Omics Technology

Chapter · October 2010

DOI: 10.1007/978-90-481-3261-4

3 authors:

Mousumi Debnath
Manipal University Jaipur
96 PUBLICATIONS **1,725** CITATIONS
SEE PROFILE

GBKS Prasad
Jiwaji University
281 PUBLICATIONS **3,460** CITATIONS
SEE PROFILE

Prakash S Bisen
Jiwaji University
579 PUBLICATIONS **7,591** CITATIONS
SEE PROFILE

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

Chapter 2

OMICS TECHNOLOGY

Abstract

Omics has become the new mantra in molecular research. 'Omics' technologies include genomics, transcriptomics, proteomics and metabolomics. Genomics had revealed the static sequences of genes and proteins. Focus has shifted to their dynamic functions and interactions.Transcriptomics, proteomics and metabolomics reveal the biological function of the gene product. The "-omic-" technologies are high-throughput technologies and they increase substantially the number of proteins/genes that can be detected simultaneously to relate complex mixtures to complex effects in the form of gene/protein expression profiles. The primary aim of omic technologies is the nontargeted identification of all gene products (transcripts, proteins, and metabolites) present in a specific biological sample. The powerful 'omics' technologies have opened new avenues towards biomarker discovery, identification of signaling molecules associated with cell growth, cell death, cellular metabolism and early detection of cancer. Omics will not only have an impact on our understanding of biological processes, but the prospect of more accurately diagnosing and treating disease will soon become a reality.

Key words

Omics, genome, Proteomics, transcriptomics, metabolomics ,systems biology, high throughput technology,genomics,human genome project,gene sequence , gene transcript, herceptin ,gleevec, molecular signature, expression profile , expression signature ,computational tools ,oncology, proteome,biomarkers, yeast two hybrid analysis ,clinical diagnostics, protein –protein interaction, mass spectroscopy, protein microarray,metabolome ,metablic profiling ,toxicogenomics,nutrigenomics,

2.1 PROLOGUE

Over the last decade, we have witnessed a fundamental change in how biomedical research is carried out and we can now assess the impact of the Human Genome Project on drug discovery and development (Bilello, 2005). The human genome sequence has revealed that sequence variations are very common. Genetic approaches have already linked a large number of genetic variants (polymorphisms) with human diseases and adverse reactions from exposure to drugs or toxicants, suggesting an important role in sensitivity to drugs and environmental agents, disease susceptibilities, and therapeutic responses. As these opportunities are transformed into reality, regulatory toxicological practice is likely to be shaped in the

future by the combination of conventional pathology, toxicology, molecular genetics, biochemistry, cell biology, and computational bio-informatics resulting in the broad application of molecular approaches to monitoring functional disturbances.

2.2 CONCEPT OF OMICS

New technologies that permit simultaneous monitoring of many hundreds, or thousands, of macro- and small molecules promise to allow functional monitoring of multiple (or perhaps all) key cellular pathways simultaneously.The new "global" methods of measuring families of cellular molecules, such as RNA, proteins, and intermediary metabolites have been termed "-omic" technologies, based on their ability to characterize all, or most, members of a family of molecules in a single analysis (Fig 2.1). With these new tools, we can now obtain complete assessments of the functional activity of biochemical pathways, and of the structural genetic (sequence) differences among individuals and species, that were previously unattainable.The terms 'Ome' and 'Omics' are derivations of the suffix *-ome*, which has been appended to a variety of previously existing biological terms to create names for fields of endeavor like genome, proteome, transcriptome and metabolome that are either speculative or have some tangible meaning in particular contexts.

Omics has become the new mantra of molecular biology. 'Omic' technologies include genomics, transcriptomics (gene expression profiling), proteomics and metabolomics. The recent availability of masses of omic data is responsible for the major growth spurt of systems biology. Pharmaceutical companies and others need to make sense out of all this omic information in order to take the next step in overcoming their innovation deficits. Systems biology provides the methods, computational capabilities, and inter-disciplinary expertise to facilitate this jump. The technology platform of genomics, transcriptomics, proteomics and metabolomics are high-throughput technologies. They increase substantially the number of proteins/genes that can be detected simultaneously and have the potential to relate complex mixtures to complex effects in the form of gene/protein expression profiles. Commercial applications are emerging at an accelerating pace as pharmaceutical and biotechnology research organizations start to combine various forms of omic data into more comprehensive computer models, and bioinformatics companies increasingly turn their attention toward offering systems biology solutions to drug developers and diagnostics companies.

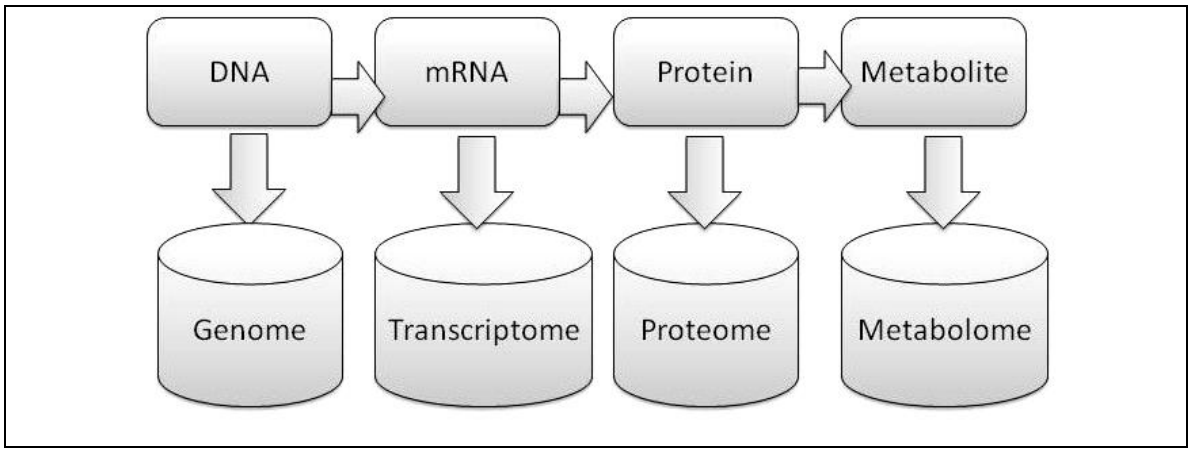

Fig 2.1: The Central Dogma and the interacting "ome" includes the study of genome, proteome, transcriptome and metabolome

2.3 GENOME AND GENOMICS

In biology the genome of an organism is its whole hereditary information and is encoded in the DNA (or, for some viruses, RNA). This includes both the genes and the non-coding sequences of the DNA. More precisely, the genome of an organism is a complete DNA sequence of one set of chromosomes; for example, one of the two sets that a diploid individual carries in every somatic cell. The term genome can be applied specifically to mean the complete set of *nuclear DNA* (i.e., the "nuclear genome") but can also be applied to organelles that contain their own DNA, as with the mitochondrial genome or the chloroplast genome. The genome contains the coded instructions necessary for the organism to build and maintain itself. Genomics is the study of an organism's genome, or genetic material. Genomics emerged as a new field of molecular biology where novel technologies were exploited in order to understand the complex, biological function of the genome. It revealed the static sequences of genes.

The molecular biology revolution and the advent of genomic technologies are facilitating rapid advances in our understanding of the molecular details of cell and tissue function. These advances have the potential to transform toxicological and clinical practice, and are likely to lead to the supplementation or replacement of traditional biomarkers of cellular integrity, cell and tissue homeostasis, and morphological alterations that result from cell damage or death. The increasing amount of genomic and molecular information is the basis for understanding higher-order biological systems, such as the cell and the organism, and their interactions with the environment, as well as for medical, industrial and other practical applications (Kanehisa *et al.*, 2006). The knowledge of full genomes has created the possibility for the field of functional genomics, mainly concerned with patterns of gene expression during various conditions. Microarray and bioinformatics are the most important tools of genomics .

A major branch of genomics concerns with gene sequencing the genomes of various organisms. In 1972, Walter Fiers and his team at the Laboratory of Molecular Biology of the University of Ghent (Ghent, Belgium) were the first to determine the sequence of a gene: the gene for bacteriophage MS2 coat protein. In 1976, the team determined the complete nucleotide-sequence of bacteriophage MS2-RNA. The first DNA-based genome to be sequenced in its entirety was that of bacteriophage Φ-X174; (5,368 bp), sequenced by Frederick Sanger (Sanger,1977). The first free-living organism to be sequenced was that of *Haemophilus influenzae* (1.8 Mb) in 1995, and since then genomes are being sequenced at a rapid pace. A rough draft of the human genome was completed by the Human Genome Project in early 2001, creating much fanfare.

