The Target E2F Family Gene *PTTG1* for Prediction of Distant Relapse-Free Survival in Breast Cancer Patients Receiving Taxane and Anthracycline-Based NACT

Yu-hao Xu

Zhejiang Chinese Medical University

Yao-qiang Du

Zhejiang Provincial People's Hospital

Qing-hui Zheng

Zhejiang Provincial People's Hospital

Qiu-ran Xu

Zhejiang Provincial People's Hospital

Xuli Meng (✉ mxlmail@126.com)

Zhejiang Provincial People's Hospital <https://orcid.org/0000-0002-5555-0232>

Research Article

Keywords: distant relapse-free survival, neoadjuvant chemotherapy, Gene Expression Omnibus, Gene Ontology

Posted Date: November 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-551089/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The breast cancer is the most commonly diagnosed cancer in recent years. The use of neoadjuvant chemotherapy (NACT) makes a significant contribution to chemotherapy in breast cancer. We aimed to develop the novel model as a predictor of distant relapse-free survival (DRFS) in breast cancer patients receiving taxane and anthracycline-based NACT.

Methods: We collected the mRNA expression datasets of patients from GSE25055 and GSE25065 in Gene Expression Omnibus (GEO). Univariate and Multivariate Cox Regression Analyses were conducted to achieve the prognostic genes that associated with DRFS. Moreover, the E2F target genes were obtained from GSEA. We obtained the intersection genes between the prognostic genes and E2F target genes, then validated in GSE32603 dataset. And we established a nomogram model based on *PTTG1* expression level and several clinical characteristics.

Results: There were 95 genes confirmed to be associated with DRFS both in GSE25055 and GSE25065. And *PTTG1* was validated as a gene related to DRFS and associated with E2F target hallmark using GSEA analysis and chosen as our target gene. The AUC of *PTTG1* model was 0.682 indicating a relatively poor prediction performance. Then a novel nomogram was conducted, the receiver operating characteristic (AUC=0.849), C-index (0.805) and calibration plots were applied to assess the effect of this model.

Conclusion: Our study found that the E2F target genes, such as the *PTTG1* may serve as a potential biomarker in breast cancer, and provided superior estimation of DRFS, which can guide the clinical practice in NACT of breast cancer.

Introduction

Recently, an estimated 2.3 million new cases, let breast cancer surpass lung cancer to be the most commonly diagnosed cancer, accounting for 11.7% of all new cancer cases [1]. Incidence rate of breast cancer has increased in the past decades, and another scared truth is the dropping average onset age [2].

The use of neoadjuvant chemotherapy (NACT) makes a significant contribution to chemotherapy in breast cancer, which could be employed as a bridge to other therapies [3, 4]. Anthracyclines and taxanes serve as the backbone of NACT regimens and are widely used clinically in breast cancer patients [5]. It has been proved that the E2F family of transcription factor are synergistically involved in regulating cell-cycle progression [6], controlling cell proliferation, apoptosis [7] and DNA repair with consequent effects on cell growth or tumor cell invasion [8–12].

PTTG1, described in the context of carcinogenesis [13–16], as a target gene of E2F family [17, 18] has been implicated in multiple cellular processes, but not limited to, cellular differentiation, apoptosis, DNA damage repair, maintains chromosomal stability and angiogenesis in breast cancer [19–21]. Quiet a lot of breast cancer patients couldn't benefit from NACT and show distinct responses to NACT even when

their clinicopathological characteristics are similar because of the molecular differences, and the absence of well-defined molecular target [22, 23]. Thus, it is in urgent need to discover a robust biomarker to predict distant relapse-free survival (DRFS). In this study, a nomogram model was constructed to predict the DRFS in patients receiving taxane and anthracycline-based NACT and for guiding clinical practice (Fig. 1).

Materials And Methods

Data collection

Datasets were used from GEO database (GSE25055, 310 breast cancer samples; GSE25065, 198 breast cancer samples), the patients were receiving taxane and anthracycline-based neoadjuvant chemotherapy as discovery cohorts, while GSE32603 (248 breast cancer samples) as a validation cohort. Gene expression of GSE25055 and GSE25065 was collected by using platform GPL96 ([HG-U133A] Affymetrix Human Genome U133A Array), and GSE32603 was using platform GPL14668 (UCSF/HAQQLAB_Human_40986_ISPY). Batch effects and other unwanted variation in high-throughput experiments were eliminated using the combat function in the sva R package or with known control probes.

