Review paper

Proteomics and genomics – toward new research tools

URSZULA DEMKOW

Department of Laboratory Diagnostics and Clinical Immunology of Developmental Age, Medical University of Warsaw

Abstract

Advances in technology, especially in molecular biology, allow for a rapid expansion of research methods. New proteomics and genomics technologies will enable scientists to discover the functions of genes and their protein products, and to get a clear picture of the complex regulatory networks that control fundamental biological processes. Novel high-throughput technology is being developed to give researchers this possibility. These include proteomic arrays, mass spectrometry and a number of high-throughput technologies. Applying genomic and proteomic methods to body fluids and tissue extracts would place valuable objective analytical power in the hands of the clinician however validation of those methods is an important issue. The array of new research tools will be of strategic importance in biological research and for the biotechnology and pharmaceutical industry, and will become available in clinical medicine to aid diagnosis and therapy.

Key words: *genomics, proteomics, research tools, diagnostics, personalized medicine.*

(Centr Eur J Immunol 2010; 35 (4):245-248)

Genomics

The human genome sequence was described in 2001, creating an opportunity for personalized medicine [1-3]. Genomics is a scientific discipline that characterizes the complete genome of an organism. Its primary approach is to determine the entire sequence of DNA and the relations between different parts of the genome. Transcriptomics is the part of the genomics that describes mRNAs encoding individual proteins.

Genomics-based devices have the potential to become first line tools to identify patients at risk for developing certain diseases or predict unusual reactions to certain drugs. This knowledge could be widely used in clinical medicine. Previously, genes for monogenic or Mendelian disorders were easily discovered [2, 4, 5]. Genomic's tools may allow to find genetic background for complex, multi-factorial disorders like autoimmune diseases, cancer, cardiovascular disorders or even schizophrenia [2, 4].

An array of high-throughput technologies dedicated to genomics research are used to characterize the biological function of genes and genomes [6-8]. For example PCR and DNA sequencing are frequently used to search for genetic variants associated with the development of a disease [8]. Innovative, high-throughput microfluidic systems and automating of the processes involved in genomics research including: extraction and purification of nucleic acids, set up of PCR reaction, automation of RT-PCR as well as the separation and direct detection of DNA and RNA, are on the development too [2, 9-11]. These advanced technologies combine microelectronics with molecular biology tools. The study of gene expression is enabled through microarrays and qRT-PCR techniques that measure RNA levels related to specific gene transcription processes, as well as other RNA-mediated processes [12]. A gene microarray [gene chip] is a miniaturized slide that carries numerous probes of nucleic acids, which are arranged in a grid pattern on the chip. Microarrays are useful because of their small size and because they can examine a very large number of genes. Microarrays can rapidly provide a detailed view of the simultaneous expression of all the genes [around 30 000] in an entire genome, and provide new insights into gene function, disease pathology and classification, and drug development [13]. The main challenge in microarray technology results from the technical complexity of the process and the large amounts of data generated in the experiment [12, 13]. Advanced array technologies include genome sequencing,

Correspondence: Urszula Demkow, Department of Laboratory Diagnostics and Clinical Immunology, Medical University of Warsaw, Marszałkowska 24, 00-576 Warsaw, Poland, e-mail: demkow@litewska.edu.pl

genotyping, transcriptome analysis but also protein analysis, functional cell microarrays and tissue microarrays [14]. Genomic microarrays enable to measure the expression of genes under different medical conditions. A small amount of body fluid has the ability to look at 30 000 genes [12-14]. Genomic microarrays may become useful in early screening of diseases such as lung cancer, which usually are not diagnosed until they are advanced and less treatable [15]. Microarray technologies under development include search for genotypes related to of cancer, allergy, autoimmune diseases, cystic fibrosis and genes having impact on drug metabolism and response [6, 16]. Specific medical applications will also include screening for inherited genetic disorders [such as amyotrophic lateral sclerosis and muscular dystrophy], combating diseases of the nervous system [such as Alzheimer's disease, Parkinson's disease and new variant Creutzfeldt-Jakob disease], cardio-vascular diseases, rare diseases, as well as for detection of infectious agents [14-17]. For example, cystic fibrosis genomic tests find genetic variations in one of the genes that causes cystic fibrosis - the most common fatal genetic disease in the world. Those tests help to diagnose cystic fibrosis in children and identify adults carrying the defective gene [18]. Microarrays are currently used to search for the antigen HLA B27 in rheumatic disease patients and HLA DQ2, DQ8 in celiac disease [19, 20].

