes of 15 base pair long. Now suppose this is one pool and this is another pool remember we have divided it into two pools and what is the difference between these two pools in one pool oligonucleotide probe is having a different sequence of primer and in other it is having another different sequence of primer.

Let us see what are these. So this oligonucleotide probes are having specific characteristics these are not some random oligonucleotide probe. These all these linkers or adapters or oligonucleotide probes which are 15 base pair long they must have these three characteristics. What are these? It must contain an A E cut site this is A E means anchoring enzyme. So, in the oligonucleotide sequence the restriction site for the anchoring enzyme which we have used previously that is GTAC must be present.

Then there should be another recognition site for another different restriction endonuclease and that restriction endonuclease is known as tagging enzyme. So, the first type of restriction endonuclease which we have used is basically anchoring restriction endonuclease and the second type of restriction endonuclease is tagging endonuclease. In the oligonucleotide probe the restriction site for both of this restriction endonuclease must be present. Along with that it the oligonucleotide probes must have one short primer sequence and that short primer sequence is different in these two pools. So, one pool of oligonucleotide is having a primer sequence consider this as A and another is having a primer sequence which is B.

So, this is the purpose of dividing it into two groups and why these two types of primer sequence are needed let us see fine. So, what we have done you can see that here we have attached one linker that is linker A and in this pool we have used another linker that is linker B. In both of the linkers what are the common things it has this two restriction endonuclease site for two different restriction endonuclease that is anchoring enzyme and tagging enzyme and in A it is having adapter A primary primer sequence and in B it is having adapter B primary sequence. So, the primer sequence are basically different. After that these attached septavidin attached tags they are treated with the tagging restriction endonuclease.

So, here you can see here is another type of tagging endonuclease which is NLA 3 is one example. So, both of this pool are basically treated with tagging enzyme. Now again these tagging restriction endonuclease has a special characteristics. What is the special characteristics? That these sequence this restriction endonuclease sorry suppose the sequence is example suppose I am giving the sequences A B C D fine. So, this tagging enzyme it does not cut over this restriction site.

It identifies the restriction site and it cuts 15 base pair downstream to the restriction site fine. So, in this tag what we are having here you can see a septavidin attached sequence with the adapter. Now, identifying this restriction endonuclease site it basically cuts downstream towards the actual mRNA sequence fine. So, in our hand what we are getting after treating with tagging enzyme. So, the tagging enzyme if you remember it was around 15 to 17 maximum base pair long.

And the tagging enzyme cuts the oligonucleotide probe is 15 to 17 base pair long and it cuts downstream to 15 base pair. So, basically it is actually separating cutting it is basically cutting the cDNA in such a way that it is separating the oligonucleotide probe from the bound streptavidin part. So, here you can see this is the part which we are getting. Similarly in this pool in the pool B we are getting this part. So, the part which was bound to the streptavidin is basically cut off from this oligonucleotide probe.

Now what happens here we are getting sticky ends and that sticky ends are basically repaired by T 4 DNA polymerase to produce blunt end tags. So, this part is basically repaired by T 4 DNA polymerase fine. So, these are the two different types of tags which we are getting. After that these two portions are mixed and ligated. Ligated how? In tail to tail orientation in such a way that one tag which is having primer sequence at one end A and another tag which is having primer sequence B at one end they are staying opposite to each other at the two ends of the mixture of the combination whereas, the in between nucleotide sequences are joined and that is known as concatomers.

So, what are the concatomers? Concatomers are similar type of DNA sequences double stranded DNA sequences which are attached in tail to tail orientation in a series. So, if the pool is like that A 1, A 2, A 3 these double stranded DNA are attached in tail to tail orientation in a series and that is known as concatomers. So, what we are doing? We are basically first creating a dye tag. So, initially two tags are basically joined to form a dye tag a tag which is having A at the end and B at another end. So, after this dye tag formation this dye tags are amplified via PCR process and for PCR if you remember we use two types of primer one is forward primer one is backward primer.

Here the forward primer sequence is our primer A sequence which has been used in the linker A and the backward primer sequence is the primer sequence B which is used in linker B. So, this is the purpose of dividing that pool of streptavidin attached cDNA into two parts and tagging them with linkers which are having two different types of primer sequence. I think it is clear now. So, this dye tags are now amplified via PCR we are having multiple dye tags now two different dye tags and now are joined by to form concatomers and those concatomers once again are treated with the anchoring enzyme. So, remember we have anchoring enzyme restriction site in this oligonucleotide probe.

So, in the dye tag we are having anchoring sequences at the both end. So, after treating it with anchoring enzymes what we are doing we are basically removing the oligonucleotide probe part or the linkers part. So, in hand what we have we have dye tags dye tags which does not have any oligonucleotide probe or linkers external sequences which have been attached. So, these are the part of our isolated mRNA sequence derived cDNA sequence. So, again from one mRNA we are basically giving rise to two pool fine one is this one double stranded cDNA sequence another is this one

double stranded cDNA sequence and they are joined and from two different pools.

Now these dye tags once again are joined to form concatomers. So, multiple dye tags are now joined in a series and those concatomers are now introduced into plasmid for cloning. So, here you can see from one mRNA what we are getting we are getting two dye tags those dye tags are joined by concatenation and then they are transferred into rather transfected into plasmids for cloning. So, this is the cloning which we are getting from the concatomer transfection into the plasmids. So, these this is our mRNA species which is coming from one mRNA.

Similarly, an another mRNA species is giving rise to these another mRNA species is giving rise to these. So, suppose in a cell there are in stressed condition there are these A B C proteins which are expressed. So, for those proteins we are having that many amount of mRNA that many numbers of mRNA. Suppose the protein A is expressed in highest number suppose there are 40 mRNA. Similarly, the protein B is expressed next in suppose 30 mRNA and this protein C is basically the expression is reduced.

So, it is considered it is having 2 mRNA. Now from this 40 mRNA which are getting double tags. So, there will be 80 dye tag 30 dye tags for B and similarly 4 dye tags for C and then via the computer based software we can analyze these different types of dye tags and quantify them. So, the number of dye tags present is basically giving the idea of number of mRNA in that specific isolated transcriptome pool and this is how we are getting the idea which protein is basically expressed in in how much abundance fine. So, what we can do in different condition we can study how a cell is expressing it is genes via assessing the mRNA pool via quantifying the mRNA pool. We can compare between a normal cell and a cancer cells proteomics or cancer cells gene expression transcriptomic pool.

We can compare between different types of diseased condition with the normal condition even in physiological condition. Suppose we are comparing and cells coming from young, coming from adult and coming from aged population and we can quantify their transcriptomes and can check how this aging process is basically affecting the cellular gene expression. Here you can see once again that suppose there are 3 types of mRNA species mRNA 1, 2, 3 and they are cloned and we are getting this clone. So, next what we will do we are sequencing them, we are getting the sequence of this dye tags then using the blast services we are basically checking this sequence is denoting which protein or which mRNA. Suppose this sequence is giving indication for mRNA tag 1 of species 1 which gives rise to a protein A.

So, what we can say this proteins are over expressed or expressed in reduced amount in a specific condition. So, this is how we can analyze and quantify different types of

transcriptomic pool and their protein expression. So, at the end what we can do with SAGE we can do a global mRNA profile of a cell at a specific condition and that global profiling of the gene expression we are doing it from mRNA. And how we were we are doing it we are creating specific tags for each mRNA and then we are basically counting those tags and this is how we are quantifying each mRNA or how many mRNA is there and then we are assessing the gene expression. So, this is all about serial analysis of gene expression these are my references. Thank you all and see you in the next class.