Exploration of Core Genes Signature and Prognostic Implication for Ductal Carcinoma In Situ of the Breast

Jiao Zhang
Zhengzhou University  https://orcid.org/0000-0002-2219-944X

Hui Lin
Wenzhou Medical University

Lei Hou
Zhengzhou University

Hui Xiao
Zhengzhou University

Xilong Gong
Zhengzhou University

Xuhui Guo
Zhengzhou University

Xuchen Cao
Tianjin Medical University

Zhennzen Liu (✉ zlyyliuzhennzen0800@zzu.edu.cn )
Zhengzhou University  https://orcid.org/0000-0002-3878-1258

Primary research

Keywords: Ductal carcinoma in situ, GEO, Prognosis, Gene signature

Posted Date: July 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-622340/v1

License: © 侵权 This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

**Background:** Following the implementation of breast screening programs, the occurrence of DCIS has risen as an early form of neoplasm. Although the prognosis is good, 20-50% of DCIS patients will develop invasive ductal carcinoma (IDC) if they are not handled. It is important to look for promising biomarkers for predicting DCIS prognosis.

**Methods:** The Gene Expression Omnibus (GEO) database provided three microarray profile datasets. The expression of genes that differed between DCIS and normal tissue was investigated. To describe the biological role and intrinsic process pathway, enrichment analysis was used. The Cancer Genome Atlas Breast Cancer Dataset was used to classify the hub genes and further verify the findings using CytoHubba and MCODE, two Cytoscape plugins. The prognostic ability of the core genes signature was determined through time-dependent receiver operating characteristic (ROC), Kaplan-Meier survival curve, Oncomine databases, and UALCAN databases. In addition, in proliferation assays, the prognostic value of core genes was verified.

**Results:** We identified 217 common DEGs, with 101 up-regulated and 138 down-regulated genes in the present study. The top genes were obtained from the PPI network (protein-protein interaction). For DCIS prognosis prediction, a novel six gene signature (including GAPDH, CDH2, BIRC5, NEK2, IDH2, and MELK) was developed. Centered on the TCGA cohort, the ROC curve showed strong results in prognosis prediction. The six core genes signature is often overexpressed in DCIS, which has a weak prognosis. Furthermore, transfected with small interfering RNAs, downregulation of core gene expression significantly inhibits breast cancer cell proliferation, implying a great potential for using core genes in DCIS prognosis.

**Conclusions:** The six core genes signature for promising DCIS biomarkers was validated in our research, which may assist in clinical decision-making for individual care.

**Background**

According to the GLOBOCAN 2020 profile produced by the International Agency for Research on Cancer, breast cancer has overtaken lung cancer as the world's most commonly diagnosed cancer, with estimated new cases of 2.3 million (1). With the advancement of drug discovery, clinical trials, and the evolution of treatment concepts, the 5-year survival rate for breast cancer has risen to over 90%. However, recurrence and metastasis continue to be the cause of treatment failure of breast cancer (2, 3). Therefore, early diagnosis and treatment become an important procedure to improve survival and reduce the burden of patients.

Ductal carcinoma in situ (DCIS) is a proliferation of neoplastic luminal cells restricted to the ductolobular system of the breast and is also considered an early stage of cancer. Patients with DCIS who would not have caused symptoms and who have significant calcification on mammography are often diagnosed (4). Following the implementation of breast screening programs, the prevalence of DCIS has risen, accounting for 20–25% of all breast cancer cases (5, 6). Then, depending on the hormone receptor expression status, traditional treatment involves mastectomy or breast-conserving surgery, radiotherapy, and, in some cases, endocrine treatment (7, 8). Long-term outcomes of DCIS after management show high rates of local control with excellent overall survival (9–11). If left untreated, DCIS is expected to advance to invasive ductal carcinoma (IDC) in 20–50% of cases. It's still difficult to know which lesions will advance to IDC and which won't. In clinical practice, the precise biomarkers to classify at diagnosis and prognosis of DCIS are unclear (12–14).
With the advancement of genome-sequencing technologies, evidence has accumulated that differentially expressed genes have a great deal of promise in the diagnosis and prognosis of DCIS. A gene microarray profile can be examined to incorporate and can be a useful tool for identifying novel biomarkers to assist in the diagnosis and personalized treatment. This research was conducted by using microarray profile datasets which were obtained through gene expression omnibus (GEO) database to perform an integrated analysis of DCIS, and six genes were identified as possible biomarkers. The Cancer Genome Atlas Breast Cancer Dataset was used to confirm the expression of these genes, and enrichment analysis was used to clarify the biological role and intrinsic mechanism pathway. Additionally, in proliferation assays, the prognostic value of core genes was verified. Our findings indicated that the six genes are suitable to be used as biomarkers in the diagnosis of DCIS.

