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 20 : Epigenetics & DNA methylation analysis

Welcome back to the lecture series in NPTEL. So, today we are going to discuss epigenetics and related mechanism which can help to diagnose the different epigenetic changes. So, the concepts over which we are going to discuss are the basic concepts of epigenetic mechanism, then different detection methods. These are the epigenetic changes DNA methylation, histone modification or different non coding RNAs how those are detected or how the epigenetic changes are detected those mechanisms we are going to discuss today. So, epigenetics basically encompasses heritable structural and biochemical alterations.

Alterations at the level of chromatin without changing the DNA sequence remember even if there are changes of chromatin structure or methylation or demethylation of DNA etcetera, but the DNA sequence remains same. So, this is basically the changes they are occurring in such a way that can be inherited. So, these are heritable structural and biochemical alterations. So, this can manipulate various types of physiological and pathological processes.

It can modulate relevant gene expression via changing the accessibility of genetic codes to the chromatin locally as well as globally. So, different epigenetic changes the examples primarily are DNA methylation, histone modification and different regulation at the level of non coding RNAs. They can change the accessibility of the genetic codes in such a way that the genes whether those will be off or on can be decided. So, this is how epigenetics can regulate the expression of different genes. Coming to DNA methylation, DNA methylation basically involves the transfer of methyl group to the C 5 carbon 5 position of cytosine within CpG dinucleotide.

So, in the DNA structure there are islands of CpG. So, this CpG dinucleotides in this CpG island they are clustered in such a way that those regions are much vulnerable to DNA methylation or demethylation by the enzyme DNA methyl transferase. And this is how they can regulate the chromatin structure, how the gene transcription will be delivered can be decided by this DNA methylation. How if there is hyper methylation in

the CpG island that in the promoters region that can lead to transcriptional silencing whereas, hypomethylation can activate the gene transcription. So, they can modulate various type of biological processes sometimes rather in different way they can regulate different pathological process as well.

Like aberrant DNA methylation in promoter regions of tumor related gene, tumor suppressor in tumor oncogenes. So, this tumor related gene if there is aberrant DNA methylation that can be suppressed. So, if there is suppression of tumor suppressor gene or if there is activation of oncogenes that can lead to development of tumor. So, this is one such mechanism by which tumorigenesis can occur. Now the very common form of methylated DNA is 5 methyl cytosine.

It is abundantly distributed in eukaryotes apart from that there is N 6 methyl adenine, 4 methyl cytosines they are also found. Now to identify the DNA methylation pattern the gold standard is basically bisulfite treatment. Now sodium bisulfite in the bisulfite treatment based method sodium bisulfite is used what it does it deaminate the un methylated cytosine to uracil, but the methylated cytosine left as it is. So, what happens in the DNA structure in the CPG island suppose those cytosines which are un methylated they are converted to uracil. So, whatever cytosines are remaining in those region are basically methylated cytosine.

Now bisulfite treatment by bisulfite treatment locus specific analysis can be done in which region the methylation or hyper methylation or hypomethylation has occurred that can be detected by bisulfite treatment. How bisulfite PCR product sequence? So, after bisulfite treatment if the product is sequence then we can detect the region, but before that before sequencing what is required amplification of the region by PCR processing. Next methylation specific PCR or MSP can also be done. So, what it what it does basically it analyses the methylation pattern specifically in the CPG island. What happens the bisulfite converted DNA is used at as template.

So, in the template there are cytosines which are only methylated and they are amplified by PCR then specific primers are used for either methylated or un methylated DNA template and they are identified the locus is identified. Similarly pyro sequencing can be done. Now pyro sequencing basically real time monitoring of the incorporation of nucleotide. So, when there is sequencing there is basically real time incorporation of nucleotide and that can be quantified. So, this is how the regions are identified where hypomethylation or hyper methylation is done.

Pyro sequencing is going to be discussed in details pyro sequencing and different types of sequencing mechanisms will be discussed in detail in the next few classes fine. So, apart from the locus specific identification of methylation genome wide DNA

methylation profiling can also be done. Those can be done by hybridization based technique microarray followed by bisulfite conversion. So, here in case of microarray you all know there are pre designed probes for methylated and un methylated CPG based on which it can be identified. Whole genome bisulfite sequencing this is this is one comprehensive methylation state analysis which includes even the low CPG density region like intergenic region, partially methylated domain, distal regulatory element these are the regions where CPG density is not so high.

But as a whole the whole genome is sequenced for identification of methylation in the genome. Onwards that can be analyzed by different sophisticated bio informatics methods. Then enrichment technique can be used. Now enrichment technique followed by bisulfite sequencing is one very important technique. Now what are the enrichment techniques what this enrichment technique actually do? Remember those regions suppose they have manifested epigenetic changes via methylation hypo or hyper methylation, but if their density or quantity is low those regions specifically to be identified and isolated for further sequencing.