As of January 2005, the complete sequence was known of about 1,000 viruses, 220 bacterial species and roughly 20 eukaryote organisms, of which about half are fungi. Most of the bacteria whose genomes have been completely sequenced are problematic disease-causing agents, such as *Haemophilus influenzae*. Of the other sequenced species, most were chosen because they were well-studied model organisms or promised to become good models. Yeast (*Saccharomyces cerevisiae*) has long been an important model organism for the eukaryotic cell, while the fruit fly *Drosophila melanogaster* has been a very important tool (notably in early pre-molecular genetics). The worm *Caenorhabditis elegans* is an often used simple model for multicellular organisms. The zebrafish *Brachydanio rerio* is used for many developmental studies on the molecular level and the flower *Arabidopsis thaliana* is a model organism for flowering plants. The Japanese pufferfish (*Takifugu rubripes*) and the spotted green pufferfish (*Tetraodon nigroviridis*) are interesting because of their small and compact genomes, containing very little non-coding DNA compared to most species. The mammals dog (*Canis familiaris*), brown rat (*Rattus norvegicus*), mouse (*Mus musculus;* Lein *et al.*, 2007) and chimpanzee (*Pan troglodytes*) are all important model animals in medical research.

Genomics has the potential to revolutionize the practice of medicine, but despite significant scientific advances, very few genomics-based tests or treatments have reached consumers.The application of this genomics revealed that gene expression profiles induced by single drugs and the ones induced by the combination of the same drugs can be entirely different. The increasing amount of genomic and molecular information is the basis for understanding higher-order biological systems, such as the cell and the organism, and their interactions with the environment, as well as for medical, industrial and other practical applications. (Kanehisa *et al.*, 2006).

The combination of genomic technologies with a knowledge of gene sequence and sequence conservation has made available markers that facilitate the correlation of genetic variation with biological outcomes, and "-omic" technologies allow efficient biochemical characterization of functional pathways--providing new markers of the susceptibility of individuals to cancer development, and of tumor susceptibility to specific therapies. The powerful -omic technologies allow efficient monitoring of gene transcripts, proteins, and intermediary metabolites, making it possible to monitor a large number of key cellular pathways simultaneously. This has enabled the identification of key biomarkers and signaling

molecules associated with cell growth, cell death, and cellular metabolism. These new markers are facilitating monitoring of functional disturbance, molecular and cellular damage, and damage-response. Improved imaging technologies have made it feasible to image some of these molecular events noninvasively. These developments promise more efficient pharmaceutical product development, safer and more efficacious drugs, and provide clinical practitioners with new and better biomarkers for cancer screening, patient monitoring, and choice of therapy (MacGregor**,** 2007).

Over the past two decades, advances in genomic technology have allowed laboratories to generate vast amounts of biological data. These data include gene sequences, protein structures, information on gene expression and metabolic pathways. Automated instrumentation has enabled large volumes of data to be generated and automatically stored in computer databases, and this data has as many different formats as there are instruments. In addition to the new information gathered from genomic technologies, pharmaceutical and biotechnology companies have large amounts of 'legacy data' - data inherited from their own and other sources on chemical structures and properties of compounds, and clinical, phenotypic and toxicological information. Most of this is stored in older types of databases designed for the particular type of data, and a major computational challenge is to integrate the new genomic information with current database systems in order to facilitate decision-making. Molecular biology and genetics has excelled in the creation of molecular diagnostics to identify and characterize disease processes. In the case of many diseases it has markedly altered clinical practice. In acute care settings it has led to increasingly accurate diagnosis and has impacted clinical treatment and outcomes. Another example is in the area of HIV disease management where viral load and resistance genotyping has altered patient treatment decisions. One example where molecular diagnostics has impacted a multifactorial disease treatment is in the treatment of breast cancer, therapy with the monoclonal antibody Herceptin (Trastuzumab®) which is effective in a proportion of patients with elevated HER-2 (*erb* B2) gene expression. The U.S. Food and Drug Administration (FDA) approved both immunohistochemistry (IHC; DAKO HercepTest®) and fluorescence *in situ* hybridization (FISH; Vysis PathVysion®) for the selection of patients for Herceptin treatment in September of 1998. This is an example of targeting a therapy to a specific genetic profile (Pegram and Slamon 2000), however Slamon and his co-workers demonstrated that amplification of the HER-2/neu gene was a significant predictor of both overall survival and time to relapse in patients with breast cancer. Similarly it took almost twenty years from the molecular understanding of the role of the *bcr-abl* fusion protein to the development of the therapeutic Gleevec. Nonetheless gene based testing technologies clearly offer a great advantage in identifying patient populations which can benefit from therapy. More importantly, the clear demonstration that one can proceed from a genetics-based disease understanding to therapeutic efficacy is a critical learning. While the impact early in the discovery phase is most significant, the application of genomics can contribute throughout the drug development process from initial discovery all the way through life cycle management.With costs of bringing a new compound to market in excess of a billion dollars, there is a concern that theoft -promised genomics-driven advances in medicine and patient

care may not materialize (Venkatesh and Harlow, 2002). While we have not seen the immediate benefit that the genomic revolution had promised, it is clear that we are at the beginning of a journey.

2.4 TRANSCRIPTOME AND TRANSCRIPTOMICS

Transcriptomics refers to the comprehensive scanning of the nearly fifty thousand currently known genes that are transcribed into RNA molecules from the three-billion-letter human genome. Each cell utilizes (expresses) different genes at different times in its development and under different physiological conditions. In general, tissues express similar sets of genes that can be used to identify those tissues in the absence of any other information. For example, the brain expresses about thirty percent of all of the known genes; those specific transcripts are different from the transcribed genome in the heart. We can therefore define molecular signatures based on expression profiles, and these profiles can then be used to automatically separate normal cells or tissues into their correct category.

This method of expression profiling can also be done with disease states and implemented diagnostic tests. For example, neurons that have one of the beginning hallmark features of Alzheimer's disease, neurofibrillary tangles, show altered gene expression patterns when compared to normal neurons. This information can be used as molecular diagnostic criteria in the absence of histopathology for tangles because it is, in fact, a surrogate marker of that cellular condition. Transcriptomics builds upon the more classical study of gene function where Northern blots are used for transcriptional analysis of individual genes. However, instead of examining one gene at a time this new technology analyses in one single go the entire transcriptome, which is the full set of all messenger RNA molecules present at any given time in a defined population of cells. This reveals an instant picture of all genes being actively transcribed, and since the technique is a least partly quantitative it also tells about the expression level of each gene. For example, by comparing such transcriptional analyses from plants exposed to different environmental conditions it is possible to quickly identify all genes involved in the adaptive process. All genes and most proteins can be regarded as instruments for making up the biochemical composition and there by the physiological identity of an organism.

The same can be done for clinical states that may be less obvious or for which there does not exist a diagnostic marker. For example, in autism we are defining the expression correlates of the disorder by sifting through the entire transcriptome of hundreds of individuals with the disorder versus hundreds without the disorder. Each individual has undergone ten hours of clinical assessment and has either been diagnosed with autism or determined to be unaffected. Once the expression signatures that are specific to autism are established and validated (through blindly diagnosing several more hundred individuals), we can then use the expression profile as the diagnostic test in order to circumvent the ten hours of clinical

assessment (contingent on the accuracy of the test, of course). This approach will also facilitate testing in a pre-symptomatic scenario to encourage early interventions that may result in better outcomes.

Another example of the power of transcriptomics is its ability to sub-classify disorders that on the surface appear to be similar. For example, drawing on the autism example above, we can identify molecular sub-classes of autism that correlate to differences in the clinical picture. For example, many children (but not all) have gastrointestinal symptoms. Some children respond to dietary modifications, but there is a large spectrum of behavioral symptom severity. By expression profiling the entire fifty thousand transcripts in a cohort of children with autism and then performing computational analyses (which allows the profiles of molecularly similar children to congregate next to one another), we can create similarity dendograms that define molecular subclasses. All available clinical data can then be overlaid and permuted according to molecular subclass to see if correlations exist. If a correlation does exist, then that profile can be used as a diagnostic test for that clinical parameter. This strategy has been shown elegantly for a number of disorders, primarily in oncology, where one can define outcome and survival rates and drug response rates based on an expression profile.

The development of microarray technologies and other high-throughput strategies are becoming increasingly important in biology because they permit to monitor expression levels of thousands of genes in only one experiment. To interpret the biology of these genetic profiles, these data must be analysed in the context of the corresponding proteins coded. This means retrieving information about their biological, biochemical or molecular function. Thus, the development of computational tools to compare and analyse these expression profiles in a suitable way for biological interpretation is needed (Subramanian *et al.*, 2005).

