DIGIT~Bio~TECH

LO1 SYSTEMS BIOLOGY AND OMICS TECHNOLOGY

Advanced level

AUTHOR:

VENTSISLAVA PETROVA

Contents

Contents 2
OMICS technologies towards improving the quality of life 3
OMICS Technologies and Molecular Medicine 3
OMICs technologies in diagnostics 4
OMICs analysis for revealing genetic architecture of common disease 9
Application of OMICs approach in disease treatment 10
OMICs technologies for disease prevention - future prospects 11
Application of OMICS Technologies in Toxicology 11
Toxicological studies via OMICs analysis 12
OMICs applications for toxicology risk assessment and regulatory submissions 16
Environmental OMICS and Biodiversity 17
Single-cell genomics 18
Omics Approaches in Industrial Biotechnology and Bioprocess Engineering 23
Omics-Guided Biotechnology 23
Nutri-omics Research 28
Genomics and nutrition 28
Trascriptomics and nutrition 30
Proteomics and nutrition 30
Matbolomics and nutrition 31
Nutrition and other OMICs 31
References 32

OMICS technologies towards improving the quality of life

OMICS TECHNOLOGIES AND MOLECULAR MEDICINE

Advances in omics technologies - such as genomics, transcriptomics, proteomics and metabolomics - have begun to enable medicine at an extraordinarily detailed molecular level. The rapidly decreasing costs of high-throughput sequencing and other massively parallel technologies, such as mass spectrometry, are enabling their use in clinical research and clinical practice. Exome and genome sequencing are already being used to aid diagnoses, particularly of rare diseases, to inform cancer treatment and progression and, in early efforts, to create predictive models of disease in healthy individuals. Numerous research efforts and companies are focusing on genome-wide profiles of genetic, gene expression and other omics data, such as the microbiome, as biomarkers for disease. These techniques have also been applied in identifying risk loci for disease, defining precise pathophysiology of the disease, as well as performing research in Occupational Environmental Health (OEH).

Ideally, different technologies would be combined both to help diagnose disease and to create a holistic picture of human phenotypes and disease. However, implementation of multi-omics data introduces new informatics and interpretation challenges. What are the ways in which integrative omics can impact medicine by helping to manage health, as well as diagnose and treat disease? In fact, to improve the treatment outcomes and reduce adverse events that matter to both the clinician and patient one of the perspective solutions is to develop the personalized medicine - customized medical treatment created to fit individuals' characteristics, needs and unique molecular and genetic profile (Fig. 1).

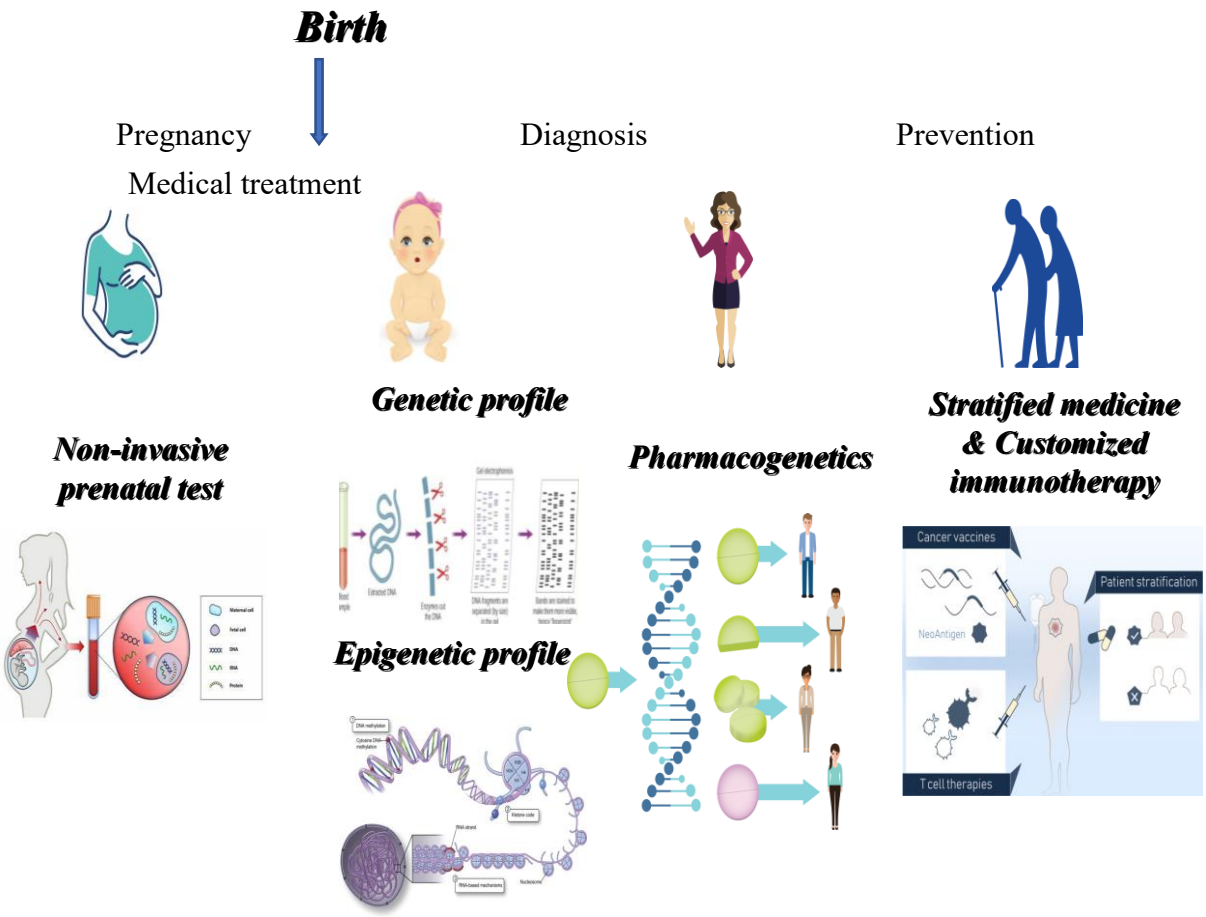

Fig. 1. Personalised medicine throughout the lifespan

1. OMICs technologies in diagnostics

PRENATAL DIAGNOSTICS

Nearly 10% of pediatric diseases and 20% of infant deaths are due to Mendelian diseases. Genetic defects can be a major threat to the baby's health. To overcome this issue, last developments in the omics technologies are employed in a person's life even before birth. Prenatal screening has become widespread

in many countries for examining not only abnormalities of a fetus but also the possible risks of diseases or abnormal conditions after birth. When a genetic predisposition exists, such as cystic fibrosis or muscular dystrophy, consulting a medical specialist or genetic diagnosis company become of great importance. The latest innovations in molecular and cellular biology allowed various genetic diagnoses for fetuses. Till recently, different invasive technologies have been applied such as amniocentesis and chorionic villus sampling. Although the procedures have been improved, there is still risk of miscarriage and serious side effects. One of the innovative attempts to avoid the side effects of conventional invasive technologies is to isolate and analyze fetal nucleated cells from maternal blood. Unfortunately, the number of fetal cells that can be isolated is about 3-5 cells per 30 ml of maternal blood. The huge jump in the field of non-invasive prenatal testing (NIPT) has been made after the advancement of omics technologies and their combination with various experimental approaches. The use of DNA amplification technology combined with NGS allowed the direct detection of the fetal DNA in maternal blood. Currently, such approach is used to determine the sex and Rh blood type, and to analyze chromosomal aneuploidies. The cell-free fetal DNA is typically released in maternal blood during the process of apoptosis of placental cells. The fragmented cell-free fetal DNA is ~200 bp long and has a half-life of ~16 min and can be detected in maternal plasma from early pregnancy (5-7 weeks). In addition to fetal DNA, fetal RNAs can also be used for NIPT. For example, fetal trisomy 21 can be diagnosed by checking the allelic ratio of PLAC4 mRNA, produced from chromosome 21 in the placenta. If a baby receives heterogeneous alleles from the parents, an unequal allelic ratio (2 : 1) suggests trisomy.

Integrating fetal genomic, transcriptomic, proteomic, metabolomic, and epigenetic data will help pushing forward the prenatal genetic diagnostics. In this way, any genetic disorder could be identified early in the pregnancy by applying non-invasive approaches. However, some concerns occur regarding raise of serious ethical issues possibilities when advanced genomic technology are used. Therefore, further studies on the impact of new technologies on society, such as possible discrimination based on genetic information, undetermined harmful effects of nanotechnologies on humans and nature, and copyright infringements due to the development of information technology should be considered (see LO8).

LAB-ON-A-CHIP TECHNOLOGY IN DIAGNOSTICS

Rapidly after the development of micro-electro-mechanical systems (MEMS), the potential use of these miniaturized platforms for various applications has been revealed. During the past few decades, interest in biological or biomedical MEMS (BioMEMS) has been significantly increased and it has found widespread applications in a various area life science including diagnostics, therapeutics, drug delivery, biosensors and tissue engineering. These integrated systems are also known as “lab-on-a-chip” (LOC) or “micro-total analysis systems” (μTAS) and provide a solid way for miniaturization, integration, automation, and parallelization of chemical processes.

Lab-on-a-chip devices integrate multiple laboratory functions on a single chip (from few square millimeters to a few square centimeters in size). These platforms provide complex chemical and/or biological analyses which can offer cheaper, faster, controllable and higher performance of bio-chemical assays at a very small scale when compared with conventional laboratory tests. They are capable of handling extremely small fluid volumes down to less than a few picoliters (few micro-droplets of whole blood, plasma, saliva, tear, urine or sweat). This fact is very important in many clinical trials where usually very small volumes of patient samples are available. Moreover, the automation with eliminating the human interfering parameters can increase the reliability in the analysis.

LOC has been used in different applications, such as: DNA extraction and purification, PCR, qPCR, molecular detection, electrophoresis, etc.

