JBUON 2016; 21(1): 61-69
ISSN: 1107-0625, online ISSN: 2241-6293 • www.jbuon.com
E-mail: editorial_office@jbuon.com

ORIGINAL ARTICLE

Integration of gene expression data and genetic variations involved in breast cancer

Xian-Xu Song[1], Xian-Dong Song[2], Mei Wang[1], Yan-Liang Han[1], Bo Liu[1], Wen Sun[3], Hong-Chun Jiang[4]

[1]Department of General Surgery, The Second Affiliated Hospital of Mudanjiang Medical University, Mudanjiang 157000, Heilongjiang, P.R. China; [2]Department of Bone Surgery , HongQi Hospital of Mudanjiang Medical University, Mudanjiang 157000, Heilongjiang, P.R. China; [3]Ji'nan Evidence Based Medicine Science-Technology Center, Jinan 250000, Shandong, P.R. China; [4]Department of Ophthalmology, HongQi Hospital of Mudanjiang Medical University, Mudanjiang 157000, Heilongjiang, P.R. China

Summary

Purpose: *The studies of transcriptome and genome involved in breast cancer are effectively promote the understanding of biological processes and the development of novel targeted therapies. Here we performed an integrated analysis of gene expression and genetic variation to disclose the molecular pathogenesis in breast cancer.*

Methods: *Gene expression profiles were applied to identify differential expression levels of genes between breast cancer and normal subjects. DNA sequencing data were extracted to analyze gene mutational information including number of mutations, number of mutated genes and their chromosomal distributions. Correlation analysis of gene mutations and differential expression was performed. Network-based approach was applied to compare the topological properties between the differentially expressed (DE) genes prone to mutation and those that were not. Two-tailed $p<0.05$ was considered as statistically significant.*

Results: *Statistical analysis showed that DE genes presented significantly positive correlation with the number of mutations ($p=1.267E^{-05}$), mutated genes ($p=0.00001$) and total genes in the genome ($p=2.489E^{-06}$). There were 81 genes, both DE and mutant, and they were distributed on chromosome 4 (N=51), chromosome 15 (N=29), and chromosome 18 (N=1). These 81 genes showed an increase in the number of genes interacting with in the protein-protein network.*

Conclusion: *Analysis of the integration of transcriptome and genome in breast cancer disclosed distinctive topology between the DE genes prone to mutation and those that were not.*

Key words: *breast cancer, differentially expressed genes, genetic mutations, single nucleotide polymorphism*

Introduction

Breast cancer is the most common malignancy among women and the second leading cause of death after lung cancer [1]. It was estimated that nearly 234,190 new cases would be diagnosed and 40,730 would die from breast cancer in the United States alone in 2015 [1]. At present, high-throughput technology has resulted in a paradigm shift in the way that researchers view breast cancer biology. As a powerful approach for expression profiling, it has been considered as a reliable tool for disclosing the molecular pathogenesis of breast cancer [2,3]. Expression profiling of breast cancer has classified this disease into various subtypes based on their gene expression pattern. For example, a recent report demonstrated that breast cancer with different estrogen receptor status could be effectively differentiated using 58 DE genes [4]. Microarray technology, based on its prognos-

Correspondence to: Dr. Hong-Chun Jiang, MM. Department of Ophthalmology, HongQi Hospital of Mudanjiang Medical University, No.1 Taiping Road, Mudanjiang 157000, Heilongjiang, P.R. China. Tel : +86 0453 8927315, E-mail : hongchunjiang@126.com
Received: 04/06/2015; Accepted: 10/08/2015

tic and predictive power, has been shown to be complementary to traditional clinicopathologic features [5].

Generally, the occurrence of cancer is suspected from the accumulation of inherited and somatic mutations in oncogenes and tumor suppressor genes. Several lines of evidence have shown that genetic events play an important role in breast cancer [6-9]. Approximately 10-15% of breast cancers are likely due to an inherited mutation, with about one third of these cases attributable to breast cancer susceptibility genes 1 (BRCA1) and breast cancer susceptibility gene 2 (BRCA2) [10,11]. Harmful mutations in either of these two cancer-susceptibility genes conferred a woman's lifetime likelihood of developing breast cancer between 60 and 85% [6,12]. Germline mutations in the TP53 tumor suppressor gene could cause the Li–Fraumeni syndrome associated breast cancer with a lifetime breast cancer risk of 49% by the age of 60 [13,14]. Meijers-Heijboer et al. [9] demonstrated that mutations in the cell cycle checkpoint kinase 2 (CHEK2) gene could result in a twofold increase of breast cancer risk in women and a tenfold increase of risk in men. Moreover, other cases of breast cancer are suspected to be attributable to additional cancer susceptibility genes with different penetrance, hormonal and environmental factors, and stochastic genetic events [6].

