# The target E2F family gene PTTG1 for prediction of distant relapse-free survival in breast cancer patients receiving taxane and anthracycline-based NACT

**Yuhao Xu**
Zhejiang Chinese Medical University
**Yaoqiang Du**
Zhejiang Provincial People's Hospital
**Qinghui Zheng**
Zhejiang Provincial People's Hospital
**Hongchao Tang**
Zhejiang Provincial People's Hospital
**Yangyang Qian**
Zhejiang Provincial People's Hospital
**Qiuran Xu**
Zhejiang Provincial People's Hospital
**Xuli Meng** ( mxlmail@126.com )
Zhejiang Provincial People's Hospital

**Research**

**Keywords:** breast cancer, neoadjuvant chemotherapy, distant relapse-free survival, PTTG1, nomogram model

**Posted Date:** May 25th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-551089/v1

## Abstract

# Background

Recent years, the breast cancer became the most commonly diagnosed cancer. 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 targets 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

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.

## Background

Recently, an estimated 2.3 million new cases help female 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, apotosis [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 including, 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 benefited from NACT and show distinct responses to NACT even when their clinicopathological characteristics are same or 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.

## Methods

# Data collection

Datasets (GSE25055,310 breast cancecr (BC) samples;GSE25065198 BC samples) were used by us for acquiring patients with breast cancer receiving taxane and anthracycline-based neoadjuvant chemotherapy from GEO database as discovery cohorts, while GSE32603 (248 BC samples) as a validation cohort. Gene expression of GSE25055 and GSE25065 was collected using platform GPL96 ([HG-U133A] Affymetrix Human Genome U133A Array), and GSE32603 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 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 GSE25055 and GSE25065. 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 related to the functional annotations of these DRFS genes with the help of clusterProfiler package, and hallmark E2F target genes was sorted out. Next, we got intersection genes of genes associated with DRFS and genes in hallmark E2F targets gene. Then, 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 related to DRFS and associated with E2F target hallmark using GSEA analysis and chosen as our target gene for further study.

# Establishment of PPI network and GSEA analysis

We downloaded a data file 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. Them we chose genes involved in GSE25055, GSE25065 and Hallmark E2F_Targets for PPI construction, whose products have

protein links with *PTTG1*. Cytoscape was used to visualize the PPI network centered on *PTTG1*. The permutaion test for this PPI network included 100000 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 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 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

# In search of 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” package in R 4.0.2 **(**Fig. 1a,b**)**. First of all, we performed univariate Cox regression analysis in GSE25055 for acquiring 2753 genes associated with DRFS, with the standard of P□0.05, and in GSE25065 was 2002. Then, 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□0.05 for acquiring 889 genes as the predictors of DRFS of patients independently, and in GSE25065 was 642. And 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. 1c,d**)** with P value < 0.05 was statistical difference **(Table. S1,2)**. From the cellular component (CC), biological process (BP), and molecular function (MF),we found the genes were mainly concerned with cell division, cell-cycle progression, controlling cell proliferation, apotosis and DNA repair. Based on the current study, we found that these functional annotations were closely related to genes of E2F family.

# Screening out the target gene in GSE32603

After the GO functional analysis, E2F family were chosen for further study. 200 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 gene set so that we obtained 3 intersection genes *CDC25B* ,*MTHFD2* 、*PTTG1*(Fig. 2a). Then we conducted heatmaps of the three genes expression profiles in GSE25055 and GSE25065 individually (Fig. 2b,c).

Next, we chose GSE32603 as a validation cohort for screening the target gene. First of all, Kaplan-Meier curve analysis of *CDC25B, MTHFD2, PTTG1* was 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 using log-rank tests with $P = 0.012$, $P = 0.008$, $P$<0.001 (Fig. 3a). Then we conducted multivariate Cox regression analysis in GSE32603. The result shows 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. 3b).

The results above proved that *PTTG1* was a gene related to DRFS,and *PTTG1* high expression 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

Then we chose the proteins interacted with *PTTG1* in the three subsets, and constructed a PPI network which is centered on *PTTG1* (Fig. 4a). We used a permutation test to identify the reliability of the network, the *P* value of all the nodes and edges were less than 0.001. As is shown in the PPI network, 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 could be seen in **Fig. S1,2.** Then 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 signifificantly enriched for E2F family target genes relative to patients from *PTTG1* low expression patients with *P* value = 0.016, normalized enrichment score (NES) > 1.5 and false discovery rate (FDR) = 0.007 (Fig. 4b) **(Table. S3)**, indicating high *PTTG1* was associated with E2F pathways in BC.

Using ROC curve to evaluate the accuracy and specificity of *PTTG1* for predicting the DRFS of breast patients, the result showed the AUC of *PTTG1* was 0.682 indicating a relative poor prediction performance (Fig. 4c).

