Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis

Prof. Arindam Ghosh

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

Indian Institute of Technology Kharagpur

Lecture 34 : Next Generation Proteomics

Hello students. Welcome back, Namaskar everybody. So, we are continuing our lecture series on comprehensive molecular diagnostics and advanced gene expression analysis. And we are into our module 7 where we have learnt advanced technique of proteomics. So, today we will be laying down the principle of next generation proteomics. Now, we are into our ongoing series within a series that is we have discussed all of these topics in our previous 3 lectures.

We have discussed the methods of quantity proteomics, how labeling methods how we can do in vitro in vivo labeling. We have discussed all of these and label free quantification techniques we have also discussed. Now, we were the list that we displayed in that slide we were left with data acquisition methods or data interpretation method that we often use in mass spectrometry based quantity proteomics. So, in today's class we will be covering these topics, this is a wide range of varied topics that we need to cover that we will be learning together.

So, we will be discussing all the methods of quantity proteomic interpretation techniques both the targeted and untargeted approaches and what are the specific methods that we need to know in these interpretation techniques. We will also be discussing the one of the special techniques that is data independent acquisition and it sub variety that is widely used in proteomics studies. We will also be discussing the basic proteomics approaches proteomics strategies that is bottom up and top down proteomics. We will be discussing shotgun proteomics and also the emerging concept of single cell proteomics alright. So, we have a lot to cover.

So, a strap in and we will start our lecture topic with the most commonly used quantitative proteomic interpretation techniques alright. So, there are four commonly used techniques that are used in LCMS based method that is combination of HPLC and mass spectrometry method that can be grossly divided into targeted approach and untargeted approach right. So, the two methods that we will be studying that are most commonly used when you are using targeted approach I will tell you what do you mean

by targeted and untargeted approach . We should know the names at least not if not the methods. So, they are selective reaction monitoring or multiple reaction monitoring often abbreviated as SRM or MRM parallel reaction monitoring PRM you have come I mean you have encountered this terms when we were discussing mass spectrometry time of flight those concepts right.

And there are untargeted approaches of data interpretation data acquisition. So, one of them is known as data dependent acquisition and the modern the latest one that is data independent acquisition. So, we need to know the basic principle and schematic of all of them alright. So, what happens in selective reaction monitoring or multiple reaction monitoring basically these are monitoring of ions that are being fragmented and as they are going through the time of flight spectra and generating the spectrum we are free to choose the ions that we want alright. So, the SRM or MRM method what it does in this method in this principle of mass spectrometry based quantification predefined precursor ions are isolated alright.

We target we will only analyze these precursor ions only alright and then after they are fragmented after they are going through the time of flight analyzer we select and monitor that only single predefined precursor ion that is derived from the precursor ok. So, mind it we are choosing a single precursor ion then we are fragment the fragmentation is happening and then we are selecting and monitoring one single product from the precursor ion ok. So, everything is very selective it is restrictive we are controlling what we will fragment and what we will select this is the basic principle of SRM right. In MRM what happens we can when it is selective we are selecting one by one while detecting while we while we are doing MRM multiple such ions can be monitored one by one ok that is the principle of multiple reaction monitoring ok. So, what is PRM parallel reaction monitoring? Mind it for all I mean if this concept appears I mean if you are finding what is going on is basically the phase where in the mass spectrometer where in a phase in the mass spectrometer where analysis is going on where ions are getting fragmented and it is travelling towards an analyzer and getting analyzed.

So, we are going we are going in depth in that phase and we are knowing how we can analyze the data right. So, PRM what happens almost similar it isolates predefined precursors ok. So, predefined precursors precursor ion we are selecting right in the same way as SRM or M, but it records all the products of a predefined precursor. So, predefined precursor ion so, we have selected the blue ion right we we are now selecting blue ion now the blue ion will fragment in SRM or MRM we are selecting one subset of the blue ion and we are analyzing it all right. In PRM all the products are analyzed all right it is the basically a spectra right a signal a characteristic peak or a curve right based on that now we will again analyze the parent compound these are all different variety of data interpretation methods.