New methods for array-based resequencing include oligonucleotide fingerprinting, iFRET technology to detect hybridization, nanoscale hybridizations in nanowells, and MALDI mass spectrometry to detect oligonucleotide composition [2, 21]. A high throughput resequencing device may become a highly competitive tool, which would create possibilities for genome-wide sequence evaluation of patient material as a routine procedure [21, 22]. Moreover monitoring the activity of a genome by measuring mRNA expression levels provides important biological insights. Another recently introduced method is RNA interference [RNAi], an effective mechanism for selective inhibition of gene expression which, has become the preferred method for inhibiting expression of targeted genes [2, 23]. As well as functional genomics applications, it also shows tremendous potential for diagnostics and therapeutics. RNAi libraries covering the entire genome are being developed to secure a functional validation of gene targets.

Recent clinical studies demonstrate that expression analysis of large gene sets can identify molecular profiles correlated to disease states, which may be used for the construction of diagnostic tools [2, 24-26].

Proteomics

Gene expression does not consistently correlate with protein expressions, and can not identify post-transcriptional and post-translational modifications, major modulators of protein function [4, 27, 28]. In the "post-genomic" era, the progress is towards examining proteins as the main effectors of physiological functions [4]. Detailed characterization of the proteome [total body proteins] is a major goal of proteomics, that analyses disease mechanisms by examining changes in the patterns of proteins in patient' body fluids and tissues. The analysis of complex protein mixtures such as serum, other body fluids or tissues by profiling hundreds of proteins in the same time, creates pattern of response characteristic for various cellular states or disease conditions. Detailed proteome analysis has become more realistic today with the high-resolution mass spectrometers capable of faster sequencing in a high-throughput fashion and with the emergence of new techniques such as microarrays. A promising area is the application of advanced mass spectrometric and other quantitative proteomic methodologies to laboratory diagnostics [28-30]. The major proteomic projects of the last decade have shaped proteome-wide sequencing, mapping, and analysis [4]. For example, the creation of the Human Proteome Organization's Human Brain Proteome Project to foster the effective international exchange of brain related proteomic data [4, 31]. Complex diseases are now rapidly investigated by novel high-throughput biochemical technologies to uncover disease activity, clinical markers, and drug targets. Such diagnostic technologies will lead to personalized medicine

Opposite to the genome, the proteome is composed of an active array of molecules constantly being modified and with special localization. Proteomic approaches are able to characterize also post-translational modifications, by which the cell quickly modifies protein functions. Protein profiling and identification techniques using advanced mass spectrometry and bioinformatics can lead to the discovery, identification, and characterization of protein biomarkers [2, 32]. Comprehensive proteomic profiling is able to identify thousands of proteins from the various clinical samples. Once tools for conducting comprehensive proteome analysis became available, much of the interest turned towards analyzing proteins for the purpose of finding novel biomarkers of diseases, such as cancer [33]. In the future, the ability to routinely identify thousands of proteins in the body fluids will be available.

Initial proteomic studies relied on 2D-gel electrophoresis, which separates proteins based on isoelectric point and molecular weight [2, 4]. This process has limited reproducibility, is complicated and not robust. Moreover weakly soluble proteins cannot be easily resolved and only a tiny portion of the proteome can be effectively stained. Another shortcoming of this method is that low-level expressed proteins can be masked by greater expression within a similar molecular weight or isoelectric point, or both. Edgar *et al.* applied this approach to the hippocampal proteome of schizophrenia [4, 34].

