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 14 : Genome Editing Technologies I

Namaskar. Welcome back to the NPTEL lecture series, where we are going to discuss the Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis. In the weeks of tools for molecular diagnostics and molecular biology and gene expression analysis, we are now going to start different types of gene editing tools. So, the concepts which we are going to cover here is basically the mechanism, how the gene editing occurs, different tools, what are the technologies available by which we can do genome editing and finally, we are going to discuss about site specific nucleases which causes which cause site specific genome editing. So, genome editing or sometimes it is called Genome Engineering as well, where DNA is inserted a segment of DNA or some sequence is inserted or deleted from a DNA or within the DNA sequence a specific fragment or segment is modified or even can be replaced by a new one that is genome editing and that is solely occurring in a living organism. Now, genome editing we have discussed a bit about the restriction endonucleases by which we can do genome editing via recombinant DNA technology, but the problem with the these are these recombinant DNA technology is the very primitive where the genome editing actually has began.

But these early genetic engineering techniques they are random, random in the sense the sequence which is modified or inserted or replaced the cutting is random. Whereas, in case of very specific editing of genome which we say site directed endonuclease or site directed editing where the site directed endonucleases help they cut in a very specific site. So, modern genome editing basically targets the insertion to a very specific location and these nucleases help by creating specific double bond double bond breaks in a specific desired location and how that break is repaired by using cells own endogenous mechanisms of repair that is by homologous recombin recombination and non homologous end joining. So, this endonucleases site directed or site specific endonucleases they cut in a very specific site, but that cut is repaired by cells own repair mechanism.

Now in these two lectures sessions we are going to discuss about different types of

endonucleases like zinc finger nucleus, mega nucleus, talons and very popular is CRISPR. CRISPR about which we have discussed a bit in our very first session. So, let us move on. Now before going directly to the nucleases let us discuss about the double strand break repair mechanisms. Now if you remember in the first weeks of the DNA repair class I told you that we were going to discuss the double strand break repair mechanisms in further classes.

So, here we are. So, as we discussed that there are two types of double strand break repair mechanism endogenously present in our body. One is homology directed repair which is definitely the process is based on homologous recombination where the exchange of DNA materials between the homologous chromosomes occurring. Now in case of genome editing how this homologous recombination is exploited? In genome editing or in genetic engineering a homologous piece of DNA is constructed and that construct is basically taking part in homologous recombination or repair. So, let us see how this double strand break repair by homologous recombination occurs.

So, as you can see that in the double strand break. So, here is our double strand break. Now this double strand break is basically converted by a converted by the exonucleases to a gap. So, you can see the break is converted to a gap here. Now this exonucleases along with creating gap they create 3 prime overhand.

So, this exonucleases basically cut the 5 prime end as well as the 3 prime end, but the excision in the or the resection rather we say it as end resection. This end resection is more in the 5 prime end. So, definitely there is a 3 prime 3 prime overhand which is created. So, this is our 3 prime overhand. So, here you can see this one is 1 3 prime overhand in the other strand this is our another 3 prime overhand.

Now these overhangs basically form here you can see these overhangs form a nucleoprotein complex. So, this is our nucleoprotein complex which is comprising of DNA repair protein that is rat 51 along with that there is nucleases and also polymerases. Now the function of this rat 51 is to find out the homologous segment of the DNA and also a protein which is RPA protein it binds to the these ends this overhangs. So, that they do not join with the complementary strands also these RPA protein they these proteins protect these overhangs from the digestion of the exonucleases. Now when the nucleoprotein complexes have found the homologous chromosome in the form of bodies own homologous chromosome or in case of genetic genome editing the engineered DNA construct when it is found there is strand invasion.

So, what is strand invasion? So, here you can see this 3 prime end this 3 prime overhand has found the homologous chromosome and it has replaced the homologous chromosomes complementary strand and inserted or rather aligned to is associated

complementary sequence in the homologous chromosome. So, basically the complementary strand sequence of the homologous chromosome is displaced forming a displacement loop. So, this is our displacement loop this part is our displacement loop. So, what you can see the 3 prime end has already invaded. So, that is the strand invasion after invasion polymerase comes in action it extends the strand of the DNA.