Microarray technologies at the transcriptomic level can be high throughput and excellent for diagnostic purposes, but often their dynamic range, informational and experimental complexities limit their utility as diagnostic tools as compared with other assessments of mRNA abundance e.g. real time RT-PCR. One of the early applications of microarray was in mode of action (MOA) studies for isoniazid (INH) an antibiotic effective in the control of tuberculosis. Studies of *M. tuberculosis* exposed to INH indicated up regulation of a number of genes relevant to the MOA. In addition to possibly increasing the number of targets by identifying genes within pathways, the generation of "signature transcriptome analysis (TA) profiles" may be useful in predicting the MOA of new chemical entities. This approach has been shown to be quite useful in the search for tumor-specific markers and in distinguishing potential therapeutic targets in neoplastic cells since this tissue is readily available from biopsy. By way of example, a number of TA studies have been performed in prostate cancer (Liotta *et al.,* 2000; Dhanasekaran *et al.,* 2001; Welsh *et al.* 2001; Rhodes *et al.,* 2002). By developing gene expression profiles of a significant number of tumors the possibility of identifying subsets of genes that function as prognostic disease markers or biologic predictors

of therapeutic response increases, thus permitting clinicians to distinguish indolent from aggressive tumors which require immediate therapeutic intervention.

Although microarray studies can reveal the relative amounts of different mRNAs in the cell, levels of mRNA are not directly proportional to the expression level of the proteins they code for. The number of protein molecules synthesized using a given mRNA molecule as a template is highly dependent on transcription-initiation features of the mRNA sequence; in particular, the ability of the promoter region is a key determinant in the recruiting of ribosomes for protein translation. A study of 158,807 mouse transcripts revealed that 4520 of these transcripts form antisense partners that are base pair complementary to the exons of genes (Katayama *et al.,* 2005). These results raise the possibility that significant numbers of "antisense RNA-coding genes" might participate in the regulation of the levels of expression of protein-coding mRNAs.

2.5 PROTEOME AND PROTEOMICS

With the avalanche of genomic information and improvements in analytical technology, proteomics is becoming increasingly important for the study of many different aspects. Since proteins serve as important components of major signaling and biochemical pathways, studies at protein levels are essential to reveal molecular mechanisms underlying growth, development, and interactions with the environment.

Proteomics is the application of evolving technologies to analyze gene products, *proteins*, on a large scale. This concerns protein expression profiles, protein modifications and protein networks in relation to cell function and biological processes e.g. development, health and disease (Macaulay *et al.,* 2005). With the mapping of the human genome, proteomics has rapidly emerged as an exciting new field of research, one that complements rather than replaces genomics. The genome holds all of an organism's genes, which carry the codes for the proteins that it needs to function. Proteomics is the integrated study of proteins and their biological functions and processes, such as hormone production, immune-system responses etc. It allows the study of protein structure and protein – protein interactions, and today, this branch of science is becoming a major tool in biomedical and drug development research.

Since proteins play a central role in the life of an organism, proteomics is instrumental in discovery of biomarkers, such as markers that indicate a particular disease.With the completion of a rough draft of the human genome, many researchers are looking at how genes and proteins interact to form other proteins. A surprising finding of the Human Genome Project is that there are far fewer protein-coding genes in the human genome than proteins in the human proteome (20,000 to 25,000 genes vs. about 1,000,000 proteins). The human body may contain more than 2 million proteins, each having different functions (http://en.wikipedia.org/wiki/Proteomics). The protein diversity is thought to be due to alternative splicing and post-translational modification of proteins. The discrepancy implies that protein diversity

cannot be fully characterized by gene expression analysis, thus proteomics is useful for characterizing cells and tissues.

To catalog all human proteins, their functions and interactions is a great challenge for scientists. An international collaboration with these goals is co-ordinated by the Human Proteome Organization (HUPO). Most proteins function in collaboration with other proteins, and one goal of proteomics is to identify which proteins interact. This often gives important clues about the functions of newly discovered proteins. Several methods are available to probe protein-protein interactions. The traditional method is yeast two-hybrid analysis. New methods include protein microarrays, immunoaffinity chromatography followed by mass spectrometry, and combinations of experimental methods such as phage display and computational methods (Shiaw-Lin *et al.,*2005). Proteomics uses various technologies such as one- and two-dimensional gel electrophoresis is used to identify the relative mass of a protein and its isoelectric point. X-ray crystallography and nuclear magnetic resonance are used to characterize the three-dimensional structure of peptides and proteins. However, low-resolution techniques such as circular dichroism, Fourier transform infrared spectroscopy and small angle X-ray scattering (SAXS) can be used to study the secondary structure of proteins. Mass spectrometry (no-tandem), often MALDI-TOF, is used to identify proteins by peptide mass fingerprinting. Protein profiling with MALDI-TOF MS can be of high use in clinical diagnostics.

Affinity chromatography, yeast two hybrid techniques, fluorescence resonance energy transfer (FRET), and Surface Plasmon Resonance (SPR) are used to identify protein-protein and protein-DNA binding reactions. X-ray Tomography is used to determine the location of labeled proteins or protein complexes in an intact cell and frequently correlated with images of cells from light based microscopes. Software based image analysis is utilized to automate the quantification and detection of spots within and among gels samples. While this technology is widely utilized, the intelligence has not been perfected yet. Because most potential molecular markers and targets are proteins, proteomic profiling is expected to yield more direct answers to functional and pharmacological questions than does transcriptional profiling.

Affinity chromatography, yeast two hybrid techniques, fluorescence resonance energy transfer (FRET), and Surface Plasmon Resonance (SPR) are used to identify protein-protein and protein-DNA binding reactions. X-ray Tomography is used to determine the location of labeled proteins or protein complexes in an intact cell and frequently correlated with images of cells from light based microscopes. Software based image analysis is utilized to automate the quantification and detection of spots within and among gels samples. While this technology is widely utilized, the intelligence has not been perfected yet. Because most potential molecular markers and targets are proteins, proteomic profiling is expected to yield more direct answers to functional and pharmacological questions than does transcriptional profiling.

The molecular biology revolution and the advent of genomic and proteomic technologies are facilitating rapid advances in our understanding of the molecular details of cell and tissue function. These advances have the potential to transform toxicological and clinical practice and are likely to lead to the

supplementation or replacement of traditional biomarkers of cellular integrity, cell and tissue homeostasis, and morphological alterations that result from cell damage or death.

Elucidation of cellular responses to molecular damage, including evolutionarily conserved inducible molecular defense systems, suggests the possibility of new biomarkers based on molecular responses to functional perturbations and cellular damage. Our improved understanding of the molecular basis of various pathologies suggests that monitoring specific molecular responses may provide improved prediction of human outcomes. Responses that can be monitored directly in the human should provide "bridging biomarkers" that may eliminate much of the current uncertainty in extrapolating from laboratory models to human outcome. More importantly the evaluation of one or a few biomarkers is severely limited by the fact that many disease syndromes share inflammatory components or involve the re-modelling of extracellular matrix proteins common to many disease processes (e.g. fibrosis). Some of the diseases have characteristic biomarkers .These biomarkers are used by the physicians for diagnosis. Glycogen synthesis kinase has been a leader in advocating asystematic, highly parallel, combinatorial approach to assemble "disease specific signatures", as well as resolving patterns in response to therapeutic agents. In Alzheimer's disease, elevations in beta secretase creates amyloid/beta-protein, which causes plaque to build up in the patient's brain, which causes dementia (Hye *et al.*, 2006). Targeting this enzyme decreases the amyloid/beta-protein and so slows the progression of the disease. A procedure to test for the increase in amyloid/beta-protein is immunohistochemical staining, in which antibodies bind to specific antigens or biological tissue of amyloid/beta-protein. Heart disease is commonly assessed using several key protein based biomarkers (Vasan , 2006). Standard protein biomarkers for CVD include interleukin-6, interleukin-8, serum amyloid A protein, fibrinogen, and troponins. cTnI cardiac troponin I increases in concentration within 3 to 12 hours of initial cardiac injury and can be found elevated days after an acute myocardial infarction. A number of commercial antibody based assays as well as other methods are used in hospitals as primary tests for acute MI. Proteomic analysis of kidney cells and cancerous kidney cells is producing promising leads for biomarkers for renal cell carcinoma and developing assays to test for this disease (Perroud *et al.*, 2006). In kidney-related diseases, urine is a potential source for such biomarkers. Recently, it has been shown that the identification of urinary polypeptides as biomarkers of kidney-related diseases allows to diagnose the severity of the disease several months before the appearance of the pathology (Rogers et al.(2003).

Proteomic technologies rely on the ability to separate a complex mixture so that individual proteins are more easily processed with other techniques. Well-known methods include low-throughput sequencing through Edman degradation. Higher-throughput proteomic techniques are based on mass spectrometry, commonly peptide mass fingerprinting on simpler instruments, or De novo repeat detection sequencing on instruments capable of more than one round of mass spectrometry. Recent successes illustrate the role of mass spectrometry-based proteomics as an indispensable tool for molecular and cellular biology and for the emerging field of systems biology. These include the study of protein-protein

interactions via affinity-based isolations on a small and proteome-wide scale, the mapping of numerous organelles, the concurrent description of the malaria parasite genome and proteome, and the generation of quantitative protein profiles from diverse species. The ability of mass spectrometry to identify and, increasingly, to precisely quantify thousands of proteins from complex samples can be expected to impact on biology and medicine **(**Aebersold and Mann, 2003). Protein correlation profiling permits the analysis of any multiprotein complex that can be enriched by fractionation but not purified to homogeneity. Andersen et al. (2003) performed a mass-spectrometry-based proteomic analysis of human centrosomes in the interphase of the cell cycle by quantitatively profiling hundreds of proteins across several centrifugation fractions.