So, in PRM what happens we select a precursor ion it fragments and now we are recording all the product of the precursor ion all right. Next this PRM ok so, mind it there might be image based question and since these are almost similar multiple choice questions might be confusing. So, you should pay attention to what is happening regarding the precursor ion and the product how we are we selecting and how what are we analyzing right. So, what happens in data dependent acquisition another process mind it is also known as information dependent acquisition mode these are all modes of a spectrometer mass spectrometer all right. So, DDA is also known as IDA the same thing in SRM MRM PRM the investigator the experimenter I mean the researcher are were selecting the precursor ions here what happens we are not selecting anything the instrument is selecting all right very important the instrument selects the precursor ions very important in SRM MRM PRM instrument is not selecting the observer is selecting right the experiment is there in such a way that we are selecting a precursor ion in here instrument is selecting a precursor based on many criteria the most important criteria is the one with highest intensity the highest intensity.

So, the mass spectrometer will automatically select the ion with the highest intensity and then it will be fragmented and then all the products of that specific ion will be analyzed ok. This is the this is one of the principles of analyzing the ions in mass spectrometry just like SRM MRM PRM very important key point to notice instrument select the ion it may be many criteria the most important criteria is the intensity. So, the basically the top N number of precursor based on the M by Z value are automatically selected by the instrument fine. Now, what is data independent acquisition? Here DIA there is no selection as such ok every ion will be fragmented all right every ion will have its product and everything is analyzed and recorded all right. So, we are not relying on any precursor ion selection very important.

So, how how is DIA different from the previous discussed methods? We are not relying on any precursor selection the machine or the mass spectrometer does not select any ion based on any value all right the entire mass range is analyzed all right and it captures data for all ions ok, but there is something. So, it gives an infinite data all right there is so much everything is already there right it is a raw data and we are now open to select what we want ok. So, predefined isolation window is something that is incorporated in data independent acquisition methods. Now, what is predefined isolation window that concept should be clear. Now, one of the most important variety of data independent acquisition that is almost always employed or deployed when we are using DIM mode is this ok termed as sequential window acquisition of all theoretical mass spectra SWOT-MS.

For all practical purposes SWOT-MS is almost always synonymous with data

independent acquisition although it is a variant, but majority of the DIA method uses this technique. So, actually what is happening everything this is the basic principle now I will discuss. So, it divides the scanning range. So, see here since all the ions are fragmented and we have got all of the information that is coming from the entire ion of the product that is to be analyzed separated or quantified right. So, we select a scanning range this is basically predefined isolation window.

So, it divides the scanning range into a series of interval at 25 Dalton. So, basically there is a molecular weight scanning range right and it obtains all debris information of all ion fragments within the scanning range through ultra high speed scanning I will make you understand what is happening. See this figure you already know this figure you have already seen what is happening in the first step with electro spray ionization or MALDI the precursor ion is being formed right. So, with bombardment the precursor ion is formed and then the precursor ion undergoes collision to form further fragmentation that fragment product and is detected this is a standard mass spectrometry procedure right nothing new in it. So, how is data independent acquisition I mean helpful and novel compared to other data driven methods you see in data dependent mode either whether we are using targeted approach or DDA we have got some control either we are selecting a precursor ion right or the machine is selecting some ion, but here we do not have anything to select there is independent this is a raw energy this is sea of information.

So, how do we go around it we select a small range all right a range of 25 Dalton this is the window and within this narrow range we scan the entire spectra all right. So, each ion each product breakdown is scanned across this entire this small range and each scanning range gives its own spectra ok you see what is happening the experimental protocol the workflow is same. So, we have got tissue sample we extract we split it into peptides and then during fragmentation there is a selection of a window. So, we have already selected a 25 Dalton window. So, we will scan all the ion that is there that is in the fragmentation chamber within this narrow range of 25 Dalton ok and they will give their characteristic spectrum then we will analyze.