Materials And Methods

Gene expression microarray data

The study used the GSE7882 (15), GSE21422 (16), and GSE59246 (4) microarray profile datasets were obtained from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) for the studies. GSE7882 is based on the GPL5326 platform (NCI Qiagen Homo sapiens 36K v3 cgh expression), included 93 DCIS tissues and 7 benign epithelium tissues. GSE21422 is based on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array), which includes 9 DCIS samples and 5 healthy tissue samples. GSE59246 is based on the GPL13607 platform (Agilent-028004 SurePrint G3 Human GE 8x60K Microarray), included 46 DCIS samples and 3 noncancerous tissues.

DEGs identification

Differentially expressed genes (DEGs) between DCIS samples and noncancerous tissues were determined using GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/). In the GEO series datasets, GEO2R can compare two or more samples group using the analysis of variance or the t-test as an R programming language-based dataset analysis tool. Adjusted P-value < 0.05 and |log FC| > 1 were set as the cut-off criteria. BioDBnet (https://biodbnet-abcc.ncifcrf.gov/db/db2db.php) was used to convert identifiers from Gene ID to Gene symbol. DEGs were used to evaluate the overlapping genes in the three microarray profile datasets using FunRich software (version 3.1.3).

Pathway enrichment analysis and Gene ontology

The Database for Annotation, Visualization, and Integrated Discovery (DAVID, version 6.8, https://david.ncifcrf.gov/) and Funrich Software (version 3.1.3) was utilized to determine the biological role of candidate DEGs and possible pathway enrichment. DAVID is a bioinformatics resource on the web that can help with gene annotation, visualization, and integrated discovery. P-value < 0.05 represented as the cut-off criterion for pathway analysis and significant function.

PPI network construction and hub genes identification

To predict candidate DEGs protein-protein interaction (PPI) network, we used Search Tool for the Retrieval of Interacting Genes database (STRING, version 11.0, http://string-db.org), with 0.400 medium confidence and confidence network edges as the product criterion. The PPI network was then built and analyzed the candidate DEGs encoding protein interactions using Cytoscape software (version 3.7.2, http://www.cytoscape.org/). CytoHubba and MCODE, Cytoscape two plugins, were utilized to explore the hub genes of the PPI network and calculated node degree which is basically the number of interconnections to filter PPI hub genes.
Validation of the identified hub genes

To predict candidate DEGs protein-protein interaction (PPI) network, we used Search Tool for the Retrieval of Interacting Genes database (STRING, version 11.0, http://string-db.org), with 0.400 medium confidence and confidence network edges as the product criterion. The PPI network was then built and analyzed the candidate DEGs encoding protein interactions using Cytoscape software (version 3.7.2, http://www.cytoscape.org/). CytoHubba and MCODE, Cytoscape two plugins, were utilized to explore the hub genes of the PPI network and calculated node degree which is basically the number of interconnections to filter PPI hub genes.

Hub genes signatures and prognosis analysis

Using Oncomine databases (https://www.oncomine.org/resource/login.html), researchers looked at the transcription levels of hub and core genes in various cancer forms. Accompany with gene expression data of 32 types of cancers, Starbase (http://starbase.sysu.edu.cn/index.php) was utilized to estimate the effect of hub genes expression level in breast cancer. Hub gene expression was evaluated by UALCAN (http://ualcan.path.uab.edu/index.html) in molecular subtypes and nodal metastasis. The importance of prognosis of the known hub genes was evaluated using the Kaplan-Meier plotter (https://kmplot.com/analysis/), an online database capable of evaluating the impact of genes on survival in various cancer types.

Cell culture

The Chinese Academy of Sciences Cell Bank (Shanghai, China) provided the T47D, MCF10A, SK-BR-3, MDA-MB-231, MCF-7, BT549, and BT474 cell lines. MCF10A was cultured in DMEM (Dulbecco’s modified Eagle’s medium)/5% horse serum (Gibco, USA) supplemented F12. T47D and MCF-7 cells were grown in 10% fetal bovine serum (FBS) supplemented DMEM. SK-BR-3, BT549, and BT474 were maintained in 10% FBS supplemented RPMI-1640 medium. Cell culture of MDA-MB-231 cell lines was prepared in 10% FBS supplemented Leibovitz’s L-15 medium in the presence of CO₂ at 37°C.

RNA extraction and reverse transcription quantitative polymerase chain reaction (RT-qPCR)

Trizol reagent (Takara, Japan) was used to extract total RNA from cultured cells. For reverse transcription, the SuperScript RT kit (Takara, Japan) was used. The SYBR Green PCR package (Takara, Japan) was used for the RT-qPCR assay. The sequences of PCR primers are described in Supplementary Table S1. The calculation for the relative expression levels was performed by using the $2^{-\Delta\Delta Ct}$ method. The Ct values are basically the average of each gene in triplicate reactions.

