

Identification of key genes involved in HER2-positive breast cancer

B.-L. ZHONG¹, L.-J. BIAN², G.-M. WANG¹, Y.-F. ZHOU¹, Y.-Y. CHEN¹, F. PENG¹

¹Department of Pathology, Second People's Hospital of Guangdong Province/Guangdong Provincial Emergency Hospital, Guangzhou, P.R. China

²Department of Pathology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, P.R. China

Abstract. – OBJECTIVE: As an invasive cancer, breast cancer is the most common tumour in women and is with high mortality. To study the mechanisms of HER2-positive breast cancer, we analyzed microarray of GSE52194.

MATERIALS AND METHODS: GSE52194 was downloaded from Gene Expression Omnibus including 5 HER2-positive breast cancer samples and 3 normal breast samples. Using cuffdiff software, differentially expressed genes (DEGs) and differentially expressed long non-coding RNAs (DE-lncRNAs) were screened. Functions of the DEGs were analyzed by Gene Ontology (GO) and pathway enrichment analyses. Then, protein-protein interaction (PPI) network of the DEGs were constructed using Cytoscape and modules of the PPI network were screened by CFinder. Moreover, lncRNA-DEG pairs were screened.

RESULTS: Total 209 lncRNA transcriptions were predicted, and 996 differentially expressed transcriptions were screened. Besides, FOS had interaction relationships with EGR1 and SOD2 separately in module E and F of the PPI network for the DEGs. Moreover, there were many lncRNA-DEG pairs (e.g. *TCONS_00003876-EGR1*, *TCONS_00003876-FOS*, *lnc-HOXC4-3:1-FOS*, *lnc-HOXC4-3:1-BCL6B*, *lnc-TEAD4-1:1-FOS* and *lnc-TEAD4-1:1-BCL6B*), meanwhile, co-expressed DEGs of *TCONS_00003876*, *lnc-HOXC4-3:1* and *lnc-TEAD4-1:1* were enriched in p53 signaling pathway, MAPK signaling pathway and cancer related pathways, respectively.

CONCLUSIONS: *ANXA1*, *EGR1*, *BCL6*, *SOD2*, *FOS*, *TCONS_00003876*, *lnc-HOXC4-3:1* and *lnc-TEAD4-1:1* might play a role in HER2-positive breast cancer.

Key words:

Breast cancer, Differentially expressed genes, Protein-protein interaction network, Module analysis, Long non-coding RNAs.

Introduction

Breast cancer is characterized by fluid coming from the nipple, a lump in the breast, dimpling of

the skin, or a red scaly patch of skin¹. It may be caused by early age at first menstruation, drinking alcohol, lacking of physical exercise, obesity, ionizing radiation, having children too late or having no children^{1,2}. In the worldwide, breast cancer is with the highest frequency (25.2%) as an invasive cancer in women² and comprises 16% of all female cancers³.

As a member of the epidermal growth factor (*EGF*) family of receptor tyrosine kinases⁴, human epidermal growth factor receptor type 2 (*HER2*) expression is associated with the poor prognosis of breast cancer⁵⁻⁷. By moderating the expression of TNF-related weak inducer of apoptosis (TWEAK) and its receptor TNF-related weak inducer of apoptosis receptor (TWEAKR), nuclear factor of activated T cells 1 (*NFAT1*) plays a role in cell invasion of breast cancer⁸. However, nuclear factor of activated T cells 3 (*NFAT3*) can blunt cell invasion of breast cancer through inhibiting Lipocalin 2 expression⁹. Breast cancer gene 1 (*BRCA1*) and breast cancer gene 2 (*BRCA2*) can be used as markers to identify patients at high risk of developing breast cancer¹⁰. By enhancing E-cadherin-mediated cell-to-cell adhesion, epithelial cell adhesion molecule (*EPCAM*) expression functions in cell proliferation, migration and invasion of breast cancer¹¹.

Long non-coding RNAs (lncRNAs) are poorly conserved and cannot code proteins, but may be used in cancer diagnostic and therapeutic¹². HOX transcript antisense RNA (*HOTAIR*) expression increases during breast cancer progression, and its expression level in primary tumor can be a predictor of eventual metastasis and death¹³. Overexpression of lncRNA breast cancer anti-estrogen resistance 4 (*BCAR4*) induces cell proliferation of breast cancer and it can function as potential target for treatment of antiestrogen-resistant breast cancer¹⁴. lncRNA metastasis-associated lung adenocarcinoma transcript 1 (*MALAT1*) upregulates in breast cancer, and may function in cancer diagnostic and therapeutic^{12,15}.

In 2013, Horvath et al¹⁶ used a whole transcriptome RNA-sequencing to detect single nucleotide polymorphisms (SNPs) between breast cancer tissues and normal breast tissues, and found some new mutations in key regulatory molecules in breast cancer. Using the same data by Horvath et al¹⁶, we screened the differentially expressed lncRNAs (DE-lncRNAs) and differentially expressed gene (DEGs) in HER2-positive breast cancer samples. Using Gene Ontology (GO) and pathway enrichment analyses, functions of the DEGs were analyzed. In addition, protein-protein interaction (PPI) network of proteins encoded by the DEGs were constructed and modules of the PPI network were screened. Furthermore, lncRNA-DEG pairs were screened, and functions of the DEGs co-expressed with each lncRNA were analyzed using pathway enrichment analyses.

