Journal of InterDisciplinary Research Applied to Medicine
JDREAM. Journal of interDisciplinary REsearch Applied to Medicine
JDREAM (2017), v. 1 i. 1, 7-20
ISSN 2532-7518
DOI 10.1285/i25327518 v1i1p7
http://siba-ese.unisalento.it, © 2017 Università del Salento

OMICS Sciences: toward omics personalized medicine

Michele Maffia[1,4], Luana Conte[2], Marco Greco[2], Maria Chiara Maffia[3], Pasquale Simeone[4], Daniele Vergara[1,5]

[1] Laboratory of Clinical Proteomic, "Giovanni Paolo II" Hospital, ASL-Lecce, Italy
[2] Laboratory of Interdisciplinary Research Applied to Medicine (DReAM),
University of Salento at the Hospital "V Fazzi", ASL Lecce, Italy
[3] University San Raffaele, Milan, Italy
[4] Unit of Cancer Pathology, Ce.S.I, Foundation University "G. d'Annunzio", Chieti, Italy Department of Neurosciences, Imaging and Clinical Sciences, University 'G. d'Annunzio', Chieti, Italy
[5] Laboratory of Physiology, Department of Biological and Environmental Sciences and Technologies,
University of Salento, Lecce, Italy

Corrisponding author: Michele Maffia
michele.maffia@unisalento.it

Abstract
The omics sciences of systems biology including genomics, transcriptomics, lipidomics, metabolomics, and proteomics, aim at understanding the biological mechanisms that give rise to the phenotype of an organism by using high-throughput technologies with the promise of great medical advances. The importance of all these sciences is that all, with the exception of genomics are context dependent. Genome is constant in time and place in each cell of an organism, but the entire complement of messenger RNA molecules, proteins and metabolites in a cell, tissue, organ or organism varies with physiological, pathological or developmental conditions (Keusch 2006). The term "omics" represents the study of biological processes as systems. It deciphers the dynamic interactions between the numerous components of a biological system to analyse networks, pathways, and interactive relations that exist among them, such as genes, transcripts, proteins, metabolites, and cells. This new scientific vision has opened the way to new research strategies and experimental technologies that have transformed the study of virtually all life processes. Expansion of the "–ome" concept was incessant and has created a host of new terms, including bacteriome, cardiome, epigenome, erythrome, immunome, microbiome, neurome, connectome, osteome, physiome, proteinome, transportome, degradome, psychome, transcriptome, and many others.
In the present review, these concepts are briefly introduced with a major focus towards proteomics.

Keywords: omics sciences, proteomics, bioinformatics, personalized medicine

1. Genomics, metabolomics, lipidomics, epigenomics

Genomics is the study of the genomes of organisms. For several years genomics was at the forefront of omics sciences, we were in the "Genomic Era". Because many diseases are intimately associated with genetic mutations, the idea that the solutions for human pathologies lie on genes has catalysed the interest of scientists for years, making genome-based analysis methods a central approach in omics science and setting the scene for the completion of the Human Genome Project (HGP), undoubtedly a major landmark event in the field of genomics after the discovery of the double-helical structure of DNA. Since the completion of the human genome project, our ability to explore genome function is increased in specificity. In fact, substantial changes have occurred in the study of genome owing to the introduction of several approaches to DNA sequencing and expression. The massive quantification of messenger RNA (mRNA), genome copy number, and single nucleotide polymorphisms (SNPs) by microarray technology has enabled to assess the expression of tens of thousands of genes shedding light on the mechanisms underlying human pathologies, providing the basis for stratifying patients and predicting outcomes in a variety of diseases. Together with microarrays, recent advances in DNA sequencing with the introduction of next-generation sequencing (NGS) technologies have made possible unprecedented extensive analyses of genome of

individuals. Presently, there are three main NGS systems: the Roche/454 FLX, the Illumina/Solexa Genome Analyzer, and the Applied Biosystems SOLiDTM System. Each one, by a different approach, seeks to amplify single strands of a fragment library and perform sequencing reactions on the amplified strands. Together with these technologies, a new generation of single-molecule sequencing technologies is now emerging offering advantages over current sequencing methods including small amounts of starting material (theoretically only a single molecule may be required for sequencing), and low cost.

An important consequence of this new emerging scenario was the creation of multidisciplinary teams and the formation of large-scale collaborative networks to handle and integrate these large amounts of data. The HGP was the first example of a large collaborative project; others include the Cancer Genome Atlas (TCGA) and the 1000 genome project. TCGA has achieved comprehensive sequencing, characterization, and analysis of the genomic changes of major types of human cancers providing also a platform for researchers to search, download, and analyse data sets generated by TCGA (http://cancergenome.nih.gov). The 1000 genome project aims to establish an extensive catalogue of human variations from 25 populations (www.1000genomes.org). The project provides an international open access resource that serves as a basis for subsequent phenotype related studies (www.1000genomes.org).

Endogenous metabolites can be seen as part of the downstream output of the genome, complementary as "upstream" changes in genes (Spratlin, Serkova, and Eckhardt 2009), and their study together with exogenous metabolite is important to understand what happens in an organism. The term "metabolome" was introduced in 1998 as the total metabolite content of a biological sample, an enormous complex and dynamic number of components that belong to a wide variety of compound classes, including nucleic acids, amino acids, sugars, and lipids.

The term metabolomics was used for the first time in 2000 by Fien, to describe the discipline devot

ed to the study of global metabolite profiles produced in biosynthetic and catabolic pathways from biological systems or originating from host-specific microbes and the intake of nutrients and pharmaceuticals, present in cells, tissues, and biofluids.

Different techniques have been developed to investigate the metabolome, distinguishing the different metabolites on the basis of their chemical and physical properties (Cacciatore and Loda 2015).

Most commonly used techniques for metabolomics are nuclear magnetic resonance (NMR), gas chromatography coupled to mass spectrometer (GC–MS) and mass spectroscopy (MS). Recently published papers describe the application of metabolomics in the study of heart disease, cancer, and other human pathologies. Another potential application of metabolomics involves the definition of biochemical pathways that contribute to drug response. Pharmaco-metabolomic signatures have also been identified for several drugs to predict individual responses to broader medical, dietary, microbiological or physiological challenges.

Data generated experimentally by metabolomics are available in metabolite databases such as the Human Metabolome Database (HMDB). HMDB (www.hmdb.ca), the equivalent of the Human Genome Project for metabolomics, is a resource dedicated to providing scientists with the most current and comprehensive coverage of the human metabolome. Created in 2004, it contains information on biological properties, ontology, spectra and physical properties of human metabolites as well as data regarding metabolite concentrations, disease associations and tissue locations. Started as an inventory of 2500 small molecules, its content has increased to 15000 entries (Liesenfeld et al. 2013).