DNA sequencing (DNASeq) technology, also known as high-throughput next-generation sequencing technology for DNA, is rapidly developing in recent years. Compared to microarray analysis and previous sequencing technologies, DNASeq allows to sequence genome-wide genetic data more quickly and cheaply with less signal noises [15]. DNASeq technology has been applied for unprecedented discoveries in various types of cancer, and as such has revolutionized the study of genomics and molecular biology. To date, this high-throughput sequencing technology has allowed in-depth study of genomic changes in over 1000 breast cancers [16]. The Cancer Genome Atlas (TCGA) database (http://cancergenome.nih.gov/), a joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), is an integrative and coordinated effort to improve the understanding of the molecular basis of cancer through the application of large-scale genome sequencing technologies, including somatic mutations, germline susceptibility variants, and single nucleotide polymorphism (SNP) [17,18].

A malignant tumor is a highly heterogeneous disease whose intrinsic characteristics are apparent not only by gene expression, but by mutational and DNA copy number profiles as well. Recently, bioinformatics and computational biology provide new insights to explore the molecular pathogenesis and therapy of cancer. Previous studies have offered an amount of preliminary data of gene expression profiles in ArrayExpress Archive and DNASeq data in TCGA involved in breast cancer [3,19,20], which have provided some useful insights to derive the prognostic and predictive signatures. However, few studies are currently considered for integrated analysis of these data to offer potential for identifying the underlying biology by comparing gene differential expressions and genetic variations.

Network-based approach, as a powerful and informative tool, provides an effective way to analyze biological and communicated systems [21], and reveals interesting topological properties of interactomes with respect to gene essentiality. The topology of networks not only sheds light on the molecular mechanisms in cancer, but also provides insight into evolutionary aspects of the genes involved [22]. A recent study indicated that cancer genes, whose mutations lead to cancer, showed an evolutionary difference from genes not mutated in cancer and played central roles in interactomes [23]. Herein, we integrated gene expression profiles with the network-based approach to examine the connectivity of genes susceptible to mutations through interconnecting to a number of DE genes involved in breast cancer.

Thus, in the present study we attempted to examine the association of genomic alterations with transcription profiles involved in breast cancer, and to study the distinction between DE genes with mutations and DE genes without mutations based on the network analysis.

Methods

Data collection and identification of DE genes

In this work, 5 microarray expression profiles of breast cancer and normal controls were extracted from ArrayExpress database [24] (http://www.ebi.ac.uk/arrayexpress/) under access number of E-GEOD-29431 [25], E-GEOD-3744 [26], E-GEOD-42568 [27], E-GEOD-50567 [28] and E-GEOD-7904 [26]. In these 5 datasets, a total of 337 samples, including 276 cases and 61 controls, were collected. The characteristics of studies are shown in Table 1. Prior to analysis, the original expression information from all conditions was carried on data preprocessing. For each gene expression dataset, we

Table 1. Characteristics of the individual studies included in the study

Accession number	*Year*	*Sample size* *Total (Cases/Controls)*	*Platform*
E-GEOD-29431	2011	66 (54/12)	Affymetrix HG-U133Plus2
E-GEOD-3744	2006	47 (40/7)	Affymetrix HG-U133Plus2
E-GEOD-42568	2013	121 (104/17)	Affymetrix HG-U133Plus2
E-GEOD-50567	2011	41 (35/6)	Affymetrix HG-U133Plus2
E-GEOD-7904	2006	62 (43/19)	Affymetrix HG-U133Plus2

preformed quality control, background correction, normalization, probe set filtering, and perfect match and mismatch correction using robust multichip average (RMA) [29] and Micro Array Suite 5.0 (MAS 5.0) [30] algorithm in Bioconductor. Each probe was mapped to one gene, and the probe was discarded if it couldn't match any genes. The value averaged over probes was selected if the gene had multiple probes.

