JBUON 2017; 22(4): 838-843
ISSN: 1107-0625, online ISSN: 2241-6293 • www.jbuon.com
E-mail: editorial_office@jbuon.com

REVIEW ARTICLE

Microarray bioinformatics in cancer- a review

Ziqi Tao[1], Aimin Shi[2], Rui Li[3], Yiqiu Wang[4], Xin Wang[5], Jing Zhao[6]

[1]Department of Science and Education, Xuzhou Central Hospital, Xuzhou 221009, China; [2]School of Public Health of Nanjing Medical University, Nanjing 211166, China; [3]Central Laboratory, Xuzhou Central Hospital, Xuzhou 221009, China; [4]Department of Surgical Oncology, Xuzhou Central Hospital, Xuzhou 221009, Jiangsu, China; [5]Department of Thoracic Surgery, Xuzhou Central Hospital, Xuzhou 221009, Jiangsu, China; [6]Department of Science and Education, Xuzhou Central Hospital, Xuzhou 221009, China

Summary

Bioinformatics is one of the newest fields of biological research, and should be viewed broadly as the use of mathematical, statistical, and computational methods for the processing and analysis of biological data. Over the last decade, the rapid growth of information and technology in both 'genomics' and 'omics' eras has been overwhelming for the laboratory scientists to process experimental results. Traditional gene-by-gene approaches in research are insufficient to meet the growth and demand of biological research in understanding the true biology. The massive amounts of data generated by new technologies as genomic sequencing and microarray chips make the management of data and the integration of multiple platforms of high importance; this is then followed by data analysis and interpretation to achieve biological understanding and therapeutic progress. Global views of analyzing the magnitude of information are necessary and traditional approaches to lab work have steadily been changing towards a bioinformatics era. Research is moving from being restricted to a laboratory environment to working with computers in a 'virtual lab' environment. The present review article shall put light on this emerging field and its applicability towards cancer research.

Key words: *bioinformatics, cancer, microarrays*

Introduction

Within the human body, thousands of genes and their products (i.e., RNA and proteins) function in a complicated web and are orchestrated both temporally and spatially. Due to this complexity, the traditional gene-by-gene approach is not powerful enough to define a global view of cellular function. The microarray technology has been designed to measure the activity of gene expression, from the complete genome in a single experiment. Genetic information contained in DNA is consistent with cells of one individual, and a source of variation within and between species [1]. Gene expression, however, varies from tissue to tissue depending on the cell types present in the tissue and its condition (e.g. disease state), giving a source of variation within and between organisms. The ability to measure the expression of multiple genes provides the researcher with a method to elucidate the mechanisms behind this process.

Within a couple of years, gene expression microarray technology has developed from profiling a selection of genes on a membrane filter to all mRNA transcripts simultaneously (known as a 'transcriptome') on a solid surface [2,3]. Current microarrays may have up to tens of thousands of

Correspondence to: Ziqi Tao, MD. Department of Science and Education, Xuzhou Central Hospital, 199 Jiefang South Road, Xuzhou, Jiangsu, 221009, China.
Tel: +86 018952170668, E-mail: fi7780@163.com
Received: 30/01/2017; Accepted: 18/02/2017

unique DNA sequences spotted to it. The underlying principle of the microarray technology is base-pair hybridization. When using a gene expression microarray, one extracts mRNA from the sample of interest, under experimental procedures makes complimentary RNA from this, labels the cRNA with a fluorescent dye and hybridizes it to a glass slide with the spotted DNA sequences. Sequence specific hybridization ensures that the mRNA mostly binds to the DNA from which it is transcribed. Colour intensities for each gene can be quantified from a laser scanner using specialized soft ware for scanning microarrays, which can be used for statistical analysis.

Microarray studies with research questions often aim at increasing the knowledge and understanding of gene functions [4]. This is usually done by investigating genes whose expression levels are correlated with experimental conditions or important phenotypes. This could also involve the identification of biological pathways affected by the expression levels of a particular gene, but also in the aspects of drug targets and drug sensitivity in therapy development. Microarray research can also address questions relating to the phenotype of a particular disease. These studies aim at understanding discovering which biological processes are related to certain aspects or subtypes of disease, or identification of disease–specific molecular markers. Such information can be of great value in unravelling the complex biological mechanisms involved in a disease. A third direction of microarray research is driven by research questions that relate to the patient. Answers to such questions could potentially improve diagnosis and treatment of disease. Microarrays can be especially useful in prognosis, as future events which are not yet clinically detectable may be predicted through measurement of gene expression activity (such as metastasis in cancer) [5].