Lipidomics, term emerged for the first time in 2003, is a sub-discipline of metabolomics that aims to define all of the lipid molecular species in a cell, including the metabolizing enzymes and lipid transporters and understand how lipids function in a biological system. In detail, lipidomics involves the identification of individual cellular lipid species, including the type and number of individual atoms in each lipid species, and their stereoelectronic interactions with other lipids and proteins. Cells use 5% of their genes to synthesize their lipids that fulfil three main functions. Lipids not only forms the bilayer matrix, not only are used as energy stor-

age, but can also act as second messengers and participate in signalling via specialized microdomains, lipid rafts, that have large amounts of lipids. The field of lipidomics is rapidly growing as demonstrated by the great utility of this approach to improve diagnostic–prognostic capabilities for human disorders, and for the identification of new classes of lipids. Similarly to what happened for genome and metabolome, an attempt to characterize the mammalian lipidome has started under the LIPID MAPS initiative (http://www.lipidmaps.org) in 2005. To accomplish this goal, a consortium of twelve independent laboratories from seven academic institutions and one company has been formed (Schmelzer et al. 2007; Fahy et al. 2009).
Early separation and identification of lipids started with GC and HPLC, but other technologies coupled to chromatographic methods, such as MS, Matrix-assisted laser desorption-ionization/time of flight (MALDI/TOF), NMR, and quadrupole–linear ion trap (QTRAP), provide now a powerful approach to the global analysis of complex lipid mixtures. Given the enormous complexity of cellular lipidomics, it has been estimated to encompass around 180 000 – 200 000 different lipid species, high-throughput technologies are needed to approach the entire lipidome of cells. MALDI can also be used to reveal the distribution of lipids in tissues with the technique of imaging mass spectrometry (IMS) obtaining information relevant to the local distribution of lipids as they occur in tissues.
The term "epigenetics" was originally coined by Conrad Waddington to describe heritable changes in a cellular phenotype that were independent of alterations in the DNA sequence. DNA methylation, the transfer of a methyl moiety from S-adenosylmethionine (SAM) to the 5-position of cytosines in certain CpG dinucleotides, represents the most studied of epigenetic processes with a great impact on gene expression. Evidence is mounting for a direct link between DNA methylation and human diseases. Chromatin changes are another central epigenetic process with a role in transcription, repair, replication, and condensation. Overall, there are now at least four different DNA modifications and 16 classes of histone modifications. Gene-specific techniques for determining DNA methylation include bisulfite sequencing, methylation-specific PCR (MSP) and quantitative MSP. The coupling of NGS platforms with established chromatin techniques such as chromatin immunoprecipitation (ChIP-Seq) represents the standard for identifying binding site locations for individual protein and histone modifications.

2. The post-genome era: Proteomics

The complete characterization of all proteins has been the goal of proteomics since its inception more than 20 years ago. Originally coined by Wilkins et al. in 1996, the term "proteome" refers to the entire PROTEin complement expressed by a genOME (Abdallah et al. 2012).
Proteomics techniques offer several advantages over genome-based technologies, as they directly deal with the functional molecules rather than genetic code or mRNA abundance. Even though there is only one definitive genome in an organism, it codes for multiple proteomes since the accumulation of a protein changes in relation to the environment and is the result of a combination of transcription, translation, protein turnover, and posttranslational modifications.
Proteins are the real-time executors of many biological functions and proteomics is the large-scale study of proteins, including their structures, localizations, post-translational modifications (PTMs), and functions.
Proteomics experiments also provide information on protein interactions and complex formation. For example, proteins interact with each other as part of large complexes that serve to execute most biological processes including signal transduction, transcription, and translation. A literature search at the start of 2013 showed there were 38031 articles published on proteomics encompassing several research strategies; today, after four years, their number is more than doubled to a value of more than 86000 articles. In both bacteria and eukaryotes, the cellular concentrations of proteins do not completely correlate with the abundances of their corresponding mRNAs. They often show a squared Pearson correlation coefficient of about 0.40, this means that about 40 % of the variation in protein concentration can be explained by knowing mRNA abundances. This

demonstrates that proteomics represents a complementary to genomics approaches.

The classic proteomics screening methodology combine two different approaches. The first one, called expression-based proteomics, has the aim to define the expression of all proteins present in biological samples. Traditionally, it is performed through the combination of several sequential steps including protein extraction, separation and identification. The general starting point is the protein separation by an electrophoresis system, one- or two-dimensional electrophoresis (1-DE or 2-DE), and the subsequent identification of digested proteins by MS. Alternatively, proteins can also be digested using a specific protease and the resulting peptides separated and analysed immediately by MS. Such approach, namely as shotgun, is considered the method of choice for the large-scale analysis of proteins. The strength of this approach is that it is unbiased; a drawback is that the outcome relies on analysis and interpretation of experimental data. By contrast, targeted proteomic using multiple-reaction monitoring mass spectrometry (MRM-MS) allows the selective detection and quantification of selected peptide ions. Such approach uses the capability of triple quadrupole mass spectrometers to act as ion filters. In a MRM-MS experiment, the precursor ion is isolated in the first quadrupole (Q1), fragmented within Q2 producing fragment ions that are monitored using Q3.

The second approach, functional proteomics, aims to define the biological role of proteins and to identify protein–protein interactions, or interactomes. Protein complexes can be purified in several ways, one very common approach is to use an affinity tag to the protein of interest and purify the interacting partners.

Proteomics has emerged more than two decades ago as a post-genomic technology with the promise to unravel the cellular mechanisms of diseases and to develop reliable markers of diagnosis or treatment. However, such studies remain challenging owing to the high degree of complexity of cellular proteomes, in particular the serum/plasma proteome, and the low abundance of regulatory proteins hidden by abundant proteins. Due to the enormous variation in protein diversity, there is currently no single methodological platform that can be used for a full characterization of the proteome.

2.1. Methods for protein separation

The separation of all the proteins contained within cells, tissues, and biofluids remains a challenging analytical problem. Existing methodologies are not adequate to completely isolate and resolve the large number of proteins present at such different levels of concentration. Proteomic approaches can be classified as either gel-based or gel-free methods that can be further subdivided in "label-free" or "label-based".

2.1.1. Gel-based Proteomics: 2-DE

The attempts to develop a 2-DE started in the late 60', but it was O'Farrell in 1975 who optimized a method on the basis that each electrophoresis separation must be done in independent parameter, to avoid protein being distributed across a diagonal rather than across the entire surface of the gel (Magdeldin 2012). In the first dimension, protein molecules are resolved depending on their isoelectric point (pI); in the second dimension, protein separation is performed based on molecular weight (Magdeldin et al. 2014).