Identifying all protein-protein interactions in an organism is a major objective of proteomics. A related goal is to know which protein pairs are present in the same protein complex. High-throughput methods such as yeast two-hybrid (Y2H) and affinity purification coupled with mass spectrometry (APMS) have been used to detect interacting proteins on a genomic scale. However, both Y2H and APMS methods have substantial false-positive rates. Aside from high-throughput interaction screens, other gene- or protein-pair characteristics may also be informative of physical interaction. Therefore it is desirable to integrate multiple datasets and utilize their different predictive value for more accurate prediction of co-complexed relationship. Many proteins participate in cellular processes as members of protein complexes of varying size. It is believed that combinatorial interactions among proteins serve as an important basis for the biological complexity of higher organisms. Therefore, increased knowledge about protein-protein interactions and protein complexes will greatly aid our understanding of protein function (Zhang *et al.*, 2004).

Advances in mass spectrometry (MS), such as surface-enhanced laser desorption ionization (SELDI), hold great promise for early ovarian cancer detection through proteomic profiling of patient serum. Because thousands of proteins and peptides can be characterized and quantified at the same time, large amount of valuable data are obtained for identifying characteristic and effective biomarkers for ovarian cancer detection. Several advanced data mining algorithms have been reported to be promising for diagnosis of early-stage ovarian cancer (Petricoin *et al.,* 2002; Wulfkuhle *et al.,* 2003; Tirumalai *et al.,* 2003; Zhu *et al.,* 2003; Yanagisawa *et al.,* 2003; Pan *et al.,* 2005). However, considerable controversy has been generated, and there remain some critical issues such as reproducibility and robustness of these methods, which make the proteomic profiling approach less congenial for achieving define diagnosis (Diamandis 2004; Baggerly *et al.,* 2005; Ransohoff, 2005). Kristiansen *et al.,* (2004) have carried out a comprehensive characterization of human bile to define the bile proteome by one-dimensional gel electrophoresis and lectin affinity chromatography followed by liquid chromatography tandem mass spectrometry. The strategy can be applicable for a detailed proteomic analysis of most body fluids. In

combination with "tagging" approaches for differential proteomics, the proteomics approach could serve to identify cancer biomarkers from any body fluid (Baggerly *et al.,* 2005).

In terms of the drug development process emphasis has been placed upon staging and progression rather than diagnosis. Recent development of highly multiplexed plate, microbead and protein microarray technologies (Liotta *et al.,* 2005) with larger numbers of analytes (with clinical relevance) gave rise to applications such as protein profiling of plasma from complex disease states. Protein microarray and assay detection systems based on this technology were assumed to emerge as the correlates of DNA microarrays in the identification, quantitation and functional analysis of proteins . In concept, protein microarray is a fairly simple process, antibodies or other affinity reagents are attached to a surface and a cell lysate or serum, is added. After a period of time to allow the antigens present to bind to their cognate antibodies, bound antigen is detected using fluorescently tagged secondary antibodies. Alternatively lysates containing fluorescently tagged or radioactively labelled proteins can be added directly to the array to permit direct detection of the analyte. Currently, "cytokine-chemokine array" strategies appear to be more robust, since a number of well characterized reagents and commercial arrays are available and many disease syndromes display characteristic patterns of cytokine and chemokine expression. Other functional protein arrays are being used to study protein-protein, protein-nucleic acid, protein-allergen, protein-carbohydrate, protein-lipid, enzyme-substrate, and protein-drug interactions. While antibody reagents are not available for most proteins, a number of initiatives have been undertaken to expand the number of antibodies, aptamers and defined proteins for use in quantitative and function based assays.

One of the most promising developments to come from the study of human genes and proteins has been the identification of potential new drugs for the treatment of disease. This relies on genome and proteome information to identify proteins associated with a disease, which computer software can then use as targets for new drugs. For example, if a certain protein is implicated in a disease, its 3D structure provides the information to design drugs to interfere with the action of the protein. A molecule that fits the active site of an enzyme, but cannot be released by the enzyme, will inactivate the enzyme. This is the basis of new drug-discovery tools, which aim to find new drugs to inactivate proteins involved in disease. As genetic differences among individuals are found, researchers expect to use these techniques to develop personalized drugs that are more effective for the individual (Arora *et al.,* 2005).

A computer technique which attempts to fit millions of small molecules to the three-dimensional structure of a protein is called "virtual ligand screening". The computer rates the quality of the fit to various sites in the protein, with the goal of either enhancing or disabling the function of the protein, depending on its function in the cell. A good example of this is the identification of new drugs to target and inactivate the HIV-1 protease. The HIV-1 protease is an enzyme that cleaves a very large HIV protein into smaller, functional proteins. The virus cannot survive without this enzyme; therefore, it is one of the most

effective protein targets for killing HIV. An interesting use of proteomics is using specific protein biomarkers to diagnose disease.

Understanding the proteome, the structure and function of each protein and the complexities of protein-protein interactions will be critical for developing the most effective diagnostic techniques and disease treatments in the future .The long-term challenge of proteomics is enormous: to define the identities, quantities, structures and functions of complete complements of proteins, and to characterize how these properties vary in different cellular contexts. One critical step in tackling this goal is the generation of sets of clones that express a representative of each protein of a proteome in a useful format, followed by the analysis of these sets on a genome-wide basis. Such studies enable genetic, biochemical and cell biological technologies to be applied on a systematic level, leading to the assignment of biochemical activities, the construction of protein arrays, the identification of interactions, and the localization of proteins within cellular compartments (Phizicky *et al.,* 2003).

2.6 METABOLOME AND METABOLOMICS

Since the emergence of genomics, several other "omic" techniques have come to the front, each promising great medical advances (Schmidt, 2004). One of the major challenges of this approach, as with other 'omic' technologies, is that the metabolome is context-dependent and will vary with pathology, developmental stage and environmental factors.

The application of the "-omic-" technologies may lead to a change of paradigms towards the application of complex mixtures in medicine and open the new field of metabolomics (Goodacre , 2005). Metabolomics (also referred to as metabonomics) is the comprehensive study of the metabolome, the repertoire of biochemicals (or small molecules) present in cells, tissue, and body fluids. Following the rapid developments in genome, transcriptome and proteome technology, there is a growing interest in metabolome research. Metabolome refers to the complete set of small-molecule metabolites (such as metabolic intermediates, hormones and other signalling molecules, and secondary metabolites) to be found within a biological sample, such as a single organism. (Oliver *et al.,*1998). It serves as a functional genomic tool for understanding lipid dysfunction in diabetes, obesity and related disorders. The study of metabolism at the global or “-omics” level is a new but rapidly growing field that has the potential to have an impact upon medical practice. The metabolic state reflects what has been encoded by the genome and modified by environmental factors (Kaddurah-Daouk, 2006).

In January 2007, scientists at the University of Alberta and the University of Calgary completed the first draft of the human metabolome (Wishart *et al.,* 2007). They have catalogued and characterized

2500 metabolites, 1200 drugs and 3500 food components that can be found in the human body (Harrigan and Goodacre , 2003).In animals and humans, metabolic profiling of blood and urine components to characterize toxicity and disease states such as inborn errors of metabolism has been ongoing since the introduction of gas-chromatography mass-spectrometry (GC-MS) in the mid-1960s . It could perhaps be said that some of the groundwork has already been done on the animal and human metabolome, although the data have not been analyzed from a bioinformatics perspective .Techniques such as NMR spectroscopy can be used for a very wide range of components of blood or urine, and advanced data analysis techniques locate the key descriptors (biomarkers) of effect. In addition to NMR, other spectroscopic methods such as MS, GC/MS can produce metabolic profiles or signatures of toxicity, disease or drug efficacy.

Thus, the possibility of globally profiling the metabolome of an organism is a genuine analytical challenge, as by definition this must also take into consideration all relevant factors that influence metabolism. Despite these challenges, the approach has already been applied to understand the metabolism in a range of animal models, and has more recently started to be projected into the clinical situation. The use of metabolomic data to predict the health trajectories of individuals will require bioinformatic tools and quantitative reference databases. These databases containing metabolite profiles from the population must be built, stored and indexed according to metabolic and health status. Building and annotating these databases with the knowledge to predict how a specific metabolic pattern from an individual can be adjusted with diet, drugs and lifestyle to improve health represents a logical application of the biochemistry knowledge that the life sciences have produced over the past 100 years (German *et al.*, 2005). Using information from the completely sequenced genomes and known metabolic pathways of various microorganisms, there are ongoing attempts to completely model the metabolism of a cell or microorganism.