Microarray and gene expression in cancer

Human cancers are diverse in their tissue of origin as well as their individual biological and genetic histories [6]. These diversities are reflected by variations in gene expression programs among human cancers. Profiling cancer-specific gene expression programs thus may provide a new basis for the classification of human cancers. With the advent of microarray technology, it became possible to analyze and understand cancer-specific gene expression profiles on a global level instead of a gene-by-gene level. Microarray technology is at the heart of this article, with particular focus on gene expression profiling of breast cancers and brain tumors.

There are two main reasons for using microarray technology in oncological research. Firstly, to understand the biology related to particular cancer types or subtypes, their gene mutations and their aberrant (downstream) biological pathways. This is largely exploratory and results from a microarray experiment can be analyzed by using pathways and gene annotations such as Gene Ontology. Secondly, to classify human cancers according a particular variable: organ type or subtype, patient's prognosis, prediction of treatment response, or site of metastasis. This can be done in two ways: a) by looking purely at the biology associated with a variable or b) classifying tumors, where the biology of the genes involved is not so important as to have reliable genes that can predict the tested variable [7,8]. These analyses correlate clinical or biological data of cancers with their molecular profiles, in order to identify reliable classifiers.

Breast cancer gene expression profiles

Perou and Botstein were the first to use microarray technology to study the biology of human cancers by their intrinsic gene expression program [9]. They were able to distinguish several breast cancer subtypes based on gene expression profiles that correlated with previously identified histological protein expression patterns [10,11].'Intrinsic' gene signatures were defined that included genes whose differential expression levels could be related to specific histological features of the breast tumors. In a series of follow-up papers, Sorlie and colleagues further refined their intrinsic gene signatures to associate 5 molecular subtypes of breast cancer (Table 1) with survival data of the patients [12-14]. The 5 subtypes defined by these researchers reflect the inherent cell biology that defines the cluster division of the breast cancer subtypes:

1. 'Luminal A' breast cancers expressing estrogen receptors (ER): this subtype is associated with a favourable prognosis.
2. 'Luminal B' breast cancers expressing ER: this subtype has a less favourable prognosis, in particular for relapse of the disease.
3. 'ERBB2' breast cancers overexpressing ERBB2 and mostly ER negative: this subtype is known for a poor prognosis.
4. 'Basal-like' breast cancers expressing basal cytokeratins 5 and 17, integrin 4 and laminin, but lacking ER, progesterone receptors (PR) and ERBB2 expression: this subtype presents with a more aggressive clinical behaviour.

Table 1. Molecular subtypes of breast cancer

Sr. no.	*Type of cancer*	*Features*
1	Luminal A	Associated with a favourable prognosis
2	Luminal B	It has a less favourable prognosis, in particular for relapse of the disease
3	ERBB2	Overexpressing ERBB2 and mostly ER negative
4	Basal-like	Expressing basal cytokeratins 5 and 17, integrin 4 and laminin, but lacking ER, PR and ERBB2 expression
		This subtype presents with a more aggressive clinical behaviour
5	Normal-like	Expressing many genes known to be expressed by adipose tissue and other non-epithelial cell types
		These tumors also had strong expression of basal epithelial genes and low expression of luminal epithelial genes

5. 'Normal-like' breast cancers, expressing many genes known to be expressed by adipose tissue and other non-epithelial cell types.