# 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. 5a).

Each patient owned his total score by adding a corresponding score for each prognostic factor in the nomogram. Patients were separated into a high-risk group and a low-risk group based on riskscores. Patients with a high-risk score exhibited significantly worse DRFS compared to the ones with a low-risk score (P□0.001), as was shown in the Kaplan-Meier (K-M) survival curve (Fig. 5b).

The calibration plots for 1-year, 3-year, 5-year distant relapse-free survival of patients indicated the nomogram-predicted outcome showed high concordance with the actual outcome (Fig. 5c,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, AUC of pathologic_response was 0.409, AUC of grade was 0.593, AUC of ajcc_stage was 0.696, AUC of N stage was 0.698, AUC of T satge was 0.701, AUC of the status of PR was 0.332, AUC of ER was 0.306, AUC of age was 0.490 (Fig. 5f).

# Discussion

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

Due to the cost and information overflow, the clinical applicability was restricted in the old days. However, 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 minging TCGA, GEO, GTEx, CCLE and other databases [28–32]. High-throughput genomic studies provided cutting-edge sights into the molecular mechanisms and identifed new potential targets of breast cancer [33–36]. Nevertheless, using a single biomarker for predicting often relatively lower predictive effects, so 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 chromatidscontinuous in the course of mitosis was chosen to be our target gene 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 sutdies 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 size of breast patients involved in our study, we aimed to gather clinicopathological characteristics of breast cancer patients for large-scale cohort studies in the future, and a series of our validation database will be verified the accuracy of this nomogram model.

# Conclusions

Our study provided a nomogram model based on *PTTG1*, a target gene of E2F family, aiming to predic 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 riskscore of specific patient, 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 and HCT contributed to the data collection. YHX and YQD performed the nomogram model analyses and wrote the manuscript. YYQ helped to perform the enrichment and network analysis. 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 in GSEA in the Hallmark E2F target.

### Ethics approval and consent to participate

This study were not involved in human participant and clinical information. All the datasets were from the GEO database.

### 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. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249.
2. A. Isik and D. Firat. Bilateral intra-areolar polythelia. Breast J. 2018;24(1):89-90.
3. N. Harbeck and M. Gnant. Breast cancer. Lancet. 2017;389(10074):1134-1150.
4. A. Kuijer, M. Straver, B. den Dekker, A. C. M. van Bommel, S. G. Elias, C. H. Smorenburg, 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. J Clin Oncol. 2017;35(24):2814-2819.
5. L. N. Chaudhary. Early stage triple negative breast cancer: Management and future directions. Semin Oncol. 2020;47(4):201-208.
6. J. Kanska, M. Zakhour, B. Taylor-Harding, B. Y. Karlan and W. R. Wiedemeyer. Cyclin E as a potential therapeutic target in high grade serous ovarian cancer. Gynecol Oncol. 2016;143(1):152-158.
7. T. Shibue and R. A. Weinberg. Integrin beta1-focal adhesion kinase signaling directs the proliferation of metastatic cancer cells disseminated in the lungs. Proc Natl Acad Sci U S A. 2009;106(25):10290-5.
8. S. Manickavinayaham, B. K. Dennehey and D. G. Johnson. Direct Regulation of DNA Repair by E2F and RB in Mammals and Plants: Core Function or Convergent Evolution? Cancers (Basel). 2021;13(5).
9. S. Y. Park, Y. R. Seo, M. J. Ko, J. H. Lee, K. S. Chun, M. J. Kim, et al. Targeting CALM2 Inhibits Hepatocellular Carcinoma Growth and Metastasis by Suppressing E2F5-mediated Cell Cycle Progression. Anticancer Res. 2021;41(3):1315-1325.
10. L. Perrotta, R. Giordo, D. Francis, H. J. Rogers and D. Albani. 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.
11. Y. Yan, H. Xu, J. Wang, X. Wu, W. Wen, Y. Liang, et al. Inhibition of breast cancer cells by targeting E2F-1 gene and expressing IL15 oncolytic adenovirus. Biosci Rep. 2019;39(7).
12. X. Zhou, S. Zhong, H. Peng, J. Liu, W. Ding, L. Sun, et al. Cellular and molecular properties of neural progenitors in the developing mammalian hypothalamus. Nat Commun. 2020;11(1):4063.
13. Z. G. Hu, C. W. Zheng, H. Z. Su, Y. L. Zeng, C. J. Lin, Z. Y. Guo, et al. MicroRNA-329-mediated PTTG1 downregulation inactivates the MAPK signaling pathway to suppress cell proliferation and tumor growth in cholangiocarcinoma. J Cell Biochem. 2019;120(6):9964-9978.
14. I. Grzechowiak, J. Gras, D. Szymanska, M. Biernacka, K. Guglas, P. Poter, 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).
15. M. A. Romero Arenas, T. G. Whitsett, A. Aronova, S. A. Henderson, J. LoBello, M. A. Habra, et al. Protein Expression of PTTG1 as a Diagnostic Biomarker in Adrenocortical Carcinoma. Ann Surg Oncol. 2018;25(3):801-807.

