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 15 : Genome Editing Technologies II

Namaskar! Welcome back to the NPTEL lecture series of Comprehensive Molecular Diagnostics and Gene Expression Analysis. Now, we were discussing different tools for the diagnostics and gene expression. And in the last class we have started the genome editing tools or technologies where we have discussed the different site specific endonucleases that is mega nucleus, then zinc finger nucleus, then talons and now we are at CRISPR. So, the concepts in this class we are going to cover are the very basic concept of what is CRISPR, origin of CRISPR in terms of the bacterial immunity system, classification or different types of CRISPR, mechanism of CRISPR mediated gene editing, how they are applied in the clinical diagnostic field and finally, like all other methods CRISPR has some limitations. So, we are going to discuss all this one by one. So, CRISPR is basically clustered c regularly interspaced short palindromic repeats.

So, this is our CRISPR. So, basically CRISPR is a DNA sequence which is commonly found in bacteria and that DNA sequence has some very specific characteristics. So, there are short palindromic sequences present in repeats. So, here you can see this blues are those this blues or light blue colors are those short repeats of palindromic sequences.

They are clustered and regularly interspaced. So, this interspaced are basically called as spacer. So, these are our spacers which are causing the regular interspacing. Now, these CRISPR or clustered regularly interspaced short palindromic repeats are associated with or CRISPR associated sequence. Now, Cas proteins are basically nucleases which can target and cleave specific sequence of a DNA in a specific manner.

Now, let us know two terms one is spacer another is protospacer. So, spacers are these which are basically the sequences which are basically causing the regular interspacing between the palindromic sequences. Now, what are those spacers? Now, as I told you CRISPR has very much originated from bacterial immunity. So, here you can see bacterial immune system is having this CRISPR sequence and is targeting the attacking virus or some what you can say or some other DNA sequences foreign DNAs and protect the bacteria from those how. Basically, what is happening when there is an invasion by virus inside a bacteria the viral genome of very specific part of those viral genome are

incorporated inside the bacterias.

So, those are the spacers those foreign DNA or viral DNA which is incorporated in this CRISPR sequence are known as spacers. So, these spacers are nothing, but the foreign DNA which has once invaded the bacteria and definitely the foreign DNA has protospacers or a sequence associated with the spacers. Now, let us see how this bacterial immune system is working based on the CRISPRs mechanism. Now, here you can see suppose this is the Fudge virus is DNA and the Fudge virus has invaded the bacterial cell. So, here you can see the viral DNA is inside the bacterial body.

Now, there are certain CAS or CRISPR associated systems which can digest or act as nucleus actually. So, they can digest DNA sequences in specific site. So, one such CAS is CAS 1 another is CAS 2 this CAS 1 CAS 2 system is basically helping the bacteria to express the CRISPR based immunity. How? CAS 1 initially it is cleaving the viral DNA and choosing a very specific segment which is incorporated in this in between the 2 palindromic sequences as spacer. So, one such sequence is incorporated here.

Similar such sequence or some other sequence is incorporated here like that and they are joined with the bacterial DNA by recombination method. Now, it remains as it is till another attack by that same DNA virus is there or same foreign DNA is intruding the bacterial cell. Now, what happens the CRISPR system is expressed or transcribed by another CAS that is CAS 2 which is causing the expression of the CRISPR sequence. So, what we are getting expression in terms of when the genetic DNA sequence is expressed by transcription there is formation of RNA. So, what we are getting? Pre CR RNA CR RNA is basically CRISPR RNA.

Now, this pre CR RNA contains the spacers as well as the palindromic sequence. So, multiple spacers till date whatever foreign DNA has invaded and incorporated between those palindromic sequences they all are transcribed by the CAS 2 mechanism when there is a similar viral attack or DNA entry. Along with that what else is transcribed? So, in the CRISPR sequence definitely you can see there are multiple CAS's they are also transcribed. So, one such CAS is CAS 9. Now, CAS 9 is responsible for activating or rather cutting the intruder DNA.

So, basically CAS 9 is responsible the CAS 9 nucleus is responsible for degrading the attacking or foreign DNA, but how this CAS 9 is activated? CAS 9 is forming a complex with CR RNA. Now, CR RNA is formed from pre CR RNA. So, basically this is a type of post transcriptional modification. Pre CR RNA is forming CR RNA by CAS 3. Now, in some types of CRISPR, CRISPR has different types which we will read later.