LOC for DNA Extraction and Purification: Often for diagnostic assays nucleic acid have to be purified from cells. The process of DNA extraction generally includes two steps: cell lysis by disrupting cellular and nucleus membrane, and DNA purification from membrane lipids, proteins, and RNA. Cell lysis on LOC devices can be achieved through several types of lysis methods: chemical, thermal, ultrasonic, electrical, mechanical or some other type of lysis. Thew well known DNA purification methods through extraction columns, using the DNA adsorption on silica beads under certain buffer conditions, are also adaptable to microfluidic techniques.

LOC Devices for PCR, qPCR, and Molecular Detection: After DNA extraction, the most frequent analysis is the PCR. The importance of PCR in genomic analysis led to the development of numerous LOC devices. Miniaturization of volume and the high surface to volume ratio defines rapid thermal transfer and more accurate PCR. This analysis requires a post-analysis and amplicons' size detection, which generally is carried out by electrophoresis. With the introduction of LOCs, DNA electrophoresis was one of the first molecular processes that was integrated on a chip. This miniaturization diminishes even more the overall process time and reduces the reagent consumption.

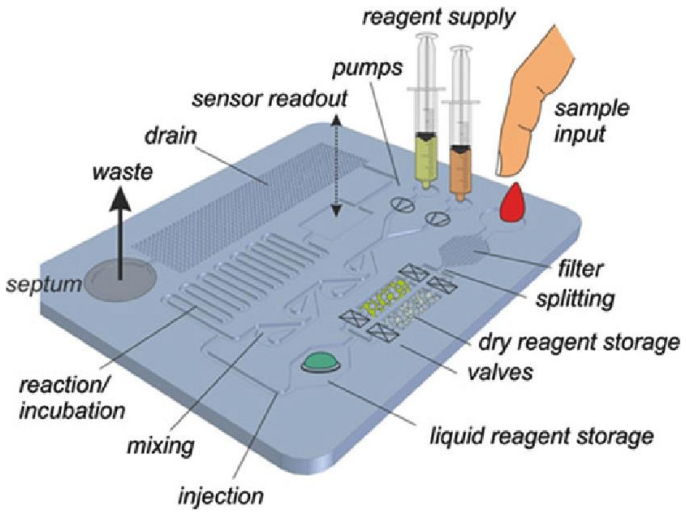

Fig. 2. Lab-on-a-Chip device

qPCR is another technique that was modified to LOC devices. Using ultra-fast pressure controller and fluorescence reader, ultra-fast qPCR microfluidic system had been developed for the molecular detection of diseases like Anthrax and Ebola. The detection efficiency is identical to commercial systems and results are achievable in less than 8 minutes being 7-15 times faster than traditionally used ones. Other advanced field is digital microfluidics that deals with emulsion and droplets within LOC devices. Ultra-low quantity of DNA can be captured within droplets, and detection limits can be enhanced with one copy number detection within LOC droplet qPCR.

Genomic Application: Due to the numerous sequencing projects, such as the Human Genome Project, the field of genomics gained a tremendous attention in recent years. Next-generation sequencing techniquess of DNA which are currently under development include sequencing-by-hybridization (SBH), nanopore sequencing, and sequencing-by synthesis (SBS), the latter of which includes many different DNA polymerase dependent strategies. Using LOC approach allows the drastic reduction in cost and duration of high-throughput sequencing.

When studying the complex DNA samples expression an integration of multiple biosensors in connection with DNA microarrays is required. The most essential characteristics of these LOCs are miniaturization, speed, and precision. This technology offers vast possibilities for rapid multiplex analysis of nucleic acid samples, including the diagnosis of genetic diseases, detection of infectious agents, measurements of differential gene expression, drug screening, or forensic analysis.

Biochemical Applications: LOC strategy is also applicable in coupling enzymatic and immunological assays on a single-channel microfluidic device. Typical example is the simultaneous measurement of glucose and insulin. Key for the effective realization of such glucose/insulin monitoring is the integration of relevant sample pretreatment/cleanup procedures essential for whole-blood analysis.

This innovative LOC approach can be applicable to the incorporation of other analyzes and additional sample handling steps, as is desired for creating miniaturized clinical instruments.

Proteomics: Proteomics is one of the great scientific challenges in the post-genome era. LOC devices are useful for creating advanced techniquess to answer complex analytical problems, such as proteome profiling. LOC devices could be applied in four key areas of proteomics: chemical processing, sample preconcentration and cleanup, chemical separations, and interfaces with mass MS. LOC being a miniaturized device could be used for separation and detection of proteins. This process is characterized with less reagent consumption, easy operation, and very fast analysis. Another benefit is that the data from the LOC devices can easily be compared to those obtained using 2D-PAGE. In order to quantitate proteins a LOC MS protein profiling device is developed. Some authors describe LOC devices that are used for direct infusion into the mass spectrometer, including Capillary Electrophoresis (CE) separation, on-chip sample digestion, and infusion or CE before MS analysis. Researchers use electrospray ionization MS and focus on the interface design between LOC and the mass spectrometer.

Biosensors: For detecting target analytes LOC biosensors are also developed. Such devices intricately link a biological recognition element with a physical transducer converting the biorecognition event into an electrical signal. There are two types of LOC biosensors: bioaffinity and biocatalytic devices. Bioaffinity devices are based on selective binding of the target analyte to a surface-enclosed ligand partner (e.g., antibody, oligonucleotide). For example, in hybridization biosensors, there are immobilized single-stranded (ss) DNA probes onto the transducer surface. When a duplex is formed it can be detected following the association of an appropriate hybridization indicator. On the other hand, in biocatalytic devices, an immobilized enzyme is used for recognizing the target substrate. For example, sensor strips with immobilized glucose oxidase for monitoring of diabetes (for more information see LO3).

Cell Research: Using LOC devices a small-scale experiment with cells could also be performed. The single cell is positioned within the microchannel in a predetermined location, making its handling and manipulation easy. The microsystem allows handling small volumes of fluid containing small quantities of analyte. LOC systems also provide precise control of the local environment around the cell, monitoring the physical or chemical nature of the surface to which the cell adheres, the local pH, and temperature. When the cell is to be exposed to a different type of stimuli, numerous and often complex fluid-handling components can be used. If the cell is a subject of further analysis, lysis and necessary analytical methods could also be introduced onto the same platform. Thus, sample loss or dilution is avoiding, which would inevitably occur if multiple devices were used.

Drug Development: The industry needs for development of new drugs and to predict their potential behavior in animals and cells triggers other analytical applications of LOC systems: e.g., analytical systems to monitor and optimize the production of protein drugs such as therapeutic antibodies; assays based on primary human cells that could predict performance in human clinical trials and pothers. These LOC systems are proven to be highly reproducible, easily manipulated, and could be used routinely.

2. OMICs analysis for revealing genetic architecture of common disease

Identifying genetic markers that are associated with a specific disease feature is key issue when estimating the disease risk. As the genetic information of an individual remains mostly unchanged, the mutated sequences can be reliable markers for certain diseases. In this regard one of the important applications of omics technology is the genome-wide association studies (GWASs). They can identify common genetic variants, generally single nucleotide polymorphisms (SNPs), associated with certain disease features. In general, SNP information is collected from two groups, a patient group and a healthy control group. Then it is statistically analyzed, and the SNPs associated with the specific disease characteristics of the corresponding patient group re identified.

For example, variations in the apolipoprotein E (APOE) gene are a well-known genetic factor that is correlated with the late-onset Alzheimer's disease (LOAD). After the assessment of 502,627 SNPs in 1086 AD patients, it was demonstrated that rs4420638 on chromosome 19 is the strongest marker differentiating AD patients from the control group. Other important success is achieved during investigation of hereditary cancers. BRCA1 and BRCA2 genes are one of the most important genetic markers for hereditary breast and ovarian cancers. These genes encode tumor suppressor proteins, which help repairing damaged DNA through homology directed repair. Some mutations in BRCA1 and/or BRCA2 significantly raise the risk of developing breast and/or ovarian cancers. For BRCA1, two mutations were detected: 185delAG and 5382insC. Approximately 55-65% of women who inherit one of the harmful BRCA1 mutations and around 45% of women who carry the BRCA2 mutation have developed breast cancer by the age of 70. Numerous other disease markers such as autosomal-dominant colorectal cancer, Li-Fraumeni multi-cancer syndrome, mucolipidosis IV, and hypertrophic cardiomyopathy, have also been successfully identified using NGS.

However, for the most common diseases such as diabetes, obesity, schizophrenia and autism, a combination of multiple genetic and environmental factors is responsible. A growing number of evidence indicates that genetic markers commonly explain only a part of the total trait variation. One of the missing crucial factors are the epigenetic variations. Various epigenetic modifications have been reported to be associated with the early stage of different cancers (abnormal growth of a tumor). Moreover, it is well known that colorectal cancer is more frequently identified in men than in women. That is why it has been suggested that estrogen may help suppress colorectal cancer. In addition, the epigenetic profiles of adjacent mucosa, and non-adjacent mucosa samples of 95 colorectal cancer patients have revealed that the promoter of the MGMT, a DNA repair gene, was frequently methylated. Up to now, thousands of genomic loci have been identified and associated with various human diseases. However, the difficulties arise when these genes have to be characterized in the context of the molecular pathophysiology and the corresponding interacting genes and pathways.

3. Application of OMICs approach in disease treatment

The main aim of personalized medicine is to decide on the most effective treatment option for an individual. In this regard, the simultaneous quantification of multiple protein markers can play an important role in subtyping tumors and selecting optimized therapies for each subtype. High-throughput, single cell-based protein quantification techniques, such as CyTOF, can be excellent alternatives for immunohistochemistry-based methods.