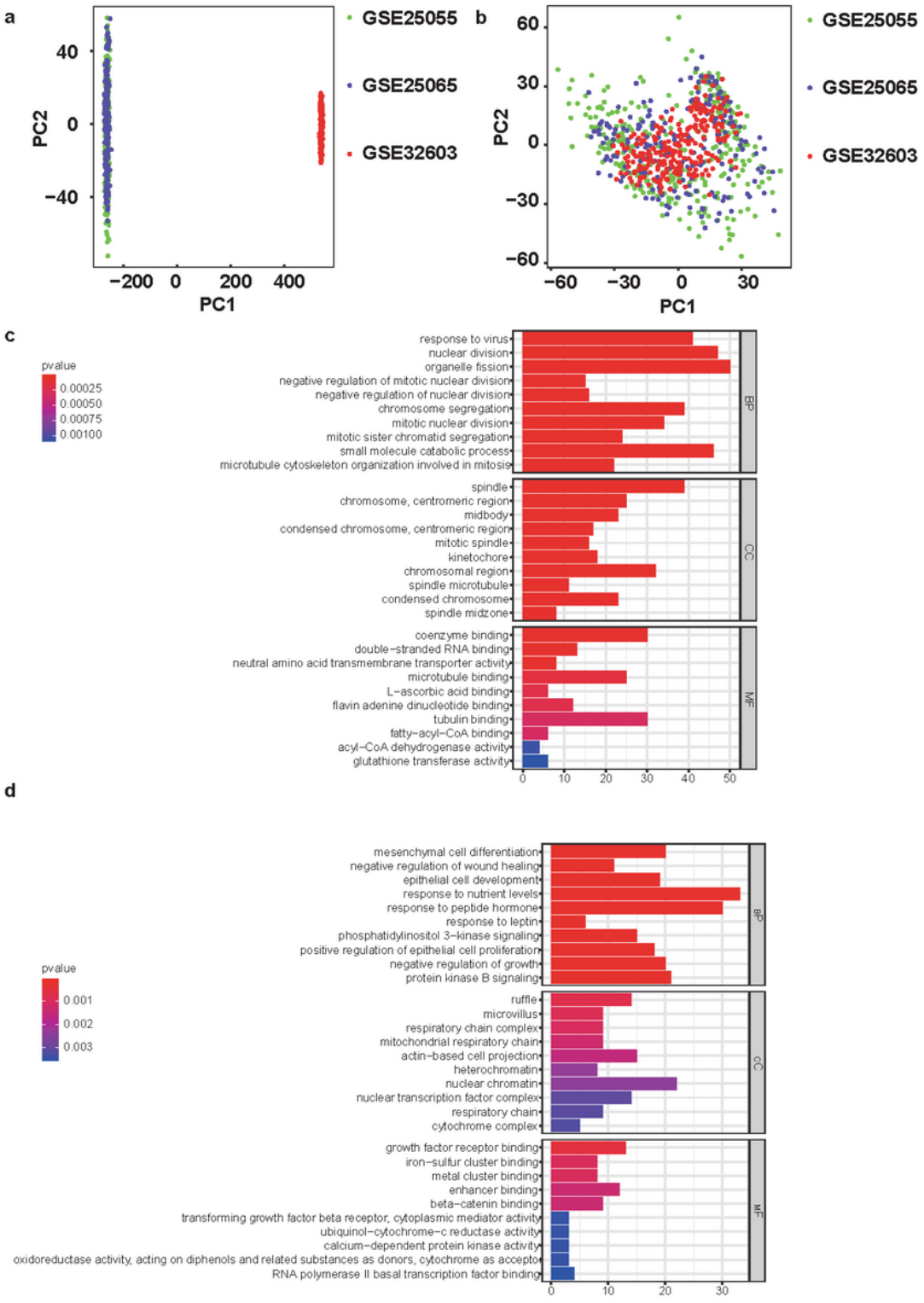
16. L. Cui, T. Ren, H. Zhao, S. Chen, M. Zheng, X. Gao, et al. Suppression of PTTG1 inhibits cell angiogenesis, migration and invasion in glioma cells. Med Oncol. 2020;37(8):73.
17. T. Zhi, K. Jiang, X. Xu, T. Yu, F. Zhou, Y. Wang, et al. ECT2/PSMD14/PTTG1 axis promotes the proliferation of glioma through stabilizing E2F1. Neuro Oncol. 2019;21(4):462-473.
18. C. Zhou, K. Wawrowsky, S. Bannykh, S. Gutman and S. Melmed. E2F1 induces pituitary tumor transforming gene (PTTG1) expression in human pituitary tumors. Mol Endocrinol. 2009;23(12):2000-12.
19. C. Meng, Y. Zou, W. Hong, C. Bao and X. Jia. Estrogen-regulated PTTG1 promotes breast cancer progression by regulating cyclin kinase expression. Mol Med. 2020;26(1):33.
20. R. Yu, A. P. Heaney, W. Lu, J. Chen and S. Melmed. Pituitary tumor transforming gene causes aneuploidy and p53-dependent and p53-independent apoptosis. J Biol Chem. 2000;275(47):36502-5.
21. Y. Zhou, K. R. Mehta, A. P. Choi, S. Scolavino and X. Zhang. DNA damage-induced inhibition of securin expression is mediated by p53. J Biol Chem. 2003;278(1):462-70.
22. C. Hatzis, L. Pusztai, V. Valero, D. J. Booser, L. Esserman, A. Lluch, et al. A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. JAMA. 2011;305(18):1873-81.
23. P. Cheng, Z. Wang, G. Hu, Q. Huang, M. Han and J. Huang. A prognostic 4-gene expression signature for patients with HER2-negative breast cancer receiving taxane and anthracycline-based chemotherapy. Oncotarget. 2017;8(61):103327-103339.
24. P. Cortazar, L. Zhang, M. Untch, K. Mehta, J. P. Costantino, N. Wolmark, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet. 2014;384(9938):164-72.
25. Y. Li, D. M. Umbach, J. M. Krahn, I. Shats, X. Li and L. Li. Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines. BMC Genomics. 2021;22(1):272.
26. W. Tan, M. Liu, L. Wang, Y. Guo, C. Wei, S. Zhang, et al. Novel immune-related genes in the tumor microenvironment with prognostic value in breast cancer. BMC Cancer. 2021;21(1):126.
27. M. Sun, X. Liu, L. Xia, Y. Chen, L. Kuang, X. Gu, 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.
28. B. Chao, X. Ju, L. Zhang, X. Xu and Y. Zhao. A Novel Prognostic Marker Systemic Inflammation Response Index (SIRI) for Operable Cervical Cancer Patients. Front Oncol. 2020;10:766.
29. C. Chen, Y. Pan, L. Bai, H. Chen, Z. Duan, Q. Si, 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.
30. Y. Wang, M. Zhu, F. Guo, Y. Song, X. Fan and G. Qin. Identification of Tumor Microenvironment-Related Prognostic Biomarkers in Luminal Breast Cancer. Front Genet. 2020;11:555865.

