



# Application of genomics and proteomics in drug target discovery

H.M. Zhang<sup>1,4</sup>, Z.R. Nan<sup>1,4</sup>, G.Q. Hui<sup>1,4</sup>, X.H. Liu<sup>2</sup> and Y. Sun<sup>3,4</sup>

<sup>1</sup>Maize Research Institute, Shanxi Academy of Agricultural Sciences, Xinzhou City, China

<sup>2</sup>College of Life Science, China West Normal University, Nanchong City, China

<sup>3</sup>Biotechnology Research Center, Shanxi Academy of Agricultural Sciences, Taiyuan City, Shanxi, China

<sup>4</sup>Key Laboratory of Crop Gene Resources and Germplasm Enhancement on Loess Plateau, Ministry of Agriculture, China

Corresponding author: X.H. Liu  
E-mail: liuxiaohong0817@aliyun.com

Genet. Mol. Res. 13 (1): 198-204 (2014)  
Received January 25, 2013  
Accepted December 2, 2013  
Published January 10, 2014  
DOI <http://dx.doi.org/10.4238/2014.January.10.11>

**ABSTRACT.** Gene medicine is making breakthroughs in health questions that have baffled humanity for centuries. To understand and utilize gene medicine, it is necessary to realize its action against targets at the molecular level. Currently, many methods can be used to discover drug targets; among these, genomic and proteomic methods are the two most important. In this study, we introduced how to discover drug targets by genomic and proteomic methods in detail. These contents are beneficial for understanding and utilizing the two methods to discover new drug targets of gene medicine.

**Key words:** Drug target; Genomics; Proteomics; Gene medicine

## INTRODUCTION

Gene medicine is making breakthroughs for health questions that have baffled humanity for centuries. At present, it is one of the medical disciplines that affect all aspects of human health, including allergies, bone growth, and cell development. With the development of the human genome project and bioinformatics, more and more gene medicine applications have come forward (Rolland, 1998; Kim, 2011). Target discovery, which involves the identification and early validation of disease-modifying targets, is an essential first step in the drug discovery pipeline (Lindsay, 2003).

Human diseases are correlated with many factors, including genetic and environmental causes. Genomics is focused on only statistical genetic information, while proteomics takes into account the genetic information of DNA or mRNA, as well as events after translation, such as the stability, structural modification (phosphorylation, glycosylation, nitrosation, ethylation, and methylation), and cellular location of protein. Therefore, the information that is obtained by proteomic methods is much greater than that by genomics, but it is more difficult to study disease at the protein level than the nucleic acid level. Today, the discovery of drug targets for specific diseases requires the understanding and use of both genomic and proteomic information.

The process of drug target discovery within the modern scientific context is quite complex. It integrates many disciplines, including structural biology, metabolomics, proteomics, and computer science. The process is generally quite tedious and expensive given the sheer number of possibilities of drug-to-target interactions *in vivo* and the necessity of successfully passing rigorous pharmacokinetic studies and toxicology assays prior to even being considered for clinical trials (Ahn and Wang, 2008).

Genomic and proteomic technologies together with computational methods have led to powerful new strategies in basic and clinical research (Ferrer-Alcon et al., 2009; Trist, 2011). It is useful to reflect on the success of these methods to develop therapeutic strategies for human diseases. In this paper, we describe how to discover drug targets by genomic and proteomic methods.

## GENOMIC METHODS FOR DRUG TARGET DISCOVERY

### Identity comparison for new nucleic acid sequences

First, the new gene sequence that was experimentally obtained needs to be compared with known drug target genes by basic local alignment search tool (BLAST) or scanning the entire genome using the functional motifs of known drug target genes to find relevant functional motifs or ligands by homologous search. It is quite suitable that this kind of strategy was used to study apoptosis and the gene families of immune response pathways. For example, tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL), a known member named apoptosis ligand 2 (Apo2L) in the TNF family, was discovered by searching homologs of the TNF family proteins in the expressed sequence tag database (Wiley et al., 1995).