This technique has broadly affected life science research and successfully used applied to the study of biological or clinical samples for the purposes of identifying novel disease-specific protein biomarkers or gaining better understandings novel protein targets for therapeutic interventions and drug developments. During these years, several advances that have enhanced resolution, detection, quantitation, and reproducibility of the technique, increasing the robustness of the entire 2-DE workflow. One of the most notable improvements was the introduction of immobilized pH gradient (IPG) gels that led to standardized procedures for 2-DE permitting higher resolution and improved reproducibility for inter laboratory comparisons. More recently, the development of 2-D differential in-gel electrophoresis (DIGE) in 1997 overcame problems of reproducibility and quantitation because allowed running test and control sample in the same gel. This method was designed in an attempt to increase sensitivity and reproducibility of 2-DE using multiplexed fluorescent dyes- labelled protein samples. 2D-DIGE is based mainly on running more than one sample (maximum 3) on a single

gel at once. Different fluorescent cyanine (Cy) dyes are used for labelling proteins from different samples

This technique enables protein detection at sub picomolar levels and relies on pre electrophoretic labelling of samples with one of three spectrally resolvable fluorescent CyDyes (Cy2, Cy3, and Cy5) (Abdallah et al. 2012). Images are subsequently imported into dedicated 2-DE image analysis softwares.

Although there has been a significant progress towards liquid chromatography and MS methods to separate and analyse proteins, 2-DE still remain a popular technique for conducting proteomic studies. Proteins of interest are excised from the gel, proteolytically digested, and identified using MS (Figure 1). In a single run, up to 1,000 – 2,000 protein species from one complex sample can be separated. In specialised laboratories, using large-gel 2-DE method, the number of protein spots detected were drastically increased up to 10, 000. Subsequently, 2-DE gel easily and efficiently couples with many other analysis and biochemical techniques; spots can be excised, proteins can be extracted and after a tryptic digestion analysed by mass spectrometry.

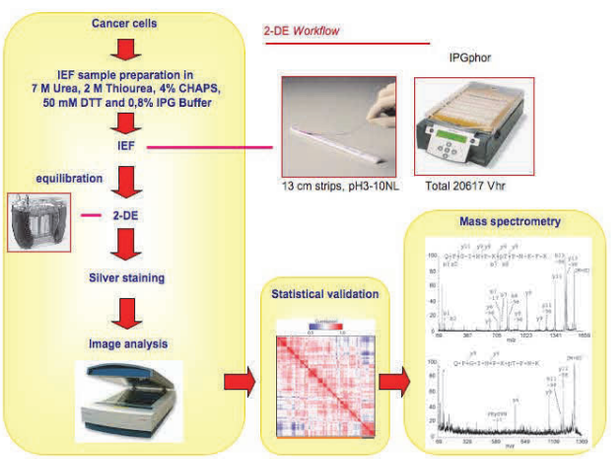

Figure 1. A schematic workflow of two-dimensional gel electrophoresis (2-DE). 2-D gel electrophoresis is an experimental technique that combines two separation methods. Biological samples are grown under different conditions and total proteins are extracted and subjected to isoelectric focusing (IEF) (first-dimension electrophoresis), where proteins are separated according to their isoelectric point (pI). After first dimension, IPG-strips are re-equilibrated to the second dimension buffer conditions, and transferred to the SDS-PAGE gels. Proteins on gels are visualized by MS-compatible stains, including Coomassie or silver staining. Software-based image analysis is then crucial for the biological interpretation of experiments. After statistical validation, differentially expressed spots from 2-DE gels are excised, and a tryptic digestion is performed to generate tryptic peptide mixtures of the proteins that are applied to MALDI- or LC-MS/MS for identification of the excised proteins. The peptide data then are compared with the entire protein database (Swiss-Prot, NCBI).

2.1.2 Gel free-based approaches

Because of problems in quantitative reproducibility and limitations on the ability to study certain classes of proteins, researchers have developed alternatives to 2-DE, for MS-based proteomics techniques.

High-resolution liquid chromatography (LC) separation coupled on line with a mass spectrometer is the central component of a gel free-approach. Complex protein mixtures are digested by trypsin into polypeptides, which are then separated by LC and analyzed by MS via an electro spray ionization (ESI) interface.

In this approach is not the protein itself, which is separated and identified. Instead, proteins are cleaved into peptides using proteolytic enzymes and, subsequently, these peptides are separated and subjected to mass spectrometric analysis. This allows the determination of the protein content of the initial sample (Baggerman et al. 2005).

For this purpose, chromatographic separations are performed using flow rates in the range of low nanoliter per minute (nano-flow liquid chromatography or nanoLC). The relative quantification of peptides usually involves either label-free or stable isotope labelling techniques to identify differences in protein abundances. The labelling methods can be classified into two main groups: chemical isotope tags and metabolic labelling. A variety of labelling approaches including, Proteolytic Labelling, Isotope-Coded Protein Labelling (ICPL), Isotope-Coded Affinity Tags (ICATs), Isobaric Tags for Relative and Absolute Quantification (iTRAQ), TandemMass Tag (TMT), 14N/15N Labelling and Stable Isotope Labelling by Amino Acids in Cell Culture (SILAC) are valuable techniques in quantitative proteomic analysis. The rationale behind each labelling strategy is to create a mass shift that distinguishes identical peptides that exhibit the same chromatographic and ionisation properties, from different samples within a single MS analysis.

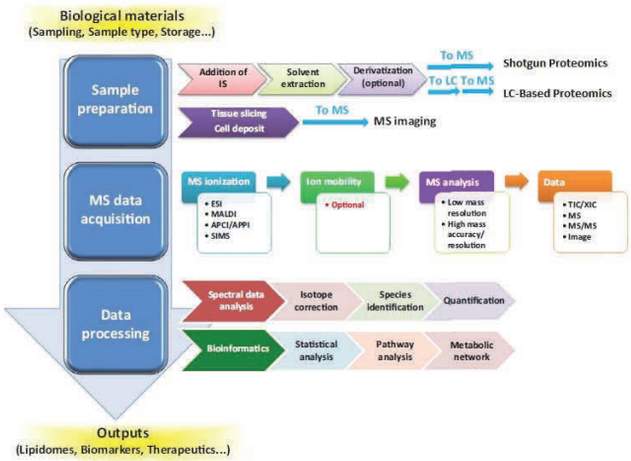

Figure 2. A schematic workflow of gel free electrophoresis (Yang and Han 2016)

3. Limitations of current proteomics approach

In recent decades gel based proteomics techniques became robust and reproducible, however two distinctive issues remains a challenge: the problems to detect low abundance and hydrophobic proteins.

