Proteomics and Bioinformatics Approaches for Breast Cancer Researches

Maryam Gholizadeh¹, Seyed Amirreza Parizadeh², Reza Pasandideh³

1. MSc Alumnus of Animal Breeding and Genetic, Department of Animal Science, Ferdowsi University of Mashhad, Iran.
2. Ph.D. Student of Animal Breeding and Genetic, Department of Animal Science, Ferdowsi University of Mashhad, Iran.
3. Ph.D. Student of Animal Breeding and Genetic, Department of Animal Science, Khuzestan Ramin Agricultural and Natural Resources University, Ahvaz.

Corresponding author email: amirreza_prz@yahoo.com

ABSTRACT: Cancer is a disease that occurs due to uncontrolled growth of abnormal cells. Breast cancer is the most common cancer diagnosed in women and consequently has been extensively investigated in terms of histopathology. Proteomics and bioinformatics are newly developing fields and becoming widely used in cancer research. Generally, one of three main steps of proteomics is separation and identification of proteins in a cell, tissue or organism, to gain objective information about protein characterization and finally, the use in database. Bioinformatics is one of the key components of proteomics technology. Cancer bioinformatics deals with the organization and analysis of the data so that important trends and patterns can be identified the ultimate goal being the discovery of new diagnostic protocols for cancer. This article reviews various proteomics and bioinformatics approaches utilized in breast cancer research.

Keywords: Bioinformatics; Breast cancer; MALDI-TOF; Proteomics; SELDI-TOF.

INTRODUCTION

Breast cancer can occur in both men and women, but it's far more common in women. The most common types of breast cancer are ductal carcinoma (85-90% of all cases) and lobular carcinoma (8% of all cases). Unfortunately, only 50% of the breast cancers are localized at the time of diagnosis (Galvão et al., 2011). Despite the availability and recommended use of mammography as a routine screening method for women 40 years of age and older, its effectiveness in reducing overall population mortality from breast cancer is still being investigated (Chung et al., 2013; Antman et al., 1999). Additionally, technical advances in breast ultrasonography and breast MRI provide physicians with a full armamentarium of supplemental diagnostic imaging modalities for earlier detection of abnormalities found on clinical breast examination or mammography. The completion of the Human Genome Project has led to a surge in the use of genomic and proteomic technologies for identification of marker for early cancer detection and molecular targeted treatments. Additionally, profiling of low molecular weight proteins in patients’ serum samples is now possible using high-tech instruments (Khan et al., 2003). This era is radically changing diagnostic pathology, as we know it today. The focus of pathology is shifting from isolated morphologic diagnostics to a complete consultation model that includes information about diagnosis, prognosis and therapy, with the use of specific biomarker data to customize therapeutics for progressive cancers (Li HJ et al., 2013). Pathologists will be required to analyze data from diverse sources and provide a comprehensive report. There is a need for a coordinated effort to maximize the yield of information that can be obtained from these valuable samples received in the pathology laboratory. Bioinformatics is an essential component of this new approach. Proteomics can be defined as the detection, identification, and quantification of all proteins present in a particular tissue, organ, and organism to provide accurate and comprehensive data about that system (Zhu et al., 2007; Baker et al., 2001). The premise for this type of proteomics is that all cells have unique identifiable characteristics related to their role in the body and that during transformation into cancer cells, the signature changes. This change then becomes a unique fingerprint of the presence and character of cancer. The ability to detect this unique fingerprint could facilitate earlier detection and treatment of tumors, thereby affecting patient outcome (Bertucci et al., 2006). The need for bioinformatics will become ever more important as new technologies increase the already exponential rate...
at which cancer data are generated. Cancer bioinformatics is one of multiple ways to concentrate bioinformatics methods in cancer, according to the specificity of disease metabolisms, signalling, communication, and proliferations. Clinical bioinformatics, an emerging science combining clinical informatics, bioinformatics, medical informatics, information technology, mathematics, and omics science together, can be considered to be one of critical elements addressing clinical relevant challenges in early diagnosis, efficient therapies, and predictive prognosis of patients with cancer. There is a need to develop cancer bioinformatics-specific methodologies or introduce new and advanced bioinformatics tools to answer the specific question of cancer (Wu et al., 2012).