These tumors also had strong expression of basal epithelial genes and low expression of luminal epithelial genes. Molecular profiles have also been associated with other known cancer genes such as TP53, BRCA1, and EGFR [15-17]. In such studies, the underlying mutation is presumed to be driving the segregation of the samples. Other prominent milestones in the application of gene expression microarrays to breast cancer involve the classification of breast cancers according clinical outcome of the patients. Van't Veer et al. [18] were the first to define a 70-gene expression signature that predicted the occurrence of metastasis in lymph node-negative breast cancer patients who had been diagnosed before 55 years of age. Similarly, a 21-gene signature was shown to predict metastasis in lymph node-negative patients with ER-positive breast cancer who had received adjuvant hormonal therapy [19]. A 76-gene signature also predicted metastasis in lymph node-negative breast cancer patients who had not received any adjuvant systemic therapy, irrespective of age and ER status [20]. Finally, a 44-gene signature has also predicted responsiveness of breast cancers to Tamoxifen therapy more accurate than the ER status of the tumors [21]. The ability of microarray technology to identify breast cancer patients who have a more or less favourable prognosis in developing metastasis could guide clinicians in avoiding adjuvant systemic therapy or, alternatively, to choose more aggressive therapeutic options. In this respect, it could also be useful to predict the site of metastasis, as recently was shown for breast cancers that metastasized to the bone [22].

Brain tumor gene expression profiles

Gene expression profiling of brain tumors has been guided primarily by their histological and pathological classification. Brain tumor gene expression profiles have been generated to investigate both the biology and the classification of brain tumors. Looking at biology, Pomeroy et al. [23] defined a gene signature that distinguished medulloblastomas from other histologically similar brain tumors and using this classification could predict their response to therapy. Importantly, this gene signature revealed that medulloblastomas are biologically distinct from primitive neuro-ectodermal tumors (PNETs), two subtypes of brain tumors that are often considered a single entity. The medulloblastoma gene expression profile implicated cerebellar granule cells as their cell of origin and revealed an unexpected involvement of the Sonic Hedgehog signaling pathway. Bredel et al. have also used gene expression profiling in the biological understanding of human gliomas by applying molecular network knowledge to the analysis of key functions and pathways associated with gliomagenesis [24]. Using a set of 50 human gliomas comprised of various histologies, they have seen via the transcriptional profiles of these tumors that integrin signalling pathway is most significant in the glioblastoma subtype, which is paradigmatic for its strong migratory and invasive behaviour. The MYC oncogene was also seen as a major network player in the biological process of gliomagenesis. More specifically, three novel MYC-interacting genes (*UBE2C, EMP1,* and *FBXW7*) with cancer-related functions were identified as network constituents differentially expressed in gliomas, as was *CD151* as a new component of a network that mediates glioblastoma

cell invasion [24]. Such biological approaches as Pomeroy et al. and Bredel et al. have extended the existing knowledge about the organizational pattern of gene expression in human gliomas, which can identify potential novel targets for future therapeutic development.

Understanding the biology is of utmost importance in brain tumors, however the classification based on its correlation with clinical parameters is also revealing important information. Classification based on histological subtype and genetic mutations as well as clinical parameters such as response to therapeutic drugs can potentially predict a patient's prognosis. French et al. have defined a 16-gene signature that predicted treatment response of oligodendrogliomas and a 103-gene signature for survival of the patients [25]. Interestingly, they were also able to define gene signatures that distinguished oligodendrogliomas with loss of 1p, loss of 19q, or loss of both chromosomal arms. Nutt et al. defined a 20-gene signature that appeared to better predict clinical outcome of patients with glioblastomas or high-grade oligodendrogliomas than classical histology [26]. This gene signature also allowed them to classify high-grade gliomas with non-classical histology. Together, these gene expression-profiling studies have shown that microarray technology may be an important tool in the molecular classification of gliomas. This technology can improve the classification of tumor subgroups as well as the correlation of patient's characteristics to make diagnoses and treatment decisions that are more informed. Perhaps most notable are the findings by French et al. that gene expression profiles not only reflect the biology and clinical behaviour of gliomas but also their underlying molecular basis. Each subtype of glioma is reflected in its pathological and histological characteristics; however, molecular profiles can further distinguish subtypes based on the underlying transcriptome. These molecular profiles are particularly important for brain tumor patients, as they are in urgent need for new treatment targets.