Materials and Methods

Microarray data

Expression profile of GSE52194 deposited by Horvath et al¹⁶ was downloaded from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) which was based on the platform of GPL11154 Illumina HiSeq 2000 (Homo sapiens). GSE52194 included a collective of 5 HER2-positive breast cancer samples and 3 normal breast samples. All the enrolled breast cancer patients were from European descent.

Sequence alignment

After GSE52194 was downloaded, the format of sequences were translated from sra into fastq, then low quality sequences were screened out by NGSQC software¹⁷ (the ratio of bases in sequences with base sequencing quality less than twenty was required to be less than 0.9, and the ratio of base N needed to be less than 0.05). Using tophat¹⁸, the remaining high quality sequences were mapped to human genome 19 (hg19). The parameter was set at --no-discordant --phred64-quals, and the other parameters were set to the defaults.

LncRNA prediction

Using cufflinks software¹⁹, the alignment results were performed transcriptome assembly. Then the assembly results of the 8 samples were integrated by cuffmerge software²⁰. According to the genome annotation information in the ucsc (GENECODE V19, Ensembl, UCSC and RefSeq), assembly results without overlapping compared to arbitrary genes were extracted. Subsequently, as-

sembled transcriptions with the length larger than 200 nt and with two or more exons were identified. Transcriptions with scores less than 100 were obtained using PhyloCSF software²¹. Moreover, hmmer software (St. Louis, MO, USA)²² was used to exclude transcriptions with E-value less than 1e-5.

DEGs and DE-lncRNAs screening

Combing with known lncRNA comments (LNCipedia1.0 database), RefSeq annotation files in ucsc website and the predicted lncRNA transcription document, DEGs and DE-lncRNAs were identified by cuffdiff software²³. The false discovery rate (FDR) < 0.05 and |logfold-change (FC)| > 1 were used as the cut-off criteria.

Functional and pathway enrichment analysis

As a tool for editing, viewing and querying, GO is used to annotate gene products²⁴. Kyoto Encyclopedia of Genes and Genomes (KEGG) has functional assignment that links genes in the genome with interacting molecules, such as a complex or a pathway, standing for a higher order biological function²⁵. GO and KEGG pathway enrichment analyses were performed for the DEGs. The *p*-value < 0.05 and involved two or more genes were used as the cut-off criteria.

PPI network and module analysis

The interaction relationships of the proteins encoded by DEGs were searched by STRING online software²⁶, and the combined score > 0.4 was used as the cut-off criterion. The PPI network was visualized using Cytoscape software²⁷. Then, modules of the PPI network were screened by CFinder software²⁸, and the parameter *k* was set to 6.

LncRNA analysis

According to the expression matrix of the DEGs and the DE-lncRNAs, their correlations were calculated and lncRNA-DEG pairs were identified. The pearson correlation > 0.95 was used as the cut-off criterion.

Results

LncRNA prediction, DE-lncRNAs and DEGs analysis

Total 209 lncRNA transcriptions were predicted, including 84 lncRNA transcriptions with more than 50% overlapping compared to known lncRNAs.

Compared with normal breast samples, there were 996 differentially expressed transcriptions,

including 92 up-regulated transcriptions (including 2 known lncRNAs, and 90 DEGs) and 904 down-regulated transcriptions (including 9 known lncRNAs, 2 predicted lncRNAs and 893 DEGs) in HER2-positive breast cancer samples.

Functional and pathway enrichment analysis

The enriched GO functions for the up-regulated genes were listed in Table 1, including macromolecular complex assembly (p -value = $7.74E-13$), macromolecular complex subunit organization (p -value = $8.37E-11$), cellular component assembly (p -value = $2.65E-10$) and non-membrane-bounded organelle (p -value = $2.69E-08$). The pathway systemic lupus erythematosus (p -value = 0) was enriched for the up-regulated genes (Table I).

The enriched GO functions for the down-regulated genes were also listed in Table I, including nuclear-transcribed mRNA catabolic process,

nonsense-mediated decay (p -value = $1.11E-16$), SRP-dependent cotranslational protein targeting to membrane (p -value = $3.66E-15$), mRNA catabolic process (p -value = $5.40E-13$) and translational elongation (p -value = $8.44E-13$). The enriched pathways for the down-regulated genes included amoebiasis (p -value = 0.000512052), NOD-like receptor signaling pathway (p -value = 0.000547511), p53 signaling pathway (p -value = 0.000615533) and small cell lung cancer (p -value = 0.001493752) (Table I).