Proteomics applications in breast cancer
Proteomics elucidates the properties of proteins, which cannot be understood by analyzing gene expressions such as post-translational modifications, compartmentalization of proteins, and formation of multi-protein complexes (Chung et al., 2007). The foundation for any biomarker discovery effort is based on identification of proteins that show differential expression between disease and control samples. Complete sequencing of human genome has led to the assembly of protein databases, which increased the speed of developments in proteomics research. Although there are about 20,000-30,000 genes in the human genome, due to alternative splicing and sequence deletions, human proteome consists of approximately a million different proteins which makes proteomic research even more difficult. Methods used in proteomics allow the validation of multiple markers at once, greatly decreasing the study time (Ornstein et al., 2000; Berman et al., 2000). There are various techniques utilized in the field of proteomics. Matrix-assisted laser desorption and ionization time-of-flight (MALDI–TOF) and surface-enhanced laser desorption and ionization time-of-flight (SELDI–TOF) are two of the methods currently being employed for this purpose. Since MALDI–TOF and SELDI–TOF analyses generate streams of data comprised of tens of thousands of data points, complex computational systems have been devised to discover subtle changes in protein expression patterns which could be diagnostic or predictive of cancer. Each of these modalities is finding its niche in gene discovery, signal pathway transduction mapping, and target selection for molecular-based treatments in a variety of cancers (Leak et al., 2002).

MALDI-TOF mass spectrometry method in research breast cancer
Matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry (MS) has been used in various studies to improve the existing method of MALDI-TOF MS, discover previously unknown multifunctional proteins, and identify new functions of already known proteins (Timms et al., 2007). MALDI-TOF MS is currently being tested in clinical trials in the field of drug development for protein biomarkers, which may help in developing disease therapy and predicting disease outcome. Biomarkers help predict disease susceptibility, outcome, treatment response, and toxicity. For example, proteins and fats may be used as biomarkers, which may suggest why some smokers get cancer but others do not; why some people have a higher incidence of cancer after exposure to a toxicant and others do not; and why some women are more prone to breast cancer than others (de Noo et al., 2006).

Method of MALDI-TOF mass spectrometry
MALDI-TOF mass spectrometry, a laser is used to ionize the sample of interest. The sample is premixed with a highly absorbing matrix compound to obtain consistent and reliable results by preventing decomposition of molecules. When the sample with matrix is bombarded with laser light, the matrix absorbs the laser energy and becomes charged. The charged matrix in turn transfers part of the charge to the target molecule (e.g., protein), thus giving a charge to the molecules (Zhang et al., 2005). This facilitates efficient energy transfer and spares the molecules from excessive or disruptive direct energy and prevents them from breaking down. The time-of-flight (TOF) analyzer is most commonly used with MALDI because of its large mass range and because it is suited to the MALDI ionization process. In the TOF analyzer, the ionized target molecules are accelerated using an electrical field. Different molecules are separated according to their size and charge and reach the detector at different times, each producing a distinct signal. Because all molecules break down into the same fragments, the same mass spectrum, known as a molecular fingerprint, is displayed. This allows for easy detection. MALDI-TOF is a very sensitive method that allows detection of very small quantities of molecules (Zeidan et al., 2008).

Limitations of MALDI-TOF MS method
One major limitation associated with MALDI-TOF MS is the difficulty in mass determination of large molecules. Also, in cases in which the mass difference between nucleotides is very small, MALDI-TOF may not be able to accurately determine the mass of such molecules (Michiels et al., 2005). Conventional MALDI-TOF MS has not been used much in quantitative studies, which involves the determination of the number of target molecules in a
sample of interest, because of lack of standardization. Also, the conventional technique of MALDI-TOF MS requires the destruction of the specimen under study. However, with many advancements made to MALDI-TOF, the specimen may be preserved for further studies (Umar et al., 2005).

Nevertheless in almost every published paper, the profiles generated by MALDI-TOF MS have been shown to yield better diagnostic sensitivities and specificities than the established cancer biomarkers in ebygon use. Because of this, the MALDI-TOF MS approaches have received extensive publicity since they promise to revolutionize early cancer detection, sub-classification, prognosis, prediction of therapeutic response, etc. However, the initial enthusiasm has been tempered somewhat by parallel reports that have identified potential problems with this approach and its clinical reliability (Diamandis., 2003; Zhang et al., 2005). These issues are not unlike those facing the gene transcript profiling community. Future validation studies will determine how close this technology is for clinical use

