IAS, Bangalore

INSA, New Delhi

NASI, Allahabad

Science Academies' Lecture Workshop on "OMICS: Basics and Applications"

A two day lecture workshop on selected topics
on
Genomics, Proteomics, Metabolomics and Instrumentation

Conducted by
Indian Women Scientists' Association (IWSA)
Vashi, Navi Mumbai.

A B S T R A C T S

Venue:
IWSA's ICICI Multipurpose Hall,
Sector 10-A, Dr. Mar Theophilus Marg, Vashi, Navi Mumbai - 400703.

Date:
8 - 9 July, 2017

Indian Women Scientists' Association (IWSA) - www.iwsa.net -Tel: 022 27661806

From Human Genome Project to Next Gen Sequencing to Genomics

Dr. Rita Mulherkar
Professor and Scientific Officer 'H' (retired), ACTREC, Kharghar, Navi Mumbai

More than a century after Gregor Mendel propounded the theory of heredity did scientists discover that DNA was the source of genetic information. Soon after that the three dimensional structure of DNA, with the iconic image of double helix was solved. Also the three letter genetic code was deciphered. Scientists discovered enzymes that could cut and paste DNA. In the 1970's two technologies transformed the field of genetics – sequencing and cloning of genes. Simultaneously, Sanger devised methods to sequence DNA using copying reactions of polymerases. In the 1980's these techniques were used to map and identify genes linked to diseases such as Huntington's disease and Cystic Fibrosis. This opened up a new era of patient management using DNA. The technology also led to the most ambitious project termed Human Genome Project (HGP) . HGP was an international project to sequence the complete human genome – all our genes together, the first draft of which was published before time in 2001. During this time the technology advanced with leaps and bounds. Sequencing machines improved drastically. Cost of sequencing DNA and time taken to sequence DNA came down. Sanger sequencing could barely sequence 25bp initially which increased to 100bp and then to ~750 bp. Then came the era of next generation sequencing using different technologies. The latest next generation sequencing machines can generate as much data in 24 hours as several hundred DNA capillary sequencers but by a single person. This opened up the field of genomics. Today genomes of many organisms including plants have been sequenced and published.

I
Instrumentation for study of genetic variations and diseases

Dr. M. Seshadri
Ex-Head, Radiation Biology & Health Science Division, BARC, Mumbai

Completion of the human genome sequencing more than a decade back led to high expectations of being able to identify the genetic basis of more than 60% of human ailments which have a genetic basis. Studies on human genome have also led to spurt in our understanding of the variations among various population with the view to understand the evolution and migration of human population. This has been based on Single Nucleotide polymorphism, micro and minisatellite variation and mitochondrial and Y chromosome variations. Genetic diseases can arise from gain or loss of whole chromosome, major/minor deletions, copy number variations and insertion/deletion of single base or base alteration. The complexity arises when even in cases where a single gene is involved in a disease such as Hemophilia A, number of mutations reported span almost the entire gene. Other diseases may be caused by a single gene mutation but number of genes involved in the disease can vary from 2 in alpha thalassemia to over hundred genes in Hearing loss. These findings have been possible because of evolution of the instrumentation alongside growth in knowledgebase. While microscopy and Sanger sequencing were the initial workhorse of investigations, post human genome project a major spurt in technology took place. While the intitial sequencing of human genome under HGP cost around few billion dollars and took almost a decade, current technology allows human genome sequencing to be done in under 2000 dollars in few days. Further progress is expected to bring it down to under 1000 dollar. The talk will cover the importance of genetic variation and advances in instrumentation which have helped in our progress and relevance of these in terms of Indian Healthcare implementation .

II
Genomics in Health and Agriculture

Dr. Rita Mulherkar
Professor and Scientific Officer 'H' (retired), ACTREC, Kharghar, Navi Mumbai

One of the objectives of Human Genome Project (HGP) was to generate comprehensive sets of reagents and data that would create kits for genomics-based research. DNA sequencing technology and the machines for sequencing evolved. Whole genomes could now be sequenced in much shorter time and at an affordable cost. Genomic maps and sequences including databases of sequence variation, clone libraries and cell lines were generated. Genomes of disease causing mcroorganisms could be sequenced. The science of genomics showed a tremendous potential for improving health. All individuals have genome sequences which are about 99.9% identical. The remaining 0.1% is responsible for the genetic variation. Based on the genetic variation in humans the concept of precision medicine has evolved. Genetic variants which contribute to human conditions were discovered. Whole-genome sequencing could identify these genetic variants. The field of medicine is changing with the field of genomics. Genomics has become a central discipline of biomedical research. Plant genomics and the genetic modification of crops have great potential for improving human health through nutritional gains and the production and delivery of vaccines and therapeutic agents. A large number of plant genomes have been sequenced. The application of plant genome analysis has contributed to our understanding of gene annotation, construction of molecular maps and designing strategies for gene cloning.

