Pathol. Oncol. Res. (2015) 21:469–475
DOI 10.1007/s12253-014-9851-1

RESEARCH

Identification of Differently Expressed Genes with Specific SNP Loci for Breast Cancer by the Integration of SNP and Gene Expression Profiling Analyses

Pengfei Yuan · Dechun Liu · Miao Deng · Jiangbo Liu · Jianguang Wang · Like Zhang · Qipeng Liu · Ting Zhang · Yanbin Chen · Gaoyuan Jin

Received: 20 April 2014 / Accepted: 14 October 2014 / Published online: 19 November 2014

Abstract This study aims to explore the relationship between gene polymorphism and breast cancer, and to screen DEGs (differentially expressed genes) with SNPs (single nucleotide polymorphisms) related to breast cancer. The SNPs of 17 patients and the preprocessed SNP profiling GSE 32258 (38 cases of normal breast cells) were combined to identify their correlation with breast cancer using chi-square test. The gene expression profiling batch8_9 (38 cases of patients and 8 cases of normal tissue) was preprocessed with limma package, and the DEGs were filtered out. Then fisher's method was applied to integrate DEGs and SNPs associated with breast cancer. With NetBox software, TRED (Transcriptional Regulatory Element Database) and UCSC (University of California Santa Cruz) database, genes-associated network and transcriptional regulatory network were constructed using cytoscape software. Further, GO (Gene Ontology) and KEGG analyses were performed for genes in the networks by using siggenes. In total, 332 DEGs were identified. There were 160 breast cancer-related SNPs related to 106 genes of gene expression profiling (19 were significant DEGs). Finally, 11co-correlated DEGs were selected. In genes-associated network, 9 significant DEGs were correlated to 23 LINKER genes while, in transcriptional regulatory network, *E2F1* had regulatory relationships with 7 DEGs including *MTUS1*, *CD44*, *CCNB1* and *CCND2*. *KRAS* with SNP locus of rs1137282 was involved in 35 KEGG pathways. The genes of *MTUS1*, *CD44*, *CCNB1*, *CCND2* and *KRAS* with specific SNP loci may be used as biomarkers for diagnosis of breast cancer. Besides, *E2F1* was recognized as the transcription factor of 7 DEGs including *MTUS1*, *CD44*, *CCNB1* and *CCND2*.

Keywords Breast cancer · Single nucleotide polymorphisms · Gene expression profile · Fisher's combined probability test · Correlation analysis · Transcription factor

Highlights 1 Eleven significant DEGs were identified in breast cancer.
2 *MTUS1*, *CD44*, *CCNB1*, *CCND2* and *KRAS* with specific SNPs were vital genes for breast cancer.
3 *E2F1* was identified as the transcription factor of *MTUS1*, *CD44*, *CCNB1* and *CCND2*.
4 Significant DEGs enriched in multiple cancer signaling pathways.

P. Yuan · D. Liu (✉) · M. Deng · J. Liu · J. Wang · L. Zhang · Q. Liu · T. Zhang · Y. Chen · G. Jin
Department of Breast Surgery, The First Affiliated Hospital of Henan University Science and Technology, Jinghua Road No. 24, Jianxi District, Luoyang City 471003, China
e-mail: liudechun2008@126.com

Introduction

Breast cancer is known to originate from breast tissue, most commonly from the inner lining of milk ducts or the lobules that supply the ducts with milk [1]. It is reported that approximately 1.2 million women suffer from breast cancer, and 50,000 people die from the cancer every year [2]. Breast cancer is more prevalent in women that 232,670 new cases are estimated to appear in women in 2014, accounting for almost 99 % of the total new cases (235,030) [3]. Despite the major advances in the operative and non-operative managements, breast cancer metastasis remains a major clinical problem that affects large numbers of patients [4]. Besides, prognosis and survival rates for breast cancer vary greatly with cancer types, stages, treatments, and geographical locations of patient [5]. With the continuous improvement of SNP detection technology, the study on the relationship between gene polymorphism and genetic predisposition has become a hot topic. A Single Nucleotide Polymorphism (SNP) is a DNA sequence polymorphism caused by single nucleotide mutation from genomic level. It is well known that the genesis and

development of breast cancer is closely related to genetic mutations and deletions. Except the alteration of DNA sequence, other abnormalities such as SNP changes were also detected to associate with breast cancer [6]. Recent studies have found that multiple SNP loci such as SNP in the regulatory region and that in the coding region are likely to cause gene mutation or the change of protein structure, which are closely related to breast cancer [7]. Zardawi and his colleagues found that high expression of *Notch1* gene caused by SNP loci (rs1137282) was a sign in early stage of breast cancer [8]. In hereditary breast cancer, the risk for population with *BRCA-1* mutation of suffering from breast cancer was 36 % –87 %, and that for *BRCA-2* mutation carriers was 45 % –84 %, which were all found to be related to SNP loci (rs1137282) mutation [9]. All these above results indicate that it is essential to elucidate the relationship between gene polymorphism and breast cancer risk for the prevention and prognosis of breast cancer.