31. M. Wu, Y. Zhao, N. Peng, Z. Tao and B. Chen. Identification of chemoresistance-associated microRNAs and hub genes in breast cancer using bioinformatics analysis. Invest New Drugs. 2021;39(3):705-712.
32. L. Xu, J. M. Shen, J. L. Qu, N. Song, X. F. Che, K. Z. Hou, et al. FEN1 is a prognostic biomarker for ER+ breast cancer and associated with tamoxifen resistance through the ERalpha/cyclin D1/Rb axis. Ann Transl Med. 2021;9(3):258.
33. L. Chen, Y. Dong, Y. Pan, Y. Zhang, P. Liu, J. Wang, et al. Identification and development of an independent immune-related genes prognostic model for breast cancer. BMC Cancer. 2021;21(1):329.
34. J. Xian, S. Wang, Y. Jiang, L. Li, L. Cai, P. Chen, et al. Overexpressed NEDD8 as a potential therapeutic target in esophageal squamous cell carcinoma. Cancer Biol Med. 2021.
35. Q. Xie, Y. S. Xiao, S. C. Jia, J. X. Zheng, Z. C. Du, Y. C. Chen, et al. FABP7 is a potential biomarker to predict response to neoadjuvant chemotherapy for breast cancer. Cancer Cell Int. 2020;20(1):562.
36. L. Zhang, J. Pan, Z. Wang, C. Yang and J. Huang. Construction of a MicroRNA-Based Nomogram for Prediction of Lung Metastasis in Breast Cancer Patients. Front Genet. 2020;11:580138.
37. C. Gong, W. Tan, K. Chen, N. You, S. Zhu, G. Liang, et al. Prognostic Value of a BCSC-associated MicroRNA Signature in Hormone Receptor-Positive HER2-Negative Breast Cancer. EBioMedicine. 2016;11:199-209.
38. Y. Guo, X. Mao, Z. Qiao, B. Chen and F. Jin. A Novel Promoter CpG-Based Signature for Long-Term Survival Prediction of Breast Cancer Patients. Front Oncol. 2020;10:579692.
39. C. Liu, S. Wang, S. Zheng, X. Wang, J. Huang, Y. Lei, 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.
40. Y. Liu, Z. Huang, G. Cheng, Y. Shou, J. Xu, D. Liu, et al. Development of a four-gene prognostic model for clear cell renal cell carcinoma based on transcriptome analysis. Genomics. 2021;113(4):1816-1827.
41. J. Shao, W. Lyu, J. Zhou, W. Xu, D. Wang, S. Liang, et al. A Panel of Five-lncRNA Signature as a Potential Biomarker for Predicting Survival in Gastric and Thoracic Cancers. Front Genet. 2021;12:666155.
42. X. Zhang, J. Wang, J. Zhuang, C. Liu, C. Gao, H. Li, et al. A Novel Glycolysis-Related Four-mRNA Signature for Predicting the Survival of Patients With Breast Cancer. Front Genet. 2021;12:606937.
43. Y. Zhao, C. Pu and Z. Liu. Exploration the Significance of a Novel Immune-Related Gene Signature in Prognosis and Immune Microenvironment of Breast Cancer. Front Oncol. 2020;10:1211.
44. D. Hanahan and R. A. Weinberg. The hallmarks of cancer. Cell. 2000;100(1):57-70.
45. Q. Huang, T. Qu, L. Qi, C. Liu, Y. Guo, Q. Guo, et al. A nomogram-based immune-serum scoring system predicts overall survival in patients with lung adenocarcinoma. Cancer Biol Med. 2021.
46. Q. Tu, C. Hu, H. Zhang, C. Peng, M. Kong, M. Song, et al. Establishment and Validation of Novel Clinical Prognosis Nomograms for Luminal A Breast Cancer Patients with Bone Metastasis. Biomed