Breitling et al. [31] and Hong et al. [32] provided a powerful, rank-based meta-analysis tool to detect DE genes by integrating multiple microarray data. In each dataset, the genes were compared and ranked using the fold change (FC) method. Then the ranks were aggregated to an overall score for across studies, obtaining a ranked gene list. In a given study, pairwise FC (pFC) was computed, and the corresponding pFC ratios were ranked. This rank product statistic is defined as

$$RP_i = (\prod_{k=1}^{K} \prod_{r=1}^{R} pFC_{irk})^{\frac{1}{R}}$$

Where k (k=1,...,K) represents a microarray study, i (i=1,..., I) represents a gene in an individual study, r (r=1,...,R) represents the rank value of gene, pFCirk is the pFC value of gene i in study k under pairwise comparison r.

In present study, we provided this rank-based method to identify the DE genes between breast cancer and normal controls combining these multiple experiments. The up- and down-regulated DE genes were identified by assimilating a set of gene-specific rank tests. Genes with a percentage of false-positives (pfp) < 0.01 and $|\log_2 FC| > 2$ were considered as DE genes between breast cancer cases and normal controls.

Chromosomal distribution of DE genes

To examine the chromosomal distribution of the breast cancer-derived transcripts, we assigned the DE genes to chromosomes on the basis of Functional Annotation Chart module in the Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/tools.jsp). The significant enrichments were identified by expression analysis systematic explorer (EASE) score with the correction of false discovery rate (FDR). The threshold of EASE score was less than 0.01.

Analysis of DNASeq data

TCGA is a cancer genomic project to catalogue genetic mutations responsible for cancer, using genome sequencing and bioinformatics, in which multiple experimental data of more than 20 different types of human cancers are characterized, including DNA changes of breast cancer. To study the whole gene variations associated with breast cancer, we analyzed DNASeq data of breast cancer extracted from TCGA. In TCGA, whole-exome capture libraries were constructed and sequenced on Illumina HiSeq flowcells, and whole-genome sequencing was done with the Illumina HiSeq sequencer. Reads were aligned to the reference human genome build hg19 using an implementation of the Burrows-Wheeler Aligner, and a BAM file was produced for each tumor and normal sample using the Picard pipeline [33]. The Firehose pipeline was used to manage input and output files and submit analyses for execution [33]. Level 2 data were applied to detect gene mutations, including single-nucleotide polymorphism (SNP), base deletion and base insertion. Somatic mutations obtained by WUSM mutation calling model were selected for our study. A total of 17 valid batches including 1577 samples (776 tumors and 801 normal subjects) were extracted. Then the genetic variant information was obtained for further analysis.

Correlation analysis of DE genes and genetic variations

Analyzing the DNASeq data obtained from TCGA, the number of mutations (mut), the number of mutational genes (G_{mut}), the number of mutations in each mutational gene (mut/G_{mut}), the percent of G_{mut} in total gene (P_{Gmut}), and chromosomal distributions of genetic mutations were extracted for correlation analysis with DE genes. Also, the total number of genes (G_{total}) in each chromosome was obtained [34]. We implemented Spearman's correlation test [35] to evaluate the correlation of the DE genes and genetic mutational information with p value<0.01 considered as significant correlation.

Comparison analysis between DE genes with mutation and DE genes without mutation based on interaction network

Table 2. Chromosomal distributions of differentially expressed genes in expression profiling

Chromosome	*Count*	*%*	*p value*
1	147	10.7	0.0540
2	90	6.6	0.514
3	106	7.7	1.76E-05*
4	77	5.6	0.00261
5	79	5.8	0.0105
6	65	4.7	0.998
7	60	4.4	0.978
8	78	5.7	2.95E-04*
9	57	4.2	0.575
10	74	5.4	0.00229
11	92	6.7	0.0514
12	84	6.1	0.0104
13	36	2.6	0.0667
14	28	2.0	0.999
15	41	3.0	0.758
16	31	2.3	0.999
17	63	4.6	0.825
18	24	1.7	0.467
19	49	3.6	0.999
20	31	2.3	0.737
21	16	1.2	0.732
22	18	1.3	0.999
X	48	3.4	0.992

* $p < 0.01$

In order to examine the connectivity of DE genes with mutation, the gene-gene interaction network was constructed using search tool for the retrieval of interacting genes/proteins (STRING, http://string.embl.de/) database [36,37]. All the nodes with degree ≥1 were reserved in the network. All DE genes involved in the network were divided into two groups: DE genes with mutation and DE genes without mutation. The two-sample t-test [38] was conducted to compare the results between the two groups based on the degree of genes in the network. The statistical significance level was set at $p<0.05$. Two-sample t-test revealed statistically significant difference in the network degree level between two groups.