The question of under-representation of hydrophobic protein and in particular membrane proteins is well known and to explain this problem very different possible mechanisms have been proposed: (i) aggregation caused by the low solubility of these protein species in the aqueous media; (ii) protein loss over the sequential steps of the 2-DE processes; (iii) precipitation at the protein corresponding pI during isoelectrofocusing phase; (iv) expression in low copy numbers; (v) difficulties to identify them by MS than hydrophilic proteins. Several fruitful strategies were considered to solve the problem. Usually, the best strategy in 2-DE experiments is to solubilize proteins from the lipid layers by detergent and chaotropic salt. This can be performed by applying a solution of 2M thiourea, 8M urea, and 4% chaps. In the most recent years, other solutions were proposed: use of different zwitterionic detergents, nonionic n-dodecyl β-D-maltoside and zwitterionic amidosulfobetaine ASB-14, and 1,2-diheptanoyl-sn-glycero-3-phosphatdiyl choline (DHPC). Two organic solvents have been recommended for miscible extraction of red blood cells membrane proteins using methanol (MeOH), 2,2,2-tri- fluoroethanol (TFE) and urea.

Despite the many solutions proposed in the recent years the problems with hydrophobic proteins, on 2D gels are widely unsolved. In fact, membrane protein solubility is low at their pI, and therefore membrane proteins tend to precipitate at their pI range. It is evident that the issue represents a built-in problem for all 2D electrophoresis systems IEF-based, for this reason IEF-free separation systems represent a natural alternative in the analysis of membrane proteins.

3.1 Protein abundance

The large number of gene and gene splice-variants that encode proteins, as well as the extensive post-translational modifications of eukaryotic proteins renders proteomic studies extremely difficult. The detection of specific, disease-related protein markers, notoriously difficult to identify, because expressed at low concentration, can be extremely challenging on a classical proteomic experiments where highly abundant proteins could obscure the rare ones. Consequently, there has been an extensive investment into developing techniques and methods capable of revealing the so named "hidden proteome". This cannot be achieved by one single approach. In fact, several methods are used for the enrichment and visualization of the low-abundance proteins and also for the depletion of the high-abundance proteins.

It was demonstrated that the 10% most-expressed gene products represented the 75% of the total protein content, and the 2/3 of less-expressed only 10% of the protein content. In this situation is simple to argue that the signal of high-abundance proteins tends to hide the signal of rare species.

Some of these technical difficulties can be bypassed loading more sample, exploiting the great capacity of 2D gels, allowing many of low abundant proteins to be detected because above the detection limit. But high-loading approach is restricted by gel crowding and is related to the strong presence of normal and modified forms of high abundance protein species. This strategy gives gel with completely saturated zones with no increased performance in the visualization of low abundant species. Possible solutions proposed the use of giant gels with greater resolution and capacity. However, this

technology is inadequate and difficult to use due to the extreme fragility of the gels employed in the analysis.

To address these issues, analytical chemists have attempted to develop pre-fractionation methods to separate large numbers of proteins. Fractionation based-methods that take advantage of proteins function or structure are extensively used, allowing the isolation of specific proteomes: glyco-proteome by lectin columns, phospho-proteome by anti-phospho-aminoacid antibodies or metal-chelating resins. However, these protocols do not resolve the challenge of signal suppression due to high-abundance species present. Immunodepletion columns, containing immobilized antibodies addressing the highest abundance proteins, were proposed as a possible solution. Though, these approaches cause the dilution of the initial sample, rendering it even more difficult to detect low-abundance proteins.

In this scenario, combinatorial hexapeptide ligand libraries have arisen as a powerful method for sample handling and are recently used to better elucidate and obtain extensive information on the protein composition of complex samples like serum, bile fluid, human urine, platelet extracts, and red blood cell lysate. The ligand libraries, designed as batch of chromatographic beads, are synthesized by modified Merrifield approach described by Lam and collaborators. The library consists of millions of affinity baits (hexapeptides) so that each bead comprises multiple copies of the same bait. The beads represent the affinity solid phase of chromatographic column. This hexapeptide ligand library is assembled in order to be able to bind each single protein species present in a given biological extract. On the basis of the saturation-overloading chromatographic principle the loaded proteins are captured by their respective specific ligand until saturation whereas the excess, unbound proteins are washed away. The proteins are captured from the peptide library by several and different combination of interacting forced: Van der Walls interactions, hydrogen bonding, structural docking, hydrophobic associations etc. Thus, high represented proteins species rapidly saturate their specific bead ligands while the excess of the same protein remains unbound. On the contrary, low-abundance proteins were concentrated by their ligand up to saturation. Washing steps eliminate the protein excess not bound to the library and are removed from the chromatographic column. The protein species bound are eluted by a single appropriate elution buffer able of destroy proteins–hexapeptides interaction or by a sequence of desorbing agents, each of them addressing a selected type of binding. These fractions can then be analyzed using well-known methods, such as SDS-PAGE, 2D electrophoresis and MS. An application of hexapeptide libraries to cellular lysates is reported in Figure 2. Proteins obtained from the human T cell lymphoblast-like cell line Jurkat were used as starting material to show the feasibility of this approach.

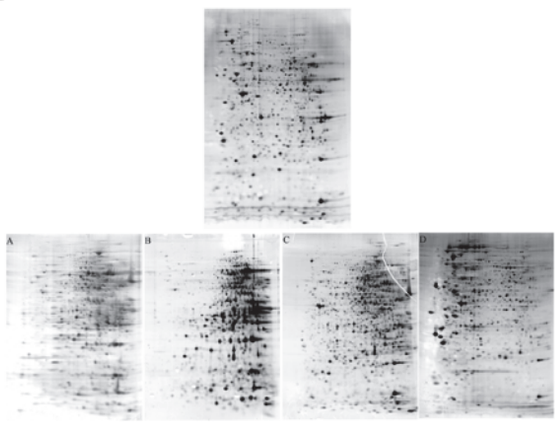

Figure 3. Figure shows the 2-DE maps (pH 3-10) obtained from initial Jurkat cell line extract (top) as compared with those obtained from: (A) 2.5 mg/ml of proteins treated with library and eluted with RS solution (2 M thiourea, 7 M urea, 4% CHAPS); (B) 5 mg/ml of proteins treated with library and eluted with RS solution (2 M thiourea, 7 M urea, 4% CHAPS); (C) and (D) 14 mg/ml of proteins treated with library and eluted with RS solution (2 M thiourea, 7 M urea, 4% CHAPS) and RSA solution (2 M thiourea, 7 M urea, 4% CHAPS, acetic acid to pH 3.3), respectively.