PPI network and module analysis

The PPI network of the DEGs had 540 nodes including annexin A1 (ANXA1), early growth response 1 (EGR1), B-cell CLL/lymphoma 6 (BCL6), c-fos (FOS) and mitochondrial superoxide dismutase 2 (SOD2), as well as had 2017 interactions (Figure 1). In addition, 6 modules (module A, B, C, D, E and F) were screened

Table I. The top 10 enriched GO functions and KEGG pathways for the DEGs.

| | ID | Name | Gene number | Gene symbol | p-value |
|----------------------|------------|---|-------------|---------------------------|-------------|
| Up-regulated genes | GO:0065003 | macromolecular complex assembly | 21 | CASC5, CENPF..... | 7.74E-13 |
| | GO:0043933 | macromolecular complex subunit organization | 21 | CENPF, HIST1H1A..... | 8.37E-11 |
| | GO:0022607 | cellular component assembly | 22 | CASC5, CENPF..... | 2.65E-10 |
| | GO:0043228 | non-membrane-bounded organelle | 26 | ASPM, BRCA2..... | 2.69E-08 |
| | GO:0046982 | protein heterodimerization activity | 15 | HIST1H2AB, HIST1H2AG..... | 2.86E-14 |
| Down-regulated genes | GO:0000184 | nuclear-transcribed mRNA catabolic process, nonsense-mediated decay | 34 | EIF4A3, ETF1..... | 1.11E-16 |
| | GO:0044391 | SRP-dependent cotranslational protein targeting to membrane | 33 | FAU, GNB2L1..... | 3.66E-15 |
| | GO:0044421 | mRNA catabolic process | 111 | S16, RPS19..... | 5.40E-13 |
| | GO:0005840 | translational elongation | 38 | APOD, FAU..... | 8.44E-13 |
| | GO:0006613 | cotranslational protein targeting to membrane | 32 | FAU, RPL10A..... | 2.22E-16 |
| | ID | Name | Gene number | Gene symbol | p-value |
| Up-regulated genes | KEGG:5322 | Systemic lupus erythematosus | 42 | HIST1H2AB, HIST1H2AE..... | 0 |
| Down-regulated genes | KEGG:5146 | Amoebiasis | 16 | ACTN1, GNA11..... | 0.000512052 |
| | KEGG:4621 | NOD-like receptor signaling pathway | 11 | BIRC3, CCL2..... | 0.000547511 |
| | KEGG:4115 | p53 signaling pathway | 12 | BAI1, CDKN1A..... | 0.000615533 |
| | KEGG:5222 | Small cell lung cancer | 13 | BIRC3, ITGA2..... | 0.001493752 |
| | KEGG:360 | Phenylalanine metabolism | 5 | AOC3, GOT1..... | 0.002558543 |
| | KEGG:5144 | Malaria | 9 | CCL2, CD81..... | 0.002919392 |
| | KEGG:5142 | Chagas disease (American trypanosomiasis) | 14 | CCL2, FOS..... | 0.003430112 |
| | KEGG:4380 | Osteoclast differentiation | 16 | CSF1, FOS..... | 0.003879054 |
| | KEGG:5200 | Pathways in cancer | 31 | BIRC3, BMP2..... | 0.006552164 |