Screening of genes related to DRFS

Firstly, we performed univariate Cox regression analysis in the GSE25055 and GSE25065 datasets. Then, in order to remove confounding bias, we conducted multivariate Cox regression analysis in above datasets. And genes associated with DRFS in both GSE25055 and GSE25065 could be sorted out. Using Gene Ontology (GO) enrichment analysis to discover the target cancer hallmark, that related to the functional annotations of these DRFS genes with the help of clusterProfiler R package, and hallmark E2F target genes were sorted out. Then we got intersection genes, included genes that associated with DRFS and genes in hallmark E2F target genes. K-M survival analysis and multivariate Cox regression analysis were performed in GSE32603 validation dataset to identify these genes. Finally, *PTTG1* was validated as a gene that related to DRFS and associated with E2F target hallmark using GSEA analysis and chosen as the target genes for further study.

Establishment of PPI network and GSEA analysis

We downloaded the STRING PPI dataset that named '9606.protein.links.full.v11.0' which containing the complete list of protein links in Homo Sapiens and selected all the proteins links involving *PTTG1*. Then we chose genes involved in GSE25055, GSE25065 and Hallmark E2F_Targets for PPI construction, whose products have protein links with *PTTG1*. Cytoscape (version 3.8.2) was used to visualize the PPI network centered on *PTTG1*. The permutation test for this PPI network included 100,000 iterations for proteins and their connections in order to improve the reliability. Then we separated the patients into high *PTTG1* expression and low *PTTG1* expression groups, and performed a pathway analysis using the Gene Set

Enrichment Analysis (GSEA) in GSE25055 and GSE25065 cohorts with the standard of P value <0.01 and False Discovery Rate (FDR) < 0.01 .

Construction and evaluation of the nomogram for predicting DRFS

Parameters such as the expression level of *PTTG1*, age, status of hormone receptor (HR) for estrogens (ER) and progesterone (PR), and the AJCC tumor-node-metastasis (TNM), staging system (AJCC stage), pathology grade and the pathological complete response were included in the analysis. A nomogram was constructed using “cph” function in the “rms” package (version: ?) for predicting 1-year, 3-year, 5-year DFS rates. C-index, calibration plots and ROC were used to evaluate the performance of the DRFS model.

Results

Filtering the target genes related to DRFS

The GSE25055 and GSE25065 datasets were used as discovery cohorts, meanwhile GSE32603 as a validation cohort. Batch effects were removed with the help of “sva” R package (Fig. 2a,b). First of all, we performed univariate Cox regression analysis in GSE25055 for acquiring 2,753 genes that associated with DRFS, with the standard of $P \leq 0.05$, and in GSE25065 was 2,002. In order to remove confounding bias and sort out independent prognostic indicator for breast cancer patients, we conducted multivariate Cox regression analysis on the basis of the result of univariate Cox regression analysis in GSE25055 with the condition of screening of $P \leq 0.05$ for acquiring 889 genes as the predictors of DRFS of patients independently, and in GSE25065 was 642. The 95 genes were finally confirmed to be associated with DRFS both in GSE25055 and GSE25065.

Gene Ontology (GO) enrichment analysis of genes for identifying potential pathways and functional annotations

To discover the potential pathways and functional annotations of the genes associated with DRFS for further study. We performed the GO functional analysis separately in GSE25055 and GSE25065 (Fig. 2c,d) with $P < 0.05$ was statistical difference. From the cellular component (CC), biological process (BP), and molecular function (MF), and we found the genes were mainly concerned with cell division, cell-cycle progression, controlling cell proliferation, apoptosis and DNA repair. Based on the above study, we found that these functional annotations were closely related to genes of E2F family.