Res Int. 2020;2020:1972064.

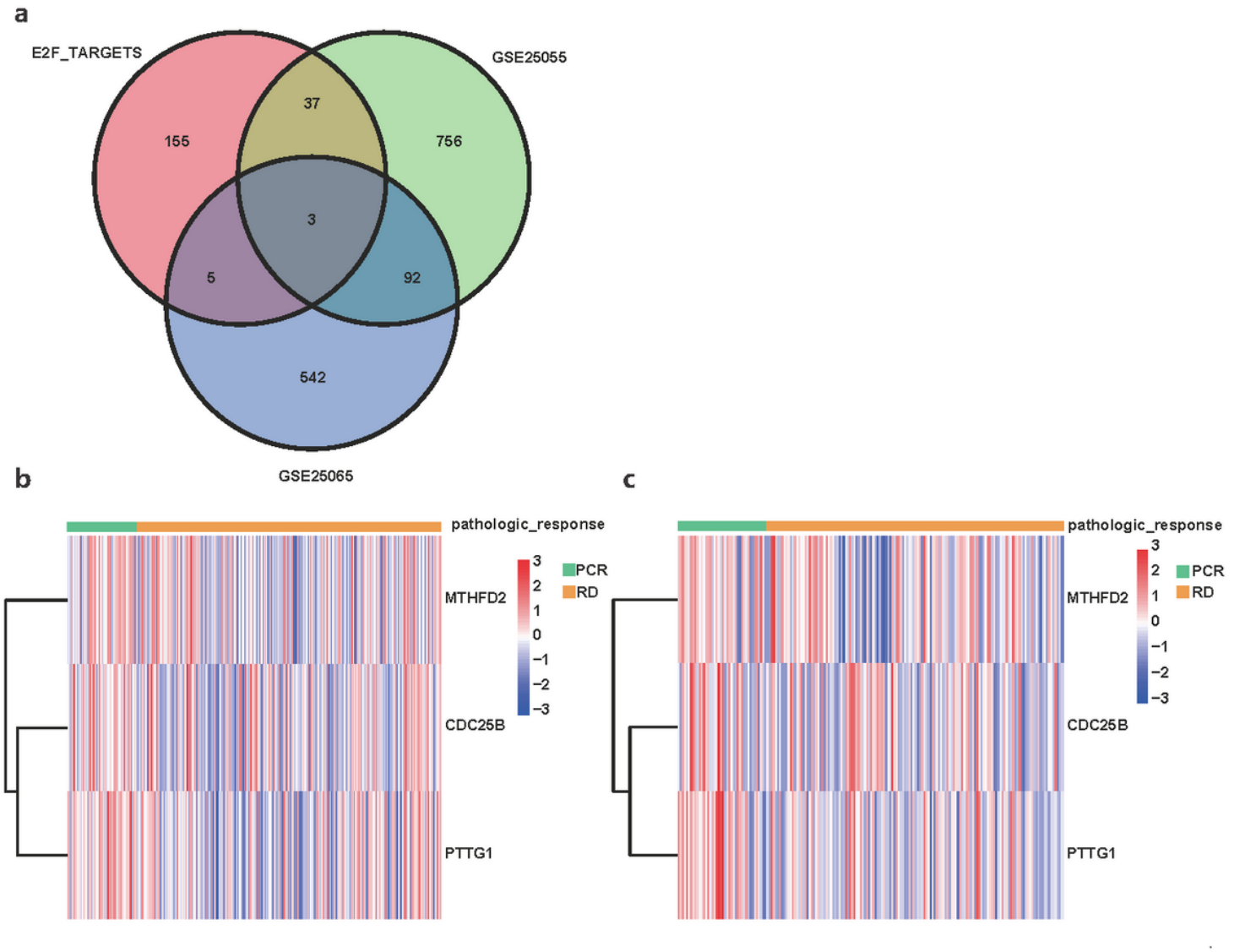
47. M. Oshi, H. Takahashi, Y. Tokumaru, L. Yan, O. M. Rashid, M. Nagahashi, et al. The E2F Pathway Score as a Predictive Biomarker of Response to Neoadjuvant Therapy in ER+/HER2- Breast Cancer. Cells. 2020;9(7).
48. C. C. Wu, T. I. Ekanem, N. N. Phan, D. T. T. Loan, S. Y. Hou, K. H. Lee, et al. Gene 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):3112-3124.
49. P. Deng, M. Tan, W. Zhou, C. Chen, Y. Xi, P. Gao, et al. Bisphenol A promotes breast cancer cell proliferation by driving miR-381-3p-PTTG1-dependent cell cycle progression. Chemosphere. 2021;268:129221.