Mechanisms of HER2-positive breast cancer

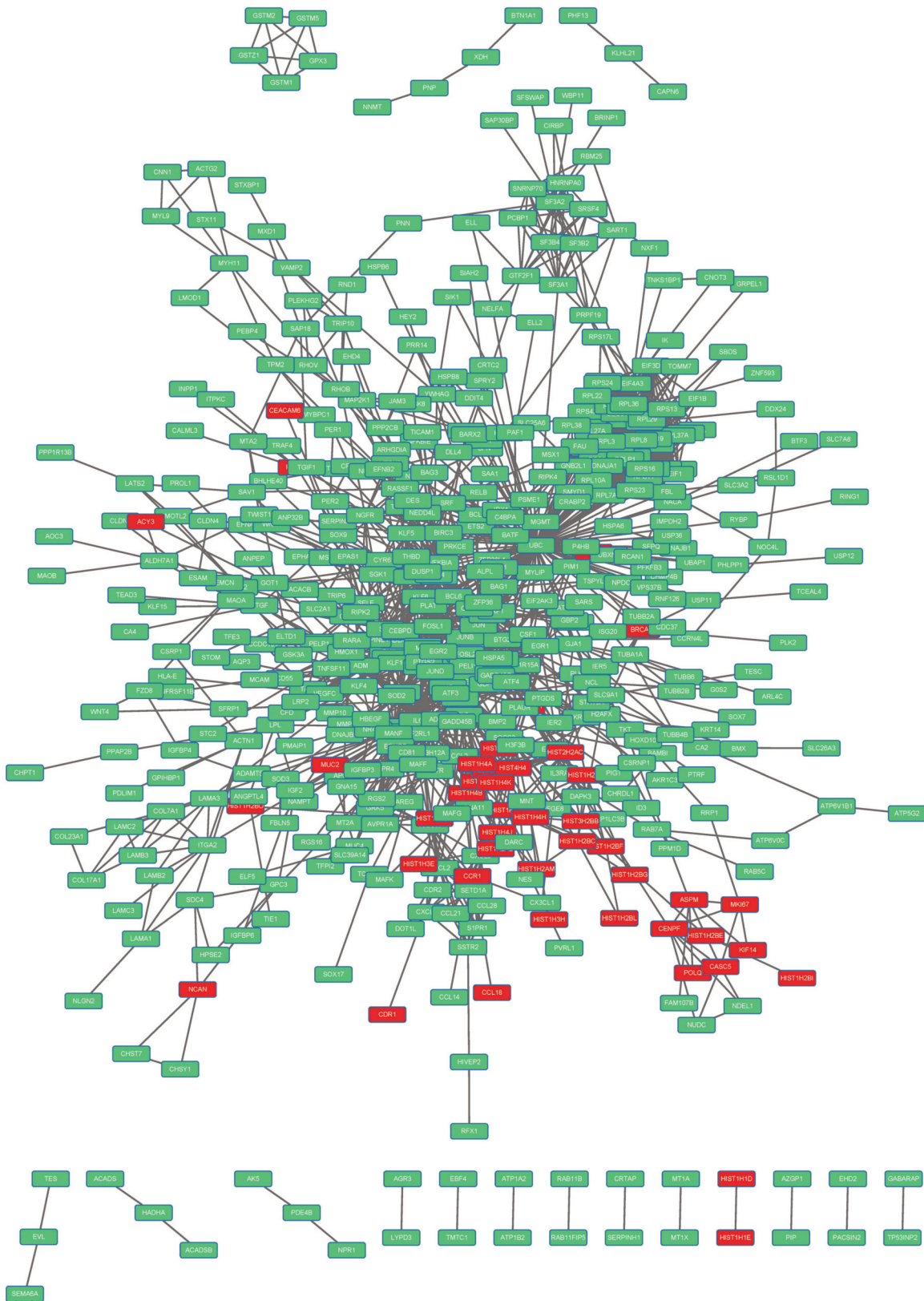


Figure 1. PPI network for the proteins encoded by the DEGs. Red and green circles stand for the up- and down-regulated genes, respectively.

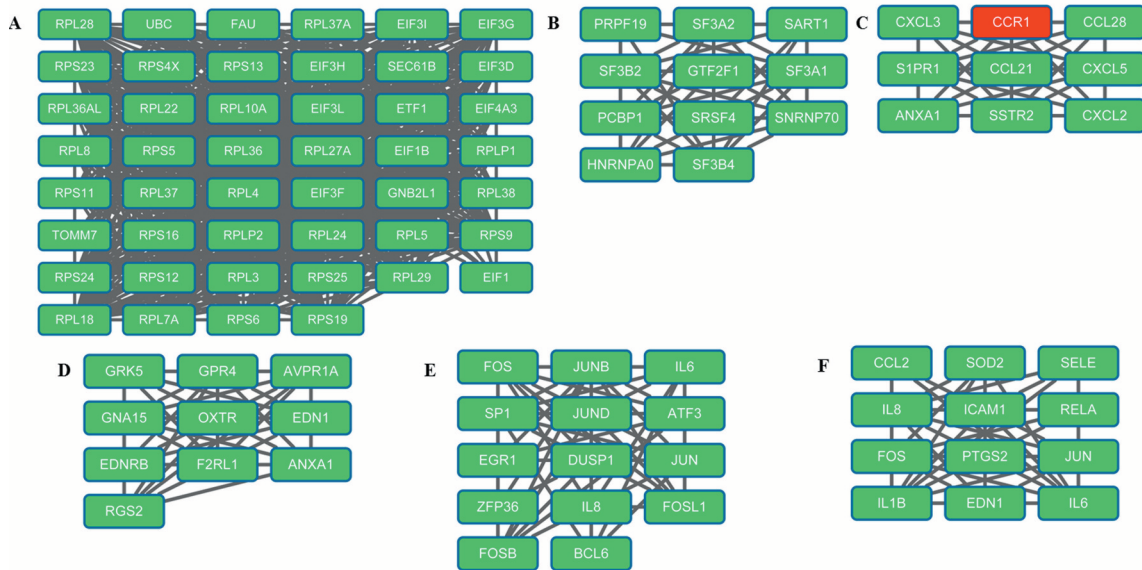


Figure 2. Module A, B, C, D, E and F obtained from the PPI network of the DEGs. Red and green circles represent for the up- and down-regulated genes, respectively.

from the PPI network (Figure 2). And the enriched GO functions and KEGG pathways for the DEGs in the modules were listed in Table II. Module A had 46 nodes and 776 interactions. The enriched GO functions for DEGs in module A included translation (p -value = $8.51E-64$) and ribosome biogenesis (p -value = $1.19E-07$). The enriched KEGG pathways for DEGs in module A included Ribosome (p -value = $2.39E-54$).

Module B had 11 nodes and 46 interactions. The enriched GO functions for DEGs in module B included RNA splicing (p -value = $6.99E-16$) and mRNA processing (p -value = $2.14E-15$). The enriched KEGG pathways for DEGs in module B included spliceosome (p -value = $3.81E-11$).

Module C had 9 nodes (such as ANXA1) and 36 interactions. The enriched GO functions for DEGs in module C included inflammatory response (p -value = $2.12E-05$) and regulation of cell proliferation (p -value = 0.008809). The enriched KEGG pathways for DEGs in module C included chemokine signaling pathway (p -value = $1.26E-06$) and cytokine-cytokine receptor interaction (p -value = $6.75E-06$).

Module D had 10 nodes (such as ANXA1) and 44 interactions. The enriched GO functions for DEGs in module D included elevation of cytosolic calcium ion concentration (p -value = $3.98E-09$). The enriched KEGG pathways for DEGs in module D included calcium signaling pathway (p -value = 0.001613) and neuroactive ligand-receptor interaction (p -value = 0.004944).