Jurkat cell line was maintained in RPMI-1640 medium containing 10% fetal bovine serum (FBS) and antibiotics. Cells were harvested by centrifugation followed by lysis via sonification in Tris buffer containing protease inhibitors. Varying amounts of cell lysate (2.5 mg/ml, 5 mg/ml, 14 mg/ml) were subjected to column chromatography over a solid-phase combinatorial ligand library (ProteoMiner, Biorad). Following washing, each individual column was subjected to two distinct elutions using a RS solution (2 M thiourea, 7 M urea, 4% CHAPS) and a RSA solution (2 M thiourea, 7 M urea, 4% CHAPS, acetic acid to pH 3.3) respectively. RSA eluted sample was precipitated with 2D Clean-Up (GE Healthcare) and resuspended in RS.

For 2-DE studies, 80 µg of proteins were dissolved in sample buffer and isoelectric focusing of protein samples was carried out by using commercial 13 cm IPG polyacrylamide strips (pH 3 to 10 NL). Separation in the second dimension was carried out in 12% SDS-PAGE gels. Silver stained gels were scanned and analyzed by the software Image-Master 2-D Platinum.

4. Bioinformatics for high-throughput platforms

Over the last decades, large-scale genome sequencing of model organisms and humans (Lander et al. 2001) has been a landmark achievement in biomedical sciences that was fueled by extraordinary advances in molecular biology and computer science. The broad application of genomics and proteomics fields as well as the diffusion of high-throughput platforms, have led to increasing the volume of available data requiring efficient algorithms along with new data management, new analysis and novel visualization techniques. This is in turn helped in redefining the synergy between biology, information sciences, mathematics, and computational techniques. This synergy is largely used by researchers to analyze the information on a large-scale, through the integration of computational algorithms, software tools and databases in an effort to address biological and medical questions.

Luscombe et al. proposed a definition on bioinformatics as a union of biology and informatics: it involves the technology that uses computers for storage, retrieval, manipulation and distribution of information related to biological macromolecules such as DNA, RNA, and proteins (Luscombe, Greenbaum, and Gerstein 2001).

Initial research applications of bioinformatics were primarily focused on analysis of biological sequence data, genome content and rearrangement, as well as for prediction of function and structure of macromolecules (Luscombe, Greenbaum, and Gerstein 2001). Over the last few decades, with the advent of high throughput "omics", it is now feasible to systematically profile a biological system at different levels of molecular and cellular organization, including its epigenome, transcriptome, metabolome, proteome, and interactome (Joyce and Palsson 2006). Overtime, then, the broad nature and scale of these data types have spawned from conventional bioinformatics to specialized branch of bioinformatics such as comparative genomics (Miller et al. 2004), functional genomics (Vukmirovic and Tilghman 2000), network biology (Barabási and Oltvai 2004) and computational proteomics (Colinge and Bennett 2007). Moreover, with the subsequent realization that the complete understanding of biology is obtained from multi level data integration, computational system biology has come into prominence.

Mass spectrometry based proteomics has undergone important advances in the scientific community, due to technological instrumentation and innovation in computational proteomics (Cox and Mann 2007) that is am intensive discipline that requires extensive analytical and data-mining support and bioinformatics approaches.

We can distinguish between qualitative proteomics and quantitative proteomics. In qualitative proteomics, most of the bioinformatics activities are focused on functional data mining of the dataset to extract the global biological theme underlying the proteome. The first step for functional interpretation of a protein list is to connect the protein identifier with its associated Gene Ontology Terms (GO_terms). Genes are then associated to hierarchically clustered, functional terms that describe the "biological process", "molecular function" or "cellular component" (Schmidt 2014). Similarly to GO term enrichment analysis, proteins can also be analyzed for pathway abundance which might be more meaningful since allow the identification of functional biological processes associated to that protein (Schmidt 2014). Comprehensive pathway databases such as KEGG, include many interaction data coming from intracellular reactions such as metabolism or signaling pathways, genetic alterations or drug studies (Kanehisa et al. 2012).

Genome-wide annotational datasets like GO, protein domain organization (PFAM, InterPro) (Mulder et al. 2007), pathways (KEGG) (Kanehisa et al. 2004) and disease mutations (OMIM) have been successfully used for functional proteomics (Pasini et al. 2006). These annotations can be further used in conjuction with statistical tests to find over/under-represented functional categories (Adachi et al. 2007). Additionally, integration with other high throughput "omics" datasets, such as microarray, may provide valuable insights into proteome expression and disease mechanisms (Graumann et al. 2008).

In quantitative proteomics, we know that mass spectrometry provide quantitative data of proteome changes in the cellular states. This information can be used on a binary level of protein changes, for instance normal vs cancer,

stimulated vs non-stimulated cells with growth factors as well as a step of cell cycle or differentiation. There are a lot of methods and approaches for proteomics datasets across multiple conditions or samples. However, in clinical applications of proteomics, biomarker discovery can be done by comparison of proteome profiles of healthy and disease profiles, using genetic algorithms (Petricoin et al. 2002) and on building classification models for disease predictions.

Microarrays have now permeated literally every field of biology and have found many applications in translational research. Large-scale microarray data are also becoming crucial for experimental biology along with computational algorithms, which find wider application in basic research, target discovery, drug discovery, biomarker identification and disease determination. Microarray Bioinformatics include, indeed, a range of inferential (Rhodes et al. 2002) and descriptive statistical methods. Reverse engineering of gene networks using gene expression data based in Bayesian statistics (probabilistic models) (Friedman et al. 2000), Boolean and Relevance networks (Butte and Kohane 2000) and graph theoretic algorithms (Tringe, Wagner, and Ruby 2004) are successfully applied to microarray data to uncover patterns in disease, especially in cancer (Lapointe et al. 2004).

Bioinformatics approach for identification of cis-regulatory sequences has gained impetus in computational prediction of the gene regulation, especially from recent discoveries in transcriptional regulation by regulatory RNA (miRNA, siRNA, piRNA) and their targets (Bentwich 2008) as well as the current appreciation of the role of epigenetics. Bioinformatics applied to gene regulation is one of the most exciting areas of computational research. One of the principle means of coordinating transcription spatially and temporally is through the presence of cis-regulatory elements. These short DNA sequence motifs, proximal to the transcriptional start site (TSS), are bound by transcription factors responsible for the recruitment of the transcriptional initiation machinery and represent principle components that act in response to a particular cellular context and extra-cellular inputs. Global elucidation of gene-specific cis-regulatory control will permit the understanding of global transcriptional networks and facilitate our understanding of the properly functioning of biological system. In addition, de novo prediction of these binding sites has become an active area of research in the field of functional genomics. A complete understanding of this molecular algorithm give great impact on biological research, essential for gaining insights into development, cellular responses to environmental and genetic alterations and the molecular basis of many diseases.