# Figures



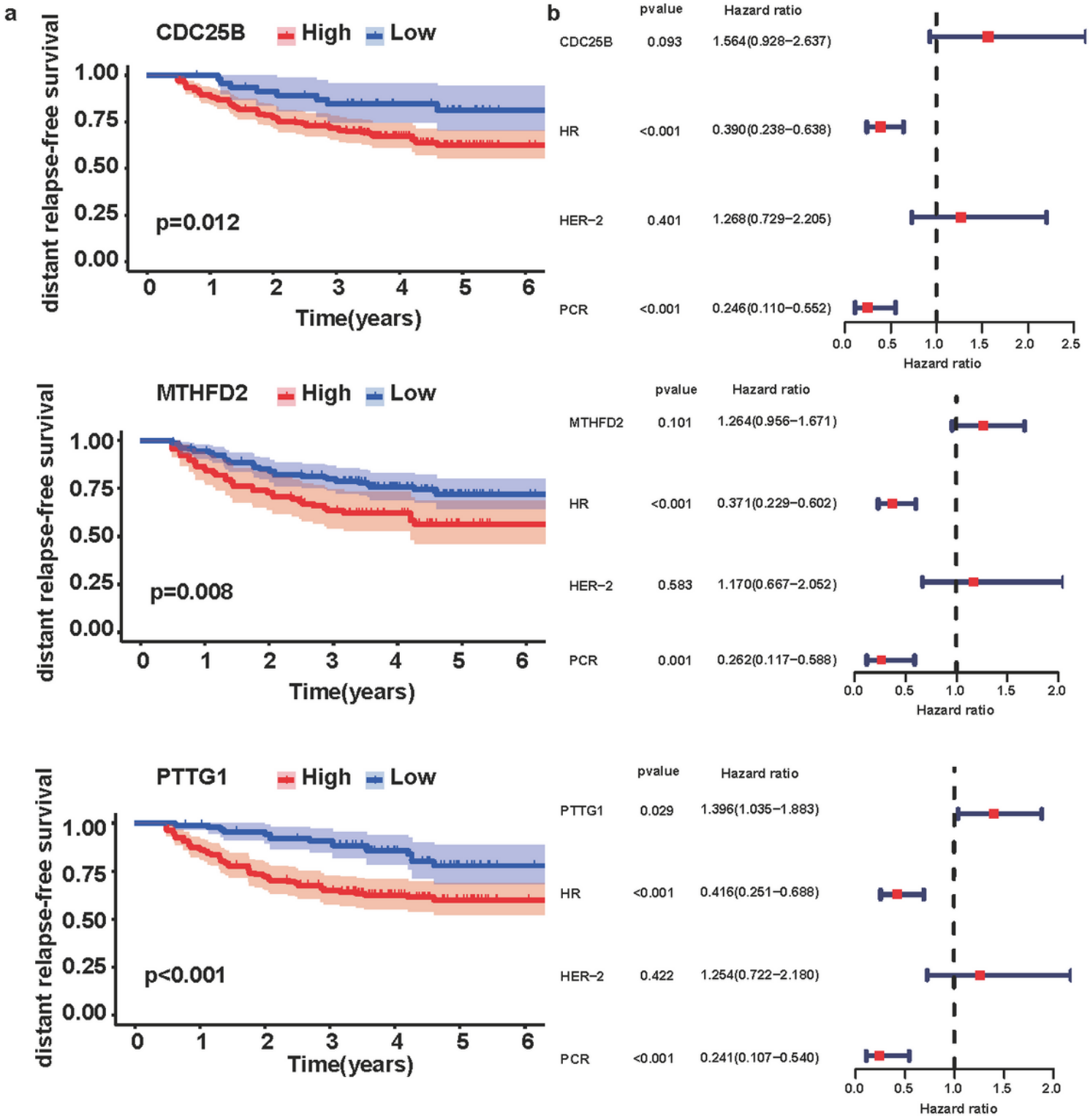

## Figure 1

a. data before removing known batch effects; b. data after removing known batch effects. c. Gene Ontology (GO) enrichment analysis of genes related to DRFS in