Recent applications in the area of oncology

In the oncology research field, microarrays are used to study diagnostics as well as the progression of disease and heterogeneity to treatment response. Cancer classifications have primarily been based on the morphological appearance of the tumor, but this has serious limitations, because histopathology is insufficient to predict disease progression and clinical outcome. To overcome this, many research groups have begun to apply microarray technology (Table 2) to identify particular pathological subgroups of disease that can predict patient survival and treatment outcomes. Disease classification not only for cancer has become an important component in downstream microarray analysis. The classification can be divided into two areas: class discovery and class prediction. Class discovery refers to redefining previously unrecognized tumor subtypes and class prediction refers to the assignment of particular tumor samples to the subclass based on a selection of significant genes [27]. Based on this classification, Beer et al. identified a set of genes that can predict survival in early-stage lung carcinoma [28]. This group also described and delineated a high-risk group that may benefit from adjuvant or supplementary therapy, whereby a pharmacological or immunological agent can be added to the treatment to increase or aid its effect or that of the antigenic response. More recently, advanced statistical tools have been applied to these class discovery and predictions in basic research. Multiple myeloma has been studied by numerous cancer research groups using microarray technologies. Claudio et al. confirmed the morphological homogeneity of multiple myeloma [29].

Results from microarray disease classification

Table 2. Microarray technology application in cancer

Sr. no	*Name of the technology*	*Applications*
1	OmniViz software SAM (Significant Analysis of Microarrays, developed by Stanford)	Identification of AML subgroups
2	PAM (Prediction Analysis for Microarrays)	Identification of class predictors to identify prognostic gene clusters for AML
3	Unique combinations of these techniques	Combinations could predict overall survival among patients within AML subgroups including that with a neutral karyotype

AML: acute myeloid leukemia

techniques also established that although multiple myeloma is morphologically homogeneous, there are underlying differences in individual tumor gene expression patterns that correlate with the heterogeneity of disease severity. Such underlying patterns include immunoglobin translocations and other structural genetic changes that both classify and impact patients'prognosis of cancer. Golub et al. [27] used sophisticated statistical methods to automatically classify new cases of acute leukaemia into those arising from lymphoid precursors (acute lymphoblastic leukemia) or from myeloid precursors (acute myeloid leukemia [AML]). More specifically and advanced in the area of AML, few studies in the recent past with very large microarray data sets were able to identify subgroups of patients with AML on the basis of molecular signatures and disease classification [30,31]. Using various advanced statistical techniques and visualization tools available today, such as the OmniViz software SAM (Significant Analysis of Microarrays, developed by Stanford) and PAM (Prediction Analysis for Microarrays), they identified 16 subgroups. Genes from these subgroups could be identified as class predictors to identify such prognostically important clusters. These subclasses of AML were featured by various chromosomal lesions such as translocations but also those with normal karyotypes. Some of these unique classes when coupled with extensive clinical data correlated with the prognosis of a poor treatment outcome and could predict overall survival among patients within AML subgroups including that with a neutral karyotype.

Conclusion

The merging of robotics, biotechnology, and computer sciences, as well as the completion of genome-sequencing efforts for several organisms, has resulted in groundbreaking changes in the way biomedical research is conducted. Biological researchers have traditionally examined functional genetic information to elucidate fundamental cellular processes and unravel the etiology of human disease. In today's post genome era, scientists are drowning in data trying to control high-throughput experimental platforms, and understand the millions of interrelations among proteins, small molecules, and phenotypes. It is now possible to manufacture high density arrays of specified DNA sequences that include every known gene of an organism on a single glass slide. Genomics, informatics, and automation will play increasingly important roles as discovery tools in the basic biological sciences, as well as in diagnostics and therapeutics within the clinical field. Many tools are continually being developed in the microarray field, in both technology and analysis, and the opportunity to apply these technologies to many different fields within bioscience is amazing. Scientists are becoming more aware of microarrays' potential to exploit their research, and, as knowledge increases, so do the awareness and possible solutions of the limitations microarrays may currently still hold.

Conflict of interests

The authors declare no confict of interests.