Transfection and Small interfering RNA (siRNA)

siRNAs of hub genes and the appropriate scrambled control were procured by RiboBio (Shanghai, China). siRNAs target sequences are listed in Supplementary Table S2. By using the manual provided by the supplier, transfection of siRNAs with different cell lines was achieved by using FuGENE HD Transfection Reagent (Promega, Madison, WI, USA).

Proliferation assays

Cell proliferation potential was assessed using MTT, Colony formation, and EdU assays. MTT assay was conducted by seeding $2 \times 10^3$ cells 96-well plates after the transfection of 24h. Cells were added 20 μL MTT (0.5
mg/mL, Solarbio) after indicated time, then incubated for 4h at 37°C followed by the medium removal, and the formosan precipitate solubilization in 150 µL of DMSO. A microplate reader was used to test the activities of the viable cells at 570 nm. While $8 \times 10^2$ cells were seeded for 2 weeks in 6-well plates in the Colony formation assay. The colonies were fixed for 30 minutes in 4% paraformaldehyde before being stained with hematoxylin and counted and compared to a control group. The EdU test was conducted following the instructions of the manufacturer while using the Cell-Light Edu Apollo488 In Vitro Imaging Kit (Ribobio). EdU-positive cells percentage was calculated under the fluorescence microscope.

**Statistical analysis**

The results were presented as the mean ± SD of at least three independent experiments. The student's t-test was used to assess significance between the experimental and control groups. P < 0.05 was considered statistically significant. SPSS version 24.0 was used to perform all calculations.

**Results**

**Identification of overlapping DEGs in DCIS**

According to the cut-off criteria of P < 0.01 and $|\log FC| > 1$ for selecting DEGs, a total of 5586, 3042, 1757 DEGs were recognized as up-regulated, and 3532, 2917, 2409 DEGs were recognized as down-regulated from GSE7882, GSE21422 and GSE59246 microarray profile datasets (Fig. 1a-c), respectively. As shown in Fig. 1d and Fig. 1e, 110 up-regulated and 107 down-regulated genes overlapped across the three datasets. The names of the overlapping genes are shown in Table 1.

**Functional enrichment analysis of overlapped DEGs in DCIS**

GO enriched functions for the 217 overlapped DEGs were involved in various cellular components (CC), including cytoplasm, cell surface, nucleus, collagen type I trimer, and catalytic step 2 spliceosome for the up-regulated genes, and protein complex, nuclear speck, lamellipodium, and focal adhesion for the down-regulated genes (Fig. 2a and Table 2). Microtubule binding, procollagen-proline 4-dioxygenase activity, extracellular matrix structural constituent, ATP binding, and transcription corepressor activity were included in the up-regulated DEGs in terms of molecular function (MF), while transcriptional activator activity, transcription factor activity, and RNA polymerase II core promoter proximal region sequence-specific binding included in the down-regulated DEGs (Fig. 2b and Table 2). For the biological processes (BP) terms, the DEGs (up-regulated) were enriched for protein autophosphorylation, blood vessel development, RNA polymerase II promoter based negative transcription regulation, mitotic spindle assembly, and response to UV, while on the other hand, negative regulation in response to DNA damages of the inherent apoptotic signaling pathway, cellular response to calcium ion, renal tubule morphogenesis, microglial cell activation, and cell morphogenesis involved in neuron differentiation were all found to be down-regulated DEGs (Fig. 2c and Table 2).

Furthermore, the DEGs’ signaling pathways were enriched. The up-regulated genes were reported to be increased association with BH3-only protein activation, NGF signaling, the C-MYB transcription factor network, the Intrinsic Pathway for Apoptosis, and ERK signaling. The genes that were down-regulated were mostly involved in TRIF-mediated TLR3 signaling, CDC42 activity regulation, MAPK targets/Nuclear events mediated by MAP kinases, Toll Receptor Cascades, and Integrin-linked kinase signaling (Fig. 2d and Table 3).
Module analysis and PPI network construction

The DEGs overlapping revealed a unique set of networks and interactions. The online database of STRING was used to filter 178 DEGs (92 up-regulated and 86 down-regulated genes) from the 217 usually altered DEGs belonging to the PPI network complex. A total of 39 DEGs were excluded from the PPI network. Furthermore, by using Cytoscape software analysis, 456 edges were identified in overlapping DEGs. The degree of the PPI network complex defaulter filter ranged from 1 to 64 (Fig. 3a).