### Analysis of gene expression profiles

The gene microarray can be simultaneously used for the parallel detection of gene ex-

pression levels between 2 groups of samples. Furthermore, multiple groups of samples might be compared among them as long as the same control is selected. Moreover, microarrays can be used to analyze different cellular mRNA or reverse transcription products that are derived from different individuals (healthy persons and patients), tissues, cell cycles, development periods, differentiation periods, pathological change conditions, or stimulation conditions (including different induction and treatments). The result of this kind of large-scale detection can help us comprehensively recognize special expression of these genes. Today, gene microarray is a very important application that is used to study gene function. Currently, many biomedical companies have accumulated a lot of data on gene expression, and they established databases of gene expression using the gene microarray technology. We can sufficiently utilize these databases to analyze gene function, tissue-specific expression, and regulatory network of a gene, and they can be used further to select candidate drug target genes.

### **Databases of model organisms**

There are over 30 model organisms for which whole genomes were sequenced, and more than hundreds of organisms are being sequenced. The obtained data resources are helpful for functional predictions of genes, especially for pathogenic organisms, and are useful for screening drug targets. For instance, this kind of whole-genome sequence information, integrated from multiple pathogenic organisms, can be used to screen race-specific genes, virulence genes, common genes of pathogenic organisms (conserved genes) specific bacterial or viral enzyme genes, and bacterial membrane-translocation proteins. According to our demands, the above genes are candidates that can be used in the design of broad drugs or “magic bullet”. Furthermore, we can use this information for early detections of some adverse reactions by homogenous analysis with the genes from human, avoiding needless studies at later stages (Zhang, 2009).

### **Hunting for disease-related genes**

The human genome is a dynamic entity. The organism has a set of complicated regulation mechanisms, which were used to regulate protein expression and the response to various stimulation factors. Pathogenic change is also a factor; thus, as long as the expression changes of a variety of proteins are realized clearly, a drug target to the corresponding disease can be found. Therefore, searching for genes that are related to diseases and understanding their structure, function, and location in regulatory networks are very significant in screening drug targets. Presently, many studies or experimental techniques can be used to identify the genes that are related to diseases, including functional cloning, location cloning, gene cloning, or transcription mapping of partial chromosomal regions, as well as physical capture, differential display, exon hunting, DNA microarray, genome-wide scanning, mutant detection, and comparative genomics. Among them, gene differential expression technology is the most popular strategy that is used for high-throughput screening.

### **Analysis of the genes related to drug action**

The genes, whose actions are related to current effective drugs, can be used as drug

targets and further be used for drug discovery. Chinese medicines, especially, have good curative effects to some difficult miscellaneous diseases. If we can confirm the targets and mechanism of drug action, this will have profound significance for developing new drugs and modernizing Chinese medicines. The gene microarray, which is used to analyze the expression profile, can be used to study the difference of gene expression in the same individual (before and after drug) and then determine which genes are differentially expressed between the two conditions. For example, Wilson et al. (1999) used a DNA microarray containing 97% of the predicted open reading frames to monitor changes in *Mycobacterium tuberculosis* gene expression in response to the antituberculosis drug isoniazid. The results showed that isoniazid induced several genes that encode proteins that are physiologically relevant to the drug's mode of action, including an operonic cluster of 5 genes encoding type II fatty acid synthase enzymes and *fbpC*, which encodes trehalose dimycolyl transferase. Other genes, not apparently within directly affected biosynthetic pathways, also were induced. These genes, *efpA*, *fadE23*, *fadE24*, and *ahpC*, likely mediate processes that are linked to the toxic consequences of the drug. Insights gained from this approach may define new drug targets.

### Screening of poisonous side effect genes

The genes leading to or directly participating in poisonous side effects are not our intended drug target genes, but they can indicate the poisonous side effects of a candidate drug, possess important reference value for drug screening, and are also an important module in developing contemporary drug-screening microarrays. For instance, cancer drugs have traditionally been identified in screens that were designed to produce broad biological end points such as cell death. A serious undesired outcome of drugs that was discovered in these screens is that the mechanism of drug action is unknown, and such drugs often have adverse side effects (Onyango, 2004). Consequently, gene screening that would indicate poisonous side effects is very important in drug design and drug discovery; this is also true for studying drug target genes. This study strategy is the same as that for drug-action genes. For example, for the construction of the *ToxBlot* arrays, approximately 2000 sequences of human or murine origin were identified as being of potential relevance to various forms of toxicity or normal cell regulatory or signaling pathways in the study (Pennie et al., 2000). Both a human and a mouse *ToxBlot* array have been constructed, comprising approximately 2400 cDNA sequences in each case and spanning about 600 genes of the relevant species. Reproducibility is ensured by each gene being represented by 4 individual spots on each array and the inclusion of 2 non-overlapping cDNAs for each gene (in duplicate) wherever possible.