In order to explore the SNP related to prognosis of breast cancer, the SNP profiling and gene expression profiling of breast cancer were integrated to identify significantly differentially expressed genes (DEGs) contained SNP. The SNP profiling composing of 38 cases from normal breast cell, which was downloaded from GEO (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/) database, was preprocessed. Then integrated with the SNPs of 17 cases from patients with breast cancer, the correlation between SNP and breast cancer were detected using chi-square test [10]. In addition, the gene expression profiling of breast cancer batch8_9, were downloaded from the TCGA (the cancer genome atlas) database and were preprocessed with limma package in R language [11]. Fisher's method [12] was applied to analyze the significant SNP and genes according to their correlations with breast cancer. Furthermore, NetBox software [13] was recruited to obtain the interaction relationship between DEGs. Then combined with the information about transcription factors from the TRED (Transcriptional Regulatory Element Database) [14] and the transcription factor binding sites predicted by UCSC (University of California Santa Cruz) database [15], the transcriptional regulatory network was constructed using cytoscape software [16]. Finally, GO (Gene Ontology) [17] and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway [18] analyses were performed to explore the functions of DEGs and genes in the constructed networks, using siggenes in R language [19].

Materials and Methods

Data Resources

The SNP profiling of 17 patients (1 male and 16 female; right breast, 7; left breast, 10; 51.1±12.1 years) with breast cancer were purchases from commercial array platforms (Affymetrix chips, Agilent arrays). The specimens were collected from the Department of Breast Surgery, the First Affiliated Hospital of Henan Science and Technology University. Informed consent was retrieved from the patients. Ethical permission was granted by the local ethics committee. The subtypes of breast cancer of these samples were identified (Table 1), according to the classification method [20]. Based on the platform of GPL6801 (Affymetrix Genome-Wide Human SNP 6.0), the SNP profiling with accession number of GSE32258 which contains 38 cases of normal breast cells was downloaded from GEO database. The gene expression profiling of batch8_9 including 38 cases of breast cancer patients and 8 cases of normal tissue was downloaded from TCGA database. The data in TCGA were divided into four levels (level 1–4), and the data of level 3 was applied in this study.

Data Preprocessing of the SNP Profiling and the Correlation Analysis of SNP with Breast Cancer

The symbols of probes in SNP profiling of GSE32258 were transformed into the symbols of SNP according to their annotation information, then the SNP from 17 patients with breast cancer were combined to recognize the relationship between SNP and breast cancer with chi-square test [10]. The specific formulation is as follows:

$$K_1 = \sqrt{\frac{\sum_{i=1}^{3} S_i}{\sum_{i=1}^{3} R_i}} K_2 = 1 \Big/ K_1 \chi^2 = \sum_{i=1}^{3} \frac{(K_1 R_i + K_2 S_i)^2}{R_i + S_i}$$

Where, S and R represent the number of cases and control samples respectively, and i represents the SNP subtypes, whose value was set as 3 for the platform of GPL6801 could test three genotypes including AA, AB and B. The p-value in chi-square test was calculated with distr package in R language [21]. Besides, combined with the annotation information of SNP from dbSNP (database of SNP, http://www.ncbi.nlm.nih.gov/snp/) and genes from GEO database (http://www.ncbi.nlm.nih.gov/geo/), the information about the related genes with the SNPs were obtained.