So, basically the segment which has been lost or lost due to DNA break or double strand break is now getting synthesized via DNA polymerase. One very important feature here is a holiday junction or holiday intermediate. Now this holiday intermediate is this. So, this is basically the crossover formed by 2 homologous chromosomes complementary strand. Now based on this holiday junctions activity the homology directed repair or DNA double strand break repair via homologous recombination can be of 2 types.

So, one is double strand break repair. So, let us see what is double strand break repair? In case of double strand break repair there is formation of 2 holiday junction. So, what is that? Here you can see these 3 prime overhang has created one holiday junction which is this along with that these has also created another holiday junction. So, there are 2 consecutive holiday junctions which are formed in double strand break repair. Now based on the cut off or the separation or the method of these double 2 holiday junctions there is crossover product or there is no crossover product, but the repair is done.

Now how it is done? So, you can see here there is no crossover product and here that is crossover product. Now what is crossover product? You can see in this region this was our the blue is our parent initial DNA which has been injured by the double strand break. So, here you can see this part is being synthesized, but this is the segment is coming from the homologous chromosome. Similarly in the homologous chromosome this part has come from the the parent or the previous DNA. So, basically there can be crossover or here you can see there is no crossover the strand is intact here the strand is intact there is no crossover.

Similarly in the homologous chromosome also the strand is as previous not it is not exchanging the materials with each other the homologous chromosomes are not exchanging the material with each other. Now how is that happening? In case of crossover product what happens the 2 holiday junctions are broken in such a way that one is one cut is on the crossing strand and the other cut is on the non crossing strand. So, if the cross is like this one cut is here another cut is here in that case what we get is the crossover product. In case of a cut between in in the region of these 2 let me delete it you will get it . So, let me delete this.

So, in case of non crossover product the cut is always on the crossing strands. So, this is how in homologous recombination double strand break repair mechanism we get 2

holiday junctions along with that based on the resolution of the holiday junctions we get a non crossover product as well as a crossover product. Now in case of double strand break repair there is formation of 2 holiday junction. The other mechanism that is strand displacement annealing in case of strand displacement annealing the initial phases are same that is the formation of holiday junction where the branch migration is is occurring the strand invasion is there DNA polymerase is filling up the lost nucleotides, but after that there is no formation of 2 holiday junction there is one single holiday junction formed in case of strand displacement annealing pathway. So, what happens here you can see it is basically very simple that the 3 prime overhang here has been invaded in the homologous complementary strand synthesized the lost segment after that these segment is basically forming the base forming the base pair with the other 3 prime end.

So, base pairing is occurring between the over other 3 prime end overhang and the newly formed strand and then the DNA synthesis occurring because this part is also synthesized based on this parent strand this part is synthesized by the DNA polymerase and ligated by ligase. So, what we are getting is a non crossover product once again in case of strand displacement annealing method. So, this is the mechanism of homologous recombination. Now in case of genome editing instead of the homologous chromosome we are constructing a DNA sequence which is similar to the complementary strand sequence or the homologous sequence. Now another method is non homologous enjoining method that is NHEJ.

Now this NHEJ is where homologous chromosome sequence is not required and this is very simple to remember that of course, there is a double strand break here you can see the double strand break. Now this double strand break is identified by multiple protein complexes like DNA PKC or protein kinase C proteins like KU 80 and 70 which forms the dimer hetero dimer they identify this double strand break binding the double strand break region and trims of the overhang. So, here you can see there is no overhang after that with the help of this ligase those two strands the broken double strands are basically bridged. So, the ends are bound and bridged the gap is repaired by the non homologous enjoining method. So, it is very easy to remember.

