Prognostic Genes of Breast Cancer Identified by Gene Co-expression Network Analysis

Jianing Tang
Wuhan University Zhongnan Hospital
Yongwen Luo
Wuhan University Zhongnan Hospital
Gaosong Wu (wugaoson@163.com)
Wuhan University Zhongnan Hospital

Research

Keywords: Breast cancer, WGCNA, CASC5, RAD51

Posted Date: April 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-23458/v1

Version of Record: A version of this preprint was published at Frontiers in Oncology on September 11th, 2018. See the published version at https://doi.org/10.3389/fonc.2018.00374.

Abstract

Background

Breast cancer is the mostly diagnosed malignance in female worldwide. However, the mechanisms of its pathogenesis remain largely unknown.

Methods

In this study, we used weighted gene co-expression network analysis (WGCNA) to identify novel biomarkers associated with the prognosis of breast cancer. Gene expression profiles were obtained from the Gene Expression Omnibus (GEO) database.

Results

A total of 5 modules were identified via the average linkage hierarchical clustering. And a module significantly with the pathological grade was screened out. 33 genes with high connectivity in the clinically significant module were identified as hub genes. Among them, CASC5 and RAD51 were negatively associated with the overall survival and disease-specific survival. Similar results were observed in the validation dataset. Protein levels of CACS5 and RAD51 were also significantly higher in tumor tissues compared with normal tissues based on the analysis of the Human Protein Atlas. Convincingly, qRT-PCR analysis of breast cancer tissues and matched paracancerous tissue demonstrated that CACS5 and RAD51 were significantly upregulated in in breast cancer compared to paracancerous tissues. Further cell proliferation assay indicated that CACS5 and RAD51 depletion decreased cell proliferation capability.

Conclusion

In conclusion, our findings suggested that CASC5 and RAD51 could serve as biomarkers related to the prognosis of breast cancer and may be helpful for revealing pathogenic mechanism and developing further research.

Introduction

Breast cancer is the most common cancer and is the second leading cause of cancer death in women worldwide. Despite the incidence of breast cancer is dramatically higher in developed countries, almost half of newly diagnosed cases and approximately 60% of deaths are occurring in developing world(1). Survival rates of breast cancer are largely different in the world. In developed countries, 5-year survival was estimated at 80%, while it is thought to below 40% for developing countries(2). Breast cancer-related deaths have been decreased in recent decades as a result of improving strategies to diagnose and

treatment. However, a minority of breast cancer patients initially diagnosed with advanced stage are nearly incurable. Some patients present early-stage disease suffer distant or locoregional recurrence(3). Thus, it is important to understand the underlying mechanisms during the initiation and progression of breast cancer.

Breast cancer is a disease with great heterogeneity. Based on the expression of the three biomarkers, estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2), molecular studies have demonstrated that breast cancer can be categorized into three subtypes: luminal (hormone receptor positive), HER2-enriched (high expression of HER2 gene), and triple-negative (ER-negative, PR-negative, HER2-negative). Luminal type of breast cancer is characterized by the expression of estrogen receptor and progesterone receptor. Patients with luminal breast cancer often have the best prognosis due to the application of endocrine treatment, while half of these patients will develop acquired resistance. Triple-negative and HER2-enriched breast cancers have more aggressive behaviors. Patients with triple-negative and HER2-enriched breast cancers are thought to have poorer prognosis, with trend to early relapse and metastatic spread to the liver, lung and central nervous system(4, 5). BRCA1 and BRCA2 are used to assess the risk of inherited breast cancer. The risks of breast cancer associated with BRCA1 and BRCA2 mutations are 65% and 45% in population, while in cancer-prone families, the risks are estimated to be 87% and 84%(6, 7).

Based on the histopathological features of breast cancer, such as tumor size, lymph node status and grade, patients could be stratified into high- and low-risk groups of recurrence and mortality(8). However, even in patients with histologically similar tumors, clinical outcomes of breast cancer patients are largely different due to the heterogeneity of breast cancer(9). Several studies have established multiple gene prognostic signatures to predict the prognosis of patients with breast cancer. Some prognostic models have been validated and are in clinical application(10–12). It is important to detect these biomarkers to support early diagnosis, therapeutic strategies determination and prognosis prediction after treatment.

WGCNA (Weighted Gene Co-expression Network Analysis) is a systems biology method to analyze the correlation patterns among genes(13). It can be used to identify candidate biomarker genes or therapeutic targets according to the interconnectedness of gene sets and the association between gene sets and phenotypes. In the present study, we used WGCNA to explore candidate biomarkers of patients with breast cancer.

Materials And Methods

Data Processing

Gene expression profiles and clinical data of breast cancer patients (GSE37751) were downloaded from the GEO database. The GSE37751 was based on GPL6244 platform (Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]). This dataset included 47 normal breast tissues and 61 breast cancer tissues. We used the Robust Multi-array Average (RMA) method to normalize raw microarray datasets,

including background correction, log2 transformation and normalization. Probes were changed into gene symbols using corresponding annotation files. the expression value of each gene was compared between cancer samples and normal samples to identify differentially expressed genes (DEGs) by Linear Models for Microarray Data (LIMMA) package. The False Discovery Rate method was used to adjust the p-value. Adjusted P-value < 0.05 and |log2 foldchange (FC)| > 1 were set as the cutoff criteria to select genes for further network construction.

Construction Of Co-expression Network

The gene co-expression networks of DEGs were constructed using R WGCNA package. In the present study, an appropriate soft-threshold power β was selected to build a weighted adjacency matrix. We calculated the network connectivity of genes by transforming the adjacency matrix into topological overlap matrix (TOM). In order to classify the genes with similar expression profiles into gene modules, the average linkage hierarchical clustering was performed according to the TOM-based dissimilarity measure with a minimum size of 30 for the genes dendrogram.

Identification Of Clinically Significant Modules

Module eigengenes (MEs) are the first principle component of each gene module and were considered as the expression patterns of all genes within a given module. Correlations between clinical trait and MEs were calculated to identify the clinically significant module. In addition, module significance (MS) of each module was calculated. The higher absolute value of module significance of a given module was considered as more biologically significant.

Pathway Enrichment And Functional Analyses

In order to identify the biological function of the clinically significant module, gene ontology and KEGG pathway enrichment analyses were performed using a functional annotation tool in the database for annotation, visualization and integrated discovery (DAVID, http://david.abcc.ncifcrf.gov/, version 6.8). Adjusted P < 0.05 was considered statistically significant.

Identification And Validation Of Hub Genes

Hub genes are thought to be highly interconnected with nodes in a module and are more likely to have functional significance. Module connectivity of each gene was measured by absolute value of the module membership (MM). And the absolute value of gene significance (GS) was also calculated to identify the relationship between genes and clinical trait. Hub genes are considered to have higher absolute value of MM and GS. In the present study, absolute MM > 0.8 and absolute GS > 0.4 were defined as hub genes in co-expression network. GSE21653 was downloaded to validated the prognostic value of

our hub genes. Survival analysis was also performed using Kaplan Meier-plotter (www. kmplot.com)(14). Immunohistochemistry data were obtained to validate protein levels of candidate hub genes using the Human Protein Atlas (http://www.proteinatlas.org)(15).

Validation In Human Breast Cancer Samples

25 paired breast cancer and paracancerous tissues samples were collected from patients undergoing modified radical mastectomy at Zhongnan Hospital of Wuhan University. All the breast tumors samples were confirmed by two pathological specialists independently. The samples were immediately frozen and stored in liquid nitrogen. We isolated total RNA from breast cancer and paracancerous tissue samples and performed qRT-PCR analysis to validate the expression of prognostic genes in human samples. This study was reviewed and approved by the Ethical Board at the Zhongnan Hospital of Wuhan University with written informed consent from all the patients

Proliferation Analysis

Breast cancer cell line (MDA-MB-231) was used to identify the biological functions of hub genes. Cells were transfected with 50 nM small interfering RNAs using Lipofectamin RNAiMAX (Invitrogen, Carlsbad, CA, USA). Cell viability was measured using Cell Counting Kit-8 (CCK8) every 24 h.