In some CRISPR there is another type of sequence which is known as T-RA-CR RNA.

So, this T-RA-CR RNA is basically trans activating CR RNA. In some types of CRISPR these trans activating CR RNA present in some types they are absent. Consider here in this CRISPR there is this trans activating CR RNA present. So, a complex is formed with taking CR RNA, trans activating CR RNA and CAS 9.

So, these complex is responsible for digesting or dissolving the attacking foreign DNA, but how CAS 9 can differentiate between which sequence which DNA to digest. It can digest bacteria's own nucleic acid for that what is required is the proto spacer adjacent motif. Remember in the very beginning I told you there is spacer and there is proto spacer. So, spacer is the attacking viral DNA segment which is incorporated in the in the CRISPR sequence. And proto spacer adjacent motif is basically the sequence which is identified by the CR RNA.

So, CR RNA binds to the sequence of the foreign DNA in the region of PAM that is proto spacer adjacent motif. So, proto spacer adjacent motif is basically a 2 to 6 base pair DNA sequence immediately following the this is our target or foreign DNA sequence. Now, this PAM site is a component of definitely the foreign DNA it can be virus, it can be plasmid and not a part of the host DNA or host bacteria. So, that is how basically CAS 9 is differentiating the self DNA from the non self DNA. Now, here you can see this is our CR RNA which is binding to the this is our target DNA site binding to the target DNA and the identification is done due to the presence of PAM.

So, when the complex of CR RNA and CAS 9 along with trans activating CR RNA when this complex can locate a PAM there the CAS 9 gets attached and activated and digests. So, basically without PAM 9 no CAS 9 or no CAS can basically cleave the target DNA. Why I am saying this I will discuss in case of genome editing this is happening naturally in case of bacterial immune system. Now, based on the PAM sequence this complex bind to the target region and that PAM is basically identified by the CAS 9 nucleus. So, from this discussion what we are getting regarding different components of CRISPR we are going to discuss, but so, here once again in this diagram you can see.

So, this is our attacking Fudge DNA or viral DNA now it is getting incorporated. So, this is our this boxes are basically the spacers or the viral DNA segment of viral DNA which has been incorporated and this is our palindromic sequence this blacks are our palindromic sequence this is the palindromic sequence. So, you can see there are multiple colors why this multiple color representation is there it shows that in this CRISPR loci multiple types of spacers can be present which indicates different types of invasion by different types of Fudge virus. That means, in a CRISPR different types of gene sequence or nucleotide sequence can be incorporated which I will discuss in the future genome editing section. So, when there is transcription this pre CR RNA or

CRISPR RNA contains all of those foreign spacers or foreign DNAs along with the palindromic repeats and when it is modified by post transcriptional modification by CAS 3.

So, we are getting multiple types of CR RNA or CRISPR RNA that can identify different types of virus. So, RNA guided targeting is basically done by CRISPR. So, different CRISPR RNA is targeted towards different sequence different foreign DNA sequences. So, this is how bacteria is gathering immunity to different types of Fudge virus or attacking DNA or nucleic acid segments via the CRISPR system. Now, if we discuss the components of this CRISPR system by the discussion.

So, we have CAS or CRISPR associated system which is an enzyme or a nucleus then trans activating CRISPR RNA, CRISPR RNA and also protospacer adjacent motif or pramside. Now, CRISPR RNA consists of 2 domains. In the first domain which is in the 3 prime end it combines with the 5 prime terminal region of the trans activating CRISPR RNA and the second domain that is located near the 5 prime end it is basically complementary to the target specific sequence. So, if we see this model here you can see this is our CR RNA, this black is basically our trans activating CR RNA and this is our this blue is the Cas9 nucleus. So, CR RNA it has 2 domains with 1 domain it is attached to this domain hope you can see.

Suppose I am indicating with the blue. So, this part is complementary base pairing with the TRCR RNA and another segment this segment is complementary base pairing with the target DNA. So, this is our target DNA fine. So, this is clear. Now, similarly TRCR RNA has also 2 domains, 1 domain for binding with the CR RNA.