So, what is happening for example, say a range of 400 to 1200 ok. So, this is the sequence this is the amount of protein sequence we need to cover or scan ok. So, this is the sample. So, we have selected 25 Dalton window. So, the entire spectrum will be scanned all the ions all the fragments will be scanned multiple times with this 25 the with this 25 Dalton window range.

So, this scanning interval 25 Dalton each time it will be scanning ok the time required to scan is 100 millisecond ok. So, in 100 millisecond this 25 Dalton window will be scanning the entire rates. Now, this 25 Dalton will vary suppose first scanning range will be from 425 400 to 425. So, we will scan the entire range what ion is present in the next

we will be scanning 425 to 450. So, we are increasing the step by 25 Dalton then we will get a spectra a characteristic signal suppose in this molecular weight range what ions are present then we will in again increase the change the window range from 450 to 475.

So, like that when there is a window or the variation molecular from 400 to 1200 which is generally the sample characteristic in proteomics 32 such slots or 32 such windows will scan the entire sample all right entire range and if we are taking 100 millisecond to scan for one window in total 3.2 second we will be scanning the complete sample all right. So, mind it this is very easy nowadays because the ions are getting fragmented. So, all the ions are getting fragmented and they are passing through multiple time of flight analyzer this specific technique this SWATH MS workflow works hand in hand with the new triple top system I have given a example. So, the triple top 5600 mass spectrometry system this.

So, triple top is MALDI TOF TOF TOF. So, in tandem 3 time of flight analyzers are connected ions are travelling right and we are scanning in 25 Dalton increment isolation window the entire spectra right and soon we get a spectral library spectral library that has record of entire thing ok and now we can go back and look for any sample. So, whether it is targeted proteomics or discovery proteomics we have got all the information over here we can now simply scan or compare this data with the help of bioinformatics to existing protein database and we can get easy information about what sample is there we can quantify we can know the nature. So, this approach of generating a spectral library from a sample with this high throughput extremely high throughput equipment is basically what we often refer to as next generation proteomics because the concept is just like next generation sequencing. So, we have in next generation sequencing you know we select a small with the big gene is broken down to small fragment we have we are designing DNA library then we are amplifying and then we are comparing with the help of bioinformatic analysis with the sequencing information.

Here also we are generating a spectral library and then we are using this high throughput equipment first we are splitting the sample into small fragments we are generating a spectral library and then we are using bioinformatic analysis to compare. So, you can see there is a great similarity between the workflow and the working principle of next generation sequencing when we are discussing DNA and RNA sequencing and next generation proteomics when we are discussing protein sequencing. So, next generation proteomics does not only encompass this SWOT-MS workflow. In fact, all the advanced techniques high throughput proteomics even protein microarray all of that fall under the domain of next generation proteomics, but this is the most advanced form of data acquisition because it has got many advantages. So, what are its advantage? So, let us first again visualize the workflow.

So, here we see we are increasing the window of scanning and each window is giving its own spectra alright. So, there are multiple spectra that is formed ok. So, m by z it is we are having a triple dimension ok. So, it is scanning since the time of flight analyzer we are already scanning the passage of ions through time we are also selecting them with the help of intensity and m by z ratio. So, when all the scans all the scans all the spectra from each range whether scanning the entire ion range the fragment ion range when they are overlap they give a very characteristic spectra which is extremely good which provides extremely good resolution which provides all the information that is there we need to compare to other data driven methods right.