Data Preprocessing of Gene Expression Profiling and Screening of DEGs

The original gene expression profiling of batch8_9 was normalized with median method of RMA (Robust Multiarray Average) algorithm of Affy package [22] to obtain the expression matrix. Limma package in R language [11] was employed to screen DEGs based on *t*-test method, and the

Table 1 The subtypes of breast cancer of 17 samples from patients

CA15-3	ER	PR	C-erbB-2	Ki-67	E-cad	Subtype
14.27	−	−	++	40 %(+)		HER-2 enriched
7.92	−	+	++	80 %(+)	+++	Luminal B
5.27	+++	++	++	20 %(+)		Luminal B
19.63	+	−	++	40 %(+)	+++	Luminal B
10.34	+++	−	++	40 %(+)	+++	Luminal B
9.73	+++	++	+++	20 %(+)	+++	Luminal B
13.14	−	−	++	30 %(+)	+++	HER-2 enriched
7	++	+++	+++	40 %(+)	+++	Luminal B
14.59	++	−	++	20 %(+)		Luminal B
8.69	−	−	++	20 %(+)		HER-2 enriched
7.25	++	−	−	20 %(+)		Luminal B
9.75	+++	+++	++	20 %(+)		Luminal B
11.2	++	++	++	30 %(+)	+++	Luminal B
9.17	+++	+	++	80 %(+)	++	Luminal B
7.93	+++	+++	++	30 %(+)	+++	Luminal B
13.75	+++	+++	++	30 %(+)	+++	Luminal B
7.12	+	−	−	10 %(−)		Luminal A

screening criteria were $|\log_2 FC \text{ (fold change)}| > 1.5$ and p-value < 0.05.

Fisher's Combined Probability Test

Fisher's combined probability test (fisher's method) [12] was applied to perform the integrated analyses of the significant disease-related SNPs and DEGs, thereafter to obtain the DEGs possessing breast cancer-related SNP loci. The integrated p-values were also obtained. Based on the screening criterion with p-value < 0.01, the significant DEGs and SNPs were integrated. Generally, one gene contains multiple significant SNP. In the present study, the SNP with maximum chi-square value (namely minimum p-value) was selected. The integrated chi-square value was calculated based on the formulation as follow:

$$\chi^2 = -2\sum_{i=1}^{k} \ln(p_i)$$

The statistic degrees comply with the chi-square distribution whose freedom was 2 k, where the k was 2.

Construction of DEGs-Associated Network and Transcriptional Regulatory Network

Comprising information from databases such as HPRD (Human Protein Reference Database), Reactome, NCI-Nature and Pathway Interaction Database (PID), as well as MSKCC (Memorial Sloan-Kettering Cancer Centre), the NetBox software is a potent tool to construct the interaction networks for human genes based on copy number alteration and sequence mutation data. Besides, it could also assemble altered genes in the constructed networks. It could identify LINKER genes under the pre-set p-value, connect all altered genes and then select network modules and calculate network modularity [13]. NetBox software was recruited to obtain the interaction relationships between DEGs and to identify statistically significant LINKER genes, so as to construct the DEGs-associated network, with the criterion that p-value < 0.05 and the shortest path threshold was 2, as well as that hypermutators was excluded. The transcription factors associated with breast cancer were identified from LINKER genes based on the information of cancer-related transcription factor deposited in TRED [14]. Then the relevant target genes were predicted according to the transcription factor binding sites provided by UCSC database [15]. Thereafter, the transcriptional regulatory network was constructed using cytoscape software (Cytoscape 3.X) [16], which could visualize the interaction relationships of genes network and transcriptional regulatory network, and the combine score >0.5 was select as the threshold for interaction relationships.

GO and KEGG Pathways Analyses

The GO annotation [17] and KEGG pathway [18] enrichment analyses were performed with siggenes in R language [19], for significant DEGs as well as the LINKER genes in the networks. The cut-off criteria were gene number ≥2 and p-value < 0.05.

Results

DEGs Selection

The gene expression profiling of batch8_9 involves of 12,042 genes from 46 samples. Calculated by the median method with RMA, the expression profile data were normalized. As shown in Fig. 1, the normalized median line was almost in a straight line, indicating a high level of normalization. With the screening criteria of $|\log_2 FC| > 1.5$ and p-value < 0.05, a total of 332 DEGs in breast cancer were identified, in which 146 were up-regulated.

Identification of SNPs Correlated with Breast Cancer

A total of 906,601 SNPs from 38 normal breast tissues were obtained after preprocessing of the SNP profiling GSE32258, then combined with 17 SNPs from patients, a total of 166 SNPs associated with breast cancer disease were screened