Furthermore, the entire PPI network was analyzed through the Cytoscape's MCODE plug-in, the most significant module and sixteen nodes were identified using degree cut-off = 2, k-core = 2, node score cut-off =0.2, and maximum depth = 100 as the criterion (Fig. 3b). Afterward, as shown in Fig. 3c and Fig. 3d, the first 27 PPI network genes were chosen using CytoHubba plug-in and nodes degree were analyzed. Twelve core candidate genes were chosen after combining the results of MCODE, CytoHubba, and nodes degree, and all of them were up-regulated DEGs, in the following order: GAPDH, CDH2, COL1A2, HNRNPA2B1, POLR2H, COL1A1, IDH2, NEK2, BIRC5, TRA2B, HNRNPH1, and MELK. They may have a significant impact on the progression of DCIS and the prognosis.

Kaplan-Meier survival analysis

The Kaplan-Meier survival plot was used to evaluate the prognostic details of the twelve core candidate genes. The twelve genes, including GAPDH, CDH2, COL1A2, HNRNPA2B1, POLR2H, COL1A1, IDH2, NEK2, BIRC5, TRA2B, HNRNPH1 and MELK, were used to plot the survival curves by uploading them to database. High expression of GAPDH (P = 7.9e-07), CDH2 (P = 0.0016), BIRC5 (P = 1.6e-08), NEK2 (P = 1.3e-06), IDH2 (P = 0.0076) and MELK (P = 9.2e-11) were correlated with poor patient's OS, while COL1A2 (P = 0.53), HNRNPA2B1 (P = 0.16), POLR2H (P = 0.051), COL1A1 (P = 0.28), TRA2B (P = 0.55) and HNRNPH1 (P = 0.46) expression were not relevant to OS (Fig. 4 and Fig. S1). Survival curves also showed that high expression of GAPDH (P = < 1e-16 and 2.9e-06), CDH2 (P = 0.011 and 0.0098), BIRC5 (P = < 1e-16 and 6e-10), NEK2 (P = < 1e-16 and 5.3e-10), IDH2 (P = 7.8e-14 and 0.00049) and MELK (P = < 1e-16 and 4.9e-14) were significantly associated with worse RFS and DMFS (Fig. 4).

Validation of prognostic effectiveness of the hub genes

Based on TCGA samples, the expression of the selected six hub genes with prognostic significance was further investigated. The findings revealed that hub gene expression was substantially increased in breast cancer tissues (Fig. S2). The six up-regulated hub genes were subjected to ROC research. These six hub genes' ROC curves all showed favorable prognostic values for DCIS. Moreover, the area under curve (AUC) of GAPDH, CDH2, BIRC5, NEK2, IDH2 and MELK were 0.8876, 0.7552, 0.7499, 0.8457, 0.7841 and 0.9664 (Fig. 5a-f), respectively.

Core gene signatures and prognostic analysis

DCIS has been linked to six core genes. Using Oncomine databases, researchers compared the transcriptional levels of core genes in pan-cancers to those in normal samples. In four to twenty-one datasets, the expression level of core genes mRNA was expressively up-regulated in breast cancer patients (Fig. S3). The increased expression of core genes in breast cancer samples has been illuminated, there were also differences in the up-regulated degree of core genes in diverse breast cancer molecular types. The expression of GAPDH, NEK2, BIRC5 and MELK was higher in triple-negative breast cancer, a subtype with the worst prognosis, CDH2, and IDH2 was higher in HER2 positive subtype. The up-regulated degree was higher in the subtype with a poor prognosis. Additionally, high expression of core genes also increases the risk of lymph node metastasis (Fig. 6a-f).
Downregulation of core genes expression inhibits the proliferation

T47D, MDA-MB-231, SK-BR-3, MCF-7, BT474, and BT549 are common breast cancer cells, MCF10A is the normal mammary epithelial cell. RT-qPCR was used to compare the core gene expressions in breast cancer and normal cell lines. The result shows that GAPDH, BIRC5, and MELK expressions were highest in MDA-MB-231, CDH2 was highly expressed in T47D, NEK2 was highly expressed in MCF-7 and IDH2 was highly expressed in SK-BR-3 (Fig. S4). Subsequently, transfection with siRNAs for the downregulation of core gene expression in the corresponding cells (Fig. S5). As shown in the figures, the cell proliferation of breast cancer was obviously inhibited when the expression of core genes was down-regulated through MTT (Fig. 7a-f), Colony formation (Fig. 8a-f), and EdU (Fig. 9a-f) assays. Data to support the use of a six-gene signature for DCIS diagnosis and prognosis prediction include GAPDH, CDH2, BIRC5, NEK2, IDH2, and MELK.

