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 52 : Metabolomics in Molecular Diagnostics

Hello, Namaskar. Welcome back to the NPTEL lecture class of Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis. So, we are continuing in the 11th module of molecular diagnostics in medicine. In today's class, we are going to discuss Metabolomics in Molecular Diagnostics. We are going to discuss the very basic concept of metabolomics, how they work, the how it works, the work flow of metabolomics and finally, how we will approach for the analysis of the metabolomics. So, metabolomics as quoted is the systematic study of unique chemical fingerprint that specific cellular processes leave behind.

So, basically this is one systematic and scientific study, study of chemical processes involving metabolites. So, metabolites as a whole are giving one chemical fingerprint. Now, metabolomics are basically dealing with the molecules which are they are size is up to 1.5 kilo Dalton or 1500 grams per mole or less.

So, this is the limit. So, these molecules very small in size, they are included in metabolomics that is why it is also known as small molecule profiling. Now, metabolome refers to that set of molecule or biomolecules which are basically obtained from biological sample at the end of one cellular process. So, this is a complete pool of metabolome which is dealt in metabolomics. Now, metabolites is basically the intermediate products in metabolism.

Now, when we talk about metabolic profile, metabolic profile is one snapshot of a specific time whereas, metabolome is one dynamic entity. So, metabolome is basically changing in time to time during different physiological pathological processes inside the body. So, what is required is actually real time metabolomics where we can study different types of metabolites in a living cell in real time. So, the molecules which we measure in metabolomics can be endogenous molecule that is produced within the body or within the organism like amino acid, lipids, different cofactors, nucleotides, sugars, hormones or exogenous which are produced external to the organism and it is introduced from outsides like drugs, toxins, different environmental, contaminants,

pesticides, herbicides. So, they are the exogenous metabolites.

So, you can see metabolomics is the end part of the biomolecules study where from DNA via transcription we get RNA from RNA we are getting different proteins via translation those proteins are forming different hierarchical structure like primary to secondary tertiary quaternary protein. Then those proteins undergo different types of cellular processes and metabolism giving rise to different small molecules and those small molecules are dealt in metabolomics. So, functional proteomics or genomics they deal up to protein from DNA up to protein synthesis and they are different characteristics. Genomics dealt with DNA and RNA only proteomics and is dealing with the protein and transcription transcriptomics is basically dealing from the transcription beginning to the end of protein synthesis. So, these are the different omics genomics proteomics transcriptomics functional proteomics or genomics and finally, our topic of today is metabolomics.

So, here you can see that genome transcriptome proteome and here is our metabolome. So, from protein via different metabolic processes what we are getting different types of small molecules which we study in metabolomics. Now for studying metabolomics the approach can be of 2 different types in a broader aspect. So, it can be a targeted metabolomics study or it can be untargeted metabolomics study. So, what is targeted metabolomics study? So, in targeted metabolomics study we have our target fixed to some specific analyte to some specific molecules molecules which are predefined pre studied which has internal standard and different types of characterization.

So, basically in targeted metabolomics what we are studying what we are measuring measuring defined set of characterized and biochemically annotated analyte. So, that is one identified and known analyte. So, for that what we need previous characterization or or analyte analytical details of that specific metabolite. So, what for we measure what for we go for targeted metabolomics? Metabolomics are basically where we are having one previously established hypothesis based on that hypothesis we want to validate our point. For that we are analyzing our known metabolite or known analyte in different different physiological or pathological condition.

So, before that what we need we need a prior knowledge of that metabolite and its related different metabolic pathways and mechanisms which are dealing with that known metabolite. And finally, for quantification what we need is internal standard and known concentration of same metabolite from based on which we prepare the sample extracted from the sample and then we measure that those specific targeted metabolites. Whereas, untargeted metabolites now it is clear that if the targeted metabolite is basically dealing with the known one untargeted metabolite is where we do not have any specific target. We are studying the sample for a global and comprehensive analysis. So, basically

whatever sample whatever analytes are present in the sample we are going to study it.

So, in the sample there can be multiple known metabolite along with that there can be multiple unknown metabolite. We are studying all of them in untargeted metabolite a untargeted metabolomics. So, here we are basically generating some hypothesis this untargeted metabolomics is basically not hypothesis driven. Here no previous hypothesis is there. We are trying to establish one hypothesis.