The dynamic interplay between genes, proteins and metabolites leads to a complex biological network that include protein-protein interactions, regulatory circuits linking transcription factors, cis-regulatory elements, signal transduction pathways and metabolic pathways. Network bioinformatics is a fascinating areas of bioinformatics and aims to study this complex biological circuit using mathematical modeling and simulation of pathways, graph-theory analysis of global network structure, application of engineering concept of network analysis as well as de novo design of networks. Protein-Protein interaction network, are typically visualized as "graph networks" where the nodes represent the protein, while an edge connecting two nodes. Protein-Protein interactions are often displayed as a networks illustrating the high degree of connectivity and the presence of particular hub proteins.

Genomics and bioinformatics are now poised to revolutionize our healthcare system by developing personalized and customized medicine. The high speed genomic sequencing coupled with sophisticated informatics technology will allow clinician to quickly sequence a patient's genome and easily detect potential harmful mutations and to engage in early diagnosis and effective treatment of diseases (Xiong 2006).

5. Clinical applications of proteomics

With the advent of proteomics several large-scale studies were launched to investigate the protein profile in different biological systems with the aim of discovering potential diagnostic and prognostic biomarkers. For example, in 2000, the National Cancer Institute (NCI) established an initiative titled the Early Detection Research Network (EDRN) which the objec-

tive to facilitate the development of biomarkers or technology that enable early detection of cancer.

Specifically, this technology offers the possibility of identifying and quantifying proteins associated with a particular disease by means of their altered levels of expression and/or PTMs between the control and disease states. This type of comparative analysis enables correlations to be drawn between the range of proteins, their variations and modifications produced by a cell, tissue and bio-fluids and the initiation, progression, therapeutic monitoring or remission of a disease state. Then, clinical proteomics should be defined as the application of proteomic analysis with the aim of solving a specific clinical problem within the context of a clinical study. The potential applications of proteomics goes from metabolic syndromes like inflammatory diseases or diabetes, to neurological disorder, cancer, dietary interventions, to drug discovery or screening (Yang and Han 2016).

As clinical proteomics consists of a variety of experimental procedures, pre-analytical variability, as well as analytical and post-analytical procedures can markedly affect a proteomic experiment. Collection of appropriate clinical specimens (e.g. urine, blood, tissue), duration of storage, number of freeze-thaw cycles, analysis of proteins and peptides of interest, data interpretation, data validation of protein dataset in a specific clinical context should effectively be standardized to reduce bias. As a result, recommendations concerning minimal information about a proteomic experiment (MIAPE) were released from the Human Proteome Organisation (HUPO). Reference materials are also expected to support both qualitative and quantitative proteomic measurements.

Despite substantial progress in the field, clinical proteomic approaches have not matured into routine diagnostic applications. As described above, proteomic analysis of blood and other body fluids and tissues is extremely difficult due to the complexity of samples and the dynamic range of concentrations of proteins in biological fluids. Major challenges exist for plasma biomarkers discovery, where the large dynamic concentration range of up to ten orders of magnitude for plasma proteins and the presence of very high abundance proteins such as serum albumin and immunoglobulins mask the lower abundance plasma biomarkers. Moreover, to validate biomarkers in a clinical setting it is required the analysis of hundreds (and perhaps more) of high-quality clinical samples. In fact, a huge amount of data and samples is necessary to ensure a bio-statistical significance. This large set of samples is in contrast with the time consuming and intensive proteomic approach that are distant from the routine of a clinical laboratory.

Improvements in this field could lead the way to the use of "omic sciences" as a diagnostic tool for screening and early detection of many pathologies. But they also could represent prognostic and predictive tools, providing information about the outcome of diseases like breast cancers; triple negative tumours have a very worst prognosis than hormone responsive ones (Crutchfield et al. 2016).

One of the major factors for successful proteomic analysis of clinical samples is the selection of an appropriate workflow. For instance, in studies that use biological fluids, samples should be pre-treated to remove high-abundance proteins (running the risk also to eliminate proteins of interest because of protein-protein interactions) or concentrated to enrich the protein fraction (in the case of urine). In biomarkers discovery, caution must be exercised in preserving samples for protein degradation, a problem that can lead to misinterpretation of data.

Working with proximal fluids (synovial fluid, pleural fluid, peritoneal fluid, ascites) it should be necessary to eliminate the contamination of mucosa and salts before sample separation by 2DE-MS, LC-MS), or capillary electrophoresis coupled to mass spectrometry (CE-MS).

With regard to clinical proteomics, among the strategies that have the highest potential to reduce the gap between proteomics and its clinical application there is the possibility to conduct a differential proteome analysis on tissue samples with the advantage to investigate the disease directly at the origin. Moreover, biomarkers present in tissue are more concentrated than those released in the blood making this biological sample suitable for specific isolation or fractionation schemas. However, researchers who work with this type of sample are well-aware that the great heterogeneity of human tis-

sues represents a well-known limit in the investigation of biomarkers. Several approaches were developed to overcome this problem with the potential to be clinically useful. 2D electrophoresis or SELDI (surface-enhanced laser desorption/ionization) have been coupled to laser capture microdissection (LCM), allowing the precise procurement of enriched cell populations from a heterogeneous tissue, or live cell culture, under direct microscopic visualization, or laser microdissection and pressure catapulting techniques (LMPC). In this last procedure, after microdissection, the sample is directly catapulted into an appropriate collection device. As the entire process works without any mechanical contact, it enables pure sample retrieval from morphologically defined origin without cross contamination.

Among the strategies that have the highest potential to reduce the gap between proteomics and its clinical application there is the possibility to conduct a differential proteome analysis directly on tissue samples with the advantage to investigate the disease directly at the origin. In these years, MALDI Imaging has emerged has another promising technique for the combined morphologic and molecular tissue analyses. In detail, MALDI Imaging allows to image/profile intact tissue sections placed onto a conductive glass side obtaining information about protein expression and localization. Because of its practical simplicity and ability to obtain reliable information from tissue section, MALDI imaging might have the potential to complement histopathologic evaluation for assisting in diagnostics, patient stratification, or predicting drug response.

Together with technological challenges, other issues that could affect proteomic results should not be underestimated including harvesting, handling and storage of samples. Considering that the proteome is dynamic over time and expression of a myriad of factors, researchers should also consider the clinical history of the patient such as age, sex, and race. To minimize these systemic problems is desirable to establish a specimen bank (biorepository). A sample, to become eligible for a biorepository, must be collected and analysed immediately because cells and proteins degradation and /or modification may affect the analysis. Moreover, sample should be subject to an accurate quality control and catalogued according to trusted, safe and standardized clinical data.