Screening out the target genes in GSE32603

After the GO functional analysis, E2F family were chosen for further study. Two hundreds target genes of E2F family were extracted in hallmark E2F targets gene set from the GSEA database for explaining the relationship between genes related to DRFS and the Hallmark E2F target genes, thus we obtained 3 intersection genes *CDC25B*, *MTHFD2* and *PTTG1* (Fig. 3a). We conducted heatmaps of the three genes

expression profiles in GSE25055 and GSE25065 individually (Fig. 3b,c). Furthermore, we chose GSE32603 as a validation cohort for screening the target genes. And Kaplan-Meier curve analysis of *CDC25B*, *MTHFD2*, *PTTG1* was also performed.

The Kaplan-Meier (K-M) survival curve showed that in GSE32603 the DRFS of high *CDC25B*, *MTHFD2*, *PTTG1* expression patients was lower than the low ones with $P=0.012$, $P=0.008$, $P=0.001$ using log-rank tests (Fig. 4a). Then we conducted multivariate Cox regression analysis in GSE32603. It indicated the expression level of *PTTG1* ($P=0.029$) and the pathological complete response ($P=0.001$) could serve as independent DRFS predictors, besides, *PTTG1* was a high-risk gene with $HR=1.396$, 95%CI: 1.035-1.883. However, *CDC25B* ($P=0.093$) and *MTHFD2* ($P=0.101$) were considered to be confounding factors which couldn't predict DRFS independently (Fig. 4b).

The above results proved that *PTTG1* was a gene related to DRFS, and high expression of *PTTG1* indicated a worse DRFS. Therefore, *PTTG1* was sorted out for further study.

Discovery of E2F-related genes and enriched pathway connected with PTTG1 in breast cancer

We chose the proteins interacted with *PTTG1* in the three subsets, and constructed a PPI network which is centered on *PTTG1* (Fig. 5a). We used a 100,000 iterations permutation test to identify the reliability of the network, both the P value of the nodes and edges were < 0.001 . The PPI network indicated that the most important biological processes were Nuclear division, Negative regulation of cell cycle process, Nuclear chromosome segregation, Regulation of chromosome organization. The results of GO and KEGG enrichment analysis seen in **Supplementary Fig. 1,2**. We separated patients into *PTTG1* high expression and low expression groups, and additionally employed a GSEA functional annotation approach. The result showed *PTTG1* high expression patients were significantly enriched in E2F family target genes that relative to patients from *PTTG1* low expression patients ($P=0.016$), normalized enrichment score (NES) > 1.5 and false discovery rate (FDR)=0.007 (Fig. 5b), indicated the high expression of *PTTG1* was associated with E2F pathways in breast cancer.

Using ROC curve to evaluate the accuracy and specificity of *PTTG1* for predicting the DRFS of breast cancer patients, the AUC of *PTTG1* was 0.682 that indicated a relative poor prediction performance (Fig. 5c).

Establishment of a nomogram for better DRFS performance

In order to predict DRFS better, a nomogram was established, including clinical characteristics and expression level of *PTTG1* in GSE25055 and GSE25065 for predicting 1-year, 3-year, 5-year DRFS of breast cancer patients receiving taxane and anthracycline-based NACT (Fig. 6a).

The patient owned his total score by adding a corresponding score for each prognostic factor in the nomogram. Patients were separated into a high-risk cancer group and a low-risk cancer group that based on the risk scores. Patients with a high-risk score exhibited significantly worse DRFS that compared with the low-risk score ($P=0.001$), the same result in the K-M survival curve (Fig. 6b).

The calibration plots for 1-year, 3-year and 5-year distant relapse-free survival of patients that indicated the nomogram-predicted outcome, showed high concordance with the actual outcome (Fig. 6c,d,e). The C-index of predicted relapse-free survival was 0.805 which showed its high prediction accuracy and a high sense of coherence. For the further assessment of the accuracy, specificity and sensitivity of the nomogram, we performed receiver operating characteristic (ROC) curve analysis for the evaluation of the prediction power of the model. The area under ROC curve (AUC) of the risk nomogram was 0.849, while AUC of the expression level of *PTTG1* was 0.682, and AUC (pathologic_response) = 0.409, AUC (grade) = 0.593, AUC (ajcc_stage) = 0.696, AUC (N stage) = 0.698, AUC (T stage) = 0.701, AUC (status of PR) = 0.332, AUC (ER) = 0.306, AUC (age) = 0.490 (Fig. 6f).