GSE25055 d. Gene Ontology (GO) enrichment analysis of genes related to DRFS in GSE25065.



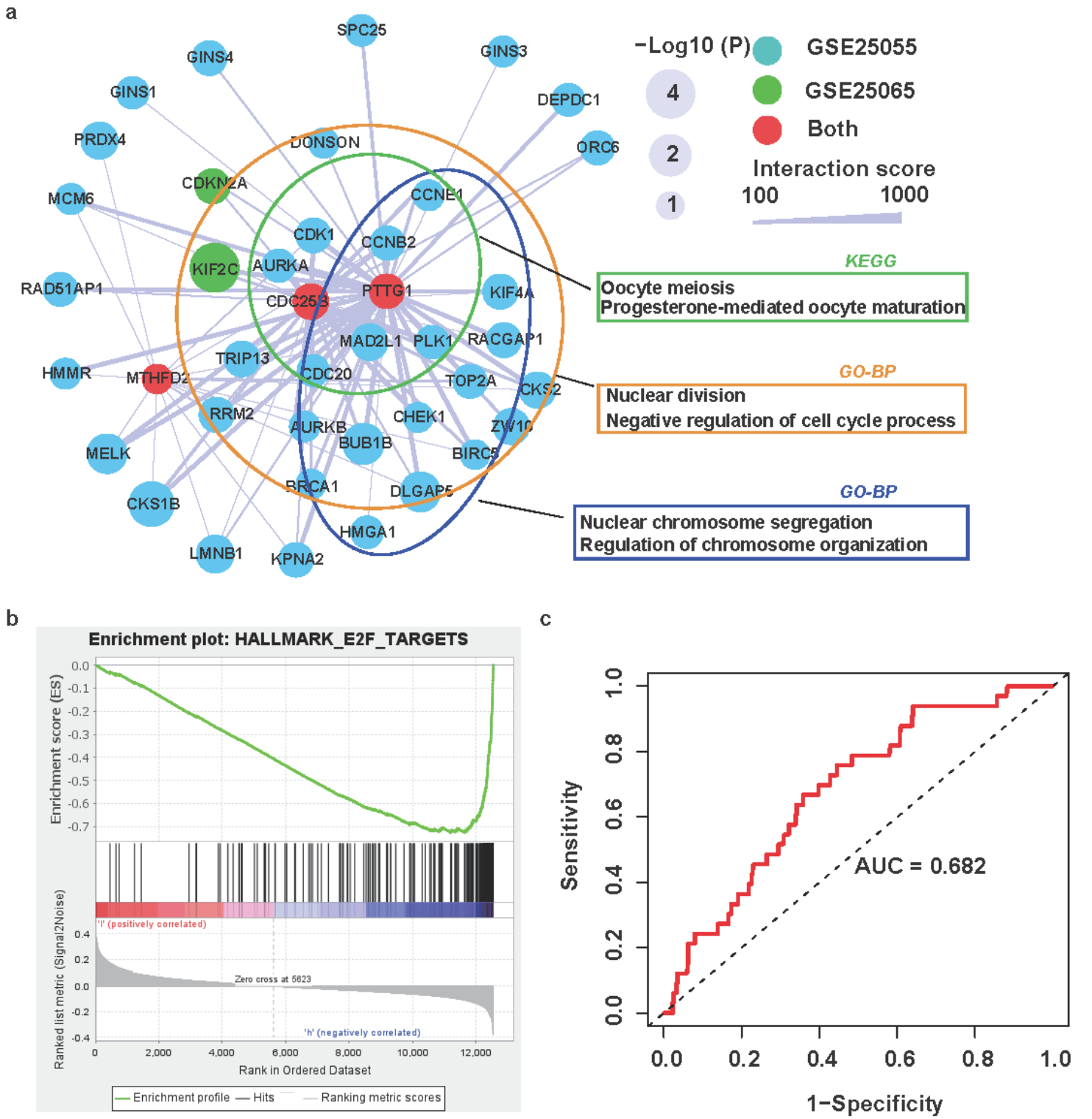

## Figure 2

a. The venn diagrams 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.