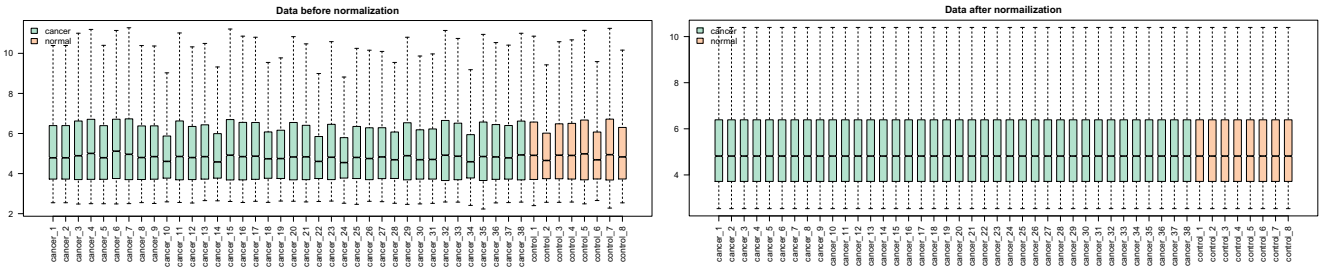

Fig. 1 The normalization of the gene expression profiling: box plots displaying the intensity log ratio distribution before and after nornolization procedures. **a**, Before normalization; **b**, After normalization. The *horizontal axis* indicates the name of the sample, and the ordinate represents the value of the expression. The *black lines* in each box-plot represent the median of data

based on chi-square test. According to the annotation of SNP from dbSNP, there were 160 SNP correlated with 106 genes (19 were significant DEGs) in the gene expression profiling.

Screening of DEGs with Breast Cancer-Related SNP Loci

The SNP with minimum *p*-value which represents a gene was integrated with DEG selected in the gene expression profiling. Tested by Fisher's method with *p*-value <0.01, 11 significant DEGs co-correlated with breast cancer were screened out (Table 2).

The DEGs-Associated Network and the Transcriptional Regulatory Network

In the DEGs-associated network constructed using NetBox software and cytoscape, 9 significant DEGs with breast cancer-related SNP were found to be associated with 23 genes in the database comprised in NetBox. In the 23 LINKER genes, there were 7 significant DEGs (in which *CDC20*, *CCNA2*, *CDC2* and *HERC5* were up-regulated and *STAT3*, *PIK3R1* and *CDKN1A* were down-regulated, Fig. 2a) identified in gene expression profiling. Based on the information deposited in TRED, other LINKER genes such as *E2F1*, *BRCA1* and *TP53* belonged to cancer-related transcription factor family. Moreover, combined with the predicted transcription factor binding sites according to UCSC database,, *E2F1* was found to regulate the transcription of 7 genes including *CD44*, *CCNB1*, *CCND2*, *ALCAM*, *IGF1R*, *MTUS1*and *PDGFRA*, which all belonged to significant DEGs with breast cancer-related SNP and were all down-regulated (Fig. 2b).

GO Annotation and KEGG Pathway Analyses

The GO annotation and KEGG pathway analyses showed that a total of 32 associated genes in the networks were significantly enriched in multiple cancer signaling pathways. Pathway analysis for 11 key DEGs showed that *KRAS* (v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog) was involved in 35 KEGG pathways (Table 3). Moreover, the SNP locus of rs1137282 was located in the *KRAS* gene, which was significantly down-regulated.

Table 2 Eleven significant DEGs (differentially expressed genes) co-correlated with breast cancer based on fisher's method

Gene	log(FC)	gene-*p*-value	SNP	SNP-*p*-value	Chisq	*p*-value
MTUS1	−1.13E+00	2.49E-06	rs3739408	0.01041984	34.9333	4.79E-07
ALCAM	−1.20551417	0.011683923	rs16851279	0.0008378	23.0685	0.000122698
MYB	−1.35E+00	0.000497047	rs3752383	0.000407968	30.8223	3.33E-06
CCNA2	2.17E+00	4.75E-09	rs769242	0.28615826	40.8341	2.91E-08
KRAS	1.12E+00	3.28E-06	rs1137282	0.003454342	36.5929	2.18E-07
PDGFRA	−1.420601453	2.22E-05	rs2228230	0.01579237	29.728	5.56E-06
IGF1R	−1.23E+00	0.000469521	rs7174918	0.025011811	22.7044	0.00014506
BLM	1.49E+00	1.56E-07	rs3815003	0.025011811	38.7212	7.95E-08
CD44	−1.457287965	2.34E-08	rs1467558	0.000407968	50.7538	2.51E-10
CCNB1	3.0165968	3.90E-13	rs1128761	5.37E-06	81.4139	1.11E-16
CCND2	−1.147007928	0.018256156	rs3217805	0.000407968	23.6151	9.54E-05

gene-*p*-value: the *p*-value for the significant DEGs selection; SNP-*p*-value, the *p*-value for the significant SNP screening; *p*-value: the integrated *p*-value according to fisher's method; Chisq: the value of chi-square test