For example, in the last decade a new approach for treating cancers has emerged - allowing the patient's immune system combat his or her cancer, called adoptive cell transfer (ACT). Glioblastoma multiforme (GBM) is the most frequent and aggressive type of malignant brain tumor. Although GBM has been treated with surgical removal, radiation treatment, and chemotherapy methods, there have been no noticeable effects. Over the past 20 years, numerous clinical trials have been performed to use the immune system to cure cancers such as GBM. Among numerous ACT approaches, particular attention has been paid on the dendritic cell (DC)- and T cell-based vaccines. They can penetrate and induce anti-tumoral response in the brain and show great potential for effectively killing various cancers including the malignant and advanced types. Autologous DCs, obtained from tumor tissue or in the form of a tumor lysate of a patient, can be grown *in vitro* in the presence of tumor-associated antigens (TAAs), a method called 'pulsing'. The pulsed, autologous DCs are reintroduced into the patient to stimulate specific, cell-mediated, anti-tumoral cytotoxicity against GBM and other malignant cancers. More revolutionary approaches, e.g. engineering specific immune cells using zinc-finger nucleases (ZFNs) and CRISPR technology, have also been attempted.

Every year, thousands of patients undergo a transplantation of organ or hematopoietic stem cells. However, mortality among such patients remains very high. A standard procedure for matching donors with recipients involves human leukocyte antigen (HLA) typing. However, it becomes clear that non-HLA factors can also considerably affect prognosis and development of graft-versus-host disease (GVHD). In order to overcome this, many omics applications can be used to determine optimal donor–recipient matches, as well as to examine different markers of rejection. For example, free DNA sequencing can be used to assess the severity of organ rejection as well as to simultaneously detect the eventual presence of viral DNA as a marker of infection. Additional omics data, such as RNA or protein expression, may also be utilized to evaluate the donor–recipient compatibility. Thus, integration between several omics' technologies may be used as a useful tool for transplant biology.

Furthermore, the microbiome has been associated with many common human diseases. Whereas causality is simple in genomic data, where DNA influences phenotypes, it is more difficult to unravel whether microbiome composition is a cause or effect of disease. Nevertheless, it is evident that patients with inflammatory bowel disease, type 2 diabetes and obesity, have different microbiome profiles from those of healthy controls. In addition, the microbiome has a great influence on immune function, which in some cases has been putatively causally related to disease in animal models. As our understanding of the microbiome advances, integrative assessment of this and other omics technologies data will help for our better understanding of human disease. Recently, it has been shown that human genetic profiles

influence overall gut microbiota composition. Additionally, interactions between human genetics and microbiomes have been shown to influence disease manifestation. Similarly, metabolic signaling between hosts and their microbiomes has become an area of active research, and there is increasing evidence that a number of metabolites secreted from gut bacteria may play a role in human disease. Therefore, it is obvious that integrated analysis across genome, metabolome, microbiome and other omics profiles will provide means for managing health and combating disease.

4. OMICs technologies for disease prevention - future prospects

Other important goal of personalized medicine is to prevent the development of various diseases before their onset. Current preventive treatments are limited to surgery - removing a tissue or an organ if a person possesses relevant cancer markers. However, this approach is not applicable to diseases like Alzheimer, Huntington's disease, and congenital muscular dystrophy. One of the recent approaches to address such issues is the nutrigenomics-based diet treatment. Omics profiling can be an effective approach in detecting large-scale or pathway-level changes, can show patient-specific trends and add statistical support through repeated measurements.

APPLICATION OF OMICS TECHNOLOGIES IN TOXICOLOGY

The use of omics tools to evaluate environmental health status is becoming more and more important in environmental management and risk assessment. Omics permit connecting molecular events and adverse effects with biological levels of organization relevant for risk assessment schemes. The analysis of the RNA transcripts level or cellular proteomics have already gained particular interest. These studies helped in better understanding the interactions between the stress factors and the organism, but also to unravel the modes of action (MoA) of various toxic substances. Therefore, from a toxicological point of view, omics techniques allow to effectively and accurately generate important information on substance-induced molecular perturbations in cells and tissues that are associated with adverse outcomes.

Cellular regulatory pathways involve a range of different (bio)molecules, which are characterized with different physicochemical properties and display complex non-linear interactions. A single-omics technique can measure biomolecules of a specific type, RNAs or proteins, and frequently these are not even the entirety of biomolecules of one type but a smaller subset of them. For example, the detection of short and long RNAs, or short peptides and proteins requires different transcriptomic or proteomic approaches. Only the use of multi-omics, also known as cross-omics approaches, permits to directly identify a significant fraction of a pathways' response to chemical exposure. Thereof, utilizing multi-omics strategies for toxicological research questions has greater importance.

Such system toxicology approach, integrating classical toxicology with quantitative analysis of large molecular networks and functional changes occurring across multiple levels of biological organization, is promising in analyzing the full spectrum of xenobiotics toxicity. A systems toxicology approach includes the following elements:

- Absorption and distribution of the xenobiotic within the biological system.
- Transformation of the xenobiotic and generation of toxic and/or non-toxic intermediates.
- Interaction of the xenobiotic and/or its products with the cellular targets.
- Modifications in the cellular molecules including genes, proteins, and lipids.
- Structural manifestations of target organ toxicity.
- Functional manifestations of toxicity, and
- Elimination of the xenobiotic and/or its intermediates from the biological system.

1. Toxicological studies via OMICs analysis

For better understanding the cellular response to chemical exposure, studies should be focused on different levels: transcriptomics, proteomics, phosphoproteomics, and metabolomics. Epigenomics should be considered as well for specific cases.

Transcriptomics. Transcriptomics main aim is the comprehensive detection of RNAa in the cell. A pathway response in transcriptomics is primarily detected via a known set of target genes that are found differentially expressed. The techniques applied to determine changes in gene expression in response to exposure to a xenobiotic include: northern hybridization, quantitative real time PCR (QRT-PCR), subtractive hybridization, serial analysis of gene expression, microarray analysis, and next generation sequencing. Depending on the purpose of the study, one or more of these techniques may be applied to define either partial or global expression profiles in the tested biological sample. However, the most commonly used gene expression profiling techniques are QRT-PCR, microarray analysis, and NGS. If the aim is to investigate the expression profile of a single or a limited number of transcripts, QRT-PCR analysis is the method of choice. QRT-PCR analysis is a relatively simple technique that includes the reverse transcription of mRNAs followed by PCR amplification of the resulting cDNAs using primers specific for the targeted gene as well as for one or more house-keeping genes. The amount of the PCR amplified gene products are than quantified. As QRT-PCR can accurately determine the expression of an individual or limited number of transcripts, it has some limitations from a system toxicology perspective. In order to study the systems-level toxicity of a given xenobiotic, it is necessary to employ a high throughput techniques that is capable of determining expression levels of the entire transcriptome. That is why the microarray and next generation sequencing are the most popular global transcriptome analysis techniques applied in toxicology research.

The microarray technique represents a major breakthrough analyses in the post-human genome sequencing era. Like Southern hybridization technique, the microarray approach is grounded on the

ability of DNA molecules to bind or hybridize in a nucleotide sequence-specific manner. Microarray can be considered as a high throughput northern blot hybridization in which almost all the genes that are expressed in a biological sample at a given time can be detected simultaneously. Therefore, it allows to detect small differences in the global gene expression profiles in a given biological sample in response to xenobiotic(s) exposure. The microarray technique involves the isolation of high quality RNA, reverse transcription of the mRNA to synthesize cDNA, cRNA synthesis and labeling to generate "probes", hybridization of the "probes" with "targets" present on the microarray, detection of the signal intensity of the hybridized probe-target complex, and analysis of the resulting data to determine the levels of expression of the genes of interest. Microarray-based analysis has been used in different areas such as environmental toxicology, drug development, and occupational toxicology. The data obtained provides a snapshot of all the genes whose expression levels are affected in response to xenobiotic exposure.

Next generation sequencing (NGS) is the most recently developed global transcriptome analysis. This technique can be applied to determine a global gene expression profile in RNA obtained from different biological samples. The key steps involved in NGS are RNA isolation, removal of abundant RNA molecules such as ribosomal RNA and globin RNA, preparation of sequencing libraries, PCR amplification, and sequencing of the libraries. The resulting data is analyzed, and the quantity of the individual transcripts is determined.

Proteomics: Transcriptomics is often the first choice for systems-level investigations as numerous well-established measurement methods have been developed during the years. However, protein alterations can be considered to be closer to the relevant functional impact of a studied stimulus. Although mRNA and protein expression are tightly related through translation, their correlation is limited. The mRNA transcript levels only explain about 50% of the variation of protein levels. This is due to the additional levels of protein regulation including their rate of translation and degradation. Moreover, the regulation of protein activity does not stop at its expression level but is often further controlled through posttranslational modifications. From toxicology point of view such examples for important post-transcriptional regulation include: the tight regulation of p53 and hypoxia-inducible factor (HIF) protein-levels and their rapid post-transcriptional stabilization, e.g., upon DNA damage and hypoxic conditions; the regulation of several cellular stress responses (e.g., oxidative stress) at the level of protein translation; and the regulation of cellular stress response through protein phosphorylation cascades. The most often applied proteomics analyzes being of key importance for toxicology include: 2D polyacrylamide gel electrophoresis, gel-free liquid chromatography, mass spectrometry and analysis of the posttranslational modifications.

2D polyacrylamide gel electrophoresis (2DGE) is used to study perturbations on the proteome based on changes in protein expression. The 2DGE technique relies on the separation of proteins based on their pH (charge) as well as their size. It has the capability to separate and visualize up to 2000 proteins in one gel. The first dimension, which is known as isoelectric focusing (IEF) separates the proteins by their isoelectric point (pI). The second dimension further separates the proteins by their mass. State-of-the-art image acquisition and analysis software allow the comparison of control and treated samples helping to identify the differentially regulated proteins by their relative intensity. A variant of 2DGE is

difference gel electrophoresis (DIGE) which is based on labeling of proteins with fluorescent cyanine dyes (Cy2, Cy3 and Cy5). Although 2DGE is a powerful tool to identify many proteins, the approach has its limitations. The main constraint is that not all proteins can be separated by IEF, such as membrane, basic, small (< 10 kDa) and large (> 100 kDa) proteins. The second limitation is that less abundant proteins are often masked by the abundant proteins in the mixture.