Discussion

Taxane and anthracycline-based neoadjuvant chemotherapy is widely used in clinical practice in breast cancer [24], the heterogeneity of molecular features of different subtypes have been explored, which leads to distinct survival outcomes [25, 26]. How to predict the long-term outcome of patients remains to be a problem. Several studies had been made to solve this problem [23, 27]. However, the accuracy and specificity of these studies was relatively poor. Thus a robust prognostic indicator was in urgent need for better risk stratification and proper treatment regimen of breast cancer patients.

In the past decade, the data was provided by public databases without cost, which led to the development of high-throughput sequencing at a high rate of speed, by mingling TCGA, GEO, GTEx, CCLE and other databases [28–32]. High-throughput genomic studies provided cutting-edge sights into the molecular mechanisms and identified new potential targets of breast cancer [33–36]. Nevertheless, using a single biomarker for predicting often relatively lower predictive effects, and a risk model for predicting the OS and DRFS of patients [37–43] and a nomogram were popular which showed superior predictive effects as a better alternative [18, 44–46]. Meanwhile E2F related genes had been selected to build a model for predict the DRFS of breast cancer patients [47].

It is widely known that one of the hallmarks of cancer is dysregulation of the cell cycle leading to disorganized and continuous cell proliferation [48]. Meanwhile, transcription factors in the E2F family are known for controlling the cell cycle, which promotes the transcription of numerous genes required for S phase entry and DNA synthesis [49]. *PTTG1*, as a target gene of E2F family, whose function are related to cell-cycle regulation and separating sister chromatids continuous in the course of mitosis was chosen to be our target genes for its bright prospect.

Using univariate Cox regression and multivariate Cox regression analyses for identifying target genes of E2F family related to DRFS. Based on the results of GO enrichment analysis and KEGG enrichment analysis, the current studies as well, *PTTG1* was chosen for further study. For better accuracy and specificity, we chose to combine *PTTG1* with clinical-pathological features for further improvement of the predictive power.

However, the molecular mechanism on how *PTTG1* effects in taxane and anthracycline-based NACT hasn't been fully revealed. Thus, we plan to conduct further experimental studies for exploring the detailed effects of *PTTG1* on the biological behavior and pathogenesis of breast cancer. Besides, all the analyses were based on the public databases, considering the limited study of breast patients, we aimed to gather clinicopathological characteristics of breast cancer patients for large-scale cohort studies in the future, and a series of our validation samples will be verified the accuracy of this nomogram model.

Conclusion

Our study provided a nomogram model based on *PTTG1*, a target gene of E2F family, aiming to predict the distant relapse-free survival in breast cancer patients receiving taxane and anthracycline-based NACT better with high accuracy and specificity. By adding a corresponding score for clinicopathological characters in the nomogram model for a risk score of specific patients, it could separate patients into different risk groups with a better risk classification, consequently, to be a coaching and guidance for individualized healthcare decisions of breast cancer patients receiving taxane and anthracycline-based NACT.

Abbreviations

DRFS

distant relapse-free survival

NACT

neoadjuvant chemotherapy

GEO

Gene Expression Omnibus

TCGA

The Cancer Genome Atlas

GO

Gene Ontology

KEGG

Kyoto Encyclopedia of Genes and Genomes

GTEx

Genotype-Tissue Expression Project

ROC curve

The receiver operating characteristic curve

Declarations

ACKNOWLEDGMENTS

Not applicable.

AUTHOR CONTRIBUTIONS

XLM and QRX substantially contributed to the conception of the work. YHX, QHZ contributed to the data collection. YHX and YQD performed the nomogram model analyses and wrote the manuscript. YHX, YQD and XLM drafted and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by grants from the National Natural Science Foundation of China (81973861), the Zhejiang Provincial Ministry Medical and Health Co-construction Major Project (20214355173), the Natural Science Foundation of Zhejiang Province in China (LQ21H200007).

AVAILABILITY OF DATA AND MATERIALS

The datasets in this study were available in Gene Expression Omnibus (GEO) in the GSE25055, GSE25065 and GSE32603 as well as the Hallmark E2F target in GSEA.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was not involved in human participant and clinical information. All the datasets were from the GEO database. All clinical data involved in this study were obtained from open databases and therefore met ethical review standards.