Now in case of genome editing different types of nucleases are directed in that specific double strand breaks or they create such double strand breaks and insert or delete or modify or replace a segment of DNA after that the repair method because there is a double strand break at the end like this suppose this is our double strand DNA now the nucleases site directed nucleases has cut in this specific region. So, what we will get a double strand break now consider a new DNA segment is inserted here, but the double strand break remains and that is being repaired by either homologous recombination or non homologous end joining. Now there are multiple this type of nucleases one very common is mega nucleus by the name it is evident mega means there is a very long

recognition sequences and because it is cutting in between the strand definitely these are endonucleases now these are commonly found in microbial species. Now based on the long recognition sequence which is more than 14 base pair it is not always possible to find out such mega nucleases those are not very common in nature. So, what we need to do we need to construct mega nucleases based on our requirement.

So, if we have chosen one target sequence which we want to modify for that we need a mega nucleus, but that is not naturally available because of the mega nucleases recognition sequence is not very common. So, how this synthetic mega nucleases are formed they can be formed by mutagenesis and high throughput screening methods. Now two different mega nucleases can be fused fused in such a way that they are recognition sequences are merged or modified in such a way that they can identify a new sequence. Generally rationally designed mega nucleases are such where the target amino acid sequence or the recognition sequence can be modified there can be some computer models based on which we can generate synthetic mega nucleases. Now one very common example is entron encoded endonuclease this entron encoded endonuclease is a mega nucleus obtained from the yeast saccharomyces cerevisis.

Now you can see the sequence the recognition sequence in this entron encoded endonuclease is very long it is an 18 base pair sequence. Now similar such sequence in nature might not be available for that reason for that reason this endonucleases are very specific. If there is these 18 sequences then only it can identify by, but the the disadvantage is it is not always available. So, it has to be genetically engineered. Now this entron encoded endonuclease it is you can see this is the 18 base pair long.

So, it can be synthesized via genetic engineering as well. Now xeroderma pigmentosa if you remember we have discussed in the DNA repair mechanisms class is one such disease originated due to the defect in the nucleotide excision repair mechanism and that is related to the gene xpc. So, recently one mega nucleus is available to alter this xpc genes expression. So, the xpc gene is basically silenced or the function is degraded by the mega nucleus. So, that the manifestation of the disease can be modified or rather reversed.

So, next moving on to another mega nucleus group. Now another endonuclease group site directed endonuclease group. Now mega nucleus amongst the site directed endonuclease here the recognition sequence is very specific. Whereas, these two endonucleases zinc finger nucleus and TALEN which is transcription activator like effector nucleus they have non specific DNA catalytic DNA cleaving catalytic domain. Now what is that? Basically these endonucleases we need to search such endonucleases when we are conducting the genetic engineering or modification we need to search some those endonucleases whose recognition site and the cleaving sites are separate or

different.

So, that suppose this endonuclease is having the recognition site here and cleaving site here. So, basically what we can do? We can cut off this part. So, the cleaving sequence now is an orphaned one which is which has no direction. So, along with that we are adding a recognition sequence of our choice. So, basically what we are doing? We are fusing a non specific DNA catalytic domain with a specific DNA sequence which can recognize peptides such as zinc fingers in case of zinc finger nucleus and TAL effectors or transcription activator like effectors.

These are the specific sequence or protein or the DNA sequence which are identified by the proteins and based on that we are choosing the recognition peptides. So, again endonucleases should be such that the recognition site and cleaving sites are different because if those are same we cannot cut it off. Then after separation it is linked to the recognition sequence of our choice that is a zinc finger motif in case of zinc finger nucleases or a TAL effectors in case of TALENs or transcription activator like effector nucleases. Now, let us discuss one by one. Zinc finger nucleus is based on the protein motif zinc finger domain.

So, basically this protein motif or a 3D structural motif is proteins hierarchical structure where the structure is forming finger like protrusions. So, here you can see these are our finger like protrusion and they are stabilized by one or more zinc ion that is why it is called zinc finger motif. Now these type of motifs are mostly present in different types of transcription factors. C So, basically they are present in this zinc finger motifs are present in the proteins that region which take part in protein DNA interaction. So, if you consider this as a whole protein and this is our DNA.