So, 1 domain is binding with the TRCR RNA another domain is basically identifying the Cas9 nucleus. Now, coming to the protospacer adjacent motif or PAM site. Now, PAM site is basically identified by the Cas9 and that is the specific what I can say that is the specific signal which is achieved or perceived by the Cas9 that where the genome editing should be done or the cut should be done. Now, the common Cas rather CRISPR associated system is Cas9 there are multiple other types. The common one Cas9 is basically originated from a bacteria Streptococcus pyogens apart from that there are multiple ones like Neisseria meningitidis, Streptococcus thermophylus.

So, there are different types of Cas which can be generated and why I am discussing this because different Cas can have different PAM site. For Cas9 a very common canonical PAM site is NGG, N can be any nucleobase followed by 2G or guanine residue. So, this is one target or signal which is identified by Cas9 and based on that it is located in the PAM site and based on that the Cas9 CR RNA and trans activating CRISPR RNA binds to the target sequence. Now, here are examples of other type of Cas9s originated from

other types of bacteria. So, NGG is the very canonical one from Streptococcus pyogens there can be NGAG, NGCG like that.

So, this is the sequence or PAM sequence based on that we can target specific CRISPR system to a specific DNA region. Now, SG RNA or single guide RNA are the combination of all these trans activating CR RNA and CR RNA in a system and we have discussed the associated Nobel prize in our very first class I think you remember it is a very popular one. So, a short nucleotide known as guide RNA is synthesized to perform the function of all these different types of RNA that is trans activating CR RNA and the CR RNA complex. So, this is a short nucleotide segment where the functions of TRC RNA and CR RNA are included. They recognize this guide RNA recognize the PAM sequence and also targets or guide the nucleus to the specific site for cutting.

So, this is how we are implementing the CRISPR system in genome editing. So, basically CRISPR system is available naturally it is present naturally in some bacteria, but based on that concept we are designing some targeting RNA or single guide RNA which can target a very specific DNA sequence can cut the sequence and can edit or modify that sequence. So, that is our CRISPR associated genome editing tools. Now, CRISPR can be of different classes and these classes are basically differentiated by the presence of signature gene. Signature gene here is the CRISPR associated system or the CASH system.

Now, the class I which contains which is comprised of type I CRISPR, type III CRISPR, type IV CRISPR this class I CRISPR is basically having several CASH proteins which forms the combined system and acts over the target DNA. Whereas, in case of class II they have this class II CRISPRs have only a single large component of CASH IX protein and in class II there is type II, type V, type VI. Now, here you can see this signature gene or the CASH are different in all this types like in type I there is CASH III, type III it is CASH X, type II CASH IX we have discussed already. Then C P F I like that the signature genes are different in different types of CRISPR. And what is the difference between class I and class II? Class II are having multiple CASH proteins combined where they are delivering the action whereas, in class II there is one single CASH protein which is conducting the function.

Now, CASH III I am giving example of CASH III which is present in type I CRISPR it encodes a large protein which shows the features of helicase as a single stranded DNA stimulated ATPase activity. So, helicase is there which is dependent on the ATP driven energy ATP hydrolysis dependent energy and targeting single stranded DNA causing unwinding as well as digestion of the RNA-DNA duplex. Also this helicase shows endonuclease activity involved in cleavage of the targeting DNA. In case of CASH IX all the functions of the effect are complex like CRRA the complex which is causing

helicase endonuclease all are present along with the cleaving quality of the system. So, basically CASH IX is a multi domain protein the multiple domains are showing different types of action.

What are those? It contains CASH I and CASH II gene. So, basically digestion of the foreign DNA and incorporation are present in the CASH IX's domain. The genes for trans activating CR RNA present in the CASH domain. Then after the digestion how the CRISPR-CR RNA is dissociated from the target that sequence is also present. So, basically that separation is done by RNA's 3.

So, the action of RNA's 3 and TRCR RNA are also present. Now, this nucleus it has two domains one is RuVC like nucleus another is HNS nucleus domain. Now, these two domains basically identify now with the help of the cellular RNA's 3 and trans activating CR RNA the processing of pre CRISPR RNA can be done. Now, two nucleases domain are present in CASH IX multi domain complex. One is RuVC like nucleus another is HNH nucleus.