III
NGS data analysis for dummies

Dr. Sanjeev Galande
Indian Institute of Science Education and Research (IISER), Pune

The race to read the information written on the DNA to decipher the code of life gave rise to various first generation sequencing chemistries. These commercial sequencers significantly reduced the cost of sequencing and made "$1,000 genome" a reality. These 'next generation' techniques generate high throughput sequencing information (NGS) that can be applied in the fields of comparative genomics, transcriptomics and ChIP-seq etc. Especially, comparative expression analysis of transcripts, identification of pioneer transcription factor binding regions and understanding the modulation of epigenetic marks genome-wide enabled us to gain unbiased insights into the gene regulation process. The computation algorithms involved in the analysis of NGS data are under constant improvement and play critical role in addressing various biological questions. I will use numerous examples to illustrate various steps in NGS analysis and data interpretation. Most importantly, since most of the biologists are not used to writing codes, I will describe simple workflows to analyze NGS data.

IV
From Genome to Epigenome: A new perspective towards understanding complex diseases

Dr. Sanjeev Galande
Indian Institute of Science Education and Research (IISER), Pune.

In Eukaryotes, nuclear DNA or the Genome is organized with help of basic proteins into an orderly packaged compact structure called chromatin. Out of all protein-coding 'genes' or DNA segments that each cell contains, only a fraction is used at any given time, and those genes that are never used are packaged

much tightly as compared to the ones that may be 'expressed' or used during the lifetime of a cell. Mapping and sequencing of genomes from large number of evolutionarily diverse species in the past decade revealed that sequence *per se* is not sufficient to understand genome function. It was appreciated that information in form of organization of the genome and its various modifications are also important – these are the molecular events contributing towards the 'epigenome'. It is therefore becoming increasingly apparent that 'epigenetic' marks reflect the functional state of the genome. The past decade has witnessed the explosion of information in biomedical sciences due to the availability of genome sequences and the development of high throughput techniques that assay DNA and histone modifications. I will summarize the technological breakthroughs and also discuss how they will enable us to study disease susceptibility and presumably take us towards personalized medicine.

V
Human microbiome: Indian perspective

Dr. Dhiraj Dhotre
Scientist, Microbial Culture Collection, National Centre for Cell Science, Pune.

The human body harbors diverse microbial communities, which are predominantly bacterial and referred to as 'Human Microbiome'. The trillions of microbes in and on our body contribute for many aspects of our physiology, from metabolizing otherwise indigestible carbohydrates and fats to making essential vitamins, maintaining our immune systems and acting as a first line of defense against pathogens. After the completion of human genome sequence in 2001 (International Human Genome Sequencing Consortium)[7], Julian Davies (2001), Relman and Falkow (2001) rightly argued that "although it (Human Genome Project) is the crowning achievement in biology, it would be incomplete until the synergistic activities between humans and microbes living in and on them are understood". They called for a "second human genome project" that "would entail a comprehensive inventory of microbial genes and genomes at the four major sites of microbial colonization in the human body: Oral, gut, vagina, and skin."In recent years, human gut microbiome research has moved from being an area of basic research to the advance therapeutics. After the realization of the potential role of the microbiome research in human health, several government agencies and major pharmaceutical companies across the globe have initiated many mega projects. Human microbiome research in India is still in its infancy and a lot needs to be done to understand and investigate the role of microbiota in shaping the overall health and disease conditions especially in the Indian scenario. Indian population harbours tremendous genetic and cultural diversity consisting of more than 6,000 communities and approximately 40,000 endogamous groups. Indian population is structured and thus endogamous group and geographical origin need to be essentially considered while making any molecular inferences. These variations make Indian population a perfect model to study the 'Genotype-Microbiome' association.

VI
Ethics of Biobanking and Genomics

Dr. Rajiv Sarin
In-Charge Cancer Genetics Unit and Professor of Radiation Oncology Tata Memorial Centre.