Statistical Analysis

Kaplan-Meier method with log-rank test and Cox regression model were used to perform survival analysis. We separated patients into low- and high- expression groups based on the median expression value of each hub gene. Student's t test and one-way ANOVA were used to compare the difference between two or more groups. We performed all statistical analysis using R software 3.6.1 and GraphPad Prism software version 7.0 (GraphPad Software, San Diego, CA, USA), a $P < 0.05$ was set as the probability value of statistical significance.

Result

Weighted co-expression network construction and key modules identification

Based on the threshold of adjusted P-value < 0.05 and |log2 foldchange (FC)| > 1, a total of 3668 DEGs (2756 up-regulated and 912 down-regulated) were selected for WGCNA analysis. Nine tumor samples without complete clinical information and 47 normal samples were removed from our subsequent co-expression analysis. To assess the microarray quality, we performed sample cluster of GSE37751 in Pearson’s correlation matrices and average linkage method (Fig. 1A**)**. Fifty-two samples were included in co-expression analysis. In this study, we selected the power of $\beta = 16$ (scale free $R^2 = 0.87$) as the soft-

thresholding to ensure a scale-free network (Fig. 1B, C). A total of 5 modules were identified via the average linkage hierarchical clustering. Turquoise module was found to have the highest association with pathological grade (Fig. 2), which was selected as the clinically significant module to be studied for further analysis.

Gene Ontology And Pathway Enrichment Analysis

In order to identify the biological functions of turquoise module, genes in this module were categorized into 3 functional groups: biological process (BP), molecular function (MF) and cellular component (CC). In the BP group, turquoise module genes were mainly enriched in cell division, mitotic nuclear division, DNA replication, sister chromatid cohesion, mitotic nuclear division and G1/S transition of mitotic cell cycle. In the MF group, they were mainly enriched in protein binding, ATP binding, histone binding, DNA binding and protein heterodimerization activity. In the CC group, they were significantly enriched in nucleoplasm, nucleus, nucleosome, kinetochore and cytosol (Fig. 3A-C). KEGG pathway analysis demonstrated that these genes were mainly involved in cell cycle, DNA replication, alcoholism, systemic lupus erythematosus and Fanconi anemia pathway (Fig. 3D). These results showed that genes in the clinically significant module were mainly involved in mitotic cell cycle process.

Identification and validation of hub genes

Based the cut-off criteria (|MM| > 0.8 and |GS| > 0.4), a total of 33 genes with high connectivity in the clinically significant module were identified as hub genes. Among them, CASC5 and RAD51 were negatively associated with the overall survival and disease-specific survival. Consistently, high expression of CASC5 and RAD51 indicated poor relapse-free survival in GSE21653 (Fig. 4). The same results were observed in Kaplan Meier-plotter database (Fig. 5). In GSE37751, CASC5 and RAD51 were upregulated in triple-negative breast cancer. Higher expression levels of CASC5 and RAD51 were associated with advanced tumor stage and grade (Fig. 6). Similar results were found in the validation set GSE21653(Fig. 7). Protein levels of CACS5 and RAD51 were also significantly higher in tumor tissues compared with normal tissues based on the Human Protein Atlas (Fig. 8). Convincingly, we performed qRT-PCR using breast cancer tissues and matched paracancerous tissue to validate their expression levels. CACS5 and RAD51 were significantly upregulated in in breast cancer compared to paracancerous tissues. Further cell proliferation assay demonstrated that CACS5 and RAD51 depletion decreased cell proliferation capability (Fig. 9). To further analyze the function of CACS5 and RAD51, GSEA was conducted to search KEGG pathways enriched in CACS5 or RAD51 highly expressed samples. Based on the cut-off criteria (FDR < 0.05), 3 functional gene sets were enriched in CACS5 highly expressed samples: “Proteasome”, “Spliceosome”, “RNA polymerase”, and “Oxidative phosphorylation”. Six functional gene sets were enriched in RAD51 highly expressed samples: “Pyrimidine metabolism”, “Spliceosome”, “Oxidative phosphorylation”, “RNA polymerase”, “DNA replication”, and “Basal transcription factors” (**Figure S1)**.

Discussion

Breast cancer is the most frequent malignant tumor in females. It is still easy to recur even after combined therapy. In the era of precise medicine, microarray has been widely used to analyze the expression changes of mRNA in breast cancer and predict the potential therapeutic targets. However, better biomarkers for cancer specific prognosis and progression were still required.

In the present study, we performed WGCNA to explore gene co-expression modules associated with progression of breast cancer. A total of 3668 DEGs were used to construct co-expression network and 5 modules were identified. Turquoise module was found to have the highest association with tumor grade, ER status and triple-negative tumor. Thirty-three genes with high connectivity were screened from the module. Among them, CASC5 and RAD51 were negatively associated with the prognosis of patients.

CASC5, also known as D40, encodes a protein that functions as a scaffold for proteins influencing the spindle assembly checkpoint during the eukaryotic cell cycle. It is required for creation of kinetochore-microtubule attachments and chromosome segregation. CASC5 is widely expressed in various cultured human cancer cell lines and primary tumors of different origins. Yuri N et al. characterized CASC5 as a member of the cancer/testis gene family, and CASC5 knockdown significantly inhibited the growth of human cancer cell lines both *in vitro* and *in vivo*(16). In poorly differentiated primary lung cancer, CASC5 expression level was significantly higher. It was the first gene in cancer/testis gene family for which expression is associated with smoking habits of lung cancer patients(17). In patients with malignancies, some of the cancer/testis genes could encode antigens on tumor cells and cause immune responses by cytotoxic T cells which were called cancer/testis antigens. However, it was not clear whether CASC5 protein elicits immune responses in breast cancer patients(18, 19). A sequence homology search of public database revealed that the CASC5 sequence was identical to a gene on human chromosome 15, AF15q14. AF15q14 is a partner that fuses to mixed-lineage leukemia genes, which were involved in the development of acute leukemias(20–22). In dataset GSE37751, CASC5 was highly expressed in breast cancer tissues compared with normal breast tissues. In addition, ROC curve indicated that CASC5 exhibited excellent diagnostic efficiency for normal and tumor tissues. One-way ANOVA and t test demonstrated that the expression level of CASC5 was higher in triple-negative tumor and associated with tumor progression. Besides, the Oncomine database also showed that the expression level of CASC5 was significantly higher in breast cancer samples. To obtain further insight of translational level of CASC5, we used the Human Protein Atlas database to examine the immunohistochemistry staining of CASC5 in both normal breast and breast cancer, and discovered that the protein level of CASC5 was significantly up-regulated in breast cancer tissues compared with normal breast tissues. Survival analysis revealed that high expression of CASC5 was associated with the worse overall survival and relapse free survival. CASC5 has the potential to be a prognostic biomarker. GSEA demonstrated that high expression of CASC5 was associated with the pathway of proteasome, spliceosome, RNA polymerase and oxidative phosphorylation. The ubiquitin-proteasome system (UPS) plays an important role in the degradation of proteins. Increasing evidence indicated that the UPS involved in development cancer development(23, 24). Proteasomes participate in many cellular degradation processes, including inflammation, cell cycle,