CONSENT FOR PUBLICATION

Not applicable

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *Cancer J Clin*. 2021;71(3):209–49. <https://doi.org/10.3322/caac.21660>.
2. Isik A, Firat D. Bilateral intra-areolar polythelia. *Breast J*. 2018;24(1):89–90. <https://doi.org/10.1111/tbj.12838>.
3. Harbeck N, Gnant M. Breast cancer. *Lancet*. 2017;389(10074):1134–50. [https://doi.org/10.1016/S0140-6736\(16\)31891-8](https://doi.org/10.1016/S0140-6736(16)31891-8).
4. Kuijter A, Straver M, den Dekker B, van Bommel ACM, Elias SG, Smorenburg CH, et al. Impact of 70- Gene Signature Use on Adjuvant Chemotherapy Decisions in Patients With Estrogen Receptor-

- Positive Early Breast Cancer: Results of a Prospective Cohort Study. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2017;35(24):2814–9. <https://doi.org/10.1200/JCO.2016.70.3959>.
5. Chaudhary LN. Early stage triple negative breast cancer: Management and future directions. *Semin Oncol*. 2020;47(4):201–8. <https://doi.org/10.1053/j.seminoncol.2020.05.006>.
 6. Kanska J, Zakhour M, Taylor-Harding B, Karlan BY, Wiedemeyer WR. Cyclin E as a potential therapeutic target in high grade serous ovarian cancer. *Gynecol Oncol*. 2016;143(1):152–8. <https://doi.org/10.1016/j.ygyno.2016.07.111>.
 7. Shibue T, Weinberg RA. Integrin beta1-focal adhesion kinase signaling directs the proliferation of metastatic cancer cells disseminated in the lungs. *Proc Natl Acad Sci USA*. 2009;106(25):10290–5. <https://doi.org/10.1073/pnas.0904227106>.
 8. Manickavinayaham S, Dennehey BK, Johnson DG. Direct Regulation of DNA Repair by E2F and RB in Mammals and Plants: Core Function or Convergent Evolution? *Cancers (Basel)*. 2021;13(5). <https://doi.org/10.3390/cancers13050934>.
 9. Park SY, Seo YR, Ko MJ, Lee JH, Chun KS, Kim MJ, et al. Targeting CALM2 Inhibits Hepatocellular Carcinoma Growth and Metastasis by Suppressing E2F5-mediated Cell Cycle Progression. *Anticancer research*. 2021;41(3):1315–25. <https://doi.org/10.21873/anticancer.14889>.
 10. Perrotta L, Giordo R, Francis D, Rogers HJ, Albani D. Molecular Analysis of the E2F/DP Gene Family of *Daucus carota* and Involvement of the DcE2F1 Factor in Cell Proliferation. *Front Plant Sci*. 2021;12:652570. <https://doi.org/10.3389/fpls.2021.652570>.
 11. Yan Y, Xu H, Wang J, Wu X, Wen W, Liang Y, et al. Inhibition of breast cancer cells by targeting E2F-1 gene and expressing IL15 oncolytic adenovirus. *Biosci Rep*. 2019;39(7). <https://doi.org/10.1042/BSR20190384>.
 12. Zhou X, Zhong S, Peng H, Liu J, Ding W, Sun L, et al. Cellular and molecular properties of neural progenitors in the developing mammalian hypothalamus. *Nature communications*. 2020;11(1):4063. <https://doi.org/10.1038/s41467-020-17890-2>.
 13. Hu ZG, Zheng CW, Su HZ, Zeng YL, Lin CJ, Guo ZY, et al. MicroRNA-329-mediated PTTG1 downregulation inactivates the MAPK signaling pathway to suppress cell proliferation and tumor growth in cholangiocarcinoma. *Journal of cellular biochemistry*. 2019;120(6):9964–78. <https://doi.org/10.1002/jcb.28279>.
 14. Grzechowiak I, Gras J, Szymanska D, Biernacka M, Guglas K, Poter P, et al. The Oncogenic Roles of PTTG1 and PTTG2 Genes and Pseudogene PTTG3P in Head and Neck Squamous Cell Carcinomas. *Diagnostics (Basel)*. 2020;10(8). <https://doi.org/10.3390/diagnostics10080606>.
 15. Romero Arenas MA, Whitsett TG, Aronova A, Henderson SA, LoBello J, Habra MA, et al. Protein Expression of PTTG1 as a Diagnostic Biomarker in Adrenocortical Carcinoma. *Ann Surg Oncol*. 2018;25(3):801–7. <https://doi.org/10.1245/s10434-017-6297-1>.
 16. Cui L, Ren T, Zhao H, Chen S, Zheng M, Gao X, et al. Suppression of PTTG1 inhibits cell angiogenesis, migration and invasion in glioma cells. *Med Oncol*. 2020;37(8):73. <https://doi.org/10.1007/s12032->