Other frequently used techniques are the gel-free liquid chromatography mass spectrometry (LC MS/MS). The LC approach employs the differences in the physiochemical properties of proteins and peptides, i.e., size, charge, and hydrophobicity. 2D-LC can be applied to fractionate protein mixtures on two columns with different physiochemical properties and maximize the separation of proteins and peptides in complex mixtures. Mass spectrometry is widely believed to be the key technology platform for toxicoproteomics. It main advantages include unsurpassed sensitivity, improved speed and the ability to produce high throughput datasets. Due to the high precision of MS, in tissues and biological matrices can be detected peptides in the femtomolar (10^{-15}) to attomolar (10^{-18}) range. This is extremely beneficial in comparative analysis where simultaneous analysis of control and treated samples are carried out. Thus, comprehensive understanding of how stimuli affect the proteome, and the subsequent identification of potential biomarkers can be achieved.

As system biology requires accurate quantification of a particular set of peptides/proteins across multiple samples, targeted approaches for biomarker quantification have been also developed. Such technique is the selected reaction monitoring (SRM). SRM measures peptides produced by the enzymatic digestion of the proteome in triple quadrupole MS. An SRM-based proteomic experiment starts with the selection of a list of target proteins, derived from preceding experiments or data search such as a pathway map or literature. This step is followed by 1) selection of the proteotypic target peptides that optimally and uniquely represent the protein target, 2) selection of a set of suitable SRM transitions for each target peptide, 3) detection of the selected peptide transitions in a sample, 4) optimization of SRM assay parameters, and 5) application of the assays to the detection and quantification of the proteins/peptides. The major advantages of the SRM technique are: 1) possibilities to monitor tens to hundreds of proteins during the same run, 2) possibilities for absolute and relative quantification, 3) highly data reproducibility, and 4) yielding absolute molecular specificity. However, some limitations of this technique also exist: 1) only a limited number of measurable proteins can be included in the same run, and 2) even with its high sensitivity it cannot detect all the proteins present in an organism.

Another MS-based targeted approach known as parallel reaction monitoring (PRM) has also been employed. It is centered on the use of next-generation, quadrupole-equipped high-resolution and accurate mass instruments (mainly the Orbitrap MS system). This technique is closely related to SRM but allows for the measurement of all fragmentation products of a given peptide in parallel.

Other key approach is the analysis of the posttranslational modifications (PTMs) occurred due to the toxic exposure. They represent an essential mechanism for regulating the cellular proteome. PTMs are chemical modifications that have an important role in regulating activity, localization and interactions of cellular biomolecules. The identification and characterization of proteins and their PTM sites are very

important to the biochemical understanding of the PTM pathways. This knowledge provides deeper insights into the possible regulation of the cellular physiology triggered by PTM. Some examples of PTMs include phosphorylation, glycosylation, ubiquitination, nitrosylation, methylation, acetylation, lipidation and proteolysis. During the past decade, MS-based proteomics has proven to be a powerful tool for the identification and mapping of PTMs. Moreover, the MS-based approaches took great advantage from the progress in MS instrumentation making the analysis more sensitive, accurate and with higher resolution for the detection of less abundant proteins.

Metabolomics: The metabolome is different to the other omics analysis, as it is dealing with a collection of chemically highly heterogeneous molecules. The metabolome is typically defined as the complete balance of all small molecule metabolites (< 1500Da) found in a specific cell, organ or organism. Thanks to initiatives like the Human Metabolome Project, the endogenous metabolome of human and animal model organisms has been mapped to a large extent. Metabolome measurements employ either nuclear magnetic resonance (NMR)-based detection or gas or liquid chromatography hyphenated to MS. NMR based approaches have the capacity to quantify the studied metabolites and constitute the gold standard for the structural elucidation of unknown metabolites. MS-based approaches surpass NMR in terms of sensitivity and can be run in an untargeted or targeted manner.

Metabolomics differs from proteomics and transcriptomics in several aspects:

(i) detects the response at very different levels of a pathway, and
(ii) simultaneously capturing molecules of a pathway that respond on extremely different time scales.

The progression of metabolomics has made it routinely and highly useful in many toxicology-related studies. It has remarkable potential in screening drug-induced cellular or organ toxicity. Rapid assessment of biological samples, together with mathematical and statistical analysis (chemometrics) of obtained data serves as a powerful tool for drug safety assessment. For example, if a drug or chemical substance can cause hepatic toxicity by inducing cellular oxidative stress, metabolomics analysis of the drug-exposed biological sample can provide direct information about metabolites that are markers of the oxidative stress. It can also identify metabolites that are considered as known biomarkers of hepatic injury, thus suggesting hepatic pathology. In addition, metabolomics analysis can also provide information about metabolites that are indirectly associated with toxicity.

Metabolomics has several advantages over conventional clinical pathology assessment. For instance, a study examining an unknown compound has shown that this compound increases serum cholesterol levels. Upon the utilization of metabolomics, it has been observed that, apart from cholesterol, the same compound also raises several phytosterols, indicating that the higher cholesterol is a result of the sterol absorption in the gut. Such findings clearly indicate that metabolomics has a much higher potential in terms of identifying exact toxicological endpoints.

Epigenomics. Main focus of epigenomics is to study the chemical modifications of the DNA and proteins organizing the three-dimensional structure of genomic DNA. Nowadays, DNA methylation is assessed by a modified genome sequencing approach (Methylome-seq). However, this approach still requires high coverage associated with still significant sequencing costs. That is why the targeted microarray-based approaches are most frequently used. DNA methylation includes two different chemical modifications with distinct functional consequences - 5-methylcytosine and 5-hydroxy-methylcytosine. Histone modifications are analyzed using Chromatin-immuno-precipitation method applying antibodies specific for the targeted modification followed by sequencing (ChIP-seq). Typically, to get a comprehensive picture, several modifications need to be detected. That requires performance of several ChIP-seq analysis simultaneously. As ChIP-seq needs more sample material than the other epigenomic approaches, such a strategy may be hard to implement in toxicological studies. ATAC-seq is an alternative approach that does not investigate chemical modifications directly, but rather one of the modifications' main consequence, DNA accessibility.

However, knowledge on the involvement of regulatory pathways and specific epigenetic modifications is still scarce. Epigenomics is thus of limited value when a pathway-based data integration approach is used. However, epigenomic data have proven highly valuable for trans-generational studies or when trying to predict long-term effects of chemical exposure based on omics data of short-term exposure.

2. OMICs applications for toxicology risk assessment and regulatory submissions

In the last decades, 'omics technologies have been extensively used in research and it was demonstrated that they are capable to provide a profound insight into the biochemistry and physiology of the cell and any harmful effect of xenobiotics. This has led to an enthusiastic approval by research toxicologists. Hopes were expressed that 'omics technologies would provide the tools to identify a set of biomarkers of adverse effects and their modes-of-action. Thus, the prediction of human effects during substance hazard assessment will be improved and a contribution will be made to the development of alternative methods to animal testing. Despite this, the translation of 'omics into the regulatory domain remains at best cautious.

The European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) has, therefore, organized a series of workshops to evaluate the possibilities and challenges of omics techniques in chemical risk assessment. The most recent workshop concluded that omics approaches contribute to answering relevant questions in risk assessment including:

(i) the classification of substances and definition of similarity,
(ii) the elucidation of the mode of action of substances, and
(iii) the identification of species-specific effects and the demonstration of human health relevance.

However, a need was identified to increase the reproducibility of omics data acquisition and data analysis, as well as to define the best practices. Furthermore, the regulatory use of any test method is closely related to the specific legal framework that is relevant for the substance under investigation. Considering the REACH Regulation (EP and Council of the EU, 2006), "omics-based data could potentially be used for regulatory submissions". To use omics in REACH dossiers, the REACH Regulation lists different requirements: specific standard information that cover specific physico-chemical, toxicological and ecotoxicological limits. These data are evaluated during hazard and risk assessment, and they are analyzed to determine risk management alternatives, respectively. The application and integration of omics tools may be useful in different levels of regulatory hazard identification and assessment contributing to:

1. Classification and labelling (C&L) of substances.
2. Weight-of-evidence (WoE) approaches to elucidate the MoA of the substance under investigation.
3. Substantiation of chemical similarity.
4. Determination of points-of-departure (PoDs) for hazard assessment.
5. Demonstration of species-specific effects and human health relevance.

The fast development of omics technologies pose several challenges for facilitating their use for hazard assessment, particularly from the regulatory submission standpoint. Sets of 'big data' must be condensed applying complex approaches and applying specific knowledge to obtain information that is pertinent for hazard assessment. Moreover, knowledge in this rapidly evolving area is not necessarily familiar to many researchers working in the regulatory settings. Accordingly, the lack of regulatory approval of omics technologies is not only linked to a lack of best practices frameworks and quality criteria, but also to a lack of confidence in the analysis and the level of uncertainty with respect to what constitutes enough data. Scientists and regulators have to first gain experience with the data obtained from this new technology and then to build confidence in its applicability.

ENVIRONMENTAL OMICS AND BIODIVERSITY

Generally, most of the environmental studies comprise culturing of isolated microorganisms or amplifying and sequencing conserved genes. However, some difficulties exist in understanding the complexity of large numbers of various microorganisms in an environment. This led to the development of new techniques allowing to enrich specific microorganisms for upstream analysis, such as single-cell isolation and analyses. Advanced tools in metagenomics and single-cell genomics (SCG) are paving the way to new methods of studying and understanding our environment.

Early environmental studies of organism diversity required samples cultivation to increase DNA quantity before genomic libraries to be constructed. At the time, sequencing of 16S and 18S rRNA was

the best option available to define environmental diversity. However, it became clear that majority of sample communities represents uncultivable species. Fortunately, polymerase chain reaction (PCR) of rRNA allows access to both cultivable and uncultivable sample diversity. Because PCR can be performed directly on environmental samples without cloning, it can be utilized to amplify targeted genes directly from the environment.