So, this is the site which takes part in DNA protein interaction and here the zinc finger motif in case of transcription factors the common zinc finger motifs location is in the protein DNA interaction site. Now the C terminal part of each finger is the specific recognition site. Now each finger the specific recognition site is formed by 3 base pairs around 3 base pairs, but if similar such fingers are combined what will happen you can see if these fingers are combined you are getting a 12 base pair long sequence. So, in that way the more zinc fingers are combined the motifs are combined the long recognition sequence can be formed. Now how this zinc finger nucleases are formed based on the zinc finger motifs.

So, what is done 2 DNA binding protein. So, these 2 DNA binding protein each containing 3 to 6 zinc finger they are combined along with that. So, this is the protein or the recognition sequence in case of the nucleus. So, where from the catalytic part is coming from another endonuclease the commonest one is Foc1 nucleus which is

extracted from the flavobacteriums. So, the catalytic part is joined here endonucleuses catalytic part is joined here. Now the very exclusive feature of this Foc1 endonuclease is that it is very much active when it dimerizes when those 2 basically join.

So, this is the reason the 2 different binding proteins are formed. So, till these 2 different proteins are separate the Foc1 endonuclease is inactive whereas, when these 2 new suppose this is our DNA here the 1 DNA binding recognition sequence along with the this is our Foc1 endonuclease 1 is joined and it is also identified by the another DNA binding protein on zinc finger protein the Foc1 2 has joined here. So, basically they are close to each other and can dimerize and get activated. So, this is the mechanism how zinc finger nucleases are engineered to create double strand break and edit the genome. Next is talons or transcription activator like effector nucleases.

Now, remember talons are artificial restriction endonuclease definitely it is causing double strand break within the DNA. So, that is the endonuclease and the recognition site which is attached to this endonuclease is basically based on the tail domains tail domains or transcription activator like effector domains they are basically the DNA binding proteins present in the plant pathogenic bacteria xanthomonas species. Now, this tail effectors are basically they are binding to promoter region of the DNA which induces the transcription of certain genes. Now, this tail regions is basically comprised of 33 or 34 amino acids tandem repeats in cluster. So, multiple tandem repeats present in a cluster that is our tail domain and along with that a nucleus is attached or it is commonly again here the nucleus is the Foc1 endonuclease.

Now, what is the speciality of this tandem repeats? This cluster is basically 30 even if it is 33 or 34 amino acid long it has in its 12th and 13th position hyper variable region. So, basically the amino acids of 12th and 13th position are variable and based on that the different recognition sites are formed. So, different tail can identify or tall effector can identify different types of sequences over a DNA based on the variability of this 12th and 13th positions amino acid and these are so, called the RVDs or the repeat variable dye residues. Again they are causing the double strand break commonly tail ends are the double strand break created by tail ends are repaired by non homologous repair mechanism. So, here you can see that this is the tail end structure here in this 12th and 13th position the amino acids can be different.

So, they are hyper variable and now based on that they bind this sequence can bind to species these are our target sequence target DNA and the dimer can bind here and remember the beauty of palindromic sequence. So, one if the 2 DNA binding proteins which we have discussed in the zinc finger nucleus is same here. So, it has 2 segments or 2 domains. So, one DNA recognition sites bound with nucleus this is our DNA recognition site bound with nucleus this is our another DNA recognition site bound with

another nucleus for 1 nucleus.

So, because it is palindromic segment. So, they can bind to the complementary strands in such a way that this 2 segments of folk into nucleus they come near to each other and can dimerize and can get activated. So, this is how tail ends work. So, in this segments we have discussed the zinc finger nucleus and the transcription activator like effector nucleases. Also we have discussed the homologous repair mechanism and non homologous repair mechanism. In the next class we are going to discuss the very well versed or well versed or popular these days is the CRISPR cache mechanism or CRISPR based genetic genome editing technology.

These are my references. Thank you and definitely see you in the next class.