**SELDI-TOF MS Proteomics in Breast Cancer**

Among the many promising strategies in clinical proteomics, protein expression profiling using SELDI-TOF MS has emerged as a successful tool for the detection and differentiation of many disease types. This technology originally described by Hutchens and Yip is a promising, rapid, approach for presymptomatic screening and preclinical early detection of pathologies (Zhang et al., 2005). Current projects are aiming to characterize patient and volunteer blood sera for biomarker profiles in breast cancer (Perou et al., 2000). SELDI-TOF MS combines the principles of retention chromatography and mass spectrometry, providing a rapid, high-throughput, sensitive screening method capable of detecting and analysing complex protein samples. This new technology makes multi-analyse discovery possible. It can rapidly perform the separation, detection and analysis of proteins at the femtomole level directly from biologic samples. SELDI is an ionization method in mass spectrometry that is used for the analysis of protein mixtures (Hu et al., 2005). SELDI is typically used with time-of-flight mass spectrometers and is used to detect proteins in tissue samples, blood, urine, or other clinical samples. Comparison of protein levels between patients with and without a disease can be used for biomarker discovery (Baggerly et al., 2004). SELDI-TOF-MS is a variation of MALDI that uses a target modified to achieve biochemical affinity with the analytic compound.

**Method of SELDI-TOF MS technology**

In SELDI, the protein mixture is spotted on a surface modified with a chemical functionality. Some proteins in the sample bind to the surface, while the others are removed by washing. After washing the spotted sample, the matrix is applied to the surface and allowed to crystallize with the sample peptides (figure 1). Binding to the SELDI surface acts as a separation step and the subset of proteins that bind to the surface are easier to analyze. Common surfaces include CM10 (weak-positive ion exchange), H50 (hydrophobic surface, similar to C6-C12 reverse phase chromatography), IMAC30 (metal-binding surface), and Q10 (strong anion exchanger). Surfaces can also be functionalized with antibodies, other proteins, or DNA. Samples spotted on a SELDI surface are typically analyzed using time-of-flight mass spectrometry (Issaq et al., 2002). A laser ionizes peptides from crystals of the sample/matrix mixture. The ions are accelerated through an electric potential and down a flight tube. A detector measures ions as they reach the end of the tube. The mass-to-charge ratio of each ion can be determined from the length of the tube, the kinetic energy given to ions by the electric field, and the time taken to travel the length of the tube. Taken together, the numerous studies suggest that SELDI-TOF MS methodology can used as a fast and robust approach to study the breast cancer proteome and enable the analysis of the correlations between proteomic expression patterns and breast cancer (Bashar et al., 2009).

**Breast cancer research with bioinformatics**

**Bioinformatics**

Bioinformatics is a synonym for computational molecular biology which is the use of computers to characterize the molecular components of living things (Ai et al., 2013). The term of bioinformatics relatively refers to the creation and advancement of algorithms, computational and statistical techniques, and theory to solve formal and practical problems posed by or inspired from the management and analysis of biological data (Sims et al., 2009). The term of bioinformatics is used to encompass almost all computer application biological sciences, but was originally coined in the mid 1980s for the analysis of biological sequence data.
Breast cancer is one of the many diseases that a cure has not yet found for, one of the keys to finding one is identifying the genetic mutations that causes the disease and this is said by scientist to be like looking for needles in a haystack, and after finding the needles or coding regions, they must find disease related sequences within them (Kihara et al., 2007). With bioinformatics this process becomes realistic and it is possible to search the haystack of 3 billion base pairs for anomalous genetic defects year of research show that genetic mutations, whether caused by exposure to environmental mutagens or inherited as defective gene copies, are inherent to the of cancer. About 5% of all incidents of breast cancer are hereditary. Women carrying defective copies of the BRCA1 and BRCA2 genes are at an increased risk perhaps as high as high as 85% of developing the disease (Lonning et al., 2005). In 1997, the national cancer institute designed the tumor gene index this was the first comprehensive index of genes involved in human cancer (National Cancer Institute). The tumor gene index is part of the NCIS larger cancer genome anatomy project (CGAP), which develops publicly available databases and technologies to support the search for cancer related genes. Through this program over 5500 genes have been identified in breast tissue and 5327 are active in cancer and more than 200 appear to be unique to breast tissue (Li et al., 2002). Recent advances in high-throughput methods had led to explosion of data and information on biological systems at different levels, from genomics, proteomics, metabolomics to whole-body physiology (Wang et al., 2011; Sims et al., 2006). Cancer research is an area where this wealth of information is exploited to create a systems perspective of cancer. Systems biology and bioinformatics have fundamentally changed cancer research, providing an integrated framework for identification and characterization of pathways that are critical to cancer, discovery of new targets for anti-cancer drugs and assessment of toxicity of existing and future therapeutics (Nagl et al., 2006). It can also help to explain the variations between individual responses to cancer treatments, thus, allowing the development of personalized and optimal treatment for each individual.

Application of High-Throughput Genomic Technology in breast cancer detection

The sequencing of the human genome together with that of model organisms has paved the way to the revolution in biology and medicine that we are experiencing today. In particular, the explosive growth in the number of new and powerful technologies within proteomics and functional genomics in combination with bioinformatics promises to accelerate the application of basic discoveries into daily clinical practice (figure 2) (Vijver et al., 2002; Ma et al., 2003; Celis et al., 1991; Baak et al., 2003).