Results

Identification and chromosomal distributions of DE genes

Across 5 datasets associated with breast cancer, a total of 1464 DE genes were identified under the criterion of pfp < 0.01 and |log2FC| > 2. Among

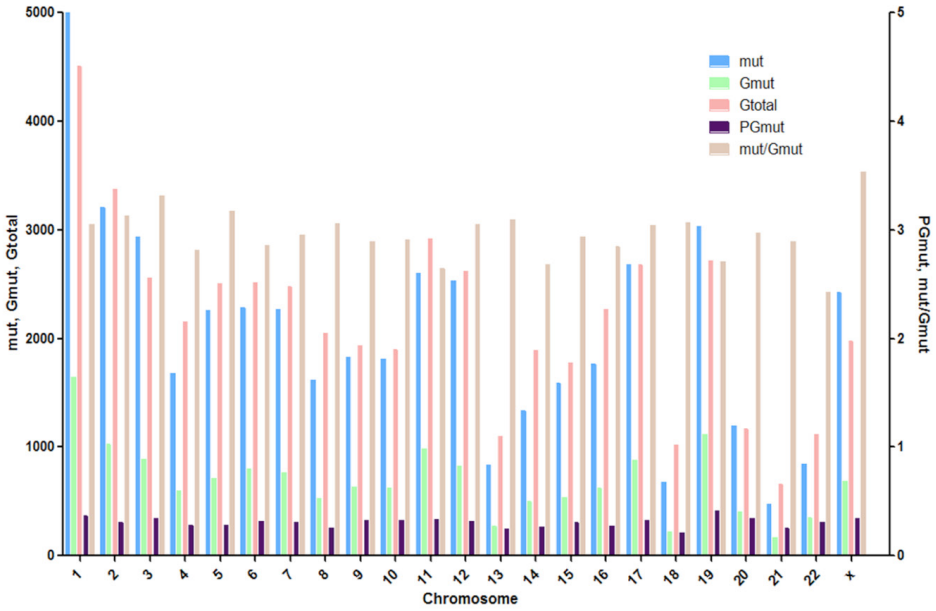

Figure 1. Chromosomal distributions of genetic mutational information shows that gene mutations exist universally in all chromosomes. Mutant genes accounted for approximately 30% of the total genes.
mut: the number of mutations, Gmut: the number of mutated genes, Gtotal: the total number of genes in chromosomes, PGmut: the percent of Gmut in total gene, mut/Gmut: the number of mutations in each mutated gene.

Table 3. The correlation of differentially expressed genes in expression profiling and genetic variations by Spearman's correlation test

		logFC	*mut*	G_{mut}	G_{total}	P_{Gmut}	mut/G_{mut}
DE genes	Correlation coefficients	0.135	0.780	0.782	0.812	0.425	0.318
	p value	0.538	1.276E-05	1.000E-05	2.489E-06	0.043	0.140

mut: the number of mutations, Gmut: the number of mutational genes, Gtotal: the total number of genes in chromosomes, PGmut: the percent of Gmut in total gene, mut/Gmut: the number of mutations in each mutational gene

these DE genes, 1038 were up-regulated and 426 were down-regulated.

On the basis of Functional Annotation Chart module in DAVID, a total of 1398 DE genes were assigned to each chromosomes (Table 2), except for the other 66 DE genes which were unknown genes. From Table 2, we discovered that these DE genes were significantly distributed on chromosome 3 (p=1.76E-05), chromosome 8 (p=2.95E-04), chromosome 10 (p=0.00229), and chromosome 4 (p=0.00261). In terms of DE genes count, chromosome 1 harbored the most DE genes (10.7%), followed by chromosome 11 (7.7%).

DNASeq data analysis

From TCGA database, 1577 samples including 776 breast cancer cases and 801 normal controls were extracted across 17 valid batches. Twenty-three chromosomes (22 autosomes and X-chromosome) were analyzed. The genetic mutational information of each chromosome including mut, Gmut, mut/G_{mut}, and P_{Gmut} is shown in Figure 1. On these chromosomes, a total of 46885 mut, 15780 G_{mut} and 49865 G_{total} were depicted. There was an average of three mutations in each mutant gene. On average, mutant genes accounted for approximately 30% of the total genes.