Importance of sequencing both noncoding rRNA and protein coding genes is growing, with specific interest in reconstructing bacterial genomes. A comprehensive sampling of all genes from all organisms existing in a complex sample is achievable via shotgun metagenomic sequencing. Using this method make possible to evaluate bacterial diversity and to detect the abundance of microorganisms in a specific environment. Moreover, it allows the reconstitution of complete genomes of uncultivable microorganisms. In addition to the discovery of new archaeal, bacterial, protozoan, algal, fungal, and eukaryotic species, this technique is crucian for the reconstruction of viral genomes. This was a critical advancement taking into account that viruses lack a shared universal phylogenetic marker such as bacterial 16S rRNA or eukaryotic 18S rRNA.

Another approach is the metatranscriptome approach which rather than using genomic DNA (gDNA) involves the collection of the entire RNA community. The latest is used to construct cDNA libraries that are sequenced. Although RNA is less stable than DNA, it gives a snapshot of the current state of the community by revealing up and down regulated genes. To date, one of the most important metatranscriptome studies are performed for the marine water communities.

1. Single-cell genomics

Metagenomic sequencing is a helpful tool for understanding environmental communities. However, assembling gene catalogs and composite genomes can be challenging. Furthermore, it is difficult to distinguish between genes that originate from the same or different organisms in genomes that are already assembled (Fig. 3). Therefore, there is considerable interest in developing a system to understand the organization of discovered genes and pathways within genomes.

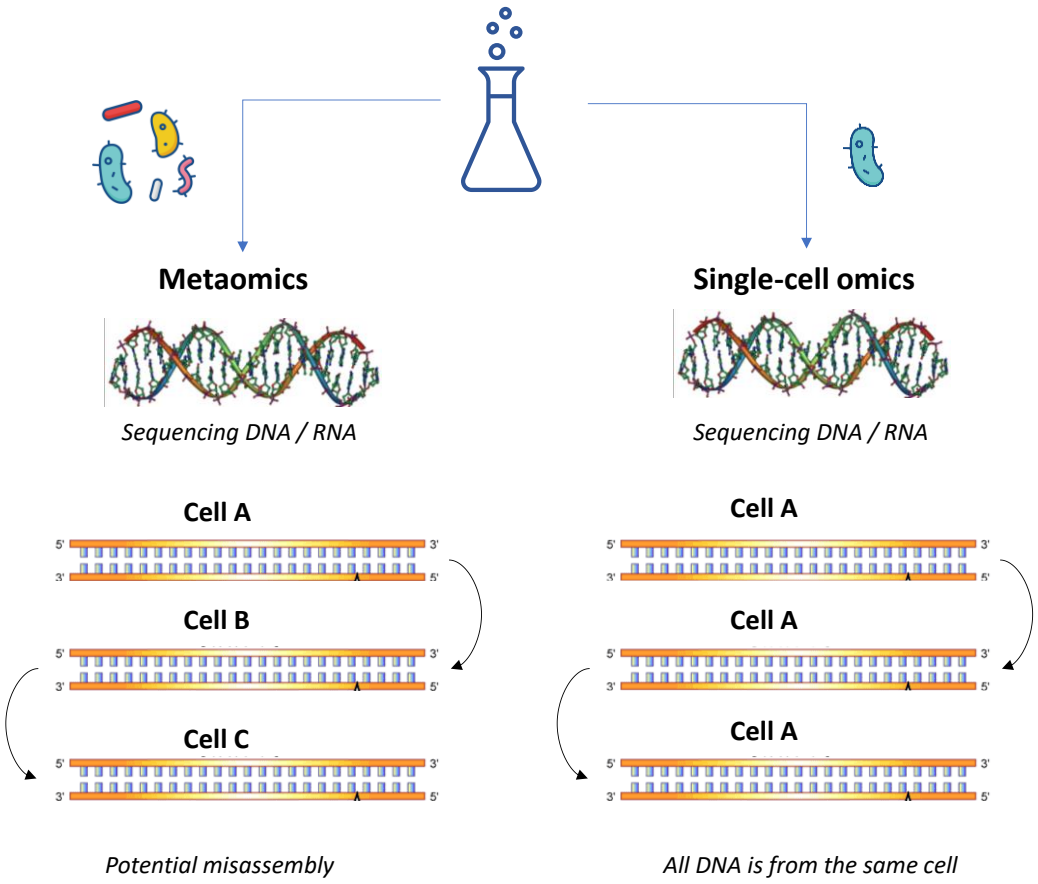

Fig.3. Metagenomics vs single cell genomics

The challenge of sample heterogeneity could be partially solved via using two techniques - scaling-up and deep-sequencing. Different approaches are required to compare metagenomic samples collected from different environments and to reveal the composition and organization of microorganism genomes in the environment. One of the approaches is to enrich the culture of a specific organisms based on their specific environmental functions. This type of targeted-metagenomic approach permits for the development of composite microbial genomes. Additionally, to the uncultivability of many organisms, other problems occur when faster growing microorganisms outgrow slower ones, subsequently introducing additional biases.

Hence, instead of culturing, a complex sample can be enriched for a target cell population using fluorescence-activated cell sorting (FACS). Using this approach helps sort cells on the basis of their shape, size, and density. Alternatively, specific organisms of interest can be enriched by handling a small

number of cells that possess specific function in the environment. With the development of whole genome amplification (WGA) techniques (e.g., multiple displacement amplification (MDA) or rolling circle amplification (RCA), the gDNA of a restricted number of cells can be amplified and sequenced. However, WGA techniques create potential deviations because the abundance of specific sequences vary and therefore are not equally amplified. Hence, gDNA amplification of defined populations with a restricted number of cells compromises the quantitative analysis of metagenomics.

Latest advancements in single-genome amplification techniques enable SCG using physical cell separation (usually by FACS on 96-well plates), cell lysis, and WGA. An amplified genome from a single cell can then be sequenced. While metagenomics is accomplished by collecting the microorganism community and extracting and sequencing total DNA, SCG separates and studies individual cells from the microorganism community. Obtained sequencing data provides quantitative information on genomic variability in microorganism populations. Gene insertions, deletions, duplications, and genome rearrangements can be studied on a single-cell level. In this way the complex metabolic pathways can be analyzed for an individual cell.

SINGLE-CELL ISOLATION AND PROCESSING

Studying the genome and its regulation at the population level is done by analyzing millions of cells at once. However, such investigations do not allow for a view of the heterogeneity in biological systems, nor do they characterize the current state of the population. That is why the SCG is gaining more and more popularity for studying both cultivable and especially uncultivable microorganisms. The process to obtain, multiply, and sequence genetic material from either live or dead cells is relatively simple. First the cell needs to be isolated and lysed and then the released genetic material is amplified and the genomic library can be constructed.

One of the approaches for single cell isolation is the micromanipulation. Cells of interest are identified, isolated, and examined microscopically. Suitable cells are isolated using a micropipette, laser source, optical tweezers, or by real-time microfluidic flow. Using microfluidics, a desired phenotype correlating to a specific genome can be selected. In addition, cells can be easily observed before capture.

Other approach is the random encapsulation. Cells are randomly selected by serial dilution. The diluted sample can then be transferred into microwells, or the cells can be encapsulated by microdroplets. Using this technique, individual cells can be separated, processed, and subjected to genomic analysis.

Flow cytometry is the most popular procedure of random encapsulation. FACS allows the isolation of cells possessing specific criteria, such as the size, shape, color, or even presence of specific nucleic acids or activities, using fluorescent dyes (Fig. 4). One of the main advantages of FACS is its high throughput, high sorting speed, and ability to distinguish live cells. Moreover, FACS can detect the

presence of cells in a droplet, reject empty droplets, and transferred the cells with desired properties into multiwell / microtiter plates. Then cells can be lysed, and genomes can be amplified in single wells.

Other possibility for single cell isolation is the microdroplet technology (Fig. 4). Microdroplets are the emulsion formed when cells in the water phase are vigorously mixed with oil. Oil droplets encapsulate cells in the water phase. Originally, PCR performed on an emulsion was thought to be much more specific than analysis of cells in the water phase alone. It was suggested that the yields of reaction products are greater and fewer chimeric byproducts are formed. Moreover, the droplets formed enclosed a limited number of template molecules (ideally only one). Latest developments in microfluidics permit to control the droplet size (nanoliter or even the picoliter), providing aqueous uniform (monodisperse) droplets in oil. One of the main advantages of the two-phase system of microdroplets over FACS is the possibility to handle single cells as separated units. Each droplet act as an independent well where cells can be grown and later screened for the products expressed. Released gDNA can be effectively amplified in microdroplets and used for sequencing-library preparation. Bacterial, yeast, plant, insect, and mammalian cells can be easily studied using microdroplet technology on agarose or other microgels. Even multicellular organisms can be encapsulated and grown in a droplet. Hypothetically, it is possible to measure any secreted molecule for which a fluorescently labeled ligand exists. Furthermore, the activities of various cell proteins can be monitored.