In metabolomics research, there are several steps between the sampling of the biological condition under study and the biological interpretation of the results of the data analysis (German *et al.*, 2002). First, the biological samples are extracted and prepared for analysis. Subsequently, different data pre-processing steps (Fiehn, 2002) are applied in order to generate 'clean' data in the form of normalized peak areas that reflect the (intracellular) metabolite concentrations. These clean data can be used as the input for data analysis. However, it is important to use an appropriate data pretreatment method before starting data analysis. Data pretreatment methods convert the clean data to a different scale (for instance, relative or logarithmic scale). Hereby, they aim to focus on the relevant (biological) information and to reduce the influence of disturbing factors such as measurement noise. Procedures that can be used for data pretreatment are scaling, centering and transformations.The ability to use metabolomics approaches for classification and mechanistic studies may influence and augment our ability to study and address the aging process scientifically and clinically (Kristal, and Shurubor, 2005).

Metabolic profiling (especially of urine or blood plasma samples) can be used to detect the physiological changes caused by toxic insult of a chemical (or mixture of chemicals). In many cases, the observed changes can be related to specific syndromes, e.g. a specific lesion in liver or kidney. This is of particular relevance to pharmaceutical companies wanting to test the toxicity of potential drug candidates: if a compound can be eliminated before it reaches clinical trials on the grounds of adverse toxicity, it saves the enormous expense of the trials (Lindon *et al.,* 2004). Metabolites are the key regulators of systems homeostasis. As such, concentration changes of specific groups of metabolites may reflect systemic responses to environmental, therapeutic or genetic interventions. Thus, the study of metabolites is a powerful tool for the characterization of complex phenotypes as well as for the development of biomarkers for specific physiological responses. Metabolic signatures provide prognostic, diagnostic, and surrogate markers for a disease state; the ability to subclassify disease; biomarkers for drug response phenotypes (pharmacometabolomics); and information about mechanisms of disease. Indeed, sophisticated metabolomic analytical platforms and informatics tools have recently been developed that make it possible to define initial metabolic signatures for several diseases (Kaddurah-Daouk.2006). Therefore, metabolomics is a valuable platform for studies of complex diseases and the development of new therapies, both in non-clinical disease model characterization and clinical settings.

Applications to preclinical drug safety studies are illustrated by the Consortium for Metabolomic Toxicology, a collaboration involving several major pharmaceutical companies. The use of metabolomics in disease diagnosis and therapy monitoring and the concept of pharmaco-metabonomics as a way of predicting an individual's response to treatment can be studied (Ellis and Goodacre, 2006). Combination drug therapies with individualized optimization are likely to become a major focus. Metabolomics incorporates the most advanced approaches to molecular phenotype system readout and provides the ideal theranostic technology platform for the discovery of biomarker patterns associated with healthy and diseased states, for use in personalized health monitoring programs and for the design of individualized interventions (German *et al.,* 2003) . It has also emerged as a key technology in preclinical drug discovery and development. The technology enables noninvasive systems assessment of untoward effects induced by candidate compounds characterising a broad spectrum of biological responses on an individual animal basis in a relatively rapid-throughput fashion, thus making it an ideal addition to early preclinical safety assessment (Lindon *et al.,* 2004). However, the implementation and interpretation of the technology and data it generates is not something that should be trivialized. Proper expertise in biological sciences, analytical sciences (nuclear magnetic resonance and/or mass spectrometry) and chemometrics should all be considered necessary prerequisites. If these factors are properly considered, the technology can add significant value as a tool for preclinical toxicologists (Goodacre, 2005).

Today, clinicians capture only a very small part of the information contained in the metabolome, as revealed by a defined set of blood chemistry analyses to define health and disease states. Examples

include measuring glucose to monitor diabetes and measuring cholesterol for cardiovascular health. Such a narrow chemical analysis could potentially be replaced in the future with a metabolic signature that captures global biochemical changes in disease and upon treatment.

2.7 INTEGRATION OF OMIC

Developing a new drug is a tedious and expensive undertaking. The recently developed high-throughput experimental technologies, summarized by the terms genomics, transcriptomics, proteomics and metabolomics provide for the first time ever the means to comprehensively monitor the disease processes at the molecular level. They increase substantially the number of proteins/genes that can be detected simultaneously and have the potential to relate complex mixtures to complex effects in the form of gene/protein expression profiles. The primary aim of "omic" technologies is the nontargeted identification of all gene products (transcripts, proteins, and metabolites) present in a specific biological sample. By their nature, these technologies reveal unexpected properties of biological systems. A second and more challenging aspect of omic technologies is the refined analysis of quantitative dynamics in biological systems. The "-omics" technologies facilitate the systematic characterization of a drug target's physiology, thereby helping to reduce the typically high attrition rates in discovery projects, and improving the overall efficiency of pharmaceutical research processes. Currently, the bottlenecks for taking full advantage of the new experimental technologies are the rapidly growing volumes of automatically produced biological data. A lack of scalable database systems and computational tools for target discovery has been recognized as a major hurdle (Fischer**,** 2005).

High-throughput analysis is essential considering data at the "omic" level, that is to say considering all DNA sequences, gene expression levels, or proteins at once (or, to be slightly more precise, a significant subset of them). Without the ability to rapidly and accurately measure tens and hundreds of thousands of data points in a short period of time, there is no way to perform analyses at this level. There are four major types of high-throughput measurements that are commonly performed: genomic SNP analysis (i.e., the large-scale genotyping of single nucleotide polymorphisms), transcriptomic measurements (i.e., the measurement of all gene expression values in a cell or tissue type simultaneously), proteomic measurements (i.e., the identification of all proteins present in a cell or tissue type), and metabolomic measurements (i.e., the identification and quantification of all metabolites present in a cell or tissue type). Each of these four is distinct and offers a different perspective on the processes underlying disease initiation and progression as well as on ways of predicting, preventing, or treating disease.

Genomic SNP genotyping measures a person's genotypes for several hundred thousand single nucleotide polymorphisms spread throughout the genome. Other assays exist to genotype ten thousand or so polymorphic sites that are near known genes (under the assumption that these are more likely to have

some effect on these genes). The genotyping technology is quite accurate, but the SNPs themselves offer only limited information. These SNPs tend to be quite common (with typically at least 5% of the population having at least one copy of the less frequent allele), and not strictly causal of the disease. Rather, SNPs can act in unison with other SNPs and with environmental variables to increase or decrease a person's risk of a disease. This makes identifying important SNPs difficult; the variation in a trait that can be accounted for by a single SNP is fairly small relative to the total variation in the trait. Even so, because genotypes remain constant (barring mutations to individual cells) throughout life, SNPs are potentially among the most useful measurements for predicting risk.

Transcriptomic measurements (often referred to as gene expression microarrays or "gene chips" are the oldest and most established of the high-throughput methodologies. Gene expression levels influence traits more directly than SNPs, and so significant associations are easier to detect. While transcriptomic measures are not as useful for pre-disease prediction (because a person's gene expression levels very far in advance of disease initiation are not likely to be informative because they have the potential to change so significantly), they are very well-suited for either early identification of a disease (i.e., finding people who have gene expression levels characteristic of a disease but who have not yet manifested other symptoms) or classifying patients with a disease into subgroups (by identifying gene expression levels that are associated with either better or worse outcomes or with higher or lower values of some disease phenotype).

Proteomics is similar in character to transcriptomics. Like transcriptomic measures, though, proteomic measures are excellent for early identification of disease or classifying people into subgroups. As with proteomics, the metabolites are measured in a very fast serial process. NMR is typically used to both identify and quantify metabolites. This technology is newer and less frequently used than the other technologies, but similar caveats apply. Measurements of metabolites are dynamic as are gene expression levels and proteins, and so are best suited for either early disease detection or disease subclass identification.

The new omics technologies seem set to fulfill their huge expectations and in combination they might prove extremely valuable in functional gene analyses. Genomic and transcriptomic studies are mostly conducted by DNA microarray technologies. Proteomics and metabolomics have no standardized procedures yet, but usually, proteome analysis is done by two-dimensional gel electrophoresis and Liquid chromatography-mass spectrometry, while metabolome analysis is conducted through Gas chromatography-mass spectrometry, Liquid chromatography-mass spectrometry and Liquid chromatography -Nuclear magnetic resonance. Usually, these technologies are applied in a “differential display” mode, i.e. by comparing two situations (e.g. diseased versus healthy) in order to reduce the complexity in data by examining only differences (Corthésy-Theulaz I *et al.*, 2005).Transcriptomics provides the tool for deciphering gene expression networks, and proteomics links these networks to protein products. The third crucial partner is metabolomics, which defines the metabolic network(s) linked to gene

expression. NMR and mass spectrometry enable the broad screen analysis of the metabolome and its transformation pathways, transcending classical targeted metabolic studies. These tools were combined to investigate the anticancer mechanisms of different selenium forms in human lung cancer cells (Aardema and MacGregor, 2002). It is, however, necessary to further standardize and automate the methods of especially proteomics and metabolomics in order to make efficient and reproducible high-throughput analyses. Omics will not only have an impact on our understanding of biological processes, but the prospect of more accurately diagnosing and treating disease will soon become a reality(Loughlin, 2007).