responses to oxidative stress, and regulation of gene expression(24, 25). The proteasome inhibitor bortezomib (N-acyl-dipeptidyl boronicacid) was approved by the FDA for the treatment of mantle cell lymphoma and relapsed multiple myeloma (MM). Alternative splicing generates different, even antagonistic products from a single gene locus. This prevalent process greatly expands the coding capacity of complex genomes(26, 27). Spliceosome plays an important role in cancer progression including the control of cell proliferation and programmed cell death, angiogenesis, metabolism of cancer cells and metastasis. Misregulation of alternative splicing patterns often cause or modify human disease, including cancer(28). RNA polymerase is required for tumorigenesis, it was reported that MYC's tumorigenic potential is regulated by polymerase I-III(29). Some studies demonstrated a reduction of oxidative phosphorylation capacity in different types of cancer cells, while other investigations revealed contradictory modifications with the upregulation of oxidative phosphorylation components and a larger dependency of cancer cells on oxidative energy substrates for anabolism and energy production. Guppy and colleagues stated that breast cancer cells generated 80% of their ATP by the mitochondrion, introducing the concept of oxidative tumors(30). It was reported that the typical “glycolytic” type of cancer cells includes enhanced glycolytic machinery confronted to a low efficiency oxidative phosphorylation system, while the “oxidative phosphorylation” type of cancer cells relies mainly on mitochondrial respiration to produce ATP from glucose and glutamine oxidation(31–33).

The protein encoded by RAD51 is known to be involved in the repair and homologous recombination of DNA. RAD51 interact with BRCA1 and BRCA2, which regulate both the intracellular localization and DNA-binding ability of this protein. Loss of these controls may be a key event causing genomic instability and tumorigenesis(34, 35). In breast cancer, RAD51 expression correlates with high-grade metastatic breast tumor and poor prognosis. Cells with RAS51 overexpress exhibit disruption of cell cycle, resistance to apoptotic signals and associated resistance to radiotherapy and chemotherapy(36, 37). In dataset GSE37751, RAD51 was highly expressed in breast cancer tissues compared with normal breast tissues. In addition, ROC curve indicated that RAD51 exhibited excellent diagnostic efficiency for normal and tumor tissues. One-way ANOVA and t test demonstrated that RAD51 was highly expressed in triple-negative tumor and associated with high-grade tumor. Oncomine database showed that the expression level of RAD51 was significantly higher in breast cancer samples. Human Protein Atlas database with the immunohistochemistry staining of RAD51 was used to further investigate the translational level of RAD51. The protein level of RAD51 was significantly up-regulated in breast cancer tissues compared with normal breast tissues. Survival analysis revealed that high expression level of RAD51 was associated with the worse overall survival and relapse free survival. GSEA demonstrated that high expression level of RAD51 was associated with the pathway of pyrimidine metabolism, spliceosome, oxidative phosphorylation, RNA polymerase, DNA replication, and basal transcription factors.

Conclusion

In conclusion, our WGCNA analysis revealed that CASC5 and RAD51 were associated with the progression and poor prognosis of breast. Enrichment analysis demonstrated that CASC5 and RAD51 might promote tumor progression via spliceosome, RNA polymerase and oxidative phosphorylation

pathways. Our results may be helpful for revealing pathogenic mechanism and developing further research.

Declarations

Ethics approval and consent to participate

The research was carried out according to the World Medical Association Declaration of Helsinki and was approved by the Ethics Committee at Zhongnan Hospital of Wuhan University.

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

The authors have no conflicts of interest

Funding

None

Authors' contributions

JT and GW reviewed relevant literature and drafted the manuscript. JT and YL conducted all statistical analyses. All authors read and approved the final manuscript.

Acknowledgements

None

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal. A 2018 Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians.
2. Coleman MP, Quaresma M, Berrino F, Lutz JM, De Angelis R, Capocaccia R, Baili P, Rachet B, Gatta G, Hakulinen T, Micheli A, Sant M, Weir HK, Elwood JM, Tsukuma H, Koifman S, GA ES, Francisci S, Santaquilani M, Verdecchia A, Storm HH. Young JL 2008 Cancer survival in five continents: a worldwide population-based study (CONCORD). The Lancet Oncology **9**:730–756.

3. O'Shaughnessy J. 2005 Extending survival with chemotherapy in metastatic breast cancer. The oncologist **10 Suppl 3**:20–29.
4. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P. Narod SA 2007 Triple-negative breast cancer: clinical features and patterns of recurrence. Clinical cancer research: an official journal of the American Association for Cancer Research 13:4429–4434.
5. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L. 2001 Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. The New England journal of medicine **344**:783–792.
6. Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE. 1994 Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. Lancet (London, England) **343**:692–695.
7. Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, Sobol H, Teare MD, Struewing J, Arason A, Scherneck S, Peto J, Rebbeck TR, Tonin P, Neuhausen S, Barkardottir R, Eyfjord J, Lynch H, Ponder BA, Gayther SA, Zelada-Hedman M, et al. 1998 Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. American journal of human genetics **62**:676–689.
8. Boughey JC, Ballman KV, McCall LM, Mittendorf EA, Symmans WF, Julian TB, Byrd D, Hunt KK. 2017 Tumor Biology and Response to Chemotherapy Impact Breast Cancer-specific Survival in Node-positive Breast Cancer Patients Treated With Neoadjuvant Chemotherapy: Long-term Follow-up From ACOSOG Z1071 (Alliance). Ann Surg **266**:667–676.
9. 2012 Comprehensive molecular portraits of human breast tumours. Nature **490**:61–70.
10. Paik S. 2007 Development and clinical utility of a 21-gene recurrence score prognostic assay in patients with early breast cancer treated with tamoxifen. The oncologist **12**:631–635.
11. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH. Bernards R 2002 A gene-expression signature as a predictor of survival in breast cancer. The New England journal of medicine **347**:1999–2009.
12. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, d'Assignies MS, Bergh J, Lidereau R, Ellis P, Harris A, Bogaerts J, Therasse P, Floore A, Amakrane M, Piette F, Rutgers E, Sotiriou C, Cardoso F. Piccart MJ 2006 Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. Journal of the National Cancer Institute **98**:1183–1192.
13. Langfelder P, Horvath S. 2008 WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics **9**:559.
14. Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z. 2010 An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res Treat **123**:725–731.

15. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F, Sanli K, von Feilitzen K, Oksvold P, Lundberg E, Hober S, Nilsson P, Mattsson J, Schwenk JM, Brunnstrom H, Glimelius B, Sjoblom T, Edqvist PH, Djureinovic D, Micke P, Lindskog C, Mardinoglu A. Ponten F 2017 A pathology atlas of the human cancer transcriptome. Science (New York, NY) **357**.
16. Urata YN, Takeshita F, Tanaka H, Ochiya T, Takimoto M. 2015 Targeted Knockdown of the Kinetochore Protein D40/Knl-1 Inhibits Human Cancer in a p53 Status-Independent Manner. Scientific reports **5**:13676.
17. Takimoto M, Wei G, Dosaka-Akita H, Mao P, Kondo S, Sakuragi N, Chiba I, Miura T, Itoh N, Sasao T, Koya RC, Tsukamoto T, Fujimoto S, Katoh H, Kuzumaki N. 2002 Frequent expression of new cancer/testis gene D40/AF15q14 in lung cancers of smokers. British journal of cancer **86**:1757–1762.
18. Chen YT, Old LJ. 1999 Cancer-testis antigens: targets for cancer immunotherapy. The cancer journal from Scientific American **5**:16–17.
19. Sasao T, Itoh N, Takano H, Watanabe S, Wei G, Tsukamoto T, Kuzumaki N, Takimoto M. 2004 The protein encoded by cancer/testis gene D40/AF15q14 is localized in spermatocytes, acrosomes of spermatids and ejaculated spermatozoa. Reproduction (Cambridge, England) **128**:709–716.
20. Chinwalla V, Chien A, Odero M, Neilly MB, Zeleznik-Le NJ, Rowley. JD 2003 A t(11;15) fuses MLL to two different genes, AF15q14 and a novel gene MPFYVE on chromosome 15. Oncogene **22**:1400–1410.
21. Hayette S, Tigaud I, Vanier A, Martel S, Corbo L, Charrin C, Beillard E, Deleage G, Magaud JP. Rimokh R 2000 AF15q14, a novel partner gene fused to the MLL gene in an acute myeloid leukaemia with a t(11;15)(q23;q14). Oncogene **19**:4446–4450.
22. Kuefer MU, Chinwalla V, Zeleznik-Le NJ, Behm FG, Naeve CW, Rakestraw KM, Mukatira ST, Raimondi SC. Morris SW 2003 Characterization of the MLL partner gene AF15q14 involved in t(11;15)(q23;q14). Oncogene **22**:1418–1424.
23. Ding F, Xiao H, Wang M, Xie X, Hu F. 2014 The role of the ubiquitin-proteasome pathway in cancer development and treatment. Frontiers in bioscience (Landmark edition) **19**:886–895.
24. Fuchs O, Neuwirtova R. 2006 [Ubiquitins, proteasomes, sumoylation and application today and in future for cancer and other diseases therapy II. Sumoylation and neddylation as posttranslational modifications of proteins and their ubiquitinylation and its significance]. Vnitrni lekarstvi **52**:619–627.
25. Lu JC, Piazza TM, Schuler LA. 2005 Proteasomes mediate prolactin-induced receptor down-regulation and fragment generation in breast cancer cells. The Journal of biological chemistry **280**:33909–33916.
26. Blencowe BJ. 2006 Alternative splicing: new insights from global analyses. Cell 126:37–47.
27. Nilsen TW, Graveley. BR 2010 Expansion of the eukaryotic proteome by alternative splicing. Nature **463**:457–463.

28. Bonnal S, Vigevani L, Valcarcel J. 2012 The spliceosome as a target of novel antitumour drugs. Nature reviews Drug discovery **11**:847–859.
29. Campbell KJ, White RJ. 2014 MYC regulation of cell growth through control of transcription by RNA polymerases I and III. Cold Spring Harbor perspectives in medicine **4**.
30. Guppy M, Leedman P, Zu X, Russell V. 2002 Contribution by different fuels and metabolic pathways to the total ATP turnover of proliferating MCF-7 breast cancer cells. The Biochemical journal **364**:309–315.
31. Rodriguez-Enriquez S, Carreno-Fuentes L, Gallardo-Perez JC, Saavedra E, Quezada H, Vega A, Marin-Hernandez A, Olin-Sandoval V, Torres-Marquez ME, Moreno-Sanchez. R 2010 Oxidative phosphorylation is impaired by prolonged hypoxia in breast and possibly in cervix carcinoma. The international journal of biochemistry & cell biology **42**:1744–1751.
32. Reitzer LJ, Wice BM, Kennell D. 1979 Evidence that glutamine, not sugar, is the major energy source for cultured HeLa cells. The Journal of biological chemistry **254**:2669–2676.
33. Jose C, Bellance N, Rossignol R. 2011 Choosing between glycolysis and oxidative phosphorylation: a tumor's dilemma? Biochimica et biophysica acta **1807**:552–561.
34. Chappell WH, Gautam D, Ok ST, Johnson BA, Anacker DC, Moody CA. 2015 Homologous Recombination Repair Factors Rad51 and BRCA1 Are Necessary for Productive Replication of Human Papillomavirus 31. Journal of virology **90**:2639–2652.
35. Brown ET, Holt JT. 2009 Rad51 overexpression rescues radiation resistance in BRCA2-defective cancer cells. Molecular carcinogenesis **48**:105–109.
36. Sarasin A, Kauffmann A. 2008 Overexpression of DNA repair genes is associated with metastasis: a new hypothesis. Mutation research **659**:49–55.
37. Schild D, Wiese C. 2010 Overexpression of RAD51 suppresses recombination defects: a possible mechanism to reverse genomic instability. Nucleic acids research **38**:1061–1070.

Supplementary File Captions

Figure S1. Gene set enrichment analysis (GSEA). (A). The gene set of “Proteasome”, “Spliceosome”, “RNA polymerase”, and “Oxidative phosphorylation” were significantly enriched in CASC5 highly-expressed samples. (B) The gene set of “Pyrimidine metabolism”, “Spliceosome”, “Oxidative phosphorylation”, “RNA polymerase”, “DNA replication”, and “Basal transcription factors” were significantly enriched in RAD51 highly-expressed samples.