Discussion

DCIS is a heterogeneous disease that describes the stage of breast cancer before it becomes invasive (17). While the majority of DCIS patients have excellent long-term results, some DCIS patients can still develop invasive breast cancer. Regrettably, current clinical methods result in overtreatment of certain women with DCIS due to confusion about which lesions are at risk of progressing to invasive cancer. As a result, the identification of novel prognostic biomarkers is important. Gene signatures based on aberrant mRNAs have recently shown great promise in cancer prognosis prediction.

The researchers looked at the gene expression profiles of 148 DCIS patients and discovered 217 common DEGs, including 101 up-regulated and 138 down-regulated genes. According to the functional enrichment review, the DEGs were mostly associated with protein autophosphorylation, cytoplasm, microtubule-binding, negative regulation of intrinsic apoptotic signaling pathway in response to DNA damage, protein complex, transcriptional activator activity, and RNA polymerase II core promoter proximal region sequence-specific binding. Signaling pathway enrichment analysis is associated with activation of BH3-only proteins and TRIF mediated TLR3 signaling. With the help of the PPI network, twelve hub genes were chosen to be studied further. While GAPDH, CDH2, BIRC5, NEK2, IDH2, and MELK were found to be negative prognostic genes in DCIS patients. ROC and signatures analysis demonstrated that the core genes can be a useful indicator for DCIS. Additionally, downregulation of core genes expression by transfected with small interfering RNAs inhibits the proliferation of breast cancer cells significantly, also suggesting a great potential of utilizing the core genes in DCIS prognosis.

GAPDH, or glyceraldehyde-3-phosphate dehydrogenase, is a housekeeping gene that is often served as an internal control in experiments. Increased GAPDH levels, on the other hand, are seen in a range of human cancers and are often linked to shorter survival times (18–21). The evidence suggests that GAPDH function mechanisms, such as its role in cell survival of tumor, angiogenesis, and posttranscriptional regulation of tumor cell mRNA, are related to deprived prognosis and increased tumor progression for the affected individual (22, 23). Surprisingly, the role and mechanism of aberrant GAPDH in DCIS remain unknown. CDH2 (Cadherin 2) as a member of the cadherin superfamily, encodes the N-cadherin protein, plays an imperative role in EMT (Epithelial-mesenchymal transition). Elevated expression of CDH2 implicated poor prognosis in various cancers such as lung cancer (24), prostatic cancer (25) and glioma (26). Especially, CDH2 was found to be overexpressed in DCIS with invasion which may be an early marker in the absence of histological signs and as a marker of a short-term local recurrence after treatment (27). BIRC5 (also known as Survivin) is an apoptosis inhibitory protein that exerts a role in both the inhibition of cell death and the promotion of cancer cell survival (28). Researches demonstrate that BIRC5
expression is significantly increased in many cancers including lung, breast cancer, and colon cancers (29, 30). BIRC5 can be used as a predictor marker in different tumors due to its aggregation. As a result, increased survivin expression could be regarded as a prognostic marker associated with increased lymph node invasion, recurrence possibility, and metastasis (31, 32).

NEK2, Never in Mitosis (NIMA) Related Kinase 2, plays a key role in regulating microtubule stabilization, centrosome separation and duplication, spindle assembly checkpoint, and kinetochore attachment (33). Accumulated evidences have shown that the level of NEK2 was up-regulated in primary tumor tissues or cancer cell lines (34–36). Furthermore, increased NEK2 overexpression was linked with advanced tumor stage, distant metastases, and lymph node invasion suggesting that it may be used to predict tumor progression and disease prognosis (37–39). IDH2, isocitrate dehydrogenase 2, performs the oxidative decarboxylation of isocitrate to α-ketoglutarate (α-KG). IDH2 is the most commonly mutated metabolic gene in cancer, and it disrupts metabolic and epigenetic regulation, promoting tumorigenesis in humans (40). Interestingly, IDH2 frequently showed overexpression rather than a mutation in the following tumors, bladder, breast, and lung cancers. According to the findings of Li et al, up-regulated wild-type IDH2 promotes proliferation and tumor formation in the lung cancer cell, and is linked to a lower overall survival rate (41). IDH2 has been related to the recurrence of DCIS and progression to invasive disease and is expressed differently in recurrent and non-recurrent DCIS. Furthermore, high wild-type IDH2 expression was linked to a poor patient outcome in DCIS (42–44). According to microarray and TCGA datasets study, MELK, maternal embryonic leucine zipper kinase, expression is higher in many cancer cells and tissues than in their counterparts (45–47). MELK expression levels are also linked to high-grade tumors, increased aggressiveness, and poor patient outcomes (48–50). MELK has been recognized as an efficient therapeutic target and prognostic factor in the treatment of cancer, according to research.