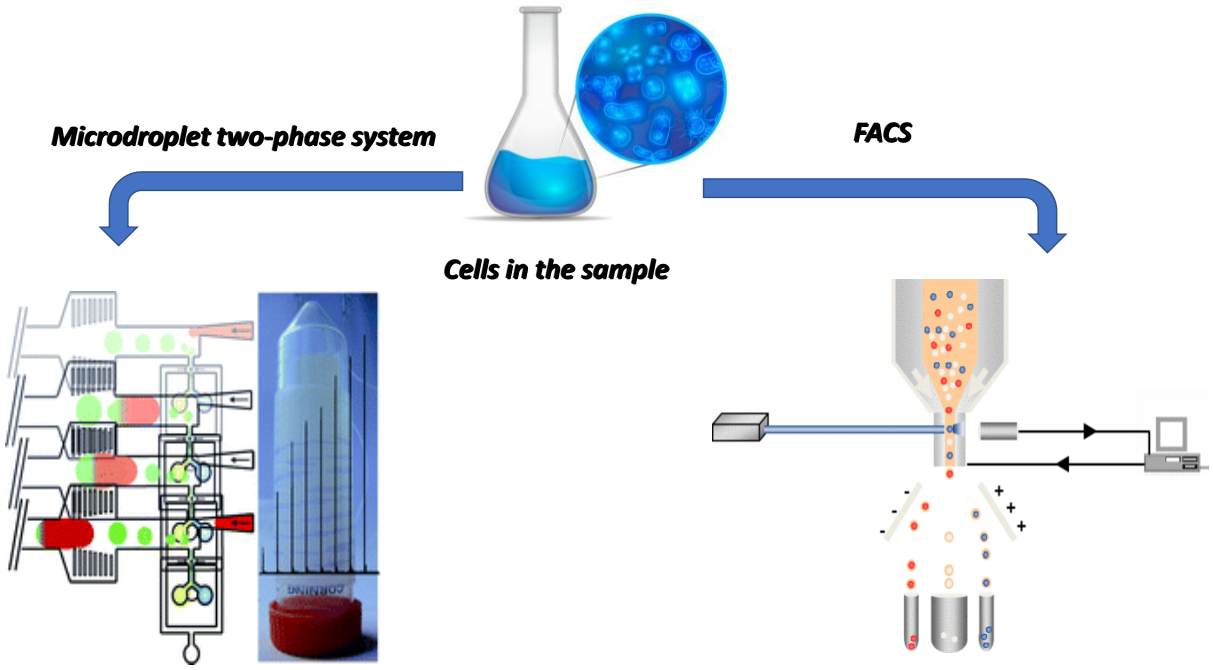

Fig. 4. Single cell isolation techniques

APPROACHES FOR STUDYING VARIOUS CELLS

Living cells can generally be classified into bacteria, archaea, and eukaryotes. Although viruses are not considered "living," they constitute an important part of life science studies. Various single-cell analysis exists that are suitable for different organisms.

Viruses are present in practically every environment. They are the most abundant and diverse biological entities. However, to isolate single viruse and sequence its genome, a suitable cultivable virus-host system must first be established. For example, eukaryotic algae are a host to an incredible diversity of viruses. However, using both FACS and single-droplet systems, viruses can theoretically be isolated without dependence on the viral-host system.

Archaea and bacteria have a similar cell structure. Only differences in cell composition and organization set these two domains apart. Bacterial cells are ordinarily 0.5 - 5.0 μm in diameter, while archaea can be larger and reach 15 μm in diameter. Basically, all methods described above can be used to separate and process single archaeal and bacterial cells. However, during FACS high currents may slow the growth of these organisms.

Eukaryotes differentiate from other life domains by their membrane bound nucleus and membrane-bound organelles. They may be multicellular or single-celled organisms. Cell size typically range from about 1 - 100 μm or they are ten-times larger than bacteria. Dilution, FACS, and droplet sorting are all appropriate methods for isolation of eukaryotic cells.

Multicellular organisms can theoretically also be isolated and processed by various methods. FACS for example can easily sort multicellular organisms. If incubated in droplets for a sufficient amount of time, single-celled organisms may synthesize a protein or ligand and have the opportunity to multiply, potentially creating a multicellular structure.

Molecular studies of our environment and ecology gain more and more knowledge for living communities. With the development of molecular tools, especially sequencing machines, it became possible to sequence whole communities from selected environments. Recent advancements in single-cell technologies facilitate the isolation and amplification of genomic material from single cells and allow structuring of numerous sequence-based libraries. Furthermore, data obtained from SCG supplements data obtained by metagenomics.

OMICS APPROACHES IN INDUSTRIAL BIOTECHNOLOGY AND BIOPROCESS ENGINEERING

Biotechnology utilizes biological processes, organisms, or systems to generate products and technologies that are improving human lives. The use of biological systems to produce bioproducts of commercial importance is a key component of the biotechnology industry. This biotechnological approach has found application in different areas: energy, material, pharmaceutical, food, agriculture, and cosmetic industries. The bioproducts made from biomanufacturing processes are usually metabolites and proteins, which are obtained from cells, tissues, and organs. The biological systems synthetizing these bioproducts can be natural or modified by genetic engineering, metabolic engineering, synthetic biology, and protein engineering. “Omics” technologies have great importance for biotechnology and metabolic engineering helping in characterizing and understanding metabolic networks. The significant amount of knowledge acquired from omics-driven experiments can be applied in the development of biotechnological tools and the advancement of metabolic engineering. This enables the manipulation of complex biological systems toward creating robust industrial biomanufacturing strategies.

1. Omics-Guided Biotechnology

“Omics” tools are used more and more in the development of biotechnological processes and the production of many vital products. Application of “omics” technologies in characterizing and understanding biological systems has permitted to select and predict phenotypes, which facilitates the optimization of biotechnological processes toward enhanced production (in quality and quantity) of commercially relevant products (Figure 5).

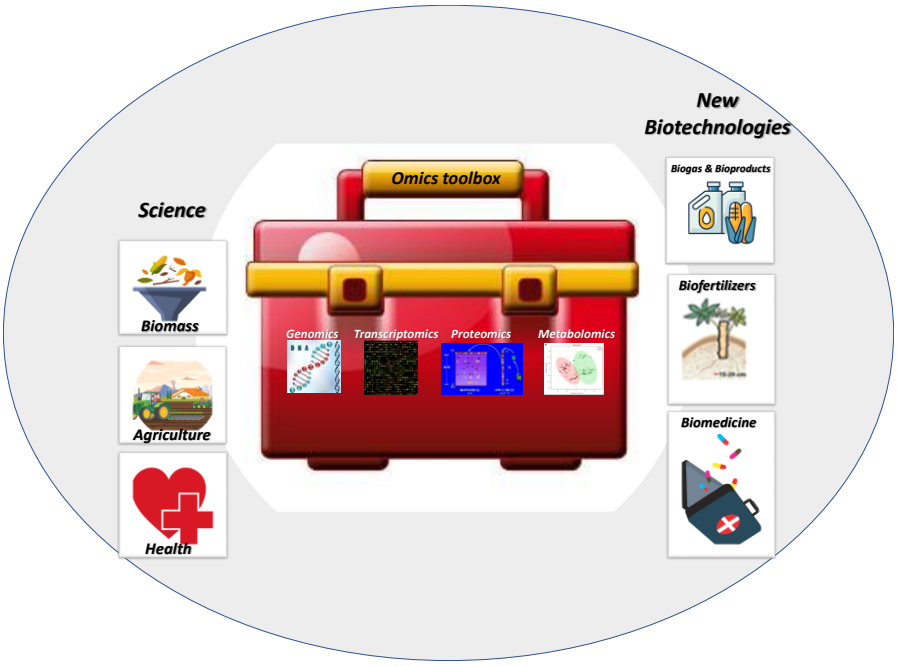

Fig. 5. Omics tools in Biotechnology

PRODUCTION OF BIOFUELS AND BIOPRODUCTS

Microbial production of natural compounds represents an attractive and more sustainable alternative to traditional petrochemicals. Its implementation has led to a growing catalog of natural products and high-value chemicals. For instant, the use of lignocellulosic biomass represents an economical approach to generate biofuels and bioproducts. However, to achieve stable conversion of low-cost raw material into value-added products at industrial levels needs systematic engineering plans. The Design-Build-Test-Learn (DBTL) cycle is becoming an increasingly implemented approach for metabolic engineering experiments. It represents a systematic and effective tool to force development efforts in biofuels and bio-based products. The DBTL cycle uses *in silico* approach to Design and Build genetic constructs in microbial hosts. Next, the information obtained from "omics" technologies, during the Test phase of the cycle, is transferred on to Learning processes (Figure 6). What is Learned is then fed back to new design cycles to achieve further strain development and optimization. Thus, a rapid optimization of microbial strains producers of any chemical compound of interest is facilitating. The weakest point in the DBTL cycle workflow is the Learning process since mathematical models are only as good as their assumptions. Therefore, both high quality and large "omics" data sets are required to improve training models, ensuring increased accuracy and reliability of the Learning process.

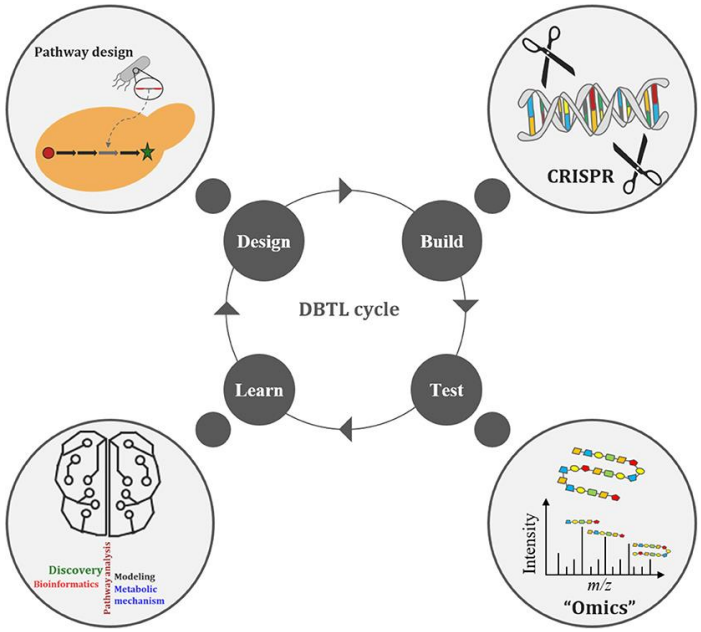

Fig. 6. DBTL cycle for optimization of biofuels and bio-based products

Often in the DBTL cycle genomic sequence information is the traditional approach utilized in the initial stages of a study. For example, barcode sequencing (Bar-seq) (a method using a short section of DNA from a specific gene or genes) can be used to study population dynamics of *Saccharomyces cerevisiae* deletion libraries during bioreactor cultivation, allowing the identification of factors that impact the diversity of a mutant pool. Single cell sequencing guided approach can be used to identify key mutated promotors for adjusting expression levels, thereby facilitating the dynamic regulation of microbial growth. Proteomics-guided approaches have been employed to engineer polyketide biosynthesis platforms for *in vitro* production of adipic acid in yeast. Additionally, metabolomics permits the assessment of pathway flux, carbon source utilization, and cofactor imbalance, which all contribute to the identification of pathway bottlenecks. Such metabolomics guided approach is already used for the characterization of the cannabinoid production in engineered *S. cerevisiae*. Cannabinoid analogs have been identified produced by several promiscuous pathway genes. Furthermore, application of metabolomics helped the design and optimization of a novel isopentenyl diphosphate-bypass mevalonate pathway in *E. coli* for C5 alcohol production. Recently, some authors have utilized the DBTL cycle to engineer *Rhodosporidium toruloides*, an oleaginous yeast species growing on lignocellulosic materials and producing the diterpene ent-kaurene, a potential therapeutic, exhibiting antimicrobial, anti-inflammatory, cardiovascular, diuretic, anti-HIV, and cytotoxic effects.