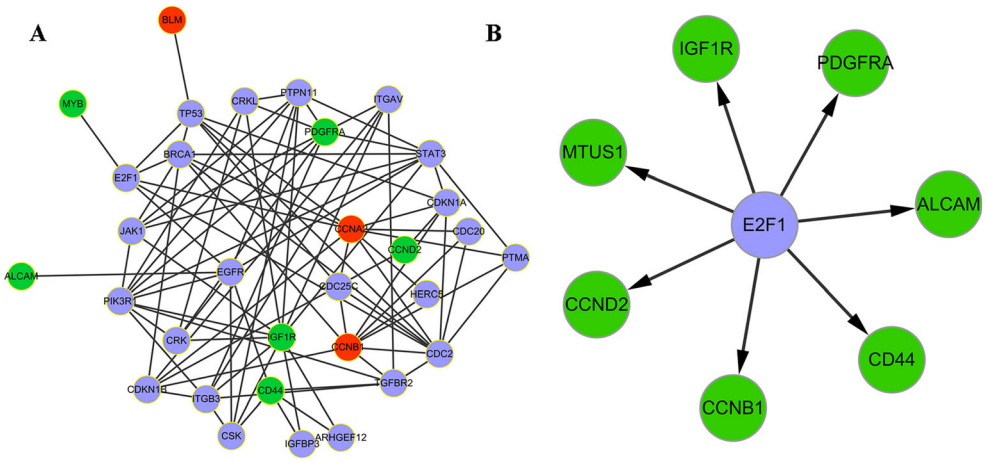

Fig. 2 DEGs (differently expressed genes)-associated network and transcriptional regulatory network. **a**, DEGs-associated network; **b**, Transcriptional regulatory network of *E2F1*. The *red circles* represent up-regulated genes, the greens represent down-regulated genes, and purples represent LINKER genes. The *arrows* represent the regulatory relationships

Discussion

Breast cancer, like other cancers, arises from an interaction between the environment and a defective gene [4]. The incidence of breast cancer is increasing rapidly in most Asian countries [23]. It is inspiring that the advance in SNP detection technology contributes to develop effective approaches for breast cancer prevention and treatment. In our study, a total of 332 DEGs in breast cancer were identified, in which 146 were up-regulated. After preprocessing of SNP microarray data and combining the SNPs from patients, totally 166 SNPs associated with breast cancer were screened and 160 SNPs were related to 106 genes (19 were significant DEGs) identified in gene expression profiling. Then after integrating *p*-values of SNPs and DEGs using Fisher's method, 11 significant DEGs co-correlated with breast cancer were screened out. In the constructed transcriptional regulatory network, *E2F1* was identified as the transcription factor to regulate the expression of 7 DEGs including *MTUS1*, *CD44*, *CCNB1* and *CCND2*, which were all with breast cancer-related SNPs and down-regulated. Additionally, 32 associated genes in the networks were significantly enriched in cancer signaling pathways, especially *KRAS*, possessing SNP locus of rs1137282, was involved in 35 KEGG pathways.

MTUS1 is a tumor suppressor gene, which was down-regulated in many cancers, such as colon cancer (CRC) [24],

Table 3 Top 5 enriched GO (Gene Ontology) terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways analyses for 32 associated genes in the networks

ID	Term	Description	Count	*p*-value
Biological Process	GO:0009987	cellular process	28	6.30E-07
	GO:0008150	biological process	28	1.37E-05
	GO:0050794	regulation of cellular process	27	5.43E-11
	GO:0050789	regulation of biological process	27	1.96E-10
	GO:0065007	biological regulation	27	8.07E-10
Cellular Component	GO:0044464	cell part	28	0.002370481
	GO:0005623	cell	28	0.002370481
	GO:0044424	intracellular part	27	0.000570814
	GO:0005622	intracellular	27	0.000571834
	GO:0005737	cytoplasm	24	0.000382581
Molecular Function	GO:0003674	Molecular function	27	0.002917227
	GO:0003824	catalytic activity	13	0.004704192
	GO:0043168	anion binding	9	0.002917227
	GO:0035639	purine ribonucleoside triphosphate binding	7	0.002917227
	GO:0032559	adenyl ribonucleotide binding	7	0.001823697
KEGG	hsa05200	Pathways in cancer	11	3.65E-07
	hsa04110	Cell cycle	8	3.65E-07
	hsa04510	Focal adhesion	8	8.54E-06
	hsa04810	Regulation of actin cytoskeleton	8	9.74E-06
	hsa05220	Chronic myeloid leukemia	7	3.65E-07

pancreas cancer [25] and ovary cancer [26]. Therefore, it is considered as the candidate biomarker for these cancers. Recently, a research demonstrated that ATIP3, a novel microtubule-associated protein product of *MTUS1*, was supposed as a biomarker for therapies of aggressive breast cancer [27]. In the present study, *MTUS1* was found to have the cancer-related SNP (rs3739408) and other relevant SNPs, suggesting that the SNP loci located in this gene may be closely related to breast cancer.