So, basically what we are doing we are doing one qualitative identification presence of different types of metabolites that is qualitative identification along with that we can relatively quantify different endogenous metabolites. Now how it is relevant? It is relevant in different types of biomarker discovery. Remember whenever we are going for a new biomarker what we need to do first we need to check whether that specific biomarker is present or not. In this way we can do another part is we are studying all possible markers present in the sample and from that 10 or 20 markers or 10 or 20 analyte which analyte can be a specific biomarker. Now that can be a new one and for that new one what we need to do in the sample we are basically checking the presence of that new some new metabolite we can measure in two different sample and can then compare.

So, that is a relative quantification and after that what we can do? We can then after the complete characterization and complete measurement and standard setting then we can go for the targeted one. So, basically initially for one unknown sample what unknown metabolite we need untargeted metabolomics. Now the problem is for this type of untargeted study we do not have any set of targets. So, there is one global metabolite extraction procedure we need to do for each and every types of metabolites we need to extract separately. After that what we will get? We will get one large data set in our hand with some known and unknown metabolites.

So, we need to design one process design one analysis and accordingly different processing step. So, this is one cumbersome procedure when we are going for untargeted metabolomics. Now targeted metabolomics of course, because the initial characterization is already done in targeted metabolites. What is the advantage? We can extensively study those identified metabolites in terms of different metabolic process in our body. So, it can be like suppose we are talking about hexokinase enzyme.

Now how hexokinase enzyme is related to some metabolic pathway we can study that. Then we need standards label standards and their internal standards and characterization. So, based on that we can study the targeted metabolites. Internal preparation should be easy for that and it is very easy when we are having one control level we can study different experimental groups based comparing with the control, but what is required for

targeted metabolites? We always need a prior knowledge for that we need internal standard for that. So, because that is one targeted metabolic metabolomics we cannot study more than 20 or 30 molecules at a at one single one single experiment and also sometimes there is a risk of missing our targeted metabolite.

Whereas, untargeted metabolite it does not have any proper target. So, when there is no prior knowledge we can do it whereas, we cannot do in targeted metabolomics. So, for even unknown samples unknown metabolites characterization we can do untargeted metabolomics. So, we can study and identify different new metabolites and can profile them. Then sample preparation it becomes complex, but we can discover some previously unidentified metabolites can find different unexpected changes in different metabolite characterization.

Again we cannot do absolute quantification because we do not have any initial reference standard. So, when there are multiple types of metabolites the smaller amount of metabolite can be missed whereas, the higher higher abundance metabolite can be detected with some bias. So, that is the problem with untargeted metabolomics. Again if we compare that the scope of targeted metabolomics is very much focused to the known one whereas, untargeted metabolite is basically handling the global profiling of the metabolites. Only a limited number like 20 at most can be studied in targeted metabolomics whereas, in untargeted metabolomics we are dealing with thousands of metabolites in a single sample.

Quantification absolute quantification can be done in targeted metabolize because there is some internal reference standard in absence of that we cannot do it untargeted metabolomics we are doing relative quantification. Because we are dealing with very small amount of target so, the sample extraction procedure is easy whereas, we need global metabolite extraction procedure in untargeted metabolomics. So, basically where we apply it targeted metabolize is for hypothesis driven studies where we validate our previously identified processes whereas, new hypothesis can be generated new biomarkers can be discovered in untargeted metabolomics. Then this point is very much important false positivity rate false because there is one internal reference standard the false positivity rate is very low in targeted metabolomics whereas, there is no reference in untargeted metabolomics for the unknown metabolites. So, the false positive rate can be high where there are some metabolic metabolites concentration variation fine.

So, these are the comparative analysis of targeted and untargeted metabolomics. Now, the very basic workflow for metabolomics is just you can assay for different samples. So, those samples should be collected and processed after that they should be separated based on the targeted or untargeted profiling. So, what are the different separation procedures it can be gas chromatography, high performance liquid chromatography, ultra

performance liquid chromatography, capillary electrophoresis. After the separated sample we will now detect different metabolites.