Proteomics advanced dramatically with the advent of mass spectrometric analysis for peptides [MALDI] [27, 35]. There are four steps in mass spectrometry. First, the ion source generates ionized proteins from the sample. Second, the mass analyzer sortes and resolves proteins based on their

mass/charge ratio. Third, the ion detector spots the ions and composes data on the ion mass/charge ratio, quantity, and time of flight [TOF], or the time it took to reach the detector. Finally, bioinformatic analysis interprets the raw data [2, 30, 31, 32]. After a mass spectrometry run, in a process called peptide mass fingerprinting, the peptides are arranged into several databases to allow protein identification. Although peptide mass fingerprinting is a method of protein identification, it often requires extensive and often complex purification, as it tenders to interpret protein match by peptide masses rather than by sequences [30, 31, 32]. Mass spectrometry has evolved to incorporate tandem mass spectrometric technology that permits effective sequencing. The MALDI-TOF is an advanced technology, a cutting-edge proteomic tool with direct amino acid sequencing and characterization capabilities [27, 30, 32]. As mass spectroscopy continues to improve, it may replace immunoassays as the best method for measuring specific analytes in biologic samples.

Surface Enhanced Laser/Desorption Ionization [SELDI], a variation of MALDI, is a new generation of mass spectrometric analysis, and offers better accuracy with built in chromatography [27, 30 31, 35]. The central technology platform is a protein chip mass spectrometer, which uses a powerful new approach [surface-enhanced laser desorption/ionisation time-of-flight mass spectrometry, or SELDI-TOF MS] to the analysis of complex protein mixtures such as serum and tissue extracts by profiling hundreds of proteins in the same time, thus creating characteristic patterns related to various cellular states or disease conditions [27, 35]. Each chip contains a unique chromatographic surface for selective protein capture. The recent emergence of methods for rapid profiling thousands of protein markers by use of mass spectrometry has raised hope for the rapid identification of novel cancer biomarkers [27, 35]. Specific questions concerns the reproducibility of SELDI-TOF, possible changes in protocols, calibration, and the ability of SELDI to detect low-abundance tumor markers. Additionally, observations that spectra can vary depending on analytical factors such as the time of processing have been noted during large profiling experiments [27, 36]. Semmes *et al.* demonstrated that relevant part of SELDI-TOF profiles can be measured reproducibly and used to distinguish a reference set of prostate cancer samples from controls [27, 37]. This study was an encouraging step toward defining the analytical reproducibility of serum proteome profiling, although it highlighted the need for rigorous calibration of instruments and adherence to standardized technical procedures [27, 37]. Although promising, the variation attributable to differences in sample collection and other sources of preanalytical bias that can be expected in routine clinical practice [27]. A more comprehensive list of issues that must be addressed to understand the effects of preanalytical, analytical, and postanalytical factors on SELDI-TOF and matrix-assisted laser desorption/ionization [MALDI]-TOF profiles has recently been proposed [27, 38]. The data provide evidence that preanalytical and analytical variation can affect profiled markers, and this result must raise awareness of the strong risk for bias in serum profiling experiments that are not carefully controlled [27, 38].

Functional study employed the profiling and sequencing properties of tandem mass spectrometric analysis. However, several studies have demonstrated the potential of this technology in determining the complexities of the dynamic proteome [4]. Sequence analysis can detect important post-translation protein modifications such as methylation, acetylation, sulfation, phosphorylation, ubiquitylation,and glycosylation [4, 39, 40]. Protein detection can be performed on microarrays, however heterogeneity and relative instability of proteins is a challenge [29, 41]. Current research focus both on protein microarray construction and molecular strategies for specific and sensitive detection [29, 41]. Antibody arrays can be used for protein expression studies and as diagnostic and discovery tools in autoimmunity [42].

One of the important tools in clinical proteomics are tissue microarrays [43]. They allow to analyze hundreds of tissue specimens in the same time. Tissue arrays investigate the distribution of proteins directly at the disease site [2, 43]. The obtained results can be assessed manually or automated and can be analyzed together with clinical data [43].