## PROTEOMIC METHOD FOR DRUG TARGET DISCOVERY

### Necessity of a proteomic method for drug target discovery

Recent advances in applied genomics helped in the target identification process because it allowed for high-throughput screening of expressed genes. However, studies have shown that there is a poor correlation between the regulation of transcripts and actual protein quantities. The reason for this is that genome analysis does not account for post-translational processes such as protein modifications and protein degradation. Therefore, the methods that were employed in

the drug target discovery process began to shift from genomics to proteomics (Okerberg et al., 2005). Analysis of the dynamic organismal proteome, as opposed to the static genome, will certainly bring a much more accurate approach to identifying not only applicable biomarkers that will aid in diagnosis but also effective remedies for diseases of varying origins.

### **Application of proteomics in drug target discovery**

Proteomics represents the systematic and broad application of technologies that have traditionally supported the field of protein biochemistry. In its most common application, proteomics is used to characterize differences in protein expression between biological specimens (Ryan and Patterson, 2002). Although proteomic technologies can be used to catalog protein differences after metabolic perturbation, its greatest therapeutic value lies in the comparison of cells from normal tissue with those representing a disease state. Such comparisons could enable the identification of disease-specific biomarkers that could be used for diagnostic or prognostic tests, or target proteins that have the potential for drug targets.

Proteomics is the large-scale study of the proteins in a cell, tissue, or entire organism (Dear et al., 2007). Its potential utilization in biomedical research is enormous; for example, it can be used to compare changes in protein levels in normal and diseased tissue (Hanash, 2003). One established technique for comparative proteomics is based on labeling proteins from normal and diseased tissues with different fluorescent dyes (Cy3 and Cy5), mixing the proteins together, and then separating them by isoelectric point and molecular weight (difference in-gel electrophoresis). Because both Cy3- and Cy5-labeled samples are run on the same gel, gel-to-gel variability is reduced (Unlu et al., 1997). An internal standard that is labeled with Cy2 is included on all gels to facilitate comparison across gels (Lilley and Friedman, 2004). Tandem mass spectrometry can then be used to identify the proteins that have changed in abundance when Cy3 and Cy5 images are compared.

In addition, proteomics can analyze biomarkers by quantifying individual proteins and show the separation between one or more protein “spots” on a scanned image from two-dimensional gel electrophoresis; for example, proteomic differences between early and advanced stages of an illness can be observed. To illustrate, Holly et al. (2006) presented a new rat model of sepsis-induced acute renal failure based on cecal ligation and puncture. They used this model to find urinary proteins that may be potential biomarkers and/or drug targets. The conclusion was that this model could be used to find changes in urinary proteins, and this approach identified a potential biomarker and drug target, meprin-1-alpha, by difference in-gel electrophoresis.

### **Challenge of industrial-scale proteomic analysis in drug target discovery**

Industrial-scale proteomic analysis offers the opportunity to conduct systematic studies into cellular and organismal protein biochemistry for the purposes of drug target discovery and academic pursuits. However, potent challenges in sample preparation, mass spectrometric instrumentation, and data analysis need to be identified and overcome for this process to be successful in a wider range of both academic and industrial settings (Ryan and Patterson, 2002). The specific nature of these challenges is determined by the biological study undertaken, the availability and type of instrumentation used, and the computational resources available to the investigator.

## PROSPECTS

The discovery of targets that are sufficiently robust to yield marketable therapeutics is an enormous challenge. Through the years, several approaches have been used with varying degrees of success. These include target-independent screening of tumor-derived cell lines (disease-dependent), reductionist approaches to identifying crucial elements of disease-affected pathways, disease-independent screening of homologs of previously drugged targets, disease-dependent “global” examination of gene transcript levels, and disease-dependent global examination of protein expression levels. These endeavors have been enabled by several major advancements in technology, most recently, the sequencing of the human genome.