Development of microarrays has permitted global measurement of gene expression at the transcript level and provided a glimpse into the coordinated control and interactions between genes (Mutch *et al.*, 2002).Additional "transcriptomic" and other "omic" analyses will further advance heart research. For example, different early heart markers can be helpful to verify the available data, add other perhaps non-Nkx2.5 expressing myocardial populations, and ultimately complete the picture of early myocardial gene expression. Protein profiling (proteome) of the heart, more comprehensive analysis of gene function by knocking down or out of candidate genes (phenome), protein-protein interaction (interactome), and protein cellular and subcellular localization (localizome) are all needed for a deeper understanding of the development, physiology, and pathology of the heart. Each technique will collectively contribute to the diagnosis, prevention, and therapy of the great number of heart diseases.

The unprecedented advances in molecular biology during the last two decades have resulted in a dramatic increase in knowledge about gene structure and function, an immense database of genetic sequence information, and an impressive set of efficient new technologies for monitoring genetic sequences, genetic variation, and global functional gene expression. These advances have led to a new sub-discipline of toxicology: "toxicogenomics". We define toxicogenomics as "the study of the relationship between the structure and activity of the genome (the cellular complement of genes) and the adverse biological effects of exogenous agents. Toxicogenomics help to monitor the expression of multiple genes, proteins, and metabolites simultaneously. It combines new technologies in genomics, proteomics, and metabolomics (Fig 2.2) with traditional tools of pathology and toxicology to study biological response to drugs and other environmental xenobiotics. The biological response to environmental exposure is so complex and involves so many interactive factors that the use of a systems biology analytical approach is required. These new technology will lead to new families of biomarkers that permit characterization and efficient monitoring of cellular perturbations, provide an increased understanding of the influence of genetic variation on toxicological outcomes, and allow definition of environmental causes of genetic alterations and their relationship to human disease. The broad application of these new approaches will likely erase the current distinctions among the fields of toxicology, pathology, genetic toxicology, and molecular genetics. Instead, a new integrated approach will likely emerge that involves a comprehensive

understanding of genetic control of cellular functions, and of cellular responses to alterations in normal molecular structure and function (Aardema and MacGregor, 2002).

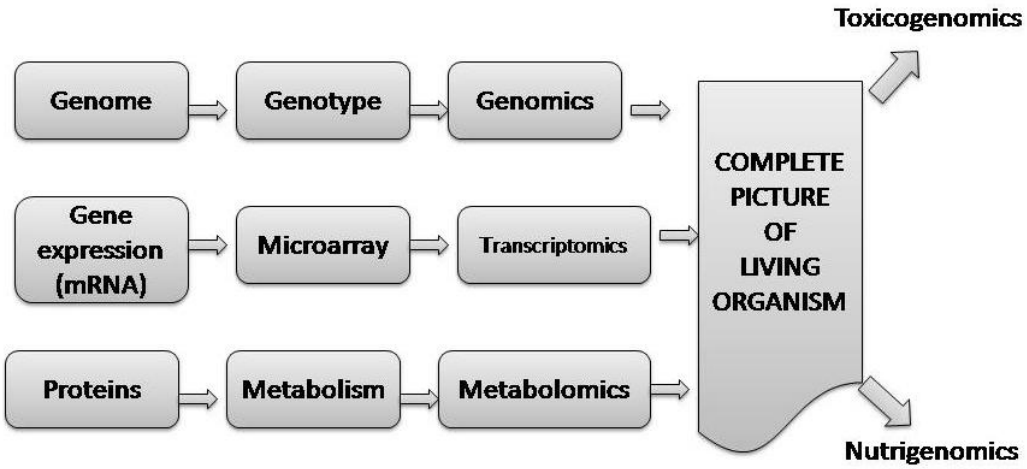

Fig 2.2 Integrating omics technology: One of the challenges of systems biology is to integrate genomics, proteomic, transcriptomic, and metabolomic information to give a more complete picture of living organisms.

Advances in proteomics and genomics contribute to the understanding of the pathophysiology of neoplasia, cancer diagnosis and anticancer drug discovery. Analysis of tumor-specific omics profiles provided a unique opportunity to diagnose, classify, and detect malignant disease; to better understand and define the behavior of specific tumors; and to provide direct and targeted therapy. These technologies however still require integration and standardization of techniques and validation against accepted clinical and pathologic parameters (Nagaraj, 2009).

Less than a decade ago, describing the complexity of chemical behavior in biological systems was severely limited because realistic models presented combinatorial and other problems beyond the capabilities of most computers. In the field of bioinformatics, for example, major advances were made not from faster statistical analysis of data after the acquisition, but from the integration of computational and data acquisition technologies. It is now possible to consider how to evaluate the vast amounts of information generated by "omic" technologies using data-mining tools made possible by rapid advances in computational storage capacity and speed. Within the systems biology framework, functional analyses at the level of gene expression (transcriptomics), protein translation (proteomics), and, more recently, the metabolite network (metabolomics) have become increasingly popular. Metabolomics experiments aim to quantify all metabolites in a cellular system (cell or tissue) under defined states and at different time points so that the dynamics of any biotic, abiotic, or genetic perturbation can be accurately assessed.

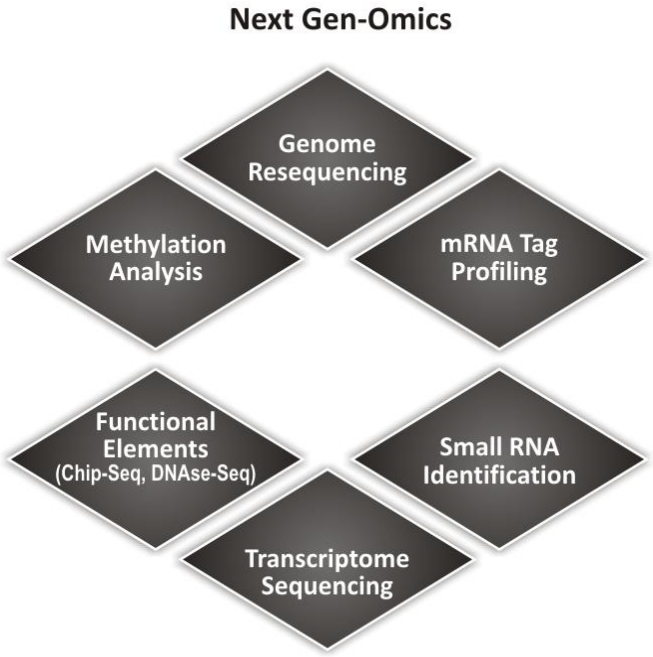

Fig 2.3 The next generation omics will include all the outcomes after integrating omics.

With the progress in genetics, biochemical disorders with high nutritional relevance were linked to a genetic origin. The next generation omics will include genome resequencing, methylation analysis, mRNA tag profiling, small RNA identification, transcriptome sequencing and also the functional elements (Fig 2.3). Genetic disorders with pathological effects are studied e.g. the few genes with pathological obesity. Other gene polymorphisms were described with consequences for human nutrition. The folate metabolism is another good example, where a common polymorphism exists for the gene that encodes the methylenetetrahydrofolate reductase (MTHFR). It was realized however, that many, possibly thousands of other gene polymorphisms might result in minor deviations in nutritional biochemistry, where only marginal or additive effects would result from these deviations. The tools to study the physiological impact were not available at the time and are only now becoming available. 'Nutrigenomics' is the application of the sciences of genomics, transcriptomics, proteomics and metabolomics to human nutrition, especially the relationship between nutrition and health. Nutrigenomics is associated with the issue of personalized nutrition, since claims are being made that differences in genotype should result in differences in the diet and health relationship. The recent advances in nutrigenomics studies are owed to the completion of human genome project and the new biomics technologies that provide means for the simultaneous determination of

the expression of many thousands of genes at the mRNA (transcriptomics), metabolites (metabolomics) and protein (proteomics) levels. The role of metabolomics in nutrigenomics (German *et al.*, 2003) requires understanding, and ultimately regulating, a multitude of nutrient-related interactions at the gene, protein and metabolic levels. These new disciplines and their attendant technologies are changing the paradigms of health research. A number of genetic variations have been shown to increase the susceptibility to diet-related diseases. These include variants that have been associated with Type 2 diabetes mellitus, obesity, cardiovascular diseases, some autoimmune diseases and cancers. Nutrigenetics aims to study these susceptible genes and provide dietary interventions for individuals at risk of such diseases.