So, those are basically enriched isolated DNA with hyper or hypo methylated sequences. So, after this enrichment technique bisulfite sequencing can be done. So, this is known as reduced representation bisulfite sequencing RRBS. In RRBS apart from this enrichment method it is also integrated because we are going to do sequencing it is also integrated with NGS next generation sequencing. So, it has enrichment technique enrichment technique with the help of restriction enzyme digestion, then bisulfite conversion and finally, sequencing by next generation sequencing NGS.

Now here you can see there is restriction enzyme digestion based enrichment technique. So, we will discuss what are the different enrichment technique in the next few slides. Before that in DNA methylation even the oxidized form of DNA methylation can also be identified. Now the enzyme TET 11 translocation enzymes TET 1, 2, 3 they are relevant in DNA demethylation. Now while doing DNA demethylation they generates oxidized form of 5 methyl cytosine like 5 hydroxymethyl cytosine, 5 formyl cytosine 5 carboxyl cytosine these are the oxidized form.

So, to identify this oxidized form what is done there is oxidative bisulfite sequencing OXBS sequencing. This is basically one modified bisulfite bisulfite technique where the 5 hydroxymethyl cytosine is basically oxidized to 5 formyl cytosine and 5 methyl cytosine remains intact. Now they can be sequenced similarly HM top sequencing that is 5 hydroxymethyl cytosine specific tethered oligonucleotide primed sequencing. This is one such high resolution bisulfite free method where you can directly read the different sequence which are primed at covalently labeled 5 hydroxymethyl sites from an in situ tethered DNA oligonucleotide. So, basically these are nothing, but one specific probe

based diagnosis probe based detection of 5 methyl 5 hydroxymethyl cytosine the oxidized form via high resolution based technique.

So, this is how oxidized form of DNA methylation can be identified. Now we are coming to different enrichment technique. Enrichment technique can be done by affinity enrichment affinity why we are talking about affinity basically this affinity is for certain proteins or antibody for methylated DNA. So, there are different transcription factor there are different epigenetic modulators which different methylation binding protein transcription factors different inhibitors inducers they can specifically bind in the methylated DNA sequence or un methylated or hypo methylated DNA sequence. So, if those specific proteins can be identified then via those proteins we also can identify the regions where the DNA methylation changes are occurring.

So, those methylated sequence are specifically captured by targeting the specific proteins which are bound in those region. Now how it is done? The very first step is genomic DNA shearing it is fragmented by sonication or different enzymatic digestion after this sonication what happens there are different fragments, but those epigenetic modulators or transcription factors they are bound to the to their specific DNA region. Now after even after this DNA shearing they remain as it is. So, if the DNA fragment is like that and this is the region where a protein is bound if we fragment. So, what will be there different fragments as well as the fragment where the protein is bound.

Now immunoprecipitation will be done by antibodies which are specific to 5 methyl cytosine. So, this region is rich in 5 methyl cytosine to where this methylated in those methylated region basically the proteins are bound. Now if the antibodies are specifically targeted to this region we can immunoprecipitate them even sometimes we can go for immunoprecipitation by targeting the proteins as well. Now after this immunoprecipitation we need to isolate this fragment how it is isolated via magnetic bit capture method those magnetic bits are basically coated with protein A g. Now the antibodies via which the immunoprecipitation is done they will be tagged upon magnetic bit via protein A g interaction.

So, via magnetic bit we can isolate these regions after the isolation it is labeled. Now a DNA control DNA sample where there is no methylation it is tagged with a separate levels consider this is Cy 3 and this region is tagged with Cy 5. So, the control and the sample they are tagged with differentially labeled fluorescent tag consider. So, after that they are identified over array platform the the sequences are identified over array platforms. So, this is how affinity enrichment is done via isolating a very specific region of methylated or un methylated DNA only.

So, if we fragment there are different multiple fragments of DNA these are not affected

by epigenetics. So, this only the small region we need to sequence or we need to identify. So, for that we need to isolate these region amongst this whole genome and affinity enrichment technique basically helps in this way. Affinity enrichment targets via specific antibody or protein base interaction to those regions. One such important affinity enrichment technique is MEDIP that is methylated DNA immunoprecipitation CHIP method.

Also for enrichment we can use restriction enzyme. Now the restriction enzymes mostly are of two different types one is methylation sensitive restriction enzyme MSRE another is methylation dependent restriction enzyme that is MGRE. Now methylation dependent enzyme is basically identifying the regions of DNA which are methylated. So, one such example is MSP 1 MSP 1 restriction in endonuclease it can identify these region 5 prime CCGG only when the cytosine is methylated. Whereas, in contrast to these methylation sensitive restriction enzyme they can identify when the cytosine is un methylated.