It is important to look for promising biomarkers for DCIS diagnosis and prognosis prediction. Identifying a panel of deregulated genes that can increase the biomarker’s sensitivity and specificity rather than identifying individual genes is a crucial point to remember when researching genes as DCIS biomarkers. The six core genes, according to our findings, may be useful prognostic and diagnostic biomarkers for DCIS. However, more research into the expression and prognostic function of the six genes at the protein level is required. The underlying mechanism of the six genes must be clarified by functional experiments.

**Conclusions**

The six core genes signature for promising DCIS biomarkers was validated in our research, which may assist in clinical decision-making for individual care.

**Declarations**

**Acknowledgements**

None.

**Authors’ contributions**

Jiao Zhang and Hui Lin devised and deliberate the study, analyzed and collected the data, established the figures and tables, and wrote the manuscript. Lei Hou, Hui Xiao, and Xuhui Guo performed the data and statistical
analysis. Xuchen Cao and Zhenzhen Liu performed the revision of the manuscript. The final manuscript owns the approval of all authors.

**Funding**

None.

**Availability of data and materials**

This published article contains all of the data analyzed or produced during this research.

**Ethics approval and consent to participate**

None.

**Consent for publication**

Not applicable.

**Competing interests**

The writers announce that the manuscript does not contain any conflicting interests.

**References**


Tables

Table 1 From three profile datasets, 217 differentially expressed genes were discovered, including 110 co-up-regulated genes and 107 co-down-regulated genes in DCIS

<table>
<thead>
<tr>
<th>Genes Name</th>
<th>DEGs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated</td>
<td>ATP2A2, POLE2, MELK, MAPKAPK2, P4HA3, BCCOR, NEK2, TRIB3, VCAN, CXCL10, PDGEA7, DHC7, HOOK1, SLC20A1, TNC, PLEK2, B3GNT2, BDH1, CTWC14, PTPN31, HNRNPH1, FNA31, PAPC, CXXD3, EPICAM, AFG3L2, SNRPN40, MTA1P, RCP1, P4HA1, CDKN3, PHTF1, GTF3C2, AB2, PABPC1, NUP120, STIP1, UBE2A, EPDC1, POLR2H, CDH2, PAX5, YAF2, HNRNPA2, B1, CUTA, MAPK8, BIRC5, IDH2, SO2, DGKZ, EMP2, CALU, DTL, PPAT, RPN2, CD44, NOP2, GALNT3, NRSAP1, RXDF5, KAP4, SSX, CT5, COL1A1, DD3, ULK3, FXO2, SLC45A4, TRA2B, KYNU, ACSL3, GPD2, TFPD1, RAB11A, ASPF, TMPO, GFR2, MYB, CREM, BCL2L11, SI1, SLC35A2, BI1, SPDL1, MTDH, EFS, ADAM12, COL1A2, MDM2, CYFIP2, REL, ACVR1B, GAPDH, TNFSF4, NUDC, D1, ICA1, SRPK1, ZSCAN12, AREG, CCUX1, AREL1, RIT1, PAIP1, RNF4, TFEC, SLC35D1, FT80, AGR2, LMNB2</td>
</tr>
<tr>
<td>Down-regulated</td>
<td>PCM1, CA3, SNX6, LIMCH1, TSC22D3, SNX2, PEPD, MYLK, HDAC7, HYAL1, RAP1B, KLF4, COL4A1, ARID4A, AOS1, PALM, PHPF11, ACKR3, TMEM218, KL, DNAM1, ITF1, MTHFD1, FYN, CD14, ADCY6, LVE1, IL15RA, CTNN1A, TME1, MEM55A, FOI, LOF2, FKBP5, PTPN11, CDUN1D4, SLC40A1, LAMA3, RAB2CO, CDH13, TAB1, ING4, SCNA4, FF01, PDKX, DCC, METC, MEOC, WDAR45, SGK2, IDH38, HBB, HDHDS, KRT1, EPHA3, ME3, EHPB1, SGCB, DGKZ, SEC22A, RUNX1, T1, UGP2, PRP5, FOS, SGC, MAP1A, NCR1, HNMT, ACSS2, ATXN3, RBP4, OGG1, CASP4, T-EAD1, SORB3, MEF2C, ENNAD1, ASMTLZ, RXH1, SDON1CD, DCPID1, ACRT3, SCUBE2, RPS5, KA2, DYNCl2, NFU1, CLU, MYO1C, ST3GAL6, TGB5, CLND1, OXTR, TL6K, TMCC1, RASSF8, ATP1A2, HEXTA, USP24, SLC41A3, EXOC3, AADAC, PPBP, CYTH3, SNCA, HSD17B14, USP47, COLGALT2, IK, PCDH19</td>
</tr>
<tr>
<td>DEGs</td>
<td>Category</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0046777~protein autophosphorylation</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0001568~blood vessel development</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0000122~negative regulation of transcription from RNA polymerase II promoter</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0090307~mitotic spindle assembly</td>
</tr>
<tr>
<td>GOTERM_BP_DIRECT</td>
<td>GO:0009411~response to UV</td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0005737~cytoplasm</td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0009986~cell surface</td>
</tr>
<tr>
<td>Up-regulated</td>
<td>GOTERM_CC_DIRECT</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
</tr>
<tr>
<td></td>
<td>GOTERM_CC_DIRECT</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
</tr>
<tr>
<td></td>
<td>GOTERM_MF_DIRECT</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
</tr>
<tr>
<td></td>
<td>GOTERM_BP_DIRECT</td>
</tr>
<tr>
<td>Down-regulated</td>
<td>GOTERM_BP_DIRECT</td>
</tr>
<tr>
<td>regulated morphogenesis involved in neuron differentiation</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0043234~protein complex</td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0016607~nuclear speck</td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0030027~lamellipodium</td>
</tr>
<tr>
<td>GOTERM_CC_DIRECT</td>
<td>GO:0005925~focal adhesion</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0001077~transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding</td>
</tr>
<tr>
<td>GOTERM_MF_DIRECT</td>
<td>GO:0003700~transcription factor activity, sequence-specific DNA binding</td>
</tr>
</tbody>
</table>