Hyperlipidemia is usually associated with atherosclerosis and coronary heart disease (Kaput ,2004). Therapy includes lifestyle changes as altrations in the patient's diet, physical activity and treatment with pharmaceuticals as statins. However, individuals respond differently to the treatment. This was attributed to genetic variations within the population. Genetic variations in genes encoding for apolipoproteins, some enzymes and hormones can alter individual sensitivity to developing cardiovascular diseases (Ordovas and Mooser, 2004). Nutrients can contribute to the development of cancers especially colon, gastric and breast cancer. Several gene variants have been identified as susceptibility genes. One example is the N-Acetyltransferase (NAT) gene. Studies have shown that the NAT2 fast acetylator genotype had a higher risk of developing colon cancer in people who consumed relatively large quantities of red meat(Mutch *et al.*, 2005) .The advent to access comprehensive sets of information (ie, genome, transcriptome, proteome, interactome, phenome, and localizome) has brought a new way of global thinking to biological questions, and analysis using these sets is increasingly the choice for many investigations. Although these approaches may appear as a substitute for the more traditional "reductionist" approach that tackles one or few genes or gene products at one time, they are, in reality, complementary and have the potential to greatly enhance the traditional approach (Abu-Issa and Kirby, 2004).

The "omic," or comprehensive, approach can yield mountains of new information in a relatively short time. Progress in applying the "omic" approach is still in its early exponential phase and does not look like it will plateau any time soon.

2.8 CONCLUSION

The technology platform of genomics, proteomics and metabolomics ("-omic-" technologies) are high-throughput technologies. They increase substantially the number of proteins/genes that can be detected simultaneously and have the potential to relate complex mixtures to complex effects in the form of gene/protein expression profiles. By their nature, these technologies reveal unexpected properties of biological systems. A second and more challenging aspect of omic technologies is the refined analysis of quantitative dynamics in biological systems. Several recent technological advances now make it possible to develop molecular profiles using genomic, proteomic, and metabolomic methods in order to identify the

effects that chemicals may have on living organisms or the environment. Although the technology continues to change and improve, conducting these types of analyses is no longer a question of capability. In summary, the new omics technologies seem set to fulfill huge expectations and in combination they might prove extremely valuable in functional gene analyses. It is, however, necessary to further standardize and automate the methods of especially proteomics and metabolomics in order to make efficient and reproducible high-throughput analyses. Omics will not only have an impact on our understanding of biological processes, but the prospect of more accurately diagnosing and treating disease will soon become a reality.

Advances in "omics" technologies have the potential to revolutionize our approach to disease diagnosis, prognostication and development of novel therapeutics. However, the promise of rapid advances in medicine "from the lab bench to the bedside" has not manifested as of yet. Indeed it appears that the translational applications of genomic-based research have preceded the development of both (i) a conceptual framework for disease understanding and (ii) effective tools that can exploit the vast amounts of data derived from these efforts. In reality great progress has been made, however understanding processes such as disease progression (or drug response) requires systematic insight into dynamic (and temporal) differences in gene regulation, interaction and function (Billelo, 2005).

Omics is a progressive and useful concept in biology. It can revolutionize the way biology is done and how we see life in the future. However, omics fields itself is undergoing change. Some omics do not have livable niche. As time goes by, practically useful omics will survive and go into the mainstream biology while some others will die out until revived in an unexpected way. The omic methods will be powerful aids to understanding biology and advancing clinical practice, but it was never going to be as easy as some led us to believe.

REFERENCES

1. Aardema, M.J., MacGregor, J.T., 2002, Toxicology and genetic toxicology in the new era of "toxicogenomics": impact of "-omics" technologies. *Mutat Res* 499:13-25.
2. Aebersold, R., Mann, M., 2003, Mass spectrometry-based proteomics. *Nature* 422:198-207.
3. Andersen, J.S., Wilkinson, C.J., Mayor, T., Mortensen, P., Nigg, EA, Mann, M., 2003, Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* 426:570-574.
4. Arora, P.S., Yamagiwa, H., Srivastava, A., Bolander, M.E., Sarkar, G., 2005, Comparative evaluation of two two-dimensional gel electrophoresis image analysis software applications using synovial fluids from patients with joint disease. *J Orthop Sci* 10: 160-166.
5. Baggerly, K.A., Morris, J.S., Edmonson, S.R., Coombes, K.R., 2005, Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J Natl Cancer Inst* 97: 307 – 309.

6. Bilello, J.A., 2005, The agony and ecstasy of "OMIC" technologies in drug development. *Curr Mol* 5:39-52.
7. Blagoev, B., Kratchmarova, I., Ong, S.E., Nielsen, M., Foster, L.J., Mann, M., 2003, A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat Biotechnol* 21:315-318.
8. Bruce, S., Kristal, Yevgeniya, I., Shurubor, 2005, Metabolomics: Opening Another Window into Aging Sci. *Aging Knowl Environ* (26), p. pe1
9. Coen, M., Ruepp, S.U., Lindon, J.C., Nicholson, J.K., Pognan, F., Lenz, E.M., Wilson, I.D., 2004, Integrated application of transcriptomics and metabonomics yields new insight into the toxicity due to paracetamol in the mouse. *J Pharm Biomed Anal* 35:93-105.
10. Corthésy-Theulaz, I., den Dunnen, J.T., Ferré, P., Geurts, J.M., Müller, M., van Belzen, N., van, Ommen, B., 2005, Nutrigenomics: The Impact of Biomics Technology on Nutrition Research. *Ann Nutr Metab* 49:355-365.
11. Diamandis, E.P., 2004, Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J Natl Cancer Inst* 96:353 – 356.
12. Ellis, D,I., Goodacre, R., 2006, Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *Analyst* 131:875-885.
13. Fiehn, O., 2002, Metabolomics the link between genotypes and phenotypes. *Plant Mol Biol* 48: 155–171.
14. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, Volckaert G, Ysebaert M., 1976, Complete nucleotide-sequence of bacteriophage MS2-RNA - primary and secondary structure of replicase gene. *Nature* 260: 500-507.
15. Fischer, H.P., 2005, Towards quantitative biology: integration of biological information to elucidate disease pathways and to guide drug discovery. *Biotechnol Annu Rev* 11:1-68.
16. German, J.B., Hammock, B.D., Watkins, S.M., 2005, Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* 1:3-9.
17. German, J.B., Roberts, M.A., Fay, L., Watkins, S.M., 2002, Metabolomics and individual metabolic assessment: the next great challenge for nutrition. *J Nutr* 132: 2486–2487.
18. German, J.B., Roberts, M.A., Watkins, S.M., 2003, Genomics and metabolomics as markers for the interaction of diet and health: lessons from lipids. *J Nutr* 133: 2078S-2083S.
19. Goodacre, R., 2005, Metabolomics shows the way to new discoveries. *Genome Biol* 6:354.
20. Harrigan, G.G., Goodacre, R., 2003, Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis. *Kluwer Academic Publishers* (Boston). ISBN xxx-xxx.
21. Hye, A., Lynham, S., Thambisetty, M., Causevic, M., Campbell, J., Byers, H.L., Hooper, C., Rijsdijk, F., Tabrizi, S.J., Banner, S., Shaw, C.E., Foy, C., Poppe, M., Archer, N., Hamilton, G.,

Powell, J., Brown, R.G., Sham, P., Ward, M., Lovestone, S., 2006, Proteome-based plasma biomarkers for Alzheimer's disease. *Brain* 129: 3042-3050.