So, one of the most important aspect is you see here whenever we are doing data dependent acquisition we are selecting a precursor ion ok. So, based on that single precursor ion we are analyzing its spectra ok, but in data independent acquisition we are not leaving anything we are scanning everything across small windows and then we are overlapping. So, one of the most important use of data driven acquisition versus data independent acquisition especially the SWATH MS is detection of post translational modification alright. So, various PTMs which may be easily missed when we are selecting a precursor and the rest is getting lost the rest of the information is getting lost, but here nothing is missed ok. So, here SWATH it will not be DIA SWATH can detect, but other data dependent acquisition modes will not be able to detect ok.

The PTM that is post translational modification. So, what are the key advantages of data independent acquisition you see very important reproducibility consistent data acquisition across sample. So, you have one sample you can just run it and you are done alright you do not need to split you do not need to select various speakers or windows based on your target experience. So, basically it is an approach of do first and then you can analyze later. So, you have got the sample you analyze everything and now you can even design or your experiment later based on the characteristic right this is not a right approach, but you can alright.

Whereas, if you design your experiment first you have got a limited sample you already take your run then when you think of something new with the same sample same type of sample you need fresh sample to select your another precursor and design the experiment right. But if you have all the information all the spectral library that is possible you can definitely now choose your own area of interest expertise and work with the same sample alright. Next the same thing that is generalizability no specific sample specific method we do not need to design. The sample the analysis the complete analysis has already been done irrespective of what experiment we need to do. So, depending on the sample we do not need to design any specific method this is very highly sensitive and very important since we are only relying on the signal and subsequent bioinformatic analysis it enables label free quantification thus reducing the cost compared to another

label based methods.

Again there is no dye there is no label so, chance of interference very low and as I discussed excellent for detection of post translational modifications right. So, with that we complete or we complete our discussion on proteomic interpretation and data acquisition methods we have discussed 4 and among them the latest and the modern the highest throughput is data independent acquisition. Now, we will be discussing the various approaches in proteomic study. So, bottom up and top down proteomics. So, what are they? So, bottom up proteomics what is it? It involves proteolytic digestion of proteins before analysis by mass spectrometry that one that we are have been seeing.

So, sample we are we might separate it first, but the major part is before we are putting the sample in mass spectrometer it is digested by trypsin into small peptides. So, the term bottom up implies the information about the constituent protein that is deconstructed individual from individually identified fragments or peptides. So, we are first from a sample of big protein we are degrading the protein to small peptides we are generating the mass spectra from each of those small segment and then we are fitting the information like a jigsaw puzzle to get the idea about the constituent protein or the parent protein that we are required to analyze. So, break analyze join this is the approach of bottom up proteomics this is the traditional proteomic approach that we all have been doing it has got its advantages and disadvantages. So, the most mature it has been used for decades now.

So, we have optimized the things it is the most mature most widely used approach in protein identification and characterization. Since it has been there for long we have learnt it we have the thing has become inexpensive. So, less sophisticated instrumentation and expertise we can easily perform bottom up proteomics with minimal experimental requirement or equipment requirement of course, we need a mass spectrometer, but the effort and the analytic method are much easier compared to the top up top down proteomics and ability to achieve high resolution separations ok. What are the disadvantages? Low percentage coverage of protein sequence right. Since we are breaking there the information about what lies in between the various sequences might be lost and this is also one very important factor why any information about post translational modification alternative splicing variants are lost.

It is inexpensive easier to perform, but there are some caveats right. So, to cater all of these there is another different approach that is top down proteomics. How is it different from bottom up proteomics? So, here in bottom up proteomics protein split analyze join here the intact protein from complex biological system are characterized very important we are not splitting there is no role of proteolytic enzyme no role of trypsin right. So, proteins are typically entire protein all right are typically ionized by electrostrianization

and then trapped in a Fourier transform ion cyclotron resonance base mass spectrometer orbitrap mass spectrometer all right. You know you have seen in PRM as well as in DDS orbitrap base mass spectrometer basically the trapped ions all right and they are analyzing each and everything ok.