The number of diagnostic tools has been steadily expanding since the advent of modern medicine. New -omics technologies will allow thousands of results per sample to be generated. With personalized medicine, therapy will be based on individual patient characteristics that become known through bioinformatics. The expected results will give the response rates close to 100%, as well as increased survival rates, improved quality of life, cost savings, and reduced morbidity and mortality. For several years proteomics research has been expected to lead to the finding of new markers that will translate into clinical tests applicable to numerous clinical samples such as serum, plasma and urine [2, 33, 41, 44]. Genome and proteome-based research offers the promise of more effective research and diagnostic tools, and targeted treatment for diseases that affect a vast majority of the population such as cancer, diabetes and cardiovascular disease [2, 45, 46].

References

1. Vostrý M. (2010): Multiplex immunoassays chips and beads. eJIFCC 20: 3.
2. U. Demkow (2010): Laboratory medicine in the scope of proteomics and genomics. eJIFCC21,3 http://www.ifcc.org/index.asp?cat=Publications&scat=eJIFCC_&suba=Vol_21_No_3&zip=1&dove=1&numero=342
3. International Human Genome Sequencing Consortium. (2001): Initial sequencing and analysis of the human genome. Nature 409: 860–921. doi:10.1038/35057062
4. Lakhan SE. (2006): Schizophrenia proteomics: biomarkers on the path to laboratory medicine? Diagnostic Pathology 1:11doi:10.1186/1746-1596-1-11