Module E had 14 nodes (such as EGR1 and BCL6) and 56 interactions (such as FOS-EGR1). The enriched GO functions for DEGs in module E included regulation of transcription from RNA polymerase II promoter (p -value = $2.10E-09$). The enriched KEGG pathways for DEGs in module E included toll-like receptor signaling pathway (p -value = $3.96E-04$) and MAPK signaling pathway (p -value = 0.006582).

Module F had 12 nodes (such as SOD2 and FOS) and 46 interactions (such as FOS-SOD2). The enriched GO functions for DEGs in module F included response to drug (p -value = $6.68E-09$) and response to molecule of bacterial origin (p -value = $1.06E-13$). The enriched KEGG pathways for DEGs in module F included toll-like receptor signaling pathway (p -value = $1.18E-06$) and pathways in cancer (p -value = $3.62E-04$).

LncRNA analysis

After lncRNA-DEG pairs (e.g. *TCONS_00003876-EGR1*, *TCONS_00003876-FOS*, *lnc-HOXC4-3:1-FOS*, *lnc-HOXC4-3:1-BCL6B*, *lnc-TEAD4-1:1-FOS* and *lnc-TEAD4-1:1-BCL6B*) were screened, KEGG pathway enrichment analyses were conducted for DEGs co-expressed with each lncRNA. Especially, co-expressed DEGs of *TCONS_00003876*, *lnc-HOXC4-3:1* and *lnc-TEAD4-1:1* were enriched in p53 signaling pathway, MAPK signaling pathway and cancer related pathways, respectively (Table III).

Mechanisms of HER2-positive breast cancer

Table II. The top 10 enriched GO functions and KEGG pathways for DEGs in module A, B, C, D, E and F of the PPI network for the DEGs.

| | ID | Name | Gene number | Gene symbol | p-value |
|----------|------------|---|-------------|---------------------|----------|
| Module-A | GO:0006412 | Translation | 42 | RPL18, PL27A..... | 8.51E-64 |
| | GO:0042254 | Ribosome biogenesis | 8 | EIF4A3, RPS19..... | 1.19E-07 |
| Module-B | GO:0008380 | RNA splicing | 10 | PRPF19, GTF2F1..... | 6.99E-16 |
| | GO:0006397 | mRNA processing | 10 | PRPF19, GTF2F1..... | 2.14E-15 |
| Module-C | GO:0006954 | Inflammatory response | 5 | CXCL2, ANXA1..... | 2.12E-05 |
| | GO:0042127 | Regulation of cell proliferation | 4 | S1PR1, ANXA1..... | 0.008809 |
| Module-D | GO:0007204 | Elevation of cytosolic calcium ion concentration | 6 | EDNRB, GNA15..... | 3.98E-09 |
| Module-E | GO:0006357 | Regulation of transcription from RNA polymerase II promoter | 10 | EGR1, BCL6..... | 2.10E-09 |
| Module-F | GO:0042493 | Response to drug | 7 | FOS, SOD2..... | 6.68E-09 |
| | GO:0002237 | Response to molecule of bacterial origin | 8 | FOS, IL6..... | 1.06E-13 |

| | ID | Name | Gene number | Gene symbol | p-value |
|----------|-------------|---|-------------|----------------------------|----------|
| Module-A | KEGG: 03010 | Ribosome | 32 | RPL18, RPL27A..... | 2.39E-54 |
| Module-B | KEGG03040 | Spliceosome | 8 | PRPF19, PCBP1..... | 3.81E-11 |
| Module-C | KEGG:04062 | Chemokine signaling pathway | 6 | CXCL5, CCL21..... | 1.26E-06 |
| | KEGG:04060 | Cytokine-cytokine receptor interaction | 6 | CXCL3, CCR1..... | 6.75E-06 |
| Module-D | KEGG:04020 | Calcium signaling pathway | 4 | EDNRB, GNA15, AVPR1A, OXTR | 0.001613 |
| | KEGG:04080 | Neuroactive ligand-receptor interaction | 4 | EDNRB, F2RL1, AVPR1A, OXTR | 0.004944 |
| Module-E | KEGG:04620 | Toll-like receptor signaling pathway | 4 | FOS, IL6, IL8, JUN | 3.96E-04 |
| | KEGG:04010 | MAPK signaling pathway | 4 | FOS, DUSP1, JUN, JUND | 0.006582 |
| Module-F | KEGG:04620 | Toll-like receptor signaling pathway | 6 | FOS, IL6..... | 1.18E-06 |
| | KEGG:05200 | Pathways in cancer | 6 | FOS, IL8..... | 3.62E-04 |

Table III. The top 10 enriched KEGG pathways for the DEGs co-expressed with lncRNA TCONS_00003876, lnc-HOXC4-3:1 and lnc-TEAD4-1:1.