17. Zhi T, Jiang K, Xu X, Yu T, Zhou F, Wang Y, et al. ECT2/PSMD14/PTTG1 axis promotes the proliferation of glioma through stabilizing E2F1. *Neuro Oncol.* 2019;21(4):462–73. <https://doi.org/10.1093/neuonc/noy207>.
18. Zhou C, Wawrowsky K, Bannykh S, Gutman S, Melmed S. E2F1 induces pituitary tumor transforming gene (PTTG1) expression in human pituitary tumors. *Mol Endocrinol.* 2009;23(12):2000–12. <https://doi.org/10.1210/me.2009-0161>.
19. Meng C, Zou Y, Hong W, Bao C, Jia X. Estrogen-regulated PTTG1 promotes breast cancer progression by regulating cyclin kinase expression. *Mol Med.* 2020;26(1):33. <https://doi.org/10.1186/s10020-020-00161-7>.
20. Yu R, Heaney AP, Lu W, Chen J, Melmed S. Pituitary tumor transforming gene causes aneuploidy and p53-dependent and p53-independent apoptosis. *J Biol Chem.* 2000;275(47):36502–5. <https://doi.org/10.1074/jbc.C000546200>.
21. Zhou Y, Mehta KR, Choi AP, Scolavino S, Zhang X. DNA damage-induced inhibition of securin expression is mediated by p53. *J Biol Chem.* 2003;278(1):462–70. <https://doi.org/10.1074/jbc.M203793200>.
22. Hatzis C, Pusztai L, Valero V, Booser DJ, Esserman L, Lluch A, et al. A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *Jama.* 2011;305(18):1873–81. <https://doi.org/10.1001/jama.2011.593>.
23. Cheng P, Wang Z, Hu G, Huang Q, Han M, Huang J. A prognostic 4-gene expression signature for patients with HER2-negative breast cancer receiving taxane and anthracycline-based chemotherapy. *Oncotarget.* 2017;8(61):103327–39. <https://doi.org/10.18632/oncotarget.21872>.
24. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet.* 2014;384(9938):164–72. [https://doi.org/10.1016/S0140-6736\(13\)62422-8](https://doi.org/10.1016/S0140-6736(13)62422-8).
25. Li Y, Umbach DM, Krahn JM, Shats I, Li X, Li L. Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines. *BMC Genomics.* 2021;22(1):272. <https://doi.org/10.1186/s12864-021-07581-7>.
26. Tan W, Liu M, Wang L, Guo Y, Wei C, Zhang S, et al. Novel immune-related genes in the tumor microenvironment with prognostic value in breast cancer. *BMC Cancer.* 2021;21(1):126. <https://doi.org/10.1186/s12885-021-07837-1>.
27. Sun M, Liu X, Xia L, Chen Y, Kuang L, Gu X, et al. A nine-lncRNA signature predicts distant relapse-free survival of HER2-negative breast cancer patients receiving taxane and anthracycline-based neoadjuvant chemotherapy. *Biochem Pharmacol.* 2020:114285. <https://doi.org/10.1016/j.bcp.2020.114285>.
28. Chao B, Ju X, Zhang L, Xu X, Zhao Y. A Novel Prognostic Marker Systemic Inflammation Response Index (SIRI) for Operable Cervical Cancer Patients. *Front Oncol.* 2020;10:766. <https://doi.org/10.3389/fonc.2020.00766>.