Correlation analysis of DE genes and genetic variations

Correlation analysis of the DE genes and genetic mutations were performed using Spearman's correlation test. The results showed that the number of DE genes had significantly positive correlation with mut (p=1.267E-05), G_{mut} (p=0.00001) and G_{total} (p=2.489E-06), but had no significant correlation with logFC (p=0.538), P_{Gmut} (p=0.043) and mut/G_{mut} (p=0.140) (Table 3).

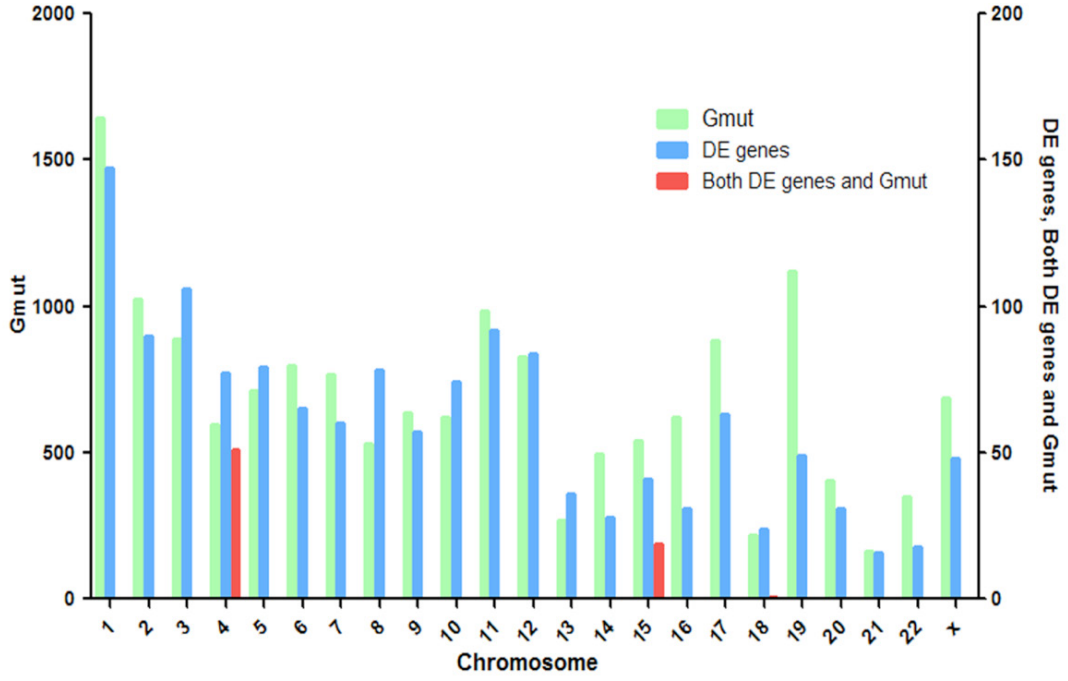

Figure 2. Genes which were both differentially expressed genes in expression profiling and mutated genes, mainly distributed on chromosomes 4, 15 and 18.
Gmut: the number of mutated genes. For other abbreviations, see text.

Comparison study between the transcription profile and genetic information was performed, and found that 81 genes were both DE genes and mutant genes. Unexpectedly, when screening their chromosomal distributions, these DE genes with mutation distributed mainly on chromosome 4 (51 DE genes), chromosome 15 (29 DE genes) and chromosome 18 (1 DE gene) in our study (Figure 2).

Interaction network analysis of DE genes with mutation and DE genes without mutation

We constructed the gene-gene interaction network to distinguish DE genes prone to mutations from those that are not. Comparison of connectivity between DE genes with mutations and DE genes without mutations were performed using two-sample t-test to observe the between-network differences. The number of interaction partners for each gene in the network was calculated. DE genes without mutations were shown to have 20.14 interaction partners, while DE genes prone to mutations had 27.17 interaction partners, on average. Statistics showed that the between-network differences were statistically significant ($p=0.472$, $t = -1.674$, $df = 1116$).

Discussion

In recent years, remarkable efforts on cancer research have been directed towards the detection of genes associated with oncogenesis. Previous studies have provided sufficient evidence that part of breast cancers are due to inherited mutation, and the differential expression of genes also plays important roles in cancer development [4,5,9,39]. However, the precise molecular basis involved in breast cancer is still unclear. In the present study, we attempted to investigate the association between expression profiles and genetic variations on genome-wide in breast cancer.