| lncRNA | ID | Name | Gene number | Gene symbol | p-value |
|----------------|-----------|------------------------|-------------|---|----------|
| TCONS_00003876 | KEGG:4115 | p53 signaling pathway | 4 | CDKN1A, SERPINB5, GADD45B, SESN2 | 0.001447 |
| | KEGG:4010 | MAPK signaling pathway | 5 | JUND, JUN, DUSP5, FOS, GADD45B | 0.044636 |
| lnc-HOXC4-3:1 | KEGG:5219 | Bladder cancer | 2 | CDKN1A, VEGFC | 0.037511 |
| | KEGG:4115 | p53 signaling pathway | 5 | CDKN1A, IGF1BP3, GADD45A, PPM1D, SESN2 | 0.038912 |
| lnc-TEAD4-1:1 | KEGG:4010 | MAPK signaling pathway | 13 | NTRK2, FOS..... | 0.033283 |
| | KEGG:5200 | Pathways in cancer | 19 | LAMB3, FOS..... | 0.00149 |
| | KEGG:5210 | Colorectal cancer | 5 | JUN, MYC, MAP2K1, FOS, TCF7L1 | 0.024433 |
| | KEGG:4010 | MAPK signaling pathway | 13 | NTRK2, FOS..... | 0.02663 |
| | KEGG:4115 | p53 signaling pathway | 5 | CDKN1A, SERPINB5, PPM1D, GADD45B, SESN2 | 0.034686 |
| | KEGG:5216 | Thyroid cancer | 3 | MYC, MAP2K1, TCF7L1 | 0.041256 |

Discussion

In this study, total 209 lncRNA transcriptions were predicted, including 84 lncRNAs transcriptions with more than 50% overlapping compared to known lncRNAs. Moreover, 996 differentially expressed transcriptions, including 92 up-regulated transcriptions and 904 down-regulated transcriptions were screened. The functions of macromolecular complex assembly, macromolecular complex subunit organization and SRP-dependent cotranslational protein targeting to membrane were enriched for the DEGs.

As a member of calcium/phospholipid-binding and actin regulatory proteins family²⁹, ANXA1 was found in the module C and D of the PPI network for the DEGs. ANXA1 may play multiple role in development, metastase, and progression of breast cancer³⁰. In MCF-7 breast cancer, ANXA1 may function as a tumor suppressor gene and regulate the proliferative of estrogens³¹. Combing with the IκB kinase complex, ANXA1 functions in activation of nuclear factor-κB (NF-κB) in breast cancer cells, and may has therapeutic potential in suppression of breast cancer metastasis³². By specific activating the NF-κB signaling pathway, ANXA1 can positive regulate expression of matrix metalloproteinase 9 (MMP9) and invasion of breast tumor cells³³. Thus, the expression levels of ANXA1 might have relation to HER2-positive breast cancer.

EGR1 functions in the repression of N-myc downregulated gene 1 (NDRG1) by combing with T-box 2 (TBX2) and plays a role in driving proliferation of breast cancer cells³⁴. EGR1 may involve in breast cancer cell development and act as an important breast cancer marker³⁵. DNazymes targeting EGR-1 may inhibit multiple important tumorigenic processes, indicating that it may be used as anti-cancer agents³⁶. Expression of EGR1 decreases in estrogen receptor (ER)-negative human breast carcinomas and may has relation to the pathogenesis of breast carcinomas³⁷. BCL6 also involved in mammary epithelial differentiation and carcinogenesis³⁸. These declared that EGR1 and BCL6 might be associated with HER2-positive breast cancer.

As a mitochondrial matrix enzyme, manganese superoxide dismutase (MnSOD) is encoded by SOD2³⁹. Overexpression of MnSOD in tumor derived cell lines can attenuate malignant phenotype and then lower metastatic potential of them⁴⁰⁻⁴². Previous study suggests that MnSOD may be a new type of tumor suppressor gene⁴³.

The c-FOS protein may function as a tumour-suppressor and may play a role in apoptosis⁴⁴. These might indicate that the expression levels of SOD2 and FOS had relation to HER2-positive breast cancer. FOS could interact with EGR1 and SOD2 separately in module E and F of the PPI network for the DEGs, suggesting that FOS might also play a role in HER2-positive breast cancer through regulating EGR1 and SOD2. There were many interaction relationships between lncRNAs and DEGs (e.g. TCONS_00003876-EGR1, TCONS_00003876-FOS, lnc-HOXC4-3:1-FOS, lnc-HOXC4-3:1-BCL6B, lnc-TEAD4-1:1-FOS and lnc-TEAD4-1:1-BCL6B), meanwhile, co-expressed DEGs of TCONS_00003876, lnc-HOXC4-3:1 and lnc-TEAD4-1:1 were enriched in p53 signaling pathway, MAPK signaling pathway and cancer related pathways, respectively. These suggested that TCONS_00003876, lnc-HOXC4-3:1 and lnc-TEAD4-1:1 might involve in HER2-positive breast cancer by mediating their co-expressed genes.

Conclusions

We conducted a comprehensive bioinformatics analysis of genes and lncRNAs which might associate with breast cancer. Total 209 lncRNA transcriptions were predicted, meanwhile, 996 differentially expressed transcriptions were screened. Besides, ANXA1, EGR1, BCL6, SOD2, FOS, TCONS_00003876, lnc-HOXC4-3:1 and lnc-TEAD4-1:1 might correlate with HER2-positive breast cancer. However, further researches are still needed to illuminate their functions in HER2-positive breast cancer.

Conflict of Interests

The Authors declare that they have no conflict of interests.