Biomedical ethics is an integral aspect of any research that is done on human subjects. It assumes special importance when the research involves studying genes or genomes in an individual, family or community or banking human tissues for future genomic research.In Biobanking, the ethical issues includes whether the donor while giving his or her informed consent has understood if the tissue sample they are donating for banking and future research will be the excess tissue left after the standard diagnostic or therapeutic procedures or will there be additional or larger procedures or biopsies done to obtain this sample and the associated risks. As the details of the research that may be done in future on stored samples is not known, the consent form for Biobanking is usually generic in which the exact nature of genomic research that may be done in future is not specified. Hence it should explain what processes will be followed to ensure that any research performed on their stored tissues in future will be in accordance with the local

and national regulations, protecting their privacy and not violating any principles of biomedical ethics.A wide range of ethical issues may arise in research on genes or genome, especially for germline genetic analysis on blood or normal tissues. These ethical issues are related to the risk of loss of privacy due to unauthorized access to data or the risk of re-identification of the research subject and their family. This may pose special concern when their genetic status predicts inherited disease risk (such as BRCA1 mutation) with the associated social stigma and discrimination by the society or in work place or by the insurers. For individuals undergoing genetic testing for a suspected hereditary condition, there is a well established process of pre-test counselling and return of results along with post test counselling. However when such disease predisposing genetic status is identified in the course of genomic research on paired tumour and normal tissues there should be a policy of return or results, how the risk of loss of privacy will be managed and what provisions for referral are made for genetic counselling and risk management. Moreover, germline analysis of a large number of genes as part of NGS multigene panels or whole exome or genome analysis greatly increases the possibility of identifying a large number of genetic Variants of Unknown Clinical Significance (VUS). These VUS are difficult to characterize and pose special challenges in counselling of the patient regarding disease risk and may create psychosocial problems. Certain technological advances such as genome editing is posing unprecedented ethical issues about which the scientific community and the society at large is yet to develop widely accepted policies. All these ethical issues related to genomic research are of greater concern when the research subject is from a vulnerable group such as minors, mentally challenged, pregnant women, illiterate and poor. The principles of biomedical ethics and special issues arising in genomics research and biobanking will be discussed.

VII

Dr. Mukesh Jaiswal

(Given last in the list)

VIII

RNAomics: an introduction to non coding RNAs

Dr. Rita Mukhopadhyaya

Molecular Biology Division, BARC, Trombay, Mumbai.

RNAomics refers to the study of non coding RNAs (ncRNA). The topic comprise of discoveries of ncRNAs, their secondary structures, classification and regulatory role in transcriptomes of model organisms and humans. As more genomes and trancriptomes came to light through high-throughput sequencing technologies (HTS), confirmation of presence of ncRNAs became convincing. Thus, in past ten years this subject got the due attention, more and more new classes of non-coding RNAs emerged and this has led to finding new regulatory RNAs. Corresponding growth in the field of RNA bioinformatics has assisted the understanding of their structure, function, tissue specific expression. The interplay between ncRNA and proteins in the RNA-focussed gene expression networks is opening avenues for new biomarker for disease diagnosis and therapy.

IX

PROTEOMICS TODAY: APPROACHES, EXPECTATIONS AND CHALLENGES

Dr. Surekha Zingde

Indian Women Scientists' Association, Vashi, Navi Mumbai.

Proteins dominated biological sciences up to the 1970s. With the advent of recombinant DNA technologies, genomics came to fore. This led to the sequence of the human genome in 2003 and a flood of genome information to date. It is however apparent that more than one protein can be made from a single gene as a consequence of variety of regulatory mechanisms at the transcription, translation and post translational

steps. Further the physiological outcome is dependent on the protein composition and the functions of these molecules. Proteomics is the science which aims to provide information about all the proteins coded for by the genome. Proteomic information deals with the number of proteins, post translational modifications, structure, interaction of the proteins with each other, their dynamic status and level of expression in healthy and diseased tissues of different origin.There are tremendous challenges in investigating over 10^6 proteins each of which are present in amounts varying in several orders of magnitude. The post translational modifications are most often difficult to identify and so also the structures and the interactions between the proteins which eventually lead to the function of the cell. Proteomics has since its inception focussed on the three technologies ie mass spectrometry, antibodies and bioinformatics to address these issues. The lecture will provide an overview of the changing technologies and their present status, and the direction in which the science of proteomics is proceeding to meet the challenges of the times. It will inform about the human proteome organization (HUPO) and its efforts to address the unknown about proteins and their functions.