The fragmentation that is achieved for tandem mass spectrometry by mostly two processes electron capture dissociation and electron transfer dissociation. Why do we need this special form of fragmentation? Because we are trying to fragment the entire protein all right. There can be separation we can separate them using chromatography prior to LCMS that is of course, done LCMS is a unit right, but we are not fragmenting the protein in bottom up proteomics the proteins are being fragmented then each fragment was being analyzed using LCMS ok and then we are joining. Here there is no fragmentation direct LC separation of the fractions of the protein and then analysis with the help of tandem mass spectrometry and then again data analysis. Advantages and disadvantages are there right main advantage is complete protein sequence can be analyzed what the disadvantages of bottom up proteomics become the advantages of top down proteomics.

So, when we are breaking down the information between the sequence might be lost depending on the activity of proteolytic enzymes. However, since we are we are subjecting the entire protein to the analysis there is no loss of information. Therefore its ability to detect post translational modification protein isoform determination of iso enzymes of protein very important we have to use top down proteomics because if we split the various iso enzymes they almost give a similar spectra right, but if we analyze the entire protein then we can actually see the difference. And elimination of time consuming protein direction since we are eliminating the step of proteolysis it saves some time. Disadvantages its expensive not only expensive its new we need to optimize it more and more.

Moreover the collision methods right they are not very efficient and they have got their own time taking steps of accumulation activation and detection alright. That is the reason it has not been achieved in a large scale alright mainly because of the lack of protein fractionation methods that are integrated with tandem mass spectrometry ok. So, intact protein fractionation we are not talking about proteolysis mind it fractionation using liquid chromatography alright. So, those are not yet optimized compared to the bottom of proteomics method I mean the fractionation of small peptides are much easier compared to fractionation of large intact proteins ok.

So, if we just look side by side. So, protein fractionation can be done see the first thing the we are taking a protein mixture protein fractionation can be done both in the cell tissue or lysate ok. In case of bottom of proteomics there are extra. So, see protein

fractionation means separation of the gross protein sample using 2D gel electrophoresis liquid chromatography isoelectric focusing we all know how it is done right. But extra step of proteolytic digestion needs to be done in bottom of proteomics and you see strong cata SCX stands for strong strong cation exchange.

So, strong cation exchange IEF. So, after we have digested there might be a need for second fractionation after which it is subjected to LCMS analysis and then data in bio bio informatic analysis using protein databases. However in case of top down proteomics there is no digestion. So, these two are cut off with directly fractionate the proteins and we analyze them using LCMS and then bio informatic analysis come by dealing with protein databases. So, this is basically a compilative analysis of top down and bottom of proteomics I have compiled figures from various journals whose reference I have given at the end, but the thing is more or less the same. Whenever you see an approach where the proteins are being split first and then they are analyzed it is bottom up.

And whenever the proteins are intact and then they are subjected to analysis using mass spectrometry it is top down. Again when we are using bottom of proteomics many information between the sequence for example, in this case is a pictorial representation showing a post translational modification PTM which has been lost final result does not give information about this, but whenever the whole sequence is intact we get idea about the post translational modification because nothing is breaking up ok. Next shotgun proteomics. So, we are now at the almost at the end of our discussion. So, what is shotgun proteomics? It is basically if you here this term we are referring to the traditional bottom of proteomics alright.

So, you have learnt about shotgun sequencing right where the analogy being shotgun randomly divides the parental gene when we are discussing shotgun sequencing. And then we are analyzing each and every fragment and then we are fitting them into jigsaw puzzle and then restructuring the whole thing right, realigning the whole sequence and then getting idea about the parent sequence. Here same thing bottom of proteomics we are dividing or splitting the sample proteins with the help of proteolytic enzyme most important being trypsin the enzymatic digestion. And then we are fractionating them with the alpha isoelectric focusing then HPLC electrospare ionization. So, again collision methods so, collision induced dissociation high energy collisional dissociation depending on what we are analyzing.