References

1. STAGE I, II I, OPERABLE I. Breast cancer treatment (pdq®).
2. STEWART BW, WILD C. World cancer report 2014. World Health Organ, 2014.
3. ORGANIZATION WH. Breast cancer: prevention and control. World Health Organ [Online], 2012.
4. Yarden Y, Sliwkowski MX. Untangling the erbb signalling network. Nat Rev Mol Cell Biol 2001; 2: 127-137.

5. GARCIA I, VIZOSO F, MARTIN A, SANZ L, ABDEL-LAH O, RAIGOSO P, GARCÍA-MUÑOZ J. Clinical significance of the epidermal growth factor receptor and her2 receptor in resectable gastric cancer. *Ann Surg Oncol* 2003; 10: 234-241.
6. NIELSEN TO, HSU FD, JENSEN K, CHEANG M, KARACA G, HU Z, HERNANDEZ-BOUSSARD T, LIVASY C, COWAN D, DRESSLER L, AKSLEN LA, RAGAZ J, GOWN AM, GILKS CB, VAN DE RIJN M, PEROU CM. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 2004; 10: 5367-5374.
7. TOVEY SM, WITTON CJ, BARTLETT J, STANTON PD, REEVES JR, COOKE TG. Outcome and human epidermal growth factor receptor (her) 1-4 status in invasive breast carcinomas with proliferation indices evaluated by bromodeoxyuridine labelling. *Breast Cancer Res* 2004; 6: R246-R251.
8. GAUDINEAU B, FOUGÈRE M, GUADDACHI F, LEMOINE F, DE LA GRANGE P, JAULIAC S. Lipocalin 2, the tnfr-like receptor tweakr and its ligand tweak act downstream of nfat1 to regulate breast cancer cell invasion. *J Cell Sci* 2012; 125: 4475-4486.
9. FOUGÈRE M, GAUDINEAU B, BARBIER J, GUADDACHI F, FEUGEAS JP, AUBOEU F, JAULIAC S. Nfat3 transcription factor inhibits breast cancer cell motility by targeting the lipocalin 2 gene. *Oncogene* 2010; 29: 2292-2301.
10. DUFFY M. Biochemical markers in breast cancer: Which ones are clinically useful? *Clin Biochem* 2001; 34: 347-352.
11. OSTA WA, CHEN Y, MIKHITARIAN K, MITAS M, SALEM M, HANNUN YA, COLE DJ, GILLANDERS WE. Epcam is overexpressed in breast cancer and is a potential target for breast cancer gene therapy. *Cancer Res* 2004; 64: 5818-5824.
12. QI P, DU X. The long non-coding rnas, a new cancer diagnostic and therapeutic gold mine. *Modern Pathol* 2012; 26: 155-165.
13. CHISHOLM KM, WAN Y, LI R, MONTGOMERY KD, CHANG HY, WEST RB. Detection of long non-coding rna in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma. *PLoS One* 2012; 7: e47998.
14. GODINHO M, MEIJER D, SETYONO-HAN B, DORSSERS LC, VAN AGTHOVEN T. Characterization of bcar4, a novel oncogene causing endocrine resistance in human breast cancer cells. *J Cell Physiol* 2011; 226: 1741-1749.
15. GUFFANTI A, IACONO M, PELUCCHI P, KIM N, SOLDÀ G, CROFT LJ, TAFT RJ, RIZZI E, ASKARIAN-AMIRI M, BONNAL RJ, CALLARI M, MIGNONE F, PESOLE G, BERTALOT G, BERNARDI LR, ALBERTINI A, LEE C, MATTICK JS, ZUCCHI I, DE BELLIS G. A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. *BMC Genomics* 2009; 10: 163.
16. HORVATH A, PAKALA SB, MUDVARI P, REDDY SDN, OHSHIRO K, CASIMIRO S, PIRES R, FUQUA SA, TOI M, COSTA L, NAIR SS, SUKUMAR S, KUMAR R. Novel insights into breast cancer genetic variance through rna sequencing. *Sci Rep* 2013; 3: 2256
17. DAI M, THOMPSON RC, MAHER C, CONTRERAS-GALINDO R, KAPLAN MH, MARKOVITZ DM, OMENN G, MENG F. Ngsqc: Cross-platform quality analysis pipeline for deep sequencing data. *BMC Genomics* 2010; 11 Suppl 4: S7.
18. TRAPNELL C, PACTHER L, SALZBERG SL. Tophat: discovering splice junctions with rna-seq. *Bioinformatics* 2009; 25: 1105-1111.
19. TRAPNELL C, ROBERTS A, GOFF L, PERTEA G, KIM D, KELLEY DR, PIMENTEL H, SALZBERG SL, RINN JL, PACTHER L. Differential gene and transcript expression analysis of rna-seq experiments with tophat and cufflinks. *Nat Protoc* 2012; 7: 562-578.
20. LI B, DUAN H, LI J, DENG XW, YIN W, XIA X. Global identification of mirnas and targets in populus euphratica under salt stress. *Plant Mol Biol* 2013; 81: 525-539.
21. LIN MF, JUNGREIS I, KELLIS M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* 2011; 27: i275-i282.
22. EDDY S. Hmmer user's guide. Biological sequence analysis using profile hidden markov models. 2003.
23. ANDERS S, REYES A, HUBER W. Detecting differential usage of exons from rna-seq data. *Genome Res* 2012; 22: 2008-2017.
24. YE H, KARP PD, NOY NF, ALTMAN RB. Knowledge acquisition, consistency checking and concurrency control for gene ontology (go). *Bioinformatics* 2003; 19: 241-248.
25. KANEHISA M, GOTO S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; 28: 27-30.
26. FRANCESCHINI A, SZKLARCZYK D, FRANKILD S, KUHN M, SIMONOVIC M, ROTH A, LIN J, MINGUEZ P, BORK P, VON MERING C, JENSEN LJ. String v9. 1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 2013; 41:D808-D815.
27. KOHL M, WIESE S, WARSCHIED B. Cytoscape: software for visualization and analysis of biological networks. *Data mining in proteomics*. Springer, 2011; pp. 291-303.
28. ADAMCEK B, PALLA G, FARKAS IJ, DERÉNYI I, VICSEK T. Cfinder: Locating cliques and overlapping modules in biological networks. *Bioinformatics* 2006; 22: 1021-1023.
29. LIM LH, PERVAIZ S. Annexin 1: The new face of an old molecule. *FASEB J* 2007; 21: 968-975.
30. WANG LP, BI J, YAO C, XU XD, LI XX, WANG SM, LI ZL, ZHANG DY, WANG M, CHANG GO. Annexin A1 expression and its prognostic significance in human breast cancer. *Neoplasma* 2009; 57: 253-259.
31. ANG EZ, NGUYEN HT, SIM HL, PUTTI TC, LIM LH. Annexin-1 regulates growth arrest induced by high levels of estrogen in MCF-7 breast cancer cells. *Mol Cancer Res* 2009; 7: 266-274.
32. BIST P, LEOW SC, PHUA QH, SHU S, ZHUANG Q, LOH WT, NGUYEN TH, ZHOU JB, HOOI SC, LIM LH. Annexin-1 interacts with NEMO and RIP1 to constitutively activate IKK complex and NF- κ B: implication in breast cancer metastasis. *Oncogene* 2011; 30: 3174-3185.
33. KANG H, KO J, JANG SW. The role of annexin A1 in expression of matrix metalloproteinase-9 and invasion of breast cancer cells. *Biochem Biophys Res Commun* 2012; 423: 188-194.