Based on these observations we conclude that proteomics can be considered as a main strategy for biomarker discovery. However, special attention has to be paid to reduce pre-analytical variables, analytical variability, and biological variation. This will require a close interdisciplinary collaboration involving clinicians, statisticians / bioinformatics, epidemiologists, chemists, biochemists and biologists.

6. Omics science: toward omic personalized medicine

The availability of human genome sequence has transformed biomedical research over the past decade. The end of the 20th century was marked by the genomics revolution. However, over the past decades it has become clear that common diseases develop as a result of multiple defects at different levels including proteins, lipids and metabolites, defects can that cannot be completely predicted by the simple analysis of genes. A systems-level approach, that integrate the results of genomics with those obtained by the analysis of metabolomes and proteomes, has enabled researchers to utilize novel strategies to tackle unexplored research questions in human diseases. The ultimate goal is to evolve an integrated omics picture of the genes, transcripts, proteins, and metabolites to fully describe cellular functioning. At this regard in 2014, a draft map of the human proteome was realised, using the high-resolution Fourier-transform mass spectrometry (Kim MS et al. Nature 2014). In that review was reported a proteomic profiling of 30 histologically normal human samples, including 17 adult tissues, 7 fetal tissues and 6 purified primary haematopoietic cells, that resulted in the identification of proteins encoded by 17,294 genes accounting for approximately 84% of the total annotated protein-coding genes in humans. The proteogenomic analysis revealed a number of novel protein-coding regions, which includes translated pseudogenes, non-coding RNAs and upstream open reading frames. This large human proteome catalogue (available as an interactive web-based resource at http://www.humanproteomemap.org) will complement available human genome and

transcriptome data to accelerate biomedical research and in particular in the diagnosing and treating of human diseases

The greatest benefits for patients are likely to be realized from the monitoring and management of early stage disease rather than from treatment of late stage disease. In this field, a comprehensive integrative omic profiles was applied with success to perform an integrated Personal Omics Profiling (iPOP) on a single healthy individual. Authors of this study combined genomic, transcriptomic, proteomic, metabolomic, and autoantibody profiles, in conjunction with routine laboratory testing, from a single individual over a 14 months period, generating an individual iPOP over the course of healthy states and two viral infections that occurred during the study interval. Peripheral blood mononuclear cells, plasma and serum were collected and results from whole-genome sequencing predicted an increased disease risks for various diseases, including hypertriglyceridemia, and type 2 diabetes (T2D). Markers associated with T2D became elevated during the course of study, in particular, following a respiratory infection. Moreover, data from omics experiments before and after viral infections allowed for the creation of a dynamic picture of this process.

Omics-based approaches will identify at risk groups supporting the implementation of risk-stratified health screening. This may lead to significant cost-savings at the societal level.

Acknowledgement

This work was supported by:
i. Apulia Regional Project Cluster Sistema T7WGSJ3;
ii. "ANGELA SERRA" Foundation for Cancer Research (Parabita - Lecce, Italy);
iii. PONa3_00334 "Research Center for Environment and Human health";
iv. PON02_00563_34847 "RINOVATIS".

Summary box

Given the complexity of cellular systems, several techniques have been developed over the years for the comprehensive analysis of molecular components. In particular, the advent of omic sciences has changed the way in which human diseases are studied making possible the simultaneous interrogation of thousands of molecular species at the system level. It was a technological revolution that modified the way in which experiments are designed, moving from mostly hypothesis- based approaches to studies that are largely hypothesis free. In this book chapter, we briefly discuss classical omics approaches including next-generation sequencing and metabolomics. More in detail, we have opted to focus our attention on proteomics, a complementary approach to genomics that over these years has led to important insights in the comprehension of cellular biological processes and human diseases. Current experimental limitations including the enormous complexity and the dynamic nature of proteomes are also discussed.

When integrated among them, omics approaches have the great potential to provide insight into the molecular alterations that drive disease pathogenesis.

References

- Abdallah, Cosette, Eliane Dumas-Gaudot, Jenny Renaut, and Kjell Sergeant (2012), Gel-Based and Gel-Free Quantitative Proteomics Approaches at a Glance. *International Journal of Plant Genomics* 2012: 494572. doi:10.1155/2012/494572.
- Adachi, Jun, Chanchal Kumar, Yanling Zhang, and Matthias Mann (2007), In-Depth Analysis of the Adipocyte Proteome by Mass Spectrometry and Bioinformatics. *Molecular & Cellular Proteomics : MCP* 6 (7): 1257–73. doi:10.1074/mcp.M600476-MCP200.
- Baggerman, Geert, Evy Vierstraete, Arnold De Loof, and Liliane Schoofs. 2005. Gel-Based versus Gel-Free Proteomics: A Review. *Combinatorial Chemistry & High Throughput Screening* 8 (8): 669–77.
- Barabási, Albert-László, and Zoltán N Oltvai (2004), Network Biology: Understanding the Cell's Functional Organization. *Nature Reviews. Genetics* 5 (2): 101–13. doi:10.1038/nrg1272.
- Bentwich, Isaac. 2008. Identifying Human microRNAs. *Current Topics in Microbiology and Immunology* 320: 257–69.
- Butte, A J, and I S Kohane (2000), Mutual Information Relevance Networks: Functional Genomic Clustering Using Pairwise Entropy Measurements. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 418–29.
- Cacciatore, Stefano, and Massimo Loda (2015), Innovation in Metabolomics to Improve Personalized Healthcare. *Annals of the New York Academy of Sciences* 1346 (1): 57–62. doi:10.1111/nyas.12775.
- Colinge, Jacques, and Keiryn L Bennett (2007), Introduction to Computational Proteomics. *PLoS Computational Biology* 3 (7): e114. doi:10.1371/journal.pcbi.0030114.
- Cox, Jürgen, and Matthias Mann. 2007. Is Proteomics the New Genomics? *Cell* 130 (3): 395–98. doi:10.1016/j.cell.2007.07.032.
- Crutchfield, Christopher A, Stefani N Thomas, Lori J Sokoll, and Daniel W Chan (2016), Advances in Mass Spectrometry-Based Clinical Biomarker Discovery. *Clinical Proteomics* 13: 1. doi:10.1186/s12014-015-9102-9.
- Fahy, Eoin, Shankar Subramaniam, Robert C Murphy, Masahiro Nishijima, Christian R H Raetz, Takao Shimizu, Friedrich Spener, Gerrit van Meer, Michael J O Wakelam, and Edward A Dennis (2009), Update of the LIPID MAPS Comprehensive Classification System for Lipids. *Journal of Lipid Research* 50 Suppl (April): S9-14. doi:10.1194/jlr.R800095-JLR200.
- Friedman, N, M Linial, I Nachman, and D Pe'er (2000), Using Bayesian Networks to Analyze Expression Data. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology* 7 (3–4): 601–20. doi:10.1089/106652700750050961.
- Graumann, Johannes, Nina C Hubner, Jeong Beom Kim, Kinarm Ko, Markus Moser, Chanchal Kumar, Jürgen Cox, Hans Schöler, and Matthias Mann (2008), Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) and Proteome Quantitation of Mouse Embryonic Stem Cells to a Depth of 5,111 Proteins. *Molecular & Cellular Proteomics : MCP* 7 (4): 672–83. doi:10.1074/mcp.M700460-MCP200.
- Joyce, Andrew R, and Bernhard Ø Palsson (2006), The Model Organism as a System: Integrating 'Omics' Data Sets. *Nature Reviews. Molecular Cell Biology* 7 (3): 198–210. doi:10.1038/nrm1857.
- Kanehisa, Minoru, Susumu Goto, Shuichi Kawashima, Yasushi Okuno, and Masahiro Hattori (2004), The KEGG Resource for Deciphering the Genome. *Nucleic Acids Research* 32 (Database issue): D277-80. doi:10.1093/nar/gkh063.
- Kanehisa, Minoru, Susumu Goto, Yoko Sato, Miho Furumichi, and Mao Tanabe (2012), KEGG for Integration and Interpretation of Large-Scale Molecular Data Sets. *Nucleic Acids Research* 40 (Database issue): D109-14. doi:10.1093/nar/gkr988.
- Keusch, Gerald T (2006), What Do -Omics Mean for the Science and Policy of the Nutritional Sciences? *The American Journal of Clinical Nutrition* 83 (2): 520S–522S.
- Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LD, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A, Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang TC, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TS, Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H, Pandey A. (2014) A draft map of the human proteome. *Nature.* May 29;509 (7502):575-81. doi: 10.1038/nature13302.