So, the counterpart of MSP 1 is HPA 2 which also recognizes the same sequence 5 prime CCGG, but it is only identify it only identifies this region when the C cytosine is un methylated. So, this is how restriction enzyme can target the restriction sites based on the methylation status of the cytosine. After this fragmentation the size based on the size of those fragments they are labeled and hybridized over the array platform for the identification. Now different comparative approaches can also be done where two complementary restriction enzymes are basically used simultaneously. One such technique is HELP, HELP is HPA 2 tiny fragment enrichment by ligation mediated PCR.

What is done here? It is very simple both of the enzymes are used here one is there are two samples one is treated with HPA 2 enzyme another is treated with MSP 1. Similarly comparative approaches like methylated CPG island amplification in combination with microarrays can be done differential methylation hybridization can be done. So, here the methylated DNA and the un methylated DNA they are tagged with different fluorescent tags. Also comprehensive high throughput arrays for relative methylation these are for identification of the sequences. Next coming to the second epigenetic change that is chromatin modification.

Chromatin modification is basically one post translational modification of histone protein which includes acetylation, methylation and specific lysine residues or code histone tails. So, these are modification over histone proteins not over DNA remember. Now increase histone acetylation is associated with transcriptional activation whereas, deacetylation leads to gene silencing. One such example is methylation of H 3 K 4 activates transcription whereas, methylation of H 3 K 9 represses the expression. And these chromatin modifications are mostly done by these enzymes histone acetyl transfer

is histone deacetylase and histone methyl transferases.

Now these histone modification they also have role over chromatin structure and they based on that different genes are expressed on or off whether they will be on or off that is decided. So, all are basically decided with whether the DNA region is accessed by the transcriptional machinery if not in that case the transcription will be off. So, this is how chromatin modification or even DNA methylation they can regulate the gene expression. Now for identification of chromatin modification the gold standard is chromatin immunoprecipitation or CHIP. Regarding the chromatin immunoprecipitation we have discussed the principle in our previous class, but here we are going to discuss how chromatin modification is helping to identify the sorry chromatin immunoprecipitation is helping to identify different chromatin modification in epigenetics.

So, here in CHIP specific antibody is directed towards the specific markers or proteins or epigenetic modulators or transcription factors which are causing histone modifications. Now CHIP chromatin immunoprecipitation can be integrated with different other techniques which can help in identification of histone modification the regions where histone modification is done and also due to this histone modification what are the DNA regions that are affected. Now if the target histone modification and the respective DNA regulatory region are specific or known in those cases CHIP can be followed by the PCR quantity PCR technique which reveal the histone modification or the ability of those modified histone in remodelling complex to the specific region. So, when the regions and histone modifications are known in those cases CHIP followed by PCR, but if the regions or the chromatin modifications are not known or undefined in those cases what we need to do we need to sequence the region those targeted for the chromatin the changes over the chromatin there are changes over gene expression. Now which regions are affected over the DNA if we need to identify that means, when that is undefined in those cases what is required sequencing is required.

So, that can be followed by CHIP sequencing or CHIP-CHIP here the CHIP is microarray. So, CHIP followed by microarray or CHIP followed by sequencing that can be NGS or pyro sequencing etcetera based on those the unidentified or undefined regions of DNA which can be modulated by histone modification can be identified. Now DNA methylation and histone modification integrated identification or integrated analysis can be done via combining different test. Now what are those CHIP bisulfite methylation sequencing which determines the methylation status of CHIP DNA, CHIP DNA which is pulled down by a specific antibody. Similarly, bis-CHIP sequencing that is bisulfite treated DNA is immunochromatin immunoprecipitate.

So, that is required for global profiling. So, remember when there is chromatin CHIP bisulfite methylation sequencing that is for some allele specific or locus specific region

for global identification what we need we need bis-CHIP sequencing. Similarly, SCAN or single chromatin molecule analysis in nanochannels it combines the epigenetic modification specific antibodies with single molecular imaging technique for high resolution assessment. So, SCAN is basically one imaging technique which can which help in identifying the DNA methylation and histone modification. Now apart from that there are different advanced method presently used like different types of 3D chromatin organization, epigenomics. So, the this can help in identification of both DNA methylation and histone modification pattern.

So, this is how we can divide the techniques evaluating histone modification. So, if we only want to know what is the type of histone modification that is whether there that is acetylation or methylation and how much it is abundance of such for that immunoblot analysis that is dot blot or western blot that will suffice. And for that what is the sample whole cell lysate can be used enriched sample for identification of specific region can be used. But if we need to identify the specific interacting protein in this modified histone region what we need to do we can go for immunoprecipitation and western blot. Then enzymatic processing of chromatin to generate the single nucleosome unit because we want to identify in that single nucleosome what is the specific protein or what is the specific interactive protein binding for that enzymatic processing is required.