Figures

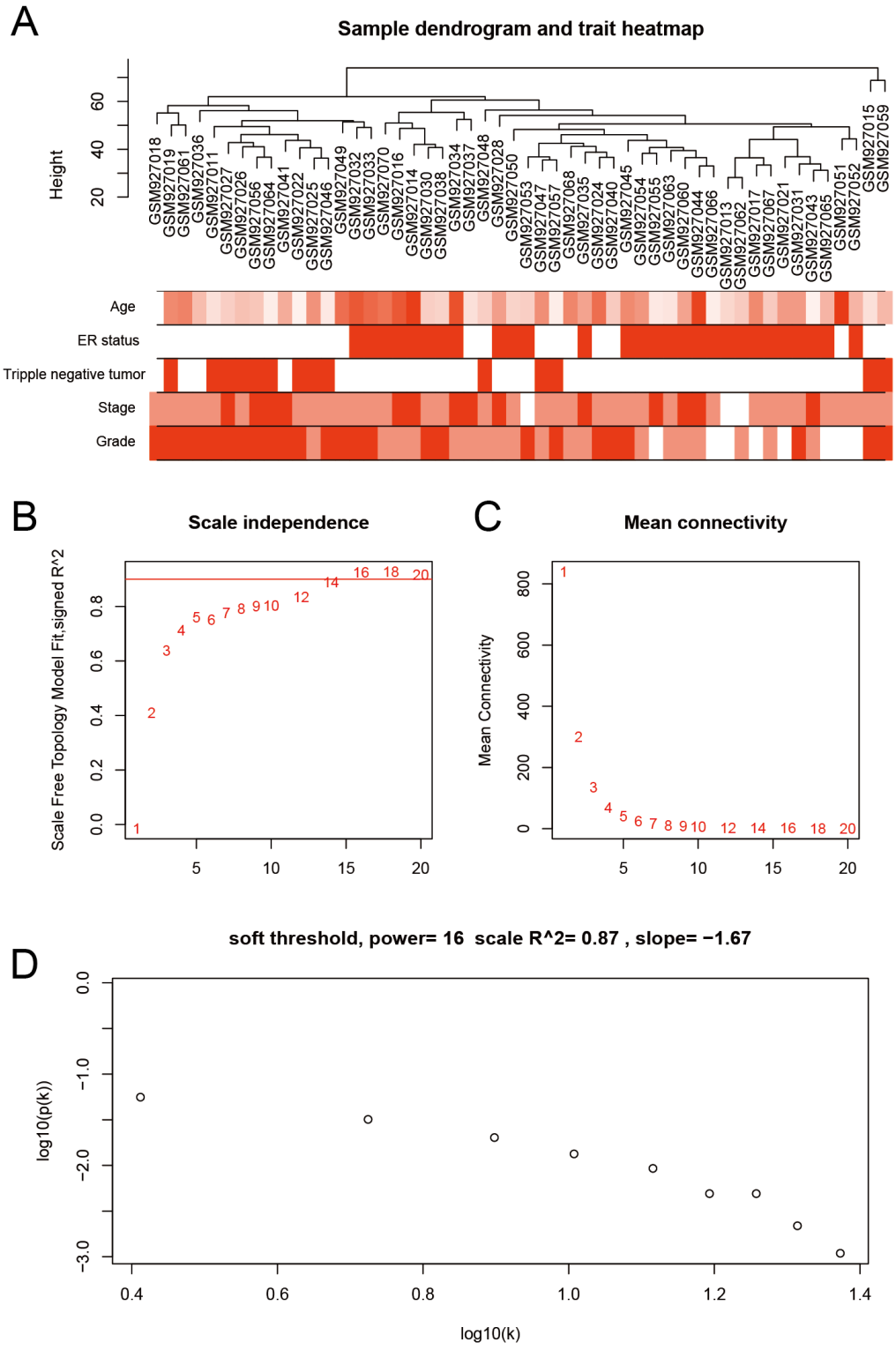

Figure 2

Clustering dendrogram and determination of soft-thresholding power in the WGCNA. (A). Clustering dendrogram. (B). Analysis of the scale-free fit index for various soft-thresholding powers (β). (C). Analysis of the mean connectivity for various soft-thresholding powers. (D). Checking the scale free topology when β = 16.

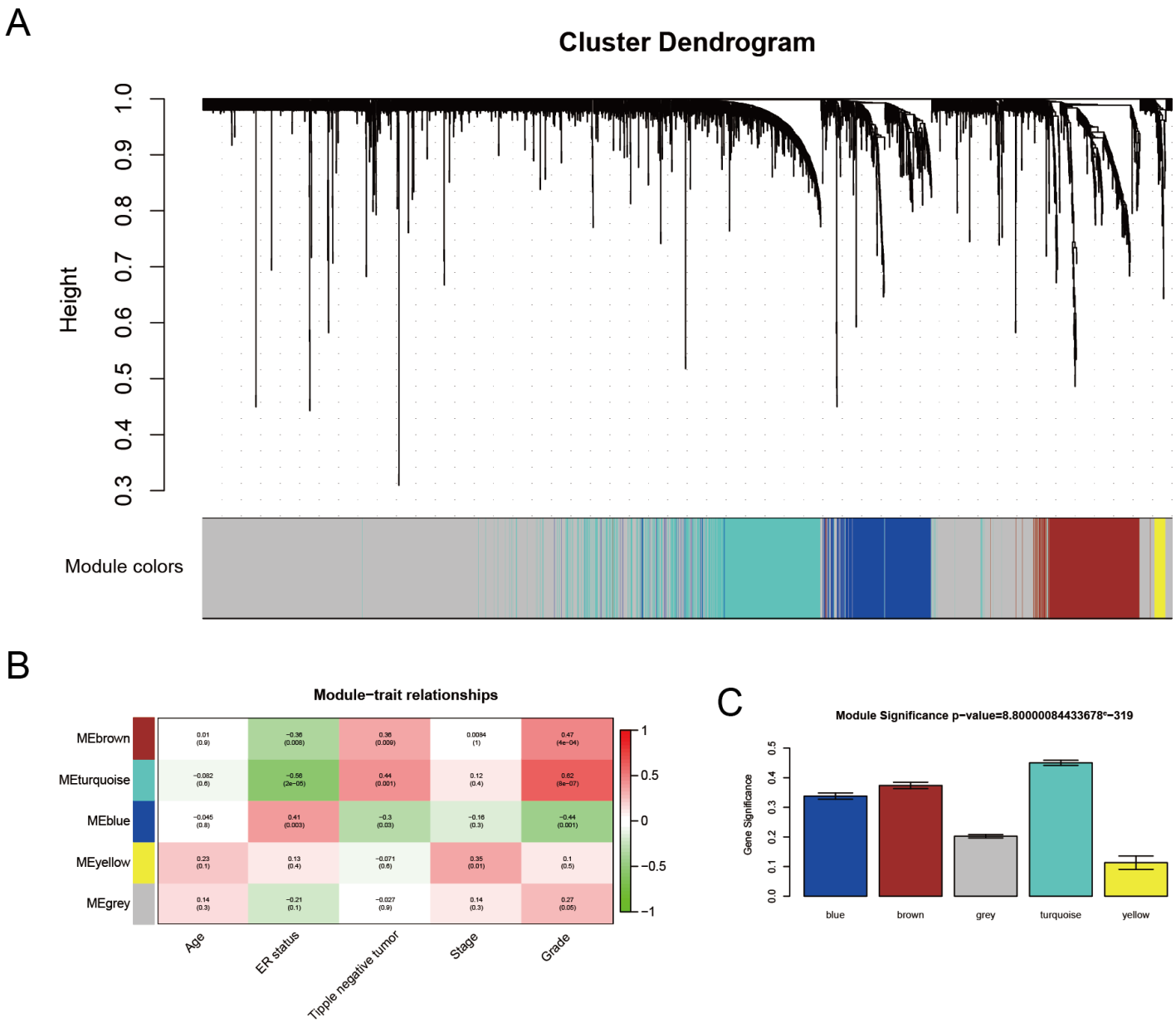

Figure 4

Identification of modules associated with the clinical traits of breast cancer. (A). Dendrogram of all differentially expressed genes clustered based on a dissimilarity measure (1-TOM). (B). Heatmap of the correlation between module eigengenes and clinical traits of breast cancer. (C). Distribution of average gene significance and errors in the modules associated with tumor grade of breast cancer.