Table 3 Signaling pathway enrichment analysis of DEGs in DCIS.
<table>
<thead>
<tr>
<th>DEGs</th>
<th>Biological pathway</th>
<th>Count</th>
<th>P-Value</th>
<th>Mapped gene names</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activation of BH3-only proteins</td>
<td></td>
<td>4</td>
<td>4.16E-05</td>
<td>MAPK8, TFDP1, BCL2L11, BID</td>
</tr>
<tr>
<td>Signalling by NGF</td>
<td></td>
<td>7</td>
<td>0.001</td>
<td>MAPKAPK2, TRIB3, MAPK8, NRAS, BCL2L11, MDM2, RIT1</td>
</tr>
<tr>
<td>C-MYB transcription factor network</td>
<td></td>
<td>5</td>
<td>0.003</td>
<td>PAX5, NRAS, MYB, COL1A2, TFEC</td>
</tr>
<tr>
<td>Up-regulated</td>
<td>Intrinsic Pathway for Apoptosis</td>
<td>4</td>
<td>0.004</td>
<td>MAPK8, TFDP1, BCL2L11, BID</td>
</tr>
<tr>
<td></td>
<td>Signalling to ERKs</td>
<td>3</td>
<td>0.005</td>
<td>MAPKAPK2, NRAS, RIT1</td>
</tr>
<tr>
<td></td>
<td>PI3K-Akt signaling pathway</td>
<td>9</td>
<td>0.01</td>
<td>FGFR2, NRAS, TNC, EFNA3, COL1A2, MDM2, COL1A1, MYB, BCL2L11</td>
</tr>
<tr>
<td></td>
<td>MicroRNAs in cancer</td>
<td>5</td>
<td>0.03</td>
<td>NRAS, CD44, TNC, MDM2, BCL2L11</td>
</tr>
<tr>
<td></td>
<td>ECM-receptor interaction</td>
<td>4</td>
<td>0.04</td>
<td>CD44, TNC, COL1A2, COL1A1</td>
</tr>
<tr>
<td></td>
<td>TRIF mediated TLR3 signaling</td>
<td>4</td>
<td>0.000</td>
<td>FOS, PTPN11, MEF2C, RPS6KA2</td>
</tr>
<tr>
<td></td>
<td>Regulation of CDC42 activity</td>
<td>5</td>
<td>0.003</td>
<td>HDAC7, KLF4, FOS, MITF, FYN</td>
</tr>
<tr>
<td></td>
<td>MAPK targets/ Nuclear events</td>
<td>3</td>
<td>0.005</td>
<td>FOS, MEF2C, RPS6KA2</td>
</tr>
<tr>
<td></td>
<td>Toll Receptor Cascades</td>
<td>5</td>
<td>0.005</td>
<td>FOS, CD14, PTPN11, MEF2C, RPS6KA2</td>
</tr>
<tr>
<td>Down-regulated</td>
<td>Integrin-linked kinase signaling</td>
<td>6</td>
<td>0.009</td>
<td>HDAC7, KLF4, FOS, MITF, FYN, FKBP5</td>
</tr>
<tr>
<td></td>
<td>Focal adhesion</td>
<td>6</td>
<td>0.025</td>
<td>COL4A1, LAMA3, FYN, TGB5, RAP1B, MYLK</td>
</tr>
<tr>
<td></td>
<td>Oxytocin signaling pathway</td>
<td>5</td>
<td>0.031</td>
<td>MEF2C, FOS, ADCY6, OXTR, MYLK</td>
</tr>
<tr>
<td></td>
<td>Pathways in cancer</td>
<td>8</td>
<td>0.037</td>
<td>FOS, COL4A1, LAMA3, MITF, ADCY6, RUNX1T1, MECOM, CTNNAT</td>
</tr>
</tbody>
</table>