X
Introduction and Basics of Mass Spectrometry for OMICS related Applications

Dr. Ajit Datar
Shimadzu Analytical (India) Pvt. Ltd., Mumbai.

Mass spectrometry (MS) is one of the most important techniques in analytical technology on which the emerging ‘‘-omics’’ approaches are based. It may provide detection and quantization of thousands of proteins and biologically active metabolites from a tissue, body fluid or cell culture etc., down to ultra-trace levels. The new developments in performance of MS technology, coupled to ease in data handling, is providing information on a better understanding of human diseases. This is leading to new molecular biomarkers, hence assisting drug targets and therapies. It is hence necessary to understand some basic concepts in Mass Spectrometry.

This would include:

1) How the large and involatile biomolecules are vaporized without decomposition and ionized so that the ions can be separated by mass analyser by using appropriate electric or magnetic field on the basis of mass to charge ratio.
2) The concept of Liquid Chromatography (LC) hyphenated to MS by using interface called as Atmosphere Pressure Ionization (API), which mainly consists of Electrospray Ionization (ESI) and Atmospheric Pressure Chemical Ionization (APCI).
3) The use of Nano-ESI with or without LC for proteomics application.
4) The use of Matrix Assisted Laser Desorption Ionization (MALDI).
5) The different mass analysers like Quadrupole, Ion trap and time of flight (TOF).
6) To understand the advantages of tandem mass spectrometry and hybrid mass spectrometry and how MS/MS or multiple stages of MS^n helps in understanding the structure of the biomarker.
7) The concept of high resolution and high mass accuracy is of great importance as the molecular weight of compounds can be measured accurately up to 4^{th} or 5^{th} decimal by MS.
8) The qualitative and quantitative analysis by using MS

In this lecture, we will briefly look at the main advances in the MS technologies, influencing genomics, transcriptomics, and proteomics, as well as metabolomics and lipidomics fields, and also the most recent MS applications to meta-omics studies.

XI

Plant secondary metabolites: structural diversity, biosynthetic pathways and biological functions

Dr. Ashok P. Giri

A. CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune.

Plants are ubiquitous sources of secondary metabolites belonging to different groups namely terpenoids, alkaloids, phenolic compounds, etc. Interestingly, plants synthesize these metabolites constitutively for endogenous functions as well as in response to biotic and abiotic stresses. Major challenge today is to identify and quantify them from various tissues, and physiological and development stages. Significant efforts are being made towards disclosing their diversity through mass spectrometry, which ultimately helping researchers to discover their biosynthetic routes and possible functions. We quote recent example of steroidal glycoalkaloids biosynthetic pathway in Solanaceous plants. The tight control of their metabolic flux in different tissues of potato and tomato revealed conserved nature of these metabolites and complexity in their accumulation. Other example is metabolic diversity detected in various species of Ocimum, which provides another illustration. Further some of these metabolites as health benefits for humans if used appropriately. We discuss our recent finding on potential dual role of eugenol in inhibiting advanced glycation end products in diabetes revealed through omics approach. Differential accumulation of several such metabolites upon biotic stresses to leaves of *Ocimum kilimandscharicum* demonstrates their potential defensive role. These results will be discussed in detail.

XII

Proteomics for understanding stress resistance mechanisms in prokaryotes

Dr. Bhakti Basu

Molecular Biology Division, Bhabha Atomic Research Centre, Mumbai.

Proteins are the cellular workhorses that fine-tune cell's physiological makeup and response to environmental challenges. Proteome, a complement of proteins produced in a cell under given set of condition, provides valuable insights into the biological pathways that impart a distinct phenotype to a cell. The field of proteomics has made rapid technological advances from separating proteins by SDS-PAGE or 2D gel electrophoresis to selective protein identification from complex mixtures by mass spectrometry. Simultaneous developments in the area of genomics and structural biology have opened up new avenues for elucidation of biological functions, pathway mapping and defining cellular mechanisms at molecular level. Here, I shall elaborate the proteomic approaches adopted to understand molecular mechanisms underlying extreme radiation and desiccation stress resistance mechanisms of *Deinococcus radiodurans*. *D. radiodurans* is gifted with an extraordinary ability to repair extensive DNA damage caused by ionising radiations, desiccation and other genotoxic stresses. Both gamma radiation and desiccation cause lethal double strand breaks, which the organism repairs efficiently by taking a short growth lag before resuming exponential growth. 2D gel based proteomic approach was used to elucidate the response of *D. radiodurans* to lethal DNA damage. Systematic investigation of proteomic modulations throughout the post-irradiation/post-desiccation recovery (PIR/PDR) revealed sequential synthesis and processing of DNA repair proteins belonging to different DNA repair mechanisms, and enhanced abundance of cellular detoxification proteins. A comprehensive picture evolved from these analyses, that uncovered multiple strategies adopted by the organism to successfully recover from lethal DNA damage, is discussed.