34. REDMOND KL, CRAWFORD NT, FARMER H, D'COSTA ZC, O'BRIEN GJ, BUCKLEY NE, KENNEDY RD, JOHNSTON PG, HARKIN DP, MULLAN PB. T-box 2 represses NDRG1 through an EGR1-dependent mechanism to drive the proliferation of breast cancer cells. *Oncogene* 2010; 29: 3252-3262.
35. LIU J, LIU YG, HUANG R, YAO C, LI S, YANG W, YANG D, HUANG RP. Concurrent down-regulation of Egr-1 and gelsolin in the majority of human breast cancer cells. *Cancer Genom Proteom* 2007; 4: 377-385.
36. MITCHELL A, DASS CR, SUN LQ, KHACHIGIAN LM. Inhibition of human breast carcinoma proliferation, migration, chemoinvasion and solid tumour growth by DNAzymes targeting the zinc finger transcription factor EGR-1. *Nucleic Acids Res* 2004; 32: 3065-3069.
37. RONSKI K, SANDERS M, BURLESON JA, MOYO V, BENN P, FANG M. Early growth response gene 1 (EGR1) is deleted in estrogen receptor-negative human breast carcinoma. *Cancer* 2005; 104: 925-930.
38. LOGARAJAH S, HUNTER P, KRAMAN M, STEELE D, LAKHANI S, BOBROW L, VENKITARAMAN A, WAGNER S. BCL-6 is expressed in breast cancer and prevents mammary epithelial differentiation. *Oncogene* 2003; 22: 5572-5578.
39. HITCHLER MJ, OBERLEY LW, DOMANN FE. Epigenetic silencing of SOD2 by histone modifications in human breast cancer cells. *Free Radic Biol Med* 2008; 45: 1573-1580.
40. OUGH M, LEWIS A, ZHANG Y, HINKHOUSE MM, RITCHIE JM, OBERLEY LW, CULLEN JJ. Inhibition of cell growth by overexpression of manganese superoxide dismutase (MnSOD) in human pancreatic carcinoma. *Free Radic Res* 2004; 38: 1223-1233.
41. WEYDERT C, ROLING B, LIU J, HINKHOUSE MM, RITCHIE JM, OBERLEY LW, CULLEN JJ. Suppression of the malignant phenotype in human pancreatic cancer cells by the overexpression of manganese superoxide dismutase1. *Mol Cancer Ther* 2003; 2: 361-369.
42. WEYDERT CJ, WAUGH TA, RITCHIE JM, IYER KS, SMITH JL, LI L, SPITZ DR, OBERLEY LW. Overexpression of manganese or copper-zinc superoxide dismutase inhibits breast cancer growth. *Free Radic Biol Med* 2006; 41: 226-237.
43. BRAVARD A, SABATIER L, HOFFSCHIR F, RICOUL M, LUCIONI C, DUTRILLAUX B. SOD2: A new type of tumor-suppressor gene? *Int J Cancer* 1992; 51: 476-480.
44. TENG CS. Protooncogenes as mediators of apoptosis. *Int Rev Cytol* 2000; 197: 137-202.