29. Chen C, Pan Y, Bai L, Chen H, Duan Z, Si Q, et al. MicroRNA-3613-3p functions as a tumor suppressor and represents a novel therapeutic target in breast cancer. *Breast Cancer Res.* 2021;23(1):12. <https://doi.org/10.1186/s13058-021-01389-9>.
30. Wang Y, Zhu M, Guo F, Song Y, Fan X, Qin G. Identification of Tumor Microenvironment-Related Prognostic Biomarkers in Luminal Breast Cancer. *Frontiers in genetics.* 2020;11:555865. <https://doi.org/10.3389/fgene.2020.555865>.
31. Wu M, Zhao Y, Peng N, Tao Z, Chen B. Identification of chemoresistance-associated microRNAs and hub genes in breast cancer using bioinformatics analysis. *Invest New Drugs.* 2021;39(3):705–12. <https://doi.org/10.1007/s10637-020-01059-1>.
32. Xu L, Shen JM, Qu JL, Song N, Che XF, Hou KZ, et al. FEN1 is a prognostic biomarker for ER+ breast cancer and associated with tamoxifen resistance through the ERalpha/cyclin D1/Rb axis. *Annals of translational medicine.* 2021;9(3):258. <https://doi.org/10.21037/atm-20-3068>.
33. Chen L, Dong Y, Pan Y, Zhang Y, Liu P, Wang J, et al. Identification and development of an independent immune-related genes prognostic model for breast cancer. *BMC Cancer.* 2021;21(1):329. <https://doi.org/10.1186/s12885-021-08041-x>.
34. Xian J, Wang S, Jiang Y, Li L, Cai L, Chen P, et al. Overexpressed NEDD8 as a potential therapeutic target in esophageal squamous cell carcinoma. *Cancer Biol Med.* 2021. <https://doi.org/10.20892/j.issn.2095-3941.2020.0484>.
35. Xie Q, Xiao YS, Jia SC, Zheng JX, Du ZC, Chen YC, et al. FABP7 is a potential biomarker to predict response to neoadjuvant chemotherapy for breast cancer. *Cancer Cell Int.* 2020;20(1):562. <https://doi.org/10.1186/s12935-020-01656-3>.
36. Zhang L, Pan J, Wang Z, Yang C, Huang J. Construction of a MicroRNA-Based Nomogram for Prediction of Lung Metastasis in Breast Cancer Patients. *Frontiers in genetics.* 2020;11:580138. <https://doi.org/10.3389/fgene.2020.580138>.
37. Gong C, Tan W, Chen K, You N, Zhu S, Liang G, et al. Prognostic Value of a BCSC-associated MicroRNA Signature in Hormone Receptor-Positive HER2-Negative Breast Cancer. *EBioMedicine.* 2016;11:199–209. <https://doi.org/10.1016/j.ebiom.2016.08.016>.
38. Guo Y, Mao X, Qiao Z, Chen B, Jin F. A Novel Promoter CpG-Based Signature for Long-Term Survival Prediction of Breast Cancer Patients. *Front Oncol.* 2020;10:579692. <https://doi.org/10.3389/fonc.2020.579692>.
39. Liu C, Wang S, Zheng S, Wang X, Huang J, Lei Y, et al. A novel recurrence-associated metabolic prognostic model for risk stratification and therapeutic response prediction in patients with stage I lung adenocarcinoma. *Cancer Biol Med.* 2021. <https://doi.org/10.20892/j.issn.2095-3941.2020.0397>.
40. Liu Y, Huang Z, Cheng G, Shou Y, Xu J, Liu D, et al. Development of a four-gene prognostic model for clear cell renal cell carcinoma based on transcriptome analysis. *Genomics.* 2021;113(4):1816–27. <https://doi.org/10.1016/j.ygeno.2021.04.005>.