Recently, the identification of DE genes between tumor samples and controls has been popularized based on the expression profiling. A total of 1464 DE genes were yielded in this study. Enormous amount of DE genes between tumor patients and normal people have been presented in previous studies involving breast cancer [4,40]. Lee et al. [41] identified unique gene expression profiles of human ductal carcinoma *in situ* and invasive breast cancer. Gene expression profiling of breast cancer presented significantly better prognostication compared with currently used clinical parameters in predicting disease outcome [42, 43].

Evidence from epidemiology suggested that genetic factors might play an essential role in the development of breast cancer [44,45]. Predisposition to certain cancers have been linked to an ever-increasing number of mutations [46]. In this work, a total of 46885 mutations and 15780 mutant genes were detected. Apparently, an average of three mutations could be observed in each mutant gene, and approximately 30% of total genes were mutant genes. There was a high average mutation rate in breast cancer. To date, multiple mutations in different genes have been associated with the development of breast cancer. Activating mutation in PALB2 was proven as important cause of hereditary breast cancer [47]. Also in 9 new cancer genes including AKT2, ARID1B, CASP8, CDKN1B, MAP3K1, MAP3K13, NCOR1, SMARCD1 and TBX3 driver mutations were found among 100 tumors [48]. More recently, using large-scale genomic analysis, Lawrence et al. [49] identified 33 novel mutated genes by analyzing nearly 5,000 patient samples across 21 cancer types. To date, genetic testing has become an important diagnostic tool for risk assessment of breast cancer patients and their families.

In the present work, Spearman's correlation test showed that differential expressions had significantly correlation with gene mutations. Recently, previous studies also demonstrated correlation between differential expression and genetic variations [46,50]. A strong correlation was found between somatic mutation frequency and gene expression level in cancers by whole-genome and whole-exome data analysis [51,52]. Hedenfalk et al. [53] also proved that heritable mutations influenced the gene expression of cancer by analyzing expression profiles of breast cancers with BRCA1 and BRCA2 mutations. A study about pseudohypoxic pheochromocytomas and paragangliomas associated with SDHB, SDHD, and VHL mutations showed that the gene expression profiles depended on tumor location as well as on the underlying mutation [46]. Consistent with a previous study [53], our findings also illustrated the correlation between expression profile and genetic variation in genome-wide scale. However, when screening the chromosomal distributions of DE genes and mutant genes, mutant DE genes were mainly distributed on chromosome 4 and chromosome 15. Changes in chromosome 4 have been identified in several types of human cancer, such as familial pancreatic cancer [54] and cervical cancer [55]. A study of Shivapurkar et al. [56] inferred that there were multiple tumor suppressor genes, the inactivation of which was important in the pathogenesis of breast cancer, on both arms of chromosome 4. Genetic variations at a susceptibility region on chromosome 15 have been linked to lung cancer risk in many previous studies [57,58]. Few studies showed changes in chromosome 15 linking to breast cancer. Changes in chromosome 4 and chromosome 15 might play important roles in the development of breast cancer, and more attention should be focused on them.

Our study also showed that the connectivity of DE genes prone to mutations involved in breast cancer was statistically significantly higher compared with not prone genes.

It has been reported that a strong correlation exists between the age of a node and its degree for a growing network, as older nodes generally have more chances to receive links in network [59]. Our study showed the degree of DE genes prone to mutations involved in breast cancer was statistically significantly higher compared with not mutational genes, that is to say DE genes prone to mutations might be older genes showing higher connectivity in network, suggesting more important roles in the complex cellular processes. This finding has been the subject of several publications, showing that genes whose mutation led to cancer played central roles in the gene network [23,60].

In summary, we dealt with the comparison between expression profiles and genetic mutations involved in breast cancer and noticed that a number of DE genes and genetic mutations were displayed in breast cancer. There was a high mutation rate (approximately 30 %) in genes of breast cancer samples. Differential expression was significantly positively correlated with genetic variations. DE genes prone to mutation were mainly distributed on chromosome 4 and chromosome 15, which should draw close attention. These DE genes which were susceptible to mutation in breast cancer exhibited an increased frequency of interactions they participate in, showing a differentiation in evolutionary aspects of these two groups.