After that antibodies are targeted towards that which is pulled down by chromatin immunoprecipitation followed by western blot analysis of those targeted proteins can be done. If we want to identify the genomic location of histones for that we need some additional methods which is added with chip chromatin immunoprecipitation. So, it can be chip can be combined with PCR chip can be combined with NGS can also be combined with microarray chip chip. So, what is that that is how much histone modification is done.

So, if there is histone modification. So, if there is histone modification what are the interacting protein and the DNA region. So, so this is our DNA region which can also be identify which is affected by histone. So, this is how these are the three these are the three approaches for identifying histone modifications. After that we are going to non coding RNA's non coding RNA's basically a group of RNA transcript that do not encode proteins like mRNA. Now, this non coding RNA can be of two categories based on their regulatory role.

So, one is housekeeping non coding RNA. So, remember in case of translation we all have we have discussed all these are RNA's like the role of rRNA tRNA they do not give rise to any proteins, but they definitely help actively in transcription in translation. Whereas, there are different types of other regulatory non coding RNA which take part in gene expression via different complex and molecular cellular processes. So, those are

the non coding RNA we are going to discuss here. So, if we divide those regulatory non coding RNA based on the nucleotide counts they are divided in small non coding RNA which are having less than 200 nucleotides those are micro RNA, si RNA, pv RNA and long non coding RNA which are having more than 200 nucleotides in length. For identification of small non coding RNA that is miRNA, siRNA or piRNA what we do for the profiling the very basic is QRT TCR also hybridization based microarray can be done then RNA sequencing high throughput RNA sequencing can also be done RNA sequencing we are going to discuss in next classes.

Now those non coding RNA how they are affecting the DNA gene expression. So, if we want to identify the target genes by the non coding RNA for that what we need some high throughput sequencing of RNA that RNA which is isolated by cross linking immunoprecipitation its clean it identifies the functional protein RNA interaction site. So, this non coding RNA so, if this is the non coding RNA if they are affecting some regions of DNA we want to identify this region or the genes that is done by heat skip. Similarly, ligar sequencing that is ligation of interacting RNA followed by high throughput sequencing is done for profiling micro RNA mRNA interaction. So, this is about the small non coding RNA for the long non coding RNA there are different profiling techniques before that you need to know something about long non coding RNA like it is transcribed by the RNA polymerase II and from independent promoters like mRNA it has caps it has poly a tail, but simultaneously it in exhibit some extensive mechanistic diversity through interaction with RNA binding proteins at specific DNA region.

So, in different DNA regions there are bound proteins RNA binding proteins they can bind this long non coding RNA. So, these proteins are bound over DNA and these are RNA binding proteins. So, over these the micro sorry the long non coding RNA can bind. Now, the detection and profiling of long non coding RNA can be done by qPCR also genome wide chromatin binding site of long non coding RNA that is CHARC chromatin isolation by RNA purification and CHARC capture hybridization analysis of RNA targets. So, the whole genome sequencing or whole identification of the whole genome where these long non coding RNAs are affecting that can be done by CHARC and CHARC.

RAP it can map the localization it can localize in fact, that specific long non coding RNA across the genome RAP is RNA antisense purification. Then there is RIP that is RNA immunoprecipitation. So, it identifies the protein RNA interaction. So, RIP or RNA immunoprecipitation if it is associated with microarray or sequencing can help in identify the protein RNA interaction and that specific site. Similarly, RNA pull down followed by LC-MS MS, LC-MS is liquid chromatography mass spectrometry.

So, it can basically identify the interacting protein. So, even by pulling down the RNA the different RNA binding protein can also be identified by different methods of proteomics like liquid chromatography or mass spectrometry. So, this is all about this class here we learn that DNA methylation, histone modification and non coding RNAs are basically the main cracks of epigenetic changes we which orchestrate gene expression as well as different cellular processes starting from biological to pathological processes. Bisulfite treatment, bisulfite sequencing is basically the gold standard for DNA methylation detection whereas, chip or chromatin immunoprecipitation is required for identification of histone modification. Now, both of this technique can be clubbed with each other can be clubbed with different sequencing method, different amplification method like PCR sequencing method like next generation sequencing, pyro sequencing for better identification of the regions where epigenetic changes are occurring. Then there are different enrichment method enriched enrichment method is required for isolation of the specific regions where the DNA DNA sorry epigenetic changes has occurred.

After that RNA profiling techniques including QRT, PCR and high throughput RNA sequencing it allows to identify or investigate over non coding RNAs and their regulatory roles. And why this is required for identification of different changes or different governing process over cellular physiology as well as different disease pathogenesis. So, along with all these techniques what is the present approach is the bioinformatic analysis of this epigenetics which help in identification of different epigenetic changes. So, these are my references. Thank you all and see you in the next class. Thank you.