41. Shao J, Lyu W, Zhou J, Xu W, Wang D, Liang S, et al. A Panel of Five-lncRNA Signature as a Potential Biomarker for Predicting Survival in Gastric and Thoracic Cancers. *Frontiers in genetics*. 2021;12:666155. <https://doi.org/10.3389/fgene.2021.666155>.
42. Zhang X, Wang J, Zhuang J, Liu C, Gao C, Li H, et al. A Novel Glycolysis-Related Four-mRNA Signature for Predicting the Survival of Patients With Breast Cancer. *Frontiers in genetics*. 2021;12:606937. <https://doi.org/10.3389/fgene.2021.606937>.
43. Zhao Y, Pu C, Liu Z. Exploration the Significance of a Novel Immune-Related Gene Signature in Prognosis and Immune Microenvironment of Breast Cancer. *Front Oncol*. 2020;10:1211. <https://doi.org/10.3389/fonc.2020.01211>.
44. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57–70. [https://doi.org/10.1016/s0092-8674\(00\)81683-9](https://doi.org/10.1016/s0092-8674(00)81683-9).
45. Huang Q, Qu T, Qi L, Liu C, Guo Y, Guo Q, et al. A nomogram-based immune-serum scoring system predicts overall survival in patients with lung adenocarcinoma. *Cancer Biol Med*. 2021. <https://doi.org/10.20892/j.issn.2095-3941.2020.0648>.
46. Tu Q, Hu C, Zhang H, Peng C, Kong M, Song M, et al. Establishment and Validation of Novel Clinical Prognosis Nomograms for Luminal A Breast Cancer Patients with Bone Metastasis. *BioMed research international*. 2020;2020:1972064. <https://doi.org/10.1155/2020/1972064>.
47. Oshi M, Takahashi H, Tokumaru Y, Yan L, Rashid OM, Nagahashi M, et al. The E2F Pathway Score as a Predictive Biomarker of Response to Neoadjuvant Therapy in ER+/HER2- Breast Cancer. *Cells*. 2020;9(7). <https://doi.org/10.3390/cells9071643>.
48. Wu CC, Ekanem TI, Phan NN, Loan DTT, Hou SY, Lee KH, et al. Gene signatures and prognostic analyses of the Tob/BTG pituitary tumor-transforming gene (PTTG) family in clinical breast cancer patients. *Int J Med Sci*. 2020;17(18):3112–24. <https://doi.org/10.7150/ijms.49652>.
49. Deng P, Tan M, Zhou W, Chen C, Xi Y, Gao P, et al. Bisphenol A promotes breast cancer cell proliferation by driving miR-381-3p-PTTG1-dependent cell cycle progression. *Chemosphere*. 2021;268:129221. <https://doi.org/10.1016/j.chemosphere.2020.129221>.

Figures

Figure 1

Overview of this study.

Figure 2

a Data before removing known batch effects. b. Data after removing known batch effects. c. GO enrichment analysis of genes related to DRFS in GSE25055. d. GO enrichment analysis of genes related to DRFS in GSE25065.

Figure 3

a The Venn diagram including three genes which are related to the DRFS of patients in the target genes of E2F family. b. Heatmap of MTHFD2, CDC25B, PTTG1 expression profiles in GSE25055. c. Heatmap of MTHFD2, CDC25B, PTTG1 expression profiles in GSE25065.

Figure 4

a Kaplan-Meier curve analysis of CDC25B, MTHFD2, PTTG1 high expression and low expression patients in GSE32603. b. Multivariate Cox regression analysis of CDC25B, MTHFD2, PTTG1 breast cancer patients in GSE32603.

Figure 5

a PPI network centered on PTTG1. The size of circle represents the degree of differential expression, and the thickness of line represents interaction score between two proteins. b. The low PTTG1 expression groups and high PTTG1 expression displayed in E2F target HALLMARK. c. The AUC of ROC curve was 0.682 showing the relatively poor predictive accuracy of survival.

Figure 6

a Nomogram for predicting the 1-, 3- and 5-year distant relapse-free survival of patients. b. Kaplan-Meier survival analysis of high risk score and low risk score patients. c. Calibration curves for predicting 1-year distant relapse-free survival. d. Calibration curves for predicting 3-year distant relapse-free survival. e. Calibration curves for predicting 5-year distant relapse-free survival. f. ROC curves representing the discriminatory ability of the nomogram.

Supplementary Files

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

- [SupplementaryFig.1.pdf](#)
- [SupplementaryFig.2.pdf](#)