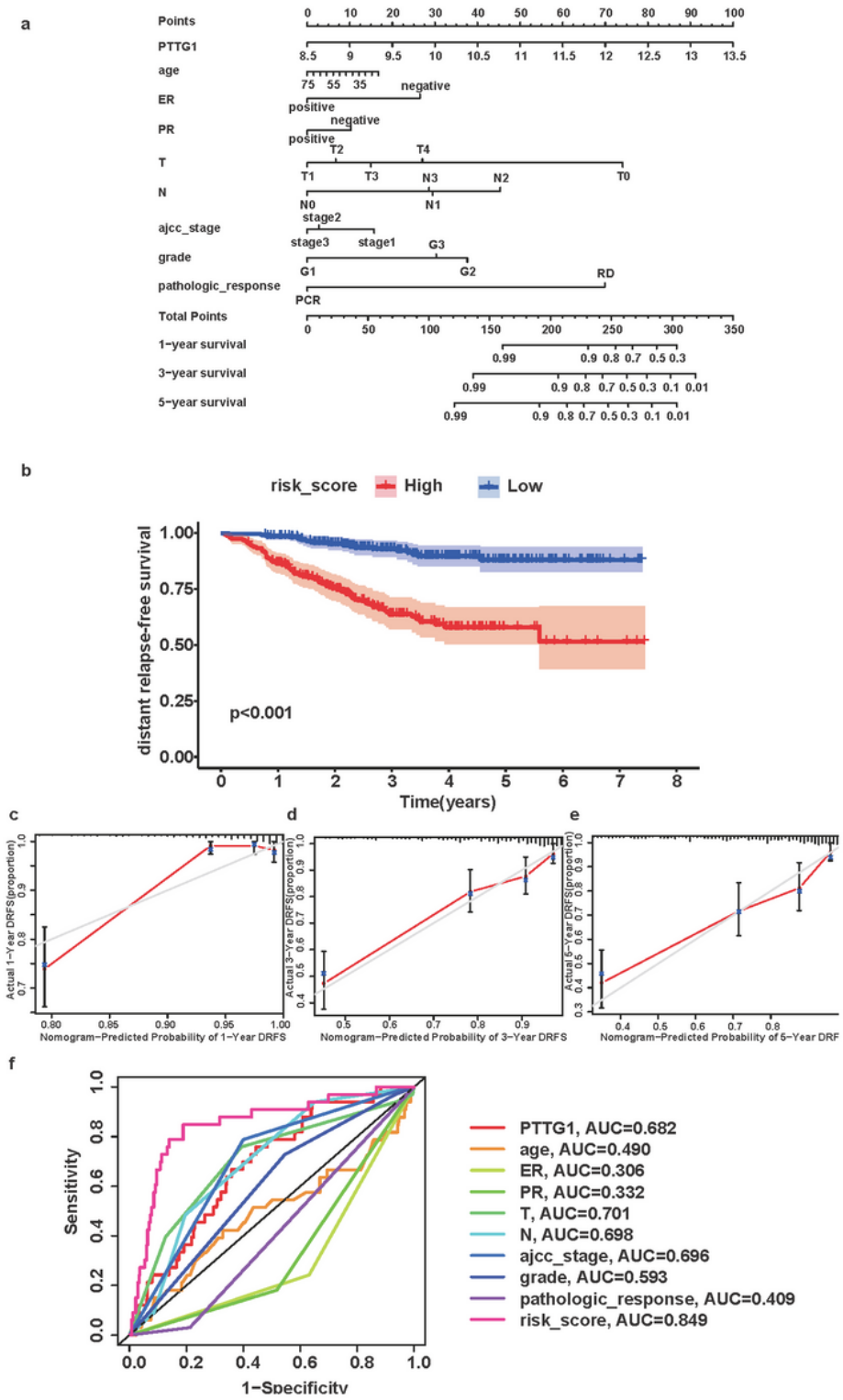

## Figure 3

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 4

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 groupsrisk 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 5

a. Nomogram for predicting the 1-,3- and 5-year distant relapse-free survival of patients; b. Kaplan-Meier survival analysis of high riskscore and low riskscore 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.

- Fig.S1.pdf
- Fig.S2.pdf
- Tab.S1.xls
- Tab.S2.xls
- Tab.S3.xls