Significant advances have been made in recent years in developing new experimental strategies for drug target discovery, biomarker discovery, and drug design (Chakraborty et al., 2009). Some approaches are still restricted to the realm of basic research; however, many of these technologies have now been applied to organismal models of disease as well as cell and tissue pathology specimens. The integration of chemical biology and large-scale analyses by genomics and proteomics, together with methods in computational biology and structural determination, provides new perspectives for addressing problems in clinical medicine and will continue to evolve rapidly to improve drug and target discovery.

## ACKNOWLEDGMENTS

Research supported by the Major Project for Genetically Modified Organisms Breeding from China Agriculture Ministry, P.R. China (#2011ZX08003-001) and the Project for the Doctoral Research Program from Shanxi Academy of Agricultural Sciences, P.R. China (#YBSJJ1106).

## REFERENCES

- Ahn NG and Wang AH (2008). Proteomics and genomics: perspectives on drug and target discovery. *Curr. Opin. Chem. Biol.* 12: 1-3.
- Chakraborty C, Hsu CH, Wen ZH and Lin CS (2009). Recent advances of fluorescent technologies for drug discovery and development. *Curr. Pharm. Des.* 15: 3552-3570.
- Dear JW, Leelahavanichkul A, Aponte A, Hu X, et al. (2007). Liver proteomics for therapeutic drug discovery: inhibition of the cyclophilin receptor CD147 attenuates sepsis-induced acute renal failure. *Crit. Care Med.* 35: 2319-2328.
- Ferrer-Alcon M, Arteta D, Guerrero MJ, Fernandez-Orth D, et al. (2009). The use of gene array technology and proteomics in the search of new targets of diseases for therapeutics. *Toxicol. Lett.* 186: 45-51.
- Hanash S (2003). Disease proteomics. *Nature* 422: 226-232.
- Holly MK, Dear JW, Hu X, Schechter AN, et al. (2006). Biomarker and drug-target discovery using proteomics in a new rat model of sepsis-induced acute renal failure. *Kidney Int.* 70: 496-506.
- Kim SW (2011). Polymeric gene delivery for diabetic treatment. *Diabetes Metab. J.* 35: 317-326.
- Lilley KS and Friedman DB (2004). All about DIGE: quantification technology for differential-display 2D-gel proteomics. *Expert Rev. Proteomics* 1: 401-409.
- Lindsay MA (2003). Target discovery. *Nat. Rev. Drug Discov.* 2: 831-838.
- Okerberg ES, Wu J, Zhang B, Samii B, et al. (2005). High-resolution functional proteomics by active-site peptide profiling. *Proc. Natl. Acad. Sci. U. S. A.* 102: 4996-5001.
- Onyango P (2004). The role of emerging genomics and proteomics technologies in cancer drug target discovery. *Curr. Canc. Drug Targets* 4: 111-124.
- Pennie WD, Tugwood JD, Oliver GJ and Kimber I (2000). The principles and practice of toxigenomics: applications and opportunities. *Toxicol. Sci.* 54: 277-283.

- Rolland AP (1998). From genes to gene medicines: recent advances in nonviral gene delivery. *Crit. Rev. Ther. Drug Carrier Syst.* 15: 143-198.
- Ryan TE and Patterson SD (2002). Proteomics: drug target discovery on an industrial scale. *Trends Biotechnol.* 20: S45-S51.
- Trist DG (2011). Scientific process, pharmacology and drug discovery. *Curr. Opin. Pharmacol.* 11: 528-533.
- Unlu M, Morgan ME and Minden JS (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18: 2071-2077.
- Wiley SR, Schooley K, Smolak PJ, Din WS, et al. (1995). Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* 3: 673-682.
- Wilson M, DeRisi J, Kristensen HH, Imboden P, et al. (1999). Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc. Natl. Acad. Sci. U. S. A.* 96: 12833-12838.
- Zhang YD (2009). *Bioinformatics*. 2nd edn. Science Press, Beijing.