The protein encoded by *CD44* is a cell-surface glycoprotein participating in various cellular functions such as cell adhesion and migration, as well as tumor metastasis [28]. It is also a receptor for hyaluronic acid (HA) and can interact with other ligands including collagens and matrix metalloproteinases (MMPs). . The gene *CD44* is convinced to be a main biomarker of early gastric cancer recurrence [29]. Moreover, the study in CRC indicated that knockdown of *CD44* strongly prevented clonal formation and inhibited tumorigenicity in xenograft model, providing the direct evidence that *CD44* is a robust marker for CRC initiation [30]. Besides, *CD44* is extensively predicted as a marker for cancer stem cells in multiple solid tumors [31]. In the researches about breast cancer, it has been illustrated that *CD44* involved in steps of adhesion to the extracellular matrix and motility. What's more, for *CD44v6* (one *CD44* variant) was an essential component for malignant transformation of the breast epithelium, *CD44* was suggested as a marker for breast cancer prognosis [32]. In this study, the key SNP (rs1467558) associated with breast cancer was located in *CD44* gene, implying that *CD44* might also be a biomarker for breast cancer detection.

The CyclinB1 protein encoded by *CCNB1* (Cyclin-B1 Human Recombinant) in breast cancer can affect cell cycle regulation, chromosome replication and centrosome separation together with *CDK1* (*Cyclin-Dependent Kinase 1*). Previous studies have also found a number of SNPs associated with breast cancer located at *CCNB1*, such as rs2049269 [33] and rs78540526 [34], suggesting that it might be the SNPs located in this gene that caused the occurrence of breast cancer.

The *CCND2* (*Cyclin D2*) gene is a direct target gene of miR-206 in breast cancer cells. The survival prognosis analysis of patients with ovarian cancer showed that polymorphism of *CCND2* is an important factor affecting the survival time of patients, and a plurality of polymorphic SNP loci were identified in *CCND2* gene [35]. The Genome Wide Association Studies (GWAS) analysis of breast cancer verified that the risk SNP locus (rs3217805) for breast cancer was located on *CCND2* [36], which was consisted with our results.

The transcription factor *E2F1* is a member of *E2F* family, which can regulate expression of numerous cellular genes and play a pivotal role in the control of action of tumor suppressor proteins [37]. In breast cancer, it was demonstrated that *E2F1* transcription factor and *PI3K* signal transduction both participated in the regulation of HA-CD44-dependent expression of SVV (survivin) [38], suggesting that *E2F1* might play a potential role in regulation of *CD44*. With regards to the relationships of other DEGs with *E2F1*, it has been uncovered that in hypoxic cells, HIF -2α, a well-characterized hypoxia-inducible transcription factor vital for tumor angiogenesis, could promote cell-cycle progression by simultaneously augmenting the expression of cell-cycle genes *CCND2* and *E2F1* [39]. However, the evidence that *E2F1* could directly regulate the expression of *CCND2* has not been confirmed, neither has that been verified for the direct regulation of *CCNB1* and *MTUS1*. Nevertheless, out results revealed that these DEGs were all regulated by the transcription factor *E2F1*, providing a clue that *E2F1* might exert its function on the DEGs through the mediation of other genes.

Furthermore, our results showed that the gene *KRAS* which was significantly down-regulated in breast cancer, involved in 35 KEGG pathways, with SNP loci of rs1137282. *KRAS* plays a critical role in normal tissue signaling, and the mutation of *KRAS* is an essential step in the development of many cancers. Previous studies found that the typical SNP locus of rs1137282 was located in *KRAS* in oral carcinom while, this locus was not detected in normal buccal cells [40]. Additionally, *KRAS* mutation is predictive for occurrence of CRC, meanwhile, the SNP locus of rs1137282 is closely related to treatment and prognosis of this cancer [41]. All these indicated that rs1137282 in *KRAS* might be the crucial SNP locus associated with breast cancer. Therefore, we speculate that *KRAS* with the SNP locus of rs1137282 can be used as a biomarker for diagnosis of breast cancer.