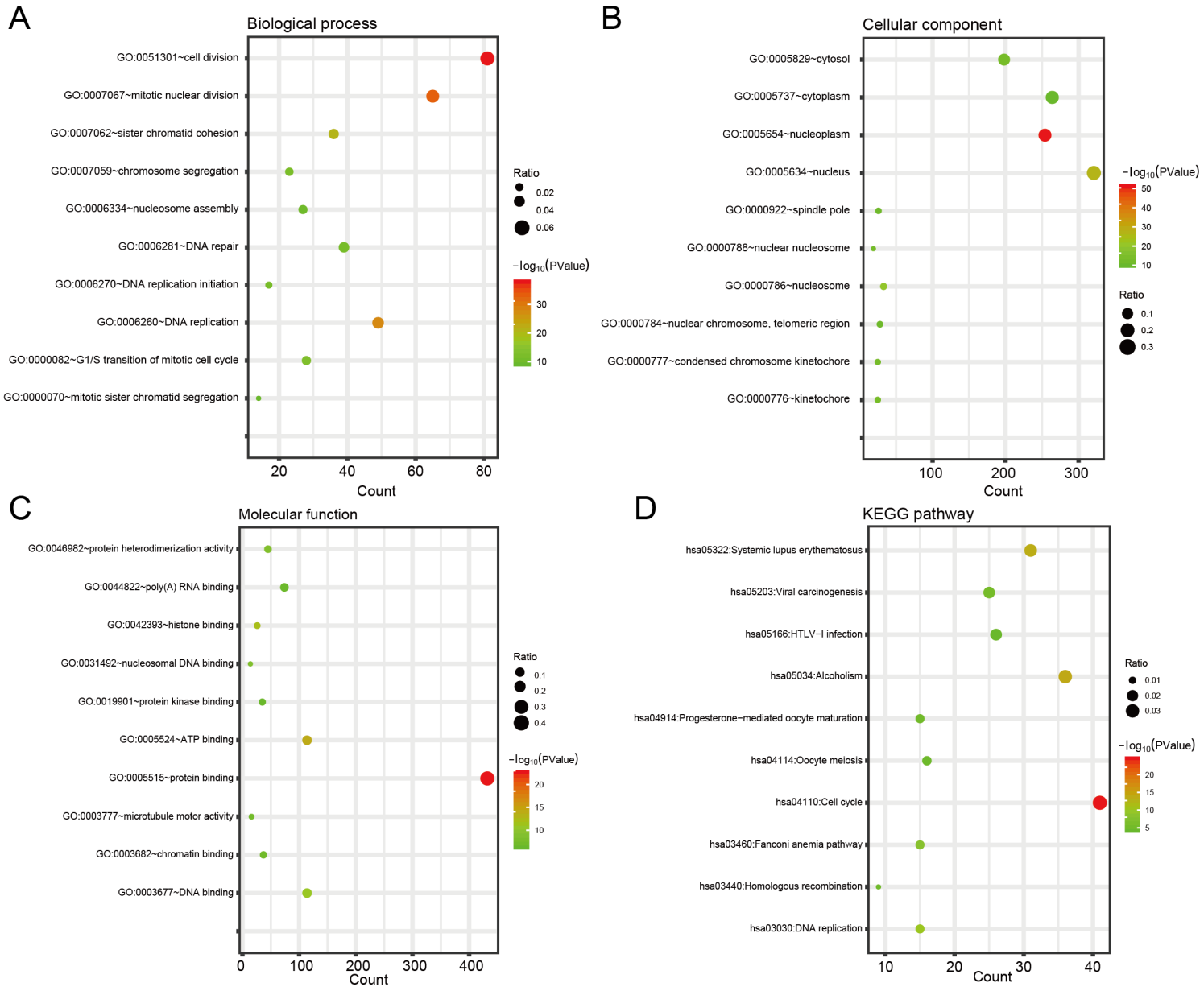

Figure 6

Gene ontology and pathway enrichment analysis of blue module genes. (A). Biological process analysis. (B). Cellular component analysis. (C). Molecular function analysis. (D). KEGG pathway analysis.

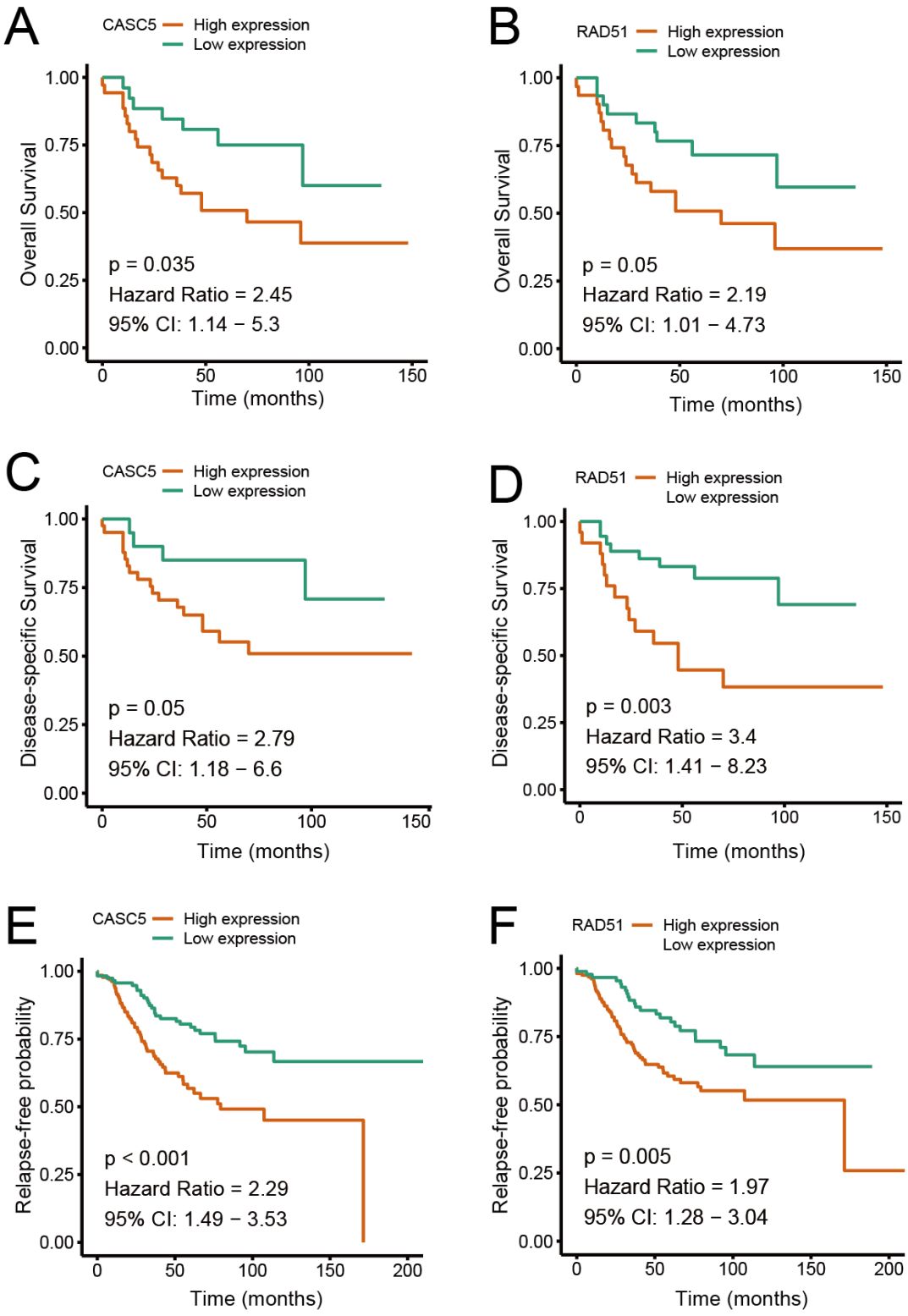

Figure 8

Survival analysis of CASC5 and RAD51 in breast cancer in dataset GSE37751 and GSE21653. The patients were stratified into high-level group and low-level group according to median expression. (A, B). Overall survival (OS) of CASC5 and RAD51 in GSE37751. (C, D). Disease-specific survival of CASC5 and RAD51 in GSE37751. (E, F). Relapse free survival (RFS) of CASC5 and RAD51 in GSE21653.