These two nucleus domain basically identify. So, if you consider this is the target DNA and this is the nucleus where this is our RuV domain and this is the HNH domain. They basically targets two specific sequences creates cut in two regions and this segment can be replaced or exchanged or deleted or here another new insert can be recruited. So, by this way CASH IX systems act and both of these nucleus are responsible for cleaving the target DNA. So, if we come to the experimental design where we are designing. Now, if we come to the experimental design where this CRISPR system is exploited in the genome editing technologies, what are the steps we need to do? First we need to identify the genomic target.

In the genomic target we need to identify the PAM site as well. So, here we have identified this target sequence where this TGG is our PAM sequence. Based on that the single guided RNA CASH IX complex or construction is formed. How it is formed? So, this is the CRISPR RNA with the target sequence, this red is the target sequence along with the hairpin loop. This is how the CRISPR RNA is basically transcribed.

They are incorporated all these CRISPR RNA sequences. So, the sequences can be multiple like this is one sequence, another sequence, another sequence all can be incorporated in a plasmid or a sequence can be multiplied like the first sequence can be multiplied in multiple copies and can be incorporated in a plasmid along with the CASH IX as well as the trans activating CR RNA. So, in a plasmid we are having all the CRISPR target sequence, the trans activating CR RNA sequence as well as the CASH IX and this plasmid can be used to transfect the host where we want to change the genome or edit the genome. So, this is how the experimental designing is conducted in case of

CRISPR associated genome editing. So, what are the different types of application in case of CRISPRs in clinical diagnostics? The very important one is diagnosis or detection of a genetic mutation based on CRISPR based assays.

What they do? They identify specific genetic mutations associated with different diseases, it can identify single nucleotide polymorphism. So, basically this CRISPR based system is actually acting as a biomarker detection portal. Also in case of infectious disease diagnosis CRISPR is very helpful, helpful while identifying the strains or presence of infectious agent by rapid detection kit, quantitative viral load can be assigned. Also antibiotic resistance can be checked via identifying different antibiotic resistance gene in bacterial population so that the antimicrobial treatment accordingly can be designed. CRISPR is also helpful in cancer diagnosis where different cancer associated mutations and alteration in the genomic DNA can be identified.

And presently CRISPR based assay are huge assays are hugely used in point of care testing and also multiplex diagnostics which we are going to discuss in the further classes. But again when we are talking about something new it has definitely some limitations. So, targeting efficiency in term of specificity as well as off target mutation. So, the PAM sequence or the PAM site is the target signal which is identified by the Cas9, but the CR RNA needs to identify a specific gene sequence. Now sometimes this target or the designing of the CR RNA can be not accurate or that complete for complete identical formation of complementary strand might not be possible.

In those case what happens an alternative or some wrong target can be digested or cleaved that is off target mutation or specificity can be modified when there is 1 or 2 nucleotide difference the CRISPR RNA might not identify that sequence. Then again the vector as discussed in just before that the vector containing the plasmid containing all the CRISPR RNA along with the other complexes the delivery can be challenging inside the host cell. Immunogenicity the because a foreign DNA is incorporated in the host there can be immunogenic responses. Then cell cycle dependency CRISPR binds the CRISPR associated system binds to the host system in specific cell phases of cell cycle. Now for those cells which are non dividing or those cells which are having very specific characteristics of cell division they might not get the opportunity to interact with the CRISPR RNA.

So, in those cases more research more modifications are required. So, in the summary we have discussed the very basic concept of CRISPR that is a DNA sequence where we are getting short palindromic repeats which are regularly interspaced by spacers which are originated from the invading foreign DNA. And along with that there is a CRISPR associated system which is generating nucleases which are targeted towards the foreign DNA causing cut over the foreign DNA. And finally, they can edit the genome by

creating a break and inserting or deleting or modifying the genome. The origin is very natural that has been discussed in bacterial immunity based on CRISPR system and also different classes of CRISPR based on different types of Cas protein we have discussed.

Then the applications and limitation of CRISPR that we are going to elaborate in our further classes, but the very basic clues we have given to you these are my references see you in the next class. Thank you.