AGRICULTURAL AND FOOD BIOTECHNOLOGY

Recent innovations in agricultural biotechnology have led to new plant varieties, engineered by recombinant DNA technology and better responding to market demands and environmental challenges. In fact, "omics" tools in agricultural biotechnology have been used to enhance desirable phenotypic characteristics (e.g., color, taste, drought tolerance, pesticide resistance, etc.). "Omics" plays role not only in improving crop quality, consistency, and productivity, but also in the development of food crops with enhanced nutritional composition. Moreover, omics-driven systems biology helps in understanding the interactions between the "OMEs" and provide links between genes and specific characteristics.

In arable land soil is more susceptible to loss of structure, organic matter, minerals, and erosion. Thus, efforts are being made, via agricultural biotechnology, to provide a constant supply of nutrients essential to the growth of crops. An integral part of this approach is the use of biofertilizers. These are preparations containing specialized microbial inoculants that can fix, mobilize, solubilize, or decompose nutrient sources. Biofertilizers are generally applied through seed or soil and enhance nutrient uptake by plants. Widespread application of this approach, however, has been hampered by fluctuating responses of microbial inoculants across fields and crops. As a result, there is a pressing need to better understand the mechanisms underlying the interdependencies between soil microbial communities and host plant productivity. Recent genomics and exo-metabolomics studies showed that specific rhizosphere bacteria have a natural preference for certain aromatic organic acids exuded by plants. This fact suggested that plant exudation characteristics and microbial substrate uptake traits interact and form specific patterns of microbial community assembly. Furthermore, the application of genomics and transcriptomics to the study of phosphate uptake by various microalgae revealed a range of Pi transporters having specific expression patterns in relation to the availability of P. Currently, "omics" approaches are being used to study complex rhizospheric intercommunications, which is crucial to the development of new biofertilizers, thus promoting stable plant growth, better crop productivity and yield.

In the related field of food biotechnology, application of transcriptomics and metabolomics demonstrated that *Bacillus pumilus* LZP02 promote the growth of rice roots by enhancing carbohydrate metabolism and phenylpropanoid biosynthesis. Further, the application of "omics" in starch bioengineering provides better understanding on the most important enzymes for its biosynthesis. This facilitates the prediction on how starch-related phenotypes can be modified and ensures further progress in the research field of rice starch biotechnology. "Omics" also help solving issues related to food quality and traceability, to protect the origin of food, and discover biomarkers of potential food safety problems. The advance investigation of wine microbiome has a great influence for wine industry helping in better understanding of factors transforming grapes to wine, including flavor and aroma. "Omics" characterization of the complex relationships between these microorganisms, the substrate and environment, is essential to shaping wine production. Finally, combining "omics" technologies with genome editing of food microorganisms can be exploited to generate improved probiotic strains, develop innovative bio-therapeutics and alter microbial community structure in food matrices.

"OMICS" TECHNOLOGIES AND COVID-19

The coronavirus disease 2019 (COVID-19) is characterized by the Severe acute respiratory syndrome coronavirus 2 [i.e., SARS-CoV-2, which binds to the ACE2 receptor in the lung and other organs]. Due to its worldwide spread a global pandemic has been announced and much of the world's economy has been slowed. Up to April 2021 there are more than 138 million confirmed cases and 2.5 million confirmed deaths globally (https://www.worldometers.info/coronavirus/). Thus, there is an urgent need for an effective countermeasure to mitigate the spread of the pandemic.

Therefore, efforts are ongoing to fast-track the development and production of safe and effective vaccines against SARS-CoV-2. Preceding knowledge of SARS and Middle East respiratory syndrome (MERS) has facilitated targeting the spike protein as the viral antigen (via the ACE2 receptor). Furthermore, the genome sequencing of the SARS-CoV-2 in January 2020 made it possible to accelerate the development of next generation mRNA and vaccine platforms that encode for the antigen. Once injected into a host, the mRNA, encapsulated into lipid nano-particles, remains in the cytoplasm. When DNA is used, encased in an attenuated adenovirus vector, it enters the nucleus. The host cell translates these genetic materials into the spike protein. It settles on the surface of the cell and provokes an adaptive immune response mediated by T cells (e.g., CD4+ and CD8+) and B cells (i.e., antibodies). These vaccines were reported to be effective against SARS-CoV-2 in recent clinical trials, which underlines the significance of genomics to this new era of vaccine development.

Since the beginning of the outbreak of SARS-CoV-2 a protein interaction was created map and using a proteomics-based approach targets for drug repurposing was revealed. By applying proteomics analysis 26 of the 29 SARS-CoV-2 proteins were cloned, affinity tagged and expressed in human cells and the associated proteins were identified. A total of 66 human proteins or host factors were discovered as potential drug targets of 69 compounds. Two sets of these pharmacological agents showed antiviral activity. Moreover, computational immunoproteomics studies were performed supporting the lab-based investigations and allowing the evaluation of diagnostic products specificity, the forecast of potential vaccines adverse effects and the reduction of animal models utilization.

Furthermore, recently a method was developed based on metabolomics that is able to distinguish COVID-19 patients from healthy controls via the analysis of 10 plasma metabolites. Additionally, the data from lipidomics study suggests that monosialodihexosyl ganglioside enriched exosomes could be involved in pathological processes. Proteomics and metabolomics investigations in COVID-19 patient sera revealed that SARS-CoV-2 infection causes metabolic dysregulation of macrophage and lipid metabolism, platelet degranulation, complement system pathways, and massive metabolic suppression. The analysis of plasma metabolomic signatures showed similar profile to those described for sepsis syndrome. Furthermore, transcriptomics results showed upregulation of genes related to oxidative phosphorylation both in peripheral mononuclear leukocytes and bronchoalveolar lavage fluid. All this information suggests a critical role of mitochondrial activity during SARS-CoV-2 infection. Understanding the clinical presentation of COVID-19 as well as metabolomic, proteomic, and genetic profiles could help in discovering specific diagnostic, prognostic, and predictive biomarkers, ensuring

the development of more successful medical therapy. Moreover, differentiating metabolic biomarkers of severe vs. mild disease states in the lung during respiratory infections could lead to the discovery of novel therapeutics that modulate symptom and disease severity.

NUTRI-OMICS RESEARCH

More than a decade has passed since the idea of nutrigenomics was first introduced. Nutrigenomics, also referred to as nutritional genomics, nutritional omics, or nutri-omics, can be defined as an area of food and nutrition research making use of profound analyses of molecules or other physical phenomena.

Considering the complexity of the human body and its various interactions with food, it is conceivable that holistic analyses of food- body interactions are an important prerequisite for better understanding the effect of dietary components. The main assumption is that the combination of different omics platforms will provide a deeper insight into the influence of food components and the mechanism of their actions.

1. Genomics and nutrition

Human genes have different variants in the population. Having in mind that response to nutrition is a multigenic process, it is not surprising that non-related individuals may respond differently. They key hypothesis proposed in recent years stated that genetic variation underlies variation in nutrition-related disorders and risk for disease. In fact, nutrigenetics tries to explain how and to what extent nutrition-related traits and disorders are influenced by genetic variation. In other words, the key questions of nutrigenetics are:

- Which genes are involved in determining a defined characteristic?
- What is the functional identity of the variation by which people differ for this characteristic?
- How can this knowledge be used for the benefit of the population?

Usually, researchers define ‘candidate genes’ for a characteristic or disorder and search in those genes for variation. This is described as the ‘hypothesis-driven’ approach because the selection of the genes is based on their presumed function. For example, the genes for lipoproteins are used as candidates for obesity and the genes for insulin signaling - as candidates for diabetes. Other attractive candidates in humans are the orthologous genes underlying animal models with a Mendelian segregation of a characteristic. A well-known example is the db/db mouse. It is a model for diabetes in which the gene

for the leptin receptor was identified. The human homologue was then tested in persons with an impaired glucose tolerance. One variant (allele) of the gene was identified and was proven that it was associated with a significantly higher frequency of diabetes occurrence in comparison to the general population.

Other methods applied is related with studying the preferred segregation of an allele from parents to an affected child (transmission disequilibrium test) or to more affected siblings (sib-pair analysis). If a gene is found to be linked to the characteristic, it remains to be proven whether the associated or linked allele itself is the risk factor, or whether it could be used only as a marker for a nearby causative variation. If the genetic variation is linked with amino acid variation in the protein, a functional test could be developed and the allelic proteins can be compared for their enzymatic activity, DNA-binding affinity, etc.

Another approach to search for risk-conferring genes for a certain characteristic or disorder is the total genome scan. Hundreds of polymorphisms are detected with a random distribution across the genome but with a known chromosomal location. For every polymorphic site it is then identified whether an association exists between the allele and the characteristic. Using this approach risk loci can be identified for many human nutrition-related disorders like obesity and diabetes.