Acknowledgements

We would like to thank a team of persons from Ji'nan Evidence Based Medicine Science-Technology Center for their technical assistance and critical reading of the manuscript.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA: Cancer J Clinicians 2015;65:5-29.
2. Bernardes N, Ribeiro AS, Abreu S et al. High-throughput molecular profiling of a P-cadherin overexpressing breast cancer model reveals new targets for the anti-cancer bacterial protein azurin. Int J Biochem Cell Biol 2014;50:1-9.
3. Maltseva DV, Khaustova NA, Fedotov NN et al. High-throughput identification of reference genes for research and clinical RT-qPCR analysis of breast cancer samples. J Clin Bioinform 2013;3:13.
4. Gruvberger S, Ringner M, Chen Y et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. Cancer Res 2001;61:5979-5984.
5. Weigelt B, Baehner FL, Reis-Filho JS. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective study of the last decade. J Pathol 2010;220:263-280.
6. Wooster R, Weber BL. Breast and ovarian cancer. N Engl J Med 2003;348:2339-2347.
7. Wooster R, Bignell G, Lancaster J et al. Identification of the breast cancer susceptibility gene BRCA2. Nature 1995;378:789-792.
8. Wood LD, Parsons DW, Jones S et al. The genomic landscapes of human breast and colorectal cancers. Science 2007;318:1108-1113.
9. Meijers-Heijboer H, van den Ouweland A, Klijn J et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet 2002;31:55-59.
10. Rosenthal TC, Puck SM. Screening for genetic risk of breast cancer. Am Fam Physician 1999;59:99-104.
11. Newman B, Mu H, Butler LM, Millikan RC, Moorman PG, King MC. Frequency of breast cancer attributable to BRCA1 in a population-based series of American women. JAMA 1998;279:915-921.
12. Brose MS, Rebbeck TR, Calzone KA, Stopfer JE, Nathanson KL, Weber BL. Cancer risk estimates for BRCA1 mutation carriers identified in a risk evaluation program. J Natl Cancer Inst 2002;94:1365-1372.
13. Malkin D, Li FP, Strong LC et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science 1990;250:1233-1238.
14. Masciari S, Dillon DA, Rath M et al. Breast cancer phenotype in women with TP53 germline mutations: a Li-Fraumeni syndrome consortium effort. Breast Cancer Res Treat 2012;133:1125-1130.
15. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 2008;9:387-402.
16. Ma CX, Ellis MJ. The Cancer Genome Atlas: clinical applications for breast cancer. Oncology (Williston Park) 2013;27:1263-1269, 1274-1279.
17. Chang H, Fontenay GV, Han J et al. Morphometic analysis of TCGA glioblastoma multiforme. BMC Bioinformatics 2011;12:484.
18. Ying H, Lv J, Ying T et al. Gene-gene interaction network analysis of ovarian cancer using TCGA data. J Ovarian Res 2013;6:88.
19. Sun B, Zhang F, Wu SK et al. Gene expression profiling for breast cancer prognosis in Chinese populations. Breast J 2011;17:172-179.
20. Nehrt NL, Peterson TA, Park D, Kann MG. Domain landscapes of somatic mutations in cancer. BMC Genomics 2012;13(Suppl 4):S9.
21. Nibbe RK, Chowdhury SA, Koyuturk M, Ewing R, Chance MR. Protein-protein interaction networks and subnetworks in the biology of disease. Wiley Interdiscip Rev Syst Biol Med 2011;3:357-367.
22. Calvano SE, Xiao W, Richards DR et al. A network-based analysis of systemic inflammation in humans. Nature 2005;437:1032-1037.
23. Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. Bioinformatics 2006;22:2291-2297.
24. Rustici G, Kolesnikov N, Brandizi M et al. ArrayExpress update--trends in database growth and links to data analysis tools. Nucleic Acids Res 2013;41:D987-990.
25. Concha A, Blanco A, Cano C, Lopez FJ, Cuadros M. Identifying breast cancer biomarkers. Dataset, Array Express database 2011. (http://www.ebi.ac.uk/arrayexpress/).
26. Richardson AL, Wang ZC, De Nicolo A et al. X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell 2006;9:121-132.
27. Clarke C, Madden SF, Doolan P et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. Carcinogenesis 2013;34:2300-2308.
28. Lisowska KM, Dudaladava V, Jarzab M et al. BRCA1-related gene signature in breast cancer: the role of ER status and molecular type. Front Biosci (Elite Ed) 2011;3:125-136.
29. Irizarry RA, Hobbs B, Collin F et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249-264.
30. Hubbell E, Liu WM, Mei R. Robust estimators for expression analysis. Bioinformatics 2002;18:1585-1592.
31. Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. Febs Lett 2004;573:83-92.
32. Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 2006;22:2825-2827.
33. Dulak AM, Stojanov P, Peng S et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. Nat Genet 2013;45:478-486.
34. Harrow JL, Steward CA, Frankish A et al. The Verte-

brate Genome Annotation browser 10 years on. Nucleic Acids Res 2014;42:D771-779.