References

1. Masuda M, Yamada T. Signaling pathway profiling by reverse-phase protein array for personalized cancer medicine. Biochim Biophys Acta 2015;1854:651-657.
2. Ewis AA, Zhelev Z, Bakalova R et al. A history of microarrays in biomedicine. Expert Rev Mol Diagn 2005;5:315-328.
3. Peeters JK, Van der Spek PJ. Growing applications and advancements in microarray technology and analysis tools. Cell Biochem Biophys 2005;43:149-166.
4. Hu N , Wang C, Clifford RJ et al. Integrative genomics analysis of genes with biallelic loss and its relation to the expression of mRNA and micro-RNA in esophageal squamous cell carcinoma. BMC Genomics 2015;26:16:732.
5. Kaifi JT, Kunkel M, Das A et al. Circulating tumor cell isolation during resection of colorectal cancer lung and liver metastases: a prospective trial with different detection techniques. Cancer Biol Ther 2015;16:699-708.
6. Marquard AM, Birkbak NJ, Thomas CE et al. TumorTracer: a method to identify the tissue of origin from the somatic mutations of a tumor specimen. BMC Med Genomics 2015;8:58.
7. Wang Y, Klijn J, Zhang Y et al. Gene expression profiles and prognostic markers for primary breast cancer. Meth Mol Biol 2007;377:131-138.
8. Chang JC, Wooten EC, Tsimelzon A et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet 2003;362:362-369.
9. Perou CM, Sørlie T, Eisen MB et al. Molecular portraits of human breast tumours. Nature 2000;406:747-752.

10. Rakha EA, El-Sayed ME, Green AR, Paish EC, Lee AH, Ellis IO. Breast carcinoma with basal differentiation: a proposal for pathology definition based on basal cytokeratin expression. Histopathology 2007;50:434-438.
11. Rakha EA, Green AR, Lee AH, Robertson JF, Ellis IO. Prognostic markers in triple-negative breast cancer. Cancer 2007;109:25-32.
12. Sørlie T, Perou CM, Tibshirani R et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869-74.
13. Sorlie T, Tibshirani R, Parker J et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100:8418-8423.
14. Abramovitz M, Leyland-Jones B. A systems approach to clinical oncology: focus on breast cancer. Proteome Sci 2006;4:5.
15. Voorhoeve PM, le Sage C, Schrier M et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Adv Exp Med Biol 2007;604:17-46.
16. Hedenfalk I, Ringner M, Ben-Dor A et al. Molecular classification of familial non-BRCA1/BRCA2 breast cancer. Proc Natl Acad Sci U S A 2003;100:2532-2537.
17. Angulo B, Suarez-Gauthier A, Lopez-Rios F et al. Expression signatures in lung cancer reveal a profile for EGFR-mutant tumours and identify selective PIK-3CA overexpression by gene amplification. J Pathol 2008;214:347-356.
18. Ko JH, Ko EA, Gu W, Lim I, Bang H, Zhou T. Expression profiling of ion channel genes predicts clinical outcome in breast cancer. Mol Cancer 2013;12:106.
19. van't Veer LJ, Paik S, Hayes DF. Gene expression profiling of breast cancer: a new tumor marker. J Clin Oncol 2005: 23:1631-1615.
20. Smid M, Wang Y, Klijn JG et al. Genes associated with breast cancer metastatic to bone. J Clin Oncol 2006;24:2261-2267.
21. Jansen MP, Foekens JA, van Staveren IL et al. Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. J Clin Oncol 2005;23:732-740.
22. Feng YM, Gao G, Zhang F, Chen H, Wan YF, Li XQ. Identification of the differentially expressed genes between primary breast cancer and paired lymph node metastasis through combining mRNA differential display and gene microarray. Zhonghua Yi Xue Za Zhi 2006;86:2749-2755.
23. Pomeroy SL, Tamayo P, Gaasenbeek M et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature 2002;415:436-442.
24. Bredel M, Bredel C, Juric D et al. Functional network analysis reveals extended gliomagenesis pathway maps and three novel MYC interacting genes in human gliomas. Cancer Res 2005;65:8679-8689.
25. French PJ, Swagemakers SM, Nagel JH et al. Gene expression profiles associated with treatment response in oligodendrogliomas. Cancer Res 200;65:11335-11344.
26. Nutt CL, Mani DR, Betensky RA et al. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. Cancer Res 2003;63:1602-1607.
27. Golub TR, Slonim DK, Tamayo P et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531-537.
28. Beer DG, Kardia SL, Huang CC et al. Iannettoni MD, Orringer MB, Hanash S. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816-824.
29. Claudio JO, Masih-Khan E, Stewart AK. Insights from the gene expression pro Peeters and Van der Spek filing of multiple myeloma. Curr Hematol Rep 2004;3:67-73.
30. Bullinger L, Döhner K, Bair E et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukaemia. N Engl J Med 2004;350:1605-1616.
31. Valk PJ, Verhaak RG, Beijen MA et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350:1617-1628