22. Joanne, T., Brindle, Henrik, Antti, Elaine, Holmes, George, Tranter, Jeremy, K., Nicholson, Hugh, W.L, Bethell, Sarah, Clarke., Peter, M., Schofield, Elaine McKilligin, David, E., Mosedale, David, J., Grainger, 2002, Rapid and non-invasive diagnosis of the presence and severity of coronary heart disease using 1H NMR -based metabonomics. *Nature Med* 8: 1439-1444.
23. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M., 2006, From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 34:D354-357.
24. Kaput, J., 2004, Diet-Disease Gene Interactions. *Nutrition* 20:26-31.
25. Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M.., Nakamura, M., Nishida, H., Yap, C., C. Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engström, P., G Mizuno, Y., Faghihi, . M.A. Sandelin, A., Chalk, A.M Mottagui-Tabar, S. Liang, Z., Lenhard, B., Wahlestedt, C., 2005, Antisense Transcription in the Mammalian Transcriptome" by the RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) and the FANTOM Consortium: in *Science* 309: 1564-1566.
26. Kristiansen, T.Z., Bunkenborg, J., Gronborg, M., Molina, H., Thuluvath, P.J., Argani, P., Goggins, M.G., Maitra, A., Pandey, A., 2004, A proteomic analysis of human bile. *Mol Cell Proteomics*: 715-728.
27. Lein, E.S., Hawrylycz, M.J., Ao N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., Chen, L., Chen, L., Chen, T.M., Chin, M.C., Chong, J., Crook, B.E., Czaplinska, A., Dang, C.N., Datta, S., Dee, N.R., Desaki, A.L., Desta ,T., Diep, E., Dolbeare, T.A., Donelan, M.J., Dong HW, Dougherty, J.G., Duncan, B.J., Ebbert, A.J., Eichele, G., Estin, L.K., Faber, C., Facer, B.A., Fields, R., Fischer, S.R., Fliss, T.P., Frensley, C., Gates, S.N., Glattfelder, K.J., Halverson, K.R., Hart, M.R., Hohmann, J.G., Howell, M.P., Jeung, D.P., Johnson, R.A., Karr, P.T., Kawal, R., Kidney, J.M., Knapik, R.H., Kuan, C.L., Lake, J.H., Laramee, A.R., Larsen, K.D, Lau, C., Lemon, T.A Liang A.J., Liu, Y., Luong, L.T., Michaels, J., Morgan, J.J., Morgan, R.J., Mortrud, M.T., Mosqueda, N.F., Ng, L.L., Ng R., Orta, G.J., Overly, C.C., Pak, TH., Parry, S.E., Pathak, S.D., Pearson, O.C., Puchalski, R.B., Riley, Z.L., Rockett, H.R., Rowland, S.A., Royall, J.J., Ruiz, M.J., Sarno, N.R., Schaffnit, K., Shapovalova N.V., Sivisay, T., Slaughterbeck,C.R., Smith, S.C., Smith, K.A., Smith, B.I., Sodt AJ, Stewart, N.N., Stumpf, K.R., Sunkin, S.M., Sutram, M., Tam, A., Teemer, C.D., Thaller, C., Thompson, C.L., Varnam, L.R., Visel, A., Whitlock, R.M., Wohnoutka, P.E., Wolkey, C.K., Wong, V.Y., Wood, M., Yaylaoglu, M.B., Young, R.C., Youngstrom, B.L., Yuan, X.F., Zhang, B., Zwingman, T.A., Jones, A.R., 2007, Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 11;445:168-176.

28. Lindon, J.C., Holmes, E., Bollard, M.E., Stanley, E.G., Nicholson, J.K., 2004, Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers* 9:1-31.
29. Liotta, L.A., Lowenthal, M., Mehta, A., Conrads, T.P., Veenstra, T.D., Fishman, D.A., Petricoin, E.F., 3rd 2005, Importance of communication between producers and consumers of publicly available experimental data, *J Natl Cancer Inst* 97:315 – 319.
30. Loughlin, M.F., 2007, Using 'omic' technology to target Helicobacter pylori Expert Opinion on Drug Discovery. 2: 1041-1051.
31. Macaulay, I.C., Carr, P., Gusnanto,A., Ouwehand, W.H., Fitzgerald, D., Watkins, N.A., 2005, Platelet Genomics and Proteomics in Human Health and Disease. *J Clin Invest* 115: 3370-3377.
32. MacGregor, J.T., 2004, Biomarkers of cancer risk and therapeutic benefit: new technologies, new opportunities, and some challenges. *Toxicol Pathol* 1:99-105.
33. Min, Jou W., Haegeman, G., Ysebaert, M., Fiers, W., (1972) Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein, *Nature* 237:82-88.
34. Mutch, D.M., Berger, A., Mansourian, R., Rytz, A., Roberts, M.A., 2002, The limit fold change model: A practical approach for selecting differentially expressed genes from microarray data. *BMC Bioinformatics* 3:17.
35. Mutch, D.M., Wahli, W., Williamson, G., 2005, Nutrigenomics and Nutrigenetics: the emerging faces of nutrition." FASEB J. 19:1602-1616.
36. Nagaraj, N.S., 2009, Evolving 'omics' technologies for diagnostics of head and neck cancer Briefings in Functional Genomics and Proteomics. 8:49-59; doi:10.1093/bfgp/elp004.
37. Nishizuka, S., Charboneau, L., Young, L., Major, S., Reinhold, W.C., Waltham, M., Kouros-Mehr, H., Bussey, K.J., Lee, J.K., Espina, V., Munson, P.J., Petricoin, E., 3rd, Liotta, L.A., Weinstein, J.N., 2003, Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays Proc. *Natl Acad Sci* U S A 25; 100: 14229–14234.
38. Oliver, S.G., Winson, M.K., Kell, D.B., Baganz, F., 1998, Systematic functional analysis of the yeast genome. *Trends Biotechnol* 16: 373–378.
39. Ordovas, J.M., Mooser, V., (2004) Nutrigenomics and nutrigenetics. Curr Opin Lipidol. 15:101-108.
40. Pan, S., Zhang, H., Rush, J., Eng, J., Zhang, N., Patterson, D., Comb, M.J., Aebersold, R., 2005, High throughput proteome screening for biomarker detection. *Mol Cell Proteom* 4:182 – 190.
41. Perroud, B., Lee, J., Valkova, N., Dhirapong, A., Lin, P.Y., Fiehn, O., Kültz, D., Weiss, R.H., 2006, Pathway Analysis of Kidney Cancer Using Proteomics and Metabolic Profiling. *Biomed Central*: 65-82.
42. Petricoin, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., Liotta, L.A., 2002, Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359:572 - 577.

43. Phizicky, E., Bastiaens, P.I., Zhu, H., Snyder, M., Fields, S., 2003, Protein analysis on a proteomic scale. *Nature* 13; 422:208-215.
44. Radwan, Abu-Issa., Margaret, L., Kirby 2004, Take Heart in the Age of "Omics." *Circ Res* 95:335.
45. Ransohoff, D.F., 2005, Lessons from controversy: ovarian cancer screening and serum proteomics. *J Natl Cancer Inst* 97:315 – 319.
46. Rima, Kaddurah-Daouk R., (2006) Metabolic Profiling of Patients with Schizophrenia. *PLoS Med* 3:363.
47. Rogers, M.A., Clarke, P., Noble, J., Munro, N.P., Paul, A., Selby, P.J., Banks, R.E., 2003, Proteomic Profiling of Urinary Proteins in Renal Cancer by Surface Enhanced Laser Desorption Ionization, and Neural-Network Analysis: Identification of Key Issues Affecting Clinical Potential Utility. *Cancer Res* 63: 6971-6983.
48. Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocombe, P.M., Smith, M., 1977, Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 24; 265:687-695.
49. Schmidt, C., 2004, Metabolomics Takes Its Place as Latest Up-and-Coming "Omic" Science. *J the Natl Cancer Inst* 96:732-734
50. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* USA 102:15545-15550.
51. Tirumalai, R.S., Chan, K.C., Prieto, D.A., Issaq, H.J., Conrads, T.P., Veenstra, T.D., 2003, Charaterization of the low molecular weight human serum proteome. *Mol Cell Proteom* 10:1096 – 1103.
52. Vasan, R.S., 2006, Biomarkers of cardiovascular disease: molecular basis and practical considerations. *Circulation* 113:2335-2362.
53. Venkatesh, T.V., Harry, B., 2002, Harlow Integromics: challenges in data integration. *Genome Biol* 3:4027-4027.
54. Wishart, D.S., Tzur, D., Knox,C., Eisner, R., Guo, A,C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M,A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block D., Hau, D.D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G.E., Macinnis, G.D., Weljie, A.M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li L., Marrie, T., Sykes, B.D., Vogel, H.J., Querengesser, L., 2007, HMDB: The Human Metabolome Database. *Nucleic Acids Res* 35:521-526.
55. Wu, S.L., Kim, J., Hancock, W.S., Karger, B., 2005, Extended Range Proteomic Analysis (ERPA): A New and Sensitive LC-MS Platform for High Sequence Coverage of Complex

Proteins with Extensive Post-translational Modifications-Comprehensive Analysis of Beta-Casein and Epidermal Growth Factor Receptor (EGFR). *J Proteome Res* 4:1155 -1170.

56. Yanagisawa, K., Shyr, Y., Xu, B.J., Massion, P.P., Larsen, P.H., White, B.C., Roberts, J.R., Edgerton, M., Gonzalez, A., Nadaf, S., Moore, J.H., Caprioli, R.M., Carbone, D.P., 2003, Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* 362:433 – 439.
57. Zhang, L.V., Wong, S.L., King, O.D., Roth, F.P., 2004, Predicting co-complexed protein pairs using genomic and proteomic data integration. *BMC Bioinformatics* 5: 38.
58. Zhang, Z., Bast, R.C Jr., Yu, Y., Li, J., Sokoll, L.J., Rai, A.J., Rosenzweig, J.M., Cameron, B, Wang, Y.Y., Meng, XY., Berchuck, A., Van, Haaften-Day C., Hacker, N.F., de Bruijn. H.W., van der Zee AG, Jacobs I.J., Fung, E.T., Chan, D.W 2004, Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res* 64:5882 – 5890.
59. Zhu, W., Wang, X., Ma, Y., Rao, M., Glimm, J., Kovach, J.S., 2003, Detection of cancer-specific markers amid massive mass spectral data. *Proc Nat Acad Sci* USA 100:14666 – 14671.