5. Homfray T, Holland T, Patton M (1995): A new X-linked mental retardation syndrome. Clin Dysmorphol 4: 289-293.
6. Rosenson RS (2010): New technologies personalize diagnostics and therapeutics. Curr Atheroscler Rep 12:184-186.
7. Bentley DR (2006): Whole-genome re-sequencing. Curr Opin Genet Dev 16: 545-552.
8. Metzker ML (2010): Sequencing technologies - the next generation. Nat Rev Genet 11: 31-46.
9. Feng X, Du W, Luo Q, Liu BF (2009): Microfluidic chip: next-generation platform for systems biology. Anal Chim Acta 650: 83-97.
10. Ho YP, Chen HH, Leong KW, Wang TH (2009): Combining QD-FRET and microfluidics to monitor DNA nanocomplex self-assembly in real-time. J Vis Exp 30 pii: 1432. doi: 10.3791/1432.
11. Sims AH (2009): Bioinformatics and breast cancer: what can high-throughput genomic approaches actually tell us? J Clin Pathol 62: 879-885.
12. Ferreira EN, Maschietto M, Silva S, et al. (2010): Evaluation of quantitative rt-PCR using nonamplified and amplified RNA. Diagn Mol Pathol 19: 45-53.
13. Redon R, Carter NP (2009): Comparative genomic hybridization: microarray design and data interpretation. Methods Mol Biol 529: 37-49.
14. O'Connor L, Glynn B (2010): Recent advances in the development of nucleic acid diagnostics. Expert Rev Med Devices 7: 529-539.
15. Singhal S, Miller D, Ramalingam S, Sun SY (2008): Gene expression profiling of non-small cell lung cancer. Lung Cancer 60: 313-324.
16. MacPhee WN, Current IA (2010): Progress in pharma-cogenetics and individualized immunosuppressive drug dosing in organ transplantation. Curr Opin Mol Ther 12: 270-283.
17. Rubinsztein DC (2008): Functional genomics approaches to neurodegenerative diseases. Mamm Genome 19: 587-590.
18. Eshaque B, Dixon B (2006): Technology platforms for molecular diagnosis of cystic fibrosis. Biotechnol Adv 24: 86-93.
19. Ferrer M, Sanz ML, Sastre J, et al. (2009): Molecular diagnosis in allergology: application of the microarray technique. J Investig Allergol Clin Immunol 19 Suppl 1: 19-24.
20. Diosdado B, van Oort E, Wijmenga C (2005): Coelionomics: towards understanding the molecular pathology of coeliac disease. Clin Chem Lab Med 43: 685-695.
21. Zheng J, Moorhead M, Weng L, et al. (2006): High-throughput, high-accuracy array-based resequencing. Proc Natl Acad Sci U S A 21: 6712-6717.
22. Barrett T, Troup DB, Wilhite SE, et al. (2009): NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Res 37: D885-D890.
23. Kiefer J, Yin HH, Que QQ, Mousses S (2009): High-throughput siRNA screening as a method of perturbation of biological systems and identification of targeted pathways coupled with compound screening.Methods Mol Biol 563: 275-287.
24. Cowin PA, Anglesio M, Etemadmoghadam D, Bowtell DD (2010): Profiling the cancer genome. Annu Rev Genomics Hum Genet 11: 133-159.
25. Auffray C, Chen Z, Hood L (2009): Systems medicine: the future of medical genomics and healthcare. Genome Med 1: 2-9.
26. Molloy RM, Mc Connell RI, Lamont JV, FitzGerald SP. (2005): Automation of biochip array technology for quality results. Clin Chem Lab Med 43: 1303-1313.
27. Stephen RM (2005): Diagnostic Proteomics: Back to Basics? Clinical Chemistry 2005; 51: 1333-1334.
28. Ideker T, Thorsson V, Ranish JA, et al. (2001): Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292: 929-934.
29. Stoevesandt O, Taussig MJ, He M. (2009): Protein microarrays: high-throughput tools for proteomics. Expert Rev Proteomics. 6: 145-57.
30. Ocak S, Chaurand P, Massion PP (2009): Mass spectrometry-based proteomic profiling of lung cancer. Proc Am Thorac Soc 6: 159-170.
31. Hamacher M, Meyer HE (2005): HUPO Brain Proteome Project: aims and needs in proteomics. Expert Rev Proteomics 2: 1-3.
32. Walther TC, Mann (2010): Mass spectrometry-based proteomics in cell biology. J Cell Biol. 190: 491-500.
33. Issaq HJ, Xiao Z, Veenstra TD (2007): Serum and plasma proteomics. Chem Rev 107: 3601-3620.
34. Edgar PF, Schonberger SJ, Dean B, et al. (1999): A comparative proteome analysis of hippocampal tissue from schizophrenic and Alzheimer's disease individuals. Mol Psychiatry 4: 173-178.
35. Conrads TP, Zhou M, Petricoin EF, et al. (2003): Cancer diagnosis using proteomic patterns. Expert Rev Mol Diagn 3: 411-420.
36. Coombes KR, Morris JS, Hu J, et al. (2005): Serum proteomics profiling – a young technology begins to mature. Nat Biotechnol 23: 291-292.
37. Semmes OJ, Feng Z, Adam BL, et al. (2005): Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. Clin Chem 2005; 51: 102-112.
38. Hortin GL (2005): Can mass spectrometric protein profiling meet desired standards of clinical laboratory practice? Clin Chem 51: 3-5.
39. Medzihradszky KF (2005): Peptide sequence analysis. Methods Enzymol 2005; 402: 209-244.
40. Wang D, Kalume D, Pickart C, et al. (2006): Identification of Protein Ubiquitylation by Electrospray Ionization Tandem Mass Spectrometric Analysis of Sulfonated Tryptic Peptides. Anal Chem 78: 3681-3687.
41. Rimini R, Schwenk JM, Sundberg M, et al. (2009): Validation of serum protein profiles by a dual antibody array approach. J Proteomics 73: 252-266.
42. Plebani M, Pittoni M, Celadin M, et al. (2009): Recent advances in diagnostic technologies for autoimmune diseases. Autoimmun Rev 8: 238-243.
43. Kononen J, Bubendorf L, Kallioniemi A, et al. (1998): Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med 4: 844-847
44. Pejcic M, Stojnev S, Stefanovic V (2010): Urinary proteomics--a tool for biomarker discovery. Ren Fail 32: 259-268
45. Manolio TA (2010): Genomewide association studies and assessment of the risk of disease. N Engl J Med. 363:166-76
46. Calzone K, Wattendorf D, Dunn BK (2010): The application of genetics and genomics to cancer prevention. Semin Oncol. 37: 407-418.