So, the very well accepted techniques for detection of the metabolites and nuclear magnetic resonance spectroscopy or NMR spectroscopy and mass spectrometry. After that there are different computer driven softwares based on that we can analyze the data and then in can we can validate it in clinical sample. Now, the raw output data can be used for metabolite identification and further process before statistical analyte and there are different many bioinformatic tool which we software and tools which can be applied for when we are getting a metabolic profile we can compare it with between healthy and the diseased one. We can compare between different disease states, we can compare the different outcomes based on their metabolite profile, we can characterized one finger print or metabolic signature for some biological processes fine. So, these are the thing we can do from the data analysis.

So, there are different projects which have dealt with metabolomics the very important one is human metabolome project or HMP which has been launched in 2005 January funded by Genome Canada. Now, the project primarily aimed for identification, quantification, indexing and storing of all the metabolites that can be found in human body in a concentration greater than 1 micromolar. Now, with that in 2007 the very first draft of this human metabolome completed which consist 2500 metabolites, 1200 drugs, 3500 food components this was the initial data and they help in different types of domains like drug evaluation of drug metabolism, then toxicology, identification of disease, then monitoring progression of the disease and this is how the human genomics the human genome project is actually indirectly linked to the metabolome project. And finally, in 2015 we got in our hand the real time metabolome profiling in which we can study in real time how this the whole metabolome is changing in a process in living human body. So, these are the different apply application of metabolomics like biomarker discovery as I told then disease stratification and sub typing.

So, based on different metabolic profile we can classify disease we different severity scoring can be done therapeutic monitoring can be done via change via studying the metabolic pro metabolomics or metabolite profile after and before the treatment, then pharmacometabolomics is dealing with the drugs dealing with the drugs how these drugs drug is internally metabolized what are the different intermediates and their metabolic profiling. So, these two are the very common examples where we utilize this metabolomics in cancer diagnosis and monitoring. So, two different types of cancers can be compared in a single type of cancer the severity based on their staging the metabolic metabolomics or metabolite metabolite profiling can be done based on that the improvement responses to drugs can be studied. Similarly, in cardiovascular disease metabolic profiling is very important because it gives an idea of how a patient of

cardiovascular disease is dealing with the lipid metabolism oxidative stress inflammation and different types of underlying risk factor how those risk factor is actually modulating this cardiovascular disease. It helps in early diagnosis and monitoring of cardiovascular condition.

So, there is an huge applied part of metabolics in clinical diagnostics. Now, there are different metabolic databases like human metabolome database. Now, this database is basically the collection collective details of different metabolites which we can avail. One common example is human metabolome database HMDB it is one electronic database which contains detailed information about different small metabolites small molecule metabolites found in human body. Those metabolites chemical data clinical data and molecular biology biochemical data they all are stored in HMDB.

HMDB contains 2,20,945 metabolites with characterization whether they are water soluble or lipid soluble. There are 8610 protein sequences including enzymes and transporters which are linked in this metabolite entry. Now, along these data are basically mass spectrometry and NMR spectroscopy derived data, data which are derived from different samples like urine blood and CSS and along with that the NMR and MS spectra of the reference metabolites are also stored. So, you can compare with the reference metabolite with your own study. Similarly, there is metabolite metabolite is a name of metabolic database.

This database is basically focused on metabolic experiment and those experiment derived data. Now this is basically acting as a repository where the studies where these experimental data is stored and also a resource where you can access or curate your own metabolic data. So, what it contains these metabolites structure reference spectra cellular location their functions experimental metadata all are stored here. And also because this is one study or experiment the best database here you can submit your own metabolic data and also browse for metabolic data which have been stored by others. Then methylene is another web based data repository where the metabolites are identified through mass analysis and there is an annotated list of known metabolites.

Their structural information cross which are cross correlated with the MS-MS which is tandem MS and LC-MS liquid chromatography mass spectrometry based data. So, these are the different metabolic databases which help in metabolomics. So, in summary we have learnt that metabolomics is studying the is a systematic study which is dealing with small molecules or metabolites in our, but metabolomes in our body which can be exogenous and endogenous. Then targeted and untargeted approach based on whether we are studying some targeted specific known metabolites or a global comprehensive profiling of the sample. Majority of the metabolomics are based on mass spectrometry and NMR based data and there are different metabolic databases like HMDB metabolites

and methylene where from we can get differ information regarding meta different types of metabolites and their different characteristics and characterization based on clinical data as well as their chemical and biological data as well.

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