**Figures**
Figure 1

Identification of the differentially expressed genes in DCIS. a-c Up-regulated and down-regulated genes are represented by purple and green points, respectively. In grey points, there are no genes with differentia expression a DEGs based on GSE7882. b DEGs based on GSE21422. c DEGs based on GSE59246. d Co-up-regulated genes from GSE7882, GSE21422, and GSE59246 microarray profile datasets. e Co-down-regulated genes from GSE7882, GSE21422, and GSE59246 microarray profile datasets.
Figure 2

Analysis for signaling pathway and biology function of DEGs in DCIS. GO enrichment of a CC, b MF, c BP, d KEGG enrichment of signaling pathway.
Figure 3

Module analysis and DEGs PPI network complex. a STRING and Cytoscape analysis based DEGs PPI network. The red nodes depict up-regulated genes, the light blue nodes represent down-regulated genes, and the lines represent the DEGs’ interactions. b The Cytoscape MCODE plug-in based identification of the most significant module was performed. c The first 27 genes were chosen using CytoHubba plug-in. The redder color represents the more front-ranking. d Genes name and high degree nodes values. The amount of additional nodes connected to a node is defined as a node degree.
Figure 4

Prognostic of core candidate genes in DCIS according to the Kaplan–Meier plotter database. a GAPDH (M33197_3_at). b CDH2 (203440_at). c BIRC5 (202095_s_at). d NEK2 (204641_at). e IDH2 (210046_s_at). f MELK (204825_at). Patients with high gene expression are represented by red lines, whereas those with low gene expression are exemplified by black lines. OS stands for overall survival. RFS stands for "relapse-free survival." Distance metastasis-free survival is abbreviated as DMFS.

Figure 5
Validation ROC curves of hub genes based on TCGA cohort. a ROC curves of GAPDH. b ROC curves of CDH2. c ROC curves of BIRC5. d ROC curves of NEK2. e ROC curves of IDH2. f ROC curves of MELK. Red represents sensitive curves; blue indicates identity lines. ROC, Receiver operating characteristic; AUC, Area under the curve; CI, Confidence interval.

Figure 6

Core genes expression overview according to the TCGA samples. Differences in candidate genes mRNA expression based on sample types, molecular subtypes, and lymph node metastasis status. a GAPDH. b CDH2. c BIRC5. d NEK2. e IDH2. f MELK.
Figure 7
Knockdown of core genes inhibits cell proliferation determined by MTT assay. a Down expression of GADPH inhibits MDA-MB-231 cell growth. b Down expression of CDH2 inhibits T47D cell growth. c Down expression of BIRC5 inhibits MDA-MB-231 cell growth. d Down expression of NEK2 inhibits MCF-7 cell growth. e Down expression of IDH2 inhibits SK-BR-3 cells growth. f Down expression of MELK inhibits MDA-MB-231 cells growth. The data were shown as mean ± SD of at least three independent experiments. Significance determination was conducted using Student’s t-test. **, p < 0.01; ***, p < 0.001.

Figure 8
Knockdown of core genes inhibits cell proliferation determined by colony formation assay. a Down expression of GADPH inhibits MDA-MB-231 cells colony formation. b Down expression of CDH2 inhibits T47D cells colony formation. c Down expression of BIRC5 inhibits MDA-MB-231 cells colony formation. d Down expression of NEK2 inhibits MCF-7 cells colony formation. e Down expression of IDH2 inhibits SK-BR-3 cells colony formation. f Down expression of MELK inhibits MDA-MB-231 cells colony formation. The data were shown as mean ± SD obtained from at least three independent experimentations. The Student’s t-test was used to determine significance ***, p < 0.001.
Figure 9

Knockdown of core genes inhibits cell proliferation determined by EdU assay. a Down expression of GADPH inhibits MDA-MB-231 cell growth. b Down expression of CDH2 inhibits T47D cell growth. c Down expression of BIRC5 inhibits MDA-MB-231 cell growth. d Down expression of NEK2 inhibits MCF-7 cell growth. e Down expression of IDH2 inhibits SK-BR-3 cells growth. f Down expression of MELK inhibits MDA-MB-231 cells growth. The data were shown as mean ± SD obtained from at least three independent experiments. Significance was determined by Student's t-test. **, p < 0.01.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTablesandFigures.docx