Mind it in shotgun proteomics the preferred analysis method are always data driven acquisition right. We compare the I mean we know we select what are the precursor ions and then or the instrument select the precursor ions and then we analyze and then we repurpose and restructure the parental protein. So, two things shotgun proteomics basically same as synonymous with the bottom of proteomics approach splitting and then

analyzing and the preferred method is data driven acquisition because we are splitting. Data independent acquisition again can be used, but we prefer whenever data independent data independent acquisition can be used for all types or of sample right. But whenever there are big samples to be analyzed data independent acquisition that gives us idea about the entire range of spectra are preferred right.

And so, see often you might find a new term or come new term that we have not discussed till now that is middle down proteomics. So, we already know what is bottom up proteomics, what is top down proteomics in bottom up there is digestion and top down proteomics there is no digestion with the help of enzymes, what is middle down? It is basically same as bottom up proteomics, but the fragments are larger in bottom up proteomics basically the fragments lie between 8 and 30 amino acids in here it is more than 30 amino acid. So, peptide separation peptide selected and then we analyze whereas, we already know what happens in top down the entire protein intact protein is selected alright. So, basically the proteoform acts as a precursor and then we analyze. Lastly we will be discussing single cell proteomics, what is single cell proteomics? Single cell proteomic profiling refers to analysis of proteins at the level of individual cells.

So, if we have got a sample from a tissue there are multiple cells which needs to be which are analyzed right and even we have got a tissue like say there are multiple cells that are analyzed averages out, but we if consider situation where there are cell to cell variability a disease in which one cell even if it is similar to another cell varies in different way. Can you name some condition where a disease pathology in which one cell will differ from another cell the adjacent cell? Stay tuned if let me see whether you can guess correctly. So, whenever there is an heterogeneity in cell I mean cellular heterogeneity and function this helps in more detailed understanding. The key steps very similar cells need to be isolated very very very important. So, we need cell sorter one of the preferred method can be fluorescence activated cell sorter or FACS ok.

There might be magnetic cell sorter there are multiple ways of sorting cell from a tissue. So, we need to select individual cell. So, from cell culture cell sorting is done and then each cell is lysed and the proteins from one single cell is digested we already know what is TMT leveling tandem mass tag. So, this is one of the method that has been shown here and then protein from each cell is quantified using LCMS tandem LCMS and then data analysis done and we get a picture from a single cell ok. So, can you tell me what are the situation that needs it most where cell to cell variability is the maximum in cancer research immunology and developmental biology very important for each cell actually varies a lot the function of each cell can vary lot from its adjacent cell ok.

So, basically if you just look into the broader application of single cell profiling. So, the work flow the principle is the same we are doing the same thing, but the source of

protein the sample is one single cell all right. So, basically these are the applications and I would like to mention studying post translational modification very important protein protein interaction distribution protein RNA distribution protein complex analysis phenotyping and proteomic profiling there can be many things that can be done from a single cell to get information about how a single cell and its proteomic function maturity every development differs from another cell. So, this is just a short overview of single cell profiling with the concept of mass spectrometer quantity proteomics being clear the sample when it is a single cell you would not have any difficulty to understand or making someone understand about single cell profiling. So, to summarize we have discussed a lot we have discussed all the quantity proteomic interpretation techniques the targeted in which we have discussed SRM, MRM and PRM the untargeted approach in which we have discussed DIA and DDA we have specially discussed SWOT-MS under as a sub variety of DIA which we have established the most important we have discussed the proteomics strategies of bottom up and top down proteomics we have discussed short gun proteomics which is synonymous to bottom up proteomics which uses data driven acquisition mode and then we have also discussed single cell proteomics the concept of it what is the what are the steps and the applications all right.

So, we have learned a lot I would request you to please review this video once again if it was overwhelming it is will become very easy for you once you have reviewed it again. So, these are my references for today's video and I thank you all for your patient listening.