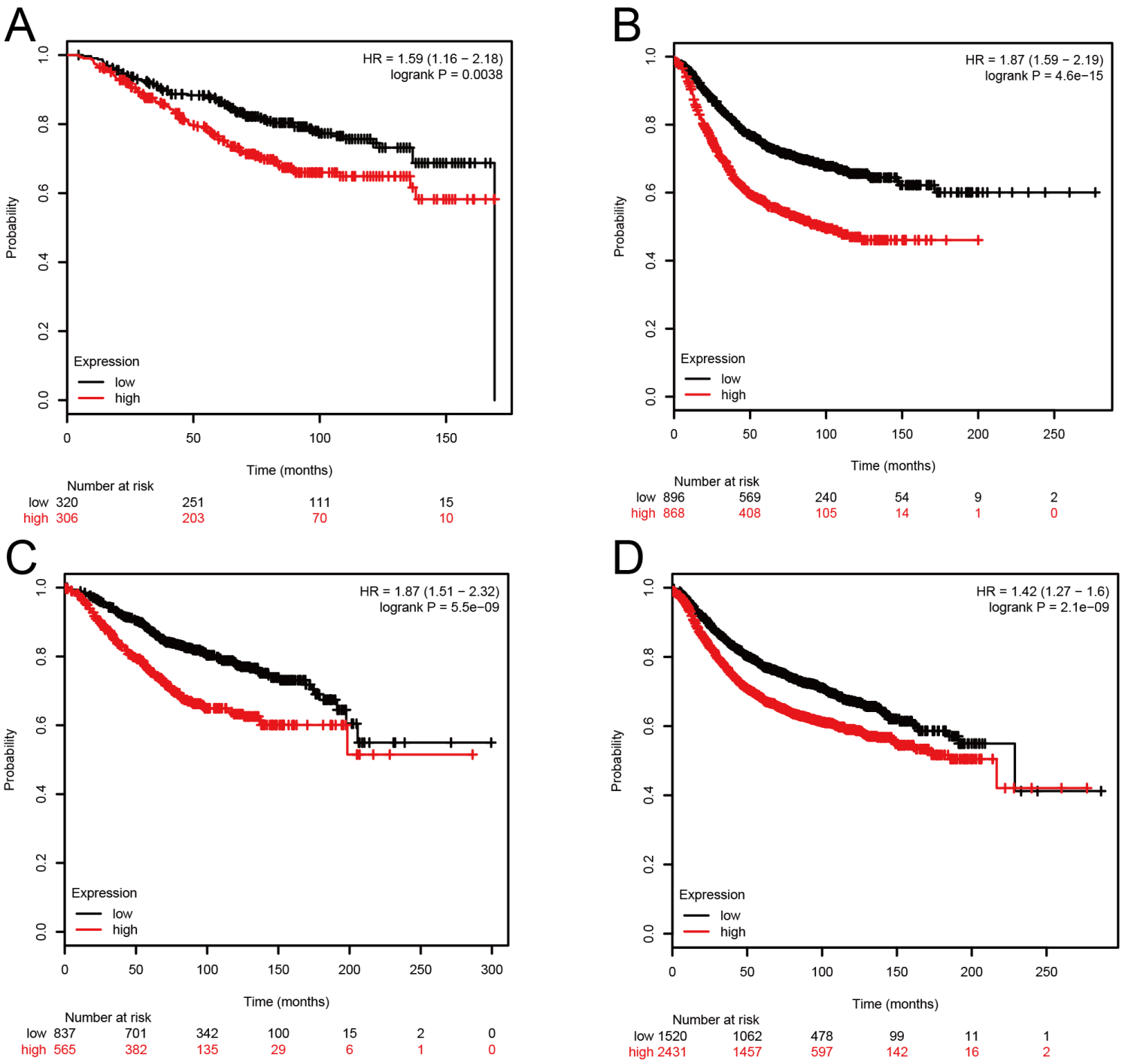

Figure 10

Overall survival (OS) and relapse free survival (RFS) of the CASC5 and RAD51 in breast cancer based on Kaplan Meier-plotter. The patients were stratified into high-level and low-level groups according to median expression. (A). OS of CASC5. (B). RFS of CASC5. (C). OS of RAD51. (D). RFS of RAD51.

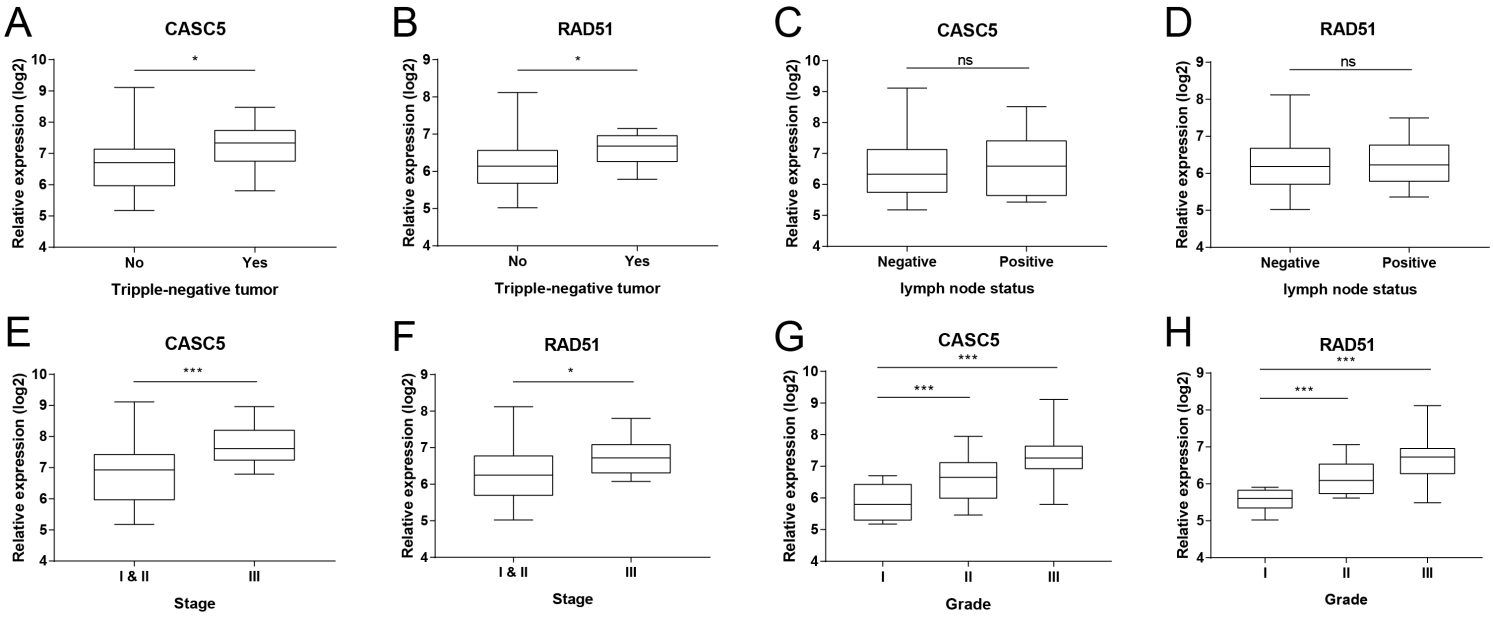

Figure 12

Expression levels of CASC5 and RAD51 in GSE37751. (A). CASC5 expression and breast cancer subtypes. (B). RAD51 expression and breast cancer subtypes. (C). CASC5 expression and lymph node status. (D). RAD51 expression and lymph node status. (E). CASC5 expression and tumor stages. (F). RAD51 expression and tumor stages. (G). CASC5 expression and tumor grade. (H). RAD51 expression and tumor grade. *, P value < 0.05; **, P value < 0.01; ***, P value < 0.001; ****, P value < 0.0001. One-way analysis of variance (ANOVA) and two-tailed Student's t-tests were used to evaluate the statistical significance of differences.