XIII
Omics to know the Ins and Outs of Endometrial Functions

Dr. Geetanjali Sachdeva
NIRRH, Indian Council of Medical Research, JM Street, Parel, Mumbai.

The “Omics era” which began with the completion of human genome project in 2000, has ushered in enormous opportunities to generate voluminous biological data. The field of reproductive biology also has been benefitted by swift path-breaking conceptual and technological advances in Omics. Omics technologies are being employed for investigating infertility, recurrent spontaneous abortions, Polycystic Ovarian Syndrome and also for genetic screening at preconception, preimplantation and prenatal stages. Clinical scientists engaged in the field of assisted reproduction are exploring a potential of these technologies in the selection of best quality oocytes and embryos with a goal to maximize the success rate of in-vitro fertilization and embryo transfer. In addition to human spermatozoa, oocytes, embryos, placenta; endometrium (inner lining of the uterus) is being investigated extensively for molecular phenotyping through “Omics” based tools in healthy and diseased conditions. Endometrium plays a critical role in the reproductive outcome as it provides anchorage and nutrition to the embryo throughout pregnancy. Any structural or functional defect in endometrium adversely affects the initiation and sustenance of pregnancy. Endometrial dysfunctions thus contribute to infertility and pregnancy losses. Endometrial defects also contribute to other gynecological disorders such as endometriosis, adenomyosis, endometrial hyperplasia and endometrial cancer. A multi-compartment tissue architecture with zonal variations and regenerative dynamicity makes the endometrium a very complex molecular web from the physiological standpoint. Moreover, it is known since long that endometrial “implantation competence” is a transient phenomenon, implying tightly regulated spatiotemporal gene expression in the endometrium. All these intricacies make “Single factor approach”, somewhat archaic, especially if the mechanisms underlying endometrial functions and dysfunctions are to be investigated. Several omics-based strides have been made towards profiling of human endometrium in the last 15 years. Our laboratory is also engaged in elucidating modifications in the molecular repertoire of the human endometrium during physiological and pathological conditions using proteomics/transcriptomics technologies. These investigations have revealed several novel factors/pathways which may be relevance in endometrial functions. In-depth studies of these pathways may unravel the causes of various endometrial pathologies

XIV
"Changes for Better Data Quality in the Western Blotting World: Stain-Free Gel Technology and Beyond"

Dr. Abhijit Dixit
Field Application Specialist, Bio-Rad Laboratories.

Synopsis: "Western blotting is a widely used powerful technique; however, its data quality is often criticized due to poor experimental design and practices. The Stain-Free Gel technology introduces a new practice of western blotting that offers quality controls and total protein loading control to improve quality without increasing the burden of the experimenters. In this talk the speaker will highlight issues with using House-keeping protein for normalization of western blots and will present to the audience a better method for normalization using Bio-Rad’s V3 Western Workflow which is now widely accepted by all International Journals"

XV

Advances in chromatography in OMICS application

Shailesh Damale
Shimadzu Analytical India Pvt Ltd

Metabolomics is a new approach that is based on the systematic study of the full complement of metabolites in a biological sample. Metabolomics has the potential to fundamentally change clinical chemistry and, by extension, the fields of nutrition, toxicology, and medicine. However, it can be difficult to separate highly polar compounds. Mass spectrometry (MS), in combination with, gas chromatography (GC), or high performance liquid chromatography (HPLC) is the key analytical technique on which emerging "omics" technologies, namely, proteomics, metabolomics, and lipidomics, are based. Although it is a powerful separation mode, one major limitation of RPLC is its inability to adequately retain extreme polar compounds. In particular, the major focus of LC-MS-based studies in recent years has been to improve the separation of water-soluble compounds. Quantitation of endogenous metabolites in biological fluids and particularly in serum has to overcome several issues. The effect of endogenous interferences of serum in multi-targeted metabolite profiling LC-MS/MS analysis need to be investigated by studying different sample preparation procedures.