- Lander, E S, L M Linton, B Birren, C Nusbaum, M C Zody, J Baldwin, K Devon, et al. (2001), Initial Sequencing and Analysis of the Human Genome. *Nature* 409 (6822): 860–921. doi:10.1038/35057062.
- Lapointe, Jacques, Chunde Li, John P Higgins, Matt van de Rijn, Eric Bair, Kelli Montgomery, Michelle Ferrari, et al. (2004), Gene Expression Profiling Identifies Clinically Relevant Subtypes of Prostate Cancer. Proceedings of the National Academy of Sciences of the United States of America 101 (3): 811–16. doi:10.1073/pnas.0304146101.
- Liesenfeld, D. B., N. Habermann, R. W. Owen, A. Scalbert, and C. M. Ulrich (2013), Review of Mass Spectrometry-Based Metabolomics in Cancer Research. Cancer Epidemiology Biomarkers & Prevention 22 (12): 2182–2201. doi:10.1158/1055-9965.EPI-13-0584.
- Luscombe, N M, D Greenbaum, and M Gerstein (2001), What Is Bioinformatics? A Proposed Definition and Overview of the Field. Methods of Information in Medicine 40 (4): 346–58.
- Magdeldin, Sameh. 2012. Gel Electrophoresis - Principles and Basics. Intech Open.
- Magdeldin, Sameh, Shymaa Enany, Yutaka Yoshida, Bo Xu, Ying Zhang, Zam Zureena, Ilambarthi Lokamani, Eishin Yaoita, and Tadashi Yamamoto (2014), Basics and Recent Advances of Two Dimensional- Polyacrylamide Gel Electrophoresis. Clinical Proteomics 11 (1): 16. doi:10.1186/1559-0275-11-16.
- Miller, Webb, Kateryna D Makova, Anton Nekrutenko, and Ross C Hardison (2004), Comparative Genomics. Annual Review of Genomics and Human Genetics 5: 15–56. doi:10.1146/annurev.genom.5.061903.180057.
- Mulder, Nicola J, Rolf Apweiler, Teresa K Attwood, Amos Bairoch, Alex Bateman, David Binns, Peer Bork, et al. (2007), New Developments in the InterPro Database. Nucleic Acids Research 35 (Database issue): D224-8. doi:10.1093/nar/gkl841.
- Pasini, Erica M, Morten Kirkegaard, Peter Mortensen, Hans U Lutz, Alan W Thomas, and Matthias Mann (2006), In-Depth Analysis of the Membrane and Cytosolic Proteome of Red Blood Cells. Blood 108 (3): 791–801. doi:10.1182/blood-2005-11-007799.
- Petricoin, Emanuel F, Ali M Ardekani, Ben A Hitt, Peter J Levine, Vincent A Fusaro, Seth M Steinberg, Gordon B Mills, et al. (2002), Use of Proteomic Patterns in Serum to Identify Ovarian Cancer. Lancet (London, England) 359 (9306): 572–77. doi:10.1016/S0140-6736(02)07746-2.
- Rhodes, Daniel R, Terrence R Barrette, Mark A Rubin, Debashis Ghosh, and Arul M Chinnaiyan (2002), Meta-Analysis of Microarrays: Interstudy Validation of Gene Expression Profiles Reveals Pathway Dysregulation in Prostate Cancer. Cancer Research 62 (15): 4427–33.
- Schmelzer, Kara, Eoin Fahy, Shankar Subramaniam, and Edward A Dennis (2007), The Lipid Maps Initiative in Lipidomics. Methods in Enzymology 432: 171–83. doi:10.1016/S0076-6879(07)32007-7.
- Spratlin, Jennifer L, Natalie J Serkova, and S Gail Eckhardt (2009), Clinical Applications of Metabolomics in Oncology: A Review. Clinical Cancer Research : An Official Journal of the American Association for Cancer Research 15 (2): 431–40. doi:10.1158/1078-0432.CCR-08-1059.
- Tringe, Susannah G, Andreas Wagner, and Stephanie W Ruby (2004), Enriching for Direct Regulatory Targets in Perturbed Gene-Expression Profiles. Genome Biology 5 (4): R29. doi:10.1186/gb-2004-5-4-r29.
- Vukmirovic, O G, and S M Tilghman. 2000. Exploring Genome Space. Nature 405 (6788): 820–22. doi:10.1038/35015690.
- Xiong, J. 2006. Essential Bioinformatics. Edited by Cambridge University Press.
- Yang, Kui, and Xianlin Han (2016), Lipidomics: Techniques, Applications, and Outcomes Related to Biomedical Sciences. Trends in Biochemical Sciences 41 (11): 954–69. doi:10.1016/j.tibs.2016.08.010.