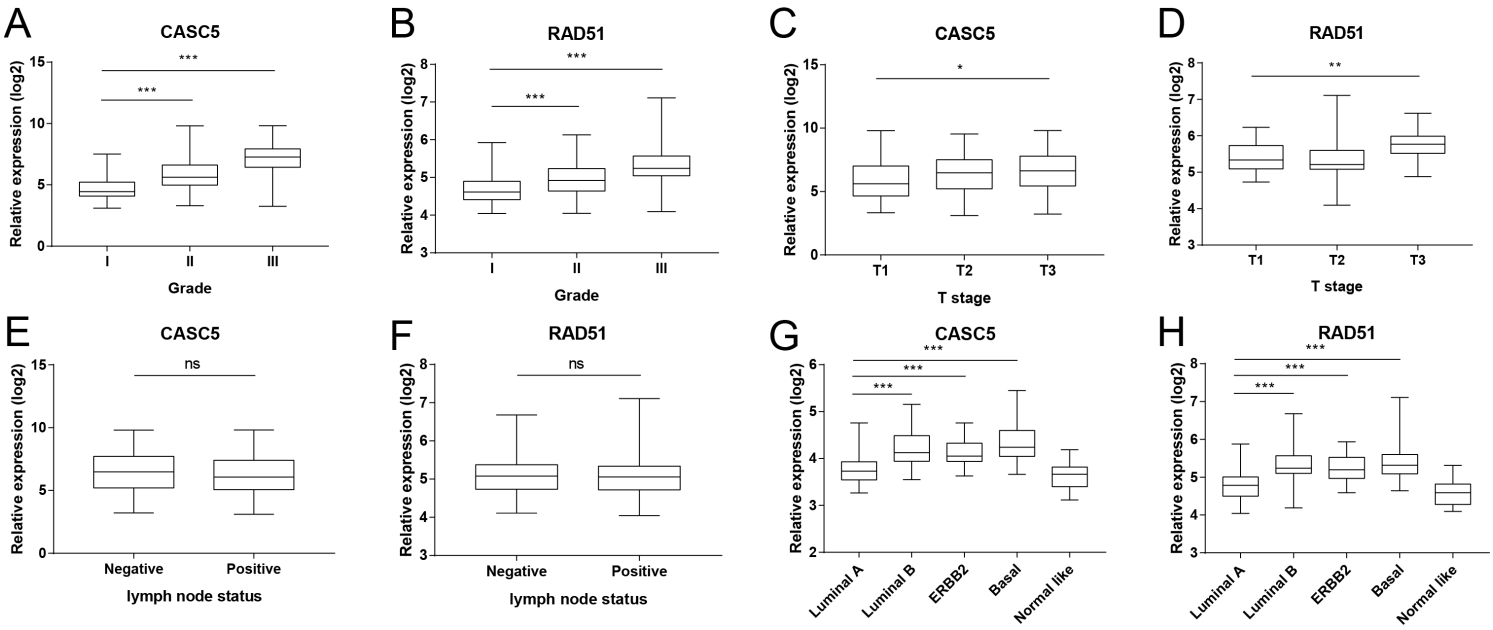

Figure 14

Expression levels of CASC5 and RAD51 in GSE21653. (A). CASC5 expression and tumor grade. (B). RAD51 expression and tumor grade. (C). CASC5 expression and tumor stages. (D). RAD51 expression and tumor stages. (E). CASC5 expression and lymph node status. (F). RAD51 expression and lymph node status. (G). CASC5 expression and breast cancer subtypes. (H). RAD51 expression and breast cancer subtypes. *, P value < 0.05; **, P value < 0.01; ***, P value < 0.001; ****, P value < 0.0001. One-way analysis of variance (ANOVA) and two-tailed Student's t-tests were used to evaluate the statistical significance of differences.

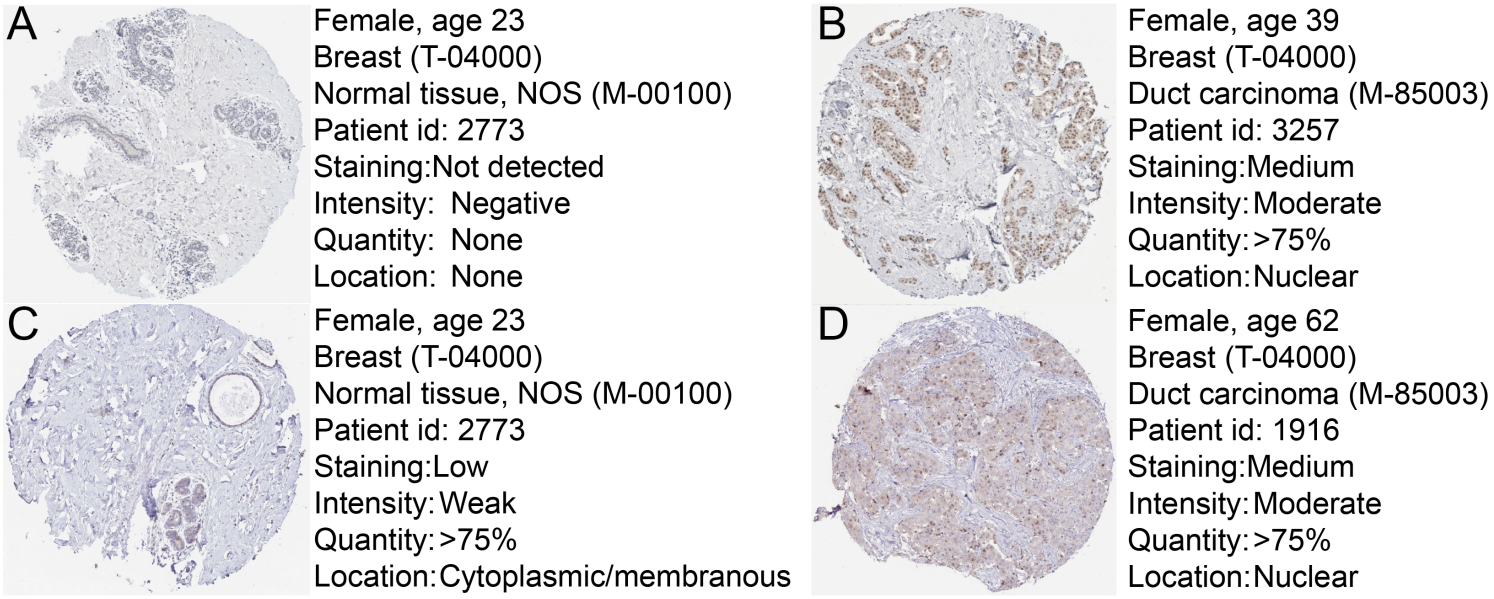

Figure 16

Immunohistochemistry of the CASC5 and RAD51 hub genes based on the Human Protein Atlas. (A). Protein levels of CASC5 in normal tissues. (B). Protein levels of CAS5 in tumor tissues. (C). Protein levels of RAD51 in normal tissues. (D). Protein levels of RAD51 in tumor tissues.

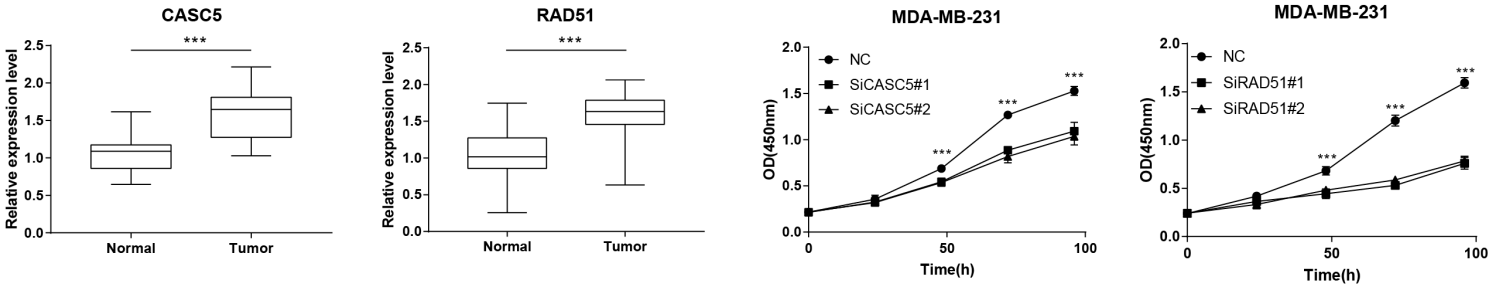

Figure 18

Experimental validation of CASC5, and RAD51. (A). Relative expression of CASC5 and RAD51 in breast cancer tissues and paracancerous tissues. (B). Cell Counting Kit-8 (CCK-8) assay. *, P value < 0.05; **, P value < 0.01; ***, P value < 0.001; ****, P value < 0.0001. One-way analysis of variance (ANOVA) and two-tailed Student's t-tests were used to evaluate the statistical significance of differences.

Supplementary Files

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

- FigureS1.tif
- FigureS1.tif