In conclusion, significant DEGs identified in breast cancer including *MTUS1*, *CD44*, *CCNB1* and *CCND2*, as well as *KRAS*, all possessed specific SNP loci and might be used as biomarkers for breast cancer diagnosis and treatment. Furthermore, the transcription factor *E2F1* might play pivotal roles in the progression of breast cancer through the regulation of these DEGs. However, the accurate transcriptional regulatory relationships need to be further convinced.

Acknowledgments This study was supported by Science and technology innovation projects in henan province department of education (NO. 4206).

Conflict of Interest The authors have declared that no competing interests exist.

References

1. Sariego J (2010) Breast cancer in the young patient. Am Surg 76(12): 1397–1400
2. Boyle P, Levin B (2008) World Cancer Report 2008. IARC Press, International Agency for Research on Cancer, Lyon
3. Siegel R, Ma J, Zou Z, Jemal A (2014) Cancer statistics, 2014. CA Cancer J Clin 64(1):9–29

4. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF (2003) Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci 100(7):3983
5. Sant M, Allemani C, Capocaccia R, Hakulinen T, Aareleid T, Coebergh JW, Coleman MP, Grosclaude P, Martinez C, Bell J (2003) Stage at diagnosis is a key explanation of differences in breast cancer survival across Europe. Int J Cancer 106(3):416–422
6. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet 42(6):504–507
7. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447(7148):1087–1093
8. Zardawi SJ, Zardawi I, McNeil CM, Millar EK, McLeod D, Morey AL, Crea P, Murphy NC, Pinese M, Lopez-Knowles E (2010) High Notch1 protein expression is an early event in breast cancer development and is associated with the HER-2 molecular subtype. Histopathology 56(3):286–296
9. Veronesi A, de Giacomi C, Magri MD, Lombardi D, Zanetti M, Scuderi C, Dolcetti R, Viel A, Crivellari D, Bidoli E (2005) Familial breast cancer: characteristics and outcome of BRCA 1–2 positive and negative cases. BMC Cancer 5(1):70
10. Satorra A, Bentler PM (2001) A scaled difference chi-square test statistic for moment structure analysis. Psychometrika 66(4):507–514
11. Sanges R, Cordero F, Calogero RA (2007) oneChannelGUI: a graphical interface to Bioconductor tools, designed for life scientists who are not familiar with R language. Bioinformatics 23(24):3406–3408
12. Whitlock M (2005) Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. J Evol Biol 18(5):1368–1373
13. Kim H-Y, Byun M-J, Kim H (2011) A replication study of genome-wide CNV association for hepatic biomarkers identifies nine genes associated with liver function. Biochem Mol Biol Rep 44(9): 578–583
14. Yang J, Chen L, Wang L, Zhang W, Liu T, Jin Q (2007) TrED: the Trichophyton rubrum expression database. BMC Genomics 8(1):250
15. Wang Y, Wang DD (2013) University of California, Santa Cruz
16. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504
17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT (2000) Gene Ontology: tool for the unification of biology. Nat Genet 25(1):25–29
18. Nakaya A, Katayama T, Itoh M, Hiranuka K, Kawashima S, Moriya Y, Okuda S, Tanaka M, Tokimatsu T, Yamanishi Y (2013) KEGG OC: a large-scale automatic construction of taxonomy-based ortholog clusters. Nucleic Acids Res 41(D1):D353–D357
19. Schwender H, Krause A, Ickstadt K (2006) Identifying interesting genes with siggenes. The Newsletter of the R Project Volume 6/5, December 2006 34:45
20. Voduc KD, Cheang MC, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H (2010) Breast cancer subtypes and the risk of local and regional relapse. J Clin Oncol 28(10):1684–1691
21. Camphausen F, Kohl M, Ruckdeschel P, Stabla T, Ruckdeschel MP (2007) The distr Package
22. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4(2):249–264