CONCLUSION

There are several obstacles to be addressed before proteomics and bioinformatics reach an optimal yield and be beneficial for the patients (Ma et al., 2003). The requirement of fresh or frozen tissue samples to protect and obtain high quality genetic material to use in high-throughput techniques limits their wide spread use (Hu et al., 2005). The facilities for immediate freezing of tissue samples are not readily available in all hospitals. Establishment of high quality sample banks with databases containing information about all clinical and histopathological characteristics of the patients will help to collect uniform samples and data in clinical trials. Proteomics and bioinformatics should be recognized as complementary fields of investigation in cancer diagnosis and strengths and weaknesses of each individual technology should be balanced to obtain maximum benefit (Baak et al., 2003). From all the work done until today, it is clear that proteomics and bioinformatics have generated a considerable amount of data for breast cancer diagnosis. However, results obtained from previous studies must be validated, refined, and extended and the relevance of these data for clinical practice still has to be established. Integration of proteomics and bioinformatics technology into clinical trials and practice could lead to individualized patient care. Biomarkers used for early detection of breast cancer can be targets of new drugs individualizing
treatment and increasing success (Ornstein et al., 2006). Mass spectrometry techniques, such as MALDI-TOF and SELDI-TOF, allow for differentiation and classification of samples for a given disease state. In breast cancer, initial studies were aimed at distinguishing benign sera, plasma, tissue, nipple fluid, or ductal lavage from its cancerous counterparts (Leak et al., 2002). By generating proteomic profiles of a given sample, the computer can be trained to recognize the differences between cancer and benign states. This technology is akin to the UPC barcodes on items we buy at a store. The scanner can recognize the difference between a can of green beans and the box of cereal because it has already been taught what the codes for those items are. Another issue is the processing of the large amount of data obtained from the use of high-throughput techniques in both proteomics and bioinformatics. Statistical data analyses are tremendously challenging. The number of measured variables always outnumbers the number of samples evaluated. In order to minimize the problems in statistical analyses, acquired data must be filtered according to the goals of the researcher. Development of methods for statistical evaluation, normalization, and filtering of the data are all areas of active research (Baggerly et al., 2004). Although the development of specialized software is continuing, there is also a need for expert statisticians in this field. Collaborations should be established between researchers across disciplines for producing, storing, analyzing, and interpreting the data obtained from various experiments. Formation of centralized databases containing information on molecular characteristics of individual tumor types will help to reach already available data and may save time and resources during research. It is generally accepted that early detection of breast cancer has great impact on patient survival, emphasizing the importance of early diagnosis. In a widely recognized model of breast cancer development, tumor cells progress through chronological and well defined stages. However, the molecular basis of disease progression in breast cancer remains poorly understood.

In the future, methods used in proteomics and bioinformatics will be useful for the discovery of tumor specific marker genes and related proteins in breast cancer, but traditional methods will be applied in daily clinical practice.

High-throughput genomic techniques allow the simultaneous measurement of thousands of DNA sequences, mRNA transcripts, peptides or metabolites giving us a holistic view of the machinations of cellular processes (Rennstam et al., 2006). This technology has rapidly become a major tool for the study of breast cancer. Gene expression profiling has been applied to many areas of research from basic science to translational studies, with the potential to identify new targets for treatment, mechanisms of resistance and to improve on current tools for the analysis of prognosis (MacBeath et al., 2000; Ma et al., 2003). However, the sheer scale of the data generated along with the number of different protocols, platforms and analysis methods can make these studies difficult for clinicians to comprehend. Similarly, computational scientists and statisticians that may be called upon to analyze the data generated are often unaware of the processes involved in sample collection or the relevance and impact of genetics and pathological characteristics. There is a pressing need for better understanding of the challenges and limitations of microarray approaches, both in experimental design and data analysis (Vantand et al., 2002). Holistic, whole-genome approaches are still relatively new and critics have been quick to highlight non overlapping results from groups testing similar hypotheses. However, it is often subtle differences in the experimental design and technology that underpin the variation between these studies. Rather than indicating that the data are meaningless, this suggests that many findings are real but highly context dependent.

Figure 2. Technologies and resources available in functional genomics.
REFERENCES

Zeidan BA, Townsend PA. 2008. SELDI-TOF proteomic profiling of breast carcinomas identifies clinicopathologically relevant groups of patients similar to previously defined clusters from cDNA expression. Breast Cancer Res. 10(3):107.