35. Artusi R, Verderio P, Marubini E. Bravais-Pearson and Spearman correlation coefficients: meaning, test of hypothesis and confidence interval. Int J Biol Markers 2002;17:148-151.
36. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res 2003;31:258-261.
37. Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res 2000;28:3442-3444.
38. O'Brien PC, Shampo MA. Statistics for clinicians. 6. Comparing two samples (the two-sample t test). Mayo Clin Proc 1981;56:393-394.
39. Ullah Shah A, Mahjabeen I, Kayani MA. Genetic polymorphisms in cell cycle regulatory genes CCND1 and CDK4 are associated with susceptibility in breast cancer. JBUON 2015;20:985-993.
40. Chen D, Yang H. Integrated analysis of differentially expressed genes in breast cancer pathogenesis. Oncol Lett 2015;9:2560-2566.
41. Lee S, Stewart S, Nagtegaal I et al. Differentially expressed genes regulating the progression of ductal carcinoma in situ to invasive breast cancer. Cancer Res 2012;72:4574-4586.
42. van 't Veer LJ, Dai H, van de Vijver MJ et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530-536.
43. Pawitan Y, Bjohle J, Amler L et al. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Res 2005;7:R953-964.
44. Chang-Claude J, Eby N, Becher H. The importance of genetic factors for development of breast cancer. Zentralbl Gynakol 1994;116:660-669.
45. McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. BMJ 2000;321:624-628.
46. Shankavaram U, Fliedner SM, Elkahloun AG et al. Genotype and tumor locus determine expression profile of pseudohypoxic pheochromocytomas and paragangliomas. Neoplasia 2013;15:435-447.
47. Antoniou AC, Casadei S, Heikkinen T et al. Breast-cancer risk in families with mutations in PALB2. N Engl J Med 2014;371:497-506.
48. Stephens PJ, Tarpey PS, Davies H et al. The landscape of cancer genes and mutational processes in breast cancer. Nature 2012;486:400-404.
49. Lawrence MS, Stojanov P, Mermel CH et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014;505:495-501.
50. da Rosa CE, Figueiredo MA, Lanes CF, Almeida DV, Marins LF. Genotype-dependent gene expression profile of the antioxidant defense system (ADS) in the liver of a GH-transgenic zebrafish model. Transgenic Res 2011;20:85-89.
51. Lawrence MS, Stojanov P, Polak P et al. Mutational heterogeneity in cancer and the search for new cancer. Nature 2013;499:214-218.
52. Nik-Zainal S, Alexandrov LB, Wedge DC et al. Mutational processes molding the genomes of 21 breast cancers. Cell 2012;149:979-993.
53. Hedenfalk I, Duggan D, Chen Y et al. Gene-expression profiles in hereditary breast cancer. N Engl J Med 2001;344:539-548.
54. Klein AP, de Andrade M, Hruban RH et al. Linkage analysis of chromosome 4 in families with familial pancreatic cancer. Cancer Biol Ther 2007;6:320-323.
55. Singh RK, Indra D, Mitra S et al. Deletions in chromosome 4 differentially associated with the development of cervical cancer: evidence of slit2 as a candidate tumor suppressor gene. Hum Genet 2007;122:71-81.
56. Shivapurkar N, Sood S, Wistuba II et al. Multiple regions of chromosome 4 demonstrating allelic losses in breast carcinomas. Cancer Res 1999;59:3576-3580.
57. Amos CI, Gorlov IP, Dong Q et al. Nicotinic acetylcholine receptor region on chromosome 15q25 and lung cancer risk among African Americans: a case-control study. J Natl Cancer Inst 2010;102:1199-1205.
58. Wu C, Hu Z, Yu D et al. Genetic variants on chromosome 15q25 associated with lung cancer risk in Chinese populations. Cancer Res 2009;69:5065-5072.
59. Fortunato S, Flammini A, Mencrer F. Scale-free network growth by ranking. Phys Rev Lett 2006;96:218701.
60. Wuchty S. Evolution and topology in the yeast protein interaction network. Genome Res 2004;14:1310-1314.