After the second world war in many Western countries, it was observed that the incidences of spina bifida gradually decreased. It was suggested that the improvement of the diet had brought forth this effect. A large epidemiologic study was started. The results obtained showed that the enrichment of the diet with vitamins was able to prevent the occurrence and recurrence of spina bifida in humans with more than 50%. Folic acid was discovered to be the active substance and now in many countries pre-conceptional folic acid supplementation is recommended as a general preventive measure. Simultaneously, it has become clear that concerning the risk for a child with spina bifida, the women in the population can be roughly divided into three groups: [i] not at risk even at a low folate intake, [b] at risk at low dietary folate but can be helped by increased folate intake, and [c] at risk despite extra folate intake. Genetic predisposition was presumed, and candidate genes taken from the folate metabolism were examined for genetic variation. This led to the detection of the 667C->T and 1298A->C alleles in the gene for methylenetetrahydrofolate reductase (MTHFR) as risk factors. However, these risk alleles are discovered to some extent in all three groups of women illustrating that genetic testing in the context of a personalized diet would be inadequate. Moreover, it became clear that not all the relative risk could be explained by these alleles. Therefore, the search for genetic variation in other genes is continued and new candidate genes may be found.

Currently, several hundreds of genes related to specific nutrition-related characteristics and disorders have already been identification via using genetic experiments in various populations and the investigation of animal and *in vitro* model systems. Finally, the nutrigenomics and nutrigenetics research should lead to the comprehensive understanding of the etiology of those characteristics and disorders with respect to the interacting genes, the dietary components, and the relative risk conveyed by both genetic variation and diet. This combined knowledge will permit the genetic profiling of every individual and thereby assess her/his risk for developing nutrition-related disorders. Based on this personal risk

profile of genetic factors, a 'personalized diet' could then be proposed by which the onset of a disorder could be prevented or at least delayed. Already some companies offer genetic testing for alleles that were found to be associated with certain characteristics. For example, the ApoE allele is tested as a risk factor for cardiovascular disorders and an allele of the alcohol dehydrogenase gene is used to predict potential alcohol sensitivity and (ab)use. However, it is important to note that the presence of certain allele may be said to increase the risk for a disorder by 100%, while in absolute terms the risk would usually still be very small like an increase from 0.001 to 0.002. Additionally, testing only one genetic factor might give an incomplete or even a wrong view of the situation. Nowadays, we have no idea of all the genetic factors and how they interact. Till more information is gathered, the advice based on simple testing probably has to remain simple, sounding like "drink less alcohol" or "eat less saturated fat".

2. Trascriptomics and nutrition

Transcriptomics is the most widely employed in food research because of the many advantages of the DNA microarray technology, including the comprehensiveness of the expression data, established protocols, and high reliability and reproducibility of the data. It has been also reported the alternations of global gene expression in response to different dietary changes such as nutritional deficiency, fasting, overeating, and ingestion of specific food factors. As one example of the performed attempts to acquire reference data a transcriptome analysis of the liver of rats treated with mild caloric restriction has been performed. When using animal experiments, it was shown that a modest reduction of food intake or altered intake pattern could be sufficient for occurrence of significant metabolic changes. Therefore, it is important to distinguish between the direct effects of the food component and the secondary effects caused by the change in eating behavior. When rats were fed a diet with 5 to 30% less food than that consumed by an ad libitum-fed group for 1 week or 1 month it has been observed restriction – dependent changes in the expression level of cyp4a14. In fact, the cyp4a14 gene was induced by even a low level of caloric restriction. This data suggests that the gene can be used as a biomarker for the beneficial effects of food on energy metabolism.

3. Proteomics and nutrition

Many food and nutrition research studies have been carried out by using proteomics approaches. For instant, the effect of mild caloric restriction described above was also examined by proteomics. Nine significantly up-regulated proteins and nine down-regulated proteins have been detected after the proteome comparison of the liver of rats treated with 30% food restriction with control ones. Ten percent restriction caused the up-regulation of 9 proteins and the down-regulation of 2 proteins. An interesting discovery was the up-regulation of prohibitin, whose involvement in the regulation of longevity was recently revealed. This result suggests that prohibitin can be applied as an effective biomarker of the beneficial effects of food factors. Although the current stage of proteomics research is much less

extensive compared with that of transcriptomics, the study described here together with other nutritional proteomics research findings indicates that proteomics is a highly promising tool for the discovery of biomarkers.

4. Matbolomics and nutrition

About ten thousand different types of major metabolites exist in animal bodies, while the number of proteins is thought to exceed 100,000. This feature of metabolites is expected to result in more comprehensive features of metabolomics analysis than proteomics ones. However, the analysis of metabolites is in fact fraught with difficulties and usually requires the use of sophisticated techniques and high skill level personnel. Another difficulty originates from the width of the metabolite's abundance. Regardless of these obstacles, metabolomics is a powerful tool in food and nutrition science.

Many investigations showed that gut and serum metabolism is influenced by the change in gut microbiome composition. The gut microbiome and its alternation in relation to the diet are essential when evaluating dietary intervention trials with metabolomic end effects. The comparison between germ-free mice colonized by human baby flora and conventional mice showed the complexity of diet modifiable microbiome/metabolome covariation. In this study, the effect of the intestinal microbiome on plasma metabolites was studied. Data revealed that more than 10% of the plasma metabolome is directly dependent upon the microbiome. Some examples for microbial dependent compounds in plasma include cinammic acid, glycine conjugated compounds, hippuric acid, and other plasma metabolites. The gut microbiome also directly affects the host's ability to metabolize lipids, carbohydrates, and proteins, and can carry out several phase II detoxification reactions. There is also evidence that gut microorganisms utilize nonnutritive phytochemicals. For example, several investigations demonstrated that levels of three microbiome-dependent diet-derived metabolites, choline, trimethylamine N-oxide, and betaine, could predict risk for cardiovascular disease in mice.

5. Nutrition and other OMICs

Many other omics platforms are also a focus of nutriomics research. It was shown that epigenetic alteration could be caused by the nutrition diet during fetal development and could affect the predisposition to lifestyle-related diseases in later life. For example, children of mothers who suffered over- or undernutrition during pregnancy have increased risk of developing obesity, diabetes, hypertension, cardiovascular diseases, etc. Many studies have proven the involvement of epigenetic modifications in such acquired susceptibility. Another promising target of omics approaches in food science is associated with the RNA transcripts that do not encode proteins. MicroRNA (miRNA) is a subtype of these non-coding RNAs. Precursors of miRNAs (pri-miRNA) are long products which are cleaved to yield mature miRNAs of 22 nucleotides in length. Mature miRNAs regulate the levels of

mRNA degradation, mRNA translation, and even gene transcription. In the case of human cells, about 1,000 miRNAs are identified and are supposed to regulate the expression of more than half of the protein-coding genes. Considering the accumulating data supporting the key role of miRNA in the development of diseases and the maintenance of health, the information on the status of miRNAs is no doubt vital for the understanding of the interaction between food components and the body. Global miRNA analysis now can be easily performed by the use of commercial miRNA arrays.

With the introduction of new technologies and acquired knowledge, the number of fields in omics and their applications in various areas are rapidly increasing in the postgenomics era. Such emerging fields - including pharmacogenomics, toxicogenomics, regulomics, spliceomics, metagenomics, and environomics - present promising solutions to combat global challenges in biomedicine, agriculture, and the environment.

References

Amer B, Baidoo EEK. 2021. Omics-Driven Biotechnology for Industrial Applications. Frontiers Bioeng. Biotech., 9, Art. 613307: 1-19.

Azizipour N, Avazpour R, Rosenzweig DH, Sawan M, Ajji A. 2020. Evolution of Biochip Technology: A Review from Lab-on-a-Chip to Organ-on-a-Chip. Micromachines, 11: 599.

Canzler S, · Schor J, · Busch W, et al. · 2020. Prospects and challenges of multi-omics data integration in toxicology. Archives of Toxicology, 94:371–388.

Horgan RP, Kenny LC. 2011. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. The Obstetrician & Gynaecologist, 13:189–195.

Joseph P. 2017. Transcriptomics in toxicology. Food Chem Toxicol., 109(1): 650–662.

Karahalil B. 2016. Overview of Systems Biology and Omics Technologies. Curr. Med. Chem., 23, 1-10.

Karczewski KJ, Snyder MP. 2018. Integrative omics for health and disease. Nat Rev Genet., 19(5): 299–310.

Kato H, Takahashi S, Saito K. 2011. Omics and Integrated Omics for the Promotion of Food and Nutrition Science. J. Trad. Complement. Medicine, 1 (1): 25-30.

Kim DH, Kim YS, Son NI, Kang CK, Kim AR. 2017. Recent omics technologies and their emerging applications for personalized medicine. IET Syst. Biol.,11 (3): 87-98.

Kodzius R, Gojobori T. 2016. Single-cell technologies in environmental omics. Gene 576: 701–707.

Mariman ECM. 2006. Nutrigenomics and nutrigenetics: the 'omics' revolution in nutritional science. Biotechnol Appl Biochem., 44(3):119-28.

Mousumi D, Prasad GBKS, Bisen PS. 2010. Omics technology. In: Molecular Diagnostics: Promises and Possibilities, Dordrech Heidelberg London, Springer, pp 11-31.

Ryan EP, Heuberger AL, Broeckling CD, Borresen EC, Tillotson C, Prenni JE. 2013. Advances in Nutritional Metabolomics. Current Metabolomics, 1: 109-120.

Sauer UG, Deferme L, Gribaldo L, et al. 2017. The challenge of the application of 'omics technologies in chemicals risk assessment: Background and outlook. Regulatory Toxicology and Pharmacology, 91: S14 - S26

Simões T, Novais SC, Natal-da-Luz T, et al. 2018. An integrative omics approach to unravel toxicity mechanisms of environmental chemicals: effects of a formulated herbicide. Scientific Reports 8:11376

Titz B, Elamin A, Martin F, et al. 2014. Proteomics for systems toxicology. Computational and Structural Biotechnology Journal, 11: 73–90.

Vlaanderen J, Moore LE, Smith MT, et al. 2010. Application of Omics Technologies in Occupational and Environmental Health Research; Current Status and Projections. Occup Environ Med., 67(2): 136–143.

Yılmaz B, Yılmaz F. 2018. Lab-on-a-Chip Technology and Its Applications. In: Omics Technologies and Bio-engineering: Towards Improving Quality of Life, pp. 145, Elsevier Inc.

Project website: www.digit-biotech.eu

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.