23. Tavassoli FA, Devilee P (2003) Pathology and genetics of tumours of the breast and female genital organs, vol 4. World Health Organization
24. Zuern C, Heimrich J, Kaufmann R, Richter KK, Settmacher U, Wanner C, Galle J, Seibold S (2010) Down-regulation of MTUS1 in human colon tumors. Oncol Rep 23(1):183–189
25. Frank B, Bermejo JL, Hemminki K, Sutter C, Wappenschmidt B, Meindl A, Kiechle-Bahat M, Bugert P, Schmutzler RK, Bartram CR (2007) Copy number variant in the candidate tumor suppressor gene MTUS1 and familial breast cancer risk. Carcinogenesis 28(7):1442–1445
26. Ye H, Pungpravat N, Huang B-L, Muzio LL, Mariggiò MA, Chen Z, Wong DT, Zhou X (2007) Genomic assessments of the frequent loss of heterozygosity region on 8p21. 3 ~ p22 in head and neck squamous cell carcinoma. Cancer Genet Cytogenet 176(2):100–106
27. Rodrigues-Ferreira S, Di Tommaso A, Dimitrov A, Cazaubon S, Gruel N, Colasson H, Nicolas A, Chaverot N, Molinie V, Reyal F (2009) 8p22 MTUS1 gene product ATIP3 is a novel anti-mitotic protein underexpressed in invasive breast carcinoma of poor prognosis. PLoS One 4(10):e7239
28. Bajorath J (2000) Molecular organization, structural features, and ligand binding characteristics of CD44, a highly variable cell surface glycoprotein with multiple functions. Proteins Struct Funct Bioinforma 39(2):103–111
29. Mayer B, Jauch K, Schildberg F, Funke I, Günthert U, Figdor C, Johnson J (1993) De-novo expression of CD44 and survival in gastric cancer. Lancet 342(8878):1019–1022
30. Du L, Wang H, He L, Zhang J, Ni B, Wang X, Jin H, Cahuzac N, Mehrpour M, Lu Y (2008) CD44 is of functional importance for colorectal cancer stem cells. Clin Cancer Res 14(21):6751–6760
31. Alvero AB, Chen R, Fu H-H, Montagna M, Schwartz PE, Rutherford T, Silasi D-A, Steffensen KD, Waldstrom M, Visintin I (2009) Molecular phenotyping of human ovarian cancer stem cells unravel the mechanisms for repair and chemo-resistance. Cell Cycle (Georgetown, Tex) 8(1):158
32. Herrera-Gayol A, Jothy S (1999) Adhesion proteins in the biology of breast cancer: contribution of CD44. Exp Mol Pathol 66(2):149–156
33. Li Y, Chen Y-L, Xie Y-T, Zheng L-Y, Han J-Y, Wang H, Tian X-X, Fang W-G (2013) Association study of germline variants in CCNB1 and CDK1 with breast cancer susceptibility, progression, and survival among Chinese Han women. PLoS One 8(12):e84489
34. French JD, Ghoussaini M, Edwards SL, Meyer KB, Michailidou K, Ahmed S, Khan S, Maranian MJ, O'Reilly M, Hillman KM (2013) Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. Am J Hum Genet 92(4):489–503
35. Song H, Hogdall E, Ramus SJ, DiCioccio RA, Hogdall C, Quaye L, McGuire V, Whittemore AS, Shah M, Greenberg D (2008) Effects of common germ-line genetic variation in cell cycle genes on ovarian cancer survival. Clin Cancer Res 14(4):1090–1095
36. Driver KE, Song H, Lesueur F, Ahmed S, Barbosa-Morais NL, Tyrer JP, Ponder BA, Easton DF, Pharoah PD, Dunning AM (2008) Association of single-nucleotide polymorphisms in the cell cycle genes with breast cancer in the British population. Carcinogenesis 29(2):333–341
37. Johnson DG (2000) The paradox of E2F1: oncogene and tumor suppressor gene. Mol Carcinog 27(3):151–157
38. Abdraboh ME, Gaur RL, Hollenbach AD, Sandquist D, Raj MH, Ouhtit A (2011) Survivin is a novel target of CD44-promoted breast tumor invasion. Am J Pathol 179(2):555–563
39. Huang L (2008) Carrot and stick: HIF-α engages c-Myc in hypoxic adaptation. Cell Death Differ 15(4):672–677
40. Wang WY, Chien YC, Wong YK, Lin YL, Lin JC (2012) Effects of KRAS mutation and polymorphism on the risk and prognosis of oral squamous cell carcinoma. Head Neck 34(5):663–666
41. Yen L-C, Yeh Y-S, Chen C-W, Wang H-M, Tsai H-L, Lu C-Y, Chang Y-T, Chu K-S, Lin S-R, Wang J-Y (2009) Detection of KRAS oncogene in peripheral blood as a predictor of the response to cetuximab plus chemotherapy in patients with metastatic colorectal cancer. Clin Cancer Res 15(13):4508–4513